Laboratory Phonology 8

Phonology and Phonetics

4-2

Editor

Aditi Lahiri

Mouton de Gruyter
Berlin · New York

Laboratory Phonology 8

edited by

Louis Goldstein
D. H. Whalen
Catherine T. Best

Mouton de Gruyter
Berlin · New York

Mouton de Gruyter (formerly Mouton, The Hague)
is a Division of Walter de Gruyter GmbH & Co. KG, Berlin.

♾ Printed on acid-free paper which falls within the guidelines
of the ANSI to ensure permanence and durability.

Library of Congress Cataloging-in-Publication Data

Conference in Laboratory Phonology (8th : 2002 ; New Haven, Conn.)
Laboratory phonology 8 / edited by Louis Goldstein, D. H. Whalen, Catherine T. Best.
p. cm. – (Phonology and phonetics ; 4/2)
"Papers from Eighth Conference on Laboratory Phonology, held in New Haven, CT"– Chap. 1.
Includes bibliographical references and index.
ISBN-13: 978-3-11-017678-0 (cloth : alk. paper)
ISBN-10: 3-11-017678-5 (cloth : alk. paper)
1. Grammar, Comparative and general – Phonology – Congresses. 2. Sign language – Congresses. 3. Language acquisition – Congresses. I. Goldstein, Louis, 1955– II. Whalen, D. H. III. Best, Catherine T. IV. Title. V. Title: Laboratory phonology eight. VI. Series.
P217.C658 2002
414–dc22
2006015539

Bibliographic information published by Die Deutsche Bibliothek

Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data is available in the Internet at <http://dnb.ddb.de>.

ISBN-13: 978-3-11-017678-0
ISBN-10: 3-11-017678-5
ISSN 1861-4191

Cover design: Christopher Schneider, Berlin.
Printed in Germany.

Table of contents

Introduction

The first Conference on Laboratory Phonology was held in 1987 at the Ohio State University. It brought together theoretical phonologists and phoneticians whose respective areas of research were, at that moment, fairly well demarcated. Phonologists studied the kinds of formal representations (and operations) that were necessary to account for speakers' knowledge of the patterning of segments and higher-level phonological units in their language: their lexical regularities and their alternations when lexical units combine into larger structures. The data were primarily transcriptions of utterances and native speaker well-formedness judgments. Phoneticians largely studied how segments are produced in context, their resulting acoustic properties, and how those acoustic properties could be used by listeners to perceive the segmental sequences. Their data were articulatory and acoustic measures of speech production and perceptual judgments. The organizers had the foresight to have phoneticians and phonologists provide commentary on each others' papers, in order to encourage communication across the divide.

But the meeting took place in a climate of growing awareness that this way of dividing up the linguistic study of the human speech system might not be the best one (and indeed this was one of the motivations for the event). Laboratory work had begun to demonstrate that some aspects of language-specific phonological knowledge could be most systematically expressed in terms of the controls for the speech production process (e.g. sparsely distributed tone or articulatory targets, overlapping organization of articulatory gestures), rather than in the linear sequencing of segments and features. This seemed intriguingly compatible with the contemporaneous development of nonlinear phonological representations. In this context, the data relevant for probing of a speaker's phonological knowledge began to be generalized beyond phonetic transcriptions and judgments of their grammaticality to more fine-grained, quantitative descriptions of speech. Thus was born *Laboratory* Phonology.

The Eighth conference on Laboratory Phonology (held in New Haven, CT, and hosted by Yale University and Haskins Laboratories) took place June 27–29, 2002, fifteen years after the first. In the intervening time, the use of laboratory data in phonology has expanded considerably, though much of theoretical phonology is still based on (and only accounts for) segmental sequences. In planning the eighth meeting, we felt that it was an opportune moment to encourage further expansion in the range of data considered in developing a theory of human phonological capacity by making signed languages, and specifically the comparison of signed and spoken languages, a focus of the meeting. There were several reasons for this choice. First, while there has been substantial work on the phonology of signed languages, not much of it had been done within the perspective of Laboratory Phonology or using its toolkit and so is unfamiliar to many in the Laboratory Phonology community. Second, because of the increased emphasis on universals in phonological theory over the previous ten years (e.g., within Optimality Theory), it seemed important to consider to what extent the same principles (or even some of the same constraints) appear to govern signed and spoken languages. Third, for those interested in gestural theories of speech and phonology (not coincidentally including the organizers), it seemed important to make systematic comparisons between speech (whose gestural basis is contentious) and sign (whose gestural basis may seem uncontroversial to *speech* researchers, but is in fact contentious within the sign research community). Finally, opening the study of phonological competence to laboratory data has necessarily meant coming to terms with variable, probabilistic, or gradient data. How to understand (and/or model) the relation between continuous variation and qualitative phonological categories has been a major focus of Laboratory Phonology since its inception (and, of course, there is no consensus on how this should be done). It seemed as though much could be learned by examining this issue in the context of a (superficially at least) very different kind of phonology – sign. The inclusion of sign phonology proved to be successful – not only were the papers significant and as relevant to the theoretical issues as we hoped they would be, but they stimulated valuable discussion between sign and speech phonologists.

This collection of papers from LabPhon 8 is organized into three sections. Section I includes papers that address the fundamental issue of integrating discrete and continuous descriptions. Alternative approaches to this issue represented here include: modular separation of categorical and continuous subsystems, dynamical grammars, and probabilistic grammars. As the papers show, this is an issue that arises in sign phonology, as well as in speech. Juxtaposition of papers on the two adds valuable perspective to each. Also included in this section are papers comparing key design properties of sign and speech phonologies and the processes of word segmentation and lexical access in the two modalities. The papers show large degrees of compatibility between sign and speech, but also some modality-specific characteristics that appear to be universal across signed languages, but different from spoken languages.

Section II includes papers on the acquisition of phonological competence. Much of the work on acquisition of phonology has assumed (tacitly at least), that the (monolingual) child's phonological environment is characterized by a single phonological system that can be characterized appropriately in traditionally discrete terms. The set of papers in this section present laboratory studies that document the richness and variation that may occur in a learner's phonological environment and demonstrate the impact of such variation in the developing child phonology, as well its consequences for phonological theory. Types of variability examined include the degree of regularity of a language's rhythmic patterns and social indexical marking of various sorts.

The papers in Section III focus on how phonological knowledge guides speech production and speech perception. A major discovery of laboratory phonology was that abstract prosodic structures exert strong influences on segment- or gesture-sized units of speech production. Papers in this section document novel cases of this and consider their implications for various theoretical proposals. Other papers demonstrate language-particular regularities in how speech units (gestures) are coordinated locally (at the level of the syllable) and discuss how this knowledge can be best captured in formal models.

* * *

The conference itself was a pleasure, even for the organizers. The details were arranged by the marvelous conference center at Yale, and, in particular, by the efforts of Susan Adler and Roberta Hudson. They helped organize a venue that met every need for making a conference successful. Most of the housing was in a single facility, increasing communication in the off hours. The room for the main conference was superb. The meeting was held in the restored lecture hall of Lindsey-Chittenden College, which is a masterwork of revival. After years of neglect, the Old Campus of Yale had received some much needed refurbishing. Now, the busts of Western history's top scientific and literary scholars – Franklin, Newton, Kant, Plato, and Cicero, Homer, Virgil, Goethe, Shakespeare, Dante – look down upon august proceedings in this stunningly re-beautified lecture hall. Although none of these worthy gentlemen was cited in the current volume, lowering their Citation Impact Factor, we hope that they approved nonetheless.

Another amazing feature of this room is the Tiffany stained-glass window that covers the length of one wall. This approximately 60 foot (18.288 meter) display includes four main scenes, with the human endeavors of Art, Science, Religion and Music being represented mostly by angels. (Some of us may have considered Music a form of Art, but this view was apparently not shared by Tiffany.) The angels of Art are labeled Form, Color, and Imagination. Presumably, the last angel gave Tiffany the go-ahead to make the angel wings look like shimmering dragonfly wings. Between Art and Science are two stranded angels, Perception and Analysis. It is not clear whether they are the intermediaries or simply the orphans of Art and Science, but in any case, they are mostly hidden by support columns. Science has three angels, Devotion, Labor and Truth, but also two human instantiations, Research and Intuition. Religion similarly has three angels, Purity, Faith and Hope, and two earthly representatives, Reverence and Inspiration. One doesn't know quite why Religion gets the flash of inspiration while Science has to settle for the guesswork of intuition, but it obviously made sense to Louis Tiffany. Between Religion and Music, also obscured by columns, is Law, who is an angel, to our befuddlement, though it surprised us less that she had only herself for company. Music's angels are Rhythm, Melody, Harmony, Verse and Voice. Art and Music also have god-like figures representing the genre as a whole, and each is gazing wistfully toward the heavens. The men who represent Science and Religion, on the other hand (yes, all four are men, perhaps explaining the following observations), are focused on the earth (although Inspiration seems to be trying to distract Reverence from his prayers, and Intuition is pointing tellingly toward some poppies…). So, as is often the case, it is far from clear what was going through Tiffany's head, but the window is nonetheless magnificent. We found it especially fitting that Voice, the only angel with her mouth open, should come last in the tableau and thus have the last word, for speech (and yes, Louie, speech encompasses the Voice of Song) is the most deeply human accomplishment of all...

Acknowledgments

So many people to thank: First and foremost, Yale University, Haskins Laboratories, and the National Science Foundation (Project 0132005) for financial support; Stephen Anderson for his support as the chair of the Yale Linguistics Department and his efforts as one of the organizers; Carol Fowler, President of Haskins and Susan Hockfield, (then) Dean of the Graduate School at Yale for their opening remarks; Casandra Murphy at Yale and Lisa Fresa at Haskins for help with various aspects of the arrangements. Many students pitched in, but the contributions of three stand out: Larissa Chen, for producing and maintaining the web site, Alexandra Diaz-Almaral for her help on the ground, as we approached crunch time, and Tiffany Gooding for her terrific work editing and formatting the manuscripts.

Louis Goldstein
Douglas Whalen
Catherine Best

Dedication

In her welcoming remarks at the conference, Carol Fowler (President of Haskins Laboratories) dedicated the meeting to Catherine P. Browman. Cathe was a pioneer in Laboratory Phonology through her work on Articulatory Phonology and earlier was a major developer of speech synthesis-by-rule. She actively participated in the first four Laboratory Phonology meetings and the last public talk of her career (due to encroaching illness) was at the Oxford meeting. She encouraged us to hold a meeting in New Haven and made every effort to attend some of the events, though in the end her health prevented it. Her ability to think in new ways about old problems, her vision of what phonology could be, her dedication to the field of Laboratory Phonology (as a way of realizing that vision), and her devotion to Haskins Laboratories remain an inspiration to so many of us. We dedicate this volume of papers to her, in gratitude.

I. Qualitative and variable faces of phonological competence

Spoken languages

Convergences and divergences
of signed and spoken languages

Signed languages

“Distinctive phones” in surface representation*

D. Robert Ladd

An analogy to alphabetic handwriting suggests that it is a priori reasonable to view the relation between “phonology” and “phonetics” as involving a mapping from symbolic categories to continuous phonetic parameters. This in turn implies the existence of an interface representation consisting of categorically distinct symbolic elements. Recent evidence suggests that “systematic phonetic” transcriptions are best seen as an informal shorthand rather than a linguistically principled representation of speech; consequently, such transcriptions are not suitable for use as a formal “level of representation” at the phonology-phonetics interface. Instead, the most plausible interface representation – and still the most widely used, forty years after the emergence of generative phonology – appears to resemble a classical phonemic transcription. It is therefore appropriate to try to bring definitions of the classical phoneme into line with current knowledge of cognitive categories generally. Phonemes may be seen as language-specific phonetic categories, which need not invariably contrast with each other absolutely, but may exhibit relations of “partial similarity” or “quasi-contrast”. Such cases could be treated as involving distinct subcategories of a single higher level category.

1. Phonetic abstraction and the phonology-phonetics interface

Every utterance is literally unique, different in physical detail from every other. Though this statement is clearly true, it is generally held to be irrelevant to linguistics, because many of the differences that are involved in the literal uniqueness of utterances are linguistically irrelevant, even imperceptible. Nevertheless, all of linguistics is ultimately based on being able to abstract away from the uniqueness of utterances and to identify certain speech phenomena as being “the same”. This paper discusses the basis of that process of abstraction.

Specifically with regard to phonetics and phonology, the abstractions that many of us most commonly deal with are transcriptions. Transcriptions are *symbolic abstractions,* representations to which operations of *discrete mathematics*, such as substitution and permutation, can apply. Any description of phonological phenomena that involves grammar-like operations on strings of transcription symbols is couched in terms of discrete mathematics. By comparison, the “gestures” of Browman and Goldstein’s articulatory phonology (e.g. Browman and Goldstein 1986) are *quantitative* abstractions, which need to be dealt with in terms of *continuous mathematics*. Browman and Goldstein have been at pains to point out that, in a great many classic cases of assimilation and deletion, gestures are not actually deleted or substituted, only modified along continuous dimensions such as amplitude and duration.

Although *as a theory of phonology* Browman and Goldstein’s ideas remain the preserve of a minority, virtually everyone agrees that at some level, scientific description of physical activity such as speaking must be expressed in quantitative terms. This leaves us with a problem. If speaking must be described in terms of continuous mathematics, then either we must banish symbolic abstractions and operations of discrete mathematics from phonology altogether, or we must accept that in some way the study of speech sounds involves an *interface* or a mapping of some sort between symbolic abstractions (“phonology”) and continuous ones (“phonetics”). The first view is apparently that of Browman and Goldstein, and has been espoused in various recent work by e.g. Coleman and his colleagues (e.g. Coleman 1998), Pierrehumbert and her colleagues (e.g. Hay, Pierrehumbert and Beckman 2003), and others. The second view – the interface view – is conservative, in the sense that it dates back at least half a century to Joos (1948), and was central to the Sound Pattern of English (SPE) formalism (Chomsky and Halle 1968). Yet it is still current, and still seems defensible. For example, it is taken for granted in a paper aimed at mathematicians by András Kornai (1994). The idea of the interface is also at least part of the reason for Halle and Bromberger’s insistence (1989) that “phonology is different”.

In a separate paper (Ladd in preparation), I have argued the case for the conservative conception of the phonology-phonetics interface on the basis of an extended analogy to alphabetic handwriting. In written language we can generally distinguish clearly between analogues to phonology – spelling rules – and analogues to phonetics – physical realization as handwriting. Analogues to phonology include things like

– *allographs* (e.g. the positionally-determined difference between <s> and <ſ> in older forms of the Roman alphabet);

– *complex segments* (e.g. Dutch <ij> as free variant of <y>);
– *alternations* (e.g. *happy/happier* or *carry/carried* in English);
– *post-lexical processes* (e.g. sentence-initial capitalization; e.g. German respelling of <ck> as <k-k> when a word is divided into syllables in choral scores or at the end of a line of type[1]).

Analogues to phonetics include things like coarticulation (subtle modifications of the shape of letters depending on the location of the connection to an immediately adjacent letter) and the trade-off between effort and intelligibility (the difference between fast sloppy writing and slow careful writing). More fundamentally, alphabetic handwriting demonstrates the essential plausibility of the conservative interface view: a handwritten word as a physical signal exhibits a lack of segmentability remarkably like what we see in speech, yet we know that it is appropriately idealized as a string of letters – that is, as a string of discrete elements chosen from a paradigmatic set.

The written language analogy thus gives considerable comfort to traditional phonologists. It makes clear that certain phenomena (like the alternation seen in *happy* and *happier*) involve operations on *representations*, which can and should be described independently of the act of writing. This makes it plausible that the analogous phenomena of spoken language phonology should be treated in the same way. For example, an abstract symbolic representation not only seems useful but may actually be necessary for describing things like the relation between *perfect* and *perfection* or *memory* and *memorial.* Proposals to do away with the phonology-phonetics interface have tended not to deal with those problems.

On the other hand, the written language analogy should make traditional phonologists fairly *un*comfortable in other ways. In particular, it makes clear that there are plenty of phenomena that are often treated in terms of operations on representations but which should actually be described as part of the physical realization of an abstract representation – as Browman and Goldstein have long maintained. For example, I may write the word *handwriting* as in (1):

(1)

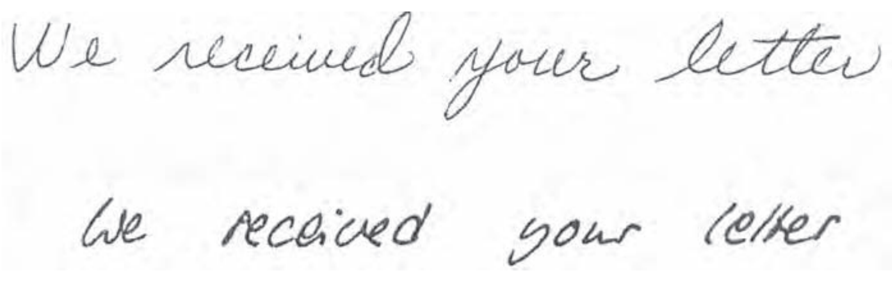
and in fact I commonly reduce *-ing* sequences to the dotted downward squiggle seen in (1), which is separated off in (2):

(2)

However, no one would say that the abstract representation of the word – the spelling – has changed. We would just say that the abstract sequence *-ing* is realized in a particular way. We would not describe (2) as a new letter, nor even think of it as a surface segment. And this fact gives less comfort to traditional non-laboratory phonologists, because it casts doubt on the legitimacy of symbolic descriptions for many phenomena of assimilation and reduction. It is clear that traditional descriptions of these phenomena – e.g. transcribing *ten past nine* as [ˌt^{h}ɛmpæsˈnain] – treat the phonetic detail in a way that is quite comparable to treating (2) as a surface segment. Note, incidentally, that there is nothing about the physical nature of (2) that prevents it from being regarded as a segment of a written signal: any "systematic graphetics" of the writing systems of the world will have to allow for segments very much like (2), e.g. the Arabic letters ض and خ.

Moreover, the written language analogy suggests that we must give a quantitative rather than a "grammatical" treatment to many features of spoken language that are clearly language-specific and clearly communicatively significant. Consider the difference between these two written versions of the same English sentence.

(3)

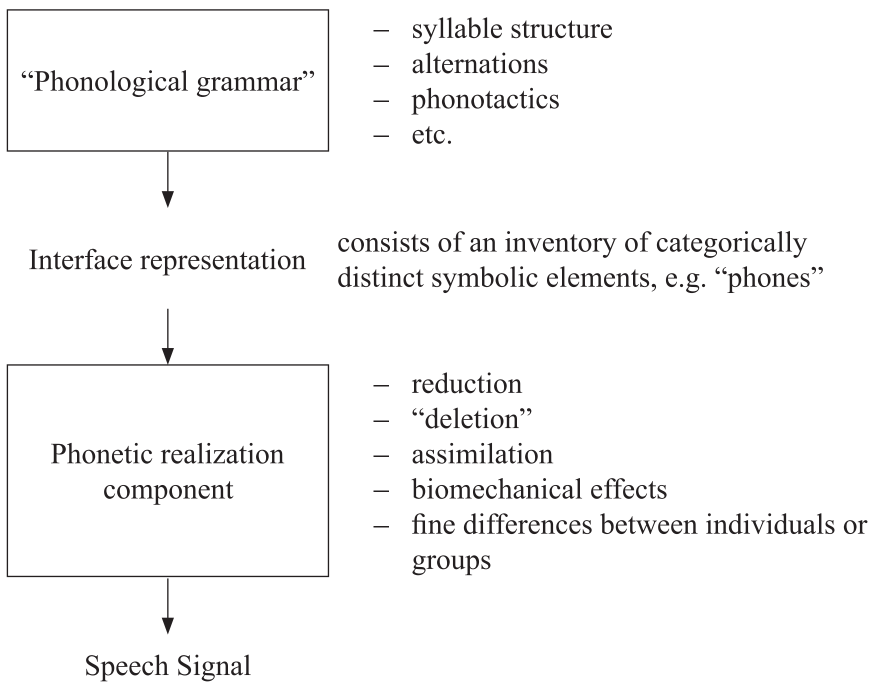


Anyone with any experience of different national styles of handwriting will have no difficulty telling which of these was written by an American and which by a German. But with the possible exception of the form of the lower-case <r>, the interface representation is exactly the same in both cases – these are both realizations of the same string of letters. The difference lies in fine details of the way the elements of that string are realized. This may seem uncontroversial when applied to handwriting, but the implications of this analogy for spoken language are startling. The differences between "typical American handwriting" and "typical German handwriting" are precisely analogous to many social and regional differences of pronunciation – for example, the fine phonetic distinctions of vowel quality that British English puts to such vicious social use. This means that such distinctions should not be described in symbolic terms: just as we would not say that Germans and Americans use different letters, so we should not say that upper class and middle class speakers of Southern British English use different segments. Instead, we will simply say that, as a matter of group-specific realization rules, different groups produce "the same" letter or "the same" segment type in slightly different ways.

In short, if we accept the allegedly conservative phonology-phonetics interface idea, it seems to lead to the rather less conservative conclusion that *there is no appropriate symbolic or grammatical characterization of many meaningful language-specific differences.* For a variety of reasons, I believe that this conclusion is correct. We need to describe phonetic realization consistently in quantitative terms, and, fully pursuing the implications of the written language analogy, we need to treat as "phonetic realization" many phenomena that have in the past been treated using grammar-like operations of deletion, substitution, and reduction. We need to accept – really accept – Browman and Goldstein's claim that there is no appropriate segmental representation of the output of the phonetic component: we cannot say that in one production of the phrase *perfect memory* the [t] is realized and in another it is deleted, and nor can we say that some sort of intermediate category is produced. The validity of this point emerges clearly from a variety of laboratory phonology work over the past 15 years (e.g. Nolan 1992), and it is embodied in the notion of "language-specific phonetic rules" proposed by Keating (1985), Port and O'Dell (1985), Pierrehumbert (1990), Cohn (1993), and others.

The foregoing considerations give us a version of the interface view that can be roughly sketched as in (4). The architecture shown in (4) is compatible with quite a range of views. For example, it is silent on the question of whether the grammar is a set of procedural rules as in SPE and Lexical Phonology or whether it is some kind of constraint-satisfaction system like Optimality Theory. By the same token, the details of the quantitative realization model are not important to the overall picture, although I regard Browman and Goldstein's work as an eminently plausible candidate for such a model. I should also make clear that I take segmental "phones" as one of a number of possible types of abstract element at the interface level. That is, nothing in (4) is intended to rule out adding things like metrical structure or feature tiers or other kinds of non-segmental devices to the interface representation; my point is that the elements of the representation are *categorically distinct symbols*.[2] The principal claim diagrammed in (4) is that, in order to describe spoken language adequately, we need to think of phonetic realization as a mapping between a categorical symbolic representation and a quantitative physical signal. Given that premise, my concern here is with how we establish the elements of the representation in the first place, and what we think about the relation between those elements and the quantitative physical realities that they represent. Exploring these questions is the goal of this paper.

(4) Phonology-phonetics interface:



2. The interface representation and the classical phoneme

Traditionally in segmental phonology and phonetics there have been two main sources of authority for positing symbolic abstractions like [t] and [m] – for saying that two physical speech events which we know to be different in physical detail count as "the same". One is the authority of the expert phonetician: A and B are the same because the phonetician hears them as the same and transcribes them as the same. Roughly speaking, this is the basis of the abstraction we call a systematic phonetic transcription. The other kind of authority for positing symbolic abstractions is the authority of the native speaker: A and B are the same because the native speaker perceives no difference between them and/or uses them as if they were the same. This, of course, is the basis of the abstraction we call a taxonomic or classical phonemic transcription.

Systematic phonetic transcription is a useful shorthand for rough observations. Indeed, this seems to be the way the IPA alphabet was conceived of in its early days – as a substitute for quantitative description, *faute de mieux* (Joos 1948). Somewhere along the line, however, systematic phonetics became one of the big ideas of twentieth century phonology, promoted by Pike (1943), Abercrombie (e.g. 1967) and others (the definitive statement of this view seems to be Laver (1994)). The central idea of systematic phonetics is that there is a principled universal set of phonetic categories that abstract away from the infinite variability of the physics of speech. This in turn implies that it should be possible to give a symbolic or non-quantitative description that *fully characterizes* the linguistically significant detail of any utterance of any language. This assumption is the basis on which Chomsky (1964) characterized systematic phonetic transcription as a level of representation in his formalization of phonology: as has been pointed out by Pierrehumbert and Beckman (1988: ch. 1), the *SPE* model (Chomsky and Halle 1968) assumes that only universal biomechanical properties of speech need to be treated in terms of continuous mathematics. As we just saw, however, there is a great deal of recent laboratory research showing that many *language-specific* phonological phenomena can only be fully described in quantitative terms. Systematic phonetic transcription appears incapable in principle of serving as a formal level of representation of speech.

If we insist on the "informal shorthand" status of systematic phonetics, yet retain the notion that there is an interface representation, then the other traditional type of symbolic representation that suggests itself for our purposes is the "classical phonemic" representation. On the face of it this does not look like a very promising idea. After all, the crushing critiques by Halle (1959) and Chomsky (1964) revealed fundamental theoretical problems with the classical phoneme, and their critiques have never been answered. But there are good reasons for considering this approach anyway.

The most obvious reason to take the classical phoneme seriously is that it still plays a large *de facto* role in phonological discussion, notwithstanding Halle and Chomsky. In practice – in speech therapy, in speech technology, in orthography design, in studies of language acquisition, and for that matter in most descriptive work within Optimality Theory – the phoneme notion continues to be applied as if there were no serious problem with it. For a theoretical construct that was discredited forty years ago, the classical phoneme is actually still doing pretty well. This has been true throughout the history of generative phonology. Only three years after the publication of SPE, Schane (1971) was already pointing out that whenever generative phonologists ignored so-called "low-level phonetic detail" – which was most of the time – their analyses generated a surface representation "almost amazingly identical to a classical phonemic representation (1971: 520)". At the very least, this means that many linguists of varying theoretical persuasions over many years have found an abstraction like the classical phonemic representation useful.

This being the case, let us go back and reexamine Halle's and Chomsky's evidence for discrediting the classical phoneme in the first place. If we do that, we see that there is an Achilles' Heel to their critiques – namely, they depend crucially on the idea that systematic phonetics is a level of representation. Note that SPE phonology assumes an interface view of phonology and phonetics that is more or less identical in overall structure to the view sketched in (4) above. Specifically, the SPE model, like the one in (4), assumes an abstract level showing the morphemic identity of phonologically conditioned allomorphs, which Chomsky called "systematic phonemic representation"; a concrete "physical phonetic" level that characterizes speech events in quantitative terms; and an intermediate level that expresses linguistically relevant phonetic distinctions in terms of abstract symbolic categories that can be manipulated by a grammar. The SPE version of (4) is shown in (5).

What Halle and Chomsky criticized about classical phonemic theory was that, in their view, it was actually a *four*-level model, with *two* representations intermediate between the morphophonological and the physical, as shown in (6).

(5) Levels of phonological representation in SPE:

underlying morphemic identity
(= "systematic phonemic representation")

↓

Phonological rules

↓

surface phonological representation
(= "systematic phonetic representation")

↓

Phonetic realization component

↓

Physical and psychophysical events
(= "physical phonetics")

(6) Classical phonemic theory according to Chomsky 1964:

systematic phonemic representation

↓

Morphophonemic rules

↓

"taxonomic phonemic representation"

↓

"systematic phonetic representation"

↓

Phonetic realization component

↓

physical phonetics

The fundamental problem that both Halle and Chomsky identified in classical phonemic theory, as represented in (6), was that it involves a mapping between one intermediate symbolic representation and another. Given the four-level picture in (6), Chomsky argued that various problems arise from positing two intermediate symbolic representations, and proposed to do away with the taxonomic phonemic level. Halle's famous argument based on Russian obstruent voicing assimilation (Halle 1959) reached the same conclusion on similar grounds.

However, the Halle-Chomsky argument works only if systematic phonetics is a formal "level of representation". If we go back to the original IPA conception of systematic phonetics as a shorthand for things that ought to be described quantitatively, and if we take seriously all the recent laboratory evidence about the non-categorical nature of assimilation and deletion and so forth, then the Halle-Chomsky argument – given the similarity of (5) to (4) – actually favors *retaining* the classical phonemic level. In effect, Halle and Chomsky were right to say that one of the intermediate representations in (6) is superfluous; with the benefit of forty years' hindsight, however, we can see that they eliminated the wrong one. I do not, of course, mean to suggest that we should simply equate the interface representation in (4) with a 1950s phonemic transcription and pick up where we left off half a century ago. My point is just that, if there is no systematic phonetics, then Halle's and Chomsky's arguments against the classical phoneme lose much of their force.[3]

3. On refining the phoneme

We may summarize the argument so far as follows. First, there are good reasons for an approach to phonology that involves an interface between a grammar-like component that deals in phonotactic structures and constraints, and a physical realization component that deals in quantitative parameters. Second, there are good reasons for suspecting that the interface representation is going to look something like a classical phonemic representation, i.e. based on an inventory of language-specific elements or categories. This means that we now need to consider how we might put the classical phoneme notion on a firmer empirical and theoretical basis.

3.1. Phonemes as phonetic categories

In my view, the key to refining the phoneme concept is to take seriously the idea of phonetic abstraction with which the paper began, and to see phonemes primarily as *language-specific phonetic categories* – language-specific abstractions away from phonetic detail. These are the "distinctive phones" of my title. Phonemic distinctions are in the first instance about *sounding different* to the native speaker, not about signaling different messages. If we see things in this light, then we are in a position to deal with some of the classic theoretical problems of the phoneme on the basis of what we know about cognitive categories more generally.

Although as a matter of theory many adherents of the classical phoneme would presumably have endorsed the proposition that phonemes must "sound different" to native speakers of a language, in practice the basis of this proposition was the ability of sound substitutions to signal differences of *meaning*. The minimal pair test was the methodological gold standard of phonemic theory: if substituting one sound for another yields a different word or sentence, then the two sounds are deemed to sound different. In this respect, traditional phonemic theory has distinctly Whorfian overtones: the minimal pair procedure in defining phonemes suggests that lexical contrast will dictate which phonetic distinctions a speaker of the language will be capable of perceiving. There are obviously good reasons for assuming that lexical contrast is *relevant* to phonological systems, but that is no reason to accept the stronger view that differences of meaning (in some fairly narrow sense) define the status of differences of sound (in some fairly loose sense defined in practice by the IPA symbol system). The problems inherent in the stronger view were extensively discussed in the heyday of classical phonemic theory in the 1940s and 1950s; I outline them briefly in the following paragraph.

One problem is what we might call *allophonic awareness***.** In cases of this sort, native speakers of a language consistently hear the difference between two phones that are supposed to be allophones of the same phoneme. The distinction between German *ich-Laut* and *ach-Laut* and pseudo-minimal pairs like *Kuchen* 'cake' vs. *Kuhchen* 'cow (diminutive)' is probably the classic example (Leopold 1948; MacFarland and Pierrehumbert 1991; Moulton 1947), but there are others. Another problem (which was actually rarely recognized as a problem until the advent of variationist sociolinguistics) is the notion of *free variation***.** This is a problem because close inspection almost always finds that the choice of "free" variants is communicatively significant in some way. An example here would be the difference between apical and uvular /r/ in several languages of Western Europe, which never makes a lexical difference but certainly makes a sociolinguistic one, and which is recognized even in lay phonetic terminology. Still another long-standing problem for classical phonemic theorists was the existence of *marginal phonemes*, for example the use of a velar fricative in the name *Bach* by speakers of English who in other respects "don't have a phoneme /x/". Do we count /x/ as a phoneme for those speakers or not? A more serious version of the same problem is what Fries and Pike (1949) referred to as *coexistent phonemic systems*. This problem is exemplified in a number of indigenous languages of Mexico, for example, whose native stop systems have allophonically determined voicing in stops but which have begun to acquire voicing contrasts as a result of massive borrowing from Spanish. Instances of these problems are widely attested in the phonology of virtually every well-studied language, but they have always been relegated to the status of interesting residual issues or have been attributed to incipient sound change or code switching, and they have never been integrated into any theory of surface representation.

These problems largely disappear if we see phonemes as phonetic types or categories, and if we assume that the formation of phonetic categories is a consequence of the whole language environment, not merely lexical contrast. Lexical contrast is obviously a major source of evidence for a language's phonetic taxonomy, but it is not the only one. All kinds of factors can play a role in sensitizing speakers of a language to differences between two phonetically distinct types: sociolinguistic distinctions, folk descriptions of sound types, exposure to neighboring languages or dialects, and paralinguistic usage of potentially lexically contrastive sounds (like the difference between an apical and a lateral click in English – the sounds of "tut-tutting" and of encouraging a horse, respectively).[4]

If we remove the obligatory lexical basis for phonetic categorization, we have no trouble accounting for the fact that babies begin to acquire the phonetic categories of the ambient language before they really have a lexicon. We have no trouble accommodating the otherwise surprising finding by Tees and Werker (1984) that English speakers who had been exposed to Hindi as babies but not subsequently were able to distinguish Hindi dental and retroflex stops as adults, even though they spoke no Hindi at all. We have no trouble understanding why English speakers are perfectly capable of distinguishing apical from lateral clicks but are likely to have trouble with the distinction between, say, voiced, nasal and aspirated lateral clicks, which are lexically distinctive in some languages.

The ideas just sketched are similar to those underlying Pierrehumbert's notion of a "fast phonological preprocessor" (Pierrehumbert 2002), which is intended to account for the categorical nature of many speech perception phenomena. In a paper that did not come to my attention until after I had prepared a final version of this paper, Pierrehumbert (2003) suggests that the likely basis for such categorical preprocessing is the statistical distribution of phonetic tokens in the input available to a listener. Distributional facts will be influenced by lexical contrast, of course, but statistical distribution is primary; in fact, Pierrehumbert explicitly suggests that the phonetic categories of a given language correspond closely to the "positional allophones" of traditional phonemic theory. These are the "distinctive phones" proposed here.

3.2. Partial similarity between categories

In addition to widening the basis on which native speakers are assumed to establish their inventory of phonetic categories, we need to address the assumption that phonemic identity is sharply defined and the same in all contexts. This assumption was summed up in the heyday of the classical phoneme in the phrase "once a phoneme, always a phoneme". According to this doctrine, if two phones contrast lexically in any context, they must always be regarded as distinct phonemes. The implication of this principle is that either native speakers will treat two sounds as the same, in which case we are dealing with instances of a single phoneme, or they will treat them as different, in which case the two sounds are utterly and absolutely different in all contexts.[5] No allowance is made for sounds being somewhat different, or sometimes different and sometimes not, or different but still somehow also the same. Yet all these kinds of uncertainties can easily be found in the way real native speakers treat real sounds.

The cases I wish to discuss here are what we can call *partial similarity* between phonemes (or, looked at the other way, cases of *quasi-contrast* between phones). A good example involves the realizations of the /ai/ diphthong in Scottish Standard English (Scobbie, Turk and Hewlett 1999). Here we find minimal pairs like *side/sighed* and *tide/tied* that arise from the use of the longer open-syllable allophone (conventionally transcribed [ae]) in the morphologically complex forms *sighed* and *tied* and the shorter pre-voiced-stop allophone (conventionally transcribed [ʌi]) in *side* and *tide.* The uncertain phonemic status of the two variants of /ai/ can also be seen in the fact that in monomorphemic but disyllabic words like *spider* and *Bible* speakers may differ among themselves which allophone they use, and any given speaker may use one allophone in some such words and the other in others. Are we dealing with one phoneme or two? In my experience, first-year linguistics students who are native speakers of Scottish Standard

English are often puzzled by the status of the variants of the /ai/ diphthong when they are introduced to the phoneme concept. There are a number of other similar cases of partial similarity or quasi-contrast that are fairly well known, including: the American English *cot/caught* contrast; the East Coast American distinction between *can* 'be able to' and *can* 'metal container' (e.g. Bloch 1948: 20); the distinction between voiced stops with and without preceding nasal in Modern Greek (e.g. Arvaniti and Joseph 2000); the marginal status of the difference between long /e/ (orthographic *e, ee, eh*) and long /ɛ/ (orthographic *ä* or *äh*) in German; and of course, the distinction between German *ich-Laut* and *ach-Laut* discussed earlier.

Let us consider in detail one specific case of partial similarity, namely the mid vowels in French and Italian. According to a traditional phonemic analysis, both languages show a contrast between higher and lower mid vowels (in Italian /e – ɛ/ and /o – ɔ/; in French /e – ɛ/, /o – ɔ/, and /ø – œ/). Some minimal pairs are given in (7).

(7) a. Italian: [ˈpɛska] 'peach' vs. [ˈpeska] 'fishing'; [ˈfɔɾɔ] 'forum' vs. [ˈfoɾɔ] 'hole',
b. French: [ʒœn] 'young' vs. [ʒøn] 'fasting'; [etɛ] 'was' vs. [ete] 'summer'; [sɔt] 'stupid (fem.)' vs. [sot] 'jumps'.

However, this contrast applies only in lexically stressed syllables. In pretonic syllables (and, in Italian, also posttonic), the contrast between higher and lower mid vowels is neutralized, and the phonetic quality in those cases is variable or indeterminate.

Even in stressed syllables, though, there is a special relation of partial similarity between the higher and lower mid vowels. Somehow these vowels do not contrast with each other as completely as most other pairs of phonemes. Trubetzkoy discussed this with respect to French in *Grundzüge* (Trubetzkoy 1969:78):

> [The members of these oppositions] are often felt only as two meaning-differentiating nuances, that is, as two distinct yet closely related phonic entities.… From a purely phonetic point of view, the difference between French i and e is not greater than the difference between e and ɛ. But the closeness of the relationship between e and ɛ is apparent to any Frenchman, while in the case of i and e there can be no question of any particular closeness.

Trubetzkoy's observations about the psychological link between the mid vowels are astute, and I think they apply equally well to Italian.

Several factors provide evidence for the "particular closeness" ("*besondere Intimität*" in the original) that Trubetzkoy talks about. First, in both languages we find speakers who do not make a distinction between the members of the minimal pairs, or who make a distinction but make it in the opposite direction. Second, especially in French, there is also some degree of predictability from the phonetic environment whether we get the higher or lower vowel – we might call this quasi-complementary distribution. Third – and this partially contradicts the predictability of the quasi-complementary distribution – the lexical distribution of the higher and lower vowels is in some cases quite variable from speaker to speaker: some words consistently have one or the other, but many other words may have either. It seems especially telling that in both languages, manuals of good usage *written for native speakers* frequently devote space to the issue of which vowel is used in which words.

This whole situation seems to demand some comment. Rigorously applied classical phonemic theory would have nothing to say except "once a phoneme always a phoneme": the existence of even one minimal pair is enough to establish the existence of the contrast. One thing that seems to argue in favor of this position is the fact that native speakers of both languages seem to have no trouble distinguishing phonetically between the higher and lower variants. And yet at the same time native speakers are not supposed to need manuals of good usage to tell them which phoneme occurs in which word: no speaker of French or Italian is in any doubt about which words contain, say, /i/ rather than /e/. There is some special relationship between the higher and lower mid vowels that somehow manages to coexist with their phonemic distinctness.

Trubetzkoy's explanation of the special relationship between the French mid-vowels was based on the fact that the distinction between them is neutralized in lexically unstressed syllables. He distinguished between "constant" and "neutralizable" oppositions, and proposed that neutralizable oppositions would have a different psychological status. This can't be right. First, some of the cases of quasi-contrast, like the Scottish *side/sighed* cases, do not involve neutralization at all. Second and more important, if the neutralizability of the opposition were the explanation for these observations, then we would expect to find the same link in any case of neutralization. For example, we would expect speakers of American English to be aware of some special relationship between /t/ and /d/ because this distinction is neutralized intervocalically by so-called "flapping"; we would expect speakers of any language with final devoicing to be aware of some special relationship between the members of voiced-voiceless phoneme pairs. I know of no evidence that this is true.

If we rule out neutralizability as the source of the "particular closeness" between the Romance mid vowels, can we find any other explanation? If pressed on this point, most phonologists would probably suggest that this is a case of sound change in progress: specifically, the pairs of higher and lower mid vowels are merging into single phonemes. Similarly, in the case of the Scottish Standard English /ai/ diphthong, we could say that the two allophones are splitting to form two separate phonemes. Yet there are obvious problems with such an explanation. First, the pace of these sound changes is glacial, apparently spanning many generations. Moreover, the situation for any given speaker is relatively clearly defined: the two sounds are different, and are used in specifiable ways, and at the same time exhibit a "particular closeness". The sound change explanation presupposes systemic instability, but in these cases the supposed instability of the system is remarkably stable. The distinction is consistently indeterminate.

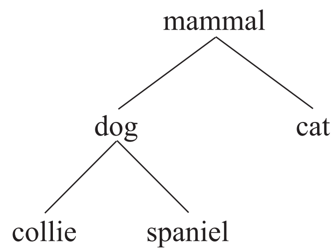
3.3. Categories and subcategories

Such phenomena of stable partial similarity or quasi-contrast can be accommodated in a theory of surface representations if we assume that, like any other system of cognitive categories, phonetic taxonomy can involve multiple levels of organization and/or meaningful within-category distinctions of various kinds. In other cognitive domains no one is surprised to find complex relations between categories, including notions like basic level categories (*cat, dog*), subcategories (*collie, spaniel*), and superordinate categories (*animal, mammal, quadruped*). I believe such notions can usefully be extended to phonetic categories as well.

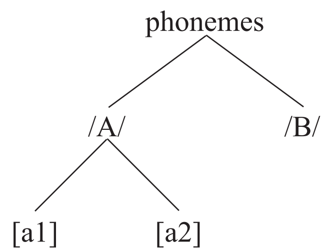
Suppose we treat the classical allophone relationship as a category/subcategory relation, which we might diagram as in (8).

The diagram in (8b) shows two phonetic types (phones) as subcategories (allophones) of a single category (phoneme). In classical phonemic theory, this category-subcategory relationship can hold if and only if the difference between the two phones never gives rise to lexical contrast (either complementary distribution or free variation). What I suggest instead is that this relationship in (8b) can hold in a much greater variety of circumstances, even including cases where the difference between the two phones is sometimes lexically contrastive.

(8) a. Categories and subcategories generally:

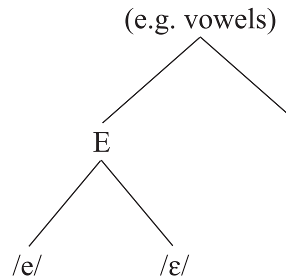


b. Allophonic variation as a category/subcategory relation:



For example, the French and Italian mid-vowels could be seen as instances of just such a category-subcategory relation, as shown in (9):

(9)



What classical phonemic theory says, in effect, is that native speakers are supposed to be aware only of the phoneme-level categories. In (8b), for example, they are supposed to be aware only of /A/, not [a1] and [a2]. Yet awareness of at least some allophonic differences is plainly a fact. If we think of the phoneme-allophone relation as a normal category-subcategory relation, this fact is not the least bit surprising. There is no paradox, for example, in the ability of speakers of English to regard collies and spaniels both as instances of the category *dog* and at the same time as clearly distinct breeds. That is, there is nothing odd in assuming that we can perceive categorical distinctions at several levels. The higher and lower mid vowels of French and Italian can be seen as distinctive phones and at the same time as subtypes of a higher category.

The other point to make in this connection is that category-subcategory relations can be stable**.** Again, in classical phonemic theory, either two phones are in contrast or they are not; either two phonemes have merged or they haven't. This being the case, any observed quasi-contrast phenomenon (like the French mid-vowels, or the Scottish English *side/sighed* distinction) can only be seen as a transitory stage from one phonemic status to another. Yet nothing about the relation shown in (8b) necessarily implies instability or transition. There is nothing to preclude a prolonged period of several generations where two sound types are distinct yet linked together in a category-subcategory relation.[6]

This suggests that we need to adjust our expectations of what native speakers should be able to do. In several of these cases of partial similarity, it appears that native speakers are aware of the distinction between the two phonetic types, but at the same time they apply the distinction quite variably in the lexicon, because somehow the distinction is felt to be only a "meaning-differentiating nuance", in Trubetzkoy's phrase. This is something we want our interface representation, and our theory of phonetic categories, to express. More generally, we want our interface representation to be able to reflect the variety of statuses that sounds can have for native speakers. Just how complicated this may turn out to be is still not clear, but it is an obvious target for empirical work.

To give an idea of what such empirical work might involve, I close this section with a brief description of a recent small-scale survey I made of the distribution of the mid vowels in Italian, based on self-reports by 15 educated native speakers of Standard Italian. For both the front and back mid vowels I emailed the respondents separate lists of approximately 75 ordinary words, including nouns and adjectives, some inflected verb forms, several function words, and a handful of productive derivational suffixes. The front vowel list was sent first, followed a week or two later by the back vowel list. Respondents were simply asked to mark each word with *a* (for *aperto,* 'open') or *c* (for *chiuso,* 'closed'), according to their own usage. The responses showed a striking asymmetry between the front and back vowels: the front vowels show great inter-speaker disagreement, while for the back vowels the speakers mostly agree which words have the higher vowel and which have the lower. Interpreting this finding via the traditional idea of phonemic contrast, we might suggest that the merger of /e/ and /ɛ/ is underway while the contrast between /o/ and /ɔ/ is still unaffected. But the native speakers I surveyed were unanimous that their intuitions were *clearer* about the front vowels than about the back vowels; that is, they found it easy to decide whether a word contained /e/ or /ɛ/, but hard to decide whether a word contained /o/ or /ɔ/, even though they gave markedly different responses in the case of the front vowels and highly consistent responses for the back vowels. An account based on classical phonemic theory and invoking incipient sound change would presumably predict that the consistent judgments should be easy and the variable ones difficult. I intend to follow up this preliminary work soon, with laboratory evidence on perception and production to supplement the self-reports.

4. Conclusion

One of the important theoretical works on phonology of the last two decades was Anderson's magisterial *Phonology in the Twentieth Century* (1985). Anderson's subtitle was *Theories of Rules and Theories of Representations,* and his central thesis was that twentieth century phonological theories could usefully be looked at according to how much descriptive work they got their rules to do and how much they got their representations to do. Classical phonemic theory, on this view, was very much a theory of representations, while SPE phonology was a theory of rules.

Superficially, the arrival of Optimality Theory as the twentieth century drew to a close suggests that Anderson was not only premature in his choice of title, but more importantly wrong in his basic idea. The real dichotomy, it might appear, is between theories of rules and theories of constraints. Instead, I would argue that at a slightly deeper level Anderson was right: one of the things that characterizes twentieth century phonology is a dichotomy between "theories of representations" and "theories of things you can do with representations" (cf. also Goldsmith 1993). "Theories of representations" include (1) classical phonemic theory, (2) systematic phonetics, (3) distinctive feature theory from Trubetzkoy and Jakobson to SPE to Clements and Hume, and (4) theories of phonological structure like autosegmental and metrical phonology and their many offshoots. "Theories of things you can do with representations" are (1) Bloomfieldian process morphophonemics, (2) much of classical SPE phonology, and now of course (3) Optimality Theory.

What OT shows us is that you can argue about what kind of things your theory allows you to do with representations, but what the entire history of twentieth century phonology suggests is that *either way you have to have a theory of representations.* The goal of this paper has been to point out some of the phenomena that such a theory will have to take into account.

Notes

* The ideas in this paper have been developing over a long period of time and consequently owe much to many colleagues. I specifically thank Jim Scobbie for long discussions and fruitful collaboration (e.g. Ladd and Scobbie 2003) that have influenced my thinking in important ways. I also thank the members of the Cognitive Science Department at Johns Hopkins University, where I spent four months in 2000 and found both the freedom and the stimulating environment to develop this line of research.

1. This practice of respelling <ck> at syllable breaks does not appear to be sanctioned by the new German orthography, but was well-established for several decades and can still sometimes be seen even in materials printed according to the new orthographic rules.
2. At least two very disparate recent lines of work (e.g. Atterer and Ladd 2004; Gafos 2002; Ladd, Mennen and Schepman 2000) have suggested that the fine temporal coordination of gesturally distinct aspects of the speech signal is phonologically significant and/or phonologically controlled. Gafos argues that this necessitates reference to quantitative abstractions in the phonological grammar; Ladd et al.'s very much more limited theoretical point concerns the nature of the phonological entities that might be relevant to the autosegmental association of intonation with the segmental structure. Here I assume that, contra Gafos, a suitably enriched notion of autosegmental association can allow us to describe these phenomena in an abstract symbolic representation that keeps the interface distinct from the actual coordination of gestures in phonetic implementation. However, I acknowledge that future work along these lines may make the interface view untenable.
3. When I presented this paper at the conference, Steve Anderson (personal communication) suggested to me that the essence of the classical phoneme – and the focus of Chomsky's critique – is the biunique correspondence between the phonemic and the systematic phonetic levels. He argued that it therefore makes no sense to talk about refining the classical phoneme if I am not also rehabilitating biuniqueness (which of course I am not, since Chomsky's critique of biuniqueness seems beyond debate, and a fortiori since I am questioning the formal status of the systematic phonetic level altogether). However, I see my proposal here as refining the basic intuition behind the classical phoneme concept, not reviving

the entire theoretical superstructure of the 1940s and 1950s. I see no contradiction in trying to profit from the intuition while at the same time getting rid of formal theoretical ideas that are manifestly based on an outdated and empirically unsatisfactory understanding of phonetics.

4. This proposition can be tested empirically. For example, one might attempt to demonstrate the existence of different perceptual boundaries in the general phonetic area of dorsal fricatives in speakers of English, Dutch, and German. If sociolinguistic significance, allophonic awareness, and exposure to other dialects and languages are relevant, one might make the following predictions. (1) English speakers will have a single category "dorsal fricative", with no obvious perceptual discontinuities; (2) German speakers will show a relatively sharp boundary between two categories corresponding to *ich-Laut* and *ach-Laut*, with the boundary relatively far forward (e.g. between [x] and [ç]); (3) Dutch speakers will have a sharp boundary between two categories, but in a different place from the boundary in German, based on the sociolinguistic distinction between so-called "soft G" (roughly /x/) and "hard G" (roughly /χ/). Note that none of those languages has any lexical contrasts among dorsal fricatives at all.
5. Obviously, this characterization of classical phonemic theory makes no allowance for neutralization. Some versions of the theory did attempt to deal with neutralization (e.g. the notion of the "archiphoneme"), but it seems fair to say that neutralization and related phenomena posed a fundamental problem for the theory. See further the discussion of examples (7a) and (7b).
6. If we allow a richer set of relations between phonetic categories, we may have a natural mechanism for splits and mergers, through the dynamics of language acquisition. We can see how it might be possible for children to acquire a subtly different grammar from that of their parents without any superficially obvious difference in language behavior. In a phonemic split, the parent would have two contrastive phones that are sub-phonemes of a single category, and the child would acquire and reproduce the categorical difference between the two contrastive phones without also acquiring the superordinate category. In a merger, the parent would have two distinct phonemes, which the child would acquire as distinct phones but would regard as sub-categories of a single higher category that the parent did not have. Again, there would be no immediately observable behavioral consequence. If subsequent generations acquired the two as sub-phonemes, this could set the stage for a more radical loss of distinction; at the same time, it would also leave open the possibility that the merger might reverse itself. The historical reversal of splits and mergers is otherwise quite mysterious.

References

Abercrombie, David
1967 *Elements of General Phonetics*. Edinburgh: Edinburgh University Press.

Anderson, Stephen R.
1985 *Phonology in the Twentieth Century*. Chicago: University of Chicago Press.

Arvaniti, Amalia and Brian D. Joseph
2000 Variation in voiced stop prenasalisation in Greek. *Glossologia* 11–12: 131–166.

Atterer, Michaela and D. R. Ladd
2004 On the phonetics and phonology of "segmental anchoring" of F0: evidence from German. *Journal of Phonetics* 32: 177–197.

Bloch, Bernard
1948 A set of postulates for phonemic analysis. *Language* 24: 3–46.

Browman, Catherine P. and Louis M. Goldstein
1986 Towards an articulatory phonology. *Phonology Yearbook* 3: 219–252.

Bybee, Joan
2001 *Phonology and Language Use*. Cambridge: Cambridge University Press.

Chomsky, Noam
1964 The nature of structural descriptions. In: *Current Issues in Linguistic Theory*. The Hague: Mouton.

Chomsky, Noam and Morris Halle
1968 *The Sound Pattern of English*. New York: Harper and Row.

Cohn, Abigail
1993 Nasalization in English: phonology or phonetics. *Phonology* 10: 43–81.

Coleman, John S.
1998 Cognitive reality and the phonological lexicon: a review. *Journal of Neurolinguistics* 11: 295–320.

Fries, Charles C., and Kenneth L. Pike
1949 Coexistent phonemic systems. *Language* 25: 29–50.

Gafos, Adamantios
2002 A grammar of gestural coordination. *Natural Language and Linguistic Theory* 20: 269–337.

Goldsmith, John A.
1993 Introduction. In: John Goldsmith (ed.), *The Last Phonological Rule: Reflections on Constraints and Derivations*, 1–20. Chicago: University of Chicago Press.

Halle, Morris
1959 *The Sound Pattern of Russian*. The Hague: Mouton.
Halle, Morris and Sylvain Bromberger
1989 Why phonology is different. *Linguistic Inquiry* 20: 5–70.
Hay, Jennifer, Janet Pierrehumbert and Mary E. Beckman
2003 Speech perception, well-formedness, and the statistics of the lexicon. In: John Local, Richard Ogden and Rosalind Temple (eds.), *Papers in Laboratory Phonology VI*. Cambridge: Cambridge University Press.
Joos, Martin
1948 *Acoustic Phonetics (Language monograph no. 23)*. Baltimore: Waverley Press.
Keating, Patricia
1985 Universal phonetics and the organization of grammars. In: Victoria Fromkin (ed.), *Phonetic Linguistics*, 115–132. New York: Academic Press.
Kornai, András
1994 Relating phonetic and phonological categories. In: E. S. Ristad (ed.), *Language Computations*, Volume 17, 21–36. Philadelphia: American Mathematical Society.
Ladd, D. R.
in prep. Could there be a phonology-phonetics interface?
Ladd, D. R., Ineke Mennen and Astrid Schepman
2000 Phonological conditioning of peak alignment of rising pitch accents in Dutch. *Journal of the Acoustical Society of America* 107: 2685–2696.
Ladd, D. R. and James M. Scobbie
2003 External sandhi as gestural overlap? Counterevidence from Sardinian. In: John Local, Richard Ogden and Rosalind Temple (eds.), *Papers in Laboratory Phonology VI*. Cambridge: Cambridge University Press.
Laver, John
1994 *Principles of Phonetics*. Cambridge: Cambridge University Press.
Leopold, Werner
1948 German ch. *Language* 24: 179–180.
MacFarland, Talke and Janet Pierrehumbert
1991 On *ich*-laut, *ach*-laut, and structure preservation. *Phonology* 8: 171–180.
Moulton, William G.
1947 Juncture in Modern Standard German. *Language* 23: 212–226.
Nolan, Francis
1992 The descriptive role of segments: evidence from assimiliation. In: G. J. Docherty and D. R. Ladd (eds.), *Papers in Laboratory Phonology II*, 261–280. Cambridge: Cambridge University Press.

Pierrehumbert, Janet

1990 Phonological and phonetic representation. *Journal of Phonetics* 28: 375–394.

2002 Word-specific phonetics. In: Carlos Gussenhoven and Natasha Warner (eds.), *Laboratory Phonology VII*. Berlin/New York: Mouton de Gruyter.

2003 Probabilistic Phonology: Discrimination and Robustness. In Rens Bod, Jennifer Hay and Stefanie Jannedy (eds.): *Probabilistic Linguistics*, 177–228. Cambridge, MA: MIT Press.

Pierrehumbert, Janet and Mary E. Beckman

1988 *Japanese Tone Structure*. Cambridge, MA: MIT Press.

Pike, Kenneth L.

1943 *Phonetics*. Ann Arbor: University of Michigan Press.

Port, Robert F. and Michael L. O'Dell

1985 Neutralization of syllable-final voicing in German. *Journal of Phonetics* 13: 455–471.

Schane, Sanford A.

1971 The phoneme revisited. *Language* 47: 503–521.

Scobbie, James M., Alice E. Turk and Nigel Hewlett

1999 Morphemes, phonetics and lexical items: The case of the Scottish Vowel Length Rule. In: *Proceedings of the XIVth International Congress of Phonetic Sciences*, Volume 2, 1617–1620. San Francisco, CA.

Tees, Richard and Janet Werker

1984 Perceptual flexibility: Maintenance or recovery of the ability to discriminate non-native speech sounds. *Canadian Journal of Psychology* 38: 579–590.

Trubetzkoy, N. S.

1969 *Principles of Phonology* (translated by Christiane A. M. Baltaxe). Berkeley: University of California Press.

The functionality of incomplete neutralization in Dutch: The case of past-tense formation*

Mirjam Ernestus and Harald Baayen

In many languages, underlyingly voiced obstruents are realized as voiceless in word-final position (Final Devoicing). Previous research has shown that this neutralization of the underlying [voice]-specification may be phonetically incomplete, with underlyingly voiced obstruents being realized as slightly voiced. Listeners are sensitive to incomplete neutralization, since they can tell apart above chance level the phonetic realizations of words underlyingly ending in voiced and voiceless obstruents. This study discusses to what extent incomplete neutralization is functional. It reports a series of experiments showing that incomplete neutralization can be induced in Dutch just by the way words are spelled, and that no minimal word pairs are required for incomplete neutralization to emerge. In addition, these experiments also show that Dutch listeners take advantage of incomplete neutralization, even when there are no task-specific reasons for them to do so. Incomplete neutralization appears to be a subphonemic cue to past-tense formation, and has a more substantial role in language processing than has been assumed thus far. Finally, our data show that Dutch listeners and speakers dynamically adapt their production and interpretation of the acoustic signal to the voicing properties of the orthographic and acoustic forms encountered previously in the experiment.

1. Introduction

In many languages, including Catalan, Dutch, German, Polish, and Russian, underlyingly voiced obstruents are realized as voiceless in word-final or syllable-final position (see, e.g., Kenstowicz 1994:75). This Final Devoicing results in the phonological neutralization of underlyingly voiced and voiceless obstruents. Example (1) illustrates Final Devoicing in Dutch. The verbs [vɛrʋɛidən] *verwijden* and [vɛrʋɛitən] *verwijten* differ phonologically only in the [voice]-specification of the stem-final alveolar plosive preceding the infinitive suffix [ən] *en*, see (1a). The [voice]-specifications of these plosives are distinctive, and the plosives must therefore be specified underlyingly as /d/ and /t/ respectively. When the stems occur word finally, the plosives are syllable-final, and they are both realized as [t] (unless regressive voice assimilation applies). The underlying phonological difference is neutralized, as shown in (1b).

(1)	a.	*verwijden*	[vɛrʋɛidən]	'widen-inf'
		verwijten	[vɛrʋɛitən]	'reproach-inf'
	b.	*verwijd*	[vɛrʋɛit]	'widen'
		verwijt	[vɛrʋɛit]	'reproach'

Previous research has shown that Final Devoicing does not always result in complete neutralization at the phonetic level (e.g., Dinnsen and Charles-Luce 1984; Port and O'Dell 1985; Slowiaczek and Dinnsen 1985; Port and Crawford 1989; Warner et al. 2004). Neutralized obstruents that are underlyingly voiced tend to have more acoustic characteristics of voiced obstruents than neutralized obstruents that are underlyingly voiceless. They tend to be shorter, to be realized with vocal fold vibration during a longer period, and to be preceded by longer vowels. In what follows, we will refer to voiceless obstruents that posses some acoustic characteristics of voiced obstruents as slightly voiced obstruents.

Although the acoustic differences between neutralized underlyingly voiced and voiceless obstruents are generally small, listeners are able to take advantage of these differences. They assign above chance level the correct spelling to the members of minimal word pairs that differ from each other only in the underlying [voice]-specification of their final obstruents, which is reflected in their spellings (e.g., Port and O'Dell 1985; Port and Crawford 1989; Warner et al. 2004). The question arises as to which extent incomplete neutralization plays a role in normal speech processing. In the spelling tasks, listeners may be forced to base their choices on acoustic features they would not rely on in normal processing circumstances. The finding that listeners take advantage of incomplete neutralization may be an artificial task effect (Warner et al. 2004).

This paper addresses the question of the extent to which the interpretation of incomplete neutralization is functional. We tested whether listeners take advantage of incomplete neutralization also when their task does not force them to do so.

The language under investigation is Dutch. Incomplete neutralization in this language has been reported by Warner et al. (2004) and by Ernestus and Baayen (in press). Warner et al. (2004) carried out a production experiment with 15 native speakers of Dutch, reading 20 minimal word pairs. They observed a significant difference in vowel duration with vowels preceding underlyingly voiced obstruents being on average 3.5 ms longer than vowels preceding underlyingly voiceless obstruents. In addition, they also found a difference in burst duration in the predicted direction (underlyingly voiceless plosives had longer bursts), both for the words with phonologically long and the words with phonologically short vowels (see their Table 2), but the difference reached significance only for the words with long vowels. In the study by Ernestus and Baayen (in press), underlyingly voiced and voiceless plosives differed in the duration of their release noise, which includes both the plosive's burst and the following period of aspiration. This study was based on 29 words ending in underlyingly voiceless plosives and 30 words ending in underlyingly voiced plosives, realized by a single speaker. In Dutch, incomplete neutralization is apparently mainly evidenced by vowel duration and the duration of the burst or release noise.[1]

We carried out two perception experiments in which native speakers of Dutch listened to the stems of pseudo verbs. The stems were all realized with voiceless final obstruents, obeying Final Devoicing. We asked the participants to write down the past-tense forms of the verbs. In Dutch, past-tense forms are created by suffixing [tə] *te* or [də] *de* to the verb stem. The suffix [tə] is added when the stem underlyingly ends in a voiceless obstruent, while [də] is added elsewhere (e.g., Booij 1995). The participants in our past-tense formation experiments, therefore, had to interpret the final neutralized obstruents as underlyingly voiced or voiceless. We showed in Ernestus and Baayen (2003) that participants base their interpretation on the bias of the gangs of phonologically similar words. That is, participants tend to add *te* to a pseudo verb when most phonologically similar words underlyingly end in voiceless obstruents, and they add *de* when most phonologically similar words underlyingly end in voiced obstruents. In the present study, we investigate whether participants also base their interpretation of stem-final obstruents on acoustic subphonemic cues to underlying voicing resulting from incomplete neutralization. If so, they should more often add *de* to a verb stem when its final obstruent is realized as slightly voiced than when it is realized as completely voiceless.

One way to obtain pseudo words with slightly voiced obstruents is to create these words out of words with completely voiceless obstruents by changing those acoustic characteristics that are different for underlyingly voiced and voiceless neutralized obstruents in Dutch. Another method is suggested by previous research which shows that orthography may induce or enforce incomplete neutralization (Fourakis and Iverson 1984; Port and Crawford 1989). We opted for this latter method. It may not only provide us with slightly voiced obstruents, but also with additional information about how slight voicing is realized in Dutch. We asked a native speaker of Dutch to read aloud pseudo verb stems that were spelled with voiceless final obstruents (e.g., *daup*) and with voiced final obstruents (*daub*).

2. Experiment I

2.1. Materials

We compiled two lists of pseudo verb stems, in which each stem is preceded by the pronoun *ik* [ɪk] 'I'. In these lists, the verb stems function as first person singular present tense forms. The first list follows the design of the experiment reported in Ernestus and Baayen (2003), and contains monosyllabic pseudo verbs ending in bilabial or alveolar plosives, or alveolar, labiodental, or velar fricatives. All final obstruents, except the velar fricatives, are spelled as voiceless in this list, and we therefore refer to this list as List –Voice. We will call the second list List Slightly Voiced. This list is identical to List –Voice, except that the 14 final bilabial plosives following sonorant segments are spelled with the letter *p* in List –Voice and with *b* in List Slightly Voiced, the 28 final alveolar plosives following sonorant segments are spelled with *t* in List –Voice and with *d* in List Slightly Voiced, and, finally, the 16 final velar fricatives are spelled with *g* in List –Voice and with *ch* in List Slightly Voiced. The remaining 115 words are spelled identically in both lists, and will be referred to as the words with constant spelling. They are all spelled with final graphemes representing voiceless obstruents, since, according to the Dutch spelling conventions, they cannot end in graphemes representing voiced obstruents. List –Voice contained an additional nineteen words, which we will not consider here. Table 1 summarizes the materials.

We expected that all final obstruents would be realized as voiceless, because of Final Devoicing. In addition, we also expected that *b*, *d*, and *g* would be realized with more characteristics of fully voiced obstruents than *p*, *t* and *ch*, since *b*, *d*, and *g* represent underlyingly voiced obstruents in Dutch, whereas *p*, *t* and *ch* represent underlyingly voiceless obstruents. That is, we expected that vowels preceding *b*, *d*, and *g* would be longer than the vowels before *p*, *t* and *ch*, and that *b* and *d* would be realized with shorter release noises than *p* and *t*. Finally, we expected that the obstruents that were spelled identically in the two lists might be more voiced in List Slightly Voiced than in List –Voice. If *b* and *d* would indeed be slightly voiced in List Slightly Voiced, this might lead to automation of the articulatory gestures for slight voicing. As a consequence, also the final obstruents of the words with constant spelling, which do not carry information in their orthography about underlying voicing, might be slightly voiced. Conversely, List –Voice may lead to automation of the articulatory gestures for voicelessness, and the final obstruents of the words with constant spellings might be realized as completely voiceless in this list.

Table 1. Summary of the materials used in Experiments I and II, with the following abbreviations: "Son" for "sonorant segment", "P" for "bilabial plosive", and "T" for "alveolar plosive". Column "Word type" indicates whether the words end in a sonorant segment and a plosive, in a velar fricative, or are words with constant spelling. Columns "List –Voice" and "List Slightly Voiced" show the spelling of the final obstruents in these lists.

Words ending in	Word type	Number of words	List –Voice	List Slightly Voiced
Son+P	Son+Plosive	14	*p*	*b*
Son+T	Son+Plosive	28	*t*	*d*
Velar fricative	Velar fricative	16	*g*	*ch*
Obstruent+P	Constant spelling	3	*p*	*p*
Obstruent+T	Constant spelling	34	*t*	*t*
Alveolar fricative	Constant spelling	57	*s*	*s*
Labiodental fricative	Constant spelling	21	*f*	*f*

2.2. Procedure

We asked a female native speaker of Dutch to read aloud the two lists. There was an interval of an hour between the recording of List –Voice and that of List Slightly Voiced. Our speaker declared afterwards that she had not noticed that the two lists were nearly identical.

We recorded her speech on a DAT (BASF master 94) by means of an Aiwa HD S100 recorder and a Sony microphone ECM MS957 in a sound attenuated room. We digitized the recorded speech at a sampling rate of 48 kHz.

2.3. Results and discussion

We first investigated whether the words spelled with voiced final graphemes had been realized with more voicing than the corresponding words spelled with voiceless graphemes by measuring the durations of the vowels. In case a vowel was followed by a sonorant consonant, we included the duration of this sonorant in our measurement, since it is in general impossible to determine exactly where the vowel ends in the acoustic signal and where the following sonorant consonant begins. We defined the beginning of the vowel as the beginning of the regular wave form with the characteristics of the vowel, and the end of the vowel (plus sonorant consonant) as the (sudden) end of the regular wave form. In addition, we also measured the durations of the final velar fricatives, the durations of the closures of the final plosives, and the release noises of these plosives, which include the bursts of the plosives and the following period of aspiration (see Figure 1 for two wave forms of release noises). We took the closure of the plosive to end at the sudden increase in amplitude at the beginning of the burst, and we assumed that the fricative and the release noise of the plosive ended where the amplitude of the wave form was nearly identical to that of the background noise.

We found a significant difference between the realizations of the graphemes *b* and *d* on the one hand and the graphemes *p* and *t* on the other hand. The graphemes *b* and *d* were on average realized with shorter release noises (mean: 96 ms) than *p* and *t* (mean: 119 ms, paired $t(42) = 4.14$, $p < 0.001$; all t tests in this study are two tailed). We observed no significant difference between the closure durations of the plosives (paired $t(42) = 0.06$, $p = 0.957$), nor between the durations of the preceding vowel plus sonorant consonant sequences (paired $t(42) = 0.31$, $p = 0.762$). This implies that the difference in release noise does not simply result from a difference in speech rate. Shorter release noises are characteristic of voiced obstruents (Slis and Cohen 1969), and our results therefore provide additional evidence that incomplete neutralization in Dutch may be evidenced by the duration of the release noise. In addition, we conclude that spelling can induce incomplete neutralization in Dutch, even if the speaker does not read minimal word pairs.

We found no significant differences between the realizations of *g* and *ch*. These fricatives did not differ in duration ($t(15) = -0.99$, $p = 0.343$), nor did the preceding vowel plus sonorant consonant sequence (t(15) = 1.52, p = 0.149). The absence of a significant difference may be due to the low number of final velar fricatives (16) in our data set. In addition, it may be due to the weak status of the voiced velar fricative in Dutch. The distinction between the voiced and the voiceless velar fricative is only functional in intervocalic positions, and only for speakers of particular variants of Standard Dutch.

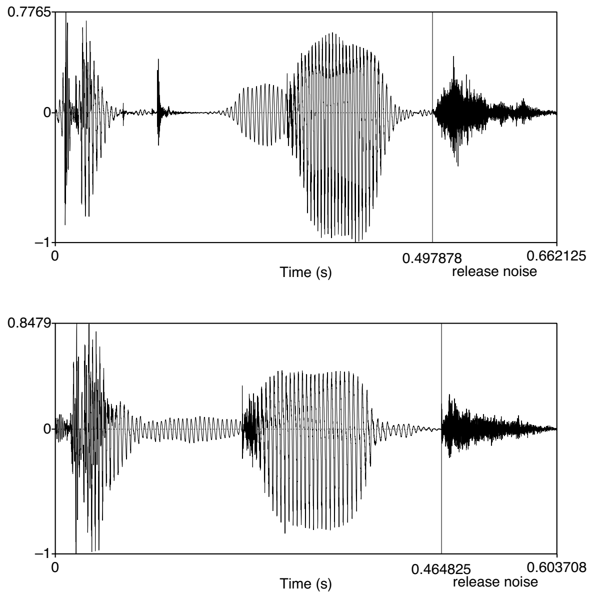


Figure 1. The wave forms of the phrases *ik duut* (upper panel) and *ik duud* (lower panel), with the vertical line in the wave forms indicating the beginning of the release noise.

We then investigated whether the words with constant spelling had been realized with different amounts of final voicing in the two lists. We measured the durations of the vowels, and the duration of the codas. If the words with constant spelling are more voiced in List Slightly Voiced than in List –Voice, we expect that their vowels are longer in List Slightly Voiced. With respect to the durations of the codas, although voiced obstruents tend to be shorter than voiceless obstruents in intervocalic position, we do not expect that the codas of the words with constant spelling are shorter in List Slightly Voiced than in List –Voice. Ernestus and Baayen (in press) found that the period of vocal fold vibration is an important cue to the voicing of final obstruents in Dutch, as it is for the voicing of obstruent clusters in intervocalic positions

(Van den Berg 1988). The period of vocal fold vibration can be longer if the coda itself is longer. Hence, our speaker may have lengthened the codas in List Slightly Voiced in order to have them sound as more voiced. We also measured the periods of vocal fold vibration in the obstruent clusters, on the basis of the spectrograms of the sound files. We expect them to be longer in List Slightly Voiced than in List –Voice.

First, we compared the words with constant spelling with respect to the durations of their vowels, by means of a linear mixed effects model (LME, Pinheiro and Bates 2000) with the pseudo word as error stratum and with the duration of the vowel as the dependent variable. We took as independent variables List (List –Voice, List Slightly Voice) and the qualitative length of the vowel, which is short for [i y u ɛ ɪ ɔ ɑ ʏ] and long for the other Dutch vowels, and reflects the fact that in the strong positions of prosodic feet these other Dutch vowels are phonetically long (Rietveld, Kerkho, and Gussenhoven 1999). This model did not provide evidence for differences in vowel duration between the two lists.

An LME model with duration of the coda as dependent variable, the pseudo word as error stratum, and List, Qualitative length of the vowel, and Type of coda (single fricative, fricative-fricative, plosive-fricative, plosive-plosive, fricative-plosive) as independent variables revealed main effects for Type of coda ($F(4,105) = 51.62$, $p < 0.001$) and List ($F(1,112) = 5.20$, $p = 0.025$), and an interaction of Type of coda by Qualitative length of the vowel ($F(5,105) = 2.76$, $p < 0.022$), after removal of two outliers. The different types of codas had average durations of 312.0 ms, 388.6 ms, 399.6 ms, 314.5 ms and 351.1 ms, respectively. Single fricatives and fricative-fricative clusters were longer after qualitatively long vowels than after short vowels (on average 2.5 ms and 19.9 ms longer, respectively), whereas fricative-plosive, plosive-fricative, and plosive-plosive clusters were on average longer after qualitatively short vowels (33.9 ms, 7.5 ms, and 35.1 ms longer, respectively). More importantly for our research question, codas were on average 10.9 ms shorter in List –Voice than in List Slightly Voiced. Codas may be longer in List Slightly Voiced because our speaker lengthened the codas in this list in order to lengthen the periods of vocal fold vibration. This hypothesis is supported by the analysis of the periods of vocal fold vibration in the obstruent clusters to which we turn now.

We analyzed also the periods of vocal fold vibration during the final obstruent clusters with an LME model with the pseudo word as error stratum. We discarded 11 items (out of 146) that were realized with no vocal fold vibration whatsoever. The removal of these outliers improved the normality of the dependent variable, which led to an improvement in the normality of the model's residuals. We took List, Type of coda, Duration of the coda, and Qualitative length of the preceding vowel as independent variables. The model revealed main effects for all factors, except the duration of the coda (List: $F(1,54) = 87.42$, $p < 0.001$; Type of coda: $F(3,68) = 30.49$, $p < 0.001$; Qualitative length of the vowel: $F(1,68) = 12.11$, $p < 0.001$). In addition, there was an interaction of List by Type of coda ($F(3,54) = 10.29$, $p < 0.001$), which is illustrated in Table 2. This table gives for each list the mean length of the period of vocal fold vibration during the different types of obstruent clusters. The differences between the lists are significant for the plosive-fricative clusters (paired $t(19) = –5.36$, $p < 0.001$), the plosive-plosive clusters (paired $t(11) = –7.29$, $p < 0.001$), and the fricative-plosive clusters (paired $t(24) = –2.85$, $p = 0.009$). The periods of vocal fold vibration are longer in List Slightly Voiced than in List –Voice, which suggests that the words of List Slightly Voiced have been realized with more final voicing than the words of List –Voice. The LME model also revealed an interaction of the duration of the coda by the qualitative length of the vowel ($F(1,54) = 4.86$, $p = 0.032$). Longer codas were realized with longer periods of vocal fold vibration after qualitatively long vowels. Finally, we also found an interaction of the duration of the coda by List ($F(2,54) = 4.78$, $p = 0.012$). Longer codas are realized with longer periods of vocal fold vibration, especially in List Slightly Voiced. This interaction supports the hypothesis that our speaker has lengthened the codas in List Slightly Voiced in order to lengthen the periods of vocal fold vibration, which may make these codas sound as more voiced. We conclude that the words with constant spelling were realized with more voicing in List Slightly Voiced than in List –Voice.

Table 2. The mean length of the period of vocal fold vibration in milliseconds during the four types of final obstruent clusters in the words with constant spelling in the two lists.

Obstruent cluster	List –Voice	List Slightly Voiced
Fricative + Fricative	11.143	11.200
Plosive + Fricative	17.120	34.875
Plosive + Plosive	21.008	40.225
Fricative + Plosive	10.396	15.496

Experiment I shows that in Dutch the spelling may affect how a speaker realizes the fine acoustic details of final obstruents. The next experiment addresses the question of whether Dutch listeners might be sensitive to these acoustic characteristics in a past-tense production task. In this experiment, participants listened to the pseudo words of either List –Voice or List Slightly Voiced, and were asked to write down the past-tense forms of these verbs. To do so, they had to interpret the final obstruents as underlyingly voiced or voiceless. This task does not require sensitivity to incomplete neutralization, since the structure of the rime of the words is a crucial determinant for the underlying voicing (Ernestus and Baayen 2003). If listeners nevertheless make use of the subtle acoustic characteristics of the words for classifying the final obstruents as voiced or voiceless underlyingly, then we may expect them to interpret the final obstruents that are slightly voiced to be underlyingly voiced more often than the obstruents that were realized as completely voiceless. This would allow us to conclude that listeners take advantage of incomplete neutralization, even when there are no intrinsic reasons for them to do so.

In Experiment II, participants listened either to List –Voice or List Slightly Voiced. That is, they heard either no slightly voiced obstruents or a high number of such obstruents. We might expect only a small effect of slight voicing in a list with many slightly voiced and few voiceless final obstruents. Speech is highly variable, and listeners can interpret subtle characteristics of the acoustic signal, such as segment duration, only by comparing them to those of other parts of the signal. Thus, listeners interpret vowels as long or short by reference to the durations of preceding segments (Nooteboom 1979; Nooteboom and Doodeman 1980; Kemps et al. 2005). In a list with relatively many slightly voiced obstruents, slight voicing is the norm, and will not be interpreted as providing much information about the underlying voice specification of the final obstruent. This means that if we nevertheless find a significant difference in the responses to List –Voice and List Slightly Voiced, we can conclude that the interpretation of incomplete neutralization is a strong intrinsic feature of Dutch.

3. Experiment II

3.1. Participants

Fifty-six native speakers of Dutch, most of them students at Nijmegen University, were paid to participate in the experiment. Twenty-eight participants listened to the phrases of List –Voice, and twenty-eight participants listened to the phrases of List Slightly Voiced.

3.2. Materials

We used the phrases recorded for Experiment I. These phrases were presented in one of three random orders to the participants, with four intervening breaks. The actual test phrases were preceded by eleven practice phrases with existing verbs, and twenty practice phrases with pseudo verbs, which had also been recorded by the speaker from Experiment I.

3.3. Procedure

The participants listened through closed headphones (Sony MDR 55) to the phrases, and wrote down as accurately as possible the past-tense forms of the pseudo verbs. The experiment was self-paced. Participants heard a new phrase only after they had indicated that they were ready by pressing a button.

3.4. Results and discussion

In the great majority of trials, participants wrote past-tense forms ending in *te* or *de*. Some past-tense forms, however, were created by vowel alternation. For instance, a number of participants produced *ties* as the past-tense form for [tas], and *bast* as the past-tense form for [bɪst]. This type of past-tense formation seems to be independent of the underlying [voice]-specification of the stem-final obstruent, and we therefore did not take these forms into account in the analysis. We also discarded responses ending in *bte*, *pde*, *dte*, *tde*, *zte*, *vte*, *gte*, and *chde*, which are illegal according to the Dutch spelling conventions. In these responses, the grapheme representing the stem-final obstruent indicates an underlying [voice]-specification that is opposite to the specification indicated by the form of the past-tense suffix, and we do not know whether the participants interpreted the stem-final obstruent as underlying voiced or voiceless. Finally, we also discarded past-tense forms the stems of which do not completely correspond to the presented stimuli. The participants probably misunderstood the words in these trials. In total, we discarded 8.58% of the produced past-tense forms (831 out of 9688).

We calculated the log of the ratio of the resulting number of *de* responses to the resulting number of *te* responses for each verb in each list. This logit is our dependent variable.

We first investigated whether the participants determined their responses on the basis of phonological gang bias, as did the participants in the study by Ernestus and Baayen (2003). Ernestus and Baayen (2003) determined the relevant phonologically gangs by means of a Classification Tree (Breiman et al. 1984). This tree had as its input all 1697 lexical stems that end in an obstruent and are present in the CELEX lexical database (Baayen, Piepenbrock, and Gulikers 1995). It grouped these stems into 11 gangs of words, such that words ending in the same rime are grouped together, and rimes with a similar preference for an underlyingly voiced specification for the final obstruent form one gang. The gang bias for a given word is the proportion of words in its gang underlyingly ending in a voiced obstruent.

Recall that the two lists contained three types of words. First, they contained words ending in a sonorant segment plus a bilabial or alveolar plosive. This plosive was slightly voiced in List Slightly Voiced. Second, the lists contained words ending in a velar fricative, which was equally voiced in the two lists. Third, the lists contained words with constant spelling, whose final obstruents were more voiced in List Slightly Voiced than in List –Voice.

We analyzed the logits using a linear model with Phonological gang bias, List (List –Voice and List Slightly Voiced), and Type of word as independent variables. We found main effects for all these independent variables (Phonological gang bias: $F(1,339) = 539.79$, $p < 0.001$; List: $F(1,339) = 11.67$, $p < 0.001$; Type of word: $F(2,511) = 46.33$, $p < 0.001$). In addition, we also found an interaction of Phonological gang bias by Type of word ($F(2,339) = 5.90$, $p = 0.003$).

As Phonological gang bias is not properly normally distributed, we also carried out non-parametric tests in order to ascertain an effect of this factor. First, Spearman correlation tests for List –Voice and List Slightly Voiced separately confirm that Phonological gang bias is a good predictor of the participants' responses (List –Voice: $r_s = 0.63$, $S = 320439$, $p < 0.001$; List Slightly Voiced: $r_s = 0.58$, $S = 365893$, $p < 0.001$). Second, a Regression Tree (Breiman et al. 1984) also supports Phonological gang bias as a predictor of the participants' responses. We conclude that listeners based their choice between *te* and *de* at least in part on the phonological gangs. The participants did not have to take advantage of the subtle cues to voicing in the acoustic signal to do this task.

The question of interest to the present study is whether the participants nevertheless took advantage of incomplete neutralization. We therefore fitted a linear model to the data with the logit of the *de* and *te* responses as dependent variable, and with List and Type of word as independent variables. List

($F(1,170) = 31.11$, $p < 0.001$) and Type of word ($F(2,170) = 60.10$, $p < 0.001$) emerged as significant predictors. There was no interaction of List by Type of word. Table 3 gives the mean proportions of *de* responses for the three types of words in the two lists.

Table 3. The mean proportion of *de* responses in List –Voice (Experiment II), List Slightly Voiced (Experiment II), and List Mix (Experiment III) for the words ending in sonorant segments plus plosives, the words ending in velar fricatives, and the words with constant spelling.

Word type	List –Voice	List Slightly Voiced	List Mix
Sonorant + Plosive	0.072	0.104	0.137
Velar fricative	0.770	0.779	0.719
Constant Spelling	0.216	0.278	0.251

The words ending in sonorant segments plus plosives received more *de* responses from the listeners to List Slightly Voiced than from the listeners to List –Voice. This is in line with our finding in Experiment I that these words were realized with more voiced like plosives in List Slightly Voiced than in List –Voice. Listeners clearly use the subphonemic cues in the signal when choosing the underlying [voice]-specifications for the final obstruents of pseudo words. This is not a trivial result, since listeners also base their choice on the phonological gangs of the words, and therefore do not need to take advantage of incomplete neutralization. Apparently, listeners also rely on incomplete neutralization when there are no compelling reasons for them to do so. In Dutch, incomplete neutralization emerges as a subphonemic cue to past-tense formation.

The words with constant spelling also received more *de* responses from the listeners to List Slightly Voiced than from the listeners to List –Voice. This result shows that longer periods of vocal fold vibration lead listeners to interpret final obstruents as voiced more often. Moreover, it shows that listeners interpret incomplete neutralization even when different interpretations do not map onto different spellings.

Finally, the words ending in velar fricatives received approximately similar responses in List –Voice and List Slightly Voiced. This suggests that these words were equally voiced in the two lists, in line with the absence of systematic acoustic differences in Experiment I.

In Experiment II, listeners heard either no slightly voiced obstruents or a high proportion of slightly voiced obstruents. In Experiment III, we investigated the effect of the proportion of slightly voiced obstruents in the list. We constructed a new list, List Mix, by taking the words from List Slightly Voiced, but replacing the acoustic realizations of the words with constant spelling with the acoustic realizations of these same words from List –Voice. The only slightly voiced obstruents in List Mix are thus the plosives following sonorant segments. We expected that these slightly voiced plosives would be more often interpreted as underlyingly voiced in List Mix than in List Slightly Voiced, since slight voicing is more exceptional in List Mix, and should be more informative.

4. Experiment III

4.1. Participants

Thirty native speakers of Dutch, who had not participated in Experiment II, took part in the experiment. Most of them were students at Nijmegen University. They were paid for their participation.

4.2. Materials

The materials are identical to List Slightly Voiced, except that the realizations of the words with constant spelling were replaced by the realizations from List –Voice. We refer to this new list as List Mix.

4.3. Procedure

The procedure was identical to that of Experiment II.

4.4. Results and discussion

We discarded the same types of responses as we did in Experiment II, which represent 9.21% of all responses.

We analyzed the data of this experiment together with the data of Experiment II, with the log of the ratio of the number of *de* responses to the number of *te* responses as the dependent variable. A linear model with List and Type of word as independent variables revealed a main effect of List ($F(2,340) = 15.71$, $p < 0.001$) and Type of word ($F(2,170) = 50.33$, $p < 0.001$). In addition, there was an interaction of List by Type of word ($F(4,340) = 5.57$, $p < 0.001$). This interaction is illustrated in Table 3, which gives the mean proportion of *de* responses for the three types of words in the three lists.

The words ending in sonorants plus plosives elicited more *de* responses in List Mix than in List Slightly Voiced (paired $t(41) = -2.517$, $p = 0.016$), even though their acoustic characteristics were exactly identical in these two lists. This difference must be due to the words with constant spelling, since List Slightly Voiced and List Mix only differ with respect to the acoustic realizations of just these words. The words with constant spelling had more final voicing in List Slightly Voiced than in List –Voice and List Mix. We conclude that more voiced realizations of the words with constant spelling diminishes the likelihood of voiced interpretations for the slightly voiced plosives following sonorant segments. This points to a list bias effect on the part of the listeners. Listeners attach more value to slight voicing if slight voicing is not the norm.

As can be seen in Table 3, the words with constant spelling elicited fewer *de* responses in List Mix than in List Slightly Voiced (paired $t(114) = -4.78$, $p < 0.001$). In contrast, the percentage of *de* responses for these words does not differ significantly between List Mix and List –Voice (paired $t(114) = -1.63$, $p = 0.105$). Recall that Lists –Voice and Mix contained exactly the same acoustic realizations of the words with constant spelling, and that in List Slightly Voiced these words had been realized with more final voicing. Hence, the listeners' reactions to the completely voiceless realizations of these words do not provide evidence for a list effect that is due to the listeners. Their interpretations for the completely voiceless realizations seem to be based mainly on the acoustic signal for these realizations themselves.

Finally, we turn to the words ending in velar fricatives. They show the same pattern over the three lists as the words with constant spelling. A linear model excluding the words ending in sonorant segments plus plosives shows no interaction between Type of word and List. Note, however, that the number of words ending in velar fricatives is relatively small, and the absence of a significant difference between the words ending in velar fricatives and the words with constant spelling is therefore not very informative.

In conclusion, the responses to the three lists indicate that slightly voiced obstruents are interpreted as underlyingly voiced more often when the list contains fewer slightly voiced obstruents. In other words, slight voicing is considered to be more informative if it is not the norm.

5. General Discussion

This study investigates whether listeners take advantage of incomplete neutralization even when they have no intrinsic reasons for doing so. In Experiment I, a speaker recorded pseudo words, and we analyzed various acoustic characteristics of the realizations that may be relevant to the perception of final voicing. In Experiments II and III, we presented the realizations to listeners, and we investigated whether listeners take advantage of the incomplete neutralization in the acoustic signal also if they are performing a task for which this is not necessary.

Our speaker recorded two lists containing the same pseudo verbs ending in obstruents. The lists differed mainly in that nearly all words were spelled with voiceless final graphemes in List –Voice, while the words ending in sonorant segments plus plosives were spelled with voiced final graphemes in List Slightly Voiced. Although our speaker was not aware of the fact that the words in the two lists formed minimal and identical word pairs, she produced the plosives represented as voiced with shorter release noises than the plosives represented as voiceless. The experiment therefore provides additional evidence that incomplete neutralization in Dutch may be evidenced by the release noise. In addition, it shows that incomplete neutralization may be induced by orthography. Finally, this speaker realizes words with incomplete neutralization, even when she does not know that she is realizing minimal word pairs. Apparently, minimal pairs are not necessary for speakers to realize words with incomplete neutralization.

A large number of words were spelled identically in the two lists. These words with constant spelling were realized with longer periods of vocal fold vibration in List Slightly Voiced than in List –Voice. This shows that our speaker varied the voicing of the final obstruents as a function of the voicing of the other final obstruents in the list. This speaker related transfer effect probably results from the automation of the articulatory gestures resulting in slightly voiced or completely voiceless obstruents.

In Experiment II, we presented the lists to two different groups of listeners. These listeners were asked to create the past-tense forms for the pseudo verbs, for which they had to interpret the final obstruents as underlyingly voiced or voiceless. Previous research (Ernestus and Baayen 2003) has shown that participants base their interpretation at least partly on the existing words belonging to the same phonological gangs as the pseudo verbs. Thus, participants do not need to take into account the subphonemic cues in the acoustic signal to underlying voicing in order to perform this task. Both groups of participants indeed based their interpretation of the final obstruents on the phonological gangs. In addition, listeners presented with slightly voiced obstruents more often opted for voiced interpretations. This shows that listeners take advantage of incomplete neutralization not only in spelling tasks that they can only perform by interpreting the incomplete neutralization. They also take advantage of incomplete neutralization when there are no intrinsic reasons for them to do so. In Dutch, incomplete neutralization emerges as a subphonemic cue for past-tense formation.

In Experiment II, listeners heard no (List –Voice) or many (List Slightly Voiced) slightly voiced obstruents. We expected that the high percentage of slightly voiced obstruents in List Slightly Voiced would attenuate the effect of slight voicing. Listeners interpret subtle characteristics of the acoustic signal by reference to the surrounding segments in the signal. If slight voicing is the norm, listeners might be less willing to base their interpretation of the final obstruent on slight voicing. In Experiment III, a new group of participants listened to a new list, List Mix, which is identical to List Slightly Voiced, except that the words with constant spelling were taken from List –Voice. The only slightly voiced obstruents in List Mix are therefore the plosives following sonorants. These plosives elicited more *de* responses in List Mix than in List Slightly Voiced. This shows that slightly voiced plosives are indeed interpreted as underlyingly voiced more often when there are fewer other slightly voiced obstruents in the list. This listener related transfer effect probably reflects the way in which speakers dynamically adapt to their conversation partner's speech habits or dialect.

The final obstruents of the words with constant spelling were interpreted as underlyingly voiced more often in List Slightly Voiced than in both List –Voice and List Mix, which is in accordance with their acoustic characteristics in these lists. The absence of a difference between List –Voice and List Mix suggests that listeners interpret completely voiceless obstruents directly on the basis of the acoustic signal without much reference to context.

Our finding that spelling induces incomplete neutralization does not mean that incomplete neutralization is simply an artefact of orthography. Incomplete neutralization has also been reported for Catalan minimal word pairs that do not reflect the difference in the underlying [voice]-

specification of the final obstruents in their spelling (Dinnsen and Charles-Luce 1984). In addition, our finding that listeners take advantage of incomplete neutralization even when different interpretations do not map onto different spellings (that is, for the words with constant spelling) provides further evidence that incomplete neutralization is not a simple artefact of orthography.

One way to incorporate incomplete neutralization into the grammar is to implement it by means of low-level phonetic rules. Speakers would realize final voiced graphemes as slightly voiced, because seeing a voiced grapheme would activate the low-level phonetic implementation rules for slight final voicing. Listeners would be affected by incomplete neutralization when interpreting a pseudo word, because incomplete neutralization would help them to build an underlying representation. When incomplete neutralization is considered as a result of phonetic implementation rules, the standard view of grammar has to be revised. Final Devoicing is a phonological rule, and, under the standard view, it would consequently apply before phonetic incomplete neutralization. As a consequence, phonetic incomplete neutralization processes would be unable to refer to the underlying distinction between voiced and voiceless obstruents, and cannot be applied. For instance, a word like /vɛrʋɛid/ (see 1) would be transformed into /vɛrʋɛit/, before it is passed to the phonetic component. The phonetic component, consequently, has no access to the underlying /d/, and we cannot explain why /vɛrʋɛid/ may be realized with slight final voicing. Several solutions have been suggested. For instance, Dinnsen and Charles-Luce (1984) and Slowiaczek and Dinnsen (1985) suggest that phonetic implementation rules may apply before phonological rules, and Port and O'Dell (1985) suggest that Final Devoicing and incomplete neutralization together form one implementation rule.

We would like to offer another solution. The mental lexicon contains auditory and visual form representations for a great many words, including inflected words (e.g., Baayen, Dijkstra, and Schreuder 1997; Alegre and Gordon 1999; Sandra, Frisson, and Daems 1999; Bertram, Schreuder, and Baayen 2000; Bybee 2001; Baayen et al. 2002; Baayen et al. 2003). Thus, with respect to the stems *verwijd* and *verwijt*, the lexicon of a Dutch speaker probably contains representations such as those listed in (2a) and (2b), respectively. The lexical storage of inflectional forms makes abstract underlying forms superfluous. The form *verwijd* needs not be stored as /vɛrʋɛid/, but can be stored as /vɛrʋɛit/, directly reflecting its pronunciation. The fact that this stem has a final [d] in the infinitive and plural forms is accounted for by the representations of these forms themselves.

(2)	a.	[vɛrʋɛit]	*verwijd*	‘widen’
		[vɛrʋɛidən]	*verwijden*	‘to widen’
		[vɛrʋɛidən]	*verwijden,*	‘widen’ (plural)
		etc.		
	b.	[vɛrʋɛit]	*verwijt*	‘reproach’
		[vɛrʋɛitən]	*verwijten*	‘to reproach’
		[vɛrʋɛitən]	*verwijten,*	‘reproach’ (plural)
		etc.		

Given lexical representations that directly reflect pronunciation, the production and interpretation of incomplete neutralization may well be primarily a lexical effect, in two ways. First, we think that the lexical representations of words contain information concerning the probability that the final obstruent is realized as slightly voiced (see, e.g., Bybee 2001 and Pierrehumbert 2003, who also assume that variability is represented in the lexical representations). The stochastic representation of the final plosive has a shorter expected release noise duration in *verwijd* than in *verwijt*. Both speakers and listeners might well make use of this stochastic information.

Independently of this possibility, we also think that incomplete neutralization is affected by lexical analogy. Lexical analogy can explain the production of existing words with incomplete neutralization. When speakers have to realize *verwijd*, they activate [vɛrʋɛit] as well as the words inflectionally related to *verwijd*, such as [vɛrʋɛidən] (see also Bybee 2001). These inflectionally related words contain [d], and they support the realization of the final obstruent of *verwijd* as [d]. This may enhance the probability that the final obstruent is realized as slightly voiced.

The stored inflectionally related word forms probably also affect the listener’s interpretation of incomplete neutralization for existing words. When listeners perceive a word, the auditory representations that best match the word are activated, including nearest neighbors such as inflectional variants (e.g., Bybee 1985). The activation levels of these neighbors are co-determined by the size of the mismatch between the acoustic signal and the auditory representations of the words. A neighbor with a voiced obstruent is more activated by a word realized with a slightly voiced obstruent than by a word realized with a completely voiceless obstruent. The final activation levels of the neighbors co-determine the interpretation of the acoustic signal. Thus, upon hearing [vɛrʋɛit], the words in (2) all become activated to some degree. If the final obstruent of [vɛrʋɛit] is slightly voiced, the inflectionally related words of *verwijd* are co-activated more than if this obstruent is completely voiceless. Hence, a slightly voiced realization makes the interpretation of the word as 'widen' more likely. If, conversely, the final obstruent is realized as completely voiceless, the meaning 'reproach' becomes more probable.

Lexical analogy can also explain the production and interpretation of incomplete neutralization for pseudo words, studied in this paper. If a speaker reads a pseudo word, the representations of existing words with similar spellings are probably activated. These orthographic neighbors may affect the speaker's realization of the pseudo word, with each activated neighbor affecting especially that part of the pseudo word to which it is similar. Thus, the reading of *dijp* may activate the Dutch words *dal* and *rijp*, and the phonetic representation of *dal* may affect especially the realization of the *d* of the pseudo word, while the phonetic representation of *rijp* affects the rime of *dijp*. If the final obstruent of the pseudo word is represented as voiced, the orthographic neighbors ending in the same orthographic rime tend to be realized with slight final voicing. As a consequence, speakers may realize also the pseudo word with slight final voicing. If, in contrast, the final obstruent is represented as voiceless, the activated orthographic neighbors ending in the same rime tend to be realized without slight voicing, and speakers realize the final obstruent of the pseudo word as completely voiceless.

As to the interpretation of incomplete neutralization in pseudo words, when a listener hears a pseudo verb with slight final voicing, especially the nearest phonological neighbors with similar subphonemic characteristics are co-activated. These neighbors tend to have inflectional paradigms supporting *de* as allomorph rather than *te*. Hence, the likelihood that the listener selects *de* increases compared to the case when the listener hears a pseudo verb without slight voicing.

We have phrased our interpretation using the metaphors of localist spreading activation modeling. We believe, however, that our data can be incorporated straightforwardly in non-localist theories based on statistical language models or non-localist artificial neural networks, if so desired. Also these approaches can easily account for the fact that incomplete neutralization is part and parcel of the grammar of Dutch.

Notes

* We would like to thank Mirjam Broersma, Taehong Cho, Anne Cutler, Delphine Dahan, Carlos Gussenhoven, Terrance Neary, and Tania Zamuner for stimulating discussion and comments, and Doug Whalen and two anonymous reviewers for their constructive criticism.

1. Interestingly, Jongman et al. (1992) and Baumann (1995) failed to find incomplete neutralization in Dutch. Jongman et al. (1992) investigated ten minimal word pairs, while Baumann (1995) restricted her study to twelve word pairs. Possibly these studies did not detect incomplete neutralization because the effects are too subtle in Dutch to surface in a small number of items, as is also suggested by Warner et al. (2004).

References

Alegre, Maria and Peter Gordon
1999 Frequency effects and the representational status of regular inflections. *Journal of Memory and Language* 40: 41–61.

Baayen, R. Harald, Ton Dijkstra and Rob Schreuder
1997 Singulars and plurals in Dutch: Evidence for a parallel dual route model. *Journal of Memory and Language* 37: 94–117.

Baayen, R. Harald, James McQueen, Ton Dijkstra and Rob Schreuder
2003 Frequency effects in regular inflectional morphology: Revisiting Dutch plurals. In: R. Harald Baayen and Rob Schreuder (eds.), *Morphological structure in language processing*, 355–390. Berlin: Mouton de Gruyter.

Baayen, R. Harald, Richard Piepenbrock, and Leon Gulikers
1995 *The CELEX lexical database (CD-ROM)*. University of Pennsylvania, Philadelphia, PA: Linguistic Data Consortium.

Baayen, R. Harald, Rob Schreuder, Nivja H. De Jong and Andrea Krott
2002 Dutch inflection: the rules that prove the exception. In: Sieb Nooteboom, Fred Weerman and Fred Wijnen (eds.), *Storage and Computation in the Language Faculty*, 61–92. Dordrecht: Kluwer Academic Publishers.

Baumann, Monika
1995 The production of syllables in connected speech. Ph.D. dissertation, Max-Planck Institute for Psycholinguistics, Nijmegen.

Bertram, Raymond, Rob Schreuder and R. Harald Baayen
2000 The balance of storage and computation in morphological processing: The role of word formation type, affixal homonymy, and productivity. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 26: 419–511.

Geert E. Booij
1995 *The phonology of Dutch*. Oxford: Clarendon Press.

Breiman, Leo, Jerome H. Friedman, Richard A. Olshen and Charles J. Stone
1984 *Classification and Regression Trees*. Belmont, California: Wadsworth International Group.

Bybee, Joan L.
1985 *Morphology: A study of the relation between meaning and form*. Amsterdam: Benjamins.
2001 *Phonology and language use*. Cambridge: Cambridge University Press.
Dinnsen, Daniel A. and Jan Charles-Luce
1984 Phonological neutralization, phonetic implementation and individual differences. *Journal of Phonetics* 12: 49–60.
Ernestus, Mirjam and R. Harald Baayen
2003 Predicting the unpredictable: Interpreting neutralized segments in Dutch. *Language* 79(1): 5–38.
in press Intraparadigmatic effects on the preception of voice. In Erik Jan Van der Torre and Jeroen Van de Weijer (eds.), *Voicing in Dutch*. Amsterdam: John Benjamins.
Fourakis, Marios and Gregory K. Iverson
1984 On the 'incomplete neutralization' of German final obstruents. *Phonetica* 41: 140–149.
Jongman, Allard, Joan A. Sereno, Marianne Raaijmakers and Aditi Lahiri
1992 The phonological representation of [voice] in speech perception. *Language and Speech* 35: 137– 152.
Kemps, Rachèl, Mirjam Ernestus, Rob Schreuder and R. Harald Baayen
2005 Prosodic cues for morphological complexity: The case of Dutch plural nouns. *Memory and Cognition* 33: 430–446.
Kenstowicz, Michael
1994 *Phonology in generative grammar*. Cambridge MA & Oxford UK: Blackwell.
Nooteboom, Sieb G.
1979 Perceptual adjustment to speech rate: A case of backward perceptual normalization. In: H. V. Deighton Van Witsen (ed.), *Anniversaries in phonetics: Studia gratulatoria dedicated to Hendrik Mol*, 276–287. Amsterdam: Institute of Phonetic Sciences Amsterdam.
Nooteboom, Sieb G. and G. J. N. Doodeman
1980 A note on speech rate. *IPO Annual Progress report*: 58–60.
Pierrehumbert, Janet B.
2003 Probabilistic phonology: Discrimination and robustness. In: Rens Bod, Jennifer Hay and Stefanie Jannedy (eds.), *Probability theory in linguistics*, 177–228. Cambridge, MA: The MIT Press.
Pinheiro, José C. and Douglas M. Bates
2000 *Mixed effects models in S and S PLUS*. (Statistics and Computing.) New York: Springer.

Port, Robert and Penny Crawford
1989 Incomplete neutralization and pragmatics in German. *Journal of Phonetics* 17: 257–282.
Port, Robert F. and Michael L. O'Dell
1985 Neutralization of syllable-final voicing in German. *Journal of Phonetics* 13: 455–471.
Rietveld, Toni, Joop Kerkhoffs, and Carlos Gussenhoven
1999 Prosodic structure and vowel duration in Dutch. In: John Ohala, Yoko Hasegawa, Manjari Ohala, Daniel Granville and Ashlee C. Baily (eds.), *Proceedings of the 14th International Congress of Phonetic Sciences, San Francisco, 1–7 August 1999*, 463–466. Berkeley: Linguistic Department, University of California.
Sandra, Dominiek, Steven Frisson and Frans Daems
1999 Why simple verb forms can be so diffcult to spell: The influence of homophone frequency and distance in Dutch. *Brain and Language* 68 (1/2): 277–283.
Slis, I H. and A. Cohen
1969 On the complex regulating the voiced-voiceless distinction ii. *Language and Speech* 12: 80–102; 137–155.
Slowiaczek, Louisa A. and Daniel A. Dinnsen
1985 On the neutralizing status of Polish word-final devoicing. *Journal of Phonetics* 13: 325–341.
Van den Berg, R. J. H.
1988 The perception of voicing in Dutch two-obstruent sequences. Ph.D. dissertation, Catholic University Nijmegen, Nijmegen.
Warner, Natasha, Allard Jongman, Joan Sereno and Rachèl Kemps
2004 Incomplete neutralization and other sub-phonemic durational differences in production and perception: evidence from Dutch. *Journal of Phonetics* 32: 251–276.

Dynamics in grammar: Comment on Ladd and Ernestus & Baayen*

Adamantios I. Gafos

The derivational view of phonetics-phonology (Ladd, this volume) expresses an intuition that seems valid, namely, that there is a distinction to be made between quantitative and qualitative aspects of phonetics-phonology. Incomplete neutralization (Ernestus and Baayen, this volume) and other phenomena like it indicate that the specific way of drawing that distinction is too rigid. At the same time, these phenomena underscore the need for a different formal language, where discrete and continuous aspects of phonetics-phonology can interact. A way of reconciling the core intuition of the derivational view with phenomena like incomplete neutralization is proposed using the mathematics of nonlinear dynamics. This allows one to integrate the continuous and the discrete without the additional postulate that phonology is derivationally antecedent to phonetics.

1. Two views of phonetics-phonology

How are the qualitative aspects of phonological competence related to their variable and continuous phonetic manifestation? This question defines the so-called 'phonetics-phonology problem' and it has been one of the central themes of laboratory phonology (Beckman and Kingston 1990: 1). It is also an instance of a broader question in cognitive science, namely, the question of how to relate the low dimensional, discrete aspects of cognition to the high dimensional aspects of performance, as shown by parallel research in vision (Haken 1990), coordination in action (Turvey 1990), agent-environment interaction (Beer 1995) and other domains.

There are two broad views on the formalization of theories aiming to address this central question. One view, firmly established with the development of generative phonology (Chomsky and Halle 1968) and subsequently elaborated and refined in important ways (Liberman and Pierrehumbert 1984;

Keating 1988, 1990; Cohn 1990; Coleman 1992), posits that the relation between qualitative and quantitative aspects of phonetics-phonology consists of a process of translation from discrete symbols to continuous physical properties of an articulatory and acoustic nature. In Ladd's words, "we need to think of phonetic realization as a mapping between a categorical symbolic representation and a quantitative physical signal" (Ladd, this volume). This is the view in the background of most current work in phonetics-phonology and cognitive science in general, e.g., see the notion of transducer in Fodor and Pylyshyn (1981) and also Harnad (1990).

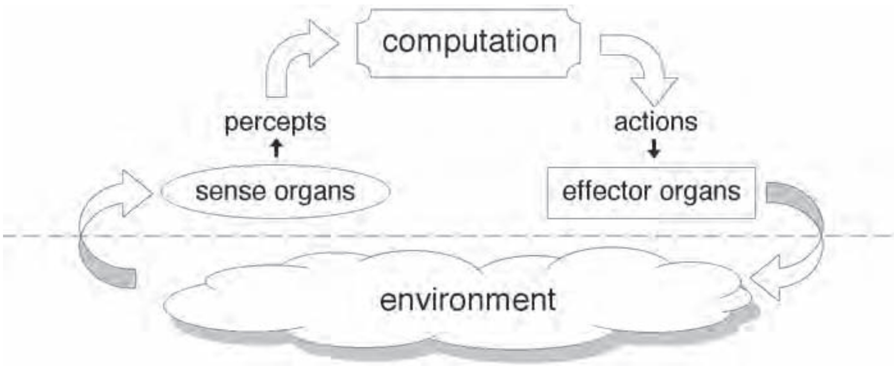
An alternative, relatively more recent and less widely explored view builds on the mathematics that can express both the discrete and the continuous aspects of complex systems, the so-called nonlinear dynamics (see Smolensky 1988 and Port and van Gelder 1995 for a proposal and a sample of applications of the dynamical view in cognitive science, respectively). In phonetics-phonology, a precedent is Browman and Goldstein's (1986 et seq.) research program. An important contribution emerging from Browman and Goldstein's work is an explicit theory of dynamically defined phonological representations. Roughly speaking, this theory implies that the atoms of phonological representations must be construed as unfolding in time (gestures) and that universal as well as language-particular principles may refer to this temporal dimension of phonological form (Browman and Goldstein 1986, 1995; Gafos 2002).

The goal in this paper is to broaden the argument for the dynamical view by focusing on a special case of the fundamental question here, the subtle context-dependency of phonological neutralization. As I discuss, in certain well-documented cases of phonological neutralization, grammatical requirements interact with variable environmental conditions (here, speakers intentions to convey contrasts). This turns out to be a problem for the derivational view of phonetics-phonology. The specific aim is to show that a dynamics model predicts this context dependency of neutralization, an aspect of the problem that has remained outside the scope of previous models.

2. The problem: final devoicing

To state the problem in most general terms, it is useful to review the three main components of cognition: perception-computation-action, as shown in (1) (Cariani 1989). For example, in a task where a listener is asked to produce the plural of a spoken word, the perceptual system identifies the singular form, say, the percept [glik], the grammar computes the plural form [gliks], and finally the output computed by the grammar is implemented as vocal-tract action.

(1) Main components of biological cognitive agents: perception, computation, production



A fundamental fact left out from this description of the perception-computation-action loop is that the cognitive system is embedded in a continuously varying environment. Moreover, all three components of the system have the remarkable capacity to deal with various sources of variability in that environment.

Consider two prototypical examples from production and perception. It is well known that the timing characteristics subserving various segmental contrasts are dependent on speech rate. For example, Summerfield (1981) shows that the VOT boundary (onset of voicing relative to oral release) between voiceless and voiced consonants changes as a function of speech rate. As rate increases, speakers' productions of voiceless and voiced consonants shift towards shorter values of VOT. In turn, listeners are sensitive to such variations, and adapt flexibly to different rate conditions. Another example is illustrated with the durational boundary between single and double consonants in examples like "topic" and "top pick". This boundary is not invariant but depends on the rate of the utterance these tokens are part of. The faster the rate, the lower the boundary value. Listeners are sensitive to this rate-dependent change in the signal. A given silence duration is judged differently depending on the rate of the utterance. Similar results hold for the distinction between /s/-/ss/ in Japanese, a language with distinctive consonant length (see Miller 1981 for a review).

So far then we see that production and perception are stable in that varying some external parameter leaves the qualitative nature of the system, the distinct categories, unaltered. These systems are also flexible, because they adapt to varying environmental requirements, such as speaking fast or slow.

Next, consider an example from the cross-linguistically common phenomenon of final devoicing. The phonological description of final devoicing or neutralization is simple. In certain languages, obstruents are voiceless syllable-finally (Bloomfield 1933: 218; Trubetzkoy 1969: 213). See (2) for representative examples from German and Ernestus and Baayen (this volume) for Dutch.

(2) *Rad* 'wheel' [ʁat] (nom.) [ʁadəs] (gen.)
inferred underlying form = /ʁad/
Rat 'advice' [ʁat] (nom.) [ʁats] (gen.)
inferred underlying form = /ʁat/
Bund 'association' [bunt] (nom.) [bundəs] (gen.)
inferred underlying form = /bund/
bunt 'colorful' [bunt] (sing.) [buntə] (pl.)
inferred underlying form = /bunt/

The situation is more complex in the phonetics of neutralization. There are two main results. First, neutralization is incomplete in that the [t] in [bunt] 'association' is not identical to the [t] in [bunt] 'colorful'. Even though both are transcribed as [t], the mean of the variable indexing voicelessness differs between the [–Voiced] and the (surface realizations of underlying) [+Voiced] consonants. The latter's mean is slightly shifted toward less extreme values of devoicing or toward more "slight voicing" in Ernestus and Baayen's (this volume) terms. Specifically, differences can be observed in the duration of the preceding vowel, in the duration of consonantal closure and glottal pulsing during that closure, and in the duration of the burst associated with consonant release. See, among others, Dinnsen (1985) for a review of other instances of incomplete neutralization, Dinnsen and Carles-Luce (1984) on Catalan final devoicing, Fougeron and Steriade (1993) on French schwa elision, and Charles-Luce (1993, 1997) on Catalan voicing assimilation and English flapping.[1]

Second, neutralization shows a subtle dependency on the communicative context. This can be illustrated with the following task, from Port and Crawford (1989). In one experimental condition, speakers are asked to read a list of words in isolation. In another condition, speakers are asked to read sentences like *Ich habe Rat(Rad) gesagt; nicht Rad(Rat)* ("I said Rat(/Rad) not Rad(/Rat)") while a German assistant, who is present in the experimental setting, is assigned the task of writing down the order of the test words in such sentences. In this second condition, then, speakers are encouraged by the context to convey the contrast more than in the word list reading condition. The observed result is a stronger version of incomplete neutralization than in the word list reading condition (where no assistant is present). This is to say that the means of the variables indexing voicing shift even more toward less extreme values of devoicing for the underlying [+Voiced] consonants (Port and Crawford 1989, see also Charles-Luce 1985).

The incompleteness of final devoicing and its systematic dependence on context are characteristic of the flexibility and stability of the phonetics-phonology system. On the one hand, there is a consistently reproducible aspect of the phonetics-phonology of German, identified with final devoicing (stability). On the other hand, the phonetics-phonology system is flexible in allowing speaker's intentions to shift the phonetic output in ways that deviate slightly from the ideal grammatical optimum (flexibility).

Consider how the derivational view of phonetics-phonology deals with stability and flexibility, in general. The symbolic constructs of phonology are by definition stable – they are mental realities abstracted from the environment (axiomatic stability). The grammar is stable because its essential constructs are symbolic in nature. Flexibility enters the life of the phonetics-phonology system in phonetic implementation, after the grammar has computed an output or 'interface representation' in Ladd's terms. In phonetic implementation, symbolic units are translated to vocal tract action under different conditions – different speech rates, styles, social contexts, etc. – and environmental variables begin to introduce their effects.

However, the incompleteness of neutralization does not fit comfortably in this view. This is illustrated in (3). Final Devoicing changes the voicing value of the final obstruent in /bund/ to [–Voiced]. This eliminates the contrast between the final consonants of /bund/, /bunt/ at the output of phonology, exactly as a 'neutralization' rule should do. Consequently, phonetic implementation, whose role is to flesh out phonology's output as vocal-tract action, is now unable to deliver the differences observed in the surface realizations of the final obstruents in /bunt/ versus /bund/.

(3) Rule of Final Devoicing, FD:
[+Voiced, –Sonorant] → [–Voiced] / $\_\_]^{\sigma}$

Underlying forms		Output of phonology		Vocal-tract action
/bunt/	→	[bunt]	►	[t], completely voiceless
/bund/	→	[bunt] (via FD)	►?	[t^+], traces of voicing

It is clear that incomplete neutralization requires some revision of the standard phonology-phonetics view. Accordingly, the incompleteness of final devoicing has led to arguments for relaxing one of the foundational assumptions of that view, the ordering of phonology before phonetic implementation. See Dinnsen and Charles-Luce (1984: 58) for Catalan, and Slowiaczek and Dinnsen (1985: 338) for Polish.

Another approach is to apply Final Devoicing at the same time as phonetic implementation (Ernestus and Baayen, this volume; Port and O'Dell 1985). As Ernestus and Baayen observe, the main problem with this proposal is that phonology becomes indistinguishable from phonetic implementation. Final devoicing is an aspect of German phonology. By moving it to phonetic implementation, final devoicing must be reformulated in a different formal language, the language of phonetic implementation, using continuous mathematics. The proposal to be fleshed out here begins with the challenge of maintaining the distinction between qualitative versus quantitative aspects of phonetics-phonology by proposing an appropriate formalization of the phonetics-phonology relation.

The second, equally important characteristic of incomplete neutralization is its systematic dependence on the communicative context. To date, I am not aware of any previous formal treatment of this effect. This phenomenon is an example of what Liberman refers to as phonological systematicities which are "modulated by … paralinguistic parameters" and which are "not well modeled as feature- or structure-changing rules" (1983: 271). The grammar output is quantitatively shifted by speakers' intentions to convey a contrast, but intentions are not the kinds of primitives that are described as being part of the grammar – they are extra-grammatical or para-linguistic.

The challenge for the derivational view of phonetics-phonology is that, on the one hand, placing final devoicing in the phonology captures the fact that final devoicing is a qualitative property of German, but it cannot account for the flexibility of the phonetics-phonology system. On the other hand, moving final devoicing to phonetic implementation would allow it to be modulated by extra-grammatical, continuous factors but loses sight of the fact that final devoicing is an aspect of German phonology.

The alternative to be proposed here is a non-derivational (parallel) way of relating discrete aspects of the grammar and continuous, environmental variables. This promises to bypass the ordering problem, under the assumption that there is a coherent way to make continuity and discreteness coexist within the same formal language, and also that there is a way to at least describe and at best derive phenomena like incomplete neutralization. The mathematics of nonlinear dynamics satisfies the first assumption, as discussed in the next section. Subsequent sections take up the issue of deriving incomplete neutralization, using basic concepts of nonlinear dynamics.

Before leaving this section, I consider whether phonological models dealing with variability can be of help with the problem faced here. In a rule-based model (Chomsky and Halle 1968), we may consider ‘variable rules’ as in Sankoff (1988) or Cedergren and Sankoff (1974). In Optimality Theory (Prince and Smolensky 1993, henceforth OT), we may consider the ‘stochastic evaluation’ method for constraint interaction as proposed in Boersma and Hayes (2000). To illustrate, in the latter model, if two constraints are sufficiently close on a rank scale, a small shift in their rank values can result in C1 >> C2 or C2 >> C1. In the specific example, the constraints are C1 = NoVoicedCoda, C2 = Faith(Voice). Their variable ranking would give rise to underlyingly voiced obstruents being produced sometimes voiceless (when C1 >> C2) and sometimes voiced (when C2 >> C1).

These models deal with a different type of variation from that addressed in this paper. They deal with variation among *discrete* alternatives. In the present case, however, it is not that the voiced obstruent is produced sometimes voiced and sometimes voiceless. Rather, the mean value of voicelessness drifts toward less extreme values, and it does so lawfully as a function of the communicative context. Hence, those models are inapplicable to this type of variation, which I will call lawful continuous variation.

There is, however, another class of models with the capacity of handling continuous dimensions, the so-called exemplar models of memory and categorization (Hintzman 1986). Recently, Pierrehumbert (2001, 2002) has developed an application of the exemplar paradigm to phonetics-phonology, with attention to variation and fine phonetic details in the realization of phonological categories. Specifically, in that application, variation in production is achieved by averaging and/or randomization over a set of memorized exemplars of a category, generating a so-called ‘echo’ of the category. The crucial observation here is that the variation involved in final devoicing has a systematic component, as changes in environmental variables result in systematic gradual drifts toward more or less voicing. This context dependency is not accounted for by an averaging and/or randomization method, as in fact noted in Pierrehumbert (2002).

3. Phonetics-phonology in a dynamical setting

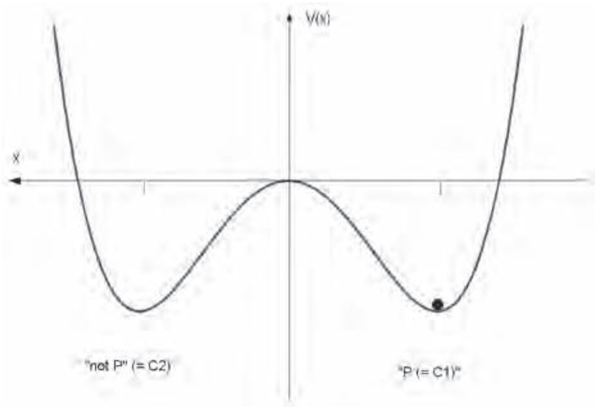
To develop a parallel view of phonetics-phonology, the essential insights of the field must be recast using the mathematics of nonlinear dynamics. Thus, phonetic categories, representations, constraints, and grammars must be given a dynamical formulation. For phonetic categories and representations, some of the foundational work in these domains has been couched in terms that are at least consistent with the dynamical approach. See Stevens (1972, 1989), Petitot-Cocorda (1985), Kingston and Diehl (1994), and references in section 1 on Browman and Goldstein's work. In this section, I focus on constraints and grammars. To anticipate, my specific proposal is that constraints are *attractors* and that grammars are *attractor landscapes*. Both notions are basic to nonlinear dynamics.

To begin, phonological constraints are formulated as competing *attractors* (Thompson and Stewart 2002: 45). Attractors define preferred modes for the macroscopic parameters of phonology. For example, constraints like "BE CORONAL" and "BE VOICELESS" state preferred values for the phonological parameters of Place of articulation and Voicing. In (4), two competing constraints C1, C2 are depicted as two attractors; attractor 1: 'have property P', attractor 2: 'not P'. Taking Voicing as an example, the system can be in two states. Either it is "Voiceless", it has property P, or it is "Voiced", it does not have property P. For illustration purposes, let us index the degree of voicing with the parameter of glottal aperture.[2] Then, the "Voiceless" state is represented with the minimum at some positive value of glottal opening and the "Voiced" state with the minimum at the some negative value of glottal opening (the actual numeric values and their signs are not crucial in the present context).

The figure in (4) represents the assumption that, in a language with a Voiceless/Voiced contrast, the Voicing parameter draws values from two recognizably distinct parts of its state space (the state space is the entire x axis). It thus describes qualitatively distinct modes of the voicing system or, in other words, it describes a dimension of macroscopic order in phonological form. For this reason, it is called an *order parameter* (Haken 1977).

Intuitively, we may interpret the behavior of an order parameter by means of a ball moving in the potential V(x) shown above. Clearly, the ball ends up in one of the two attractors, the macroscopic observables of the system. The attractor landscape shown there is known as the 'anharmonic oscillator' and it is described by the potential function $V(x) = (-1/2)*(x^2) + (1/4)*(x^4)$).

(4) Phonological constraints as competing *attractors*

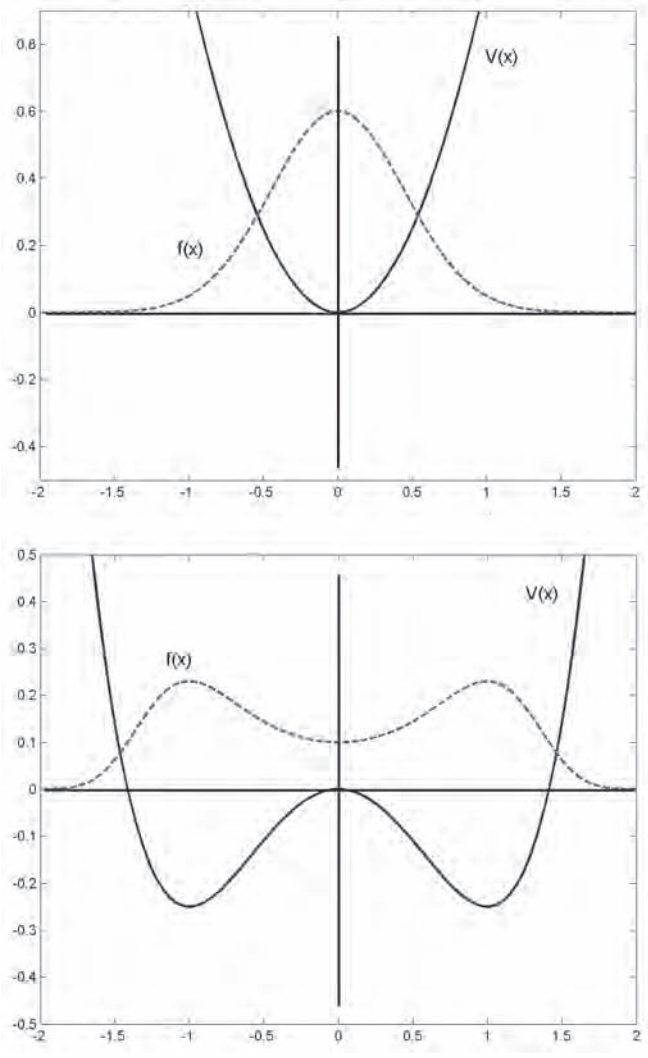


Given that macroscopic order is expressed via order parameters and constraints referring to these, what is the relation between these parameters and traditional symbols? Specifically, what is the symbol [+/–Voiced] in the dynamical formulation of the voicing distinction? In the dynamical formulation, the symbol is inseparably linked with its phonetic substance. It is not derivationally antecedent to that substance and therefore it does not need to be translated to that substance. Eco, who has studied the foundational notion of symbol closely, writes: "One cannot speak of a form without presupposing a matter and linking it immediately (neither before nor after) to substance" (1984: 23).

Next, how is the stability of macroscopic order achieved in a dynamical formulation of phonetics-phonology? Attractive modes are *dynamically stable*, that is, they exhibit small fluctuations around their mean states (the two minima shown above). Fluctuations are inevitable due to noise. Noise is inevitable because complex systems described by low-dimensional dynamics are coupled to various subsystems at a more microscopic level. In our case, the control of voicing, the microscopic level corresponds to the neuronal, aerodynamic and myodynamic subsystems (see Titze 1988).

Following Haken (1977), I describe noise as a small, random perturbation force pushing the representative point of the system x, the position of the ball, back and forth randomly. Randomness introduces stochasticity and consequently we can only compute the *probability* for finding x within a given interval of values of x. This probability is described by the probability distribution function f(x) multiplied by the length of the interval. Two probability distribution functions corresponding to two different potentials are shown in (5). The potential to the left is monostable, that is, it has a single attractor, and the one to the right is bistable.

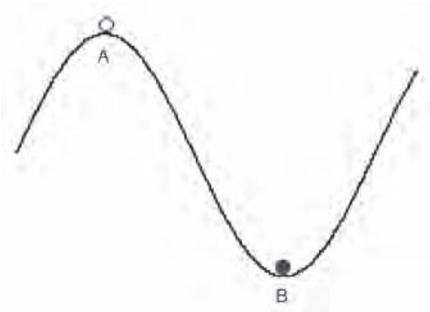
(5) V(x) and probability distribution function f(x) for two potentials



It can be seen that the probability to find the system around the mean state(s) of the attractor(s) is quite high. The probability to find the system at some other point decreases quickly as we move away from the mean states but it may not be zero. In short, the preferred modes of order parameters, the attractors, are resistant to noise in a probabilistic sense.

Noise is inherent to the process of modeling a phenomenon in dynamical terms and it can be used to generate predictions. Specifically, noise has a differential effect on the order parameters depending on the strength of the attractor. To illustrate, imagine the ball in the well of a strong attractor. As a classic example, consider the ball at point B below. Here, noise has a small effect in causing minute perturbations around the mean state.

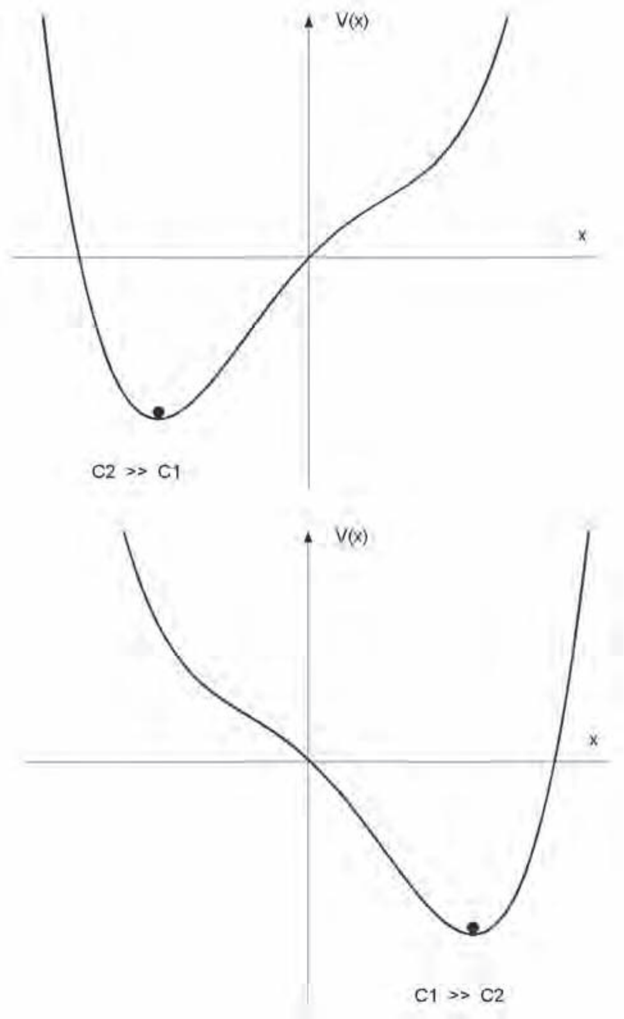
(6) Unstable (A) and stable (B) equilibria



Now, imagine what happens when the ball is put at point A. Due to random fluctuations, the ball ends up falling at the left or the right side. A is an unstable point. This illustrates that fluctuations can have dramatic effects at highly unstable regions of the state space (see Benus, Gafos and Goldstein, to appear, for an application of this to modeling suffixal variation in Hungarian vowel harmony). In dynamics, then, it is possible to exploit noise to discover the stable attractors of the system. The consequences of noise can be measured by the variance or standard deviation of some essential variable x around the attractive state. The more stable the attractor the smaller the deviation from the attractive state.

I now turn to a fundamental insight on grammars, namely, the idea that the qualitative aspects of linguistic form are the result of constraint optimization, and specifically the notion of constraint ranking. Both of these derive from OT (Prince and Smolensky 1993). In the proposed model, constraint ranking is modeled as reorganization of the attractor landscape. This is illustrated in the figures below, which show two qualitatively different reorganizations of the attractor landscape in (4).

(7) Constraint ranking as reorganization of the attractor landscape – compare with 4



Such dramatic reorganization in landscapes can be effected formally by changes in so-called ‘control’ parameters that enter the mathematical model underlying the phenomenon of interest. In the examples illustrated below, these changes are brought about by adjusting the control parameter k in the potential function V(x) = k*x + (–1/2)*(x^2) + (1/4)*(x^4), which determines the tilt and direction of the potential (see Tuller et al. 1994 for an application to perception). Thus in (7), the potential to the right representing the C1 >> C2 ranking corresponds to k = 1, and the potential to the left representing the C2 >> C1 ranking corresponds to k = –1. The potential in (4) corresponds to k = 0, where the two constraints are unranked. It can be seen then that the grammar shift from C1 >> C2 to C2 >> C1 or vice versa implies an intermediate stage where the two constraints are unranked (since to go from 1 to –1, k must pass through 0). Thus, grammar change necessarily IMPLIES an intermediate stage of variation. This corollary of the dynamical formulation of constraint re-ranking seems consistent with the course of sound change (e.g., Lass 1997: 287ff., Sommerstein 1977: 250–251). Moreover, as in the stochastic Optimality Theory model in Boersma and Hayes (2000), it is possible to model fine, probabilistic variation in constraint ranking by smoothly varying the control parameter k. As k modifies the attractor landscape, the probability distribution function over that landscape changes accordingly, thereby modulating the probabilities of the different states the system may reside in (recall the discussion around 5). However, I cannot illustrate in detail these consequences of the proposed dynamic model for constraint ranking here.

In what follows, I briefly discuss one definitional property of nonlinear dynamics that is of critical importance in modeling complex systems in general and phonetics-phonology in particular. This is the property is non-linearity. A system exhibits non-linearity when large or discontinuous changes can be observed in the behavior of that system as some control parameter varies smoothly. Examples in natural systems abound (Haken 1977; Winfree 1980). One such example from biological coordination is briefly mentioned here. Kelso (1984, 1995) observed that when adults are asked to move their index fingers in an anti-phase pattern (both fingers move to the left or the right at the same time), they can perform this task over a wide range of cycling frequencies. But as frequency is increased, subjects show a spontaneous shift to an in-phase pattern, that is, to a pattern where the fingers move toward each other or away from each other at the same time (such qualitative change is commonly referred to as a bifurcation by mathematicians, or as a phase transition by physicists). In this example, then, gradual changes in cycling frequency drive the coordination system from one stable mode of coordination to another, anti-phase to in-phase. The phenomenon has been modeled in detail using nonlinear dynamics by Kelso and colleagues. For a recent review, see Wing and Beek (2002).

To return to phonetics-phonology, the formulation of constraint ranking given above exploits the property non-linearity. The systems in (4) and (7) are qualitatively different. They correspond to distinct Optimality Theoretic grammars, "C1, C2 unranked" in (4) and "C1 >> C2", "C2 >> C1" in (7). What makes this formulation of constraint ranking particularly relevant to phonetics-phonology is that it comes with a handle for driving the system from one qualitative state to another, as a consequence of varying the control parameter k. So from smooth, continuous variation in some control parameter, distinct grammars can emerge. In nonlinear dynamics, then, continuity and discreteness coexist and interact within a unified framework. By contrast, in a derivational phonetics-phonology, there is no way to express this interplay between continuity and discreteness. In such a model, variation in continuous or environmental parameters cannot affect the discrete aspects of phonetics-phonology. Phonologists working on the phonetic bases of phonological patterns have encountered (instances of) this limitation repeatedly. Steriade (1997) has expressed this most accurately and succinctly: "phonetic implementation has to live with prior decisions taken in the phonology" (1997: 3). To generalize the same, in the derivational model, the continuous aspects of phonetics-phonology are enslaved by the discrete dimensions of the system. But as Browman and Goldstein have pointed out, there are clear cases of bi-directional interaction between the discrete and the continuous, or between the macro- and micro-levels of description in their terms (see Browman and Goldstein 1995).

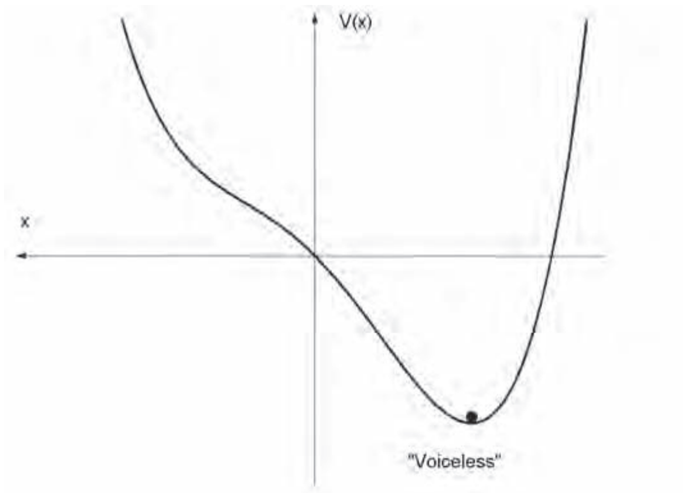
Next, I consider how the concepts introduced here can be applied to our specific problem, the incompleteness of neutralization and its dependence on the communicative context.

4. Grammar dynamics

A first step in a dynamical model of a natural system is mapping the macroscopic observables to attractors of a hypothesized model underlying that system (Kelso, Ding, and Schöner 1992).

Consider the specific phenomenon addressed here, a language with syllable-final devoicing. The relevant macroscopic observables are that coda obstruents are voiceless and that onset obstruents can be voiced or voiceless. To spell out these language-particular properties in dynamical terms, a grammar potential function must be specified that contributes attractors at appropriate values of voicing. Since coda obstruents can only be voiceless, the grammar potential in the coda environment must contribute a single attractor at a value of voicing corresponding to voiceless obstruents. Let us assume that degree of voicing is indexed with the parameter x of glottal aperture, a tract-variable in Browman and Goldstein's dynamical representations. Then, as shown in (8) for the coda position, the grammar attractor appears at the right side of the x axis, at some positive value of glottal opening characteristic of voiceless obstruents. Below, I explain how to derive the specific grammar potential function from basic assumptions.[3]

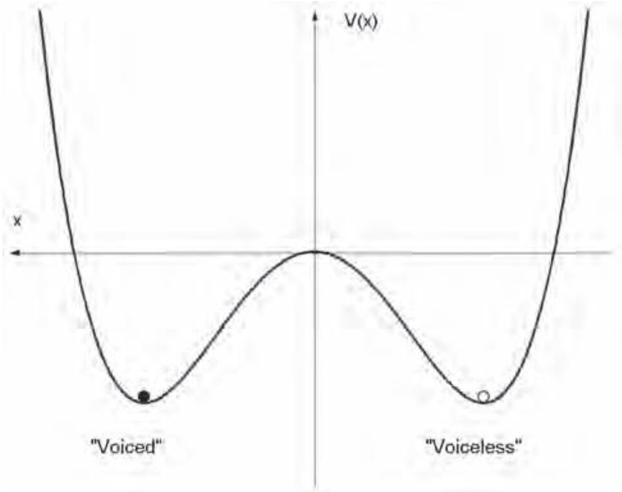
(8) Coda potential in a dynamical model of final devoicing – one attractor present



Onset obstruents can be voiced or voiceless, so the potential in that environment must be bistable, as shown in (9), with attractors at voicing values appropriate for voiced and voiceless consonants.

Let us compare this dynamical model of coda devoicing to an OT grammar for the same phenomenon, "NoVoicedCoda >> Faith(Voice)". Consider what happens to a ball when it is placed within the landscapes specified in (8), (9). We will interpret the initial coordinate of the ball on the x axis as the voicing value in the Input (where Input is as defined in OT). In coda position, there is a unique attractor. As a consequence of the grammar dynamics, the ball ends up in the Voiceless state and this is the only stable state where the ball can end up. In other words, irrespective of the Input voicing value, the Output voicing value is always voiceless, cf. "Richness of the Base" in OT.

(9) Onset potential in a dynamical model of final devoicing – two attractors present



For the potential in (9), on the other hand, there are two attractors and the attractor the ball ends up in is a function of its initial position; when the Input voicing value is in the vicinity of the "Voiced" / "Voiceless" attractor, the Output ends up in the "Voiced" / "Voiceless" state, cf. the notion of Faithfulness in OT. In effect, the dynamical statement of coda devoicing captures the essential properties of the corresponding OT grammar. However, as will be shown later on, the dynamical formulation allows us to model the grammar's interaction with context and ultimately derive phenomena like incomplete neutralization.

I now describe the grammar dynamics formally. As in any (autonomous) dynamical system, grammar dynamics is defined by a differential equation of the general form dx/dt = G(x), where G(x) is a nonlinear function of x. Intuitively, this equation embodies the 'dynamic law' obeyed by the system. A proposed dynamical model of some phenomenon is a good model to the extent that aspects of the phenomenon in question correspond well with qualitative properties of its mathematical formulation (see section 6). As a working hypothesis, I assume that the 'tilted' anharmonic oscillator provides a first approximation for the grammar dynamics: G(x) = dx/dt = – k + x – x^3. The crucial choice here is that this polynomial has to be cubic (the largest exponent of x is 3). This is because we need at least two distinct attractors, one for the voiceless and another for the voiced state. It can be shown that a polynomial of degree less than three, allows for at most one attractor (Arnold 2000).

Given –dV(x)/dx = dx/dt and G(x) = – k + x – x^3, we can compute by integration the potential for the grammar dynamics V(x) = k*x + (–1/2) * (x^2) + (1/4) * (x^4), up to some constant term C which can be ignored as it does not affect the discussion or the qualitative results of the simulations. This V(x) is the potential shown in the (8) above. A similar method allows us to derive the potential V(x) shown in (9).

With the formal aspects of the model specified, we are now in a position to sketch how the grammar is linked to environmental variables and to situate our grammar in communicative context. We know that G(x) has a stable point at the grammatically required value of $x = x^0$ ("Voiceless"). We also know that the observed value of voicing is modulated by extra-grammatical parameters. Voicing is modulated by orthography, as shown in Ernestus and Baayen's work (this volume), and by intentions as shown in Port and Crawford's (1989) work. In what follows, I use intention as the extra-grammatical parameter, without loss of generality.

The basic fact of interest is that intentions can shift the preferred grammar modes. How can we formulate this in a principled way? The core idea to be fleshed out is that intentions contribute to the grammar an attractor corresponding to the intended form. The intention to communicate a lexeme with a final voiced consonant, in particular, is defined as a part of a dynamics that attracts the order parameter toward the intended voicing. In turn, intentions are constrained by the grammar dynamics, namely, by how forms 'should be produced' in specific contexts. Overall, then, grammatical requirements sometimes compete and sometimes cooperate with variable environmental conditions (intentions). The phonetic output is the result of this combination of grammar dynamics and intentional dynamics. Incomplete neutralization will follow as a special case of this interaction.

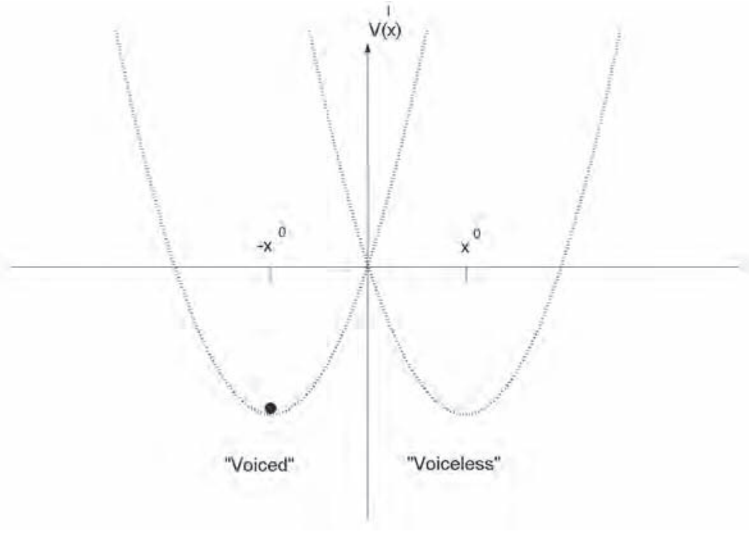
5. Intentional dynamics

To situate grammar in communicative context, we need an appropriate dynamic formulation of intentions.

Informally, intentions are communicative goals. Let us assume a communicative act wherein the speaker's goal is to convey the lexeme *Rad* 'wheel' as opposed to *Rat* 'advice'. Intentional dynamics adds an attractor at the required value of voicing $\{-x^0, x^0\}$, where x^0 '=' [–Voiced], $-x^0$ '=' [+Voiced]. The potential $V^I(x)$ for these two values is shown below. Note that intentions are mutually exclusive. One can't intend *Rad* and *Rat* at the same time – viz. the ball can only be in one of the two attractors, as in (10).

I now describe the formal model for intentions. The dynamics of intentions in the context of a grammar G is modeled by the equation $dx/dt = G(x) + I(x)$, following Schöner and Kelso (1988) on coordinated movement by humans. Intuitively, the 'dynamic law' obeyed by the combined system is given by a linear combination of the grammar dynamics G(x) and the intentional dynamics I(x). I(x) is the simplest function that specifies an attractor at the (intentionally) required value of voicing. That is, $I(x) = \text{intent} * (x^{REQ} - x)$. In this function, 'intent' is a linear term representing the relative strength of the intentional contribution. The higher the value of 'intent', the stronger is the intention. The term x^{REQ} takes values from $\{-x^0, x^0\}$, that is, the values for glottal aperture corresponding to [+Voiced] and [–Voiced].

(10) Dynamical model of "Voiced" and "Voiceless" intentions



Given these assumptions, the contribution to the grammar dynamics that adds an attractor at the required value of voicing is given by the potentials shown above. To derive these potentials, we start with $-dV(x)/dx = dx/dt = G(x) + \text{intent} * (x^{REQ} - x)$, and by basic calculus, we compute the part of the potential that corresponds to the intentional dynamics $V^I(x)$

= (1/2) * intent * (x^2) – intent * x^{REQ} * x, up to a constant C which can be dropped since it is of no qualitative significance in the context of this discussion and the simulations. It is this $V^{I}(x)$ that is shown in the graph above.

I now sum up the essential ingredients of the proposal, in (11). There is a parameterization in terms of an order parameter and a control parameter, in (11a, b) respectively. Order parameters describe the macroscopic form of phonology and grammar principles refer to such parameters (see Gafos 2002 on gestural coordination relations). In our example, the control parameter is intentional strength. As shown in (11c), there is also an 'interface', the hypothesized model relating these two parameters, dx/dt = G(x) + intent * (x^{REQ} – x), where G(x) = – k + x – x^3. Crucially, however, this 'interface' does not translate symbols to continuous signals. Rather, it states a dynamic linkage, in the form of a testable relation, between a grammatical (order) parameter and an extra-grammatical (control) parameter. The linkage is dynamic because the two parameters it relates are interdependent and changing quantities, as seen in section 2.

(11) Nonlinear dynamics as the linkage between the qualitative and the quantitative

a	x (degree of voicing)	order parameter (grammatical)
b	intent (degree of intentional strength)	control parameter (non-grammatical)
c	dx/dt = – k + x – x^3 + intent*(x^{REQ}–x)	the 'interface'; the dynamic linkage between the order and control parameters

In short, this is the core proposal of this paper: an alternative conception of the 'phonetics-phonology interface' where dynamics offers a non-derivational way of relating qualitative and quantitative aspects of phonetics-phonology.

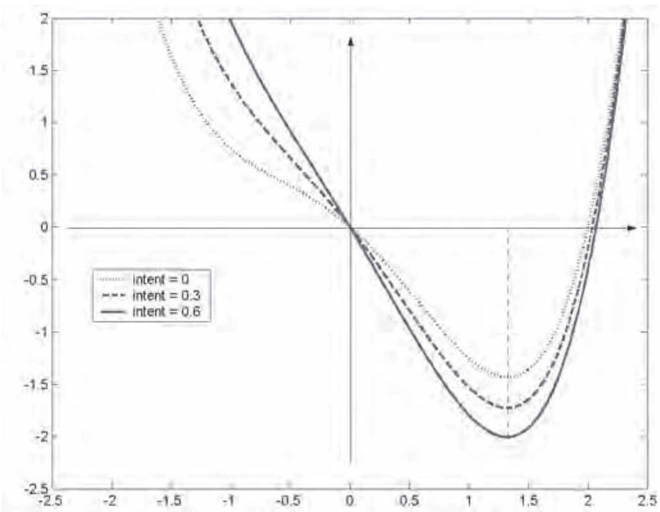
6. Simulations of grammar in varying intentional contexts

I now simulate the combined dynamics, grammar with intentional information. The parameters manipulated in the simulations are intention and its associated strength. Intention is categorically either Voiceless or Voiced, corresponding to the underlying value of the final obstruent in examples like

Rat, *Rad*. Intentional strength is a scalar variable, which varies continuously in the interval [0, 1]. A value closer to 0 corresponds to a context where the speaker's intention to communicate the contrast between *Rat* and *Rad* is weak, as would be the case in the word-list reading, assistant-absent condition. Higher values correspond to communicative contexts with stronger requirements for expressing the contrast as would be the case in the assistant-present condition.

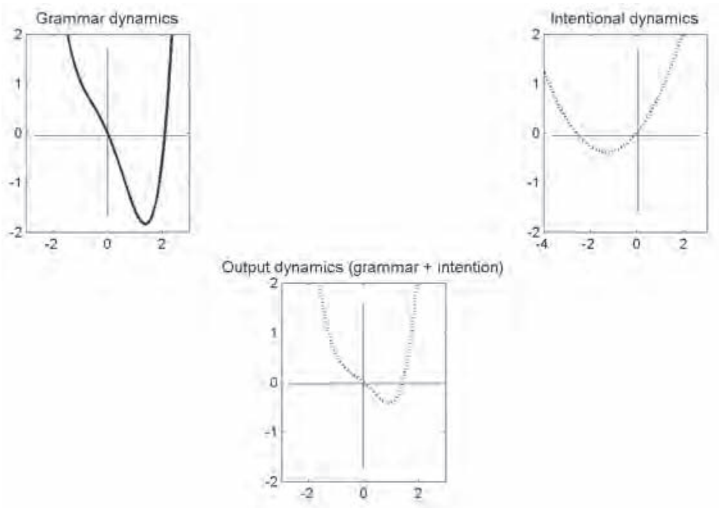
Consider first the case where the intention is a Voiceless obstruent, *Rat* 'advice'. The intentionally required voicing value coincides with the grammatically prescribed value. They are both Voiceless. In this case, then, we have cooperation of intentional requirements and grammar dynamics. As the figure below illustrates, there is no qualitative change in the resulting dynamics, indicated by the fact that the stable point remains fixed at the same value of x (x^0 '=' [–Voiced]).

(12) Grammar dynamics as modified by intentional information [–Voiced]



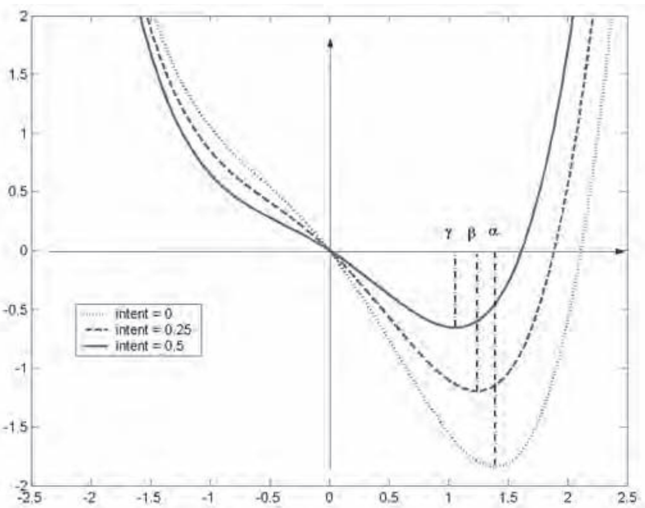
Consider now the more interesting case where the intention is a Voiced obstruent, *Rad* 'wheel'. Here, the grammar dynamics contributes an attractor at the voiceless end of the x axis (the right side) and the Voiced intention contributes an attractor at the voiced end of the x axis (the left side). In this case, then, the intentionally required value does not coincide with the grammatically prescribed value. We have competition between grammar and intention. An instance of this competition is shown in (13) below.

(13) Competition between grammar and intention, when intention is Voiced



The result is that the Voiceless attractor drifts toward less extreme values. This scaling of the system's dynamics is shown more clearly in the figure in (14).

(14) Grammar dynamics as modified by intentional information [+Voiced]



It is observed that, as intentional strength increases, the potential is gradually pulled away from the [–Voiced] minimum toward more voicing (α→β→γ). This is incomplete neutralization.

The effect of communicative context is directly captured in this model by the factor of intentional strength, and its effects on the dynamics. Overall, then, the two facts about neutralization, its incompleteness and its dependence on the communicative context, can be derived using basic concepts and tools of dynamics.

In simulations with this model not shown here, when the intentional strength for Voiced obstruents is increased beyond some relatively high value (> 0.78), the system changes discontinuously so that the only stable mode appears all the way at the other end. That is, the attractor is now at the Voiced end of the voicing continuum. The model then predicts a bifurcation, a qualitative change in the system's dynamics, as a result of a continuous increase in intentional strength. Indeed, if necessary, German speakers can produce Voiced obstruents as voiced in the neutralizing context (*Rad* as [ʁad]).

To sum up, the present model combines two seemingly incompatible ideas from Ernestus and Baayen's (this volume) paper. The first is that "incomplete neutralization seems to be part and parcel of the grammar" (46). In the model, this is reflected in the way intentions parameterize the grammar. The second idea is that "incomplete neutralization may well be primarily a lexical effect" (45). This is reflected by identifying intentions with basic lexical forms. The intention for *Rat* is identified with an attractor at the voiceless end, whereas that for *Rad* with an attractor at the voiced end (of the order parameter, Voicing). As a consequence, intentions attract the order parameter toward the intended 'lexical' voicing. For voiced obstruents, specifically, incomplete neutralization follows.

7. Conclusion

The view of a phonological component preceding a phonetic implementation component is one way of expressing the intuition that phonetics-phonology is a system with qualitative and quantitative aspects (Ladd, this volume). However, it may not be the only way. A look at other complex systems may provide clues for alternative design methodologies. Given the preeminent view of language as a 'biological object' (Chomsky 2000), biological systems are the natural candidates. In theoretical biology (Waddington 1970; Pattee 1973), organisms described at the macroscopic level exhibit low-dimensional qualitative properties of considerable simplicity. At the microscopic level, the physicochemical processes of molecular biology are vastly detailed and continuous. Here, the temporal precedence metaphor of qualitative before quantitative clearly fails. It does not make sense to say that the qualitative aspects of a living organism are related by precedence to their quantitative manifestations. The qualitative and quantitative coexist as two mutually dependent parts of a coherent whole.

Down to the more concrete level of analytical tools, the view of language as fundamentally biological suggests the use of the mathematics employed by leading physicists (Haken 1977) and biologists (Yates 1984) to study complex systems. As a small step in that direction, I hope to have shown some of the promise of nonlinear dynamics in providing a powerful formal method for addressing the central issue behind the phonetics-phonology 'interface', the issue of the relation between qualitative and quantitative aspects of phonetics-phonology.

The proposal is that it is both necessary and promising to do away with the temporal metaphor of precedence between the qualitative and the quantitative, without losing sight of the essential distinction between the two. This leads to the alternative non-derivational conception of the term 'interface' as a dynamic linkage between the two interdependent aspects of a unified system.

Notes

* Preparation of this paper was partly supported by a NYU Goddard fellowship, NYU Research Challenge Fund grant N5006, and NIH grant HD-01994 to Haskins Laboratories. The simulations reported were done in S-Plus 6 and Matlab 6.1. Thanks to Robert Port, Luigi Burzio, Paul Smolensky, Carol Fowler, Louis Goldstein, René Kager and Wim Zonneveld for useful feedback, to Janet Pierrehumbert, Donca Steriade, and Alice Turk for cogent comments and questions after this paper was presented at the Laboratory Phonology meeting, and to Arto Anttila, Richard Kayne and Anna Szabolsci at NYU for useful discussions. Any errors are my responsibility.

1. Fourakis and Iverson (1984) ascribe the incompleteness of neutralization to "hypercorrection under linguistically artificial conditions [AG: orthography in word list reading]" (149). But Catalan, a language where incomplete final devoicing has been documented, lacks an orthographic distinction between word-final underlying voiced and voiceless stops (see references in the text). See also Charles-

Luce (1985: 318–319), Port and Crawford (1989: 258–259), and Ernestus and Baayen (this volume) for related discussion.

2. The issue of identifying the right parameter for voicing is a difficult one. In our example, we have a number of choices. Voicing can be identified with glottal aperture, glottal tension, larynx lowering or some other parameter that may be a combination of these. Ultimately, the right choice will depend on the specific language, but this issue is orthogonal to the argument made here.
3. Note that I do not examine the independent issue of why grammars develop properties like final devoicing. See Steriade (1997) for a proposal on this, and for an OT analysis of numerous laryngeal neutralization phenomena.

References

Arnold, Vladimir I.
2000 Nombres d'Euler, de Bernoulli et de Springer pour les groupes de Coxeter et les espaces de morsification: le calcul de serpents. In: Éric Charpentier et Nicolas Nikolski (eds.), *Leçons de Mathématiques d'Aujourd'Hui*, 61–98. Paris: Cassini.

Beckman, Mary E. and John Kingston
1990 Introduction. In: J. Kingston and M. E. Beckman (eds.), *Papers in Laboratory Phonology I: between the grammar and the physics of speech*, 1–16. Cambridge: Cambridge University Press.

Beer, Randall D.
1995 Computational and dynamical languages for autonomous agents. In: Robert F. Port and Timothy van Gelder (eds.), *Mind as Motion*, 121–148. Cambridge, MA: MIT Press.

Benus, Stefan, Adamantios Gafos and Louis Goldstein
in press Phonetics and phonology of transparent vowels in Hungarian. In: *Proceedings of the Berkeley Linguistics Society*.

Bloomfield, Leonard
1933 *Language*. New York: Holt, Rinehart and Winston.

Boersma, Paul and Bruce Hayes
2000 Empirical tests of the gradual learning algorithm. *Linguistic Inquiry* 32: 45–86.

Browman, Catherine P. and Louis Goldstein
1986 Towards an articulatory phonology. *Phonology* 3: 219–252.
1995 Dynamics and articulatory phonology. In: Robert F. Port and Timothy van Gelder (eds.), *Mind as Motion*, 175–193. Cambridge, MA: MIT Press.

Cariani, Peter A.
1989 On the design of devices with emergent semantic functions. Ph.D. dissertation, State University of New York at Binghampton.
Cedergren, Henrietta and David Sankoff
1974 Variable rules: performance as a statistical reflection of competence. *Language* 50: 233–55.
Charles-Luce, Jan
1985 Word-final devoicing in German: effects of phonetic and sentential context. *Journal of Phonetics* 13: 309–324.
1993 The effects of semantic context on voicing neutralization. *Phonetica* 50: 28–43.
1997 Cognitive factors involved in preserving a phonemic contrast. *Language and Speech* 40: 229–248.
Chomsky, Noam
2000 *New horizons in the study of language and mind*. Cambridge: Cambridge University Press.
Chomsky, Noam and Morris Halle
1968 *The sound pattern of English*. New York: Harper and Row.
Cohn, A. C.
1990 Phonetic and phonological rules of nasalization. Ph.D. dissertation, University of California, Los Angeles (*UCLA Working Papers in Phonetics* 76).
Coleman, John
1992 The phonetic interpretation of headed phonological structures containing overlapping constituents. *Phonology* 9: 1–44.
Dinnsen, Daniel A.
1985 A re-examination of phonological neutralization. *Journal of Linguistics* 21: 265–279.
Dinnsen, Daniel A. and Jan Charles-Luce
1984 Phonological neutralization, phonetic implementation and individual differences. *Journal of Phonetics* 12, 49–60.
Eco, Umberto
1984 *Semiotics and the philosophy of language*. Bloomington, IN: Indiana University Press.
Ernestus, Mirjam and Harald Baayen
(this vol.) The functionality of incomplete neutralization in Dutch: the case of past-tense formation.
Fodor, J. A. and Z. W. Pylyshyn
1981 How direct is visual perception? Some reflections on Gibson's 'ecological approach'. *Cognition* 9: 139–196.

Fougeron, Cécile and Donca Steriade
1997 Does deletion of French schwa lead to neutralization of lexical distinctions? In: *Euro-Speech 1997: Proceedings of the 5th European conference on speech communication and technology* (University of Patras, Greece), Vol. 7, pp. 943–946.

Fourakis, Marios and Gregory K. Iverson
1984 On the 'incomplete neutralization' of German final obstruents. *Phonetica* 41: 140–149.

Gafos, Adamantios
2002 A grammar of gestural coordination. *Natural Language and Linguistic Theory* 20: 269–337.

Haken, Hermann
1977 *Synergetics, An introduction*. Heidelberg: Springer-Verlag.
1990 Synergetics as a tool for the conceptualization and mathematization of cognition and behavior – How far can we go?. In: H. Haken and M. Stadler (eds.), *Synergetics of Cognition* (Springer Series in Synergetics, Vol. 45), 2–31. Heidelberg: Springer-Verlag.

Harnad, Stevan
1990. The symbol grounding problem. *Physica D* 42: 335–346.

Hintzman, D.
1986 "Scheme abstraction" in a multiple-trace memory model. *Psychological Review* 93: 411–428.

Keating, Patricia A.
1988 Underspecification in phonetics. *Phonology* 5: 275–292.
1990 Phonetic representations in generative grammar. *Journal of Phonetics* 18: 321–334.

Kelso, Scott J.A.
1984 Phase transitions and critical behavior in human bimanual coordination. *American Journal of Physiology* 15, R1000–R1004.
1995 *Dynamic patterns: the self-organization of brain and behavior*. Cambridge, MA: MIT Press.

Kelso, Scott J.A., Mingzhou Ding, and Gregor Schöner
1992 Dynamic pattern formation: a primer. In: A. Baskin and J. Mittenhal (eds.), *Principles of organization in organisms*, 397–439. Santa Fe, NM: Santa Fe Institute; and Reading, MA: Addison-Wesley.

Kingston, John and Randy Diehl
1994 Phonetic knowledge. *Language* 70: 419–54.

Ladd, D. Robert
this vol. "Distinctive phones" in surface representation.

Lass, Roger
1997 *Historical linguistics and sound change*. Cambridge: Cambridge University Press.

Liberman, Mark Y.
1983 In favor of some uncommon approaches to the study of speech. In: Peter F. MacNeilage (ed.), *The Production of Speech*, 265–274. New York: Springer-Verlag.

Liberman, Mark Y. and Janet Pierrehumbert
1984 Intonational invariance under changes in pitch range and length. In: M. Aronoff and R. T. Oehrle (eds.), *Language Sound Structure*, 157–233. Cambridge, MA: MIT Press.

Miller, Joanne L.
1981 Effects of speaking rate on segmental distinctions. In: P.D. Eimas and J.L. Miller (eds.), *Perspectives on the study of speech*, 39–74. Hillsdale, NJ: Erlbaum.

Pattee, Howard H.
1973 The physical basis and origin of hierarchical control. In H. H. Pattee (ed.), *Hierarchy Theory: The Challenge from Complex Systems*, 71–108, New York: Braziller.

Petitot-Cocorda, Jean
1985 *Les catastrophes de la parole. De Roman Jakobson à René Thom*. Paris: Maloine.

Pierrehumbert, Janet
2001 Exemplar dynamics: word-frequency, lenition and contrast. In: Bybee, J. and P. Hopper (eds.), *Frequency effects and the emergence of linguistic structure*, 137–157. Amsterdam: John Benjamins.
2002 Word-specific phonetics. In: C. Gussenhoven and N. Warner (eds.), *Laboratory Phonology VII*, 101–139. Berlin/New York: Mouton de Gruyter.

Port, Robert and Michael L. O'Dell
1985 Neutralization of syllable-final voicing in German. *Journal of Phonetics* 13: 455–471.

Port, Robert and Penny Crawford
1989 Incomplete neutralization and pragmatics in German. *Journal of Phonetics* 17: 2257–2282.

Port, Robert and Timothy van Gelder (eds.)
1995. *Mind as motion: Explorations in the dynamics of cognition.* Cambridge, MA: Bradford books, MIT Press.

Prince, Alan and Paul Smolensky
1993 *Optimality Theory: Constraint Interaction in Generative Grammar*, unpublished manuscript, Rutgers University and University of Colorado, Boulder.

Saltzman, Elliot
1995 Dynamics and coordinate systems in skilled sensorimotor activity. In: Robert F. Port and Timothy van Gelder (eds.), *Mind as Motion*, 149–174. Cambridge, MA

Sankoff, David
1988 Variable rules. In: U. Ammon, N. Dittmar and K. J. Mattheier (eds.), *Sociolinguistics: An international handbook of the science of language and society*, 984–997. Berlin/New York: Mouton de Gruyter.

Schöner, Gregor and J. A. Scott Kelso
1988 A dynamic pattern theory of behavioral change. *Journal of Theoretical Biology* 135: 501–524.

Slowiaczek, Louisa M. and Daniel A. Dinnsen
1985 On the neutralizing status of Polish word-final devoicing. *Journal of Phonetics* 13: 325–341.

Smolensky, Paul
1988 On the proper treatment of connectionism. *Behavioral and Brain Sciences* 11: 1–74.

Sommerstein, Alan H.
1977 *Modern phonology*. London: Edward Arnold.

Steriade, Donca
1997 Phonetics in phonology: the case of Laryngeal neutralization. Los Angeles: UCLA ms. [Published in UCLA Working Papers in Phonology.]

Stevens, Kenneth N.
1972 The quantal nature of speech: evidence from articulatory-acoustic data. In: E. David and P. Denes (eds.), *Human Communication: A Unified View*, 51–66. New York: McGraw-Hill.
1989 On the quantal nature of speech. *Journal of Phonetics* 17: 3–45.

Summerfield, Quentin
1981 Articulatory rate and perceptual constancy in phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance* 7: 1074–1095.

Thompson, J. M. T., and H. B. Stewart
2002 *Nonlinear dynamics and chaos*. [2nd edition.] West Sussex, UK: John Wiley and Sons.

Titze, I. R.
1988 The physics of small amplitude oscillation of the vocal folds. *Journal of the Acoustical Society of America* 83: 1526–1552.

Trubetzkoy, N. S.
1969 *Principles of phonology*. Berkeley and Los Angeles: University of California Press. [translation by C. Baltaxe of *Grundzüge der Phonologie*. 1958. Goettingen: Vandenhoeck and Ruprecht.]

Tuller, Betty, Pamela Case, Mingzhou Ding, and J. A. Scott Kelso
1994 The nonlinear dynamics of speech categorization. *Journal of Experimental Psychology* 20 (1): 3–16.
Turvey, M. T.
1990 Coordination. *American Psychologist* 45: 938–953.
Waddington, C. H. (ed.)
1970 *Towards a theoretical biology* (Volume 3). Chicago: Aldine Publishing Company.
Winfree, Arthur T.
1980 *The geometry of biological time*. New York: Springer-Verlag.
Wing, Alan M., and Peter J. Beek
2002 Movement timing: a tutorial. In: Wolfgang Prinz and Bernhard Hommel (eds.), *Attention and Performance XIX: Common mechanisms in perception and action*, 203–226. Oxford University Press.
Yates, F. E. (ed.)
1984 *Self-organizing systems: The emergence of order*. New York: Plenum Press.

The statistical basis of an unnatural alternation

Janet B. Pierrehumbert

The /k/-/s/ alternation in word pairs such as electric, electricity is not phonetically natural, and is learned by English speakers as a generalization over morphological relatives. Data gathered in an open-response experimental task show that it is productive before the suffix -ity. Lesser productivity is found for the same materials in a backformation task. Outcomes are analyzed as the result of a cognitive process of statistical inference. Abstract descriptions make crucial use of phonological variables. Cognitive preferences for certainty and for robust, redundant descriptions are argued to jointly determine the universe over which operational statistics are estimated.

1. Introduction

This paper presents an experimental study of the productivity of the /k/-/s/ alternation exhibited in derivational pairs such as *electric*, *electricity*. It is exemplified in numerous word pairs in English involving various suffixes, including -*ity*, -*ism*, and -*ist*. The Collins on-line English dictionary (distributed in 1990 through the ACL Data Collection Initiative) includes 108 clear examples of words ending in these suffixes in which a stem-final /k/ softens to /s/. The largest group, and the main topic of this paper, is words formed with -*ity*. The dictionary contains only twelve words with stems ending in /k/ which fail to soften before one of these three suffixes (e.g. *anarchy*, *anarchism*, *York*, *Yorkist*). All involve affixes other than -*ity*.

The productivity of the alternation is disputable. First, there are very few forms which would support extension of the alternation beyond an orthographic -*ic* followed by one of the triggering suffixes. For the suffixes just listed, the only common examples listed in the Collins are *Greek*/*Grecism*; *opaque*, *opacity*; *reciprocal*/*reciprocate*, *reciprocity*; and *pharmacology*, *pharmacist*. (A number of potentially relevant pairs, such as *caducous*, *caducity*; *cecum*, *cecity*; *paucal*, *paucity*; *raucous*, *raucity* would only be known to very erudite speakers.) Second, as Myers (1999) also notes, the

/k/-/s/ alternation as presently found in English is not natural (in the sense of Anderson 1981).

Velar Softening is not phonetically natural because the evident phonetic pressures on a /k/ in the target position would not produce /s/. If the suffix vowel is /ɪ/ (as transcribed in the dictionary) then coarticulation and lenition would yield an aspirated palatal approximant rather than the alveolar fricative /s/ (see Lavoie 2001). The alternative possibility for the vowel, /ə/, provides still weaker phonetic motivation for /s/. /s/ differs from the phonetically expected outcome by its maximal vocal fold abduction and its precise tongue shaping, which directs a jet of air against the teeth. These are active adjustments which cannot be characterized as accommodation to a following vowel. Thus, understanding the alternation of /k/ with /s/ requires recourse to some version of the concept of Structure Preservation in phonology (see Kiparsky 1985) which states that lexical alternations stay within a language's system of phonological categories. Since an aspirated palatal approximant is not a contrastive category in the English lexicon, it cannot be the outcome of a morphophonological rule, either. The reanalysis involved in lexicalizing the phonetically expected approximant as the lexically contrastive segment /s/ reveals the role of abstract cognitive factors, over and beyond phonetic ones. Guion (1998) also discusses the role of perceptual structuring for the typologically related change /k/ → /tʃ/.

Velar Softening is also unnatural because it is phonologically opaque. Though it originates historically in fronting and spirantization of the velar stop before a non-low front vowel, suffixes with such vowels on the surface do not in general trigger the softening of /k/ to /s/ in the synchronic phonology. /k/ never softens to /s/ before -*y*, as *smoke*, *smoky*. On the other hand, -*ize*, beginning with a low vowel, does trigger softening, because -*ize* formerly had a nonlow front vowel. In Chomsky and Halle (1968), this historical ordering is recapitulated in the extrinsic ordering of rules in the synchronic phonology. The phonological opacity created by such orderings is precisely one reason that the psychological validity of the Chomsky-Halle model became a matter of widespread dispute. The finding that the vowel shift is only partially productive (c.f. Jaeger 1984; McCawley 1986) also calls into question the productivity of the rule of Velar Softening, which is ordered before it.

Understanding productivity is important because it provides a crucial line of evidence about cognitive abstractions. The failure of an alternation to generalize suggests that no abstract generalization over the forms exhibiting the alternation has been formed. If the alternation is aggressively and reliably extended, even to forms which differ substantially from attested ones, it follows that a very broad abstraction has been formed. For example, the reliable and aggressive extension of the regular English plural pattern indicates that it abstracts away from many properties of the word. If the situation lies somewhere in the middle, then the exact pattern of productivity can yield insights about the exact character of the abstraction that is formed.

Phonotactics is the area in which most research has been done on the availability of lexical patterns for use in novel forms. Numerous studies, reviewed in Pierrehumbert (2003), indicate that the type frequency (frequency in the lexicon) of a phonological pattern affects the likelihood and perceived well-formedness of novel words containing that pattern. This dependence is gradient; frequent sequences are readily extended to new words, rare sequences are avoided, and moderately frequent sequences fall in between. For example, Hay, Pierrehumbert, and Beckman (2004) found that the perceived well-formedness of novel words containing nasal-obstruent clusters (such as /strɪnfi/ and /zæmpɹ/) was a gradient function of the frequency of the cluster. The frequency for a tautomorphemic cluster was estimated as its frequency in trochaic monomorphemic words with a lax front vowel in the CELEX monomorphemes. (see Baayen, Piepenbrock and Gulikers 1995, regarding CELEX; Hay, Pierrehumbert and Beckman 2004, regarding monomorphemes).

This choice of universe for estimating frequencies was opportunistic, and obscures a central issue in understanding the relation of lexical frequencies to pattern productivity. This issue is taken up with Figure 1. Figure 1 shows a partial lattice of heterosyllabic N.O clusters. The atoms on the bottom are individual heterosyllabic phoneme clusters. The nodes above the atoms are some of the various available natural classes of such clusters. As is well-known, natural classes can be formed using partial descriptions of phonological patterns. For example, the sequence /nt/ is an element of the set of clusters of /n/ followed by any stop; it is also an element of the set of clusters containing a homorganic nasal and stop. The cluster /np/ belongs to the former set but not the latter; the cluster /mp/ belongs to the latter set but not the former. The lattice is organized from specific (on the bottom) to general (at the top). Each node is labeled with the probability of the indicated descriptor with respect to the universe of N.O clusters, as estimated from counts in the CELEX monomorphemes. Clearly, the less specific the description, the more cases it encompasses and the larger the natural class it describes. Thus, the probabilities go up as we follow the lines up the lattice, but the exact way they go up depends on exactly what is lumped together in each class.

The topmost case, any nasal followed by any obstruent, is taken to define the universe for the probabilities which are indicated below each node. If the universe were larger, the probabilities shown on the figure would all be smaller, but their rankings would remain the same.

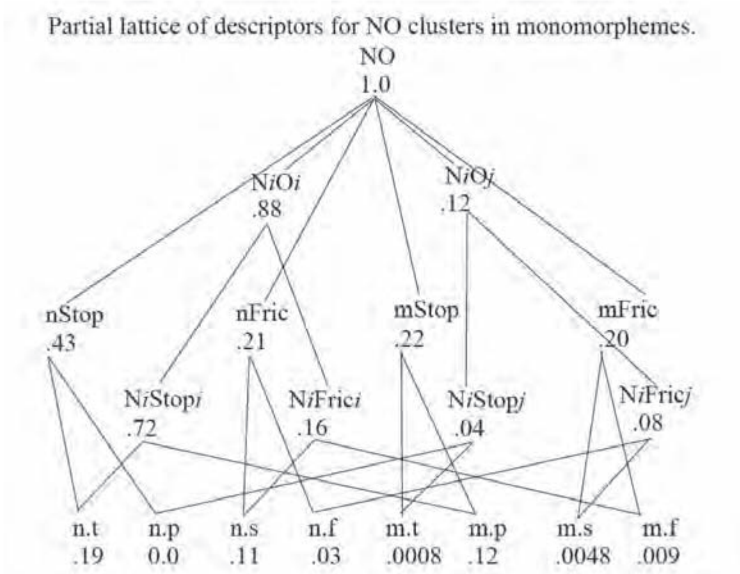


Figure 1. Partial lattice of probabilities for the universe of heterosyllabic nasal-obstruent clusters in monomorphemic words of English. Counts are established with respect to the Celex monomorphemes, as discussed in Hay et al. (2004).
Atoms at the bottom of the lattice represent specific nasal-obstruent sequences. For the sake of legibility, only eight atoms with a total P = 0.54 are shown. The remaining sequences, with a total P = 0.46, have been omitted. Superordinate nodes represent classes of nasal-obstruent sequences. Italic indices are used for convenience to indicate homorganicity or nonhomorganicity. (Actual phonological structures for homorganic consonants have feature sharing.) Capital N denotes any nasal. Probabilities of superordinate categories include frequencies of atoms which are not shown at the bottom, but which are properly described by the superordinate note. Superordinate nodes /ŋStop/ and /ŋFric/ are not shown.

The question raised by Figure 1 is: Of all the probabilities which may be defined using partial phonological descriptions of a pattern, which are relevant to productivity and perceived well-formedness? For example, if the description N_iO_i (p = .88) versus N_iO_j (p = .12) were the relevant level of abstraction then the feature [+/– continuant] would have no importance for the evaluation and productivity of these clusters, and inhomorganic nasal-fricative clusters would seem every bit as bad as inhomorganic nasal-stop clusters. However, results in Hay et al. (2004) clearly show differential outcomes for nasal-stop and nasal-fricative clusters, indicating that this degree of generalization is too great. The next level down shows two alternative ways to break out the cases. One fixes the nasal consonant and generalizes over the following the obstruent. The other generalizes over place. It separates homorganic from inhomorganic clusters regardless of place, but it maintains information about the continuance of the obstruent. If the first alternative were the cognitively relevant description, then /np/ would be as acceptable as /nt/. This is false. The second alternative also groups /nt/ in a class with other clusters, namely /mp/ and /ŋk/ but not /np/. It is closer to the true state of affairs, since it captures both a strong effect of homorganicity on nasal-stop clusters and a weaker effect on nasal-fricative clusters. However, the lattice structure does not in itself say that one line of generalization is more relevant than the other. Although any phonologist would sensibly prefer one line of generalization to the other, there is no explicit formal account of what this "sensibleness" consists of. Still less is there an explanation of why subjects operated at a detailed level of description, rather than applying a simple overarching generalization about all N.O clusters.

The same issue arises in a different guise in dealing with morphophonological alternations. When such alternations are language particular, they must be learned from examples. There is by now abundant evidence that the productivity of an alternation depends on its type frequency (as well as on other factors). Alternations found in extremely few types, such as irregular conjugations for auxiliaries, are not productive no matter how frequently the irregular forms may be used. However, the universe of examples relevant for any given alternation, and the types of formal generalizations which are made over these examples, is not well understood.

In comparison to phonotactics, morphophonology provides both challenges and opportunities in addressing this issue. It is challenging because many morphophonological alternations are generalizations over word pairs rather than merely over words. For phonotactics, set theory provides a convenient hierarchy of abstraction over words, taking the shape of a lattice of partial descriptions as in Figure 1. For word pairs, in contrast, the proper formal toolkit is not as evident. Is it partial descriptions of the base which are relevant? Or partial descriptions of the complex form? Or relations of partial descriptions of the base and the complex form? The research literature contains case studies which argue for all of these possibilities. As discussed in Myers (1999) and below, the /k/-/s/ alternation needs to be formalized with respect to word pairs; but this need does not in itself define the relevant universe of word pairs. For determining the pronunciation of a novel form *clemicity* given the base *clemic*, the pair *conic*, *conicity* is plainly relevant. But how about *Turk*, *Turkism* (involving a different affix)? Or *ferocious*, *ferocity* (for which no bare form of the base exists)? Or *morbid*, *morbidity* (illustrating preservation of a final stop before the same affix)?

In short, the expectation (based on results in phonotactics and in psychology) that implicit knowledge of morphophonology is stochastic does not in itself define what probabilities will be relevant. Probabilities can be estimated for any formal description that can be tabulated. Many of the "analytic biases" mentioned in Steriade's (2002) original commentary on this session can be viewed as claims about what statistics are available to the cognitive system. For example, in discussing Goldrick (2002), she suggests an analytic bias to the effect that voicing pairs should alternate alike. This is equivalent to the claim that statistics on formal descriptions which abstract over place, but not voicing, are highly available in the formation of phonological grammars.

The investigation of unnatural alternations provides special opportunities in understanding how people form abstract generalizations, because it sidesteps one of the most recalcitrant problems of phonology. This is the relationship of frequency to the phonetic foundations of phonological systems. Under the rubric of markedness theory, scholars have long observed that phonetic simplicity is related to frequency and to default phonological behavior. Quantitative phonetic models have now gone some distance towards elucidating the articulatory and psychoacoustic basis for more and less common segment types, and similar arguments can also be made about phonological sequences. For example, the tendency for languages to favor homorganicity between a nasal and a following obstruent is agreed to be founded in the tendency towards gestural overlap between successive consonants. In the light of such research, there is a risk of confusing correlation and cause when interpreting experimental findings such as Hay, Pierrehumbert and Beckman (2004). The high correlation ($r^2 = 0.65$) they report between lexical log frequency and perceived well-formedness could in principle arise from a concealed factor, namely markedness. Possibly, the phonetically simpler clusters are judged to be better because they are simpler and they are also used more often in words because they are simpler. This would lead to a correlation between lexical frequency and perceived well-formedness, even in the absence of any ability to learn lexical frequencies.

Arguing against this viewpoint are reported findings of dissociations between universal markedness and stochastic generalizations within specific languages. The phoneme /t/ is rare in Arabic despite being unmarked (see counts in Frisch, Pierrehumbert and Broe 2004). Whalen (in press) observes that clicks are common in languages which have them, despite being marked. The learnability of language-specific sequential statistics is shown by comparative studies such as Cutler and Otake's (1998) findings on NO clusters in Japanese versus Dutch. Thus, empirically observed frequency effects are not in general reducible to markedness effects. However, present knowledge is very incomplete and this is an important issue for further research. Work on unnatural processes can make a contribution to this research by permitting an examination of frequency effects in a area where the phonetic foundation is poor. Since the /k/-/s/ alternation is neither pervasive in English nor ubiquitous across languages, any frequency effects which are observed can be presumptively attributed to the the learning from experiences with words of English.

2. Methods

The experiment uses a wugs paradigm, pioneered in Berko's (1958) experiments with children. In this paradigm, subjects are taught a novel stem and they use it as the base for a complex form. "Here is a wug. Look, now there are two of them. There are two ?????". This paradigm has been widely used to investigate the competition between regular and irregular inflectional forms; see, for example, Bybee and Pardo (1981) and Albright and Hayes (2003). Here, it is extended to derivational morphology, an extension also made in Zuraw's (2000) study of Tagalog morphology.

Most early studies of derivational morphology, such as the studies of the English Vowel Shift presented in Jaeger (1984), and McCawley (1986), use concept formation tasks or judgments of words presented in pairs, rather than the wugs paradigm. These tasks have the potential drawback of priming awareness of the regularity being studied through the very design of the stimulus materials. The wugs paradigm, with its open-response format, is more conservative. The materials for this study did not provide any examples of a /k/-/s/ alternation and the subjects were unaware that this alternation was being investigated.

Two related experiments are reported. In both, the stimuli were two sentence paragraphs. The first sentence introduced a target word. The subject's task was to supply the morphologically related word missing in the second sentence. For one group of subjects, the first sentence introduced a base adjective, and the task was to create an abstract noun. The instructions mentioned the variety of means in English for turning adjectives into nouns, such as affixing *-ity* (as in *virgin bride*/*virginity*) and affixing *-ness* (as in *bright*/*brightness*). Subjects were told to make a noun in any way they wished, and to respond as soon as an idea occurred to them. For the second group, the format of the materials was reversed, and the subject's job was to backform the base adjective from an abstract noun. Subjects were young adults recruited through Northwestern University and Ohio State University. Some of the subjects were members of a subject pool comprised of students in introductory linguistics courses, and others were paid $8. There were 10 subjects in the noun formation task and 7 in the backformation task. No subject did both tasks.

For both groups, the instructions and the materials were presented entirely orally. There were 64 items: 16 baseline items (extant words with an established nominal form ending in *-ness* or *-ity*); 16 fillers (existing and novel words that present some uncertainty between *-ity* and *-ness*), and 32 novel target words. None of the example items, baseline items, or fillers involved /k/, /s/, or an alternation between /k/ and /s/. The target word in each stimulus was the last or next-to-last word in the sentence. The same set of bases figured in both the noun formation and the backformation task. A full listing of baseline, target, and filler words can be found in the appendix, including IPA transcriptions for nonwords used in filler and target items.

Of the 16 baseline items in the experiment, eight were words for which there is an established noun in *-ness* and no established noun in *-ity*, such as (1).

(1) When Anna discovered a new doughnut shop, she was very happy.
For her, a warm doughnut means ?????.
(ANSWER: happiness, *happyity)

Eight were words for which there was an established form in *-ity*, with the *-ness* form, if any, having an inappropriate meaning.

(2) Bob's short-term bonds were among his most liquid assets. After he got arrested, he was able to post bail because of his high ?????.
(ANSWER: liquidity, *liquidness)

The sixteen filler items were evenly divided between real words and non-words. For both, -*ness* and -*ity* forms were possible, as indicated by pilot testing.

(3) My brother has always been very frugal. Reusing aluminum foil is just one symptom of his ?????.
(ANSWER: frugalness OR frugality)

(4) Anthropologists working in Manuka found all the hallmarks of a caustive society. In fact, it became a textbook example of ?????.
(ANSWER: caustiveness OR caustivity).

There were four different types of target items, differing in their prototypicality as hosts for a /k/-/s/ alternation. Eight items were Latinate pseudowords, ending in the phonetic form of the suffix -*ic*, /ɪk/). In the pronunciation used by the experimenter, this is a front schwa, as in the well-known minimal pair *roses* /ɹozɪz/, *Rosa's* /ɹozəz/. All Latinate pseudowords were polysyllabic, and some suggested existing words through their prefixes or stems.

(5) Halley's comet is a very interponic comet. Its orbital period varies because of its ?????.

The second set of eight target items, the semi-Latinate set, had a main stress on the initial syllable and a secondary stress on the last syllable. These items also ended in /k/. The (unreduced) vowel in the last syllable was /ɛ/, /æ/ or /a/. Thus, the last syllable clearly differed from -*ic*.

(6) Before Pierre stood an electrifyingly hovac sculpture. In his entire career as curator, he had never before seen such a perfect example of ????? .

The third set, the non-Latinate bases, were monosyllabic pseudowords (in some cases with a prefix *over*- or *under*-).

(7) Inside, the light was so dim it was entirely mork. We couldn't read the instructions in the ?????.

The stimuli also included a fourth set of target items, non-Latinate pseudowords ending in /s/. These are not eligible for velar softening, but are included because they are needed in the backformation task to ensure a balance between /...knɛs/ and /...snɛs/ forms.

In the backformation condition, the two sentence paragraphs were reworked so that the first sentence introduced a complex noun, and the second sentence had a missing adjective. The semantic content of each paragraph was minimally modified so as to maintain the contexts for the forms. For example, the reversed version of (5) is (8) and the reversed version of [6] is (9).

(8) The period of Halle's comet varies because of its interponicity. It is a very ????? comet.

(9) In Pierre's entire career as a curator, he had never before seen such a perfect example of hovacity. It was an electrifyingly ????? sculpture.

Stimuli were block randomized in eight blocks of eight, and read aloud to each subject individually. Subjects repeated the target adjective out loud during a pause after the first sentence. Their pronunciation was corrected if necessary. Intersubstitutions of /æ/ and /ɛ/ in words such as *bowdec* and *hovac* were, however, accepted, as some speakers have apparently merged these vowels. The responses were transcribed as they occurred. The entire session was recorded, and the recordings were used to resolve the few uncertainties in transcription.

To score the noun formation data, the subset of responses in which subjects selected the affix -*ity* after a stem ending in /k/ was extracted. The frequency with which /k/ is softened to /s/ is computed on this subset of "hits". The size of this subset differed considerably across subjects due to the open response format. The hits for the backformation task are the responses in which all and only the affix -*ity* was removed from a noun ending in /sɪti/. The frequency with which the bare stem was produced with final /k/ (as opposed to /s/ or some other consonant) was computed on this set.

Subjects in both groups understood the instructions and generally succeeded at the task. Debriefing after the noun formation task revealed that only one subject was able to guess before the end of the experiment that the /k/-/s/ alternation was being investigated. In the backformation task, no subject guessed what was being scored in the experiment.

3. Results

In the noun formation condition, subjects produced -*ity* and -*ness* responses about equally often. 80% of baseline items predicted to have -*ity* did indeed have it. 82% of baseline items predicted to have -*ness* had -*ness*. 53% of fillers were produced with -*ity* and 47% with -*ness*.

The ten subjects in the noun formation task produced a total of 71 hits. Of the 10 subjects, eight produced examples of velar softening, with the number of examples ranging from 4 to 13 per subject. These results indicate that velar softening is productive for most educated adults. Results by target type are shown in Table 1. For Latinate hits, softening applied nearly 100% of the time. It was somewhat less reliable for the semi-Latinate hits, but the sample size is not big enough to be confident of a difference. In the few cases in which -*ity* was attached to a non-Latinate target, softening never applied. Though the number of such cases was small, the lack of softening is so readily confirmed by native speaker judgments that I will view it as a fact which needs to be explained. Thus, the main effects which require explanation are the high productivity of softening for the semi-Latinate stems, a group for which there is no critical mass of extant forms, and the lack of softening in the non-Latinate stems (given that softening was observed in the Latinate and semi-Latinate stems).

Table 1. Outcomes in the noun formation task.

Target Type	Hits	Softening Before -*ity*	% Softening
Latinate targets	30	28	93
Semi-Latinate targets	36	30	83
Non-Latinate targets	5	0	0

Subjects were generally successful on the backformation task. There were only two errors on the baseline items (*inanity* → *inate* and *profanity* → *profound*). Over all items, subjects produced a form with a bare stem 86% of the time. (Other responses involved either addition of a suffix, stem truncation, or lexical intrusions.) A total of 68 hits reflect an implicit choice to preserve the surface /s/ or to undo velar softening to yield /k/. Results for these forms are shown in Table 2. In interpreting this table, recall that the materials did not include non-Latinate targets ending in -*ity*. The non-Latinate targets all involved /k/ or /s/ before -*ness*.

Table 2. Outcomes in the backformation task.

Target type	Hits	Backformations to /k/	% /k/ Responses	% /s/ Responses
Latinate targets	32	6	18	82
Semi-Latinate targets	36	5	13	87

The fact that some reversals of /s/ to /k/ occur is evidence of the psychological reality of the /k/-/s/ alternation. However, the rate of back-formation of /k/ from /s/ is much lower than the rate of velar softening in the noun formation task. Two subjects out of seven were responsible for all cases of backformation of /s/ to /k/. These backformed at rates of 33% and 75%, respectively. (Both produced examples of /k/ for both Latinate and semi-Latinate bases). The finding that backformation to /k/ is less frequent and reliable than softening of /k/ to /s/ requires explanation.

4. Discussion

The experiment showed that the /k/-/s/ alternation is productively applied in an open-response task. However, it not completely productive; it fails to apply to non-Latinate stems and for backformation, there is large variability across subjects. My goal will be to explain this exact pattern of productivity as a reflex of statistical learning over patterns in the lexicon. In exploring this issue, I will make several simplifying assumptions. One is that the relevant probabilities can be approximated over word pairs involving the exact affix in the experiments, *-ity*. This assumption is made because the relevance of exceptions to velar softening involving other affixes is unclear; notably, Ohala's (1974) experimental study of *-ism* reports a much lower (30%) productivity level for Velar Softening, possibly as a consequence of the rather many exceptions involving this suffix. A second assumption is that any given word pair either is, or is not, in the universe over which a probability is defined. I consider only in passing models in which the importance of a word pair is weighted on a scale by its similarity to the current target.

A straightforward extension of previous experimental studies of phonotactics would seek to identify stochastic constraints on word form to which the products of *-ity* affixation must conform. This extension would be in the spirit of Optimality Theory, as well as of many prior studies in prosodic morphology, in seeking to explain morphophonological alternations as the result of constraints that are generally true of the language. However, it is not the case that the consonant preceding word-final /ɪti/ (or the morpheme *-ity*), is usually /s/. In the Collins on-line dictionary, only 25% of words ending in /ɪti/ end in /sɪti/. /l/ is more common than /s/, although even /l/ does not achieve a majority of the forms. Simple frequency matching on the surface forms would predict that subjects would tend to substitute /l/ for /k/ (or for any other consonant!) but only at a rate of about 25%. Surface statistics do not explain the extremely high rate of substitution observed for the Latinate and semi-Latinate stems ending in /k/, or the failure of other consonants to be affected in the same way. Similarly, the responses by two subjects in the back-formation task also appear to reflect implicit knowledge of specific morphological relationships. Without knowledge of such relationships, there would be no reason to backform /s/ to /k/, since /s/ is both more faithful to the stimulus and more common in word-final position.

As a result, knowledge of the alternation must be a generalization over morphologically related word pairs. This conclusion echoes the treatment of velar softening as a derivational rule in Chomsky and Halle (1968). In Optimality Theory, constraints generalizing over word pairs have been used since McCarthy and Prince (1995) proposed them in order to overcome the limitations of constraints over word forms in explaining the behavior of reduplication. Generalizations over word pairs also figure in the non-OT literature on computational morphology, notably Skousen (1989), Daelemans et al. (1999), Ernestus and Baayen (2002), Baayen (2003), and Albright and Hayes (2003). These works all share the assumption that variable outcomes in morphophonology are related to conditional probabilities defined on word pairs.

The acknowledgment that the alternation is learned as a generalization over word pairs goes far towards explaining the amount of variation observed across individuals. Individuals differ both in the size and the contents of their vocabularies. To know a relevant example of an alternation, they must know both words in the pair. Furthermore, they must view them as related to each other. Not everyone infers a decomposition of *Mediterranean* on the basis of words such as *medium* and *terrestrial*. The difficulties of assessing such implicit semantic relationships mean that most computational studies, including the present one, rely on phonological matching in large dictionaries and on morphological analyses by linguists. They probably overestimate the pool of relevant word pairs known to the subject pool.

In the following discussion, I will be particularly concerned with the claim, advanced in the analogical models of Skousen (1989) and Baayen (2003), that the productivity of morphophonological alternations in new forms is determined by a general statistical inference. Unlike Chomsky and Halle (1968), these models draw no fundamental distinction between morphological derivation and back-formation. They therefore make precise predictions in both directions. The pronunciation of an unknown form is inferred from a related known form in the light of a universe of known word pairs exemplifying the same relationship. The models set up analogies in which the unknown variable in the analogy may be either a base or a derived form: they pose equally questions such as *conic* : *conicity* :: *clemic* : *?* and questions such as *conicity* : *conic* :: *clemicity* : *?* . An analogical approach does predict that an alternation may display different rates of productivity for morphological derivation and back-formation, because the probability of A given B, and the probability of B given A are mathematically distinct and often have different values. Exact predictions about the outcome probabilities depend on the exact assumptions about the universe of generalization over which the probabilities are estimated. Thus, the key issue is how the universe of generalization is established, and why the cognitive system takes some universes of generalization to be the operational ones, as opposed to others which are equally available from a mathematical point of view.

The following discussion is premised on three hypotheses about the operational level of generalization, which bring together some of the threads of the literature on morphological processing by people and by machines.

HYPOTHESIS 1: All other things being equal, the cognitive system prefers generalizations which yield more certainty about the outcome to those which yield less certainty.

This claim, a plain language statement of the information-theoretic proposals of Daelemans et al. (1999), means that descriptions of the data which are associated with extreme probabilities are more relevant than ones which characterize the outcome as a random choice. The extreme probabilities of 1.0 and 0.0 provide complete certainty; an outcome with probability 1.0 is the only one possible, and one with probability 0.0 is absolutely impossible. For a two-way choice, a probability of 0.5 represents complete uncertainty, providing no information either way.

Typically, hypothesis 1 will tend to favor generalizations based on small sets of words over generalizations based on bigger sets, as smaller sets tend to be more homogeneous (to exhibit more uniform outcomes) and bigger sets tend to be more heterogeneous (to exhibit more diverse outcomes). But this is not always the case, as the study of learning of English verb morphology by Derwing and Skousen (1994) indicates. They successfully apply Skousen's (1989) AML approach (Analogical Modeling of Language), which anticipates the conclusions of Daelemans et al. (1999) by providing an algorithm for automatically growing smaller analogical sets to bigger ones exactly when the homogeneity of the universe is not compromised. Albright and Hayes (2003) also adopt information-theoretic weighting of generalizations over examples.

HYPOTHESIS 2: All other things being equal, the cognitive system prefers generalizations based on larger sets of examples to those based on smaller sets.

This preference is justified because increasing the sample size increases the reliability of the estimate of the probability of a pattern. In a morphological analyzer directed towards automatic part-of-speech tagging, Mikheev (1997) brings together premises 1 and 2 by assigning rule scores based on the lower edge of the 90% confidence interval for the probability associated with the rule. For rules which positively specify the nature of the outcome, this number increases both with the estimated probability and with the size of the sample from which the probability is estimated.

A further claim made in Mikheev (1997) is:

HYPOTHESIS 3: All other things being equal, longer phonological descriptors are preferable to shorter ones.

This claim is directly at odds with the assumption of classical generative grammar that the structural descriptions in rules should be as simple and general as possible. Notably, the Velar Softening rule in Chomsky and Halle (1968) is maximally general, targeting all biphonemic sequences with certain distinctive features. It states that in derived environments,

(10) [–anterior, –continuant, <–voice >] → [–back, <+anterior>]
/ ___ [–back, –low, –cons]

(10) groups together the /k/-/s/ alternation and the alternation of /g/ with /dʒ/. It neglects potential conditioning by the specific suffix involved, and by the etymological class, length, or structure of the stem. It thus predicts softening in neologisms such as *taskism*. It presupposes that all surface ex-

amples of the /s/ variant before low or back vowels (such as *focus*, *foci*) involve vowel shifting from underlying nonlow front vowels. In the interests of unifying the analyses of pairs such as *opaque*, *opacity*; *critic*, *criticize*; *analogue*, *analogy*, Chomsky and Halle thus pursued an aggressive program of generalization and abstraction and developed a complex theory of rule interaction.

Mikheev's hypothesis receives support from recent experimental work in phonetics and psycholinguistics. Numerous results indicate that cognitive representations are more redundant than was imagined in the early days of generative grammar; see discussion in Baayen (2003), Broe and Pierrehumbert (2000), and Bybee (2001). These results bear on the present data in suggesting that phonological characteristics which are true of all word pairs exhibiting an alternation would be maintained in the general template for that alternation even if they are redundant with respect to the universe over which the generalization has been learned.

Given these hypotheses, the high rate of velar softening for Latinate and semi-Latinate stems can be readily explained. Extant Latinate stems provide a core of multisyllabic stems ending in /ɪk/, for which velar softening occurs at P=1.0 before the affix -*ity*. By hypothesis 1, this is a highly relevant generalization. Expanding the set to include all multisyllabic stems ending in /k/ (e.g to also include the few pairs such as *opaque*, *opacity*, *reciprocate*, *reciprocity*) increases the sample size without compromising the reliability of the generalization, since the probability of softening is still 1.0. However, dropping the shared prosodic features, such that the relevant universe is simply words ending in /k/ with a counterpart in -*ity*, simplifies the description without expanding the sample size. The sample size remains the same because there are no monosyllabic words ending in /k/ with a derivative in -*ity*. By hypothesis 3, then, the generalization stops short of this simplification. Hence, its structural description is not met in novel forms such as *bleckity*, and it fails to apply. A general implication of Mikheev's claim is that morphological alternations would generalize to forms which lie in the cracks between existing forms, so to speak, but not to forms which break entirely new ground.

This analysis does not explain why productivity of velar softening was higher for Latinate than for semi-Latinate forms. If the small observed difference proves replicable, it will provide an example of a prototypicality effect. The Latinate words are extremely similar to the core of the distribution for the alternation, the many forms involving stems in /ɪk/. The semi-Latinate forms are less similar. A trading relationship between similarity and frequency is widely attested, and as discussed by Dell (2000) is a hallmark of connectionist models. However, the same trading relationship is also characteristic of stochastic analogical models. In all of these approaches, the behavior of prototypical examples is observed only part of the time for nonprototypical examples, because the nonprototypical examples can also be captured by the pattern of a competing generalization. For the present case, the competing generalization is either a generalization over multiple affixes (a complexity which exceeds the scope of this paper), or a default, namely that of maintaining the base form consonant without modification. This default is one to which we will return in connection with the backformation data.

Turning now to the backformation data, a generalization at the same level relates words of four or more syllables ending in /sɪti/ to words of two or more syllables ending in any consonant. This universe includes everything in the universe for the noun formation task, plus pairings such as *diverse*, *diversity* and *porous*, *porosity*. The probability of /k/ in the stem given /s/ in the derived form is 0.42 over this universe. The minority of /k/s is due to the greater number of examples such as *porosity*, *immensity*, in which the surface /s/ really does relate to a base /s/. The probability of 0.42 represents considerable uncertainty, very unattractive according to hypothesis 1. A second problem is that the observed rate of backformation to /k/ was much lower.

Some light may be shed on this low rate by considering more carefully the competing outcome, in which the consonant in the backformation is identical to the corresponding consonant in the suffixed from. The universe just discussed did not include pairs such as *liquid*, *liquidity* or *concave*, *concavity*, because these do not end in /sɪti/. At a superficial level of description, introducing such pairs into the universe would lead to extremely heterogenous outcomes, since all the various consonants except /s/ that appear in -*ity* words are faithfully copied from the stem. However, such pairs may be characterized in a homogeneous way by introducing variables. All examples except for the /k/-/s/ pairs involve identity between the stem-final consonant and the consonant appearing before -*ity*. Just as the alpha notation of Chomsky and Halle (1968) can enforce matches between structural description and output, indexing can also enforce points of identity in the reversed direction. The pairs *liquidity*, *liquid* and *concavity*, *concave* are then both examples of the abstract pairing (C_iɪti, C_i). The need for such variables is extensively discussed in Marcus (2001) on the basis of other results on phonological learning and productivity.

Figure 2 shows the observed probability of maintaining the same C in the stem, for complex words ending in -*ity*, as a function of the universe of lexical items used to make the estimate. The leftmost point on the graph corresponds to the universe just discussed, in which the probability of maintaining the same C is 0.58 (that is, 1.0 minus the 0.42 probability of /k/). Successively bigger universes are various supersets of this universe. As shown, the probability increases monotonically as the description of the universe is expanded to include more and more cases. Given that the probability of maintaining the same consonant is already above 0.5 for the smallest set, enlarging the set steadily increases certainty about the outcome.

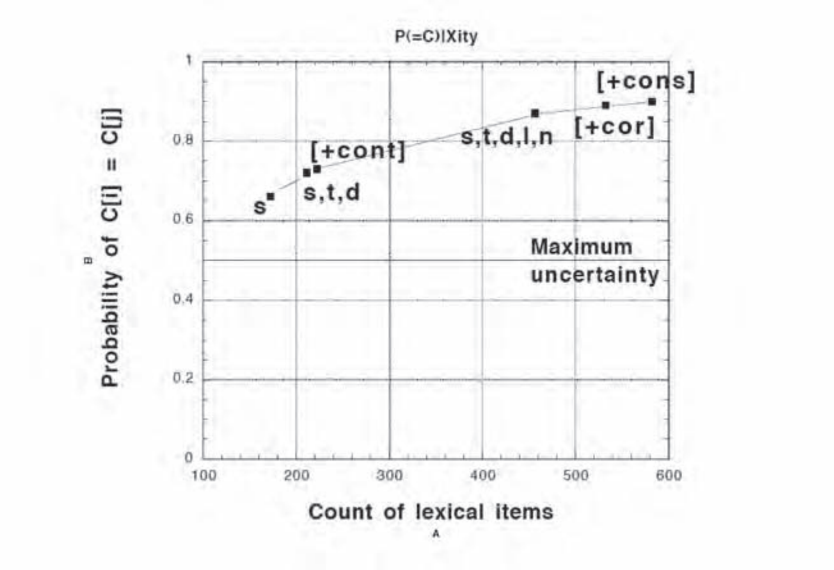


Figure 2. Rate of faithful pairings between the final phoneme preceding the suffix *-ity* and the final phoneme of the stem when this stem occurs as a independent word. The universe over which the rate is computed varies along the x-axis. 1) Complex words ending in /sɪti/, however spelled. 2) Complex words ending in voiceless coronal obstruents (/s/, /t/,/ʃ/,/θ/) plus /ɪti/. Points 1) and 2) appear superimposed because the sets differ by only one word pair. 3) Complex words ending in any coronal obstruent plus /ɪti/. 4) Complex words ending in any fricative (whether coronal or not) plus /ɪti/. 5) Complex words ending in any coronal consonant (including stops, fricatives, and sonorants) plus /ɪti/. 6) Complex words ending in any true consonant (excluding glides) plus /ɪti/. 7) All complex words in which stripping /ɪti/ yields an identifiable stem.

Insofar as enlarging the universe of comparison includes more and more examples of nonalternating (faithful) consonant pairings, this line of reason-ing about backformation yields a bias towards a representation of the base form that is faithfully reflected in the complex form. We may compare it to the OT principle of lexicon optimization proposed in Prince and Smolensky (1993), according to which languages select underlying representations which are maximally harmonic with the surface representations. In the present case, the /s/ candidate is faithful, whereas the /k/ candidate is not. Given that the /s/ candidate is otherwise unproblematic, it would always win over the /k/ candidate. The /k/ candidate could only win if the surface representation of a base form with /k/ were also available to the learner, which was not the situation in the backformation experiment.

Unamended, OT lexical optimization predicts a single outcome, but actually the outcomes are variable. For five individuals, the results are consistent with the suggestion that the maximally inclusive universe of Figure 2 is relevant. This universe predicts 11% /k/s per individual, or 1.7 examples. This is surely within the statistical error of zero, especially if one allows for a low vocabulary level or for the confidence interval calculation proposed by Mikheev (1997). The two individuals who did backform to /k/ produced five examples (out of 15 hits) and six examples (out of eight hits), respectively. The two subjects differed in that one removed only the -*ity*, whereas the second often removed additional material. For *nodacity*, she responded *nodal* and for *runomicity*, *runate*, leaving only eight responses with just the -*ity* removed. Thus, one way of understanding these two subjects is that both worked with a comparatively narrow universe, but one had more active competition from other morphological interpretations of the input forms. Although this explanation is not complete, the statistical inference model at least provides ways by which vocabulary level, morphological awareness, and individual decision-making traits can manifest themselves in variable outcomes.

A second difference between the OT accounts and the present one is that the pressure towards faithfulness which can be read into Figure 2 depends critically on monotonicity of the graph and on the statistics of the most narrowly described universe. Faithfulness is not relevant in all situations, but only in those in which it resolves uncertainty. The outcomes for the noun formation task provide a case in point. Although this task was an exact counterpart to the backformation task, the productivity of the alternation proved to be entirely different; there was at best a sporadic penchant for a faithful outcome. This difference can be explained by noting that in the forward direction, the /k/-/s/ alternation is statistically reliable; expanding the universe of description would only weaken a certain inference. For the backformation task, in contrast, expanding the universe increases certainty.

Given this line of argument, it is important to determine why the analogy set based on words in -*icity* /ˈɪsɪti/ did not appear to be active or relevant in the backformation experiment. This descriptor is more specific than those used to calculate the figure, and the set it describes has a baseform ending in /k/ with P = 1.0. This generalization reliably covers both words ending in the suffix -*ic* and words ending in /ɪk/ which are not synchronically decomposable, as in (11).

(11) eccentric/eccentricity (*eccenter)
plastic/plasticity (*plast)
public/publicity (*puble)
rustic/rusticity (*rust)
toxic/toxicity (*tox)

If this set were operational, the experimental results should have displayed a strong distinction between the Latinate targets (which would backform to /k/) and the others (for which /s/ would be the more probable outcome).

The failure of words in -*icity* to support a reliable pattern of backformation casts doubt on the form of hypothesis 3. Even if moderately long descriptions are preferable to short ones, very long descriptions might not be preferable to long ones. Consider the joint effect of hypotheses 2 and 3. Extremely detailed descriptions tend to be statistically unstable, since they pertain to so few cases that their statistics are not robust across individual differences in vocabulary; this point is developed in more depth in Pierrehumbert (2001). The joint pressure towards detailed descriptions and large sample sizes could mean that the best entry level for generalizations over phonological patterns is moderately detailed – more detailed than the simplest descriptions that Mikheev dismisses, but still more broadly applicable than the worst cases considered by Pierrehumbert. These entry-level descriptions are then further generalized when the generalization increases certainty about the outcome, as discussed above.

5. Conclusion

In conclusion, the /k/-/s/ alternation was found to be highly productive in noun formation, and some evidence of its psychological reality is also found in backformation. An approach based on statistical inference over word pairs enjoys considerable success in explaining the outcomes. A key assumption is that the universe of comparison grows more general to provide a critical mass of examples and to reduce uncertainty in the predicted outcome. A second key to success is the idea that word pairings can involve variables, so that pairings in which the same consonant is preserved no matter what its character can act together in influencing the outcome.

Explaining why /k/ is preserved in forms such as /blɛk/ – /blɛkɪti/ requires the assumption that in abstracting over a universe of examples, the cognitive system prefers to maintain somewhat rich and redundant descriptions. Abstractions are not simplified beyond what is required by differences amongst the examples. At the same time, the preservation of /s/ in some backformations such as /mtɹpanɪsɪti/ – /mtɹpanəs/ indicates some limits on how fine-grained and detailed abstractions can be. The behavior of these cases suggests that arbitrary phonological descriptors are not actually relevant to forming the universe for the morphophonological inference. Instead, there may be a privileged degree of granularity in analysis, reminiscent of the basic level of categorization which privileges the concept DOG over DALMATION or ANIMAL. Initial generalizations made at this level can then be adjusted upward or downward to achieve more certainty about the outcomes.

Acknowledgments

I'm very grateful to the Spring 2001 students in Laboratory Phonology II who participated in the initial development of the stimuli for this experiment as a laboratory exercise: Martin Armstrong, Jungwon Choy, Erin McMahon, Jessica Peterson, Rebecca Schwarzlose, and Xanthe Wigfall.

Appendix

(1) Baseline Items

NESS-WORDS: calm, calmness; correct, correctness; happy, happiness; icy, iciness; kind, kindness; light, lightness; lonely, loneliness; sweet, sweetness.

ITY-WORDS: abnormal, abnormality; captive, captivity; inane, inanity; liquid, liquidity; profane, profanity; valid, validity; virile, virility.

(2A) Word Fillers

USED WITH -NESS IN BACKFORMATION TASK: arid, frugal, human, odd.

USED WITH -ITY IN BACKFORMATION TASK: arboreal, binomial, liberal, real.

[2B] Nonword fillers

USED WITH -NESS IN BACKFORMATION TASK

Spelling	*IPA*	*Spelling*	*IPA*
clipid	klˈɪpɪd	clipidness	klˈɪpɪdnəs
demarte	dəmˈaɹt	demarteness	dəmˈaɹtnəs
flader	flˈedɹ	fladerness	flˈedɹnəs
mastive	mˈæstɪv	mastiveness	mˈæstɪvnəs

USED WITH -ITY IN BACKFORMATION TASK

Spelling	*IPA*	*Spelling*	*IPA*
bordal	bˈɔɹdl	bordality	bˌɔɹdˈælɪti
caustive	kˈɔstɪv	caustivity	kˌɔstˈɪvɪti
justical	dʒˈʌstɪkl	justicality	dʒˌʌstɪkˈælɪti
tromucal	trˈamjukl	tromucality	trˌamjukˈælɪti

[3A] "Latinate" targets ending in /ɪk/

Spelling	*IPA*	*Spelling*	*IPA*
clemic	klˈɛmɪk	clemicity	klˌɛmˈɪsɪti
criotic	kɹˌajˈatɪk	crioticity	kɹˌajətˈɪsɪti
extric	ˈɛkstɹɪk	extricity	ˌɛkstɹˈɪsɪti
hytronic	hˌajtɹˈanɪk	hytronicity	hˌajtɹənˈɪsɪti
interponic	ˌɪntɹpˈanɪk	interponicity	ˌɪntɹpanˈɪsɪti
malatonic	mˌælətˈanɪk	malatonicity	mˌælətənˈɪsɪti
phynomic	fˌajnˈomɪk	phynomicity	fˌajnəmˈɪsɪti
runomic	ɹˌunˈamɪk	runomicity	ɹˌunəmˈɪsɪti

[3B] "Semi-Latinate" targets ending in a secondary stressed syllable not construable as *-ic*.

Spelling	*IPA*	*Spelling*	*IPA*
bowdec	bˈodˌɛk	bowdecity	bˌodˈɛsɪti
hovac	hˈovˌæk	hovacity	hˌovˈæsɪti
nodac	nˈodˌæk	nodacity	nˌodˈæsɪti
pavoc	pˈævˌak	pavocity	pˌævˈasɪti
solvoc	sˈalvˌak	solvocity	sˌolvˈasɪti
stanorac	stˈænəɹˌæk	stanoracity	stˌænəɹˈæsɪti
strenoc	stɹˈɛnˌak	strenocity	stɹˌɛnˈasɪti
trylec	tɹˈajlˌɛk	trylecity	tɹˌajlˈɛsɪti

[3C] "Non-Latinate" targets ending in /k/

Spelling	*IPA*	*Spelling*	*IPA*
bleck	blˈɛk	bleckness	blˈɛknəs
mork	mˈɔɹk	morkness	mˈɔɹknəs
over-glique	ˌovɹglˈik	over-gliqueness	ˌovɹglˈiknəs

shruk	ʃɹˈʌk	shrukness	ʃɹˈʌknəs
snilk	snˈɪlk	snilkness	snˈɪlknəs
toque	tˈuk	toqueness	tˈuknəs
twake	twˈek	twakeness	twˈeknəs
under-grack	ˌʌndɹgɹˈæk	under-grackness	ˌʌndɹgɹˈæknəs

[3D] "Non-Latinate" targets ending in /s/

Spelling	*IPA*	*Spelling*	*IPA*
blarse	blˈaɹs	blarseness	blˈaɹsnəs
deploose	dəplˈus	deplooseness	dəplˈusnəs
dwess	dwˈɛs	dwessness	dwˈɛsnəs
druss	dɹˈʌs	drussness	dɹˈʌsnəs
jace	dʒˈes	jaceness	dʒˈesnəs
melse	mˈɛls	melseness	mˈɛlsnəs
queece	kwˈis	queeceness	kwˈisnəs
under-dass	ˌʌndɹdˈæs	under-dassness	ˌʌndɹdˈæsnəs

References

Albright, Adam and Bruce P. Hayes
2003 Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition* 90: 119–161.
Anderson, Stephen R.
1981 Why phonology isn't natural. *Linguistic Inquiry* 12: 493–539.
Baayen, R. Harald
2003 Probabilistic approaches to morphology. In: Rens Bod, Jennifer Hay and Stefanie Jannedy (eds.), *Probabilistic Linguistics*, 229–288. Cambridge, MA: MIT Press.
Baayen, R. Harald, R. Piepenbrock and L. Gulikers
1995 *The CELEX lexical database (release 2) CD-ROM*. Philadelphia, PA: Linguistic Data Consortium.
Berko, Jean
1958 The child's learning of English morphology. *Word* 14: 150–177.
Broe, Michael and Janet Pierrehumbert
2000 Introduction. In: Michael Broe and Janet Pierrehumbert (eds.), *Papers in Laboratory Phonology V: Acquisition and the Lexicon*, 1–8. Cambridge: Cambridge University Press.
Bybee, Joan
2001 *Usage-Based Phonology*. Cambridge: Cambridge University Press.

Bybee, Joan and Elly Pardo
1981 Morphological and lexical conditioning of rules: Experimental evidence from Spanish. *Linguistics* 19: 937–968.
Chomsky, Noam and Morris Halle
1968 *The Sound Pattern of English*. New York: Harper and Row.
Cutler, Anne and T. Otake
1998 Assimilation of place in Japanese and Dutch. In: *Proceedings of the Fifth International Conference on Spoken Language Processing*, 1751–1754. Sydney.
Daelemans, Walter, Jakub Zavel, Ko van der Sloot and Antal van den Bosch
1999 TiMBL: Tilburg memory-based learner version 2.0 reference guide. http://citeseer.ist.psu.edu/article/daelemans99timbl.html.
Dell, Gary S.
2000 Commentary: Counting, connectionism, and lexical representation. In: Michael Broe and Janet Pierrehumbert (eds.), *Papers in Laboratory Phonology V: Acquisition and the Lexicon*. Cambridge: Cambridge University Press.
Derwing, Bruce and Royal Skousen
1994 Productivity and the English past tense: Testing Skousen's analogy model. In: S. D. Lima, R. L. Corrigan and G. K. Iverson (eds.), *The Reality of Linguistic Rules*, 193–218. Amsterdam: John Benjamins.
Ernestus, M. and R. Harald Baayen
2002 The functionality of incomplete neutralization in Dutch: The case of past-tense formation. Paper presented at: *Eight Conference on Laboratory Phonology*. New Haven, CT.
Frisch, S., Janet Pierrehumbert and Michael Broe
2004 Similarity avoidance and the OCP. *Natural Language and Linguistic Theory* 22: 179–228.
Goldrick, M.
2002 Probabilistic phonotactic constraints and sub-segmental representations: Evidence from speech errors. Paper presented at: *Eight Conference on Laboratory Phonology*. New Haven, CT.
Guion, S.
1998 The role of perception in the sound change of velar palatalization. *Phonetica* 55: 18–52.
Hay, Jennifer, Janet Pierrehumbert and Mary E. Beckman
2004 Speech perception, well-formedness, and the statistics of the lexicon. In: John Local, Richard Ogden and Rosalind Temple (eds.), *Papers in Laboratory Phonology VI*, 58–74. Cambridge: Cambridge University Press.

Jaeger, Jeri J.
1984 Assessing the psychological reality of the Vowel Shift Rule. *Journal of Psycholinguistic Research* 13: 13–36.
Kiparsky, Paul
1985 Some consequences of lexical phonology. *Phonology Yearbook* 2: 85–138.
Lavoie, Lisa M.
2001 *Consonant Strength: Phonological Patterns and Phonetic Manifestations.* New York: Routledge.
Marcus, Gary
2001 *The Algebraic Mind: Integrating Connectionism and Cognitive Science (Learning, Development and Conceptual Change)*. Cambridge, MA: MIT Press.
McCarthy, John and Alan Prince
1995 Faithfulness and reduplicative identity. ROA-60, Rutgers Optimality Archive, http://roa.rutgers.edu.
McCawley, J. D.
1986 Today the world, tomorrow phonology. *Phonology Yearbook* 3: 27–45.
Mikheev, A.
1997 Part-of-speech guessing rules: Learning and evaluation. *Computational Linguistics* 23: 405–423.
Myers, James
1999 Lexical phonology and the lexicon. ROA-330-0699, Rutgers Optimality Archive, http://roa.rutgers.edu.
Ohala, John J.
1974 Experimental historical phonology. In: J. M. Anderson and C. Jones (eds.), *Historical Linguistics II: Theory and Description in Phonology*, 353–389. Amsterdam: North Holland.
Pierrehumbert, Janet
2001 Why phonological constraints are so coarse-grained. In: James M. McQueen and Anne Cutler (eds.), *Spoken Word Access Processes (SWAP)*. Psychology Press.
2003 Probabilistic Phonology: Discrimination and Robustness. In: Rens Bod, Jennifer Hay and Stefanie Jannedy (eds.), *Probabilistic Linguistics*. Cambridge, MA: MIT Press.
Prince, Alan and Paul Smolensky
1993 Optimality theory: Constraint interaction in generative grammar. *Technical Report TR-2, Rutgers University Cognitive Science Center*.
Skousen, Royal
1989 *Analogical Modelling of Language*. Dordrecht: Kluwer Academic Publishers.

Steriade, Donca
2002 Commentary. Paper presented at: *Eight Conference on Laboratory Phonology*. New Haven, CT.

Whalen, Douglas H.
in press How the study of endangered languages will revolutionize linguistics. In: Piet van Sterkenburg (ed.), *Linguistics Today*. Amsterdam: John Benjamins.

Zuraw, Kie
2000 Patterned exceptions in phonology. Ph.D. dissertation, UCLA.

Modeling intonation in English: A probabilistic approach to phonological competence*

Audra Dainora

This paper presents an empirically based autosegmental model of intonation in which tone sequences are states, and transitions between states are expressed as probabilities. The model's parameters are estimated using a large corpus of radio news stories read in standard American English by professional announcers. Empirical results include comprehensive statistics on the distribution of pitch accents, phrasal tones, boundary tones, intermediate phrases, and tone sequences within this corpus. The findings lead me to argue for the inclusion of L+H in the tonal inventory of English and that pitch accents and boundary tones are not chosen independently. The method developed may be applied to probabilistic linguistic phenomena in a variety of speaking styles and languages.*

1. Introduction

Increasingly, empirical research suggests phonological competence involves probabilistic knowledge of how sounds pattern together. Coleman and Pierrehumbert (1997) favored probabilistic generative grammars after finding that the well-formedness of a word depends on the frequency of its parts. Neologisms that combine nonoccurring syllables with high-frequency syllables may be perceived as better formed than words created from acceptable but low-frequency syllables; such results do not conform with standard optimality theory or standard generative theory. Subsequent work by Frisch, Large, and Pisoni (2000), Treiman et al. (2000), Hay, Pierrehumbert, and Beckman (2004), and others explores whether people use probabilities in language processing by examining the impact of phonological frequency on acceptability judgments.

If processing involves probabilistic aspects of language, the relevant statistical properties must be documented to fully develop theories of competence. I report how applying a statistically based methodology to a large data set exemplifies one type of phonological competence: the way tones pattern together to form intonational tunes. A complete model of intonational structure requires (1) a grammar that describes the building blocks of an intonational phrase and the rules and constraints involved in the construction of well-formed phrases, (2) estimation of the frequencies associated with the use of different tones and tunes, and (3) understanding why speakers choose one tone or tune over another.

The seminal work of Pierrehumbert (1980), Ladd (1983), Beckman and Pierrehumbert (1986) and others established the grammar for English intonation. This paper extends these models by adding the second component: estimating the probabilities of various intonational phrases and their tones from data. Pierrehumbert's finite-state grammar of intonation (shown in Figure 1) is nonstochastic in that it does not contain information about transition probabilities between tones. Which tunes occur more often? Does choice of pitch accent affect which other pitch accents occur? Does choice of pitch accent affect choice of phrasal or boundary tones or vice versa? How many pitch accents are expected in an intermediate phrase? How many intermediate phrases are expected in an intonational phrase?

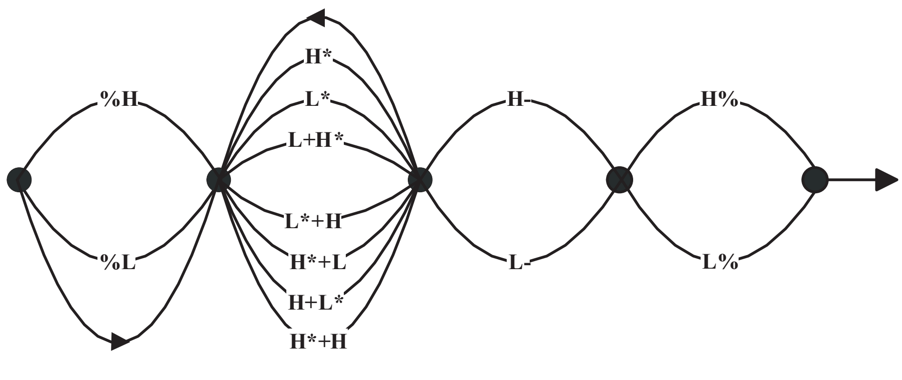


Figure 1. Tunes in English in the Pierrehumbert model (adapted from Pierrehumbert 1980: 13) are formed from an optional initial boundary tone, one or more pitch accents, a phrasal tone (called phrasal accent), and a boundary tone.

The third component – understanding all factors that generate intonational patterns – is a goal that research can approach only asymptotically. Relevant factors may include semantic content, syntactic structure, discourse structure, phonetic factors, and, finally, random error in the tone production process.

Probabilistic models have several advantages over nonstochastic ones. Interaction between various factors causing intonational choice may be too complex to model perfectly, but a probabilistic approach allows a partial identification of the combined effect of these factors without isolating individual contributions. Additionally, probabilistic grammars can account for certain types of phenomena by showing that their existence or nonexistence is the direct result of particular distributional patterns, thus limiting the use of rules or constraints to essentials. Looking for phonological patterns involves both assigning probabilities and finding data for which entropy is not maximal – sequences of tones in which the frequency of the sequence differs from the product of the frequencies of the tonal components. By illustrating the utility of the probabilistic approach I aim to increase awareness of a valuable tool for studying phonological phenomena.

Section 2 describes the corpus and methodology. Section 3 establishes facts regarding the distribution of tones in one large corpus of English monologue narratives and uses the data to estimate parameters for a probabilistic model of intonation, presented in Section 4. Section 5 discusses the model's implications for intonational theory and points out areas in which a probabilistic approach provides new insights. Section 6 concludes with suggestions for future research.

2. Data and methodology

The analyses concern how tones are related on the level of symbolic representation – which patterns exist in the distributions of pitch accents, edge tones, and intermediate phrases. The identities of pitch accents in various structural positions are considered, comparing the position of *nuclear accent* (the final pitch accent of an intonational phrase) with that of *prenuclear accent* (any pitch accent preceding this final pitch accent). With intonational phrases, focus is on the nuclear accent, the phrasal tone, and the boundary tone – referred to as the *nuclear tune* – since those three tones form the minimal requirement for an intonational phrase. Examining such sequences allows direct comparison of all intonational phrases in the data, despite variation in numbers of pitch accents and intermediate phrases contained within.

My focus on the nuclear tune reflects a hypothesis that the bulk of semantic work in intonational meaning is carried by the final three (or four) tones in each intonational phrase. Ladd (1996: 208) suggested "we want to be able to treat the final [...] sequence [of pitch accent, phrasal tone, boundary tone], which occurs only once, separately from the accent or accents that precede it."

```
              H*     H*L-L%          H*        H*     H*L-L%
(1)   a.  I read it to Julia.     b. I wanted to read it to Julia.
                                                  (Ladd 1996: 208)
```

A declarative contour, such as in (1), could thus be specified as zero or more H* pitch accents followed by an H*L-L% sequence.

2.1. Data

This study is based on 1,207 intonational phrases (over 6,300 tones) from the Boston University Radio Speech Corpus (Ostendorf, Price, and Shattuck-Hufnagel 1996), which contains news broadcasts read by professional radio announcers. Informal observation suggests the newscasters used a prosodic style that includes numerous pitch accents, especially H* and L+H*, as well as frequent phrasal boundaries, which serve to highlight important information and delineate long sentences. Ostendorf, Price, and Shattuck-Hufnagel (1995: 3) characterize the style as "natural but controlled," noting "there is evidence that these newscasters use more clear and consistent indications of prosodic structure than nonprofessional read speech."

The analyses are based on lab news data from Speakers F1A and F2B and radio news data from Speaker F1A. The data are labeled with prosodic markers using the ToBI transcription system (Beckman and Ayers 1994). Dainora (2001: 23–26) provides expanded description of the data set. Data analysis preparation included collapsing downstepped and nondownstepped pitch accents into a single category, given Dainora's finding (2001: 46–69) that high pitch accents and downstepped pitch accents in the corpus are drawn from the same distribution – the frequency change in pairs of peak tones labeled downstep appears to be the right-hand tail of a normal distribution; the drop in pairs not so labeled appears to be the left-hand tail of the same normal distribution. Thus sequences such as H*!H*L-L% and H*H*L-L% are treated as identical. Similarly, downstepped and nondownstepped phrasal tones were collapsed.

2.2. Methodology

I study relations between tones using tools common in probability theory and computational linguistics (see e.g. Charniak 1993, Manning and Schütze 2000, and Goldsmith 2002). For each possible three-tone sequence, I compute the *unconditional frequency* distribution. The unconditional frequency is estimated by computing the fraction of tones in a particular category that have a particular value. I also compute the *conditional frequency* distribution for a variety of possible subsequences. Conditional frequency is computed similarly to unconditional frequency but calculates the ratios of outcomes considering only the subset of phrases meeting a particular requirement. The conditional probability of *X* given *Y* refers to the chance that event *X* will occur given that event *Y* has occurred. This can be written mathematically as:

$$P(X \mid Y) = \frac{P(X \& Y)}{P(Y)}$$

For example, the conditional probability of H% given H*L- is .41, which is computed by dividing the number of phrases ending in H*L-H% (276) by the number of phrases ending in H*L-T% (674), where T% can be any boundary tone.

Statisticians have devised methods for combining information in several conditional probabilities into a single number summarizing the relevant information. One method is *pointwise mutual information*, which is obtained by dividing the probability of two tonal sequences occurring together (in any order) by the probabilities of each sequence occurring; thus:

$$I(x, y) = \frac{P(x \& y)}{P(x)P(y)}$$

Traditionally researchers have applied a log transformation as a matter of mathematical convenience (e.g., Manning and Schütze 2000: 68), which gives:

$$I(x, y) = \log \frac{P(x \& y)}{P(x)P(y)}$$

Pointwise mutual information is useful with binary events (e.g., boundary tones, which have only two values: H% and L%). When many tonal sequences are being considered, pointwise mutual information may produce a large array of numbers. To find a single number, one uses *mutual information,* a weighted average of pointwise mutual information for every possible combination of tonal sequences. The weights used in the weighted average are given by the probability of each combination of sequences occurring. The formula for mutual information, where the summation is over all *x,y* pairs, is:

$$\sum P(X \& Y) \log \frac{P(X \& Y)}{P(X)P(Y)}$$

3. Estimating the parameters

The probabilistic grammar approach assumes both common and uncommon observations inform the underlying grammar. One must be able to separate frequently occurring phenomena from infrequently occurring phenomena in a quantifiable way. A standard approach is for the model to have *parameters*: unknown constants estimated from data. The parameters for a probabilistic model are estimated to determine the distribution of tones in this corpus.

3.1. Distribution of tunes

How often does each of the tunes occur in the data set? Table 1 lists the twenty possible sequences of nuclear tunes – the final three tones of each intonational phrase – and their frequency of occurrence.

The data set displays 18 of 20 possible nuclear tunes occurring. H*L-L% and H*L-H% represent over half the sequences, and the next most common sequences are L+H*L-H% and L+H*L-L%. These four tunes comprise almost 80% of the nuclear tunes in the data set. If one considers the distribution of nuclear tunes conditional on the existence or nonexistence of a prenuclear accent, the distribution facts remain largely the same. The only nuclear tune for which frequency of appearance changes is H*L-H%, which occurs more often in phrases lacking prenuclear accents than in phrases with such (28% versus 20%).

Does the distribution change when both the presence and type of prenuclear pitch accent are taken into account? Of 100 possible types of sequences containing at minimum a prenuclear pitch accent, a nuclear pitch accent, a phrasal tone, and a boundary tone, 44 occur in the data. Most common are H*H*L-L% and H*H*L-H%, which comprise 41% of the data. Next most common are H*L+H*L-H% and H*L+H*L-L%, which comprise an additional 17%. (Table 5.2 in Dainora [2001: 80–81] shows a complete frequency distribution for these tunes.) Given that 18 of the 20 three-tone sequences occur, but only 44 of the 100 four-tone sequences occur, this suggests constraints on how prenuclear and nuclear accents combine.

Table 1. Frequency distribution of last three tones of the intonational phrase

Nuclear tune	Occurrences	Overall frequency of occurrence (%)
H*L-L%	398	33
H*L-H%	276	23
L+H*L-H%	158	13
L+H*L-L%	116	10
L*L-H%	75	6
H+!H*L-L%	57	5
H*H-L%	47	4
L+H*H-L%	23	2
H+!H*L-H%	20	2
L*L-L%	14	1
H*H-H%	9	1
L*+HL-L%	4	<.5
L+H*H-H%	3	<.5
L*H-L%	2	<.5
L*H-H%	2	<.5
L*+HH-L%	1	<.5
L*+HL-H%	1	<.5
H+!H*H-L%	1	<.5
H+!H*H-H%	0	0
L*+HH-H%	0	0

A major contributing factor of the absence of some tunes is the rarity of certain individual tones. For example, H+!H* constitutes only 6% of pitch accents in this sample. Consequently, tunes that contain H+!H* are relatively rare; of 36 possible four-tone sequences containing H+!H*, 24 do not appear in the sample. This illustrates the power of the probabilistic approach; a simple recognition of the rarity of H+!H* does much to explain the frequencies of all 36 tunes that contain it, without invoking any explicit constraints prohibiting the unattested sequences. In general, probabilistic models enable parsimonious explanation of numerous phenomena in the data.

Analyses of this type may provide a way of addressing unresolved issues in prosody research. For example, one controversy concerns whether intonational meaning arises from individual tones or from the tune as a whole. Consider the fall-rise contour L*+HL-H%, which indicates "up-in-the-airness" for Bolinger (p.c., in Ward and Hirschberg 1985: 751), making it possible to use the tune while responding to questions such as in (2a) but not as in (2b).

(2) a. What interesting people came to the party?
b. Who came to the party?
(Adapted from Ward and Hirschberg 1985: 752)

While Ward and Hirschberg (1985) consider the tune as the unit signaling speaker uncertainty with respect to a scale, Pierrehumbert and Hirschberg (1990) view the tones of the tune as three separate components, each with its own meaning. If the whole-tune approach is correct, one would expect a tune such as L*+HL-H% to be more common than a tune (such as L*+HH-H%) that has similarly common components but no special meaning of its own. Unfortunately, this data set provides only one example of L*+HH-L%, no examples of L*+HH-H%, and the pitch accent L*+H occurs only six times overall. While the probabilistic approach should shed light on whether meaning emerges from the pitch accent or the tune, in this case the data are insufficient to the task: either the nature of the data (radio announcer speech) is unlikely to provide relevant examples, or the data set is too small to contain a statistically useful number of examples of these potentially rare tunes. Dainora (2002) contains additional discussion of compositionality and intonational meaning.

3.2. Intermediate phrases

Next, consider the distribution of intermediate phrases. Most intonational phrases contain only one or two intermediate phrases, as illustrated in Table 2. For example, row 1 indicates a 65% chance that an intonational phrase contains only one intermediate phrase (with the intermediate phrase and intonational phrase overlapping) and a 35% chance that an intonational phrase contains additional intermediate phrases. In principle, the probability of starting a new intermediate phrase could depend on how many phrases have preceded. The data, however, show extremely consistent numbers: .35, .27, .28, .29. (Beyond this point, there are too few examples for accurate estimation.) Consequently, I simplify the model by estimating a single parameter (.33), which is the probability of a new phrase starting regardless of the number of previous phrases. Note the elegant simplicity the probabilistic approach allows. A simple rule predicts that shorter phrases will be more common and that extremely long sequences of intermediate phrases may never occur in any sentence; the grammar does not prohibit such, but a long sequence of relatively improbable events transpiring is unlikely. Thus, probabilistic analysis reveals how data features may be explained without additional grammatical constraints.

Table 2. Frequency of distribution of number of intermediate phrases in intonational phrases

Number of intermediate phrases	Occurrences in sample of 1207 phrases	Frequency of occurrence (%)	Intonational phrase moves to another intermediate phrase (%)
1	790	65	35
2	305	25	27
3	81	7	28
4	22	2	29
5	8	1	11
6	1	<.5	0
7	0	0	no data

3.3. Distribution of pitch accents

Analysis of all intermediate phrases in the data set reveals that the majority contain only one or two pitch accents, as shown in Table 3.

Column 4 shows the likelihood that a speaker accents additional syllables in the phrase. An intermediate phrase with one pitch accent is very likely to move to a second pitch accent, but a phrase with two or more pitch accents is only about 31% likely to add another accent. The data are distributed much the way they would be if tones were generated randomly. When a coin is tossed, the probability of "heads" is the same for each sequential flip regardless of how many flips have preceded. Similarly, the probability of a subsequent pitch accent being added to an intermediate phrase is unaffected by its sequence position, after the second accent. This does not prove that a random process generated the data; a complex set of rules could produce data that look very much like what would be produced randomly. The nature of such a rule set is unclear – the research focus has been on determining which words may be accented or deaccented and predicting the identity of such accents, not on accounting for variability in the number of accents.

Table 3. Frequency distribution of number of sequential pitch accents within all intermediate phrases

Number of pitch accents in a row (within an intermediate phrase)	Occurrences within 1775 intermediate phrases	Frequency of occurrence (%)	Intermediate phrase moves to another pitch accent (%)
1	690	39	61
2	732	41	33
3	262	15	26
4	67	4	26
5	15	1	38
6	9	1	0
7	0	0	no data

Table 3 data are detailed in Table 4, which compares the distribution of pitch accents in *final intermediate phrases* (the last intermediate phrase of each intonational phrase) versus *nonfinal intermediate phrases* (any preceding phrase).

Table 4. Frequency distribution of number of sequential pitch accents within nonfinal and final intermediate phrases

Nonfinal intermediate phrase				Final intermediate phrase			
Number of pitch accents	Occurrences within 568 nonfinal intermediate phrases	Frequency of occurrence (%)	Intermediate phrase moves to another pitch accent (%)	Number of pitch accents	Occurrences within 1207 final intermediate phrases	Frequency of occurrence (%)	Intermediate phrase moves to another pitch accent (%)
1	303	53	47	1	387	32	68
2	201	35	24	2	531	44	35
3	52	9	19	3	210	17	25
4	9	2	25	4	58	5	18
5	2	<.5	33	5	13	1	1
6	1	<.5	0	6	8	1	0
7	0	0	no data	7	0	0	no data

Nonfinal intermediate phrases most likely contain one pitch accent (53%), whereas final intermediate phrases most likely contain two pitch accents (44%). Columns 4 and 8 indicate the likelihood that at a certain accentual position the speaker will produce another pitch accent instead of producing a phrasal tone and subsequently ending the phrase. The values in column 8 steadily decrease; the more accents an intermediate phrase contains, the more likely it is to finish. Column 4 values do not exhibit this pattern. Thus, I refine the Table 3 finding that the probability of moving to a new pitch accent was steady from the second pitch accent onward; Table 4 clarifies that pitch accent transition probabilities are relatively constant in nonfinal intermediate phrases, but that in final intermediate phrases, the probability of transitioning to a new tone declines rapidly, making long sequences of tones in final phrases rare.

The longest strings of pitch accents within an intermediate phrase in the data set contain six tones. The H*L+H*H*H*H*H* sequence shown in (3) is one example.

```
               H*      L+H*        L-H% H*       H*L-H%
(3)         The case of accused police killer Albert Lewin

            H*   L+H* H*       H*                  H*    H*L-L%
         is going back down to Massachusetts Superior Court.
            (Ostendorf, Price, and Shattuck-Hufnagel 1996, file F1as08p2)
```

Once a speaker starts a phrase with a particular pitch accent, will he keep using the same type of accent throughout the phrase? Examination of intermediate phrases containing two or more pitch accents shows that 50% contain strings of identical accents. The three phrases containing five identical accents are all of the type H*H*H*H*H*. The one phrase containing six identical pitch accents also contains only H* pitch accents. In intermediate phrases containing exactly two pitch accents, the accents are the same 58% of the time. The apparent tendency of pitch accents to be identical, however, masks the true cause – the preponderance of H* accents, shown in Table 5.

Next, consider different combinations of pitch accents. Table 6 shows the conditional probability of a pitch accent occurring given an earlier pitch accent in the string. The data shown are for all intermediate phrases.

Table 5. Frequency distribution of all pitch accents in data set

Pitch accent	Occurrences	Overall frequency of occurrence (%)
H*	2335	70
L*	141	4
L+H*	676	20
L*+H	20	1
H+!H*	166	5

Table 6. Number of occurrences, (conditional probabilities), and {pointwise mutual information} for pitch accent pairs in all intermediate phrases (Tone 1, vertical axis; Tone 2, horizontal axis)

Tone	H*	L*	L+H*	L*+H	H+!H*
H*	784 (66%) {0.052}	48 (4%) {–0.341}	283 (24%) {–0.029}	6 (1%) {0.122}	61 (5%) {–0.187}
L*	17 (52%) {–0.202}	12 (36%) {1.853}	1 (3%) {–2.094}	1 (3%) {1.911}	2 (6%) {–0.025}
L+H*	132 (50%) {–0.239}	26 (10%) {0.515}	85 (32%) {0.311}	0 (0%) {no data}	19 (7%) {0.154}
L*+H	5 (45%) {–0.327}	4 (36%) {1.853}	2 (18%) {–0.302}	0 (0%) {no data}	0 (0%) {no data}
H+!H*	46 (65%) {0.027}	0 (0%) {no data}	10 (14%) {–0.558}	0 (0%) {no data}	15 (21%) {1.224}

These conditional probabilities reveal disparate relations between pitch accents. For example, comparing the conditional probabilities for H*L* and L*L* shows that H* has only a 4% chance of being followed by L*, whereas L* has a 36% chance of being followed by L*. Average mutual information is a relatively high positive number (.0355), indicating accent type choice for adjacent accents is related; the identity of a pitch accent provides considerable information about the identity of an adjacent pitch accent.

I also examined whether the probability that a given pitch accent will follow another changes when they are in the final intermediate phrase of the intonational phrase. Conditional probabilities for pitch accents pairs in nonfinal and final intermediate phrases are very similar. Excluding the two rare tones L*+H and H+!H* (which appear in only five accent pairs), the correlation between the two sets of data achieves 96%. Knowing that pitch accent pairs have very similar distributions in nonfinal and final intermediate phrases allows for the creation of a simpler intonational model without sacrificing explanatory power.

3.4. Nuclear and prenuclear accents

Consider whether speakers use the same pitch accents in nuclear accent position as they do in prenuclear accent position, as indicated by Beckman and Pierrehumbert (1986: 286). Table 7 compares the last pitch accent of the intonational phrase with other pitch accents occurring in the same final intermediate phrase.

Table 7. Frequency distribution of nonfinal and final pitch accents in final intermediate phrases

	Nonfinal pitch accents		Final pitch accents	
Tone	Occurrences	Overall frequency of occurrence (%)	Occurrences	Overall frequency of occurrence (%)
H*	899	74	730	60
L*	31	3	94	8
L+H*	214	18	299	25
L*+H	10	1	6	0
H+!H*	60	5	78	6

There are significant differences between the frequencies of different accents in the two positions. While H* constitutes 74% of tones in final prenuclear position, it constitutes only 60% of nuclear accents. Based on 1,207 phrases, this result is significant at the .01 level. Since there are fewer H* accents in nuclear position, this difference must be made up by the other tones: both L+H* and L* are significantly more common in nuclear position than in prenuclear position. L+H* makes up 25% of nuclear accents but only 18% of prenuclear accents. L* is more than twice as common in nuclear position than in prenuclear position – 8% versus 3%.

Would transitional probabilities between pitch accent pairs suggest that prenuclear accents differ from nuclear accents? Table 8 shows transitional probabilities between pairs of penultimate and ultimate pitch accents in final intermediate phrases and between pairs of nonultimate pitch accents in the same phrases.

Table 8. Number of occurrences and (conditional probabilities) for pitch accent pairs in final intermediate phrases (Tone 1, vertical axis; Tone 2, horizontal axis)

	Pairs of nonultimate pitch accents					Pairs of penultimate and ultimate pitch accents				
Tone	H*	L*	L+H	L*+H	H+!H*	H*	L*	L+H	L*+H	H+!H*
H*	218 70%	6 2%	68 22%	2 1%	19 6%	363 62%	35 6%	154 26%	3 1%	31 5%
L*	0 0%	2 50%	1 25%	0 0%	1 25%	16 59%	9 33%	0 0%	1 4%	1 4%
L+H*	38 63%	4 7%	13 22%	0 0%	5 8%	71 46%	21 14%	52 34%	0 0%	10 6%
L*+H	2 100%	0 0%	0 0%	0 0%	0 0%	2 25%	4 50%	2 25%	0 0%	0 0%
H+!H*	10 67%	0 0%	2 13%	0 0%	3 20%	29 64%	0 0%	7 16%	0 0%	9 20%

The two groups share similar values, yet the transitional probability for one pitch accent pair stands out. L* is followed by H* only 17 times in the data set (in both nonfinal and final intermediate phrases), and in 16 of these, H* is the last pitch accent of the intonational phrase. A low pitch accent is almost never followed by a high pitch accent unless the high tone is a nuclear accent.[1] This could indicate a significant difference between prenuclear and nuclear position and suggest a model in which nuclear accent is structurally separate. With relatively few prenuclear-prenuclear pairs in this data sample, and since the other pairs with sufficient data for reasonable estimates share similar probabilities, further investigation is needed to validate this result.

Nuclear accent is defined here as the last pitch accent of the intonational phrase, but other researchers such as Beckman and Ayers (1994) consider nuclear accent to be the last pitch accent of an intermediate phrase. I examined the pitch accent inventories and tone pair transitional probabilities for prenuclear and nuclear accents using this alternative definition and found results very similar to those in Tables 7 and 8.

While the differences between prenuclear and nuclear accents noted are statistically significant, they are not very large numerically; thus, I do not include them in the model. While Ladd (1996: 211) suggests a model that treats prenuclear and nuclear accents as structurally different, the probabilistic evidence here does not argue strongly for or against such a decision; further consideration of the functional contributions of these accent positions may suggest otherwise.

3.5. Distribution of phrasal and boundary tones

Does the choice of pitch accents have any bearing on the choice of the subsequent phrasal tone or boundary tone? Table 9 presents statistics for the last pitch accent, phrasal tone, and boundary tone in each intonational phrase.

Table 9. Frequency of occurrence of edge tone given stated pitch accent (ratio of observed to expected probability) and {pointwise mutual information} for pitch accent-edge tone pairs

Tone	H-	L-	H%	L%
H*	8% (1.05) {0.051)	92% (12.66) {–0.004}	39% (0.87) {–0.144}	61% (1.11) {0.104}
L*	4% (0.59) {–0.528)	96% (13.13) {0.032}	83% (1.84) {0.608}	17% (0.04) {–1.161}
L+H*	9% (1.19) {0.173}	91% (12.53) {–0.015}	54% (1.19) {0.175}	46% (0.35) {–0.170}
L*+H	17% (2.29) {0.827}	83% (11.43) {–0.107}	17% (0.37) {–0.995}	83% (0.01) {0.417}
H+!H*	1% (0.18) {–1.738}	99% (13.54) {0.063}	26% (0.57) {–0.564}	74% (0.14) {0.303}

Nuclear pitch accent is a significant determinant of boundary tone: high boundary tones are far more likely following certain pitch accents than others. For example, 83% of boundary tones following L* are high, while only 26% of boundary tones following H+!H* and 17% of boundary tones following L*+H are high.[2] In the middle are H* (at 39% high boundary tones) and L+H* (at 54% high boundary tones). The difference among these probabilities is statistically significant at the 1% level.

These results are confirmed by calculating the information shared between pitch accents and boundary tones. Average mutual information is relatively high (0.0377), thus the identity of a nuclear accent significantly determines the identity of a boundary tone and vice versa. Mutual information between nuclear accents and phrasal tones is relatively low (0.0038) however; this indicates that knowing the identity of a given pitch accent says little about the identity of the phrasal tone and vice versa.

My tests of whether phrasal tones predict boundary tones showed that high boundary tones follow low phrasal tones about half the time (46%) but succeed high phrasal tones relatively rarely (16%). These results are summarized in Table 10.

Table 10. Number of occurrences, (frequency of occurrence), {frequency of occurrence of phrasal tone before boundary tone}, and [pointwise mutual information] for phrasal tone-boundary tone pairs

Tone	H%	L%
H-	14 (1%) {16%} [–1.041]	74 (6%) {84%} [0.426]
L-	530 (44%) {46%} [0.050]	589 (49%) {54%} [–0.043]

If the phrasal tone is high, there is some confidence the boundary tone will be low, but if the phrasal tone is low, very little can be said about which boundary tone will occur. Thus, phrasal tone can be an important predictor of boundary tone, but only in the less common case of a high phrasal tone. Consistent with this, the mutual information shared between them is moderate (0.0150). (This low mutual information may simply reflect that the L*H-H% yes-no information question contour and the other contours that include H-H% are more characteristic of dialogue.)

Since nuclear accent and phrasal tone are independently strong predictors of boundary tone, it is no surprise that in combination they provide substantial information about which boundary tone will be chosen. For example, if the first two tones in a sequence are high, the boundary tone is almost certain to be low. Conversely, if the sequence begins L+H*L-, a high tone is not at all unexpected in the final position. Other similar predictions are evident in Table 5.12 of Dainora (2001: 93). Overall, the results reinforce the findings of Tables 9 and 10. The data indicate that some tones are not chosen independently and that the identity of the boundary tone in particular is strongly tied to other parts of the phrase.

4. An empirically based model of intonation

Figures 2 and 3 present the empirically based model of intonation. Figure 2 applies a first order Markov approach, while Figure 3 uses a second order Markov model.

4.1. First order Markov model

I describe this model as "empirically based" for two reasons. First, it includes estimates of transition probabilities between states, enabling specific predictions regarding the likelihood of various tone combinations. Each probability is a model parameter representing the probability of transitioning from one state to another; these numbers are estimated from the corpus data. Second, the data themselves guided a number of choices made in developing this model. For example, test results showed significant differences in the length of pitch accent sequences in nonfinal and final intermediate phrases. To capture this distinction, I have separated these phrases in the model. Note that these are important distinctions from the Pierrehumbert model in Figure 1.

Let us examine the figure to clarify how the model works. This hierarchical model contains two stages with the bottom level comprising individual tones and the top level comprising phrases. As each branch is described, the probability of following that branch is given in parentheses. From the starting point, the intonational phrase begins with an initial %H boundary tone (.003), or the boundary tone is skipped (.997). Moving rightward on the figure, there is the choice node for nonfinal intermediate phrases; the black dot (and the other dot in the figure) do not indicate actual states in the model but are a visual tool to simplify the figure. Each intonational phrase contains a positive number of nonfinal intermediate phrases (.35) or skips directly to the final intermediate phrase (.65). These probabilities, like those that follow, are conditional probabilities; they represent the probability of one or another occurrence conditional on having arrived at that state.

	H*	L*	L+H*	L*+H	H+!H*
H*	.66	.04	.24	.01	.05
L*	.52	.36	.03	.03	.06
L+H*	.50	.10	.32	.00	.07
L*+H	.45	.36	.18	.00	.00
H+!H*	.65	.00	.14	.00	.21

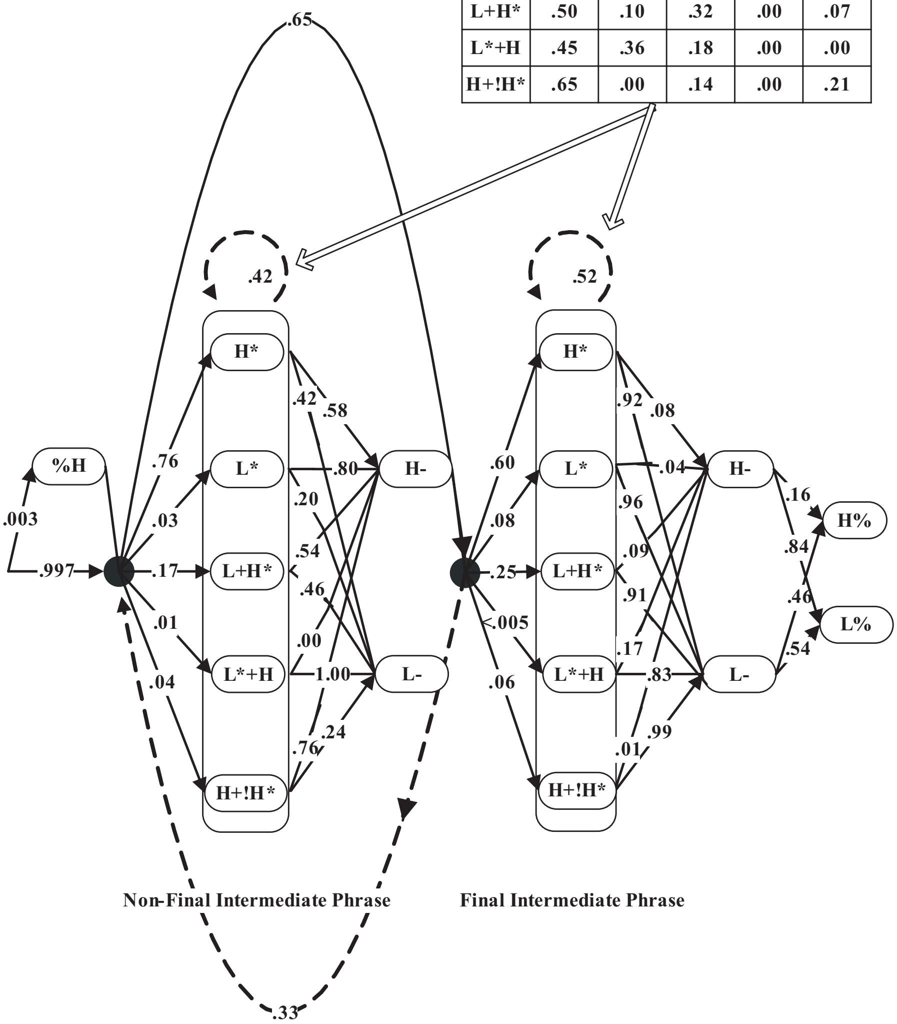


Figure 2. A probabilistic model of intonation in American English

Consider the case where there is at least one nonfinal intermediate phrase. The intermediate phrase begins with one of five pitch accents: H* (.76), L* (.03), L+H* (.17), L*+H (.01), or H+!H* (.04). From the pitch accent position, there is a 42% chance that the model transitions to an additional pitch accent; this relatively low likelihood indicates that long sequences of pitch accents will be rare in the model.

Inset on the top right of the figure is a matrix showing the probability that a given pitch accent moves to another pitch accent. For example, an H* moves to another H* 66% of the time and to L* 4% of the time. I am assuming a stationary model: transitions between pitch accents are the same whether the data are the first pitch accent moving to the second, or the fourth pitch accent moving to the fifth. Because there are relatively few long pitch accent sequences, it is difficult to produce a powerful formal test of the hypothesis that the transition process is stationary. Nonetheless, the data are broadly consistent with a stationary model. The model incorporates the simplifying assumption that conditional and unconditional pitch accent frequency distributions are the same for nonfinal and final intermediate phrases (see Section 3.3). I estimate these values using pitch accents from all intermediate phrases.

When the model reaches the final pitch accent of the first intermediate phrase, it transitions into one of two phrasal tones. Specifically, H* moves to H- (.58) or L- (.42), L* moves to H- (.80) or L- (.20), L+H* moves to H- (.54) or L- (.46), L*+H moves to H- (.00) or L- (1.00), and H+!H* moves to H- (.76) or L- (.24).[3] Note the strong tendency for nonfinal intermediate phrases to contain H-, which supports the claims of Pierrehumbert and Hirschberg (1990: 302) that a high phrasal tone indicates that the phrase is "part of a larger composite interpretive unit with the following phrase." In this data set, the phrasal tone H- is almost never used in final intermediate phrases, but it is the dominant phrasal tone in nonfinal intermediate phrases.

At the end of the first nonfinal intermediate phrase, the model gives two choices. The first is the recursive option to begin a new nonfinal intermediate phrase; the probability associated with this option is .33. The second alternative is to move to the final intermediate phrase. If the recursive option is selected, the option to add another nonfinal intermediate phrase is offered again. This process repeats itself until the second option is chosen. At some point – perhaps after skipping the nonfinal intermediate phrase entirely, perhaps after one or more such phrases – the speaker begins the final intermediate phrase. This phrase consists of a sequence (possibly of length 0) of prenuclear accents, followed by a nuclear accent, a phrasal tone, and a boundary tone. First, one of five pitch accents is chosen: H* (.60), L* (.08), L+H* (.25), L*+H (less than .005), or H+!H* (.06). The chance that the model will move from one pitch accent to another pitch accent is .52. Again, I assume a stationary model of pitch accent behavior. If the model does not move to another pitch accent, the current pitch accent is the final pitch accent of the intonational phrase, and there is a transition to the final phrasal tone: H* moves to H- (.08) or L- (.92), L* moves to H- (.04) or L- (.96), L+H* moves to H- (.09) or L- (.91), L*+H moves to H- (.17) or L- (.83), and H+!H* moves to H- (.01) or L- (.99). Last, there is a transition from the final phrasal tone to the boundary tone, marking the end of the intonational phrase: H- moves to H% (.16) or L% (.84), and L- moves to H% (.46) or L% (.54).

4.2. Second order Markov model

Figure 3 presents a second-order Markov model of the nuclear tune.

The complete intonational model consists of Figure 2 with the rightmost section replaced by Figure 3. Each of the twenty branches in Figure 3 represents a three-tone nuclear tune. To obtain the overall probability of a tune, one multiplies the transition probability between each tone in the branch. The probabilities here are estimated using the last three tones of every intonational phrase in the data, including phrases that do and do not contain prenuclear accents. Because these probabilities are estimated from the data, the overall probability of each tune given by this calculation will exactly match the frequency of the tune in the actual data (Table 1). To look at one example, each nuclear tune is 8% likely to begin with L*. Conditional on the L* pitch accent, the next tone is 96% likely to be L-. Conditional on the nuclear tune beginning L*L-, the tune is 84% likely to end with H%. Multiplying .08 × .96 × .84 = .0645. Thus, the nuclear tune L*L-H% is approximately 6% likely to occur.

Consider the probability model estimate based on the first-order Markov model in Figure 2. Here the three probabilities to be multiplied are .08 × .96 × .46 = .0354, or approximately 4%. The first-order Markov model predicts this tune will occur only about half as often as it actually occurs in the data. While the advantage of lower order models is that they have fewer parameters to estimate, higher order models have the ability to capture additional features of the data. Since by construction the second-order model always captures the nuclear tune sequence data perfectly, it has a clear advantage in terms of accuracy.

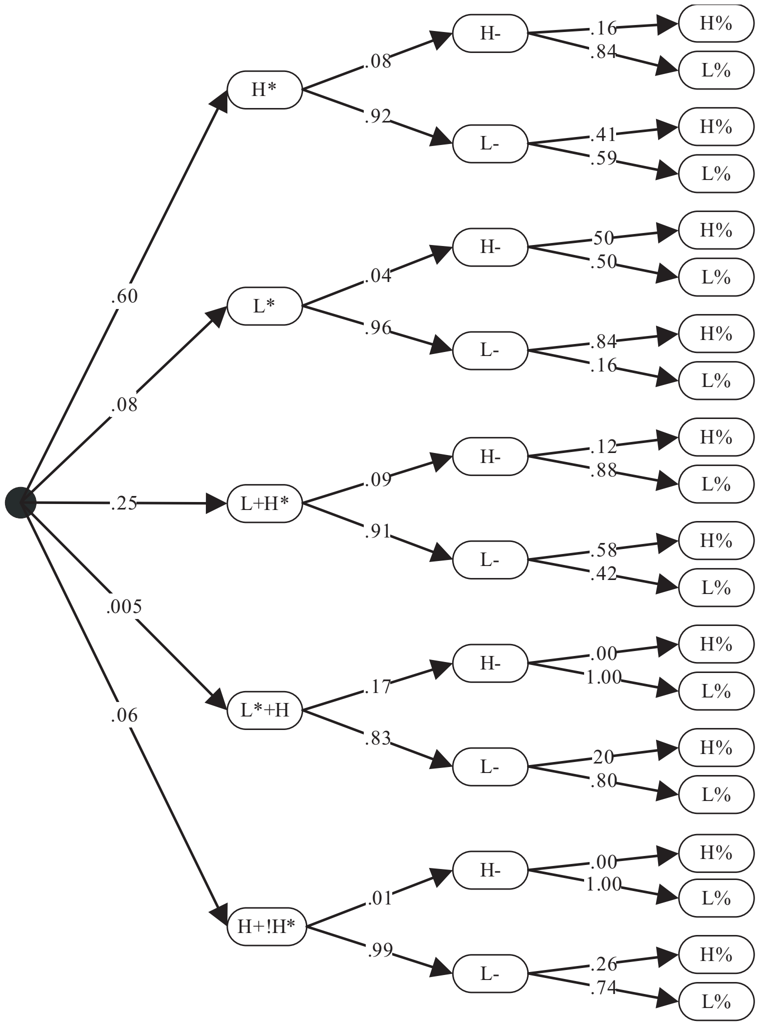


Figure 3. Second-order Markov model showing interactions between pitch accents and boundary tones

5. Implications

Of the many parameters estimated in Section 3, a few stand out for their implications for intonational theory. Detailed here are two particular examples:

the identity of strings of pitch accents and the interdependence of pitch accents and boundary tones.

5.1. Strings of pitch accents

One issue raised in the intonational literature concerns the status of strings of pitch accents. Ladd (1996: 211) discusses a model in which prenuclear accents tend to be identical, with the selection of prenuclear pitch accents being a single linguistic choice. He notes that "In a limited number of cases, there may be prenuclear accents of a different type" (1996: 296). Similarly, Cruttenden (1997: 65) observes that the majority of prenuclear tonal sequences contain identical pitch accents, but that a minority contain non-identical accents.

I test these claims by examining sequences of prenuclear accents occurring immediately before the final accent, phrasal tone, and boundary tone of intonational phrases. I find no tendency for prenuclear pitch accents to be identical. Identical sequences are common only because the dominance of the H* tone (which constitutes about 70% of all prenuclear pitch accents in this sample) causes identical sequences of H* to occur frequently. Identical sequences of other pitch accents are exceedingly rare. Of the 287 final intermediate phrases containing two or more prenuclear accents; 158 of them contain a sequence of identical prenuclear accents. Of these, only 7 contain a pitch accent other than H*, and all 7 of these sequences are of length 2.

To test whether prenuclear accents might tend to be similar, even if not every tone in the sequence is identical, I examine pairs of adjacent pitch accents to see whether identical pairs are more common than would be predicted by chance. The 287 final intermediate phrases contain 402 pairs of prenuclear pitch accents. Consider the most common pitch accent H*. When H* is followed by another pitch accent, the second pitch accent is 70% likely to be H*. When L+H* is followed by another pitch accent, the following tone is H* in 64% of the cases. H* occurs about 71% of the time following the other three pitch accents. The second pitch accent of a pair is about 70% likely to be H* regardless of whether the first accent is H* or not. There is no particular tendency to observe pairs of identical pitch accents, only a tendency to see a great many H* tones. Similar results are obtained from other pitch accents and for longer sequences. I conclude that a model in which strings of identical prenuclear pitch accents are the norm and nonidentical strings a rarity clearly can be rejected.

5.2. Differentiating H* and L+H*

One controversial area in the phonological analysis of tones is whether H* and L+H* are contrastive. Some researchers argue that H* is distinct from L+H* because of the different meanings they connote. Pierrehumbert and Hirschberg (1990: 289, 296) propose that the H* accent is used by the speaker to indicate that the accented item is new information while the L+H* accent is most commonly used for contrast to indicate that "the accented item – and not some alternative related item – should be mutually believed." Other researchers are not convinced that H* and L+H* are contrastive, as they can be difficult to distinguish, especially in utterance initial position. The majority of labeling conflicts in the Boston University Radio Speech Corpus were between H* and L+H* (Ostendorf, Price, and Shattuck-Hufnagel 1995: 13). Cruttenden (1997: 65) mentions that the forms and the meanings of the two accents are often difficult to distinguish: "At best, L+H* is said to be more contrastive." Bartels and Kingston (1995) find that subjects tend to use accent peak height to distinguish contrastive interpretations from nonconstrastive interpretations rather than the depth of dip of the contour or a combination of peak timing and rise onset timing.

Whether H* and L+H* are distinctive is addressed here by looking at how they combine with other tones. Consider phrases that begin with H* or L+H* and end in L-L%. Of these, 23% begin with L+H*. This is consistent with the idea that H* is the default and L+H* is a less common noncontrastive variant. Now consider the set of phrases that begin with H* or L+H* but end in L-H%. Of these, 36% begin with L+H* – more than 1.5 times as many as when the phrase ends in L-L%. A difference of means test reveals a *p*-value on this difference of .000003, indicating it is highly unlikely that such a result would be obtained by chance. It is difficult to see why if H* and L+H* are noncontrastive, L+H* would be so much more commonly used in phrases that end in L%.

Could these findings result from labeler error or from arbitrary labeling? The overwhelming evidence of the statistical test makes it clear the labeling is not random. If it were, phrases beginning with L+H* would be just as likely to end in L% as phrases beginning with H*. Moreover, labeler error is an unlikely culprit. What basis could labelers have to show a consistent but not universal tendency to label ambiguous tones H* when the phrase ends in L%, but L+H* when the phrase ends in H%, especially since in the cases under consideration the middle tone is always L-. To exhibit such bias, labelers would have to look two tones ahead and then sometimes, but not always, adjust their labeling of the first tone of the nuclear sequence based on the third tone. The evidence of the data is quite clear. H* is used differently in nuclear sequences from L+H*.

6. Conclusions

I build on the finite-state grammar of Pierrehumbert (1980) and others to produce a model of intonation in American English. This paper provides quantitative evidence on how pitch accents, phrasal tones, boundary tones, and prosodic phrases are used in one particular corpus, thus paving the way for future research on statistical approaches to phonological competence. Furthermore, estimations of the probabilities of tones produced a more complete phonological model that incorporates those theoretical claims that match the data best. The approach developed here may be applied to a wide variety of linguistic problems. An important next step is to model other speaking styles, dialects, and languages, in addition to American radio announcer speech. Such analyses will be revealing in themselves. More significantly, this method makes comparative analyses possible, such that differences among speech corpora may be quantified. The probabilistic modeling technique may also be expanded along other dimensions. While this analysis is limited to intonation, adjustments to allow for analyses of other phonetic or phonological data, or pragmatic and semantic information, are easy to imagine. This may allow for a higher degree of precision to areas of inquiry traditionally characterized by scrutiny of small data samples from limited numbers of speakers.

Notes

* This work was supported in part by the University of Chicago Language Laboratories and Archives. I thank Amy Dahlstrom, Mirjam Ernestus, Martine Grice, the late Karen Landahl, Stefanie Shattuck-Hufnagel, Donca Steriade (the discussant), and especially John Goldsmith for valuable insights and discussion. The final version benefited greatly from excellent suggestions made by Mary E. Beckman, an anonymous referee, and editor Douglas H. Whalen.

1. Three-quarters of these L*H* pairs occur before L-L%, thus creating the surprise-redundancy contour.
2. The low frequency in the data set of the fall-rise contour (17% of nuclear tunes that begin with L*+H) could be interpreted as evidence against the approach

taken by Ward and Hirschberg (1985), among others, of treating L*+HL-H% as a nondecomposable tune, as discussed in Section 3.1.

3. The value for the L+H*H- transition is given as .00, which accurately represents the data in the corpus in which that transition occurs 0 times. If the model were being implemented, one would incorporate smoothing in order to assign a positive value to all transitions.

References

Bartels, Christine and John Kingston

1995 Salient pitch cues in the perception of contrastive focus. In: Rob A. van der Sandt (ed.), *Focus and Natural Language Processing,* Vol. 1: *Intonation and Syntax,* 1–10. Heidelberg: IBM Working Papers.

Beckman, Mary and Gayle Ayers

1994 Guidelines for ToBI labeling, version 2.0. Columbus: Ohio State University, ms.

Beckman, Mary and Janet Pierrehumbert

1986 Intonational structure in Japanese and English. *Phonology Yearbook* 3: 255–310.

Charniak, Eugene

1993 *Statistical Language Learning.* Cambridge, MA: The MIT Press.

Coleman, John and Janet Pierrehumbert

1997 Stochastic phonological grammars and acceptability. *Proceedings of the 3rd Meeting of the ACL Special Interest Group in Computational Phonology (12 July 1997)*, 49–56. Somerset, NJ: Association for Computational Linguistics.

Cruttenden, Alan

1997 *Intonation.* Cambridge: Cambridge University Press.

Dainora, Audra

2001 An empirically based probabilistic model of intonation in American English. Ph.D. dissertation, Department of Linguistics, The University of Chicago.

2002 Does intonational meaning come from tones or tunes? Evidence against a compositional approach. In: Bernard Bel and Isabelle Marlien (eds.), *Proceedings of the Speech Prosody 2002 conference (11–13 April 2002)*, 235–238. Aix-en-Provence: Laboratoire Parole et Langage.

Frisch, Stefan A., Nathan R. Large and David B. Pisoni

2000 Perception of wordlikeness: Effects of segment probability and length on the processing of nonwords. *Journal of Memory and Language* 42: 481–496.

Goldsmith, John
2002 Probabilistic models of grammar: Phonology as information minimization. *Phonological Studies* 5: 21–46.

Hay, Jennifer, Janet Pierrehumbert and Mary E. Beckman
2004 Speech perception, well-formedness and the statistics of the lexicon. In: John Local, Richard Ogden, Rosalind Temple (eds.), *Papers in Laboratory Phonology VI*. Cambridge: Cambridge University Press.

Ladd, D. Robert
1983 Phonological features of intonational peaks. *Language* 59: 721–759.
1996 *Intonational Phonology*. Cambridge: Cambridge University Press.

Manning, Christopher D. and Hinrich Schütze
2000 *Foundations of Statistical Natural Language Processing.* Cambridge: The MIT Press.

Ostendorf, Mari, Patti J. Price and Stefanie Shattuck-Hufnagel
1995 The Boston University Radio News Corpus. Philadelphia: Linguistic Data Consortium, ms.
1996 Boston University Radio Speech Corpus LDC96S36. CD-ROM. Philadelphia: Linguistic Data Consortium.

Pierrehumbert, Janet B.
1980 The phonology and phonetics of English intonation. Ph.D. dissertation, Massachusetts Institute of Technology.

Pierrehumbert, Janet and Julia Hirschberg
1990 The meaning of intonational contours in the interpretation of discourse. In: Philip R. Cohen, Jerry Morgan, and Martha E. Pollack (eds.), *Intentions in Communication,* 271–311. Cambridge, MA: The MIT Press.

Treiman, Rebecca, Brett Kessler, Stephanie Kneewasser, Ruth Tincoff and Margo Bowman
2000 English speakers' sensitivity to phonotactic patterns. In: Michael Broe and Janet Pierrehumbert (eds.), *Papers in Laboratory Phonology V: Acquisition and the Lexicon,* 269–283. Cambridge: Cambridge University Press.

Ward, Gregory and Julia Hirschberg
1985 Implicating uncertainty: the pragmatics of fall-rise intonation. *Language* 61: 747–776.

The diachrony of labiality in Trique, and the functional relevance of gradience and variation

Daniel Silverman

At some point in the history of Trique, round vowels began to spread rightward across velars, eventually turning these velars into labio-velars. This spreading did not occur when the consonant was alveolar. In this study I consider the diachronic interaction of certain phonetic, cognitive, and functional pressures on the Trique system which may be responsible for the asymmetrical development of this sound change, and I provide psychoacoustic experimental results which support my approach. I further propose that sound changes of the Trique sort can only be compellingly accounted for within a theory of enriched representations that incorporates the probabilistic components of language use, including (though not limited to) the distribution of variants in the acoustic/articulatory space.

> One wonders whether the habit of constantly operating with graphic notations does not make some linguist(s) deaf to the gradual shifts which any painstaking observation can reveal. If one has been taught, not only that phonological systems are made up of discrete units, but also that these units are basically the same in all languages, and that even if a discrete unit may well appear under the form of different allophones, these allophones can be listed and identified, so that they, in a sense, partake in the discreteness of the phonemes, one can hardly avoid concluding that no change can take place except by means of jumps from one unit or allophone to another. Only those who know that linguistic identity does not imply physical sameness, can accept the notion that discreteness does not rule out infinite variety and be thus prepared to perceive the gradualness of phonological shifts.
>
> André Martinet, 1975: 25

1. Introduction

Trique is a Mixtecan language of the Otomanguean group, spoken by about 23,000 people in the states of Oaxaca, Guerrero, and Puebla, Mexico (Grimes

2003). There is an interesting distributional asymmetry in Trique which is the focus of this study. Whenever the high round vowel [u] precedes a velar consonant, a labial glide immediately follows. However, the glide is not present when the consonant that immediately follows the round vowel is alveolar. So, we find sequences like [ukʷɑ] and [utɑ], but never [ukɑ] or [utʷɑ]. Longacre (1957) attributes this present-day asymmetry to a sound change. Historic *ukɑ has become present-day [ukʷɑ], for example, [ʒukʷɑhɑ] 'fish', [rugʷi] 'peach'. However, *utɑ has remained [utɑ], for example, [utɑh] 'annoint', not [utʷɑh]. In this paper I suggest that Trique trans-velar labial spreading may be historically rooted in the greater likelihood of labial coarticulation in the velar context, as opposed to the alveolar context, since such coarticulation enhances the acoustic distinction between the two contrastive values. Trans-alveolar labial spreading cannot be similarly motivated, since labial spreading here would partially undo the increased acoustic distinction which trans-velar labial spreading created. I further report the results of a psychoacoustic study which supports this phonetic account of the sound change. When subjects were asked to identify the sound sequences [udɑ], [udʷɑ], [ugɑ], and [ugʷɑ] in various degrees of white noise, they least often confused [udɑ] and [ugʷɑ] with each other. I then consider some theoretical implications of these results. First, I consider probability matching. Language users seem to learn variable linguistic patterns by calculating the perceived probability of occurrence of the variants, and largely matching this variation in their own productions (Labov 1994 *pace* Gallistel 1990, Liberman, 2002). I then consider the interaction of these cognitive factors with phonetic and functional influences on the sound change. Since certain variants are more likely to be perceived unambiguously – that is, since certain phonetic variants (over others) of a given word are more acoustically distinct from *other* words – then, as a consequence of probability matching, it is *these* variants that are more likely to be produced as listeners become speakers, and so a sound change may be set in motion. Based on the experimental results, I propose a diachronic scenario for the Trique sound change. The stability of word meaning was the decisive factor in determining the relevant phonological categories. So, for example, during the early stages of the sound change, [dugɑh] may have varied with [dugɑh], and eventually became [dugʷɑh] 'to twist'. Since the overall meaning of the word did not change, all variants along the [g]-[gʷ] continuum may have been regarded as categorially non-distinct by learners. In this sense, phonetic realizations may be categorized together as long as meaning remains stable, regardless of phonetic gradience or token-to-token phonetic variation. In other words, any emergence of phonological categories may be at least partly parasitic on perceived lexical semantic identity, rather than on the specific physical similarities or differences among variants or alternants. Finally, I consider a multiple trace or exemplar model approach to gradience and variation. I conclude that it is probably only in a theory such as multiple trace, which proposes quantitatively enriched representations, that sound changes of the Trique sort might be accounted for.

2. The diachrony – and limits – of Trique labial spreading

As mentioned, whenever the high round vowel [u] precedes a velar consonant, a labial glide immediately follows. Some examples are provided in (1). Data are from Hollenbach 1977, including forms from both the San Juan Copala dialect and closely related San Andrés Chiquihuaxtla. Words are usually disyllabic; syllables are typically CV; the relevant sequences are underlined; tones are not indicated.

(1) Trique trans-velar spreading:

nukʷɑh	strong	dukʷɑ	possessed house
dugʷɑh	to twist	zugʷi	(name)
ʒugʷɑ	to be twisted	dugʷe	to weep
dugʷɑne	to bathe (someone)	rugʷi	peach

The glide is not present when the consonant that immediately follows the round vowel is alveolar. In (2) are some examples of this pattern. (Throughout Mixtecan, labial consonants are quite rare, and in Trique this seems to be especially true (Silverman 1993).

(2) Trique round vowel - alveolar sequences:

rune	large black beans	utɑh	to anoint
utʃe	to get wet	utʃi	to nurse
utɑ	to gather	dunɑ	to leave something
rudɑʔɑ	stone rolling pin	ʒutʃe	hens, domestic fowl

This asymmetry seems to be due to a sound change (Longacre 1957, 1962, Gudschinsky 1959, Longacre and Millon 1961, and Rensch 1976). Among the root-final syllables that Longacre (1957) reconstructs for Proto-Mixtecan is *kɑ. This sequence largely survives in the three main branches of Mixtecan, which include Mixtec, Cuicatec, and Trique. In Mixtec, the pattern basically survives in full. In Cuicatec, both [kɑ] and [kuː] are found. In Trique, kɑ survives except when the penult possesses [u]. In this context, we now know, Trique possesses a labial glide at stop offset. Longacre states that "(The) T(rique) gʷ cluster is largely a development of g in the situation u…a" (p.17). He further lists Proto-Mixtecan *kɑ as becoming Proto-Trique *kʷɑ in the context of a preceding [u]: "*kɑ > […] (u)kʷɑ" (p.33).

The prevalence of [ukʷɑ] forms over [ukʷ] forms with other final vowels seems due to a number of factors. *kɔ merged with *kɑ, and subsequent lexicalized compounding innovations have lead to labial spreading in the relevant contexts: *kɔ 'snake, lizard' became intermediate [kɑ] and, combining with [ʒu] 'animal', becomes [ʒukʷɑ] 'snake'. This means that both *ukɑ and any *u+kɔ forms became Trique [ukʷɑ]. Among the other non-round final vowels which, in theory, could have followed *uk were *i, *ɨ, and *e. But *ki and *kɨ were rare in Proto-Mixtecan, and barely survive into Trique, and *ke is not even reconstructed for Proto-Mixtecan. Regarding *ɑ, in a few cases, including some words with [uʔw__ʔ/h], and [ugʷ__h], *ɑ has raised to [e]. It has sometimes raised to [i] in the situations [ɑw__], [ɑʔw__ʔ/h], and [ugʷ__(ʔ)] (Longacre 1957: 44). These idiosyncratic innovations may have lead to the broader contexts of labial spreading found today, such that the u-velar-w sequence may now very sporadically precede [i] or [e] as well. Labialization is never found before [o], because CuCo forms are largely absent from the language. Hollenbach provides only two such forms, and neither possesses a velar as the second consonant ([guno] 'to hear', [uno] 'to sow'). CuCʷu is absent in Trique, perhaps because it does not make for a robust acoustic contrast with CuCu. Indeed, it is rare that a language has lexical contrasts involving Cʷu and Cu (though they may be present at the post-lexical level, e.g. English "hoodoo" versus "who'd woo").

The mid round vowel [o] is largely absent in penults. Hollenbach writes that "Although all five long vowels occur in nonultimas [penults –D.S.], it is almost possible to reduce the number of contrasts in this position to three […] /e o/ are uncommon in non-ultimas. They occur mainly when the ultima vowel is itself mid" (p.42f). Therefore, the absence of labio-velars following [o] may be a consequence of the near absence of [o] in this context, and is not due to an asymmetrical application of the spreading process itself.

Significantly, in Proto-Mixtecan, *k and *k^w were not contrastive in the context of a preceding round vowel (Longacre 1957). Consequently, the trans-velar labial spreading innovation did not induce homophony; the change was purely contrast-maintaining, indeed, contrast-enhancing.

The table in (3) breaks down CVCV sequences into sixteen logical labial classes, only seven of which are actually documented in Longacre's Trique word list (consisting of about 500 items). "C^w" represents any labial(/ized) consonant, while "V^w" represents any round vowel.

(3) Trique disyllabic root classes with respect to the distribution of labiality

$C_1 V_1 C_2 V_2$ Classes:	# of subclasses:	
C V C V	72	
C^w V C V	6	
C V^w C V	11	(C_2 is never velar)
C V C^w V	17	(C_2 is a plain labial in 10 subclasses, a labialized velar in 7)
C V C V^w	20	
C^w V^w C V	0	
C^w V C^w V	0	
C^w V C V^w	0	
C V^w C^w V	31	(V_1 is always [u]; C_2 is virtually always a labialized velar or [w]; very rarely [m])
C V^w C V^w	15	(V_1 and V_2 are identical in all but one entry)
C V C^w V^w	0	
C^w V^w C^w V	0	
C^w V^w C V^w	0	
C^w V C^w V^w	0	
C V^w C^w V^w	0	
C^w V^w C^w V^w	0	

Although the prevalence of words within each class is not indicated, the "totals" column lists the number of subclasses within each root class. Each subclass is different from all others in terms of at least one consonant or vowel. Vocalic length, phonation, nasality, and tone distinctions are pooled, however. The totals, along with the parenthesized commentary, convey the strict limitations on the distribution of labiality in Trique. In particular, labial spread is only present when V_1 is [u] and C_2 is velar.

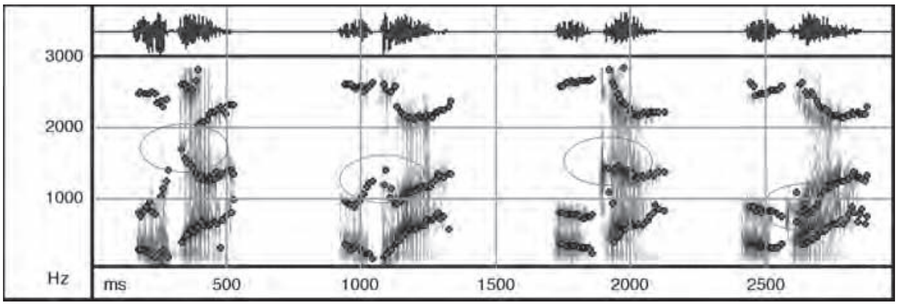
3. Phonetic underpinnings

Why should a labial glide have evolved in the *uk context? The answer I wish to pursue includes phonetic, functional, and cognitive components, which diachronically interact. Consider some phonetic facts first. The tongue and lips are independent articulators; both may be active simultaneously. In the historical context under investigation, it is reasonable to assume that the lip-rounding gesture characteristic of [u] may have variably perseverated into the dorsal closure characteristic of [k]: [k̹]. General perserveration of rounding has been documented in New York English, for example (Bell-Berti and Harris 1982). Other vocalic gestures in addition to labiality have been shown to both perseverate and anticipate in this fashion (Öhman 1966, Bell-Berti and Harris 1979, for example). Moreover, given the tongue-backing gesture required of the preceding vocoid, the distance traversed by the dorsum to implement the stop closure is comparatively short, thus increasing the likelihood of some variable rounding "spill-over". That is, given the short time frame of dorsal closure implementation, minor timing variations may lead to rather pronounced acoustic distinctions among variants. Specifically, persistence of lip-rounding through the dorsal closure may lead to the perception of a labialized velar.

This might seem like a good point of departure for the evolution of [ukʷɑ] in Trique. Unfortunately, the proposal suffers from a fatal flaw. The problem is that certain other consonants may just as readily be produced with perseverative labiality as may [k]. Thus, for example, an alveolar stop involves a (typically rapid) tongue-tip raising gesture, which can be achieved largely independently of the dorsal backing gesture characteristic of [u]. We might thus predict little-to-no asymmetry in the diachronic comportment of *ukɑ and, say, *utɑ. Yet Trique clearly has not treated these two patterns in a parallel fashion: *utɑ ↛ [utʷɑ].

Instead, the spreading asymmetry may serve to enhance the acoustic distinction between the two contrastive values. Since, according to Longacre's reconstructions, *ukʷ sequences were absent in the proto-language, spreading labiality across the velar *increased* the acoustic distinction between the velars and the alveolars, without inducing homophony. Accompanying trans-alveolar spreading, by contrast, would serve to *diminish* the acoustic distinction. This is suggested spectrographically in (4), where waveforms, wideband spectrograms, and formant tracks are provided for the sequences [udɑ], [udʷɑ], [ugɑ], and [ugʷɑ], spoken by a native speaker of New York English (the author).

(4) Waveforms, spectrograms, and formant tracks for the four sequences spoken in New York English.



[udɑ] [udʷɑ] [ugɑ] [ugʷɑ]

Observe the F2 onset values after stop release in the circled regions. F2 onset values are: [udɑ] 1700 Hz, [udʷɑ] 1200 Hz, [ugɑ] 1500 Hz, [ugʷɑ] 900 Hz. Thus, [ugʷɑ] and [udɑ] are maximally distinct.

To summarize the proposal, by considering the acoustic and consequent functional benefit of spreading labiality across velars – a pattern which might be present due to the variation inherent in speech production – and the counter-functionality of spreading labiality across alveolars, we might motivate the Trique sound change. This proposed mechanism of sound change is not speaker-induced (through an effort to enhance the distinctness among contrastive elements) as has been suggested by some researchers. For example, Kingston (2002) concludes that "[S]peakers *exert themselves* to convey contrasts in ways that are entirely unexpected if they couldn't optimize their pronunciations to ensure that contrasts are conveyed," and that "*Speakers must be altruists*" (emphasis mine). Instead, I propose a listener-based account by which contrasts might be enhanced passively, evolving over generations of speakers, due to the communicative success of some tokens, and the communicative failure of others. Note especially that the labial glide has not suddenly popped out of the ether in an effort on the part of the speaker to enhance the distinction among the relevant contrastive configurations. Instead, labiality was already loitering in the neighborhood, so to speak, and was passively harnessed to play a new, functionally beneficial role.

4. Experiment

A laboratory condition may serve to recapitulate elements of the hypothesized historical scenario in "speeded-up" form by introducing white noise into the speech signal, and having listeners report on their perception. Although only the author's speech was employed, subsequent investigation of three other native speakers of English revealed largely comparable F2 onset values.

4.1. Subjects and methods

The subjects for this double-blind study were 10 University of Illinois students in linguistics, all native English speakers. Sound files consisting of the four relevant phonetic sequences were digitally recorded in the Department of Linguistics' phonetics lab at a sampling rate of 22,050 Hz: [udɑ], [ud^{w}ɑ], [ugɑ], [ug^{w}ɑ].

Each file was overlaid with four levels of white noise, with each noise level increased in amplitude from the previous level. Including a no-noise level, this resulted in a total of four continua with five noise levels each, for a total of twenty sound files. Stop closure durations for the four forms were: [udɑ] - 50 msec, [ud^{w}ɑ] - 51 msec, [ugɑ] - 40 msec, [ug^{w}ɑ] - 54 msec. Second vowel durations from stop release (thus including the glides) were: [udɑ] - 210 msec, [ud^{w}ɑ] - 213 msec [ugɑ] - 213 msec, [ug^{w}ɑ] - 202 msec. Pitch tracks and intensity contours were comparable across stimuli.

Using PsyScope, subjects listened with headphones in a quiet room to 1000 trials – 50 of each of the 20 sound files – in randomly generated blocks of 100, with a 2 second inter-trial interval, and untimed rests between blocks. Using the keyboard, subjects reported which sound sequence they heard ([udɑ], [ud^{w}ɑ], [ugɑ], or [ug^{w}ɑ]). Subjects were encouraged to guess if they were undecided.

4.2. Results

In order to eliminate any floor or ceiling effects, the no-noise condition (5% incorrect responses) and maximum-noise condition (60% incorrect responses) were not pooled. The highest total of pooled errors is 792 (26%

incorrect responses), for the [udʷɑ]-[ugʷɑ] distinction (F2 onset difference minimal at 200 Hz), and the lowest total of pooled errors is 32 (1% incorrect responses), for the [udɑ]-[ugʷɑ] distinction (F2 onset difference maximal at 700 Hz). Overall, these results suggest that the F2 distinctions among the four sound sequences are good predictors of confusability: by and large, the more similar the F2 onset values, the more confusable; the less similar the F2 onset values, the less confusable. It further suggests that stop closure duration and second vowel duration had little effect on listeners' classifications.

A confusion matrix, also excluding the no-noise and maximum-noise conditions, is provided in (5). Correct responses are bold-boxed. Percentages reflect the number of responses out of 1500 (500 each at the middle three noise levels; not all stimuli were responded to).

(5) Confusion matrix

perceived → presented ↓	udɑ	udʷɑ	ugɑ	ugʷɑ
udɑ	1208 81%	40 3%	145 10%	17 1%
udʷɑ	223 15%	812 54%	71 5%	291 19%
ugɑ	355 24%	47 3%	964 64%	43 3%
ugʷɑ	15 1%	501 33%	14 1%	879 59%

There are two notable trends in the pattern of directional errors. First, labialized stops are more often misperceived as non-labialized (323 errors), rather than vice versa (147 errors). Second, velars are more often misperceived as alveolars (918 errors), rather than vice versa (524 errors). Both of these asymmetries might be due to a response bias induced by phoneme frequency factors. For example, according to Fry (1947), [t] occurs with more than twice the frequency of [k] in the English spoken in Southern Britain. A similar account might be offered for the pattern of correct responses. However, two subjects, to the exclusion of the other eight, quite regularly reported hearing [udɑ] when presented with [udʷɑ]. It is not clear why these two subjects – and *only* these two subjects – responded in this fashion, but their idiosyncratic performance might be an alternative reason for aspects of the observed pattern of errors.

In (6), the confusion matrix is re-arranged according to increasingly distinct F2 oppositions (Levels 2 through 4). Estimated absolute F2 onset differences are parenthetically noted. As presented in (6), it becomes clear that the presence versus absence of the glide is readily perceived, but confusion increases between forms which differ solely in terms of the stop's place of articulation, especially when labiality is present. The degree of confusion thus does indeed correlate well with the degree of F2 similarity.

(6) F2-based confusion matrix

perceived ↘ presented ↓	Level 1 Correctly answered	Level 2 Nearest F2	Level 3 Mid F2	Level 4 Furthest F2
udɑ	udɑ 81%	ugɑ (200 Hz) 10%	udʷɑ (500 Hz) 3%	ugʷɑ (700 Hz) 1%
udʷɑ	udʷɑ 54%	ugʷɑ (200 Hz) 19%	ugɑ (300 Hz) 5%	udɑ (500 Hz) 15%
ugɑ	ugɑ 64%	udɑ (200 Hz) 24%	udʷɑ (300 Hz) 3%	ugʷɑ (500 Hz) 3%
ugʷɑ	ugʷɑ 59%	udʷɑ (200 Hz) 33%	ugɑ (500 Hz) 1%	udɑ (700 Hz) 1%

A repeated measures ANOVA confirmed a main effect for F2 similarity, $F(3, 27)=158.6$, $p<.001$. Pairwise comparisons with Bonferroni adjustment revealed a significant difference between Levels 1 and 2, and between Levels 2 and 3 ($p<.001$). The difference between Levels 3 and 4 was not significant ($p>.05$), even when including the idiosyncratic responses of the two aforementioned subjects, suggesting that when F2 differences surpassed a certain value, the rate of misperception leveled off.

5. Discussion

Before embarking on a discussion of certain theoretic implications of Trique-type sound changes, I want to briefly consider a mechanism by which one sound might gradually change into another. Martinet's take on this matter, encapsulated in the quote which opens this paper, implicates the importance of the gradience and variation which are inherent in speech production: the gradual nature of some sound changes can be explained by considering the gradient and variable nature of speech production itself. Similar proposals have made, for example, by Paul (1886: 43), Hockett (1968: 83) Anttila (1972: 53) Ohala (1989), and Janda and Joseph (2001). But given this inherent gradience and variation, again, what is the actual mechanism by which they might induce sounds to change? To answer this question, I briefly turn to the foraging behavior of rats and ducks.

5.1. Probability matching

Gallistel (1990: 352) reports on a study in which rats in a T Maze were rewarded with food 75% of the time at one end, 25% of the time at the other. When provided with feedback, these rats matched the probability of reward – running to the one end 75% of the time, the other end 25% of the time – despite the fact that they would receive a higher rate of reward if they ran to the one end 100% of the time (61.5% versus 75%). Experimental variations on the rat-in-a-T-maze theme have been performed, yielding similar results. For example, in a somewhat less controlled experimental setting, Harper (1982) reports that two experimenters standing by a pond, set apart from each other some distance, threw food to ducks at two different rates. Very quickly, the ducks were able to calculate the distinct rates of feeding, and match their foraging time near each experimenter accordingly, spending more time at the location of greater payoff, and switching to the location of lesser payoff for a percentage of time that matched the lower yield. These ducks did not necessarily receive any food before matching their behavior to the probability of payoff. Rather, some were able to predict the payoff before any reward was received.

Comparable statistical calculations seem to underlie certain aspects of human linguistic behavior: even though certain variants are better than others at communicating speakers' lexical semantic content to listeners, when listeners become speakers they largely match their own variation of production to that which they perceive, including both "better" variants (more distinct from other words) and "worse" variants (less distinct from other words). (See, for example, Preston and Yeni-Komshian 1967, Poplack 1980a,b, Hudson and Newport 1999). But if listeners were able to perfectly match the probabilities present in the speech that is produced around them, then, *ceteris paribus*, sounds would never have the opportunity to change in the proposed fashion. Rather, perfect reproduction would yield perfect diachronic stability. Since sounds do change over time, we may assume instead that listeners match their speech to their own perceptions of these ambient productions. When two different words are acoustically similar, some tokens of the one word may be misperceived as the other word, or may simply remain unanalyzed (that is, be thrown out, ignored). Since perception is demonstrably imperfect, then reproduction is imperfect as well, and so a sound change may gradually ensue. Since probability matching is based on listeners' perceptions, sound patterns may slowly take on new characteristics in the direction of the "better" – less ambiguous – tokens. (The misunderstanding of ambiguous tokens would be extremely difficult to document in a systematic way, but Labov [1994] discusses many anecdotal cases, and suggests that the phenomenon is far more prevalent than innocent language users would like to think.) Most systems of reproduction are imperfect, and in speech (re-)production, the facts of probability matching suggest that one locus of this imperfection may lie in the realm of perception. However, the mimetic abilities of speakers are imperfect as well. Due to the inherent imperfection of speech (re-)production, there will inevitably be *some* stray tokens which end up confusable with other words. Strays which are too similar to another word are more likely to be passively factored out (thrown out, ignored) because they remain uninterpreted by the listener, and thus serve to induce and maintain an acoustic buffer between one word and others. Other strays, however, might be slightly *more* distinct from other words than most tokens are. Such strays may become more prevalent in the system, due to their perceptual and consequent functional advantages: one generation's strays may evolve into a later generation's norm.

Consider how probability matching may play a role in sound changes of the Trique sort. There is inherent gradience and variation in speech production, thus [ukɑ...uk̹ɑ...ukʷɑ], and [utɑ...ut̹ɑ...utʷɑ] are among the possible variants. If Longacre's reconstructions are accurate, then in the proto-stages, productions leaned heavily toward [ukɑ] and [utɑ]. However, stray [uk̹ɑ...ukʷɑ]-like variants rendered the [u]-velar-V sequences more distinct from their [u]-alveolar-V counterparts. Therefore, these variants were more likely communicated unambiguously to listeners. Ambiguous tokens – specifically, [utʷɑ]-like variants – were sometimes miscategorized (misinterpreted as an unintended word), but also were sometimes uncategorized (simply ignored), and hence were not added to the pool of tokens over which probabilities were calculated, and so [utɑ] survived largely intact. That is, the variation engaged in by elders was largely matched by learners, but nonetheless, due to the greater likelihood of unambiguous perception of certain variants over others – [ukʷɑ] over [ukɑ]; [utɑ] over [utʷɑ] – learners' calculated probabilities may have differed slightly from their elders', in that the variants which contrast more sharply with other words were more often perceived correctly, hence, in turn, more likely produced. Consequently, listeners were more likely to perceive [ukʷɑ] and [utɑ] as unambiguously belonging to different categories, and hence, as the generations proceeded, speakers were increasingly more likely to produce [ukʷɑ] and continue to produce [utɑ] in their own speech, as a consequence of probability matching. It is quite likely that the very words at greatest risk of homophony – words in dense lexical neighborhoods – would lead such a shift. Indirect support of this hypothesis comes from studies by Port and Crawford (1989) and Charles-Luce (1993). Both reports find that in semantically ambiguous contexts, speakers are less likely to implement genuinely neutralized variants of potentially homophonous forms than they are in semantically unambiguous contexts. Charles-Luce writes (1993: 41), "This is not to suppose that this is conscious behavior. It may be quite automatic and learned through experience with communication," although Charles-Luce seems to give experience with speaking at least as much of a role to play as experience with listening. It is also possible that token frequency factors influence such changes as well. Certain frequent words might lead the change, only to be followed by others as the change diffuses through the lexicon (e.g. Bybee 2001).

Consider, then, an impressionistic formulation of the proposed mechanism, portrayed in (7). This scenario demonstrates how very minor phonetic tendencies, coupled with the sporadic lexical semantic ambiguities they might induce or eschew, may eventually have far-reaching consequences for the phonological system. Moreover, it is consistent with the fact that sound change is probabilistic, and not deterministic. Not every Mixtecan language underwent the sound change that Trique did. There simply exists a probability that any given sound change will take hold in any given speech community. Probabilities may be affected by, among many other factors, the language-specific system of contrasts, and the contrastive values' functional load: in Trique the introduction of labio-velars was contrast-enhancing, since spreading did not induce homophony. In some other language, a prevalence of contrastive labialized velars might very well passively induce the curtailment of such a sound change (see Öhman 1966, Manuel 1990, 1999 for suggestive synchronic evidence).

(7) Diachronic forces at work

[ukɑ......uk̹ɑ......uk^{w}ɑ]		[utɑ......ut̹ɑ......ut^{w}ɑ]	
less distinct from [utɑ]	more distinct from [utɑ]	more distinct from [uk$^{(w)}$ɑ]	less distinct from [uk$^{(w)}$ɑ]
less likely perceived unambiguously	more likely perceived unambiguously	more likely perceived unambiguously	less likely perceived unambiguously
less likely produced	more likely produced	more likely produced	less likely produced
∴ gradual move towards [uk^{w}ɑ]		∴ stability of [utɑ]	

5.2. Multiple trace theory

The speech signal is rife with phonetic detail to which listeners are demonstrably sensitive as they listen and learn, since they largely recapitulate in their own speech the very variation which they perceive. Indeed, exactly because gradience and variation are conventionalized in the observed manner, we have clear behavioral evidence that it is part of speakers' phonological knowledge. According to multiple trace theory, also known as "episodic" or "exemplar" theory (Gluck and Bower 1988, Goldinger 1997, 1998, Johnson 1997, Kruschke 1992, Nosofsky 1986, 1988, Pierrehumbert 1994, 1999, 2001a,b, Steels 2000, Bybee 2001, Lotto 2000, Wright 2003), emergent perceptual categories are defined as the set of all experienced instances of the category, such that variation across exemplars actually contributes to the categorical properties themselves. It is knowledge of – and sensitivity to – this variation that surely influenced Martinet's assertion that "discreteness does not rule out infinite variety."

With both probability matching and multiple trace theory to work with, we are now able to draw some preliminary theoretical conclusions regarding the diachrony of labial spreading in Trique. The conventions established by speech communities betray a nuanced mastery of the phonetic variation internalized by individual speakers that is demonstrably a part of these speakers' linguistic knowledge. The exquisite articulatory control that speakers display in their productions is best evidenced by the fact that they are able to largely match the variation present in the ambient pattern. On this view, learners' articulatory talents may be harnessed largely in service to *copying* or *imitating*, not *modifying* (improving upon or otherwise) the ambient speech pattern. The facts of probability matching are thus consistent with the hypothesis that categorical phonological (phonetic) targets may not exist. Rather, consistent with multiple trace theory, the target of phonological acquisition may be the gradience and variation itself. But still, speakers' mimetic talents are not perfect. Stray tokens are inevitable, and it is the functional benefit of certain of these strays which might ultimately take hold in a system and come to permeate the lexicon. If language theorists insist on maintaining a distinction between speakers' phonetics and phonology, then we might say that genuine strays are phonetic, while all variation that is probability-matched is phonological in origin.

6. Conclusion

In this study I have employed the laboratory in an attempt to recapitulate the forces responsible for a sound change which took place in Trique. On the working assumption that some sound changes are a consequence of listeners misinterpreting the words intended by speakers, the proper laboratory conditions may reflect real-world historical patterns in compressed form. The operative assumption in the present experiment has been that noise introduced into the speech signal might induce a "speeded-up" rate of misperception in certain contexts, and thus reflect one origin of real-world sound change. Given that language learners largely (though imperfectly) match the variation they perceive, the sorts of perceptual errors induced in the present study might only reflect the culmination of a slow, generation-to-generation accretion of such errors, rather than offering any major insights into the online processing of natural speech. The gradience and variation inherent in speech production may be the fodder for these sorts of sounds changes: the more distinct the variant from an acoustically similar word, the more likely that it will be interpreted correctly, and so the more likely the system will wend towards this value. In the present experiment, the least confusable forms ([udɑ] and [ugʷɑ]) are exactly those which actually seem to have evolved in Trique from more confusable forms ([udɑ] and [ugɑ]).

I emphasize that the proposals advanced herein should not be interpreted in teleological terms. Language is not inexorably headed toward an optimal state. Language evolution is unguided and passive, just as in the evolution of species. (See also Keller 1994, and Croft 2000 for extended discussion.) Speakers are probably no more "altruistic" than are stowaway rats on a garbage scow. Indeed, for every case of contrast-enhancing sound change (typically, as in the Trique case, found in pre-vocalic or stressed contexts, where there is greater opportunity for contrast-enhancing variation), we might encounter a case of contrast-merging sound change (typically found in pre-consonantal or stressless contexts, where there is less opportunity for contrast-enhancing variation).

Finally, I note that the present psychoacoustic findings do not bear directly on the issues of phonological categorization or probability matching, since no meanings were associated with the sound sequences. Nonetheless the findings may be seen as consistent with the sorts of diachronic scenarios that are likely, given the facts of probability matching, and given the theoretical assumptions of multiple trace theory.

Acknowledgments

Warm thanks to "my loyal subjects", to Margaret Russell for running the experiment, Kyle Chambers for help with PsyScope, Molly Mack for help with statistics, Wouter Jansen for answering my queries, and Devin Casenhiser for help with *everything*. Also, many thanks to Doug Whalen, Keith Johnson, and another LabPhon reviewer for detecting a kernel of something interesting in the terribly rough and rushed first draft of this paper.

References

Anttila, Raimo
1972 *An Introduction to Historical and Comparative Linguistics.* New York, MacMillan.

Bell-Berti, Fredericka and Katherine S. Harris
1979 Anticipatory coarticulation: some implications from a study of lip-rounding. *Journal of the Acoustical Society of America* 65: 1268–1270.
1982 Temporal patterns of coarticulation: Lip rounding. *Journal of the Acoustical Society of America* 71, 449–454.
Bybee, Joan
2001 *Phonology and Language Use.* Cambridge Studies in Linguistics 94, Cambridge University Press.
Charles-Luce, Jan
1993 The effects of semantic content on voicing neutralization. *Phonetica* 50: 28–43.
Fry, Dennis B.
1947 The frequency of occurrence of speech sounds in Southern English. *Archives Néerlandaises de Phonétique Experimentales* 20.
Gallistel, Randy
1990 *The Organization of Learning.* Cambridge MA, MIT Press.
Gluck, Mark A. and Gordon H. Bower
1988 Evaluating an adaptive network model of human learning. *Journal of Memory and Language* 27: 166–195.
Goldinger, Stephen D.
1997 Words and voices: perception and production in an episodic lexicon. In: Keith Johnson and John W. Mullennix (eds.), *Talker variability in speech processing.* San Diego: Academic Press. 33–56.
1998 Echoes of echoes? An episodic theory of lexical access. *Psychological Review* 105.2: 251–271.
Grimes, Barbara F.
2003 *Ethnologue: Languages of the World*, 14th Edition. Dallas: SIL International.
Gudschinsky, Sarah C.
1959 Proto-Popotecan: A Comparative Study of Popolocan and Mixtecan. Indiana University Research Center in Anthropology, Folklore and Linguistics Memoir, 15. Bloomington: Indiana University.
Harper, D. G. C.
1982 Comparative foraging in mallards: ideal free ducks. *Animal Behaviour* 30: 575–584.
Hockett, Charles F.
1968 *The State of the Art.* Janua Linguarum. Series minor number 73. The Hague: Mouton.

Hollenbach, Barbara E.
1977 Phonetic versus phonemic correspondence in two Trique dialects. In: W.R. Merrifield (ed.), *Studies in Otomanguean Phonology,* 35–68. Dallas: Summer Institute of Linguistics.
Hudson, Carla L. and Elissa Newport
1999 Creolization: could adults really have done it all? In A. Greenhill et al. (eds.), *Proceedings of the 23rd Annual Boston University Conference on Language Development: Vol. 1*., 265–276. Somerville, MA: Cascadilla Press.
Janda, Richard D. and Brian D. Joseph
2001 Reconsidering the canons of sound change: towards a Big Bang theory. Manuscript of paper presented at the International Conference on Historical Linguistics, Melbourne.
Kingston, John
2002 Keeping and losing contrasts. Presentation at the 28th Annual Meeting of the Berkeley Linguistics Society.
Kruschke, John K.
1992 ALCOVE: an exemplar-based connectionist model of category learning. *Psychological Review* 99.1: 22–44.
Labov, William
1994 *Principles of Linguistic Change, Internal Factors.* Oxford: Blackwell.
Liberman, Mark
2002 Simple models for emergence of a shared vocabulary. Paper presented at Laboratory Phonology VIII. New Haven, CT.
Longacre, Robert E.
1957 Proto-Mixtecan. *International Journal of American Linguistics* 23.4.III. Indiana University Research Center in Anthropology, Folklore and Linguistics. Bloomington: Indiana University.
1962 Amplification of Gudschinsky's Proto-Popolocan-Mixtecan. *International Journal of American Linguistics* 28: 227–242.
Longacre, Robert E. and René Millon
1961 Proto-Mixtecan and Proto-Amuzgo-Mixtecan Vocabularies: A Preliminary Cultural Analysis. *Anthropological Linguistics* 4.3: 1–44.
Lotto, Andrew
2000 Language acquisition as complex category formation. *Phonetica* 57: 189–196.
Manuel, Sharon
1990 The role of contrast in limiting vowel-to-vowel coarticulation in different languages. *Journal of the Acoustical Society of America* 88: 1286–1298.

1999 Cross-language studies: relating language-particular coarticulation patterns to other language-particular facts. In: William J. Hardcastle and Nigel Hewlett (eds.), *Coarticulation: Theory, Data and Techniques*, 179–198. Cambridge, UK: Cambridge University Press.

Martinet, André
1975 The internal conditioning of phonological changes. Published in *La Linguistique* 24.2: 17–26 [1988].

Nosofsky, Robert M.
1986 Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General* 115: 39–57.
1988 Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 14: 700–708.

Ohala, John J.
1981 The listener as a source of sound change. Papers from a Parasession on Language and Behavior, Chicago Linguistics Society, 178–203.
1989 Sound change is drawn from a pool of synchronic variation. In: Leiv E. Breivik and Ernst H. Jahr (eds.), *Language Change: Contributions to the Study of its Causes,* 173–198. Mouton de Gruyter, Berlin.

Öhman, Sven E. G.
1966 Coarticulation in VCV utterances: spectrographic measurements. *Journal of the Acoustical Society of America* 39: 151–168.

Paul, Hermann
1886 *Principles of the History of Language* (translated from 2nd edition by H.A. Strong, 1970). College Park: McGroth Publishing Company.

Pierrehumbert, Janet B.
1980 The phonology and phonetics of English intonation. Ph.D. dissertation, MIT.
1994 Knowledge of variation. Papers from the Parasession on Variation, 30th Meeting of the Chicago Linguistics Society, Chicago Linguistics Society, 232–256.
1999 What people know about sounds of language. *Studies in the Linguistic Sciences* 29.2: 111–120.
2001a Stochastic phonology. *GLOT* 5.6: 1–13.
2001b Exemplar dynamics: word frequency, lenition, and contrast. In: Bybee, Joan and Paul Hopper (eds.), *Frequency and the Emergence of Linguistic Structure,* 135–157. Amsterdam: John Benjamins.

Poplack, Shana
1980a The notion of the plural in Puerto Rican Spanish: competing constraints on (s) deletion. In W. Labov (ed.), *Locating Language in Time and Space*, 55–67. New York: Academic Press.

1980b Deletion and disambiguation in Puerto Rican Spanish. *Language* 56 (2): 371–385.

Preston, M. S. and G. Yeni-Komshian

1967 Studies on the development of stop consonants produced during the second year of life. *Annual Report, Neurocommunications Laboratory, Johns Hopkins University School of Medicine* 3: 211–222.

Rensch, Calvin R.

1976 *Comparative Otomanguean Phonology.* Bloomington: Indiana University.

Silverman, Daniel

1993 Labiality in Mixtecan – a unified treatment. In: Daniel Silverman and Robert Kirchner (eds.), *UCLA Occasional Papers in Linguistics* 13: 109–123.

Steels, Luc

2000 Language as a complex adaptive system. In: Marc Schoenauer, Kalyanmoy Deb, Guenter Rudolph, Xin Yao, Evelyne Lutton, Juan Julian Merelo, Hans-Paul Schwefel (eds.), *Lecture Notes on Computer Science. Parallel Problem Solving from Nature.* PPSN-VI. 17–26.

Wright, Richard

2003 Factors of lexical competition in vowel articulation. In John Local, Richard Ogden, and Rosalind Temple (eds.), *Papers in Laboratory Phonology VI*, Cambridge University Press. 75–87.

I. Qualitative and variable faces of phonological competence

Spoken languages

Convergences and divergences of signed and spoken languages

Signed languages

Effects of language modality on word segmentation: An experimental study of phonological factors in a sign language*

Diane Brentari

This paper analyzes the word-segmentation strategies used in signed and spoken languages. The claim is that there is a strong "modality effect" in the word segmentation strategies used in these two types of languages. The experimental results show that both signers and non-signers use place of articulation and movement more heavily than handshape to make word segmentation judgments; however, signers are more sensitive to handshape than nonsigners are for making such judgments. This work shows that there are strategies for segmenting visual language input that are different from those used in segmenting auditory language input, regardless of language exposure. From the experimental evidence presented here and from the work on word segmentation in spoken language, one can conclude that viewers *use a word-sized unit itself to segment strings into words, which is argued to be due in large part to the visual/gestural nature of sign languages. In contrast,* listeners *depend most heavily on the syllable in their word segmentation strategies, which is argued to be due the auditory/vocal nature of spoken languages. This work can at least partially explain the variation in the well-formedness constraints on found in signed and spoken languages, which capitalize on this modality effect.*

1. Introduction

From the structural point of view, sign languages are more like each other than they are like the surrounding spoken languages in their locale, and they constitute a distinct typological class among languages with respect to phonology and morphology in a number of ways (van der Hulst 2000, Brentari 2002). An important question is how much of this similarity is due to modality effects – i.e., the fact that sign languages use the hands, face and body to produce visible signals while spoken languages use the vocal tract to produce audible signals – rather than to the relatively short period that sign languages have existed. One can address the impact of modality on language structure in a number of ways. One way is to do cross-linguistic work on as many sign languages as possible, both on comparatively younger and older sign languages, and to study these languages from as many different genetic sources as possible. This type of work is underway in a variety of forms, asking a range of questions about general typological features (Zeshan 2000), morphological characteristics (Mathur 2001; Aronoff et al. 2003), and sign language genesis (Senghas 1995; Kegl, Senghas and Coppola 1999). Studying the time course of language acquisition in signed and spoken languages has also been a way to uncover modality effects or lack thereof (Meier 1990; Petitto and Marentette 1991). Another way to approach this question is in the laboratory, by constructing psycholinguistic experiments that illuminate modality issues. This approach has been used by Singleton, Morford, and Goldin-Meadow (1993), Supalla (1990), Emmorey (2001 and references therein), to name a few. Finally, from the perspective of theory, there is a body of work that has investigated where modality effects might be responsible for the structures that represent signed and spoken languages at the level of phonology (e.g., Stokoe 1960; Liddell and Johnson 1989; Sandler 1989; Brentari 1998), syntax (e.g., Padden 1983; Lillo-Martin 1991; Neidle et al., 2000) and morphology (e.g., Supalla 1982; Mathur 2001; Liddell 2003). In this chapter, the focus is on the modality question and its relation to phonology – particularly, its effect on word segmentation. The analysis utilizes what has been discovered about phonology in sign languages over the last 40 years to examine the phonological well-formedness constraints in signed and spoken languages. These phonological constraints established for American Sign Language (ASL) are used to construct an experimental task that will help disambiguate the effects of language exposure (in this case, exposure to ASL or English) from modality effects.

This chapter is organized as follows. First, the rationale for studying word segmentation in sign languages is given and predictions made for the outcome of the empirical study. In Section 1.1, the use of cue conflict as a design for studying word segmentation is motivated, and in Section 1.2 enough information is provided about the structure of signed words so that the properties under investigation will be well understood. Second, the empirical study is presented and the results discussed. Finally the broader implications of the findings are discussed, especially as they bear upon discussions of word segmentation in the literature on spoken languages.

1.1. Background on Word Segmentation

Word segmentation is the competency that allows language users to break up a string of an uninterrupted language signal into shorter, more manageable pieces that are subjected to further analysis; it helps us identify word breaks (Mattys and Jusczyk 2001). We have all experienced watching a sign language or listening to a spoken language in an unfamiliar language setting, and having the sensation that the signal is continuous, without breaks; this is the lack of the ability to segment the string into words. Word segmentation differs from lexical access, which is the competency that allows us to identify a word with a particular meaning, because it is not necessary to know the meaning of the word-sized units to segment the string; it is sufficient to recognize where the words begin and end. Babies are able to this well before they have a lexicon, between 6–9 months of age (Mattys and Jusczyk 2001).

Individuals segment words by utilizing phonetic properties, such as pause length, and by the phonological properties of individual languages. In spoken languages, this might be the allophonic variation that appears in different parts of the word. For example, in English, it includes the ability to recognize that if one hears an aspirated /t/ ([t^h]), it must be the beginning of a word (Jusczyk, Hohne and Bauman 1999). In a sign language, it might be the ability to recognize that two handshapes, such as the W-handshape and the H-handshape in ASL, do not appear together in a single word, and hence there must be a word break between them (Friedman, 1977, Mandel 1981). The study of word segmentation is important because it is a key component in acquiring a language, and because this skill seems to involve both an innate competence and an acquired skill. To what extent is it a part of our innate abilities (i.e., competence) and to what extent is it tied to exposure to properties of a particular language or a particular type of language (i.e., performance)? Is it a general type of learning that could apply, for example, to both strings of linguistic units or strings of music, or is it a special type of language learning? And most relevant for the study reported here, how much of word segmentation depends on language modality – i.e., whether it is produced by the hands and body in space or by the vocal tract?

An experimental design using cue conflict has been useful in studying word segmentation effects in spoken languages in order to determine their strength and importance with respect to one other in making such judgments; it allows a ranking of strategies to emerge from most resilient to least resilient (Cutler et al 1986, Jusczyk et al. 1993; Suomi et al 1997; Vroomen et al 1998; Houston, et al 2000). In spoken languages, for example, word stress might be put into conflict with allophonic behavior in English or with vowel harmony in Finnish, in order to determine which of these aspects of the phonology is more important for listeners. In the next section, the properties that will be put into conflict in the sign language experiment are explained. This technique will be employed here to investigate questions of word segmentation in a sign language. In particular, we want to know if adults without exposure to a sign language have a systematic strategy for segmenting signed strings, and if they do, whether it is the same as or different from the word segmentation strategies used by signers. We will address whether non-signing individuals are equipped to do this purely on the basis of intuitions about how a signal produced by the hands, arms, face and body might do this. We will examine the experimental results to determine the degree to which users of spoken languages rely on learned, spoken language strategies to segment signed strings. In the experiment reported on here, one group of participants consists of individuals who do not know a sign language. To insure that the signers and non-signers are being presented the same task, nonsense words must be used, rather than actual signs, because having lexical access to the signs would give signers an advantage.

One of the questions concerning the differences in word segmentation judgments between signers and non-signers is the extent to which each group is able to use the different types of simultaneous information in the signed signal. Both signed and spoken languages have paradigmatic (simultaneous) information and syntagmatic (sequential) information used in creating contrast between words, but it has been argued that sign languages make more extensive use of simultaneous information in their phonology while spoken languages make more extensive use of sequential information (Brentari 2002; Meier 2002). It would, therefore, be plausible to expect that signers might be more sensitive to simultaneous information in making word segmentation judgments.

1.2. Sign Language Phonological Structure

There is general consensus that there are five basic phonological parameters (or feature classes) of sign language phonology: handshape, place of articulation, movement, non-manual behaviors, and orientation. We will not directly test orientation or non-manual behavior in the experiment (although we control these factors to prevent experimental confounds), so we will not discuss them at length here. Our experiment will place handshape, place of articulation, and movement parameters in conflict with one another to assess the strength and importance of each for signers and non-signers in making word segmentation judgments. In general, we predict that exposure to the language-particular phonology of ASL will affect word segmentation judgments in signers, but not non-signers. Some useful information about these three parameters of ASL phonology is explained in the following paragraphs.

Handshape is the particular form the hand assumes when producing a sign; there are approximately 40–45 distinctive handshapes in ASL, specified by approximately 13 distinctive features (Brentari 1998). Fingers are divided into those that move or have physical contact with a place of articulation, called the *selected fingers*, and those that do not, the *non-selected fingers*. The selected fingers have features of *joint configuration, quantity* and *point of reference*. Quantity and point of reference express the number of fingers and where on the hand they are located (the 1-handshape and the V-handshape in APPLE, and NERVE differ in quantity: Figure 1), while the 1-handshape and the I-handshape, made with the single pinkie finger (Figure 2), differ in point of reference. Joint configuration of the selected fingers specifies the joints of the fingers that are [flexed] or extended when articulating a specific handshape. The handshapes in CANDY and APPLE differ in joint configuration (Figure 1). When the fingers are extended, this is not expressed in the representation because, based on its phonological behavior and appropriate use of underspecification theory (Archangeli, 1984), 'extended' is the default specification, for joints in ASL (Brentari 1998: 162).

Figure 1. CANDY – 1st position and 2nd position (left); APPLE – 1st and 2nd position (middle); NERVE – 1st position and 2nd position (right).

Handshapes have been divided into marked and unmarked sets (Battison 1978), based on the selected fingers and the joints that are involved (Figure 2). The unmarked set uses either all of the fingers in a range of joint configurations (i.e., B, A, S, C, O, 5) or solely the extended index finger (i.e., 1); the marked handshapes use either other fingers – e.g., I (i.e., the pinkie finger), H (i.e., the index and middle fingers), W (i.e., the index and middle fingers), or other joint configurations of the fingers, such as the X-handshape that occurs in APPLE (Figure 1). The handshape in APPLE is considered marked with respect to joint configuration, since it has the 1-handshape, but it is bent at the interphalangeal joint. NERVE (Figure 1) is marked both because it has marked selected fingers (the index and middle finger) and because it is bent.

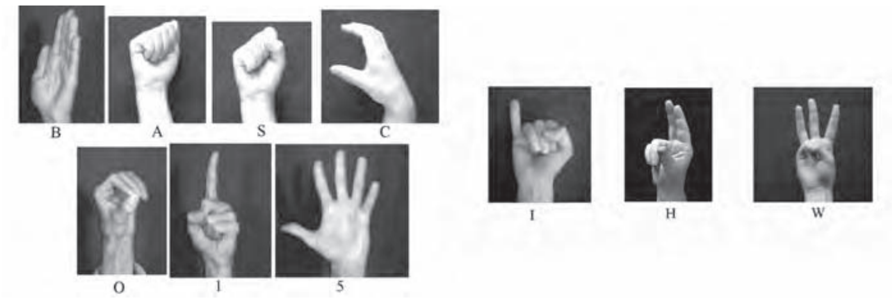


Figure 2. Unmarked ASL handshapes (left), and three marked handshapes (right).

Handshape will be manipulated in the experimental design by being divided into marked and unmarked groups. This is a fundamental distinction because even signs with two handshapes typically include one or both unmarked handshape(s), such as occurs in the following lexicalized compounds: FACE^STRONG=RESEMBLANCE (handshapes: 1→S), MIND^DROP=FAINT (handshapes:1→5), WATER^RISE=FLOOD (handshapes: W→5).[1] Handshapes in monomorphemic forms may change allophonically in aperture ([open]←→[closed]; see DESTROY, Figure 3), but not in selected fingers (Friedman 1977, Mandel 1981). These allophonic handshape sequences have been argued to be tied to segmental units (Sandler 1986; Liddell and Johnson 1989; Perlmutter 1992; Brentari 1998). The handshape aperture change itself is also discussed further below, because this dynamic change is classified as a movement (specifically, a local movement) within the phonology.

Place of articulation has distinctive regions of the *body* – head, torso, arm, and the non-dominant hand (also called H2, to distinguish it from the dominant signing hand, or H1) – as well as the three dimensional planes (e.g., the *horizontal plane*, the *vertical plane*, and the *midsagittal plane*; see Figure 3). In monomorphemic signs, while there may be allophonic changes within a distinctive region, there is typically just one distinctive place of articulation, (Sandler 1987). For example, in DESTROY (Figure 3), the two hands move allophonically from the ipsilateral to the contralateral sides of the horizontal plane and back again, but the region remains the horizontal plane throughout the sign's production. Other examples are signs such as DEAF, MEMBER, NAVY, all of which have a change in allophonic places of articulation within a region, but do not change distinctive place of articulation from one region to another. In polymorphemic forms, however, such as the lexicalized compounds mentioned above (MIND^DROP=FAINT, WATER^RISE =FLOOD, FACE^STRONG=RESEMBLANCE), there may be two distinctive places of articulation. In the present experimental design, the number of distinctive places of articulation is manipulated. Items contain either 1 or 2 distinctive places of articulation. Changes in place of articulation, like those of handshape, have been argued to be tied to segmental units (Sandler 1987, Brentari 1998).

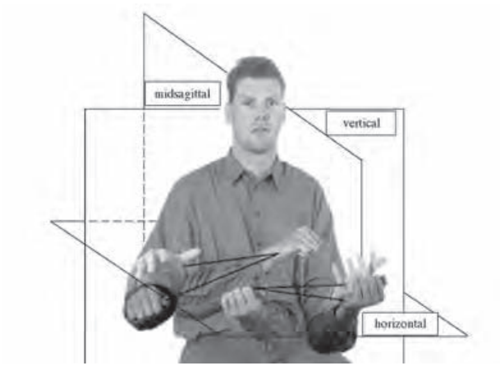


Figure 3. The ASL sign DESTROY, and the three dimensional planes utilized for places of articulation (those not on the body) are superimposed on the figure.

The Movement parameter is divided into path and local movements (Brentari 1998). Path movements are articulated by the elbow or shoulder; local movements are articulated by the wrist or hand. All movements may also have a shape (circle, straight, and arc). All of the signs in Figure 1 have a local movement, namely an orientation change, specifically [supination] of the wrist. The sign DESTROY has a sequence of two, straight path movements, the second of which is a complex movement. The second movement of DESTROY has, simultaneously, both a [straight] path movement and a change from an [open] to [closed] variant of the 5-handshape, which is a local movement (Figure 3).

Movements have been argued to be the "vowels" in ASL phonology (Liddell 1984) and also the syllable nuclei of the signed word, based on the way that they behave in the phonology (Brentari 1990a, Perlmutter, 1992). For example, all well-formed signs must have a movement, just as virtually all words have a syllable nucleus, and the notion of sonority can be calculated based on the size and excursion of movements of joints (Corina 1990; Brentari 1990a, 1998; Sandler 1993), analogous to the way that sonority in spoken languages can be calculated based on the size and excursion of mandibular movements (MacNeilage and Davis 1993; MacNeilage 1998). APPLE, NERVE, and CANDY have one syllable; DESTROY has two syllables. While most signs are monosyllabic, there are also a number of monomorphemic words, such as DESTROY (also REMOVE, GOVERNMENT, MONOPOLIZE, APPOINTMENT, LOCK) that are disyllabic. See Brentari (1998: 47) for a discussion of these forms. Moreover, lexicalized compounds, such as those mentioned earlier, are also often disyllabic. Repetitions of movement and a combination of circle+straight movement are permissible; other movement combinations are not permissible in a single disyllabic sign, and no ASL sign has more than two syllables (Uyechi 1996). The experimental design manipulated the number of movements (one vs. two) and the quality of the changes in 2-movement forms; the stimuli include permissible monosyllabic items, permissible disyllabic items, and non-permissible disyllabic items. The distinctive and allophonic properties of the three parameters, and the ways that they were manipulated in the empirical study are given in Table 1.

Table 1. The distinctive and allophonic properties of handshape (HS), place of articulation (POA), and movement (M).

	HS	POA	M
distinctive properties	selected fingers and joint configurations	body regions and dimensional planes	shapes: arc, circle, straight
allophonic properties	aperture changes	areas within a given region	repetitions
properties manipulated in the experiment	1. combinations of marked and unmarked HSs 2. changes permissible within a word and those occurring only across a word boundary	1. number of distinctive regions (1 vs. 2)	1. combinations of path and local movement 2. mono- and disyllabic forms

An important generalization can be stated regarding all three parameters discussed here. Namely, within a word, allophonic changes in handshape, place of articulation, and movement are permitted, but a change from one distinctive handshape, place of articulation, or movement to another is not permitted, except in special cases, such as compounds. Even in compounds, there is a tendency toward unmarked combinations of handshapes and permissible combinations of movements. It is predicted that the phonological structure of ASL will influence the word segmentation judgments of signers, but not of non-signers.

1.3. Modality Effects

Modality effects are defined here as the predispositions in signed and spoken languages to organize linguistic material in a manner that capitalizes on the relative strengths of the particular phonetic systems involved. A fundamental assumption employed in this work is that signed and spoken languages use different organizational strategies to present phonological information. These differences affect not only the substance of individual features and feature classes (e.g., the difference between having a [nasal] feature on vowels vs. a [circle] feature on movements), but also our intuitions about where word breaks might be found. For example, concerning the use of sequential (syntagmatic) vs. simultaneous (paradigmatic) organization in the two types of languages, Meier (2002) and Brentari (2002) have argued that factors such as the following allow sign languages to transmit more simultaneous information in the linguistic signal at a given moment in time: (i) the relatively fast speed of light vs. the relatively slow speed of sound (Bregman 1990), and (ii) the differences in so-called 'tone fusion,' which is the relatively short time needed between two auditory signals in order to perceive them as discrete, versus 'flicker fusion,' which is the relatively long time needed between two visual signals in order to perceive them as discrete (Chase and Jenner 1993).

The primary motivation of the experiment is to investigate whether the different organizational patterns to which speakers or signers are exposed affect word segmentation judgments a different language modality. Are we, regardless of language experience, capable of using the patterns of organization in a different modality effectively without prior experience? In the following sections two examples of modality differences in the phonology of signed and spoken languages are explained, and their potential relationship to the following empirical study is spelled out.

1.3.1. The segment in signed and spoken languages

It has been argued in that one modality difference between signed and spoken languages is the relative importance of segments (van der Hulst 2000, Brentari 2002, Channon 2002a, 2002b). With respect to the prominence of phonological units in a features geometry, the position of the segment – understood here as timing unit – is fundamentally different in signed and spoken languages. The difference can be stated as follows: In a spoken language hierarchy of units, segments are autonomous from features, dominate features, and can create minimal contrast (i.e., segments >> features). In a sign language hierarchy of phonological units, feature information predicts and dominates segmental units (i.e., features >> segments). This can be seen in the use (or lack of use) of segmental orders to create new words, and in the unit upon which minimal pairs are based. Both phenomena are described below.[2]

Regarding segment order, the degree of variability in order and content of segments in spoken words is greater than in signed words. 'Teen' and 'neat' are different, unrelated words in English, using the same phonological segments in different orders. In ASL, this type of contrast is severely limited (Channon 2002a, 2002b). Except in the case of compounds, the order of handshapes in a sign is largely predictable; the only handshape that can precede or follow a [closed] handshape is its [open] variant and the only place of articulation that can precede or follow a specified place is an allophonic place within that region. Most counter-examples to this generalization – i.e., pairs of signs, similar to 'neat'/'teen' with contrasting order in ASL are signs with the opposite meaning, such as CATCH (Hs1→Hs2) vs. THROW (Hs2→Hs1); IMPROVE (POA1→POA2) vs. GET-WORSE (POA2→POA1), showing that segment order might have a morphological rather than a phonological use. Since changes in handshape and place of articulation are argued to be segmental properties, as discussed earlier, by manipulating these parameters in the experiment, we will be able to determine the effects of the segment in making word segmentation judgments in our groups.

The unit upon which minimal pairs are based also shows a lack of reliance on the segment. Minimal pairs in ASL involve the unit of the entire word, rather than the syllable or the segmental unit.[3] For example, in ASL, the surface difference between the signs CANDY and APPLE (Figure 1) is based on the presence vs. absence of a [flexed] feature located at the joints structure – APPLE has a joints structure, given in bold in Figure 4a, and CANDY has none. CANDY has no specification for joints because the index finger is extended, which is the default setting for joints (Brentari 1998:107).

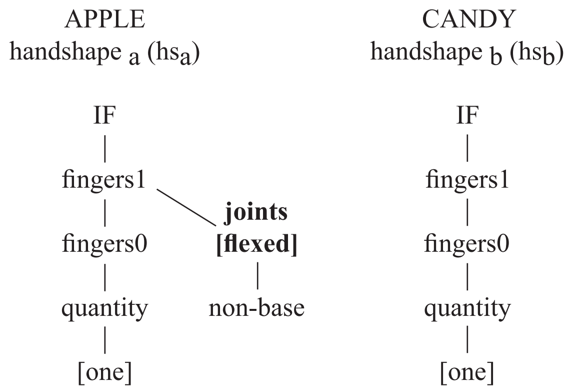


Figure 4. Representations of APPLE (left) and CANDY (right) in the Prosodic Model (Brentari, 1998), showing the placement of the branch structure responsible for creating a minimal pair in bold.

In the Prosodic Model, the concept of a consonant is taken to be the combination of handshape and place of articulation specified for the sign; these are the *inherent features* (IF) of the sign. This idea was first proposed in Chinchor (1978), but it was developed within a contemporary framework of feature geometry and autosegmental phonology in Brentari (1998). In Figure 4, only the handshape information is given, because it is the only structure relevant to this particular contrast. In the Prosodic Model representation, the joints structure spans the entire word domain – in this case, a monosyllabic word domain – not the segmental domain as expected in a spoken language (e.g., the initial consonant in [tin] 'teen' and [din] 'Dean'). In previous work on sign language phonology, a model more similar to that of spoken languages was used. Consider the representation in Figure 5a, which is a representation in the Hold-Movement model of sign language phonology (Liddell 1984; Liddell and Johnson 1989). In this model, periods of stasis were considered the consonantal units of ASL. A sign such as APPLE or CANDY would have three segments (stasis-movement-stasis), and the difference in handshape would be expressed on all three segments. Since the “joints/no joints” distinction is expressed on each of the three segments (hs_s represents the handshape in APPLE with the finger joints flexed, and hs_b represents the handshape in CANDY with the finger joints not flexed) these two forms cannot therefore be considered a minimal pair in the Hold-Movement model. In contrast, in Figure 5b, the handshapes (and their features) of CANDY and APPLE are represented only once in the representation. The important point here is that in the Hold-Movement representation, it is not possible to call APPLE and CANDY a minimal pair in ASL, yet this is a contrast accepted as such by native signers. There is only one specification for handshape in the representations in Figure 5b, which allows these to be called a minimal pair.

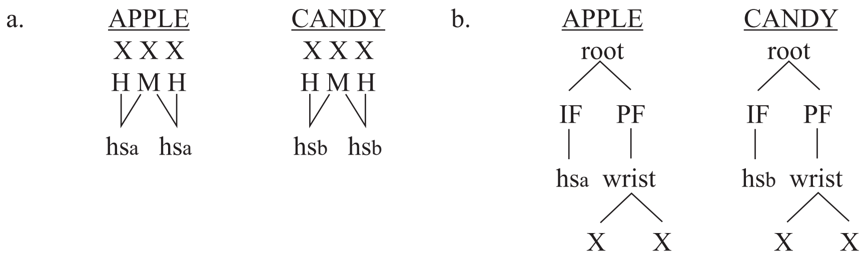


Figure 5. a. The Hold-Movement Model (Liddell and Johnson, 1989) and
b. the Prosodic Model (Brentari, 1998) representations of APPLE and CANDY, with hs_a and hs_b substituted for the complete representations of their respective handshapes

1.3.2. The syllable in signed and spoken languages

The syllable has been demonstrated through experimental evidence to be a fundamental unit in word segmentation in spoken languages (Jusczyk et al. 1993, Jusczyk, Hohne and Bauman 1999; Mattys and Jusczyk, 2001). In adults, there is evidence that when properties such as vowel harmony, which are expressed at the word-level, are put into conflict with stress, the syllabic cue of stress is more heavily relied upon (Suomi et al 1997; Vroomen et al. 1998). Moreover, words in most spoken languages have binary metrical feet, where feet often are defined in terms of syllables (Halle and Vergnaud 1987; Hayes 1995). All of these factors indicate that there is at least a statistical preference for words to be longer than one syllable in spoken languages, and that spoken word segmentation relies more heavily on syllabic than segmental or word-level properties.

This contrasts with the relatively low use of the syllable in sign languages. The majority of ASL monomorphemic signs are one syllable (93%, based on the Stokoe et al.,1965, dictionary). However, within the 7% of disyllabic lexical items, there are four types (Brentari, 1998: 186): (i) those that repeat the movement in some form, with no change in handshape (e.g., CHILDREN, COUGH, MILITARY, LEATHER); (ii) those that repeat the movement in some form, with an allophonic change in handshape on one movement only (e.g., DESTROY, EXPENSIVE, NOTE-DOWN); (iii) those with two distinctive movements, with an allophonic change in handshape (e.g., GOVERNMENT, REMOVE, JUMP); and (iv) those that repeat the movement in some form with a contrastive change in handshape (e.g., BACKGROUND, SOCIAL-WORK, NOTRE-DAME). As argued in Brentari (1990, 1998) and Perlmutter (1992) the syllable is the unit that best captures the co-ordination of the timing of handshape with movement in such forms. For example, a handshape change such as the one in DESTROY is co-extensive with an individual movement; it does not extend throughout an entire disyllabic sign. By manipulating this set of disyllabic forms in ASL in an empirical study, word-domain effects can be disambiguated from syllable effects in order to determine if the strong reliance on syllables by users of spoken languages (in speech) carries over into their word segmentation judgments of signs, and to see if signers employ the syllable in their own word segmentation judgments.

2. Experimental Method

As we have explained above, ASL utilizes the unit of the word in capturing minimal pairs and in the distribution of distinctive handshapes, movements and places of articulation. ASL also utilizes the syllable to organize the timing of handshape changes across a single movement, but the number of monomorphemic disyllabic forms is rather small with respect to the lexicon as a whole. The segment is used relatively little in ASL, as we have seen in the previous section. In spoken languages, the syllable has been shown to be the unit most heavily relied upon in making word segmentation judgments, but segmental information, such as allophonic distribution, and word-level effects, such as vowel harmony, play a role as well.

The hypotheses of the experiment are formulated in order to investigate differences in language experience and language modality, and the role of phonological units in word segmentation judgments. First, we hypothesize that sign experience will not play a dominant role in word segmentation judgments. That is, although experience with spoken languages predisposes speakers to use the syllable or segment to make word segmentation judgments in spoken languages, we expected both speakers and signers to use word level information for word segmentation judgments in a sign language. Thus, a modality effect was expected, whereby nonsigners (speakers) adapt to the new language modality by shifting to a word-level segmentation approach. Second, we nonetheless expected experience to affect phonological contributions to word segmentation. Specifically, we hypothesized that signers would utilize more simultaneous phonological cues than nonsigners. We also expected language-particular phonological constraints (distribution of sequences of handshapes, places of articulation, and movements) to affect word segmentation by signers but not by nonsigners.

2.1. Participants

Two groups of participants were employed in this study: 13 Deaf native ASL signers (18–50 years), and 13 hearing, non-signing individuals (15–30 years). A "native" signer learned ASL from his/her parents as a first language; our participants were also profoundly deaf, went to residential schools for the Deaf, have Deaf spouses, and have used ASL as their primary language throughout their lives. They would be expected to show the strongest effects of the grammar on word segmentation. The non-signing participants had never had any exposure to a sign language, not even to fingerspelling. They were native speakers of English, born and raised in the Mid-West (Illinois, Indiana, Ohio, and Michigan). Both signing and nonsigning subjects had completed high school.

2.2. Stimulus Materials

All stimuli were 'pseudo-signs' (i.e., nonsense signs), with a total of 168 items in all. There were 6 movement conditions x 5 handshape conditions x 2 POA conditions. Here "condition" means a particular combination of properties of that parameter, which will be described in detail below. The items were recorded on a digital Canon OpturaPI camcorder and digitized and compressed at 30 frames per second using Adobe Premier 6.0 software. The signer who produced the stimulus items on video was a Deaf, life-long user of ASL, went to a residential school for the Deaf, and is a member of the Deaf Community. The items were practiced before recording to insure a neutral non-manual expression across each entire form with no intervening eye blinks, and to insure that there was a 80–120 ms pause between the two movements of the disyllabic forms, consistent with pauses between words in a phrase (Brentari, Poizner and Kegl, 1995).[4] If a 2-handed sign was included, the non-dominant hand was present in the production throughout the entire item; that is, across both movements.[5]

In all, there were 28 cells, in which handshape, place of articulation, and movement cues were placed in conflict with each other in order to determine the most resilient word segmentation strategies in each group of participants. Each of the cells has six items – three items with one, and three with two, distinctive places of articulation.

Handshapes were separated into marked (abbreviated HSm) and unmarked (abbreviated HSu) groups as described earlier. B, A, S, C, O, 1 and 5 were the unmarked handshapes, and the rest were considered marked handshapes. There were five handshape conditions. Fifty-four items were phonotactically permissible signs from the point of view of handshape: (i) 24 had one handshape, and (ii) 30 had one set of selected fingers with an aperture change ([open]←→[closed]). Another 114 items were not permissible monomorphemic signs: (iii) 36 forms in which both handshapes were unmarked (i.e., the index or all fingers), (iv) 36 with one unmarked handshape (groups (iii) and (iv) are possible compounds, but not monomorphemic signs), and (v) 36 items in which both handshapes were marked, combinations which do not occur in single signs of any type.

There were two place of articulation conditions. In each of the 28 cells, three items had one place of articulation and three have two places of articulation (i.e., head, torso, H2, arm, and the three dimensional planes). In all, 84 items had one place of articulation and 84 had two places of articulation. The forms with two places of articulation may occur in compounds, but not in monomorphemic signs.

Regarding movement, there were six conditions. Sixty items were permissible structures: (i) 30 had one movement and (ii) 30 had two movements that were permissible in monomorphemic signs. The other 108 items are not permitted in monomorphemic signs or compounds: (iii) 24 combinations of non-permissible local movements (e.g., combinations of handshape changes and orientation changes), (iv) 30 illicit combinations of two path movements (e.g. straight+arc), (v) 24 combinations of a path movement and a handshape change, and (vi) 30 combinations of a path movement and an orientation change. The organization of the stimuli is shown in Table 2; bold typeface indicates that a form is given in Figure 6 as a example stimulus item. By putting cues in conflict in this way, we can directly evaluate the following word segmentation factors.

Table 2. Distribution of items in stimulus set. HSu=unmarked handshape; HSm= marked handshape; M=movement; POA=place of articulation; OR=orientation; Δ=change;. Grey cells indicate physically impossible forms. Bold typeface and underlining indicates places from which the sample stimulus items are extracted; see Figure 6.

Handshape types → Movement types ↓	1HS	HSm+HSm	HSu+HSu	HSu+HSm	1HS+aperture Δ
1 MOV	1 POA (3) 2 POA (3)	1 POA (3) 2 POA (3)	1 POA (3) 2 POA (3)	1 POA (3) 2 POA (3)	1 POA (3) 2 POS (3)
ORΔ+HSΔ		1 POA (3) 2 POA (3	1 POA (3) 2 POA (3	1 POA (3) 2 POA (3	1 POA (3) 2 POA (3
path+path	1 POA (3) 2 POA (3	1 POA (3) 2 POA (3	1 POA (3) 2 POA (3	1 POA (3) 2 POA (3	1 POA (3) 2 POA (3
path+HSΔ		1 POA (3) 2 POA (3	**<u>1 POA</u> (3)** 2 POA (3	1 POA (3) 2 POA (3	1 POA (3) 2 POA (3
path+ORΔ	**<u>1 POA</u> (3)** 2 POA (3)	1 POA (3) 2 POA (3	1 POA (3) 2 POA (3	1 POA (3) 2 POA (3	1 POA (3) 2 POA (3
2 M	1 POA (3) 2 POA (3	1 POA (3) 2 POA (3	1 POA (3) 2 POA (3	**<u>1 POA</u> (3)** 2 POA (3	1 POA (3) 2 POA (3)

Figure 6. Sample stimulus items. 1 HS x 1 POA x 2 Ms (1 path+1 ORΔ; left); 2 HS (5[open] 5[closed]) x 1 POA x 2 Ms (path+ aperture change; center); 2 HSs (8O) x 1 POA x 2 Ms (repeated path; right).

First, specifically with respect to group, if signers and nonsigners show a general difference in the items that are permissible sequences of handshape or movement, we will see an effect of ASL phonological constraints on signers' word segmentation judgments. For handshape, this is column 5, for movement, this is row 6. If, however, both groups' word segmentation patterns are the same for these conditions, we will conclude that there is less of an effect of the phonological constraints, and more of a more general modality effect. Second, specifically with respect to the units, if either group responds to specific sequences of handshapes or place we will see an effect of the segment. If we see either group respond to specific sequences of movements, this will be an effect of the syllable.

2.3. Procedures

The task was run on a Dell 4600 computer with an 18-inch computer monitor. A program was constructed specifically for this experiment that randomized the items for each subject, presented the instructions in English, inserted 'READY' printed in the middle of the screen as a prompt before each item, and registered the responses in individual files for each participant according to item number. Instruction was presented in printed English to be read by all participants, and they were also signed in ASL to the Deaf participants. Participants were instructed to watch the item and to decide if the form looked like one sign or two signs, based on their best intuitions. They were told that these stimuli were not real ASL signs. Items were shown only once. Participants were told that accuracy was important, but not speed, and that they were to consider their choice as long as needed to make as accurate a selection as possible. There were three practice items, after which further clarification could be provided. Participants initiated the presentation of items by hitting the space bar, and the mouse was used to register their responses by clicking on one of two 2-inch x 3-inch squares on the computer screen, with '1-sign' or '2-signs' written inside, both of which were presented simultaneously on the screen after each item. No reaction times were recorded. The average session duration for both groups was approximately equal, 44 minutes.

2.4. Results

An analysis of variance (ANOVA) was conducted on the between subjects factor of Group (2: signers vs. nonsigners) and repeated measures of Stimulus Value (2: judgments of one vs. two signs) and Parameter (3: handshape, place of articulation, and movement). The results are given in Figure 7, reported as percentage of 2-sign judgments. There were significant main effects of Stimulus, $F_{(1,24)} = 42.5$, $p < .001$, and Parameter, $F_{(3,73)}$=55.8, p < .001. There were also significant interactions of Stimulus x Parameter, $F_{(3,72)}$ =6.78, $p < .001$, and of Group x Parameter, $F_{(3,72)} = 2.9$, p < .04.

An analysis of variance (ANOVA) was also conducted on each of the parameters of handshape, place of articulation, and movement alone, using the between subjects factor of Group (2: signers vs. nonsigners) and repeated measures of Stimulus Value (2: judgments of one vs. two signs) and Parameter. For handshape, there was a main effect of Parameter, $F_{(4,96)} = 78.4$, $p < .001$, a significant interaction of Stimulus x Parameter, $F_{(4,96)} = 14.2$, $p < .001$, and a significant interaction of Group x Parameter, $F_{(4,96)} = 3.5$, $p < .05$. For movement, there was a main effect of Parameter, $F_{(5,120)} = 7.04$, $p < .001$, and a significant interaction of Stimulus x Parameter, $F_{(5,120)} = 135.9$, $p < .001$

There were no other significant statistical effects; however, the difference between the means of 1- and 2-sign judgments for each of the parameters differed between the two groups. The difference between the means of 1- and 2-sign judgments for handshape was 24% (signers) versus 13% (nonsigners); for place of articulation it was 14% (signers) versus 17% (nonsigners); and for movement it was 36% (signers) versus 33% (nonsigners.). If we take these differences in mean percentages between 1- and 2-sign judgments as an indication of how heavily each of the parameters was relied upon, we see that signers depend on movement, handshape, then place of articulation (in descending order) while nonsigners depended on movement, place of articulation, then handshape.

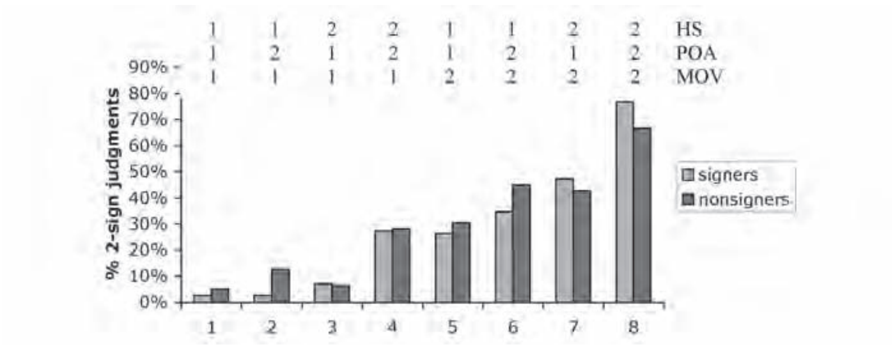


Figure 7. Mean percentages of 2-sign judgments for one- and two- value stimuli in eachparameter.

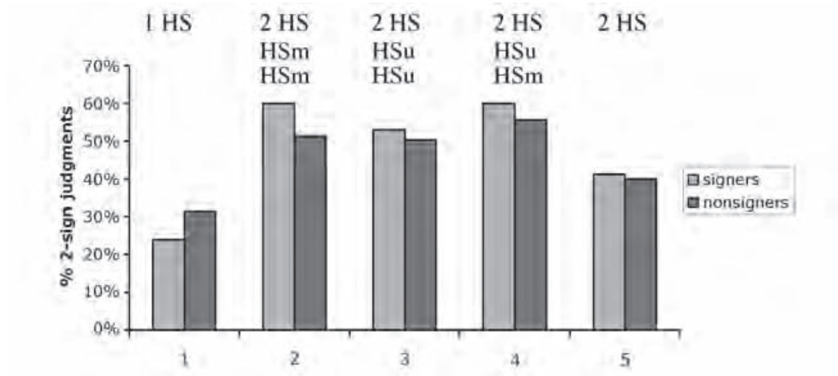


Figure 8. Mean percentages of 2-sign judgments across all handshape conditions.

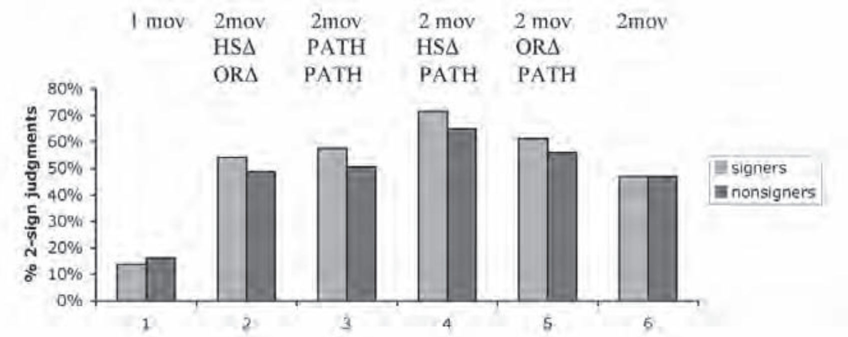


Figure 9. Mean percentages of 2-sign judgments across all movement conditions.

3. Discussion

In this section, we will address our results in terms of group comparisons, effects that relate to the use of phonological units, and finally, the degree to which the parameters of handshape, place of articulation and movement were relied upon in making word segmentation judgments. It was hypothesized that both groups would rely on the same word-level strategy to make word segmentation judgments in a sign language; that is, we expected a strong modality effect. Second, it was hypothesized that signers would be able to utilize a greater number of simultaneous cues in making word segmentation judgments. Third, we expected language-particular phonological constraints on sequences of handshapes, places of articulation, and movements to be used by signers but not nonsigners in making word segmentation judgments.

3.1. Group comparisons

The first hypothesis concerning group was that signers and nonsigners would use the same overall strategies. This hypothesis was supported. There was no significant Group x Stimulus Value interaction for any of the parameters; however, there were some results that did not achieve statistical significance, but will be discussed in the section below on sign parameters. The second hypothesis concerning group was that signers would be more sensitive to a greater amount of simultaneous information. This hypothesis was supported by the significant interaction between Group x Handshape Parameter, showing that signers make significantly greater use of handshape information in word segmentation judgments than nonsigners do; this is the only parameter which shows this interaction. Despite native exposure to languages in different modalities, both groups relied on the same strategies overall in making word segmentation judgments in a sign language.

Surprisingly, the signing participants seem not to use ASL phonological constraints in making word segmentation judgments. This is inferred from the lack of difference in signing and nonsigning participants in column 5 in Figure 8 (for handshape) and column 6 in figure 9 (for movement). These are the items that had permissible sequences – either aperture changes for handshape, or types of repetition or circle+straight combinations in movement.

3.2. Sign Language Unit of Word Segmentation is the Word

As discussed earlier, the stimuli were constructed to address the question of potential differences in the unit of analysis used by signers and nonsigners in the experimental task. The permissible handshape sequences and permissible place of articulation sequences were included to determine segmental-level effects. The permissible movement sequences were included to determine syllable-level effects. Importantly, the set of permissible disyllabic forms in column 6 of Figure 9 were employed to disambiguate word-domain effects from syllable effects, in order to determine whether the strong reliance on syllables by users of spoken languages carries over into their word segmentation judgments of signs, and to see whether signers employ the syllable in their own word segmentation judgments, given that the phonological constraints concerning the timing of handshape changes are based on the syllabic unit (Perlmutter 1992; Brentari 1998).

With regard to the segment, there was no statistical evidence that signers or nonsigners treated the permissible 2-handshape sequences differently than the non-permissible 2-handshape sequences. There was, instead, a strategy of '1 value = 1 word'.

Regarding the syllable, both groups relied most heavily on movement cues to make word segmentation judgments. As described earlier, the syllable is possibly the most important unit used in making word segmentation judgments in spoken languages when other cues are in conflict with it, even though it has been shown that the segment in English (Cutler et al. 1986), the syllable in French (Mehler et al. 1981), the word as the domain of vowel harmony in Finnish (Suomi, McQueen and Cutler 1997), and the metrical foot in English, Dutch, Finnish (Vroomen, Tuomainen, and de Gelder 1998) are all used in some capacity to segment words.[6] Finnish is the most relevant to the discussion here because it is the only spoken language for which both the word domain cue of vowel harmony and the syllable domain cue of rhythm were put in conflict in an experiment very similar to the ASL study reported here (Vroomen, Tuomainen, and de Gelder 1998).[7] When the rhythmic effects of stress conflicted with those of vowel harmony, native Finnish subjects relied more heavily on stress cues, associated with the syllable.

A key question for the sign segmentation task presented here is whether the reliance on movement cues is a word effect or a syllable effect. I would argue that it is a word effect, for the following reasons. First, when deciding which unit is relevant, the one that captures the most generalizations should be selected. In this case, the strategy used by participants across all three phonological parameters is best captured by the word domain. The generalized strategy is '1 value = 1 word'; it did not matter if this value involved a segmental unit, such as handshape or place of articulation, or a syllabic unit, such as movement. Second, the pattern of performance by both groups was not based on a regularly alternating pattern. Crucially, neither set of participants used an alternating pattern to make word segmentation judgments, such as the 'strong' unit of trochaic stress, or any other pattern that included a repetitive sequence. In order for the syllable to be the relevant unit, there would not only have to be a sequential pattern involved, but a sequence of a particular sort. Instead, like word segmentation based on vowel harmony, *every change* in value triggered the perception that it signaled the beginning of a new word.

3.3. Sign Language Parameters

In this section, the sign language parameters that are used most heavily to make word segmentation judgments in our groups are examined and discussed in terms of the existing ASL literature. First, concerning movement and place of articulation, Corina and Hildebrandt (2002) found that combinations of movement and place of articulation also generated the strongest "similarity" responses in both signing and non-signing subjects; that is, forms with the same movement and place of articulation are judged to be most similar in both signing and non-signing subjects (see also Corina and Knapp, this volume). This would be consistent with the results presented here, since the place of articulation and movement parameters showed a great deal of similarity between groups.

Second, the heavier reliance on handshape by the signing group in the study may be explained in two ways. First, handshape can be considered the most arbitrary and the most categorical of the sign language parameters, and therefore, it may be accessible only to individuals with experience with a sign language. From the point of view of phonological description and analysis, handshape is much more "well-behaved" as a linguistic entity, than is movement or place of articulation. And, in this context, one can interpret "well-behaved" as being analyzable in a discrete fashion and able to conform most closely to the formalism of binary oppositions and hierarchical representations. There is also a great deal more consensus about representations of handshape – in particular, how they should be represented in a feature geometry and what parts of the handshape are distinctive within a sign language grammar – than there is about place of articulation or movement (see, for example, van der Hulst 1995; Sandler 1996; Brentari 1998). It is also true that it is comparatively easier to list a reasonably small number of handshapes based on their discrete component parts than it is to do this with movement or place of articulation. In the psycholinguistic literature, handshape is the parameter that shows the strongest divergence between the 'companion gestures' (McNeill 2000) that speakers use as they talk and the manual gestures of a sign language. In a study that compared three groups of subjects – hearing people using gestures that accompany speech, hearing people who were asked to gesture without speech, and home signers – Singleton, Morford, and Goldin-Meadow (1993) and Goldin-Meadow, McNeill, and Singleton (1996) found that the biggest difference between gestures that accompany speech and productions of home signers was that the home signers used a greater variety of handshapes in their output. This suggests that as gesture becomes more linguistic, handshape becomes more differentiated. Independent evidence also comes from a recent experiment on categorical perception (Emmorey, McCullough and Brentari 2002), in which handshape was found to exhibit categorical perception effects for signers (but not for non-signers), while place of articulation failed to exhibit such effects. The strong linguistic nature of handshape in sign languages is supported by these experimental findings. Second, handshape may be the most difficult parameter for non-signers to perceive because the joints involved are relatively distal (those closer to the extremities of the body), and as such are smaller in size and least likely to be used by the untrained eye in a task of word segmentation.[8]

To conclude this section, it is worth commenting on difficulties that have been reported when children attempt to learn artificial Manually Coded English systems of signs (MCE; e.g., Gustason, Pfetzing, and Zawalkow 1980). The work on MCE by Supalla (1990, 1991), Stack (1999), and Supalla and McKee (2002) shows that artificial systems fail to be learned by Deaf children because they follow no discernable phonological system, either signed or spoken. The experimental work just presented suggests that another difficulty of MCE systems is that they lack the strong cues to segmentation that are due to a modality effect. Those cues are important even to persons who lack exposure to a sign language.

In closing, we must note that despite the fact the instrumental measurements on sign languages and spoken languages must be carried out with different equipment and consideration for different types of sensory effects, the question of how much the medium infiltrates the code in language cannot be answered until both signed and spoken languages are studied with comparable tasks and experimental designs.

Notes

* I am grateful to Robin Shay for her help in constructing the stimuli, to the participants of LabPhon 8 for the discussion of these issues at the conference, and to Catherine Best and an anonymous reviewer for their helpful comments on an earlier version of this paper.

1. Lists of compounds from Klima and Bellugi (1979) and Liddell and Johnson (1986) were consulted.
2. This is not an absolute, but rather a matter of statistical tendency, since there are spoken languages that use tone or vowel harmony values to signal a minimal pair, and in ASL there are a few minimal pairs that use the segment.
3. I am considering only forms from different morphological paradigms, so FLY and FLY-THERE would not be a minimal pair.
4. Supalla (1990) demonstrated that neither pause length between two parts of a signed string nor the number of hands producing a sign influences subjects' judgments.
5. This would facilitate a judgment of '1 sign', since well-formed words in ASL have one specification for the non-dominant hand (Brentari 1998).
6. Often these adult experiments were targeting a combination of word segmentation and word identification or lexical access.
7. For an excellent summary of vowel harmony as a word-based phenomenon, see Krämer (2001).
8. I am grateful to Onno Crasborn for this suggestion.

References

Archangeli, Diana
1984 *Underspecification in Yawelmani phonology and morphology.* Doctoral dissertation, MIT. Cambridge, MA.

Aronoff, Mark, Irit Meir, Carol Padden and Wendy Sandler
2003 Classifier constructions and morphology in two sign languages. In: Karen Emmorey (ed.) *Perspectives on classifier constructions in sign languages*, 53–84. Mahwayh, NJ: Lawrence Erlbaum Associates.

Battison, Robbin
1978 *Lexical borrowing in American Sign Language.* Linstok Press, Silver Spring, MD. [reprinted 2003]

Bregman, Albert S.
1990 *Auditory scene analysis*. Cambridge, MA.: MIT Press.

Brentari, Diane
1990a *Theoretical Foundations of American Sign Language Phonology*, Doctoral dissertation, University of Chicago , published 1993, University of Chicago Occasional Papers in Linguistics, Chicago, Illinois.
1990b Licensing in ASL Handshape Change. In Ceil Lucas (ed.) *Sign Language Research: Theoretical issues,* 57–68. Washington, DC: Gallaudet University Press.
1998 *A prosodic model of sign language phonology*. Cambridge, MA: MIT Press.

2002 Modality differences in sign language phonology and morphophonemics. In: Richard Meier, Kearsy Cormier and David Quinto (eds.) *Modality in Language and Linguistic Theory*, 35–64. Cambridge, UK: Cambridge University Press.

Brentari, Diane, Howard Poizner and Judy Kegl
1995 Aphasic and Parkinsonian Signing: Differences in Phonological Disruptions. *Brain and Language* 48: 69–105.

Channon, Rachel
2002a *Signs are single segments: Phonological representations and temporal sequencing in ASL and other sign languages.* Doctoral dissertation, University of Maryland, College Park, MD.
2002b Beads on a string? Representations of repetition in spoken and signed languages. Richard Meier, Kearsy Cormier and David Quinto (eds.) *Modality in Language and Linguistic Theory*, 65–87. Cambridge, UK: Cambridge University Press.

Chase, Chris and Annette R. Jenner
1993 Magnocellular visual deficits affect temporal processing of dyslexics. *Annals of the New York Academy of Sciences* 682 (June 14, 1993): 326.

Chinchor, Nancy
1978 The Syllable in American Sign Language: Sequential and simultaneous phonology. Unpublished manuscript.

Corina, David
1990 Reassessing the Role of Sonority in Syllable Structure: Evidence from a Visual-Gestural Language. *Papers from the 26th Annual meeting of the Chicago Lingustic Society: Vol. 2: Parasession on the Syllable on the syllable in phonetics and phonology,* 33–44. Chicago Linguistic Society, University of Chicago, Chicago, Illinois.

Corina, David and Ursula Hildebrandt
2002 Psycholinguistic investigations of phonological structure in American Sign Language. In: Richard Meier, Kearsy Cormier and David Quinto (eds.) *Modality in Language and Linguistic Theory*, 88–111. Cambridge, UK: Cambridge University Press.

Cutler, Anne, Jacques Mehler, Dennis Norris and Juan Segui
1986 A language-specific comprehension strategy. *Nature* (London) 304: 159–160.

Emmorey, Karen, Steven McCullough and Diane Brentari
2003 Categorical Perception in American Sign Language. *Language and Cognitive Processes* 18: 21–45.

Friedman, Lynn
1977 Formational properties of American Sign Language. In: Lynn Friedman (ed.) *On the other hand: New perspectives in American Sign Language*, 13–56. New York: Academic Press.
Goldin-Meadow, Susan, David McNeill and Jenny Singleton
1996 Silence is liberating: removing the handcuffs on grammatical expression in the manual modality. *Psychological Review* 10*:* 34–55.
Gustason, Gerilee, Donna Pfetzing and Esther Zawalkow
1980 *Signing Exact English*. Los Alamitos, CA: Modern Signs Press.
Halle, Morris and J.-R. Vergnaud
1987 *An essay on stress.* Cambridge, MA: MIT Press,
Hayes, Bruce
1995 *Metrical stress theory:principles and case studies.* Chicago: University of Chicago Press.
Houston, Derek M., Peter Jusczyk, Cecile Kuijpers, Riet Coolen and Anne Cutler
2000 Cross-language word segmentation by 9-month olds. *Psychonomic Bulletin and Review:* 7.
van der Hulst, Harry
1995 The Composition of Handshapes. *University of Trondheim, Working Papers in Linguistics,* 1–17 Dragvoll, Norway: University of Trondheim.
2000 Modularity and modality in phonology In. Nancy Burton-Roberts, Philip Carr and Gerard Docherty (eds.) *Phonological knowledge: Its nature and status,* 46–63. Oxford: Oxford University Press.
Jusczyk, Peter W., Anne Cutler and Nancy J. Redanz
1993 Preference for the predominant stress patterns of English words. *Child Development* 64: 675–687.
Jusczyk, Peter W., Elizabeth Hohne and Angela Bauman
1999 Infants' Senstitivity to allophonic cues for word segmentation*. Perception and Psychophysics* 61(8): 1465–1473.
Kegl, Judy, Anne Senghas and Marie Coppola
1999 Creation through contact: Sign Language Emregence and Sign Language change in Nicaragua. In Michele De Graff (ed.) *Comparative Grammatical change: The Intersection of Language Acquistion, Creole Geneiss, and Diachronic Syntrax*. Cambridge, MA: MIT Press.
Klima, Edward and Ursula Bellugi
1979 *Signs of Language*. Cambridge, MA: Harvard University Press.
Krämer, Michel
2001 *Vowel Harmony and Correspondence Theory*. Doctoral dissertation, Heinrich-Heine-Universität, Düsseldorf.

Liddell, Scott
1984 THINK and BELIEVE: Sequentiality in American Sign Language. *Language* 60: 372–392.
2003 *Grammar, Gesture, and Meaning in American Sign Language*. Cambridge, UK: Cambridge University Press.
Liddell, Scott and Robert E. Johnson
1986 American Sign Language Compound Formation Processes, Lexicalization, and Phonological Remnants, *Natural Language and Linguistic Theory* 4: 445–513.
1989 American Sign Language: The Phonological Base, *Sign Language Studies* 64, 197–277.
Lillo-Martin, Diane
1991 *Universal Grammar and American Sign Language.* Dordrecht: Kluwer.
MacNeilage, Peter
1998 The frame/content theory of evolution of speech production. *Behavioural and Brain Sciences* 21: 499–546.
MacNeilage, Peter and Barbara Davis
1993 Motor explanations of babling an dearly speech patterns. Bénédicte de de Boysson-Bardies, Scania de Schonen, Peter Jusczyk, Peter MacNeilage,John Morton (eds.), *Developmental Neurocognition: Speech and face processing in the first year of life*,, 341–352. Dordrecht: Kluwer.
Mandel, Mark
1981 *Phonotactics and morphophonology in American Sign Language.* Doctoral disseration, University of California, Berkeley, California.
Mathur, Gaurav
2001 *Verb Agreement as Alignment in Signed Languages.* Doctoral dissertation, Massachusetts Institute of Technology.
Mattys, Sven and Peter W. Jusczyk
2001 Phonotactic cues for segmentation of fluent speech by infants. *Cognition*, 78: 91–121.
McNeill, David
2000 *Language and Gesture*. New York: Cambridge University Press.
Mehler, Jacques, J.-Y. Dommergues, Uli Frauenfelder and Juan Segui
1981 The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior* 20: 298–305.
Meier, Richard
2002 Why difference, why the same? Explaining effects and non-effects of modality upon linguistic structure in sign and speech. In: Richard Meier, Kearsy Cormier and David Quinto (eds.) *Modality in Language*

and Linguistic Theory, 1–25. Cambridge, UK: Cambridge University Press.

Neidle, Carol, Judy Kegl, Dawn Maclaughlin, Benjamin Bahan and Robert G. Lee
2000 *The Syntax of American Sign Language*. Cambridge, MA: MIT Press.

Padden, Carol
1983 *Interaction of Morphology and Syntax in American Sign Language.* Doctoral dissertation, University of California, San Diego. Published, 1988, New York: Garland.

Perlmutter, David
1992 Sonority and syllable structure in American Sign Language. *Linguistic Inquiry* 23: 407–442.

Petitto, Laura and Paula Marentette
1991 Babbling in the manual mode: Evidence from the ontogeny of language. *Science* 251: 1493–1496.

Sandler, Wendy
l986 The spreading hand autosegment of American Sign Language. *Sign Language Studies* 50: 1–28.
1987 Assimilation and feature hierarchy in American Sign Language. *CLS.23 (2). Parasession on Autosegmental and Metrical Phonology,* 266–278. University of Chicago, Chicago, IL.
1989 *Phonological representation of the sign.* Dordrecht: Foris.
1993 A sonority cycle in American Sign Language. *Phonology* 10(2): 243–279.
1996 Representing handshapes. International Review of Sign Language *Linguistics* 1: 115–158.

Senghas, Anne
1995 *Children's Contribution to the Birth of Nicaraguan Sign Language*. Doctoral dissertation, MIT, Cambridge, MA.

Singleton, Jenny, Jill Morford and Susan Goldin-Meadow
1993 Once in not enough: Standards of well-formedness in manual communication over three different time spans. *Language* 69: 683–715.

Stack, Kelly
1999 *Innovation by a child acquiring Signing Exact English 2*. Doctoral dissertation, UCLA.

Stokoe, William
1960 *Sign language structure: An outline of the visual communication systems of the American deaf. Studies in Linguistics, Occasional Papers 8*. (available from Silver Spring, MD: Linstok Press)

Stokoe, William, Dorothy Casterline and Carl Croneberg
1965 *A Dictionary of American Sign Language on Linguistic Principles*. Linstok Press, Silver Spring, Maryland.

Suomi, Kari, James. M. McQueen and Anne Cutler
1997 Vowel harmony and speech segmentation in Finnish. *Journal of Memory and Language*, 36: 422–444.
Supalla, Samuel
1990 *Segmentation of Manually coded English: Problems in the mapping of English in a visual-gestural mode*. Doctoral dissertation, University of Illinois, Champaign-Urbana.
1991 Manually Coded English: The modality question in signed language development. In: Patrica Siple and Susan Fischer (eds.) *Theoretical Issues in Sign language Research, vol. 2: Psychology*, 85–109. Chicago: University of Chicago Press.
Supalla, Samuel and Cecile McKee
2002 The role of manually coded English in language development of Deaf Children. In: Richard Meier, Kearsy Cormier and David Quinto (eds.) *Modality in Language and Linguistic Theory*, 143–165. Cambridge, UK: Cambridge University Press. .
Supalla, Ted
1982 *Structure and acquisition of verbs of motion and location in American Sign Language.* Doctoral dissertation, University of California, San Diego, California.
Uyechi, Linda
1996 *The Geometry of visual phonology,* CSLI, Stanford, California.
Vroomen, Jean, Jurki Tuomainen and Beatrice de Gelder
1998 The roles of word stress and vowel harmony in speech segmentation. *Journal of Memory and Language* 38: 133–149.
Zeshan, Ulrike
2000 *Sign language in Indopakistan. A description of a Signed Language*. Amsterdam: Benjamins.

Phonology, phonetics and the nondominant hand

Wendy Sandler

Many studies have shown that the phonology of sign language has much in common with that of spoken language, including the existence of meaningless phonological features and feature classes, phonological constraints, rules of assimilation, and more. At the same time, the phonetics of the two language modalities is shaped in part by articulators with very different anatomical and physiological characteristics. For these reasons, sign language offers an interesting vantage point from which to view the relation between phonetics and phonology. In this chapter, the behavior of the nondominant hand in the phonology and the prosody of sign language is analyzed. This articulator, anatomically identical to another articulator in the system (the dominant hand), has no parallel in spoken language, yet it plays a significant role in aspects of phonological organization that are similar in the two modalities.

1. Introduction

Natural sign languages have been found to be similar to spoken languages in significant ways at all levels of structure, despite the radically different physical channel of transmission: the hands, face, and body and vision, instead of the vocal tract and audition (Sandler and Lillo-Martin 2006). Because the physical system underlying the phonetics and phonology of sign language is beyond any doubt qualitatively different from that of spoken language, it is at once surprising and intriguing that similarities have been found at the phonological level of structure. I wish to make two related claims: (1) There is a phonological level of linguistic organization common to signed and spoken language (2) phonological organization does not derive from phonetics alone. This is clearly a complex issue with important implications, and it has been approached from various perspectives in the sign language literature (e.g., Brentari 1998; Crasborn 2001; Sandler 1989; Sandler and Lillo-Martin 2006; Uyechi 1996). The present investigation proposes to examine it in a restricted way, but one that is potentially especially revealing – namely, from the point of view of the nondominant hand, an articulatory element with no parallel in spoken language.

The nondominant hand (henceforth, h2) is anatomically identical to the primary articulator of sign language, the dominant hand. A priori, one might anticipate that each hand could participate equally to provide meaning in sign language utterances, the way they do in some kinds of co-speech gesture (Enfield 2004; McNeill 1992). For example, each hand might be configured to represent a different entity iconically, and together, they could convey the spatial relation between the entities. As in gesture, the shapes and relations of the hands might also be metaphorically extended, representing the interaction of events or concepts. If a system like this were to exist in sign language, it would mean that each hand assumes some kind of meaningful word-like status.

Stokoe's (1960) demonstration that the handshapes of American Sign Language are like phonemes, formationally significant but meaningless, provides a diametrically different characterization of h2. Like the dominant hand, h2 is shown to function not as a word or morpheme, but as an articulator, assuming different shapes and articulating different locations and movements – themselves also meaningless – in a phonological system. Stokoe's work served to bring sign language into the arena of general linguistic investigation. In so doing, it opened the door to new and important questions.

For example, once the hands are thought of as articulators in a phonological system, the nondominant hand appears to be anomalous when compared to the spoken language articulatory system, because it is an anatomical copy of the dominant hand. Spoken language has no such dual articulator. The primary articulator of spoken language, the tongue, is unitary, as are all other elements in the articulatory apparatus of the medium. One might expect sign languages to exploit the articulatory potential of the nondominant hand, promoting it to equal or near equal status to that of the dominant hand in the phonology. Contrarily, the system might bear only one dominant articulator, the dominant hand or h1, relegating h2 to a subordinate role in the phonology.

Yet another possibility is that the biology of language, having a perfectly good articulator at its disposal, will not let h2 off the phonological hook so easily, and will instead exploit it for a function that is phonological but different from that of h1 – for augmenting the rhythmic properties of prosody, for example.

In this chapter, the behavior of the nondominant hand in the phonology, morphology, and prosody of sign language is singled out. This articulator, uniquely available to the sign modality, reveals linguistic properties at each of these levels of organization that are independent of modality. This state of affairs is argued here to be incompatible with a theory in which phonetics and phonology are one and the same, and incompatible as well with a theory in which all of phonology is directly derived from phonetics.

Distinguishing between phonetics and phonology has proved to be a challenging enterprise (see, e.g., Kingston and Beckman 1990) and the present paper does not aim to resolve the issue. Instead the goal is to demonstrate that there is phonology in sign language that is recognizably similar to spoken language phonology, despite the fact that the articulators are not comparable in any meaningful sense. As the articulators and their properties mold the phonetic system, it follows that the phonetic systems cannot be the same in the two modalities. And if it can be shown that phonological organization is similar in specific areas of spoken and signed languages where the phonetics is different, we must conclude that the phonology does not derive from the phonetics alone.

For the sake of the discussion, I make the following general assumptions about phonetics and phonology. First, phonetic processes are gradient while phonological processes are discrete and categorical (e.g., Cohn 1993). Second, phonetic processes are mandated mainly by physical production mechanisms while phonological processes are also linked to higher levels of grammatical structure, such as the lexicon and the syntax. Most of the attention in the present study is given to structure and processes related to the nondominant hand that are phonological in the sense just described.

The exposition begins with the lexicon, where h2 plays a clearly subordinate and restricted phonological role (§1), then proceeds to classifier constructions, in which h2 has the status of a morpheme and does not conform to the same restrictions (§2). In §3, I review research demonstrating that the nondominant hand plays a systematic role in the demarcation of prosodic constituents. The point of the overview is to show that a single phonetic articulator is recruited by a diverse range of subsystems in the grammar of sign language in ways that that clearly have no parallel in the phonetics of spoken language. The claim that there are significant phonological similarities in the two modalities despite the different phonetic foundation will rely mainly on two of these levels, lexical and prosodic. It's at these two levels that the linguistic behavior of h2 has been most clearly worked out, and can best exemplify the specific claims that I wish to make here. The data come from two sign languages: American Sign Language (ASL) and Israeli Sign Language (ISL).

2. The phonology of the nondominant hand in the lexicon

There is a broad consensus that there is only one primary articulator in the lexical phonology of sign language, the dominant hand (Brentari 1990, 1998; Brentari and Goldsmith 1993; Perlmutter 1991; Sandler 1989, 1993a; van der Hulst 1996; van Gign, Kita and van der Hulst in press).[1] This means that the nondominant hand plays only a minor role in lexical representations. It represents a meaningless phonological element, and its shape and behavior are so strictly constrained as to make it largely redundant.

The formational elements in sign language words are subject to phonological constraints, among them constraints on the nondominant hand. The constraints I am about to describe have been attributed to various domains, such as the 'sign' or the morpheme. I am attributing them here to the lexeme. What is relevant for our purposes is that these constraints hold within the lexicon, and that together they characterize the typical sign language word. One constraint on structure is monosyllabicity: most sign language words have only one movement; that is, they are monosyllabic (Coulter 1982; Sandler 1993c).[2] The Selected Finger Constraint (Mandel 1981) requires a maximum of one specification for selected fingers (on the dominant hand) in a lexeme. This means that a handshape is defined in terms of the fingers that are selected in its articulation. The handshape  selects the index finger and thumb; the shape  selects the index and middle finger, etc. Among the constraints on lexemes are two that have special relevance here, as they restrict the specification of the nondominant hand: the Dominance Condition and the Symmetry Condition (Battison 1978). These constraints are paraphrased below.

> *The Dominance Condition*
> If the hands of a two-handed lexeme do not share the same specification for handshape, then one hand must be passive while the active hand articulates the movement, and the specification of the passive handshape is restricted to be one of a small set:[3]   

> *The Symmetry Condition*
> If both hands move independently, then both hands must be specified for the same handshape and the same movement (whether performed simultaneously or in alternation), and the specifications for location and orientation must be either identical or mirror-image.

Lexemes are further constrained by specification for a single major body area (Battison 1978), called place of articulation in the model used here. Places of articulation include the head, trunk, and nondominant hand (h2). Normally, h1 moves from one setting to another (e.g., high to low, contralateral to ipsilateral, or proximal to distal) with respect to the place. The Dominance Condition is relevant for lexemes in which only one hand, h1, articulates, and h2 is a place of articulation (Sandler 1989, 1993a). An example of a sign in which h2 is a place of articulation is (ISL) AT-THAT-MOMENT, pictured in Figure 1b. Figure 1a is a schematic representation of such signs. In this schematic example, HC stands for the category of Hand Configuration; Ls are location positions and M is a movement position on a skeletal tier. In the sign, the dominant hand is configured in a particular shape and orientation, represented in a complex feature hierarchy (Sandler 1987, 1989). In the schema here, an icon is used for simplicity. The hand moves from one location to another, on or near a single major body area, such as the head, the trunk, or in the present example, the nondominant hand, labeled [h2] in the schematic representation. The single major body area is labeled 'Place', for place of articulation. Location features specify further refinenements of the place category, such as [proximal], [high], [contact], etc., called settings. In the sign pictured, the dominant hand moves from the first location, a point above the nondominant palm ([proximal]), to the second location, contact with the palm.

The Symmetry Condition refers to lexemes in which h2 is essentially a copy of the dominant hand, h1. In such signs, h2 is simply represented as a member of the Hand Configuration class (Sandler 1989, 1993a). An example is (ISL) CAT, shown in Figure 2b. In this sign, the head is the place of articulation and the settings are [ipsilateral] to the signing hand, [low] (i.e., side of mouth area) for both locations, [contact] for the first location and [proximal] for the second location. The change in setting describes the path traversed by the hands. Signs in which h2 functions like a copy of h1 may be represented as in Figure 2a. The h2 node is associated to the same feature complex as h1, abbreviated here by the handshape icon , and articulates the same locations and movement.

In both types of two-handed words, the nondominant hand is underspecified. In the dominance type (Figure 1), the hand must either have one of only a few unmarked handshapes, or it is redundantly marked for the same shape as h1. In the example, the h2 handshape is the unmarked shape, . The notion of markedness assumed here is that of Jakobson (1968), and underspecification is seen as a device for expressing relative markedness: the less specified, the less marked (see Sandler 1995, 1996 for a treatment of handshapes in this framework). In the symmetrical type (Figure 2), h2's shape, and the locations and motion it articulates, are all completely unspecified, assuming those of h1 by default.[4]

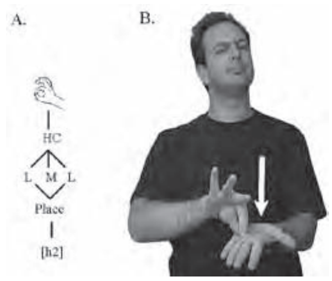


Figure 1. A) Partial schematic representation of sign with h2 as a place of articulation: AT-THAT-MOMENT (ISL).
B) Illustration of the sign.

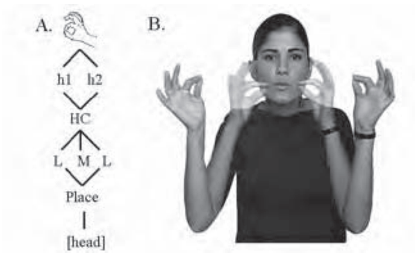


Figure 2. A) Partial schematic representation of sign with two symmetrical articulators: CAT (ISL).
B) Illustration of the sign.

Furthermore, in the phonology and morphology of ASL, h2 patterns with h1 in symmetrical signs, and with the place of articulation class in the other type of two-handed sign (Sandler 1989, 1993a). For example, the place constraint applies to h2 in dominance (Figure 1) type signs: if h2 is the place of articulation, then no other place of articulation may be specified in the lexeme (Perlmutter 1991), just as there may be no other place besides the head in the sign in Figure 2. Similarly, in symmetrical signs (Figure 2 type), h1 and h2 behave identically under assimilation of hand configuration in compounds.

In order to understand the behavior of the hands in sign language phonology, let's look at assimilation in ASL compounds. The form of lexicalized compounds is often reduced, in part by total regressive assimilation of the Hand Configuration (Liddell and Johnson 1986; Sandler 1987, 1989). If the second base sign happens to be a symmetrical two-handed sign, i.e., if h2 is part of the HC class, then both hands assimilate. Figure 3 shows the compound MIND+DROP, which means FAINT. Figure 4 shows (schematically) how the rule works (see Sandler 1987; Sandler and Lillo-Martin 2006 for full representations).

In this lexical phonological rule, the entire hand configuration is involved, and it spreads discretely to the beginning of the compound and no further. It is to be distinguished from post-lexical coarticulatory (phonetic) processes, which are gradient and/or non-categorical. For example, Corina (1993) describes coarticulation between words that adds a single finger from the handshape of one sign to the handshape of an adjacent sign. Coarticulation of this sort between signs is not categorical, since only part of the shape assimilates. It is non-structure preserving in the sense that it may create shapes that don't exist in the handshape inventory. I note that no instrumental measurement of the sign language phenomena under discussion has been undertaken. Thus, claims about gradience and discreteness rely on human judgment. This shortcoming is not quite as dire as it may seem, however, for the following reasons: (1) unlike the articulators of speech, sign language articulators are large and slow relative to speech articulators; (2) sign language articulators are directly observable by the eye; and (3) all relevant phenomena have been scrutinized on videotape many times in slow motion ranging from 30% of normal speed down to frame-by-frame viewing.[5] The fact that this method can distinguish between non-categorical handshape coarticulation between words and categorical handshape assimilation in compounds encourages us to accept results achieved this way, at least until a better method can be satisfactorily implemented.

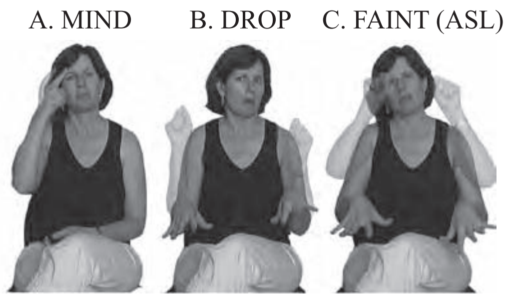


Figure 3. ASL lexicalized compound with Hand Configuration assimilation.

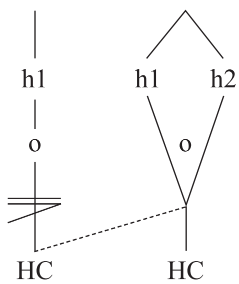


Figure 4. Total HC assimilation in a symmetrical two-handed compound.

Because the total HC assimilation rule in compound words is lexical, structure preserving, and categorical. rather than post-lexical, non-structure preserving, and non-categorical, it is presented here as an example of a phonological rule.[6] Phonology in both modalities, then, recognizes the category 'word' and is affected by discrete assimilation rules. Furthermore, regardless of modality, phonology systematically affects whole classes of features (e.g., Clements 1985). In the kind of assimilation presented here, the whole hand configuration class, including the selected fingers, their position, and the orientation of the hands, assimilates.[7] Yet despite these phonological similarities, the phonetic articulatory systems of the two modalities have nothing in common that meets the eye or ear. The specific shape taken by coarticulation in each system is determined by the physical nature and dynamic properties of the articulators. But the fact that HC assimilation is categorical and linked to a higher level of structure – the word – makes it phonological. These phonological properties don't derive from the phonetics, and they are present in both modalities.

Returning to the nondominant hand in lexical signs, we may describe its role as largely redundant. It is underspecified, and it behaves like h1 in symmetrical signs or like a place of articulation in dominance signs. Of the sign languages that have been studied, vanishingly few minimal pairs have been attested in which the presence or absence of h2 is contrastive. In fact, h2 is so redundant that it can often be omitted, by a process called Weak Drop (for ASL, Brentari 1998; Padden and Perlmutter 1987; for Sign Language of the Netherlands, van der Kooij 2002; for ISL, Levy 2001). Specific phonological analyses vary, but there is a consensus on the following claim: although phonetically there are two manual articulators, there is phonologically only one major articulator in lexical signs: h1, the dominant hand (see the references at the beginning of this section).

It is likely that a combination of motoric, perceptual, and cognitive factors underlies the dramatic subordination of h2 to h1 within lexemes in the sign language lexicon. Discovering what these factors are and how they interact is worthy of future research. But whatever they are, the end result is a lexicon in which the form of two-handed words is severely restricted and the specification of the nondominant hand is largely redundant.

The question of why h2 appears at all in lexical words is worthy of attention. Part of the answer is surely phonetic: the nondominant hand is there, and it is subject to motor patterns that are dictated by bimanual coordination. Yet due to phonological constraints, its role within the lexicon is minimal. Keeping in mind that these constraints hold only on words, an explanation suggests itself: the redundant properties of h2 in lexical words are significant in sign language processing. It is reasonable to speculate that the redundancy itself signals to the child acquiring sign language or to the addressee that a two-handed articulation so formed has the status of a lexeme or lexical word. Words may be distinguished by such constraints from other linguistic elements in sign language, such as classifier constructions, to which we turn now.

3. The nondominant hand as a meaningful element: Classifiers

I have said that h2 plays only a minor role in the representation of lexemes. This does not mean, however, that h2 is insignificant throughout the lexicon. In all established sign languages studied to date, an elaborate system of classifier constructions exists, in which h2 has a more independent status. These structures, often invoked to express events of motion and location, spatial relations among concrete referents, or the handling of objects (Supalla 1982, 1986), involve a set of handshapes that function as classifiers. Classifier handshapes typically classify referents in terms of semantic category (e.g., HUMAN, SMALL ANIMAL, VEHICLE, etc.), size and shape (SMALL-ROUND-OBJECT, FLAT-OBJECT, etc.), or the dimensions of the handler of an object (and by extension, of the object being handled). These combine with different paths and manners of movement, and with locations. In this system, each hand, instead of being a phonological element, may represent a morpheme by its configuration.

Classifier constructions are most clearly reminiscent of the contribution of gesture in the formation of sign languages. For example, like iconic gestures (McNeill 1992), the hands can take on the shape of objects being described and can mimic their relative locations and the kinds of motion they undergo. Also as in gestures (Enfield 2004), the nondominant hand in classifier constructions can serve as the ground for the dominant hand, the figure (Supalla 1982). Yet these structures are not pantomimic analogs. Rather, they are comprised of a finite list of handshapes and movements, they are rule-governed (Supalla 1982, 1986), and they pose a challenge for the child acquiring sign language (Slobin et al. 2003; Supalla 1982, 1986). Nevertheless, linguists often treat the system separately from the rest of the language, because of the formal structure of classifier constructions, which is quite different from that of lexical words. To begin with, each of the main components – handshape, location, movement – usually has meaningful morphological status.

This is in direct contrast with ordinary words of sign languages, in which each of these categories is strictly phonological, and, by definition, meaningless. When the morphemes combine, they do not create lexemes, but rather expressions that translate as full sentences, such as 'A small animal is sitting on a log,' or 'A vehicle drove over a hill.'

Distributional and prosodic properties of classifier constructions provide more reasons for regarding them as distinct from lexemes or words (Sandler and Lillo-Martin 2006). Although a classifier construction might span only one syllable, giving it the appearance of a word (Brentari 1995), a single classifier might also span several prosodic constituents (such as intonational phrases) without re-articulation of the handshape (Aronoff et al. 2003). The handshape remains constant across a sequence of movements traversing several locations and representing a chain of events in which the classifier's referent participates. An event in which a car, for example, drives uphill, turns right, then left, then parks along an incline is likely to be represented in this way. In this respect, too, classifier constructions are unlike lexical words. The individual morphemes in the classifier subsystem, each a minimal pairing of form and meaning that recombines productively with other morphemes in the system, must be assumed to be independently listed in the lexicon, like other morphemes. But crucially, each is a bound morpheme and cannot constitute a word by itself. Their meanings, their phonological forms, and their prosodic properties suggest that the morphemes of the classifier system are combined post-lexically (Sandler and Lillo-Martin 2006).

Of special interest to us here is a particular anomaly of this subsystem that provides additional evidence for the claim that the Symmetry and Dominance conditions on the nondominant hand are lexical and phonological and not imposed by the phonetics. In classifier constructions, the nondominant hand (h2) can function as an independent classifier (Supalla 1982), and as such can freely break the phonological constraints that are strictly enforced on h2 in words.

In Figure 5, h2 is configured as an AIRPLANE classifier, and the dominant hand as an UPRIGHT-HUMAN. It is taken from an ISL utterance meaning, 'a person approaches an airplane'. In this figure, the dominant, moving hand has one handshape, and the nondominant, static hand has a different handshape, , and one that is marked. In a lexical word, this combination of configurations in an otherwise similar structure would be ruled out by the Dominance Condition.

The Symmetry Condition is violated in classifier constructions as well. In discourse context, Figure 6 means something like 'a person proceeds forward, dragging a dog squirming behind' in ASL.[8] One hand represents an upright human and the other a legged-creature – a different configuration on each hand, each hand moving in a different shaped path. The structure is not a possible word of ASL or ISL, ruled out by the Symmetry Condition.

Figure 5. ISL classifier construction: 'a person approaches an airplane'.

Figure 6. ASL classifier construction: 'a person proceeds forward, dragging a dog squirming behind'.

Configured and functioning as a classifier, the nondominant hand can exhibit a great deal of independence in the discourse. For example, it can remain in the signing space throughout a discourse segment to background the referent it represents (see Brentari and Crossley (2002) for references). Miller (1994) describes discourse-regulation devices such as placing a classifier to mark a location and directing a non-classifier with respect to it.

This discussion does not mean to imply that there is no relation between lexical words and classifier constructions. A considerable number of words in any sign language lexicon are believed to have originated as classifier constructions and to have become lexicalized. An example is ISL WRITE, shown in Figure 7.

Figure 7. ISL lexical word: WRITE.

Presumably, WRITE originated as a classifier construction in which the dominant hand, h1, is configured as the handler of a small manipulable object, and h2 represents a flat object, the piece of paper being written on.

As with any other words and unlike the classifier constructions from which it evolved, each of the formational categories in WRITE (handshape, location, movement) is meaningless (Stokoe 1960), and the form behaves in all phonological and morphological respects like a word.[9] However, the grammar treats words differently from classifier constructions.

Sign language classifier constructions are in some respects unique to the sign modality (Schembri 2003). However, even this system bears grammatical and functional similarities to verbal classifiers of some spoken languages, such as Cayuga and Digueño (Aronoff et al. 2003), and it would be a mistake to deduce that sign language morphology is modality specific. On the contrary, other more familiar lexical morphological processes such as verb agreement, temporal aspect inflection, and derivational processes are widely found in these languages as well, and with more familiar formal instantiation, templatic as well as affixal (e.g., Aronoff, Meir and Sandler, 2005; Padden 1988).

Classifier constructions have been introduced into the present discussion for two reasons. First, they underscore the possibility that the redundancy of h2 in words plays a role in the system, namely, to mark words as words, distinct from classifier constructions that are formed at the post-lexical level. Second, they show that the constraints on h2 in words are not phonetic, as they are freely violated in classifier constructions. Instead, they are related to a higher level of linguistic structure – the morphosyntactic word – and are therefore phonological.

4. A grammatical role for h2: the delineation of prosodic constituents

The next question to be addressed is this: does the phonological redundancy of h2 in words mean that the nondominant hand plays no grammatical role in the phonology of sign languages? Interestingly, the answer is 'no'.

An investigation of the prosodic structure of Israeli Sign Language conducted in large part with Marina Nespor (Nespor and Sandler 1999; Sandler 1999b) revealed that h2 functions as a delineator of boundaries of two prosodic constituents: the phonological word and the phonological phrase. A brief review of those results will demonstrate how this phonetic element is recruited by the prosodic phonology.

Throughout the following discussion, Nespor and Vogel's (1986) theory of prosodic phonology is assumed. The prosodic hierarchy shown below is the same as theirs, except that here their clitic group category is subsumed by the next higher category, phonological word (as explained below).

Prosodic hierarchy (following Nespor and Vogel 1986)
mora < syllable < foot < phonological word < phonological phrase < intonational phrase < phonological utterance

In our study, 30 Hebrew sentences were created with the purpose of determining whether prosodic constituents and intonation exist in ISL, and if so what their properties are. Based on patterns known from spoken languages, the target stimuli were designed to elicit simple, declarative sentences; longer, more complex strings like sentences with relative clauses or sentential complements; yes-no questions; and wh-questions. Three trained native signer consultants were asked to read each sentence, internalize it, put the paper aside, and sign the sentence in natural ISL to another native signer seated by the camera. All 90 sentences (30 x 3 signers) were recorded on videotape and subsequently glossed with the help of a native signer consultant, one sentence per coding sheet. Each observable facial articulator (eyebrows, eyelids, cheeks, nose, and mouth) was listed down the lefthand side of the sheet, as were head and body. In addition, rhythmic and prominence properties of the hands, such as size, speed, number of iterations, pauses and holds were also listed. Coding consisted of describing the action of each articulator on its line, and drawing a line tracing its scope vis à vis the glossed text at the top of the page. All coding was done by a research assistant together with a native signer consultant, by viewing the data on videotape repeatedly, in slow motion, and recording the activity of hands, face, and body next to the relevant coding category listed beneath the gloss. The extent of each articulation was indicated by drawing a solid line opposite the articulator label under the words in the gloss that were characterized by it (see above references for a sample).

The coding revealed that the sentences were divided rhythmically by the hands and head into prosodic constituents, specifically, phonological words, phonological phrases and intonational phrases, and that the scope of the facial expression adhered to that rhythm (in particular, changing dramatically at intonational phrase boundaries).[10] After noticing that the nondominant hand often behaves differently in connected signing than might be predicted from the citation form of signs, we tracked and coded its behavior as well by adding a line for h2 on the coding sheet and describing its behavior there, drawing a line for each exemplar to show the scope of the behavior as with the other categories. We now turn to results of the prosody study that are relevant to the nondominant hand.

4.1. The phonological word

By phonological word I mean a morphosyntactic word plus any surrounding words (typically function words) that are part of the same stress group. The category is similar to that defined in Nespor and Vogel (1986), but different in one respect: the category 'phonological word' is collapsed with the category 'clitic group'. Thus, for our purposes, *Mary's*, in an English sentence like *Mary's on the phone*, is a single phonological word, formed from the two morphosyntactic words, *Mary* and *is.* The formation of phonological words from more than one morphosyntactic word is typically achieved by reducing a function word and assigning it weak stress, relative to the adjacent content word to which it is grammatically related. Thus, the process makes reference to the lexical category of the words and the grammatical relation between them.

In Israeli Sign Language (ISL), if a symmetrical sign is followed by a pronoun in the prominent (final) position of a phonological phrase, the pronoun can cliticize to its host through coalescence to form a single phonological word. The nondominant hand articulates only the host sign, while the dominant hand smoothly articulates the host and the clitic pronoun in the same time span. Figure 8 is extracted from a sentence meaning 'The shop around the corner went bankrupt'. Use of the deictic sign THERE is a typical sign language device, locating a concrete noun in space at first mention, establishing a locus for potential spatial referencing later in the discourse.

A. SHOP B. THERE C. SHOP-THERE

Figure 8. ISL SHOP, THERE, and the cliticized form SHOP-THERE.

In 8a, the sign SHOP is a symmetrical two-handed sign and the deictic pronoun THERE 8b is one-handed, normally signed with the dominant hand. In the cliticized phonological word, the dominant hand signs only half of the sign SHOP, and then changes hand configuration to that of the sign THERE, while moving forward to complete that sign – a coalescence process shown in 8c.

What is of interest here is the behavior of h2. As the dominant hand coalesces, blending two signs, the nondominant hand simply completes the host sign, SHOP, over the same temporal span. The nondominant hand articulates a single syllable (i.e., one movement), and the dominant hand also articulates a single movement, though with a sequence of two different handshapes. In a noncliticized version of these two words, two movements (syllables) would be required to sign SHOP and then THERE. Under coalescence, the cliticized THERE loses its syllabicity, a phenomenon found also in spoken language clitics, e.g., in English aux contraction (Selkirk 1984). In the ISL case specifically, the effect is to create a monosyllable over the domain of the phonological word. The process occurs not only with THERE, but also with both personal and possessive pronouns. It does not occur between two content words, which indicates that the process has access to the grammatical distinction between the two kinds of words.

The effect is to make the cliticized form more like a typical word in its phonological form (Sandler 1999a, 1999b), since sign language words are typically monosyllabic (Coulter 1982). In this process, then, h2 serves to mark the boundaries of the phonological word. As a post-lexical phenomenon, coalescence is non-structure preserving – it violates the Symmetry Condition which requires two active hands to be symmetrical in shape, movement, and location.[11] But as a phonological process, it is discrete. The nondominant hand demarcates precisely the boundary of the host plus clitic, and no more.

4.2. The phonological phrase

The next level up in the prosodic hierarchy is the phonological phrase, projected from the heads of syntactic phrases such as NPs, VPs, and AdjPs (see Nespor and Vogel 1986, for a formal definition and explanation). Though not always isomorphic with the syntactic boundaries of those phrases, phonological phrases are linked to them and, like intonational phrases, they are linked to the syntax. The algorithm below for forming phonological phrases is adapted from Nespor and Vogel's:

> The domain of a phonological phrase consists of a …lexical head X, and all [phonological words] on its non-recursive side up to the [phonological word] that contains another head outside of the maximal projection of X.

The phonological phrase constituent is identifiable by minor rhythmic breaks. For example, the square brackets divide the following sentence into phonological phrases that would be likely to occur at a normal to slow rate of speech: *[The very tall] [construction worker] [carefully walked] [under the ladder].* If the phrases are shorter, under certain circumstances, Nespor and Vogel show that phrases may be restructured together, becoming non-isomorphic with the individual syntactic phrases projected from each lexical head, e.g., [He ate] [a hearty lunch] versus [He ate lunch], the latter restructured from [He ate] [lunch]. Restructured or no, under normal circumstances, prosodic constituency does not disrupt syntactic constituency, so that the following divisions are impossible: **[The very] [tall construction] [worker walked under] [the ladder].* Instead, prosodic constituency may be thought of as interpreting the syntax.

Phonological evidence for the existence of this prosodic constituent is found in external sandhi rules whose application is restricted to the domain of the phonological phrase. An example is French liaison (Nespor and Vogel 1986; Selkirk 1986). The underlying final consonant, normally deleted, is pronounced before vowel-initial words if the two words are in the same phonological phrase. Liaison does not apply between words across a phonological phrase boundary.[12] In the sentence, *Les enfants [sont^allés] φ à l'école.* taken from Nespor and Vogel (1986), there is liaison (signified by the symbol ^) between *sont* and *allés* within a phonological phrase; i.e., the [t] of *sont* is pronounced. But there is no liaison between *allés* and *à* (i.e., the final consonant of *allés* is *not* pronounced) because a phonological phrase boundary (φ) intervenes.

ISL utterances are also divided into phonological phrases. Final phonological phrase boundaries are marked phonetically by holds, reiterations of the last sign, or pauses (Nespor and Sandler 1999; Sandler 1999b). The sentence below is divided into two intonational phrases (each marked with an I index), the first containing three phonological phrases, and the second containing two phonological phrases.[13]

[I-TELL HIM]$_{\varphi}$ [BAKE CAKE]$_{\varphi}$ [TASTY]$_{\varphi}$]$_{I}$ [[ONE FOR ME]$_{\varphi}$ [ONE FOR SISTER]$_{\varphi}$]$_{I}$
'I told him to bake a tasty cake, one for me and one for my sister'.

Confirmation for the existence of the phonological phrase constituent was found in an external sandhi rule involving h2, called Nondominant Hand Spread (NHS). Unlike French liason, this sandhi rule does not involve sequential segments. Rather, the spread of the nondominant hand from the triggering two-handed sign is simultaneous with the signing of other words by the dominant hand. An example is the phrase, BAKE CAKE from the sentence above.

Figure 9 illustrates NHS in this sentence. In it, the nondominant hand from the sign BAKE spreads to the end of the phonological phrase by remaining in the same configuration as in the source sign, BAKE, throughout the next sign, CAKE, which is a one-handed sign. The end of the phonological phrase is marked by a hold – holding the hand in position at the end of the last sign. Precisely at the onset of the next phonological phrase, $[\text{TASTY}]_{\varphi}$, the sandhi stops, and the hand assumes a neutral shape. The illustration shows the signs BAKE^CAKE in Figures 9b and 9c with NHS. Figure 9a shows the sign HIM in the phonological phrase that precedes BAKE^CAKE, and Figure 9c shows the sign TASTY in the phonological phrase that follows it.

A. B. C. D.

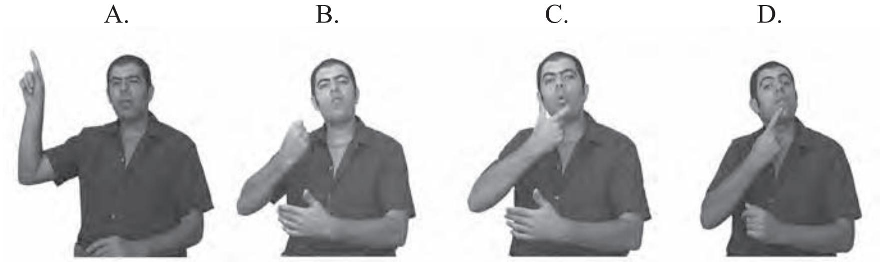


Figure 9. A. HIM; B. BAKE; C. CAKE; D. TASTY. Nondominant Hand Spread in the phonological phrase, $[\text{BAKE CAKE}]_{\varphi}$ blocked from spreading across the boundary either to $[\text{HIM}]_{\varphi}$ in the preceding phrase, or to $[\text{TASTY}]_{\varphi}$ in the following phrase.

In our corpus, the spread of h2 is only to the edge of the phonological phrase, as seen clearly in this example. Importantly, two other conceivable explanations for the blocking of NHS spread are ruled out here. If there were a two-handed sign in the following phrase, the presence of a specified h2 there might conceivably be blamed for the fact that NHS stops where it does. However, the sign in the phrase that follows the one with NHS, TASTY, is one-handed, so it is not implicated. The other conceivable explanation is the presence of a higher (and stronger) constituent boundary, that of the Intonation Phrase. This possibility is also ruled out in the present example, because, according to both syntactic and prosodic criteria established in the Nespor and Sandler (1999) study, the IP boundary occurs after the next phonological phrase, $[\mathrm{TASTY}]_{\varphi}]_{I}$. In our corpus, there were no exceptions to the blocking effect of the phonological phrase boundary. Furthermore, the spread is discrete and not gradient. Repeated slow motion viewings revealed that NHS extends clearly to the edge (right, left, or both) of the prosodic constituent. From this behavior, together with the fact that phonological phrases are linked to syntactic phrases, it appears that NHS is a phonological process, and not a phonetic process of coarticulation.[14] As a post-lexical process, it is, like coalescence, non-structure preserving: the result of NHS is the presence of two places of articulation during the production of one sign (here, the sign 'CAKE'). At the lexical level, only one place of articulation is licensed (Battison 1978).

I've suggested that h2 participates in signs, despite its redundancy, partly because of the physiology of the system, which includes motoric organization of bimanual coordination. The behavior of h2 in the 'cake' sentence (Ex. 9) is a good example of this. The signer's nondominant hand is present in the signing space, in a neutral configuration and location, even when it is participating neither in a two-handed sign nor in Nondominant Hand Spread. The neutral configuration for this particular signer is one in which the index finger is slightly prominent and the rest of the fingers are loosely curled.[15] The transition from the handshape of BAKE to this neutral configuration, assumed precisely at the onset of the next phonological phrase, difficult to convey in still pictures, is striking on the videotape.

5. Conclusion

We've now traced the nondominant hand in a full circle through the grammatical system of sign language. In the lexical words of sign languages, there is only one major articulator, the dominant hand, and the nondominant hand plays a subordinate and largely redundant role in which its handshape, place of articulation, and movement are severely restricted. The prosodic system exploits this seemingly redundant element to demarcate prosodic constituents at different levels of the prosodic hierarchy, as we have seen. A meaningful role for the nondominant hand is seen in the classifier subsystem, where it has the status of a morpheme and enjoys almost as much articulatory freedom as the dominant hand.[16] At the discourse level, a level that was touched on briefly in §2, h2 can be quite useful in rhetorical devices of sign languages, tracking referents and backgrounding portions of the discourse while the dominant hand simultaneously continues to 'talk', using lexical words in syntactic constructions.

It should be quite clear by now that the nondominant hand is a phonetic articulator utterly unique to sign language, that there is no corresponding element in spoken languages. At the two extremes of the grammar – the classifier subsystem and the discourse level – the nondominant hand is an equipotential articulator, representing entities and concepts like its physiological twin, the dominant hand. No individual phonological element in spoken languages functions systematically and cross-linguistically as a full morpheme. In addition to the modality specific range of functions performed by h2 in sign language grammar, the physical and dynamic properties of this articulator are also obviously different from those of any spoken language articulator. Insofar as such properties form an integral part of phonetic systems, we must say that the phonetics of h2 is unlike those of any articulator in spoken language. We will return to this point below.

The linguistic roles that h2 plays, however, do have direct counterparts in spoken language. Let's first consider words in the lexicon. Just as spoken words have structural constraints, so do signs. In most signed words, h2 must either be an articulator that is symmetrical with h1, or it must be a place of articulation in which handshape is underspecified. Spoken words are constrained differently, e.g., in terms of the number and types of consonants that may appear initially and finally. While the specific constraints are different, the fact that phonological elements and their co-occurrence are constrained within words is well known from spoken languages. Presumably, this kind of predictability about the shape of a linguistic entity has acquisition and processing advantages in both modalities.

Another way in which sign language phonology behaves like that of spoken language despite very different phonetics is in signaling prosodic constituents. In many languages, assimilation rules that cross word boundaries, external sandhi rules, have been shown to respect particular prosodic constituent edges. One can think of this merging of words that stops at the boundary of a prosodic constituent as a way of binding together words within that constituent. In spoken languages, where linear structure is prominent, sandhi normally occurs between adjacent segments. In sign languages, in which structure at all levels is more simultaneous, sandhi can merge more than just adjacent segments; it can merge whole stretches of words.[17] This is the effect of coalescence and of NHS. The seemingly redundant articulator of sign language, the nondominant hand, can unite a host and clitic, and it can spread simultaneously across whole words within a phonological phrase, with the effect of binding together the words in the constituent. The phonetics of external sandhi processes is starkly different in the two modalities – affecting adjacent sequential segments in spoken language, and simultaneously affecting an entire word or more than one word within a prosodic constituent in sign language – but the phonological role is arguably quite similar.

This investigation, considering only one element in the structure of sign language, reveals both universal and modality-specific properties. The specific feature pool and classes, and the details of phonological processes are not universal; they differ in the two modalities. Furthermore, spoken languages tend to have sequential organizing properties and phonological processes that affect sequentially arranged elements, while sign languages have a good deal more nonlinear or simultaneous structure and processes.[18]

But there are also significant properties that this brief investigation of the nondominant hand in sign language shows to be universal. The existence of features, feature classes, and processes that systematically change underlying form are linguistic universals. For example, in the model assumed here, h2 may either be represented as a member of the HC class or of the Place of Articulation class, and in each case it behaves like other members of its class in the morphophonology.[19] This indicates that languages universally organize features into classes (in the sense of Clements 1985), regardless of modality. In addition, the very fact that the surface forms of signs can differ systematically and discretely from the underlying forms, i.e., that there are phonological rules, is also universal. The present investigation has demonstrated three such rules involving h2: assimilation in compounds, coalescence and Nondominant Hand Spread. Concomitantly, we see that when words are strung together in sentences, constraints on canonical word form are relaxed. In other words, there is a distinction between lexical and post-lexical levels in both modalities.

We have also seen that the existence of underspecified or default forms is a universal property of language. The nondominant hand must share the specification of h1 in symmetrical signs; it gets its specifications by default. In signs in which h2 is a place of articulation, it may only be specified for one of a few unmarked handshapes, another form of underspecification (Sandler 1995, 1996; van der Kooij 2002).

Finally, the segmentation of the language stream into prosodic constituents that are linked to morphosyntactic constituents is universal. These constituents are marked by cues related to rhythm and prominence, and their cohesion is often reinforced by sandhi rules that may not cross the prosodic constituent boundary.[20] The present discussion points to two such phenomena, both of them sandhi-type rules instantiated by the nondominant hand. Coalescence characterizes the phonological word constituent (host+clitic), while Nondominant Hand Spread characterizes the phonological phrase.

This investigation has three interrelated messages to convey about phonology, phonetics, and the relation between them. First, we've seen evidence that at least some phonological restrictions of sign language do not derive directly from the phonetics of that system. In the classifier subsystem and for rhetorical effect, the Symmetry and Dominance conditions are freely violated in the same language, as we saw in §2 and §3. These constraints must not be dictated solely by phonetics, but answer instead to a higher authority: the morphosyntactic entity, 'lexeme.' Second, significant aspects of phonology were shown to be common to languages in two phonetically different modalities, implying dissociation between phonetics and phonology in a different way. For example, assimilations within prosodic constituents delimit the scope of a prosodic domain that is linked to syntactic structure in both modalities – but the phonetic means recruited by spoken and signed language to achieve this end are unrelated. Third, in both modalities, there is a linguistic component, phonology, that is characterized by a relation to higher levels of grammatical structure. The sign language examples of this relation discussed here were (a) constraints in the lexicon that don't hold post-lexically (§1, §2), and (b) processes that characterize prosodic constituents linked to morphosyntax (§3).

The investigation reported here does not imply that there is no relation between phonetics and phonology, of course. Instead, it may help us be more explicit about what the relation might be. We might begin by drawing a distinction between inherent phonetic properties and phonetic organizing principles. While the former are distinct in spoken and signed language, the latter may share common ground.

Insofar as the physical and dynamic properties of specific articulators are integral to phonetics, effects that are directly related to these properties are expected to be different in the two modalities. And indeed, we see that certain assimilatory processes are instantiated simultaneously by two anatomically independent but identical articulators in sign language – an impossibility for spoken language. Both the articulatory behaviors and the physical apparatus behind them are so different in spoken and signed language that any attempt to derive the specific assimilatory behaviors found in each modality via the same mechanism seems doomed.

However, the organizing principles of phonetics may derive from a common base, a possibility with potentially interesting theoretical consequences. Cheek (2001) argues that handshape coarticulation in ASL – which she distinguishes from the handshape assimilation found in compounds – bears the following similarities to coarticulation in speech: it is gradient; it is affected by rate; and it is explained by principles of economy. A similar possibility worth investigating was raised by an anonymous reviewer: that general principles underlying rhythmic motor behavior influence assimilatory effects within prosodic constituents in both modalities. Furthermore, the claim that certain principles unify phonetics and phonology (e.g., Flemming 2001; Ohala 1990), may well be valid across modalities. For example, it is likely that phonetic coarticulation is the source of the phonological sandhi rules in both French and ISL. The present study implies that the articulatory path from one to the other cannot be the same, but the organizing principles behind them may be shared.

In sum, this investigation has highlighted some architectural similarities in the phonology and prosody of spoken and signed languages. It has done so by documenting the behavior of the nondominant hand in sign language, an articulator with no phonetic counterpart in speech.

Notes

1. The models of the nondominant hand presented in the works referred to here vary. However, there is a general consensus that the nondominant hand is not an independent articulator in the phonology of sign language.
2. A syllable is defined as one movement, either (a) along a path, (b) internally through handshape or orientation change, or (c) the two simultaneously.
3. Battison's set of unmarked shapes was larger than the one shown here. The set I assume here abstracts away from small differences in the degree of closure and spread between the fingers which are either phonologically predictable or noncontrastive (Sandler 1995, 1996).
4. In the case of symmetrical two-handed signs, the representation reflects the least marked possibility for two-handed signs: one in which only one handshape is represented.
5. So far only Cheek (2001) has done instrumental studies of coarticulation between words. See discussion in §4.
6. Not all post-lexical rules are phonetic. However, rules and constraints associated with the lexicon and word formation are assumed here to be phonological and not phonetic. See Cohn (1993) for discussion.

7. The hand configuration assimilation rule supports an additional claim about ASL phonology, namely, hierarchical representation of feature classes, in which handshape dominates orientation. The representation is motivated by the fact that orientation in compounds may assimilate alone, but if handshape assimilates, orientation must assimilate as well (Sandler 1987, 1989). Such a representation is similar in principled ways to the representation of feature classes proposed to characterize spoken language phonology (Clements 1985). This requirement breaks down post-lexically: handshape may coarticulate without orientation at that level (Corina 1993; Sandler 1993b), providing further evidence for a distinction between lexical and post-lexical levels.
8. There are some differences between the classifier systems of ASL and ISL (Aronoff et al. 2003), but they are not relevant to the analysis presented here.
9. Another difference between classifier constructions and their verbal word counterparts is that the latter may undergo aspectual inflection, while classifier constructions may not (Sandler and Lillo-Martin 2006).
10. We concluded from this and other kinds of analysis that facial expression corresponds to intonation in sign language (Nespor and Sandler 1999; Sandler 1999b; developed in Sandler and Lillo-Martin 2006).
11. This lexical-post-lexical distinction relies on Kiparsky (1982, 2000).
12. According to Selkirk (1972), the prosodic behavior of liaison is most consistent in informal registers.
13. In our data, the cues to a phonological phrase boundary – hold, reiteration or pause – were in complementary distribution, suggesting that they perform the same function, but we cannot yet predict under which conditions each will occur. The reader is referred to Nespor and Sandler (1999) for discussion of the intonational phrase markers.
14. NHS is optional, and other factors confound the tally, so that the number of clean examples in our data, though exceptionless, is small (9). However, if phonological phrase boundaries that coincide with intonational phrase boundaries are not ruled out, and are instead tallied together with the spreads that coincide only with the lower phonological phrase boundary alone, then the number of spreads that are stopped by some prosodic boundary increases greatly, reinforcing the claim that the process is discrete.
15. The neutral location for the nondominant hand in running signing is close to and in front of the body (coincidentally at a location near that specified for the sign BAKE).
16. As a ‘secondary classifier’, h2 depicts the ground, while the dominant hand depicts the figure (Supalla 1982). For this reason, and probably due also to motoric constraints, the dominant hand typically moves and changes configuration more than the nondominant hand in classifier sequences.

17. Simultaneity at all levels of structure is described in detail, and both the reasons and the implications connected with this kind of structuring examined, in Sandler and Lillo-Martin (2006).
18. In addition to widespread 'simultaneous' structure, there is also significant sequential structure in sign language phonology and morphology. See Liddell (1984), Liddell and Johnson (1989), Sandler (1989; Sandler 1993c), and Sandler and Lillo-Martin (2006) for specific arguments and analyses.
19. Detailed arguments for this model and for the representation of h2 either as a member of the HC class or as a member of the place of articulation class are presented in Sandler (1993a). For alternative analyses in which h2 is represented in a unitary way regardless of sign type, see van der Hulst (1996) and Brentari (1998). The models are compared in Sandler and Lillo-Martin (2006).
20. As is the case with the other properties mentioned here, the bulk of the evidence for such constituents is presented elsewhere (Nespor and Sandler 1999; Sandler 1999a, 1999b).

References

Aronoff, Mark, Irit Meir, Carol Padden and Wendy Sandler
2003 Classifier constructions and morphology in two sign languages. In: Karen Emmorey (ed.), *Perspectives on Classifier Constructions in Sign Languages*, 53–84. Mahwah, NJ: Lawrence Erlbaum Associates.

Aronoff, Mark, Irit Meir and Wendy Sandler
2005 The paradox of sign language morphology. *Language* 81/2: 301–344.

Battison, Robbin
1978 *Lexical Borrowing in Americal Sign Language*. Silver Spring, MD: Linstok Press.

Brentari, Diane
1990 Theoretical Foundations of ASL Phonology. Ph.D. dissertation, University of Chicago.
1995 Sign language phonology. In: John A. Goldsmith (ed.), *The Handbook of Phonological Theory*, 615–639. Cambridge, MA: Blackwell.
1998 *A Prosodic Model of Sign Language Phonology*. Cambridge, MA: MIT Press.

Brentari, Diane and Laurinda Crossley
2002 Prosody on the hands and face: Evidence from American Sign Language. *Sign Language and Linguistics* 5: 105–130.

Brentari, Diane and John A. Goldsmith
1993 Secondary licensing and the nondominant hand in ASL phonology. In: Geoffrey Coulter (ed.), *Current Issues in ASL Phonology (Phonet-*

ics and Phonology 3), 19–41. New York, San Francisco: Academic Press.

Cheek, Adrianne
2001 The Phonetics and Phonology of Handshape in American Sign Language. Ph.D. dissertation, University of Texas, Austin.

Clements, G. N.
1985 The geometry of phonological features. *Phonology Yearbook* 2: 223–250.

Cohn, Abigail
1993 Nasalization in English: phonology or phonetics. *Phonology* 10: 43–81.

Corina, David P.
1993 To branch or not to branch: Underspecification in ASL handshape contours. In: Geoffrey Coulter (ed.), *Current Issues in ASL Phonology (Phonetics and Phonology 3)*. New York, San Francisco: Academic Press.

Coulter, Geoffrey
1982 On the nature of ASL as a monosyllabic language. Paper presented at: *Annual Meeting of the Linguistic Society of America*. San Diego.

Crasborn, Onno
2001 Phonetic Implementation of Phonological Categories in Sign Language of the Netherlands. Ph.D. dissertation, University of Leiden.

Enfield, Nick
2004 On linear segmentation and combinatorics in co-speech gesture: a symmetry-dominance construction in Lao fish trap descriptions. *Semiotica* 148-1/4: 57–123.

Flemming, Edward
2001 Scalar and categorical phenomena in a unified model of phonetics and phonology. *Phonology* 18: 7–44.

Jakobson, Roman
1968 *Child Language, Aphasia, and Phonological Universals*. The Hague: Mouton.

Kingston, John and Mary E. Beckman
1990 *Papers in Laboratory Phonology I*. Cambridge: Cambridge University Press.

Kiparsky, Paul
1982 Lexical phonology and morphology. In: In-Seok Yang (ed.), *Linguistics in the Morning Calm*. Seoul: Hanshin.
2000 Opacity and cyclicity. *The Linguistic Review* 17: 351–367.

Levy, Doron
2001 Weak Drop in Israeli Sign Language. Unpublished manuscript, Tel Aviv University.

Liddell, Scott K.
1984 THINK and BELIEVE: Sequentiality in American Sign Language. *Language* 60: 372–392.
Liddell, Scott K. and Robert Johnson
1986 American Sign Language compound formation processes, lexicalization, and lexical phonological remnants. *Natural Language and Linguistic Theory* 4: 445–513.
1989 American Sign Language: The phonological base. *Sign Language Studies* 64: 197–277.
Mandel, Mark
1981 Phonotactics and Morphophonology in American Sign Language. Ph.D. dissertation, University of California, Berkeley.
McNeill, David
1992 *Hand and Mind: What Gestures Reveal about Thought*. Chicago: University of Chicago Press.
Miller, Christopher
1994 Simultaneous constructions in Quebec Sign Language. In: Mary Brennan and Graham H. Turner (eds.), *Word-order issues in Sign Language-Working papers*, 89–112. Durham, UK: International Sign Linguistics Association.
Nespor, Marina and Wendy Sandler
1999 Prosody in Israeli Sign Language. *Language and Speech* 42(2-3): 143–176.
Nespor, Marina and Irene Vogel
1986 *Prosodic Phonology*. Dordrecht: Foris.
Ohala, John J.
1990 There is no interface between phonetics and phonology: a personal view. *Journal of Phonetics* 18: 153–171.
Padden, Carol
1988 *Interaction of Morphology and Syntax in American Sign Language*. New York: Garland Publishers.
Padden, Carol and David Perlmutter
1987 ASL and the architecture of phonological theory. *Natural Language and Linguistic Theory* 5: 335–375.
Perlmutter, David
1991 Representing the non-dominant hand. Paper presented at: *The Linguistic Society of America Summer Institute*. Santa Cruz, CA.
Sandler, Wendy
1987 Assimilation and feature hierarchy in American Sign Language. *Chicago Linguistics Society* 23: 266–278.
1989 *Phonological Representation of the Sign: Linearity and Nonlinearity in Sign Language Phonology*. Dordrecht: Foris.

1993a Hand in hand: The roles of the nondominant hand in sign language phonology. *The Linguistic Review* 10: 337–390.
1993b Linearization of phonological tiers in ASL. In: Geoffrey Coulter (ed.), *Current issues in ASL phonology (Phonetics and phonology 3)*, 103–129. San Francisco, New York: Academic Press.
1993c A sonority cycle in American Sign Language. *Phonology* 10: 243–279.
1995 Markedness in the handshapes of sign language: A componential analysis. In: Harry van der Hulst and Jeroen van de Weijer (eds.), *Leiden in last: HIL phonology papers 1*, 369–399. The Hague: Holland Academic Graphics.
1996 Representing handshapes. *International Journal of Sign Linguistics* 1: 115–158.
1999a Cliticization and phonological words in a sign language. In: T. Alan Hall and Ursula Kleinhenz (eds.), *Studies on the Phonological Word*, 223–254. Amsterdam: John Benjamins.
1999b The medium and the message: Prosodic interpretation of linguistic content in Israeli Sign Language. *Sign Language and Linguistics* 2: 187–215.

Sandler, Wendy and Diane Lillo-Martin
1999 Natural sign languages. In: Mark Aronoff and Janie Rees-Miller (eds.), *The Handbook of Linguistics*, 533–562. Oxford: Blackwell.
2006 *Sign Language and Linguistic Universals*. Cambridge, UK: Cambridge University Press.

Schembri, Adam
2003 Rethinking “classifiers” in sign language. In: Karen Emmorey (ed.), *Perspectives on Classifier Constructions in Sign Languages*. Mahwah, NJ: Lawrence Erlbaum Associates.

Selkirk, Elisabeth O.
1972 The Phrase Phonology of English and French. Ph.D. dissertation, MIT.
1984 *Phonology and Syntax: The Relation between Sound and Structure*. Cambridge, MA: MIT Press.
1986 On derived domains in sentence phonology. *Phonology* 3: 371–405.

Slobin, Dan I., Nini Hoiting, Marlon Kuntze, Reyna Lindert, Amy Weinberg, Jennie Pyers, Michelle Anthony, Yael Biederman and Helen Thumann
2003 A cognitive/functional perspective on the acquisition of “classifiers”. In: Karen Emmorey (ed.), *Perspectives on classifier constructions in sign languages*, 271–296. Mahwah, NJ: Lawrence Erlbaum Associates.

Stokoe, William C.
1960 *Sign Language Structure*. Linstok Press.

Supalla, Ted
1982 Structure and Acquisition of Verbs of Motion in American Sign Language. Ph.D. dissertation, University of California, San Diego.
1986 The classifier system in American Sign Language. In: Colette G. Craig (ed.), *Noun Classification and Categorization*, 181–214. Philadelphia: John Benjamins.
Uyechi, Linda
1996 *The Geometry of Visual Phonology*. Stanford: CSLI Publications.
van der Hulst, Harry
1996 On the other hand. *Lingua* 98: 121–143.
van der Kooij, Els
2002 Phonological Categories in Sign Language of the Netherlands: The Role of Phonetic Implementation and Iconicity. Ph.D. dissertation, University of Leiden.

Lexical retrieval in American Sign Language production

David P. Corina and Heather P. Knapp

Processes involved in the lexical retrieval of signs in naturally-occurring manual languages of the deaf are poorly understood. In the current experiments, we assess the time-course of semantic and phonological form retrieval in a visual-manual language, American Sign Language (ASL), using a sign-picture interference task. We find that native signers of ASL exhibit longer naming latencies in the presence of semantically-related sign distractors than in the presence of unrelated distractors at early, but not late, stimulus onset asynchronies (SOAs). In contrast, phonologically-related stimuli produce naming facilitation both early and late in the naming process, but effects vary by degree and type of phonological relatedness. A detailed analysis of the phonological effects provides support for independently-motivated models of sign structure that give prominence to movement and location properties of signs. This study represents the first systematic investigation of lexical retrieval during sign language production, and provides important insights into the commonalities underlying the lexical organization of signed and spoken languages.

1. Introduction

The small yet rapidly-growing field of signed language psycholinguistics has developed largely around studies exploring issues in syntactic, morphological, and single sign comprehension. The present paper extends this body of literature to an under-represented research domain, psycholinguistic processes underlying the retrieval and production of American Sign Language (ASL) signs.

Models of spoken language lexical organization (e.g., Dell, 1986, 1988; Dell and Reich, 1981; Garrett, 1975, 1976; Glaser, 1992; Levelt, 1989; Levelt, Reolofs, and Meyer, 1999; Morton, 1969) have attempted to characterize the nature and temporal ordering of semantic and phonological form retrieval during word production. The issue of whether a word's semantic and phonological form representations are retrieved simultaneously or in two successive, non-overlapping stages during word production is important because such ordering provides a window into the psychological structure of stored lexical items. For example, whether lexical concepts are *retrieved* (1) discretely, (2) in overlapping fashion, or (3) simultaneously may reveal whether semantic concepts and their corresponding phonological forms are mentally *represented* as (1) isolated yet complementary word components, (2) dynamic units interacting during real-time language processing, or (3) holistic entities whose components are linked arbitrarily in linguistic theory (de Saussure, [1906–1911] 1983) but inextricably in representation and practice.

Current models of the time-course of lexical retrieval are based solely upon observations from spoken language production. Recent data suggesting linguistic and processing differences between speech and sign language representations (for a recent review see Emmorey, 2002) prompt an assessment of whether the organization of the mental lexicon implied by semantic and phonemic retrieval patterns holds true for human language generally, regardless of the sensory-articulation channels through which a particular language is perceived and produced.

For example, sign language and speech morphological and phonological structure differ with respect to the temporal ordering of sub-lexical segments. In spoken languages, lexical segments are produced and perceived sequentially, whereas in sign languages, the sub-lexical segments of hand configuration, movement, and place of articulation segments occur contemporaneously (e.g., Sandler, 1993; Brentari, 1998). Indeed, even the existence of a level of representation corresponding to the segment in signed languages has been questioned (e.g., Wilbur, 1998). However, in the psycholinguistic domain, on-line lexical processing studies have had difficulty establishing sub-lexical phonological effects on sign recognition (but see Dye and Shih, this volume, for a report of phonological priming effects in BSL). This is true despite an abundance of evidence for semantic processing effects (e.g., Hildebrandt and Corina, 2002). Finally, growing evidence for both similarities and differences between speech and sign at the neural level have been observed during sentence processing tasks (e.g., Neville, 1998).

Therefore, an assessment of the patterns of semantic and phonological form retrieval during sign production informs not only sign language psycholinguistics, but speaks more broadly to the forces that shape the psycholinguistic and neurological organization of human language. While it is possible that lexical representations are entirely abstract and independent of their surface form, it is more likely that they are constrained by (or even inextricably linked to) the sensory and production systems in which they are most commonly encountered. These findings could thus offer insights about modality-free phonological universals, and/or modality-specific constraints on phonology in language.

In this paper we assess the relative order of semantic and phonological form retrieval in sign language production by adapting theory and practice from lexical retrieval studies of speech. We provide a brief review of the methodologies historically applied to the question of the temporal order of lexical concept and form retrieval in speech, introduce the basic structure of word forms in signed languages, and report the results of a series of experiments on the time course of lexical retrieval in American Sign Language (ASL). Through these experiments, we extend the findings of spoken language studies both to another language and to another language modality, and explore differential attributes of lexical-semantic and phonological form retrieval during lexical sign production.

Two methodologies have historically been applied to the question of the relationships between meaning and form retrieval in lexical selection – speech error analyses and picture naming manipulation. One common goal of these paradigms is to chart the relative order of semantic and phonological activation in word retrieval and production.

1.1. Speech error data

An analysis of speech error data (e.g., Fromkin, 1971; Garrett, 1984, 1990) reveals that naturally occurring word substitutions fall largely into two classes: (1) those that are semantically based (i.e., *finger* for *toe*) and (2) those that are form based (i.e., *envelope* for *elephant*) (Garrett, 1990, pg. 160), also called *malapropisms* (Fay and Cutler, 1977). Errors that are both semantically- and form-based (i.e., the substitution of *lobster* for *oyster*) are rare (but see Dell and Reich, 1981 for a discussion of the prevalence of mixed errors in speech). The processing glitch that results in a word substitution is presumed to take as input the aspect of a lexical representation that is active during a given processing window. Thus, the prevalence of errors that are clearly semantically-based *or* form-based, coupled with the scarcity of errors that are semantic+form-based, strongly suggests that representations of meaning and form are not simultaneously active during speech production, but are retrieved (and thus vulnerable to error) in two separate, independent steps.

Speech errors are not limited to whole words. A variety of sub-lexical errors are well-attested, including anticipations, perseverations, sound substitutions and deletions (e.g., Fromkin, 1971; Garrett, 1975). Several regularities are observed across sub-lexical errors. Typically, segmental class is preserved (e.g., consonants are substituted for consonants, and vowels for vowels), word onsets are more prone to error than rhymes, and consonant errors are more frequent than vowel errors. Additionally, place features are the most vulnerable to errors (Jaeger, 1992).

1.2. Sign error data

As with slips of the tongue, semantic- and form-based errors are attested in sign languages, while dual (semantic+phonological) errors are rare. For example, in a recent report of production errors in German Sign Language (Deutsche Gebärdensprache [DGS]), 38/40 sign-substitution errors are semantically based; only one is semantically+form-based (Hohenberger, Happ, and Leuninger, 2002). Sign language form-based errors manifest primarily as anticipations, perseverations, and harmony errors[1] that occur for whole words (n=25) and sub-lexical units (n=76).

Interestingly, there is marked disparity in the frequency with which the individual sub-lexical form parameters are subject to error, with hand configuration being most susceptible (82.5%), and place of articulation and movement being less vulnerable to error (8.8% each) (Hohenberger et al., 2002). These data are consistent with previous slip literature (Klima and Bellugi, 1979; Newkirk, Klima, Pederson, and Bellugi, 1980) in which 65% of form-based errors were limited to hand configuration errors, while 13% and 11% were attributed to location and movement, respectively. They are also consistent with data reported for sign language formational paraphasias, in which aphasic signers' handshape errors were much more common than location or movement errors (Corina, 2000; Corina, Poizner, Bellugi, Feinberg, Dowd, and O'Grady-Batch, 1992), and with data from sign language acquisition studies demonstrating that deaf, signing children acquire control over the phonological parameters of location and movement prior to mastering hand configurations (e.g., Conlin, Mirus, Mauk, and Meier, 2000; Knapp, 2000; Marentette and Mayberry, 2000; Siedlecki and Bonvillian, 1993).

In sum, lexical and sub-lexical errors are attested in both sign and speech, with sub-lexical errors exhibiting regular patterns (and frank asymmetries) in both modalities. However, while speech and sign error data are of enormous value in describing natural language production in real time, they are largely uncontrolled. The corpus of naturalistic speech error data is, by definition, limited to utterances that escape detection by the speaker prior to their articulation, and that moreover fail to correspond to the apparent (or a sensible) lexical target. The shortcomings of this corpus are evident: Researchers miss out on an unknown number of 'errors in progress' that are caught and corrected by the speaker (see Baars, Motley, and MacKay, 1975 for a discussion of a proposed *output editor*), and of the errors that do escape an output editor, only those that are so glaringly at odds with the narrative flow as to alert the interlocutor are analyzed. It is interesting to note that differences in the output editor have been claimed to play a role in the paucity of sign-based slips of the hand, relative to speech-based slips of the tongue (e.g., stranding errors). Hohenberger et al. (2002), for example, have suggested that the longer duration of articulation of a sign may allow for the editor to capture mispronunciations earlier. In addition, the availability of the major articulators to visuo-kinesthetic monitoring may affect the frequency with which errors escape detection prior to output.

1.3. Picture naming data

One experimental technique adapted by Schriefers, Meyer, and Levelt (1990) that provides additional insights into the lexical retrieval process is the picture naming interference paradigm (see also Glaser, 1992; Lupker, 1979; Glaser and Dunglehoff, 1984). In this paradigm, participants name pictures of common objects as quickly as possible while ignoring auditory or visual distractor stimuli. The distractor stimuli are typically presented at specified times during the naming window and are linguistically related to the picture's name in one of a variety of ways.

Specifically, distractors may be presented just prior to, simultaneously with, or just subsequent to the presentation of the target picture. In addition, the distractors typically are semantically related, phonologically related, or unrelated to the target word. By subtracting the length of time participants need to name pictures in the presence of unrelated distractors from their naming latencies in the presence of semantic and phonological distractors, it is possible to measure and compare naming latencies associated with each distractor type at several points in time during the naming process.

Over the past two decades, picture naming interference studies have yielded two major results: First, the retrieval of lexical concepts can be delayed relatively early, but not relatively late, in the picture naming process (see, for example, Glaser and Duengelhoff, 1984; Levelt, Schriefers, Vorberg, Meyer, Pechmann and Havinga, 1991; Schriefers, Meyer, and Levelt, 1990). This suggests that the activation of semantic representations may be temporally constrained during the lexical retrieval process, and is most evident at the beginning of the time-course of word production. Second, retrieval of the phonological code can be hastened by the presence of phonologically similar words (Lupker, 1982; Schriefers, Meyer, and Levelt, 1990; Starreveld and La Heij, 1995), syllables (Jeschniak, Schriefers, and Hantsch, 2000), morphemes (Zwitserlood, Bölte, and Dohmes, 2000), and phonemic segments (Meyer and Schiefers, 1991) both early and late in lexical retrieval.

That semantic properties of words are accessed relatively early in the naming process is uncontroversial, as are findings that the phonological form of words can be accessed relatively late in naming. That phonological form can also be accessed early in naming, however, is less clear. Although Levelt et al. (1991), and Schriefers, Meyer, and Levelt (1990) found no effects of early phonological interference in picture naming, others, such as Damian and Martin (1999); Jescheniak and Schriefers (2001); Jescheniak, Schriefers, and Hantsch (2003); Meyer and Schriefers (1991); and Starreveld (2000), have found evidence for substantial early phonological facilitation, and importantly, have begun to identify the conditions under which these early effects can be found.

In general, it appears that early phonological facilitation in the picture naming interference paradigm varies as a function of the design of the experiment, the nature of the stimuli, and the modality of the distractor (for a detailed account see Starreveld, 2000). For example, while orthographic distractors have been found to produce naming facilitation early in word retrieval (e.g., Rayner and Springer, 1986; Starreveld and La Heij, 1996; Damian and Martin, 1999), auditorily-presented distractors produce more variable naming-time effects. For example, while Damian and Martin (1999), Meyer and Schreifers (1991), and Staarveld (2000) do find early phonological facilitation for auditory distractors, Schriefers et al. (1990) do not.

Broadly, then, we may conclude that as measured by the picture naming interference paradigm, some aspects of a spoken word's conceptual representation are retrieved early in naming, while the particular phonological patterns associated with that concept are retrieved simultaneously with, or subsequent to, semantic retrieval.

Crucially, while picture-naming studies to date have covered a modest range of languages (e.g., English, Dutch, French, and German), all have been spoken languages. No published study has utilized the picture naming interference paradigm to assess the stages of lexical selection and form encoding in the service of sign language production. This is not surprising. It is only in the last several decades that linguists and psycholinguists have begun to lay bare the phonological structure of signed languages, and concurrently, the psychological and neurological processes that underlie their perception and production. It is an open question as to whether and how the modality in which a language is perceived (auditory/visual) and produced (vocal/manual) affects its linguistic structure, psychological representation, and neurological organization.

1.4. Sign language structure

Two popular and parallel misconceptions, even among educated psychologists and linguists, are that (1) sign languages are artificial communication systems constructed *in toto* by hearing people for the deaf, and thus are composed of manual versions of the dominant culture's spoken words, or (2) sign language signs are elaborate but originally iconic gestures that have become conventionalized with use, but have not lost their transparency. In fact, the thousands of signed languages formerly and currently in existence across the world are naturally-occurring, complex languages with full semantic, syntactic, morphological, and phonological systems (e.g., Klima and Bellugi, 1979; Liddell and Johnson, 1989; Stokoe, Casterline, and Croneberg, 1976) whose signs share a conventionalized but largely arbitrary relationship with the objects they signify.

Although it is easy to understand that individual signs have associated lexical concepts, less obvious is that signs also have phonological form analogous to words. Each sign in a signed language can be described in terms of four major articulatory parameters: the configuration of the fingers on the dominant and non-dominant hands during the production of the sign (e.g., the /5/ handshape), the orientation of the hands in space (e.g., /upright/), the location of the sign on the body or in the space immediately around the body (e.g., /chin/), and the movement pattern associated with the sign (e.g., /twisting/). Each of the four major parameters is characterized by a closed set of linguistically meaningful elements, commonly (and largely interchangeably) referred to by sign language linguists as *primes*, *cheremes* (Stokoe, 1960; Stokoe, Casterline, and Croneberg 1976), or *phonemes*. Parameter inventories vary widely across signed languages largely as a function of historical relatedness, rendering less-related languages mutually unintelligible (e.g., McKee and Kennedy, 2000; Woodward, 1996, 2000).

One issue of sign phonology that is most relevant to this study is the marked difference in the temporal unfolding of phonological units in sign compared to speech. Sign structure has often been described as exhibiting far more structural simultaneity than has been attributed to spoken languages. That is, while a spoken word unfolds over time, exposing a great deal of phonological information about individual segments as they are sequentially produced (and perceived), signs appear as concurrent instantiations of values of each parameter (a simultaneously-produced hand configuration, place of articulation, and movement). Thus, a great deal of phonological information is available quite early in a sign's production and is maintained throughout the duration of the sign (e.g., Corina, 1993; Meier, 1993; Sandler, 1993). These differences have documented processing consequences (Emmorey and Corina, 1990) that further motivate a careful investigation of the temporal properties underlying sign production in the context of picture naming.

2. Method

We conducted a sign language lexical retrieval study using the picture naming interference paradigm. Deaf, native-signing Gallaudet University undergraduate students were asked to name pictures of objects as quickly as possible while ignoring distractor signs presented at one of three temporal offsets with the objects: –130 ms, 0 ms, and +130 ms. These stimulus onset asynchronies (SOAs) were calculated by adjusting the SOAs reported in Schriefers, Meyer, and Levelt (1990) (–150 ms, 0 ms, +150 ms) to accommodate the relatively faster recognition of signs than spoken words reported in a previous study (Emmorey and Corina, 1990).[2]

2.1. Participants

Twenty-eight Gallaudet University undergraduate students (13 female) ranging in age from 17–35 years old (M_{age} = 21 years +/– 4 yr) participated in this study. All were native signers who were initially exposed to sign naturally in their homes by deaf parents and siblings and subsequently attended residential schools for the deaf, where ASL was the primary form of communication. Each participant was paid 10 dollars. Due to unforeseen circumstances, subjects were not assigned equally across conditions: Seven participated in the –130 ms condition, twelve in the 0 ms condition, and fifteen in the +130 ms condition. Six of the –130 condition subjects also participated in the +130 ms condition.

2.2. Materials

Target signs were elicited with 61 still video images of common, everyday objects. Interfering stimuli (IS) consisted of video images of 169 common ASL signs produced in real time by a deaf research assistant who is a native, fluent signer of ASL. These interfering signs were chosen because they are common in the ASL lexicon, and because they shared a semantic or phonological relationship with at least one target object (e.g., cheese/CRACKER [a semantic pairing]; apple/KEY [a phonological pairing in which two parameters, hand configuration and movement, are shared]). These signs were videotaped, then digitized, edited, and normalized to 1000 ms with Media 100 digital video system (v. 6.0). Objects and signs were digitally overlaid, with objects appearing at full strength and interfering signs appearing at 60% transparency. The result of this composition is of a full colored object appearing with a highly visible but semi-transparent signer (Figure 1).

cat/PIG cat/STORY cat/HAMMER

Figure 1. Examples of Target/Interfering Stimulus (IS) pairs using the target CAT. The subject must name the target while ignoring the IS. (left). IS (PIG) is semantically related to CAT (both animals); (middle). IS (STORY) is phonologically related (same hand configuration); (right). IS (HAMMER) is semantically and phonologically unrelated.

All stimuli were edited carefully so that a succinct and accurate representation of the intended sign was immediately available. The beginning of a sign was defined as three frames prior to the beginning of a contact or path movement. For example, in a contacting sign like THINK, in which the index finger approaches and touches the forehead, we edited the onset of the sign to 3 frames prior to the touch of the forehead, thus capturing the approach to contact. In a zero-space sign such as TRIP, in which a bent-V handshape makes a circular path movement about a vertical plane, we edited the onset of the sign to begin three frames prior to the beginning of the first circular path movement.

2.3. Procedures

Trials were run using Psyscope version 1.2.5 on a Macintosh G4 Powerbook running OS 8.6. Throughout the duration of the experiment, participant responses were videotaped and later coded for accuracy. Participants sat in front of a computer monitor with their wrists resting on a plexiglass platform and their fingers resting lightly on the computer keyboard. They were instructed to name the objects that appeared on the screen as quickly and accurately as possible, while ignoring the concurrently-appearing signer.

Individual trials were begun by the participant pressing the computer key pad. The stimulus image remained on the screen until the participant raised her hands to sign the name of the object, at which point an infrared beam running between the bracketed ends of the platform was tripped by the hands crossing the beam. Responses were recorded by Psyscope, and response times were calculated as the difference between the onset of the object stimulus and the time of the beam trip (stimulus onset asynchrony [SOA]). After signing the object name, participants quickly pressed a Psyscope response-box button to mark the end of a trial, and returned their hands to the resting position.

Target objects were presented six times per SOA: once alone, and once each in the presence of a 60% transparent interfering sign that was phonologically, semantically, phonologically+semantically, or not related to the object sign. Semantic category members and semantic associates were collapsed for this analysis. The linguistic relationships between target and distractor pairs were confirmed by the first author, who is a fluent ASL signer, and a native signing research assistant.

Semantically related pairs were judged as being target sign/interfering pairs that were either semantic associates (e.g., apple/TREE) or members of a semantic category (e.g., bed/DRESSER). Phonologically related signs were those that shared one to three formational parameters (handshape, place of articulation, or movement) with the intended target. For example, the target sign FLOWER and interfering sign NUMBER share values for hand configuration; the target sign NEWSPAPER and interfering sign BIRD share values for handshape and movement but differ in place of articulation; and the target sign CHEESE and interfering sign MOVIE share the parameters of handshape, place of articulation, and movement, but differ in orientation. This fourth phonological parameter, orientation, was not systematically varied in this study for the purpose of assessing its effects on lexical retrieval speed, because its effects on phonological processing are much less well-established than the other parameters. Phonologically+semantically related pairs (e.g., bike/CAR) constitute somewhat unusual lexical forms and are not considered in present manuscript. Unrelated items were pairings that shared no semantic or form relation to one another.

3. Results

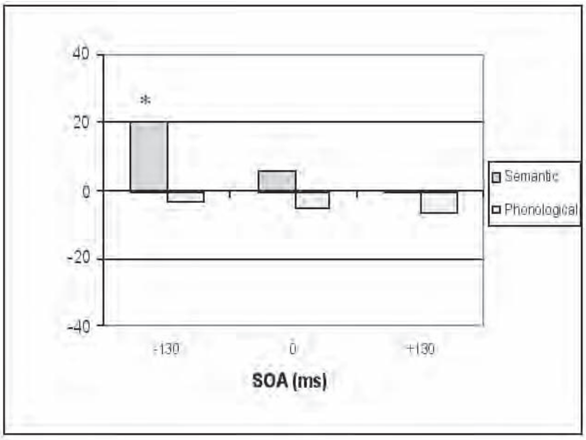


Figure 2. Mean naming latency differences, by linguistic condition and stimulus onset asynchrony (SOA), relative to naming time in the presence of unrelated interfering stimuli (IS). “Semantic” bars represent target naming with semantically related distractors. “Phonological” bars represent target naming with phonologically related distractors.

The effect of distractor type on object naming is typically computed as the difference between mean picture naming latency in the presence of linguistically related and unrelated distractors. The unrelated distractors serve as an interference baseline condition against which the other IS types can be evaluated. As the number of participants differed between SOA levels, we evaluated each SOA level independently. For the data reported in sections 3.1 and 3.2, we tested whether the pairwise differences between conditions (semantic-unrelated and phonological-unrelated) were each statistically different from zero. Two pairwise comparisons were made at each SOA level.

Looking at latency patterns alone (Figure 2), semantic distractors slowed naming early in the lexical retrieval process relative to their effect later in the process, while phonological distractors exhibited the opposite pattern. Linguistic distractors thus appear to have different effects when presented early and late in lexical retrieval.

3.1. Early lexical retrieval (–130 ms SOA)

Specifically, when the interfering stimuli (IS) appeared 130 ms prior to the onset of the target picture, naming was slowed for semantically related IS, relative to the unrelated condition. The mean difference between naming with a semantic IS vs. naming with an unrelated IS was 20.49 ms +/– 8.0 ms (s.e.), $t_6 = 2.54$, $p = .044$. Phonologically related IS had quite variable effects on naming at this SOA. The difference was not statistically different from zero ($M_{\text{difference}} = -3.22$ +/– 7.09 ms).

3.2. Late lexical retrieval (0 ms SOA and +130 ms SOA)

When the distractor sign appeared simultaneously with the picture's onset or 130 ms after the picture's onset, mean naming time differences between the semantic and unrelated conditions did not differ significantly from zero (0 ms SOA: $M_{\text{difference}} = 6.06$ +/– 9.53; +130 SOA: $M_{\text{difference}} = -.64$ +/– 8.25 ms). Similarly, the effects of phonologically related distractors on picture naming latency at these SOAs (relative to the unrelated baseline condition) appeared moderately facilitatory, although this facilitation was not statistically significant (0 ms SOA: $M_{\text{difference}} = -4.92$ +/– 11.20 ms; +130 SOA: $M_{\text{difference}} = -6.52$ +/– 7.94 ms).

3.3. Phonological patterns: Number of shared parameters

In ASL and other signed languages, two signs are considered to be phonologically related if they share a hand configuration value, a place of articulation value, a movement value, or any combination of the three. While conducting the phonological analysis, we observed that effects of phonologically related IS on picture naming varied with the strength of the phonological relationship between the object name and the IS (Figure 3): both the direction and magnitude of the effect of phonologically related IS on naming latency change with the number of phonological parameters shared by the target sign and IS. Specifically, we observed different amounts of interference between IS that shared one, two, and three parameters with the target. We conducted post-hoc analyses to further explore these observations. Three contrasts were computed at each SOA level: one parameter vs. unrelated, two parameters vs. unrelated, and three parameters vs. unrelated.[3]

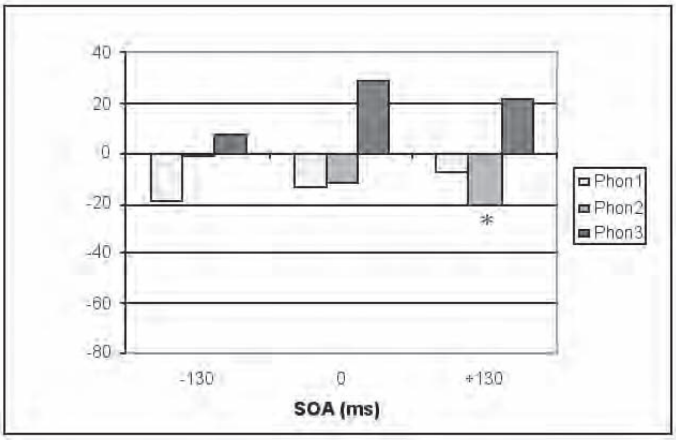


Figure 3. Mean naming latency differences for phonologically related IS, by degree of relatedness, relative to naming time for unrelated IS. Phon1, Phon2, and Phon3 bars represent target naming with IS that share one, two, and three phonological parameters, respectively, with the target.

3.3.1. One parameter

IS that share one phonological parameter with object names show a consistent but non-significant pattern of phonological facilitation across SOAs, relative to naming in the unrelated condition. Mean differences for –130 ms, 0 ms, and +130 ms were –18.60 +/– 16.53 ms; –12.60 +/–24.13 ms; and –6.85 +/– 16.52 ms, respectively; all p's > .05.

3.3.2. Two parameters

IS that share two phonological parameters with object names show no demonstrable effect on naming time when presented prior to or simultaneous with objects, relative to the unrelated baseline condition (–130 ms: $M_{difference}$ = –.89 +/– 9.26 ms; 0 ms: $M_{difference}$ = –11.49 +/– 13.83, p's > .05). However, when presented 130 ms after the object's onset, the IS resulted in faster object naming than did unrelated IS ($M_{difference}$ = –20.14 +/– 9.27 ms, t_{14} = 2.17, p = .048, Bonferroni uncorrected).

3.3.3. Three parameters

IS that share three phonological parameters with object names appear to interfere with naming across SOAs ($M_{differences}$ = 8.18 +/– 18.61 ms; 29.57 +/– 26.34 ms; and 22.27 +/– 16.37 ms, respectively). However, no difference in naming resulting from the presence of three-parameter distractors (relative to unrelated ones) is significantly different from zero.

3.4. Phonological patterns: Category of shared parameters

In order to better understand the interactions between these phonologically related target/sign pairs, we additionally quantified interference effects by the exact nature of the shared phonological relationship (Figure 4). This was possible only for pairs that had sufficient numbers for statistical analysis, and consequently was limited to signs and IS pairs overlapping in two parameters (handshape and location, handshape and movement, or movement and location). Three contrasts were computed at each SOA level: Handshape+location vs. unrelated, handshape+movement vs. unrelated, and movement+location vs. unrelated.[4]

Effects of IS sharing handshape+movement or handshape+location with the target were small and statistically nonsignificant. Specifically, mean naming time differences for IS sharing handshape+movement with the target were

21.10 +/– 23.58 ms, –5.69 +/– 20.07 ms, and 10.60 +/– 13.54 ms, relative to the unrelated condition, for SOAs of –130 ms, 0 ms, and +130 ms, respectively. Mean naming time differences for IS sharing handshape+location with the target were –4.86 +/– 11.51 ms, 0.41 +/– 21.38 ms, and –23.33 +/– 15.01 ms, relative to unrelated IS, for SOAs of –130 ms, 0 ms, and +130 ms, respectively. All p-values were > .05.

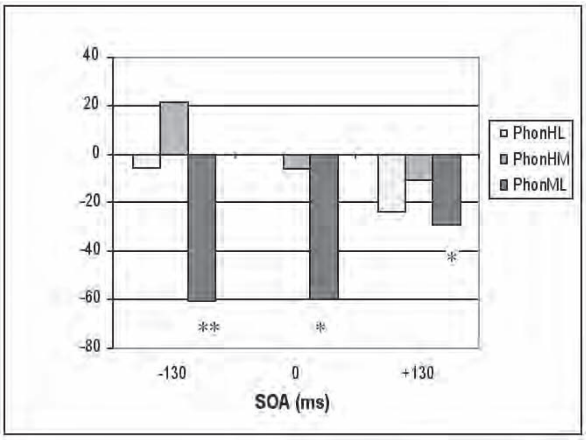


Figure 4. Mean naming latency differences as a function of the phonological parameters shared by target objects and IS, by SOA relative to naming time in the presence of unrelated IS. PhonHL, PhonHM, and PhonML bars represent target naming time for IS that share hand configuration+location, hand configuration+movement, and movement+location with the target, respectively.

However, the parameter combination movement+location (ML) facilitated naming at all stimulus onset asynchrony levels. ML overlap between target and IS resulted in faster naming at each SOA level. Specifically, (1) when ML distractors preceded the target by 130 ms, object naming was 60.73 +/– 12.32 ms faster than when an unrelated distractor preceded the target (t_6 = –4.84; p = .003); (2) when ML distractors occurred simultaneously with the target, naming was 59.63 +/– 27.67 ms faster than when an unrelated IS preceded the target (t_{11} = –2.12; p = .05, Bonferroni uncorrected); (3) when ML distractors appeared 130 ms after the target, object naming was 28.42 +/– 10.14 ms faster than when an unrelated distractor preceded the target (t_{14} = –2.80; p = .015).

4. Discussion

By adapting the well-known picture naming interference paradigm for use with sign language stimuli, we have demonstrated that the semantic and phonological processes involved in sign lexical retrieval are differentiable, and may adhere to the interference patterns that have been used to argue for seriality of semantic and phonological processing in spoken language production (e.g., Schreifers, Meyer, and Levelt, 1990). In addition, we have shown some of the first evidence for the importance of phonological structure in the online mediation of word production in a signed language.

As is consistent with the spoken language literature, we find evidence that lexical-semantic representations of signs are active early in the naming process, that these representations are susceptible to disruption by prior activation of semantically-related signs, and that this disruption has a detrimental effect on naming latency. Importantly, we observe that such interference may be obtained when the IS items themselves are within-modality primary linguistic representations (i.e., signs, rather than written words).

Phonological interference, however, presents a more intriguing story: while at first-glance, phonological interference effects appear to be absent or negligible at all SOA levels, a closer examination of these data reveals patterns of facilitation that provide insights into the phonological representations underlying sign production that are worthy of further exploration.

First, we see much larger interference effects when signs and targets share three phonological parameters compared to when they share only one or two parameters. Interestingly, the degree of phonological overlap modulated the valence of the interference effects, such that three-parameter distractors slowed naming, while one- or two-parameter distractors facilitated naming. Second, we see that among distractor/target pairs sharing two phonological parameters, the nature of the shared parameters heavily influences the direction and magnitude of interference. Specifically, distractors that share movement and location with targets speed picture naming at all SOA levels, whereas those that share handshape and location or handshape and movement have smaller and more variable influence.

We look first at the issue of degree of phonological overlap between distractor and target signs. In cases in which target signs and interfering signs shared one or two parameters, we observed non-significant naming facilitation at –130 ms and 0 ms SOAs. However, significant phonological facilitation effects were present at the +130 ms SOA level. This pattern of results appears consistent with the view of seriality of lexical retrieval processes originally described in Schreifers, Meyer, and Levelt (1990).

There are at least two possible explanations for the observed patterns of interference in the cases of interfering signs and targets sharing three phonological parameters. First, it may be that under the constraints of the experiment, participants mistook these highly-related but non-identical interfering signs for the intended targets, leading to a momentary re-evaluation of the prepared response. Alternatively, these IS may have induced lexical neighborhood effects, such that the recognition of the interfering sign (a close phonological neighbor of the target sign) may have actively inhibited the activation of the target (Elman and McClelland, 1988; Goldinger, Luce, and Pisoni, 1989).

Importantly, the differential interference patterns that we see from interfering signs that share one, two, and three formational parameter values with the target sign are suggestive of sub-lexical processing effects in sign picture naming. This suggestion of sub-lexical effects motivated a post-hoc analysis of the observed phonological facilitations.

We now turn to a closer examination of the particular patterns of observed phonological interference. Recall that each individual sign in a signed language is composed of values from each of four possible phonological parameters – a specification in the domains of hand configuration, place of articulation, movement, and hand orientation. A breakdown of phonological interference effects along these parameter lines (excluding orientation, which was not systematically manipulated in the present study) reveals striking differences between interference patterns promoted by shared handshape + location and handshape + movement, on the one hand, and shared location + movement, on the other. Specifically, phonological facilitation effects observed in the overall analysis are, in fact, being carried almost entirely by the combination of movement and location specifications. No other combinations contribute significantly to this pattern. Moreover, the robust facilitation driven by the combination of shared movement and location is seen at all SOA levels. In contrast to our findings based upon the analysis of absolute number of parameters shared between target and distractor (i.e., one, two, or three), these new data present a challenge to Levelt and colleagues' original conceptualization of lexical retrieval as a strict serial process.

While studies comparing the effects of auditory versus written word interference in picture naming show effects of early shared phonological overlap in both modalities, it is generally agreed that conditions of within-modality interference (i.e., visual-visual) are more apt to lead to early phonological effects[5]. This has been attributed to the presence of entire orthographic word-forms throughout the naming task, as opposed to the seriality that is required for the unfolding of acoustically presented speech targets. In this respect, the American Sign Language sign signals may be thought of as holding an intermediate status between speech and orthographic signals. On the one hand, signs are articulated over time, like spoken words. On the other hand, signs maintain their compositional structure of bundled, simultaneously-articulated parameter values throughout this articulatory unfolding (Corina, 1993; Corina and Sandler, 1993; Klima and Bellugi, 1979; Sandler, 1993). This well-known characteristic of sign structure has documented effects on sign recognition, such that proportionally less time is required to recognize a sign (34%) than a spoken word (83%) (Emmorey and Corina, 1990), and may underlie the phonological interference effects that we observe throughout the time-course of picture naming.

It is interesting that linguistic theories of the structure of signed languages provide arguments that particular combinations of parameters may be privileged in the composition of a sign. Notably, segmental theories of ASL structure have argued that the abstract properties of movements and locations may serve as the building blocks of the ASL syllable, while relegating handshape to a separate status. In this literature (e.g., Brentari, 1998; Corina and Emmorey, 1993; Corina, 1993; Sandler, 1986) the existence of phonological processes and particulars of word formation motivate the isolation of these units, and thus lead to coherent descriptions of sign language phonological structure.

Differential processing effects of the formational parameters of sign on word production are consistent with a growing body of literature revealing location and movement as more resilient to disruption than hand configuration (e.g., Bellugi and Siple, 1974; Corina, 1998; Corina, Poizner, Bellugi, Feinberg, Dowd, and O'Grady-Batch, 1992; Hohenberger, Happ, and Leuninger, 2002; Klima and Bellugi, 1979; Tartter and Fischer, 1982). They are further bolstered by a recent study investigating perceptual similarity judgments in American Sign Language that draws parallels between sign similarity and rhyming in speech (Hildebrandt and Corina, 2002). In this study, participants viewed a running display of five phonologically possible non-signs: a central target surrounded by four flankers that were related to the target by combinations of shared parameters (handshape + movement, handshape + location, movement + location, or unrelated). When asked to determine which of the flankers was perceived as most similar to the target, native deaf signers overwhelmingly preferred the combination of movement and location as highly similar, but gave little weight to shared hand configurations.

In the present picture-naming study, we have documented differential and significant effects of lexical-semantic and form-based phonological interference on sign production in native signers of American Sign Language. As is seen in studies of spoken languages, robust but transient semantic interference effects are detectable relatively early during the sign naming process. This pattern contrasts with the presence of facilitatory sign phonological effects that are seen throughout the sign naming process, which are more robust and consistent than has been reported for phonological effects on spoken word naming. This fact is a confirmation that the processes underlying word retrieval and production are highly consistent across languages, regardless of a language's natural perceptual and production channels. These psycholinguistic facts are consistent with functional neuroanatomical studies that have shown great similarity between neural regions underlying higher level aspects of lexical retrieval and production in signed and spoken languages (Corina, McBurney, Dodrill, Hinshaw, Brinkley, and Ojemann,1999; Corina, San Jose-Robertson, Guillemin, High, and Braun, 2003; Emmorey, Grabowski, McCullough, Damasio, and Bellugi, 2003; MacSweeney et al., 2002; Petitto, Zatorre, Gauna, Nikelski, Dostie, and Evans, 2000). In addition, these data emphasize the importance of sub-lexical structures in sign formation, and particularly the combination of sub-lexical location and movement, which may hold a special status in the mediation of sign language processing in native Deaf signers.

Notes

1. Hohenberger, Happ, and Leuninger (2002) define harmony errors as those that have two possible sources, preceding the error or following the error. Thus, it is impossible to determine whether the error is anticipation or a perseveration.
2. Specifically, in a psycholinguistic gating study by Emmorey and Corina (1990), average sign duration was found to be approximately 703 ms, with isolation occurring within 239 ms (i.e., 34% of the sign). In contrast, in data reported by Grosjean (1980) in which average word duration was measured at 400 ms, word isolation occurred within 330 ms (83% of the word). See also Emmorey (2002) for further discussion. In the current experiments, these timing differences between speech and sign were taken into account during our calculations of SOA. In the present case we assumed average word and sign lengths to be 250 ms and 500 ms, respectively. Based on the previous findings that words can be isolated

within 83% of the processing time for the form (208 ms) and signs within 34% (170 ms), the SOA used by Schriefers, Meyer, and Levelt (1990) (150 ms) represents 72% of the time required to isolate the word. Thus, adopting this same percentage results in a sign recognition offset of 122 ms, or roughly 4 video frames at 33 ms/frame. The smallest divisible unit that digital video permits results in an SOA of 130 ms.

3. Note that for these multiple post-hoc tests, the Bonferroni-corrected alpha level at or below which differences should be considered statistically significantly different from zero is p = .02.
4. As in Section 3.3, for these multiple post-hoc tests, the Bonferroni-corrected alpha level at or below which differences should be considered statistically significantly different from zero is p = .02.
5. A reviewer questioned whether the observed patterns of interference in this study might be due not only to the linguistic relationship between targets and distractors, but to some amount of physical similarity between distractor signs and the objects that they name. While a small number of ASL signs in the lexicon bear a physical resemblance to the items they signify, this iconicity has been shown to have little consequence for ASL acquisition or production. For example, data from language acquisition (Orlansky and Bonvillian, 1984), neuroimaging (Emmorey, Grabowski, McCullough, Damasio, Ponto, Hichwa, and Bellugi, 2004), and language breakdown (Corina, Poizner, Bellugi, Feinberg, Dowd, and O'Grady-Batch, 1992) suggest that iconicity does not drive sign language mediation in native users of ASL. In our study, a liberal coding of all stimuli for iconic properties (i.e., signs bearing some resemblance to either the named object [CIGAR] or to the action used when manipulating the object [SPOON]) yields an iconicity proportion of approximately 30%. However, this rate is highly consistent across semantic, phonological, and unrelated conditions. Any effects of iconicity would thus be factored out of our data, which are comparisons to a baseline unrelated condition. Therefore, within the constraints of the current paradigm it is unlikely that we are seeing effects of iconicity on naming time. However, this is an interesting issue that is worthy of further study.

References

Baars, B. J., M. T. Motley, and D. G. MacKay
1975 Output editing for lexical status from artificially elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behavior* 14: 382–391.

Bellugi, U., and P. Siple
1974 Remembering with and without words. In F. Bresson (ed.), *Current Problems in Psycholinguistics.* Paris: Cetnre National de la Recgercge Scientifique.

Brentari, D. A.
1998 *Prosodic model of sign language phonology.* Cambridge (MA): MIT Press.

Channon, R.
2002 Beads on a string? Representations of repetition in spoken and signed languages. In R. P. Meier, K. Cormier, and D. Quinto-Pozos (eds.), *Modality and Structure in Signed and Spoken Languages.* Cambridge: Cambridge University Press.

Conlin, K. E., G. Mirus, C. Mauk, and R. P. Meier
2000 Acquisition of first signs: Place, handshape, and movement. In Chamberlain et al. (eds.), *Language Acquisition by Eye*, 51–69. Mahwah, NJ: Erlbaum.

Corina, D. P.
1993 To branch or not to branch: Underspecification in ASL handshape contours. In G. R. Coulter (ed.), *Phonetics and Phonology: Volume 3, Current Issues in ASL Phonology*, 63–94. San Diego, CA: Academic Press.
1998 The processing of sign language: evidence from aphasia. In B. Stemmer and H. A. Whitaker, (eds), *Handbook of neurolinguistics*, 313–329. New York: Academic Press.
2000 Some observations regarding paraphasia and American Sign Language. In K. Emmorey and H. Lane (eds), *The Signs of Language Revisited: An Anthology to Honor Ursula Bellugi and Edward Klima*, 493–507. Mahwah, NJ: Lawrence Erlbaum Associates.

Corina, D. P., and K. Emmorey
1993 Lexical priming in American Sign Language. Paper presented at the Linguistic Society of American Conference, Philadelphia, PA.

Corina, D. P., S. L. McBurney, C. Dodrill, K. Hinshaw, J. Brinkley, and G. Ojemann
1999 Functional roles of Broca‘s area and supramarginal gyrus: evidence from cortical stimulation mapping in a deaf signer. *Neuroimage* 10: 570–581.

Corina, D. P., H. Poizner, U. Bellugi, T. Feinberg, D. Dowd, and L. O‘Grady-Batch
1992 Dissociation between linguistic and nonlinguistic gestural systems: A case for compositionality. *Brain and Language* 43(3): 414–447.

Corina, D. P., L. San Jose-Robertson, A. Guillemin, J. High, and A. R. Braun
2003 Language lateralization in a bimanual language. *Journal of Cognitive Neuroscience* 15(5): 718–730.

Corina, D., and W. Sandler
1993 On the nature of phonological structure in sign language. *Phonology* 10: 165–207.
Damian, M. F., and R. C. Martin
1999 Semantic and phonological codes interact in single word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 25: 345–361.
Dell, G.
1986 A spreading-activation theory of retrieval in sentence production. *Psychological Review* 93: 283–321.
1988 The retrieval of phonological forms in production: Tests of predictions from a connectionist model. *Journal of Memory and Language* 27: 124–142.
Dell, G., and P. A. Reich
1981 Stages in sentence production: An analysis of speech error data. *Journal of Verbal Learning and Verbal Behavior* 20: 611–629.
Dye, M., and S.-I Shih
this vol. Phonological priming in British Sign Language.
Elman, Jeffrey L., and James L. McClelland
1988 Cognitive penetration of the mechanisms of perception: Compensation for coarticulation of lexically restored phonemes. *Journal of Memory and Language* 27(2): 143–165.
Emmorey, K.
2002 *Language, cognition, and the brain: Insights from sign language research.* Lawrence Erlbaum and Associates: Mahwah, NJ.
Emmorey, K., and D. Corina
1990 Lexical recognition in sign language: Effects of phonetic structure and morphology. *Perceptual and Motor Skills* 71(3): 1227–1252.
Emmorey, K., T. Grabowski, S. McCullough, H. Damasio, L. L. Ponto, R. D. Hichwa, and U. Bellugi
2003 Neural systems underlying lexical retrieval for sign language. *Neuropsychologia* 41(1): 85–95.
2004 Motor-iconicity of sign language does not alter the neural systems underlying tool and action naming. *Brain and Language* 89(1): 27–37.
Fay, D., and A. Cutler
1977 Malapropisms and the structure of the mental lexicon. *Linguistic Inquiry* 8: 505–520.
Fromkin, V. A.
1971 The non-anomalous nature of anomalous utterances. *Language* 47: 27–52.

Garrett, M. F.
1975 The analysis of sentence production. In G. Bower (ed.), *Psychology of learning and motivation*, 33–177. New York: Academic Press.
1976 Syntactic processes in sentence production. In R. Wales and E. Walker, (eds.), *New Approaches to Language Mechanisms.* Amsterdam: NorthHolland.
1984 The organization of processing structure for language production: Applications to aphasic speech. In D. Caplan, A. R. Lecours and S. Smith (eds.), *Biological perspectives on language.* Cambridge, MA: MIT Press.
1990 Sentence processing. In D. Osherson (ed.), *Invitation to Cognitive Science*, 133–175. Cambridge, MA: MIT Press.
Goldinger, S. D., P. A. Luce, and D. B. Pisoni
1989 Priming lexical neighbors of spoken words: Effects of competition and inhibition. *Journal of Memory and Language* 28(5): 501–518.
Glaser, W. R.
1992 Picture naming. *Cognition* 42: 61–105.
Glaser, W. R., and F. J. Düngelhoff
1984 The time course of picture-word interference. *Journal of Experimental Psychology: Human Perception and Performance* 10(5): 640–654.
Grosjean, F.
1980 Spoken word recognition processes and the gating paradigm. *Perception and Psychophysics* 28(4): 267–283.
Hildebrandt, U. C., and D. P. Corina
2002 Phonological similarity in American Sign Language. *Language and Cognitive Processes* 17(6): 593–612.
Hohenberger, A., D. Happ, and H. Leuninger
2002 Modality-dependent aspects of sign language production: Evidence from slips of the hands and their repairs in German Sign Language. In R. P. Meier, K. Cormier, and D. Quinto-Pozos (eds.), *Modality and structure in signed and spoken language.* Cambridge: Cambridge University Press.
Jaeger, Jeri
1992 Phonetic features in young children's slips of the tongue. *Language and Speech* 35(1–2): 189–205.
Jescheniak, J., H. Schriefers, and A. Hantsch
2003 Utterance format affects phonological priming in the picture-word task: Implications for models of phonological encoding in speech production. *Journal of Experimental Psychology: Human Perception and Performance* 29(2): 441–454.

Jescheniak, J. D., and H. J. Schriefers
2001 Priming effects from phonologically related distractors in picture word interference. *Quarterly Journal of Experimental Psychology* 54(2): 371–382.
Klima, E., and U. Bellugi
1979 *The Signs of Language.* Cambridge, MA: Harvard University Press.
Knapp, H. P.
2000 Acquisition of hand configuration and place of articulation in American Sign Language. Unpublished master's thesis, University of Texas at Austin, Austin, TX.
Levelt, W. J. M.
1989 Models of word production. *Trends in Cognitive Sciences* 3: 223–232.
Levelt, W. J. M., A. Roelofs, and A. S. Meyer
1999 A theory of lexical access in speech production. *Behavioral and Brain Sciences* 22: 1–38.
Levelt, W., H. Schriefers, D. Vorberg, A. S. Meyer, T. Pechmann, and J. Havinga
1991 The time course of lexical access in speech production: A study of picture naming. *Psychological Review* 98: 122–142.
Liddell, S. K., and R. E. Johnson
1989 American Sign Language: the phonological base. *Sign Language Studies* 64: 195–277.
Lupker, S. J.
1979 The semantic nature of response competition in the picture-word interference task. *Memory & Cognition* 7: 485–495.
1982 The role of phonetic and orthographic similarity in picture-word interference. *Canadian Journal of Psychology* 36(3): 349–367.
MacSweeney, M., B. Woll, R. Campbell, P. K. McGuire, A. S. David, S. C. Williams, J. Suckling, G. A. Calvert, and M. J. Brammer
2002 Neural systems underlying British Sign Language and audio-visual English processing in native users. *Brain* 125(7): 1583–1593.
Marentette, P. F., and R. Mayberry
2000 Principles for an emerging phonological system: A case study of acquisition of early ASL. In Chamberlain et al. (eds.), *Language Acquisition by Eye*, 71–90. Mahwah, NJ: Erlbaum.
McKee, D., and G. Kennedy
2000 Lexical comparison of signs from American, Australian, British, and New Zealand sign languages. In Emmorey, K. and Lane, H. (eds), *The Signs of Language Revisited: An Anthology to Honor Ursula Bellugi and Edward Klima.* Mahway, NJ: Lawrence Erlbaum Associates, Inc.

Meier, R. P.
1993 A psycholinguistic perspective on phonological segmentation in sign and speech. In G. R. Coulter (ed.), *Phonetics and Phonology, Vol. 3: Current Issues in ASL Phonology*, 169–188. San Diego, CA: Academic Press.

Meyer, A. S., and H. Schriefers
1991 Phonological facilitation in picture-word interference experiments: Effects of stimulus onset asynchrony and types of interfering stimuli. *Journal of Experimental Psychology: Learning, Memory, & Cognition* 17: 1146–1160.

Morton, J.
1969 Interaction of information in word recognition. *Psychological Review* 76: 165–178.

Newkirk, D., E. S. Klima, C. C. Pedersen, and U. Bellugi
1980 Linguistic Evidence from Slips of the Hand. In V. Fromkin (ed.). *Errors in Linguistic Performance: Slips of the Tongue, Ear, Pen and Hand*, 165–198. New York: Academic Press.

Neville, H.J., D. Bavelier, D. Corina, J. Rauschecker, A. Karni, A. Lalwani, A. Braun, V. Clark, P. Jezzard, and R. Turner
1998 Cerebral organization for language in deaf and hearing subjects: Biological constraints and effects of experience. *Proceedings of the National Academy of Sciences, USA* 95(3): 922–929.

Orlansky, M. D., and J. D. Bonvillian
1984 The role of iconicity in early sign language acquisition. *Journal of Speech & Hearing Disorders* 49(3): 287–292.

Petitto, L.A., R. J. Zatorre, K. Gauna, E. J. Nikelski, D. Dostie, and A. C. Evans
2000 Speech–like cerebral activity in profoundly deaf people processing signed languages: implications for the neural basis of human language. *Proceedings of the National Acadamy of Sciences, USA* 97(25): 13961–13966.

Rayner, K., and C. J. Springer
1986 Graphemic and semantic similarity effects in the picture-word interference task. *British Journal of Psychology*, 77(2): 207–222.

Sandler, W.
1986 The spreading hand autosegment of ASL. *Sign Language Studies* 50, 1–28.
1993 Linearization of phonological tiers in ASL. In G. R. Coulter (ed.), *Phonetics and Phonology: Current Issues in ASL Phonology*, 103–127). San Diego, CA: Academic Press.

de Saussure, F.
1983 *Course in General Linguistics* London: Duckworth. [Tranlations by R. Harris. Original work published from notes of C. Bally and A. Sechehaye, Lectures delivered at the University of Geneva, 1906–1911.]
Schriefers, H., A. S. Meyer, and W. J. M. Levelt
1990 Exploring the time course of lexical access in language production: Picture-word interference studies. *Journal of Memory and Language* 29: 86–102.
Siedlecki, T., and J. D. Bonvillian
1993 Location, handshape, and movement: Young children's acquisition of the formational aspects of American Sign Language. *Sign Language Studies* 78: 31–52.
Starreveld, P. A.
2000 On the interpretation of auditory context effects in word production. *Journal of Memory and Language* 2(4): 497–525.
Starreveld, P. A., and W. La Heij
1995 Semantic interference, orthographic facilitation and their interaction in naming tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 21(3): 686–698.
1996 Time-course analysis of semantic and orthographic context effects in picture naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22(4): 896–918.
Stokoe, W. C.
1960 *Sign Language Structure* (2nd edition 1978). Silver Spring, MD: Linstock Press.
Stokoe, W., D. C. Casterline, and C. G. Croneberg
1976 *A Dictionary of American Sign Language on Linguistic Principles*. Washington, D.C.: Galluadet University Press.
Tartter, V., and S. Fischer
1982 Perceiving minimal distinctions in ASL under normal and point-light display conditions. *Perception and Psychophysics* 32: 327–334.
Wilbur, R.
1998 Syllables and segments: Hold the movement and move the holds! In G. R. Colter, (ed.), *Phonetics and Phonology, Volume 3: Current Issues in ASL Phonology.* San Diego: Academic Press.
Woodward, J.
1996 Modern Standard Thai Sign Language, influence from ASL, and its relationship to original sign language varieties in Thailand. *Sign Language Studies* 92: 227–252.
2000 Sign languages and sign language families in Thailand and Viet Nam. In K. Emmorey and H. Lane (eds.), *The Signs of Language Revisited:*

An Anthology to Honor Ursula Bellugi and Edward Klima. Mahway, NJ: Lawrence Erlbaum Associates, Inc.

Zwitserlood, P., J. Bölte, and P. Dohmes
2000 Morphological effects on speech production: evidence from picture naming. *Language and Cognitive Processes* 15(4-5), 563–591.

Phonological priming in British Sign Language

Matthew W. G. Dye and Shui-I Shih

Models of lexical access seek to explain how incoming language data is mapped onto long-term lexical representations. The experiment reported here aims to provide insight into which elements of language input are used for mapping onto a sign language lexicon. Rather than using the organs of the vocal tract, sign languages use the arms, hands, body and face to create meaning, combining handshapes, locations and movements to create meaningful words (signs). This study aims to determine whether these parameters are also used in lexical access processes. Twelve deaf native and twelve deaf non-native signers of British Sign Language (BSL) were presented with a primed lexical decision task. They were required to make a lexical decision about a target sign after viewing a preceding prime that was phonologically related to the target. Analysis of the data suggests that native signers use phonological information in signs in order to access their mental lexicon. Moreover, it appears that the salient parameter in the input is a combination of location and movement – only when both these parameters are shared by prime and target is facilitatory priming observed. There was no evidence that non-native (deaf) signers used phonological parameters to access their lexicon, despite a high degree of success in the lexical decision task at a speed comparable to that of native signers. These findings are discussed in relation to sign language acquisition and the development of phonological theories of signed languages.

1. Phonological priming in British Sign Language

This study aims to establish whether deaf signers use the phonological parameters of signed languages to access their mental lexicon, and whether models of lexical access derived from spoken languages are of a type that can explain such a process. Sign languages are the natural languages of the deaf communities around the world and use the body, hands, and face to communicate, rather than the vocal organs. Each sign language possesses phonological, morphological and syntactic structures that exploit movement through space to create meaningful utterances (see Sutton-Spence and Woll 1999, for a discussion of these processes in British Sign Language). Models of lexical access for spoken languages, such as cohort theory (Marslen-Wilson and Welsh 1978) and the neighborhood activation model (Goldinger, Luce and Pisoni 1989), are intended to explain how a phonetic input is recognized as a token of a lexical item and how stored information about that item is retrieved. Although contemporary models differ in certain respects, in others they are in broad agreement. They all conceive of lexical recognition as the summation of evidence for the hypothesis that a given phonetic input is a token of a certain lexical item. This evidence is usually instantiated as activation of a node representing a lexical item, with that activation being derived from bottom-up (phonetic) and top-down (contextual) information. The initial stage of lexical access is feature extraction, with phonetic features in the input stream being extracted. If a lexical node is associated with the input features, then it receives activation; if not, then its activation is attenuated. Here we report a primed lexical decision study using a sign language, designed to investigate whether phonetic features related to lexical representations based upon handshape, location and movement are extracted and used to access the mental lexicon.

1.1. Sign language phonology

Current models of phonology are based heavily upon studies of American Sign Language (ASL), and have the three main parameters of handshape, location and movement as a common denominator. A sign can be conceptualized as consisting of a handshape, held in a specific location and executed by performing a particular movement. For example, the BSL sign NUMBER is executed by tapping the chin twice with a clenched fist. If the sign were articulated on the forehead, then the meaning would change to STUPID. Indeed, such sign pairs have been used to determine the phonemes within sign languages. Figure 1 shows three such minimal pairs in BSL: FAKE and ATTITUDE differ in their handshape, LONDON and WHO in terms of location, and PERSON and ITALY involve different movements.

It is also possible to combine parameters that occur in the language to create an acceptable sign with no meaning in that language or a sign that is not acceptable. For example, taking the sign ITALY and changing the handshape created the sign shown in Figure 2a.

Figure 1. Six BSL signs representing three minimal pairs (arranged in columns): FAKE–ATTITUDE (handshape), LONDON–WHO (location) and PERSON–ITALY (movement). The white arrows and symbols indicate movement. A single arrow denotes a simple path movement; 'c' indicates a rotation of the wrist, 'O' indicates a large circling motion; alternating '>' and '<' represents a small horizontal oscillation during a downward movement. Wrist rotations are performed at the same time as downward movements. The word in capitals is the nearest translation equivalent in English.

Figure 2. (a) ITALY signed with a W-handshape creates a possible sign in BSL that does not occur in the language. (b) A B-handshape and a 1-handshape making an orchestral conducting motion results in a sign that is not acceptable in BSL, because the language requires two-handed signs to share the same handshape.

Although this is a possible sign in BSL, it does not occur. The sign in Figure 2b is a combination of parameters existing in BSL, but it violates constraints on what signs can look like, so would not be acceptable to a native BSL signer.

1.2. Phonological processing in deaf signers

This parameter-based approach to sign language phonology was pioneered by William Stokoe (Stokoe, Casterline and Croneberg 1965), and has been superceded by more complex phonological models that take into account the temporal and non-linear properties of signs and the morphological processes that operate upon them (Brentari 1998, this volume; Sandler 1999, this volume). Nevertheless, handshape, location and movement are still used within these accounts, and they have been used to provide evidence for phonological coding in the working memory of deaf signers (Wilson and Emmorey 1997; 1998; 2001; 2003). In addition, Emmorey and Corina (1990) used a gating task to investigate the use of phonetic information by native signers of ASL in a sign recognition task. Their results indicated that first the location and orientation of the sign were extracted, followed by the handshape and finally, movement. It was isolation of the movement parameter that led to the identity of the sign being determined. More recent work by Hildebrandt and Corina (2002) presented deaf signers with a target sign along with four other signs which varied in the parameters they shared with the target. The signers rated signs that shared both location and movement as being more similar than signs sharing other combinations of parameters. Thus, we have linguistic evidence for the validity of sign language phonology as a concept, and psycholinguistic evidence that phonological parameters such as handshape, location and movement are used in working memory and word (sign) recognition processes.

It is important to note that not all deaf signers are native users of their sign language. In the UK, only around 5–10% of deaf community members are native BSL signers (Dye and Kyle 2000). Most signers are born to hearing parents, and acquire BSL later in life at school or through friends. Research by Mayberry and her colleagues (Mayberry 1993; Mayberry and Eichen 1991; Mayberry and Fischer 1989) has shown that acquiring a sign language late results in an impoverished ability to process that language. Of particular interest here, late acquirers make more phonological substitution errors in a sentence repetition task, whereas native signers make more semantic substitution errors. This is particularly evident in deaf people for whom a sign language is acquired late and as a first language, i.e. they did not acquire a spoken language such as English prior to learning the sign language, although it is also seen in those who acquired a sign language as young children (Mayberry 1993). The preponderance of phonological substitution errors for these late learners suggests that there is a bottleneck early on in language processing, which contrasts with the efficient and more automatized language processing of native signers.

1.3. Models of spoken word recognition

Logogen theory (Morton 1969), cohort theory (Marslen-Wilson and Welsh 1978) and the neighborhood activation model (Goldinger, Luce and Pisoni 1989) all conceive of lexical access as the summation of evidence for the listener's hypothesis that a given phonetic input is a token of a certain lexical item. This evidence is usually instantiated as activation of a node representing a lexical item, with that activation being derived from bottom-up (phonetic) and top-down (contextual) information. The initial stage is feature extraction, with phonetic features in the input stream being extracted. If a lexical node is associated with the input features, then it receives activation; if not, then it is deactivated (cohort theory) or no activation is passed on to it (neighborhood activation model). Marslen-Wilson (1990) states three basic assumptions which all current theories of lexical access share: (1) the activation metaphor is appropriate for computing the goodness-of-fit between input and stored lexical representations; (2) perceptual processing of the input is based upon competition between simultaneously-active lexical representations; and (3) decisions about the input are based upon the relative levels of activation of the competing representations. In the work reported here, we focused upon the predictions made by cohort theory.

In cohort theory (Marslen-Wilson and Welsh 1978), which addresses spoken languages, the onset of the input determines the initial set of candidate items. Thus, given the input [pat], the onset [p] determines the initial set of possible lexical items, i.e. all items beginning with [p]. Further information restricts the possible set of items further, until one item emerges as the 'winner'. Cohort theory predicts that a target will be recognized more quickly when preceded by a prime with a similar onset than when preceded by a prime with a differing onset. The input activates units for similar words as well as the presented word. The residual activation in similar word units will only remain for a short period of time as it is subject to decay. However, if one of these similar words is presented before activation decays to resting levels, then it will have a 'head-start' and reach criterion level more quickly. As the initial cohort of possible words is determined by the onset of the input, this should occur only if prime and target have the same onset.

Slowiaczek and Pisoni (1986) tested this prediction using a lexical decision task, and found that onset similarity between prime and target failed to facilitate response to the target. In fact, they observed inhibition when a large number of initial phonemes overlapped. However, for an identification-in-noise task, significant facilitation of response was observed as a function of phonological overlap. Slowiaczek and Pisoni concluded that the lexical decision may have taken place in a different level of the system, possibly using abstract lexical representations, and/or the identification in noise task involved more reliance upon the processing of phonetic input due to stimulus degradation. Using spoken word input and a lexical decision task, some authors have reported facilitation of response latency when prime and target overlap at onset (Goldinger et al. 1992), yet this has not been replicated by others (Marslen-Wilson 1993; Praamstra, Meyer and Levelt 1994; Radeau, Morais and Dewier 1989). Other studies have even found inhibition of response (Marslen-Wilson 1990; Slowiaczek and Pisoni 1986). Slowiaczek and Pisoni (1986) proposed revising cohort theory by adding inhibitory connections between units. Such lateral inhibition would mean that the more activated a unit became, the more it would inhibit phonologically similar neighbors. This would reverse the prediction of cohort theory, predicting that onset similarity of prime and target phonemes leads to inhibition of response to the target.

1.4. Aims and hypotheses of current study

Cohort theory, as applied to sign languages, states that the phonological parameters of signs are used to access lexical representations. If the location parameter defines the initial cohort (as suggested by Emmorey and Corina 1990), then the sign cohort should initially consist of all signs with a given location. Further information – handshape and then movement – would serve to attenuate the activation of mismatching cohort members and reinforce the activation of congruent cohort members. Within a priming paradigm, cohort theory predicts that prime-target similarity will facilitate response to the target when prime and target share the same location, as location determines the initial cohort of competitors. Lexical decisions in cohort theory are based on relative activation levels (a criterion difference must be reached for recognition to occur). If the target belongs to the same cohort as the prime, then it will already possess some residual activation from processing of the prime. This residual activation is the source of the facilitation.

This study tested the predictions of cohort theory for lexical decisions in a sign language, BSL, using a phonological priming design and a lexical decision task. Two groups of deaf signers (those who learned BSL from birth, and those who acquired it after the age of 6 years) were presented with pairs of signs and were required to make a lexical decision about the second sign in each pair. The type of phonological overlap between the prime and target of each pair was manipulated, and reaction time and accuracy measures were recorded. It was predicted that response times to targets that shared location with their primes would be facilitated. Furthermore, if this facilitation were induced by real sign primes only, this would suggest that priming occurs between sign representations in the lexicon. However, priming by non-signs would indicate that priming could occur at the level of the parameter, i.e. pre-lexically. Furthermore, based upon the work of Mayberry (1993), this facilitation is hypothesized to occur only for native signers who acquired BSL as a first language before the age of 6 years. If facilitative priming were restricted to native signers alone, it would suggest that the ability to efficiently process signs in terms of their phonological parameters must be acquired early in life.

2. Method

2.1. Participants

There were two groups of participants: 12 native signers of BSL who learned the language at home from their family and 12 non-native signers of BSL who came from hearing families and acquired BSL later in life. All participants were profoundly deaf from birth. There are no standardized measures of signing skill in BSL, so a self-rating procedure was employed. Native signers rated themselves as excellent or very good signers; non-native signers selected either good or very good signer to describe their competence. All participants were paid for their involvement.

2.2. Stimulus materials

All signs used in the study were recorded using a Panasonic digital camcorder (Model No. NV DS25B). Five deaf native signers served as models. They wore plain light-colored clothes and were filmed in front of a dark blue background to maximize contrast. Signs were edited into Video for Windows™ movie files using QuickEditor 6.0 video-editing software. Each movie file was 384 pixels wide and 288 pixels high, with a frame rate of 25 fps.

Two types of stimuli were used: signs and non-signs. Signs were in the lexicon of BSL and were mono-morphemic. They covered a range of grammatical classes, the majority being nouns, verbs and adjectives. Non-signs were not in the BSL lexicon, although they were permissible signs within the language, i.e. they conformed to the rules of sign formation in BSL (see Figure 2). The sign models themselves generated the non-signs, and confirmed that they were permissible in BSL.

These signs were arranged into prime–target pairs, such that there were an equal number of sign–sign, sign–non-sign, non-sign–sign and non-sign–non-sign pairs. For the sign–sign and non-sign–non-sign pairs, there were eight possible types of phonological overlap between the signs. They could share no parameters (hlm), handshape (Hlm), location (hLm), movement (hlM), handshape and location (HLm), handshape and movement (HlM), location and movement (hLM) or handshape, location and movement (HLM). When signs shared handshape, location and movement, they were identical, and the same movie file was used for each member of the pair. Examples of stimulus items for sign–sign and sign–non-sign pairs are given in Figure 3 and Figure 4, respectively. Identity pairs were not possible for sign–non-sign and non-sign–sign combinations, so the HLM condition was omitted for these. Within each pair type, and for each type of phonological overlap, 10 different sign pairs were recorded. There were therefore a total of 300 pairs of signs, with each sign model producing 60 pairs.

Previous studies have used the same stimulus repeatedly as a target, observing the differential effects of varying primes on that target. Such a procedure would have required a between-subjects design in order to avoid repetition of the target sign, and the resulting priming due to its repeated presentation. The scarcity of native signers for the present study necessitated a within-subjects design, and therefore different targets were used on each trial. The order of signs within a pair was fixed for all participants.

Figure 3. Sign–sign pairs used in the priming study, arranged in columns. The white arrows and symbols indicate movement. A single arrow denotes a simple path movement; double arrows indicate a double tap; 'c' indicates a repeated rotation of the wrist, '+' indicates a plus sign being outlined; '^^^^^' represents a repeated left-to-right movement with trilling of the fingertips. The word in capitals is the nearest translation equivalent in English. The movements in CHAIR and OLD were considered to differ on the basis of number of hands employed. The handshape for BINOCULARS used in the study was the same as for the sign GROUP.

A further 150 signs were recorded to be used as distractors within a sign recognition task. The relationship between these distractor signs and the recognition targets was not manipulated. Another 30 signs were filmed for use as practice items, and two signs (YES and NO) were used for response key practice.

As the non-signs were not naturally occurring, the duration of articulation of target signs was measured across all sign pairs and entered into an ANOVA to examine whether the duration of signs differed from that of non-signs. This revealed a significant two-way interaction between prime lexicality and target lexicality: F (1, 296) = 7.01, p < .01. In conditions where prime signs were real BSL signs, non-sign targets (M = 2004 msec) took longer to articulate than real sign targets (M = 1665 msec). When prime signs were non-signs, then the articulation time of non-sign targets (M = 1975 msec) and sign targets (1905 msec) did not significantly differ. In an attempt to control for the effects of articulation duration on response latencies, the clip duration of each target sign was used as a covariate in reaction time analyses.

Figure 4. Sign–non-sign pairs used in the priming study. Movement is denoted using the same symbols as in *Figure 3.* No translation equivalent is given for the non-signs, as they have no meaning in BSL.

2.3. Design

A primed lexical decision was interleaved with a sign recognition task. In the primed lexical decision task, deaf native signers and deaf non-native signers observed sign pairs (a prime sign followed by a target sign), and were required to make lexical decisions about the target sign in each pair. There was an interval of 50 milliseconds between offset of the prime and onset of the target. Measures were taken of response accuracy and response latency (milliseconds from onset of target) of each lexical decision. The priming task began with practice trials, followed by presentation of 300 experimental trials divided into ten blocks, with one pair from each condition in each block. Blocks were presented in the order 1–10 or 10–1, counterbalanced across participants in each experimental group. The same sign model was not used in consecutive blocks.

Following each priming block was a recognition task, using stimuli seen as primes in the previous priming block, plus an equal number of novel distractor items. For each participant the proportion of correct responses was calculated. The purpose of this recognition task was to encourage participants to pay attention to the first item of each pair in the priming task and thereby ensure processing of the prime. A pseudo Latin square procedure was used to assign primes from different conditions to the recognition blocks, such that each condition was equally represented across the recognition blocks. Each recognition block used the same signer as in the previous priming block.

2.4. Apparatus

Stimuli were presented to participants using specially written software (Sign Prime v6.2) on an IBM PC with a 15" diameter CRT display. Participants responded using a specially designed response box, with three buttons: red, green and black. For half of the participants, the red button was used to issue a NO response, the green for YES, and the black to initiate the next trial. For the remainder, the roles of the red and green buttons were reversed.

2.5. Procedure

The procedure consisted of four different phases: key press practice, lexical decision practice, a primed lexical decision task and a recognition test. The initial key press practice was intended to familiarize participants with the response procedures. Participants were presented with sixteen trials, each one being the presentation of the BSL sign for either YES or NO (eight times each). Half of the participants were required to press the green button if they saw the sign YES, and the red button if they saw the sign NO. This was reversed for other participants. The black button initiated the next trial. The number of trials was set to a minimum of 10, with the phase stopping after 5 consecutive correct responses had been made.

This was followed by a practice session, where participants were told that they would see a sequence of two signs in the middle of the computer screen. Their task was to decide if the second sign was a real BSL sign or a made-up nonsense sign. If they thought that it was a real sign, they were instructed to press the 'yes' button they had been trained on, and to press the 'no' button if they thought it was a nonsense sign. Pressing the black button initiated the next trial. After each response, the participants were told whether they were correct or incorrect, and whether their response was early, good or late. If a response was made within the first 20% of the target clip, the response was considered to be anticipatory, and classified as early; if the response occurred after 100% of the target clip had been viewed, then it was classified as late. All other responses were classified as good. This was done in order to encourage participants not to respond too early, or spend too long considering the lexical status of the target. In all, there were twelve practice trials, with pairs sharing none of the three phonological parameters. There were equal numbers of sign–sign, sign–non-sign, non-sign–sign and non-sign–non-sign pairs. After the practice trials had been completed, participants observed twelve signs presented individually in a prime recognition task. Six of these signs had been used as primes in the preceding trials, and six were novel distractor signs. Participants were required to indicate which of the signs had been presented in the previous lexical decision block, and which were new.

The procedure for the actual primed lexical decision task was the same as for the practice, except that no feedback on performance was given. There were ten blocks of lexical decision trials, with a prime recognition task after each block.

3. Results

3.1. Prime recognition

In order to ensure all subjects were paying attention to the primes, hit rates and false alarm rates were calculated for each subject from the prime recognition trials. Native and non-native signers did not differ significantly in their hit rate, $t(22) = 0.12$, p = n.s., or false alarm rate, $t(22) = 0.48$, p = n.s. D-prime scores were calculated for each subject, revealing one non-native signer with a score more than 2 standard deviations below the group mean (d-prime = 0.18, hit rate = .57, false alarm rate = .51). This subject was removed from all further analyses. D-primes were high for all other subjects (native mean = 1.65, non-native mean = 1.63).

3.2. Accuracy of lexical decision

Accuracy data are reported in Table 1. The effects of prime lexicality (sign or non-sign), target lexicality (sign- or non-sign) and age of acquisition (native or non-native) were assessed using a 3-way mixed ANOVA. No main effects or interactions were statistically significant. Subjects were as accurate at rejecting non-signs as they were at identifying real BSL signs. Lexicality of the preceding prime did not affect this, nor did the age at which they acquired BSL. Both native and non-native signers responded with a high degree of accuracy, suggesting equivalent vocabulary knowledge.

Accuracy data were much lower than average for two native and one non-native signer, and these outliers were removed from subsequent reaction time analyses.

Table 1. Accuracy of lexical decision for native and non-native signers, expressed as percentage correct (SD).

Group	Prime-target lexicality			
	Sign –sign	Sign –non-sign	Non-sign –sign	Non-sign –non-sign
Native signers (n = 12)	92 (15)	91 (18)	92 (14)	93 (13)
Non-native signers (n = 11)	91 (15)	90 (18)	90 (15)	90 (12)

3.3. Time required for lexical decision

An initial analysis examined the effects of prime lexicality, target lexicality and age of acquisition on the time taken to make a lexical decision. Reaction time data were only analyzed for correct responses, and are reported in Table 2 for native signers and Table 3 for non-native signers. Reaction times were measured as the time between the onset of the target sign and the subject's response. Collapsing the data across sign parameters, three-way mixed ANOVAs – by subjects and by stimuli – revealed significant main effects of prime lexicality, F_1 (1, 18) = 110.92, p < .001 and F_2 (1, 259) = 11.90, p < .005, target lexicality, F_1 (1, 18) = 20.80, p < .001 and F_2 (1, 259) = 35.39, p < .001, and age of acquisition, F_1 (1, 18) = 4.49, p < .05 and F_2 (1, 259) = 97.29, p < .001. None of the interactions approached significance. When primes were real BSL signs, subjects responded to targets 65 milliseconds more quickly on average. All subjects accepted real BSL signs more quickly than they rejected non-signs, on average by 80 milliseconds. Native signers responded more quickly than non-native signers by 151 milliseconds. They were over 10% faster than non-native signers in making their decisions. Equivalent accuracy scores militate against this being considered a greater speed-accuracy trade-off on the part of native signers.

As noted in section 2.2, the duration required to articulate each sign clip varied and was not controlled for in stimulus selection. Because of this, the analysis was repeated using clip duration as a covariate in a 'by stimuli' analysis. The effects of introducing the covariate differed for native and non-native signers, so they are reported separately here. For native signers, clip duration was a significant covariate: F (1, 269) = 692.04, p < .001. It resulted in the effect of prime lexicality being non-significant, although the effect of target lexicality remained: F (1, 269) = 15.66, p < .001. For non-native signers, clip duration was also a significant covariate: F (1, 269) = 119.95, p < .001. Its introduction into the ANOVA model did not remove the significant effects of either prime lexicality or target lexicality: F (1, 269) = 13.90, p < .001 and F (1, 269) = 10.05, p < .01 respectively. For native signers, as expected, it took longer to correctly reject non-signs than it did to correctly identify real BSL signs. This was also true for non-native signers, with the added finding that a non-sign prime resulted in slower processing regardless of the lexical status of the target.

The next step in the analysis was to examine the effects of the parameters shared by prime and target signs on response times. Initially, signs sharing no parameters (hlm) were compared with those sharing all parameters (HLM, or identical signs). These analyses were therefore confined to sign–sign and non-sign–non-sign pairs. For these pair types, there were no significant differences in reaction time for the native or the non-native signers (all F < 1). Again, clip duration was introduced into the models as a covariate. This revealed a marginally significant effect for native signers for sign-sign pairs only, F (1, 16) = 3.96, p = .067. Thus there was some evidence for identity priming for real BSL signs in the native signing group. For these subjects, seeing a sign caused it to be processed faster on a subsequent presentation. This advantage did not extend to non-signs, suggesting that the source of the processing benefit is a sign-to-sign interaction in the lexicon and not due to the previous activation of similar parameter units.

Table 2. Mean reaction time (milliseconds) and standard deviations for correct lexical decision for native signers (n = 10).

Parameters shared by prime and target	Prime-target lexicality			
	Sign –sign	Sign –non-sign	Non-sign –sign	Non-sign –non-sign
None	1217 (43)	1245 (49)	1288 (33)	1297 (51)
Handshape	1210 (23)	1257 (46)	1257 (40)	1257 (47)
Location	1210 (31)	1267 (33)	1321 (44)	1322 (41)
Movement	1232 (34)	1313 (38)	1285 (31)	1358 (41)
Handshape and Location	1224 (35)	1360 (46)	1272 (45)	1396 (54)
Handshape and Movement	1215 (40)	1293 (42)	1249 (38)	1315 (38)
Location and Movement	1175 (25)	1324 (43)	1254 (42)	1368 (46)
All	1095 (29)	---	---	1257 (42)

Table 3. Mean reaction time (milliseconds) and standard deviations for correct lexical decision for non-native signers (n = 10).

Parameters shared by prime and target	Prime-target lexicality			
	Sign –sign	Sign –non-sign	Non-sign –sign	Non-sign –non-sign
None	1325 (38)	1387 (36)	1445 (38)	1457 (38)
Handshape	1277 (29)	1356 (47)	1411 (38)	1447 (51)
Location	1243 (28)	1361 (48)	1463 (40)	1420 (38)
Movement	1297 (31)	1390 (44)	1406 (38)	1441 (48)
Handshape and Location	1340 (26)	1491 (41)	1349 (34)	1468 (49)
Handshape and Movement	1257 (38)	1417 (43)	1339 (39)	1371 (41)
Location and Movement	1253 (26)	1413 (39)	1340 (44)	1443 (37)
All	1230 (39)	---	---	1391 (51)

Finally, examining the effects of shared parameters on non-identical prime-target pairs tested for the presence of phonological priming. Here the critical analyses are for sign–sign and non-sign–sign pairs. Sign-on-sign effects in the absence of non-sign-on-sign effects would indicate priming at the level of the whole sign. However, the co-occurrence of non-sign-on-sign effects would indicate that the priming occurs at a lower, parameter-based, level.

For both native and non-native signers, ignoring the effects of target articulation duration on response latency, for the sign-sign pairs there were no significant main effects of sharing handshape, location or movement, and no significant interactions (all $F > 1$, but $p > .05$). However, the introduction of clip duration as a covariate suggested that the effect of articulation duration was masking some effects. For native signers, the covariate revealed a significant main effect of shared location, $F(1, 73) = 6.90$, $p < .05$, as predicted in light of Emmorey and Corina's (1990) findings. Native signers required less time to correctly identify a BSL sign as a real sign when it was preceded by a prime sharing location ($M = 1203$ msec) than when the prime did not share location ($M = 1219$ msec). For native signers, there was also a marginally significant interaction between shared location and movement, $F(1, 73) = 3.54$, $p = .064$. In light of Hildebrandt and Corina's (2002) study of sign similarity judgments, planned comparisons were performed to examine the source of the interaction. These revealed that reaction times were significantly faster when prime and target shared both location and movement ($p < .05$). It is important to note that this analysis did not include data from identical sign–sign pairs. Thus there is an observable effect due to shared location and movement, which cannot be attributed to the prime and target being identical (see Figure 5). The main effect of shared location and the interaction between shared location and movement did not approach significance for any pair type except the sign–sign pairs, suggesting that the priming effect was again occurring at the level of sign-to-sign interactions in the lexicon. For non-native signers, introducing articulation duration as a covariate led to a significant main effect of shared movement for sign–sign pairs, $F(1, 72) = 5.24$, $p < .05$. Targets sharing movement with their prime were responded to more quickly than when they did not share movement. However, this effect was also significant for non-sign–sign pairs, $F(1, 62) = 5.63$, $p < .05$, with the effect in the same direction. This would suggest that any priming is taking place at the level of individual parameter units – phoneme units in cohort theory – and is not lexical in origin.

4. Discussion

The study reported here tested the predictions of cohort theory using a signed language – British Sign Language (BSL). Furthermore, by comparing native and non-native signers of BSL, we hoped to see whether early acquisition of a sign language was important for fluent and automatic lexical access. Whilst lexical decision accuracy was the same for native and non-native signers, suggesting an equivalent vocabulary knowledge and familiarity with the language, native signers made faster decisions. In addition, whereas lexicality of the target signs affected the response latency of both groups, only non-native signers were influenced by the lexicality of the prime signs. For both groups, sign targets were correctly identified as real signs more quickly than non-signs were correctly rejected as being meaningless. This mirrors findings from spoken language studies, and is taken to reflect an exhaustive search of the lexicon being required before a possible non-sign can be rejected. Non-sign primes resulted in slower processing of the targets by non-native signers. This may reflect less automatic, and more effortful, processing of non-signs by non-natives. That is, extra resources are directed towards processing of non-sign primes, given the recognition task requirements, leaving fewer resources for processing of the target.

The group and lexicality effects suggest, therefore, that lexical access in non-native signers is accurate but requires greater effort and is less automatized. The phonological priming effects suggest that the process is not only slower for non-native signers, but that the process is different. For native signers, lexical decisions were faster when a sign sharing both location and movement preceded the target sign. This was the case whether or not the signs also shared handshape. Importantly, this facilitative priming occurred only for prime-target pairs that contained two real BSL signs. When the prime was a non-sign, the phonological priming effect was not observed. This suggests that the effect takes place at the level of the lexicon. Signs within the lexicon that share both location and movement are connected, such that activation of one sign leads to activation of the others. As predicted by the results of Hildebrandt and Corina (2002) and Corina and Knapp (this volume), and to some extent by Emmorey and Corina (1990), it is a combination of a sign's location and movement parameters that is used as the initial input for lexical access. For non-native signers, this location-plus-movement priming effect did not occur. Rather, there was facilitative priming based upon shared movement. However, this priming was observed for both sign–sign and non-sign–sign pairs. As a result, it is not possible to attribute the effect to solely lexical processes. If this were the case, then non-signs should not produce priming, as they are not represented in the lexicon. Possible explanations include priming in pre-lexical processes, i.e. it is the movement parameter representation that is primed, or non-lexical effects due to the saliency of movement for a deafened individual. Whatever the reason, lexical access in non-native signers appears to be a qualitatively different process to that used by native signers.

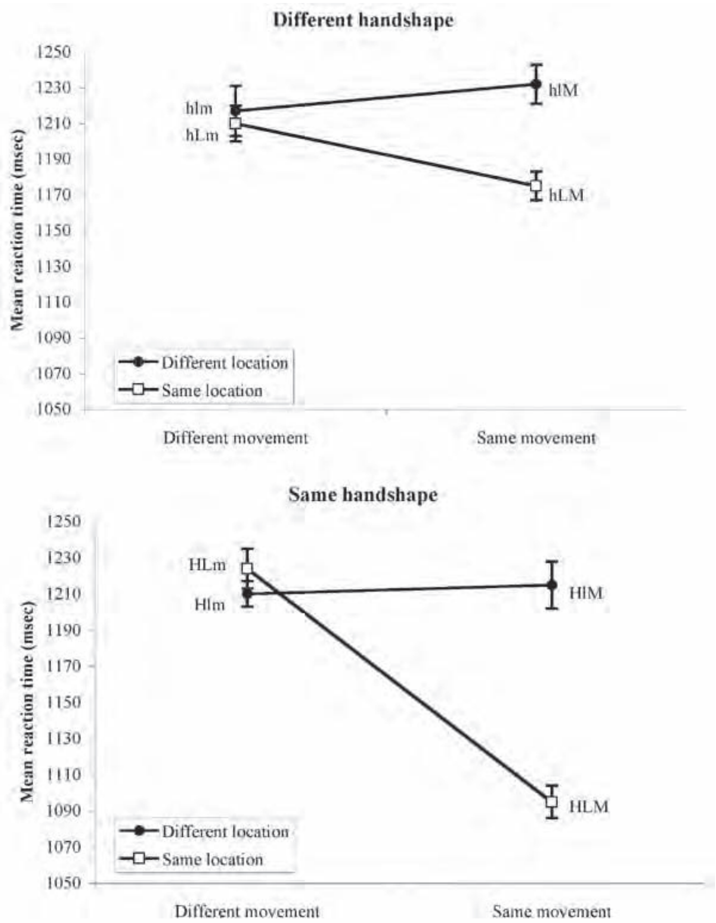


Figure 5. Reaction times for native signers viewing sign–sign pairs. When the target sign shared location and movement with the prime, native signers were faster to make a lexical decision. This was true whether the signs also shared handshape (i.e. were identical) or the signs differed in their handshape (e.g. NUISANCE–CAKE).

Models of lexical access that conceive of the process as the accumulation of phonetic evidence for a listener's hypothesis, such as cohort theory, can be used to explain lexical access processes in viewers of a sign language. The data reported here suggest that native signers use a combination of location and movement as the initial input in access to the sign lexicon. Non-native signers appear to use a qualitatively different process.

Acknowledgments

This research was supported by a graduate research scholarship awarded to the first author by the University of Southampton, UK. We wish to thank Pete Dargie for his effort and ingenuity in creating the SignPrime software necessary for this study, David Corina, Mike Tanenhaus and Elissa Newport for helpful discussions, and two anonymous reviewers for their constructive comments.

References

Brentari, Diane
(1998) *A Prosodic Model of Sign Language Phonology*. Cambridge, MA: MIT Press.

Dye, M. and J. Kyle
(2000) *Deaf People in the Community: Demographics of the Deaf Community in the UK*. Bristol: Deaf Studies Trust.

Emmorey, Karen and David P. Corina
(1990) Lexical recognition in sign language: Effects of phonetic structure and morphology. *Perceptual and Motor Skills* 71: 1227–1252.

Goldinger, Stephen D., P. A. Luce and David B. Pisoni
(1989) Priming lexical neighbors of spoken words: Effects of competition and inhibition. *Journal of Memory and Language* 28: 501–518.

Goldinger, Stephen D., P. A. Luce, David B. Pisoni and J. K. Marcario
(1992) Form-based priming in spoken work recognition: The role of competition and bias. *Journal of Experimental Psychology: Learning, Memory and Cognition* 18: 1211–1238.

Hildebrandt, Ursula C. and David P. Corina
(2002) Phonological similarity in American Sign Language. *Language and Cognitive Processes* 17: 593–612.

Marslen-Wilson, William D.
(1990) Activation, competition, and frequency in lexical access. In: Gerry T. M. Altmann (ed.), *Cognitive Models of Speech Processing:*

Psycholinguistic and Computational Perspectives. Cambridge, MA: MIT Press.
(1993) Issues of process and representation in lexical access. In: Gerry T. M. Altmann and Richard Shillcock (eds.), *Cognitive Models of Speech Processing: The Second Sperlonga Meeting*, 187–210. Hove, UK: Lawrence Erlbaum Associates.

Marslen-Wilson, William D. and A. Welsh
(1978) Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology* 10: 26–63.

Mayberry, Rachel
(1993) First-language acquisition after childhood differs from second-language acquisition: The case of American Sign Language. *Journal of Speech and Hearing Research* 36: 1258–1270.

Mayberry, Rachel and E. B. Eichen
(1991) The long-lasting advantage of learning sign language in childhood: Another look at the critical period for language acquisition. *Journal of Memory and Language* 30: 486–512.

Mayberry, Rachel and Susan D. Fischer
(1989) Looking through phonological shape to lexical meaning: The bottleneck of non-native sign language processing. *Memory and Cognition* 17: 740–754.

Morton, J.
(1969) Interaction of information in word recognition. *Psychological Review* 76: 165–178.

Praamstra, P., A. S. Meyer and Willem J. M. Levelt
(1994) Neurophysiological manifestations of phonological processing: Latency variation of a negative ERP component timelocked to phonological mismatch. *Journal of Cognitive Neuroscience* 6: 204–219.

Radeau, Monique, J. Morais and A. Dewier
(1989) Phonological priming in spoken word recognition: Task effects. *Memory and Cognition* 17.

Sandler, Wendy
(1999) A prosodic model of sign language phonology. *Phonology* 16: 443–447.

Slowiaczek, Louisa M. and David B. Pisoni
(1986) Effects of phonological similarity on priming in auditory lexical decision. *Memory and Cognition* 14: 230–237.

Stokoe, William C., Dorothy C. Casterline and Carl G. Croneberg
(1965) *A Dictionary of American Sign Language on Linguistic Principles*. Washington, DC: Gallaudet College Press.

Sutton-Spence, Rachel and Bencie Woll
(1999) *The Linguistics of British Sign Language*. Cambridge: Cambridge University Press.
Wilson, Margaret and Karen Emmorey
(1997) A visuospatial "phonological loop" in working memory: evidence from American Sign Language. *Memory and Cognition* 25: 313–320.
(1998) A "word length effect" for sign language: Further evidence on the role of language in structuring working memory. *Memory and Cognition* 26: 584–590.
(2001) Functional consequences of modality: Spatial coding in working memory for signs. In: V. Dively, M. Metzger, S. Taub and A. M. Baer (eds.), *Sign Languages: Discoveries From International Research*, 91–99. Washington, DC: Gallaudet University Press.
(2003) The effect of irrelevant visual input on working memory for sign language. *Journal of Deaf Studies and Deaf Education* 8: 97–103.

I. Qualitative and variable faces of phonological competence

Spoken languages

Convergences and divergences
of signed and spoken languages

Signed languages

Phonetic implementation and phonetic pre-specification in sign language phonology

Harry van der Hulst and Els van der Kooij

In this chapter we will propose that the set of phonological features needed for sign languages is much smaller than what is usually proposed or assumed. Even though it has been recognized (since Stokoe's seminal work) that phonological features must capture only those properties of signs that are distinctive in the language, all subsequent models for sign language phonology typically encode a lot of phonetic detail that, on closer study, isn't really distinctive (in a phonological sense). In this chapter, we argue that the non-distinctive nature of these phonetic properties is due to two sources: (a) phonetic predictability (to be accounted for in terms of phonetic implementation rules) and (b) iconicity (to be accounted for in terms of lexical pre-specification). The two routes in (a) and (b) allow us to 'clean up' the phonology which, as a result, can be shown to be quite restricted and non-random, i.e. in accordance with structural principles that appear to play a crucial role in spoken language phonology as well. A case study involving the notion place of articulation is provided. Our claims are based on a study of signs from Nederlandse Gebarentaal (NGT – Sign Language of the Netherlands), and, in particular on a database (SignPhon) that contains over 3000 signs, provided with a detailed phonetic/phonological encoding.

1. Introduction

In this article it will be argued that the set of phonological features needed for sign languages is much smaller than is usually proposed or assumed. Reduction is possible by appealing to the following principles:

a. Strict adherence to the requirement that phonological properties be distinctive.
b. Specification of *semantically driven* phonetic properties in the lexical representation.

Even though it has been recognized in the sign language phonology literature since the seminal work of Stokoe (1960) that phonological features must capture only those properties of signs that are distinctive in the language, subsequent models of sign language phonology typically encode a lot of phonetic detail that, on closer study, may not be distinctive. A first criterion for distinctiveness is that some property can be distinguished perceptually from other phonetic properties. However, not every distinction that humans can perceive is by that fact alone de facto or even potentially distinctive. It is well known that many perceptible distinctions (often called *paralinguistic* or *extralinguistic*) only play a role in the expression of pragmatic, sociolinguistic or affective values. To be called distinctive it is necessary that some property distinguishes one lexical item (i.e. one meaning bearing unit) from another in at least one language. For instance, the phonetic property of voicing distinguishes the /b/ in *bak* 'container' from the /p/ in *pak* 'package' in Dutch and is accordingly referred to as the distinctive feature [voice]. Typically, of course, distinctive features distinguish many pairs of lexical items in many languages. Distinctive features thus distinguish meanings in the lexicon, while being meaningless units themselves.[1]

In this article it will be claimed, first, that phonetic properties that are non-distinctive must be accounted for by phonetic implementation rules. This idea is neither new nor should be controversial, as it merely promotes adherence to a fundamental characteristic of phonological analysis. A second claim is that certain phonetic properties of signs, although unpredictable in terms of phonetic implementation, still should not qualify as exponents of distinctive features because they are, as we will argue, *semantically driven*. It is proposed here that these semantically driven phonetic properties can be accounted for by specifying them directly in the lexical representation of signs. The proposal will be referred to as *semantically driven phonetic pre-specification*.

With respect to phonetic predictability (cf. 1a), this approach is non-controversial and traditional. The reason that phonologists tend to grant phonological status topredictable properties is mainly due to the fact that analyses of newly studied languages (let alone languages in a new modality) naturally focus first on accurate phonetic characterizations before a rigorous phonological analysis can start. Apart from (sometimes anecdotal) remarks about the predictability of the position of the thumb or of non-selected fingers, and so on, most 'phonological' feature proposals for sign languages do not seem to be based on systematic, large-scale investigations into distinctiveness. In section 3.3, we discuss several examples of phonetic properties that have been granted phonological status incorrectly, being predictably based on the presence of other formal elements.[2] Claiming that certain properties are predictable entails a commitment to spelling out the contextual factors that determine the precise phonetic exponents of the reduced set of phonological features. Therefore, the proposal for a reduced set of features comes with a proposal for a set of *phonetic implementation rules*.

The second route (specified in 1b) that allows us to reduce the set of phonological features is new and potentially controversial. It has long been recognized that many signs in sign languages have a property often referred to as 'iconic': aspects of the form of signs reflect aspects of the shape or action of referents.[3] The present study, however, will not be restricted only to the so-called iconic aspects but will include other types of semantic motivation of the phonetic shape of signs as well.[4]

If one understands language at any given point in time to result from conflicting forces, and focuses on the phonetic form of language, it seems obvious that semantic motivation is, in principle, in conflict with phonology. Whereas phonology implies a finite and minimal set of discrete meaningless building blocks[5] that are arranged in an autonomous, compositional structure, then semantic motivation (especially iconic motivation) may create an infinite array of gradual, holistic shapes that mirror the shapes of referents (object or action, directly or via metaphor). In short, phonology flourishes where meaning is absent.

Both phonology (through phonetic implementation) and semantic motivation determine the overall phonetic event of a sign. In addition, many aspects of the phonetic event are determined by factors such as the emotional state of the signer, the size of his or her articulators and all the situational factors that are part of the communicative setting (formality, distance to signee and so on). Crasborn (2001) offers an extensive discussion of the effects of these para- or extralinguistic factors on the phonetic shape of signs, and the variability that results. In doing a phonological analysis, researchers routinely abstract away from all extralinguistic factors, which are assumed to fall within the realm of 'sociolinguistics' (broadly defined). This chapter argues that it is important to also separate out the impact of the phonology and semantic motivation in a principled way.

The interplay between phonetic shape and semantic motivation could, in principle, be twofold: mediated via phonological structure or direct. In Figure 1, the circle in the center represents a phonetic form.

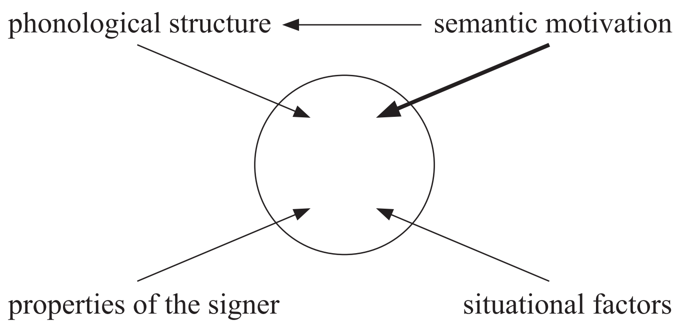


Figure 1. Diagram of the interplay between phonetic shape, phonological structure and semantic motivation in sign languages.

The view that semantic motivation of phonetic shape operates via the phonological structure is essentially that of Boyes Braem (1981). That is, the phonetic exponents of certain phonological features can be meaningful, leaving open the crucial possibility that these features also occur in signs where they do not carry meaning. For example, a feature [round] (indicating curved fingers: Boyes Braem 1981) may be semantically driven in, for instance, the ASL sign CUP, while it occurs in an arbitrary way in the ASL sign TO DECIDE. However, in NGT (*Nederlandse Gebarentaal*, Sign Language of the Netherlands) at least one can find several examples of semantically driven phonetic properties of signs that *never* occur as (arbitrary) distinctive properties. It is the latter kind of property (represented by the bold arrow in Figure 1) that one must be reluctant to give the status of phonological feature.

In early work on sign language phonology, it seemed crucial to discount the importance of iconicity in order to validate the claim that sign languages have dual patterning (independence of form and meaning) and thus a phonology. This was important because, prior to Stokoe's work, the prevailing idea was that the form of signs was determined by, and therefore not independent of, meaning. Hence, sign languages were believed to lack 'dual patterning', which had often been claimed to be a defining property of human language. Spoken languages, on the other hand, were said have dual patterning, emphasizing the Saussurian dictum that form and meaning are arbitrarily related. Thus, while on the whole spoken language phonology was taken to be arbitrary, sign language forms were taken to be iconic, or more generally meaning-driven.[6] Then, when Stokoe revealed arbitrary phonological composition in signs, it seemed best to sign researchers to deny the relevance of iconicity, which, as we have seen is fundamentally incompatible with the notion of phonological compositional structure.

The position advanced here is that due recognition of semantic forces which determine the phonetic shape of signs is not incompatible with an account of their phonological structure. The proper question is: how can *both* semantic motivation and phonological compositionality be accounted for in a harmonious, consistent and cognitively plausible way? The answer, we suggest, lies in the hypothesis that the lexical structure of signs contains a specification of semantically-driven phonetic properties *alongside* a phonological structure. The phonetically pre-specified properties can be unique for a sign or have a certain degree of productivity, as evidenced in the formation of new signs. Pre-specified phonetic properties logically take precedence over the phonetic implementation rules that spell out the phonetic properties of the phonological structure where a conflict arises. An example of such a conflict is provided in section 3.3.

It must be stressed that allowing pre-specification of phonetic properties conflicts with the traditional idea that all non-predictable properties are phonological (i.e. distinctive). After all, a semantically driven property may very well be the only difference between two signs that are otherwise identical. Hence, the danger should be recognized of using phonetic pre-specification as a wild card to be used to 'protect' the phonology from an excess of phonological features. The semantic motivation at the basis of pre-specification needs careful argumentation in the context of a theory about the interface between phonetic properties and semantics. One way to investigate the cognitive reality of recurrent form-meaning associations experimentally is to set up a sign recognition test in which signers would have to provide a meaning for signs consisting of recombinations of form elements from meaningful signs (i.e., pseudo-signs). Alternatively, signers could be asked to make up new signs given a certain description of an (imaginary) object or action. It could then be determined whether the form elements of the resulting new signs match the extra-phonological form-meaning pairings normally found in the lexicon.

In this section, the two strategies for reducing the number of elements in sign language phonology have been outlined, i.e. strict adherence to the idea that phonological features are distinctive and not predictable, and the proposal to specify semantically motivated phonetic properties in the lexical representation.

In the next section the phonological model and feature set that emerged from this reductionist approach is briefly sketched mainly to provide the background for the case study involving a proposal for the representation of Place of Articulation or Location properties that will be provided in section 3.

The claims made in this article are based on a study of signs from NGT and, in particular on a database (*SignPhon*; Crasborn et al. 1998. Crasborn 1998) that contains over 3000 NGT signs in isolation or citation forms, provided with a detailed phonetic/phonological encoding. An extensive report of this study can be found in Van der Kooij (2002). An accessible description of the Signphon database can be found in Crasborn, van der Hulst, van der Kooij (2001) and on http://www.let.kun.nl/sign-lang/signphon2.html.

2. Model

The reductionist model that is laid out in this section, the Dependency Model, is restricted by various structural principles, specifically binary branching and head-dependency relations. Heads express relative salient and/or stable information, whereas dependent nodes contain more moveable (e.g. in terms of spreading behavior) or dynamic information. The features that are used in this model are unary and are hierarchically organized into class nodes (such as *Selected Fingers* or *Orientation*). In the representation of signs in Figure 2, phonological under-specification is used, i.e. phonological features that are predictable from other, more basic features are left unspecified. These theoretical aspects, which cannot be motivated in the context of this brief article, were first introduced and examined in van der Hulst (1993) and have been further supported and developed in the form of several variants of the basic model in subsequent work. For a full recent version of the model (and relevant references), we refer to van der Kooij (2002:280).[7]

As for the manual node, the major cut is that between the *(active) articulator* and *location* (or *passive articulator*). The former then splits into *orientation* and *handshape*, a grouping that was adopted from Sandler (1989). Handshape proves to be the most detailed class node, as shown by the fact that it divides into *Finger Selection* and *Finger Configuration*, both with their own detailed internal structure (not shown here). The articulator node is taken to be the head of the whole sign segment because, intuitively, the articulator is the most central and obligatory unit, and because it is the most highly structured node. Location, as we will see in a moment, is not an obligatory aspect of signs.

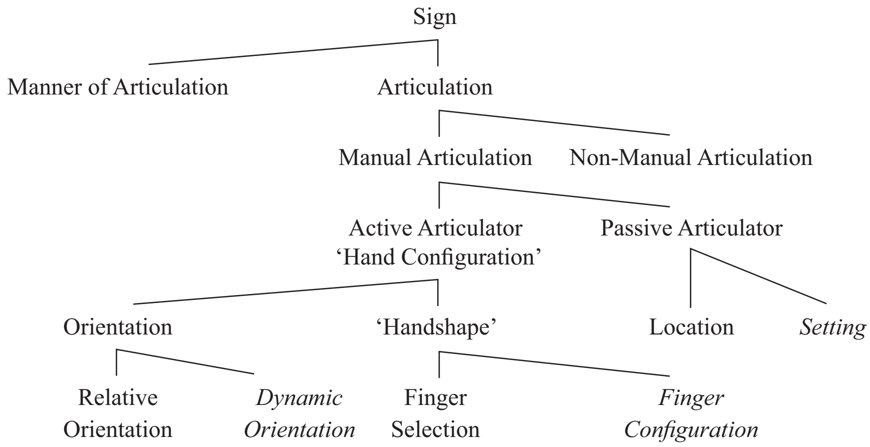


Figure 2. A partial version of the Dependency Model of signs.

Movement, in this model, is represented in terms of branching nodes that are dependent on the head nodes that determine the scope of the movement, as shown in Table 1.

Table 1. Specification of different types of movement in the Dependency Model

Head node:	*Dependent node branches:*	*Resulting movement:*
Location	Setting	Path Movement
Relative Orientation	Dynamic Orientation	Orientation change
Finger Selection	Finger Configuration	Hand-internal change

In this respect, the present model deviates from models that represent movement in terms of a 'skeletal position' M(ovement) that is sandwiched between two static skeletal positions (cf. Liddell & Johnson 1989, Sandler 1989, Perlmutter 1992).

In the Dependency Model, *path movement*, a movement of the whole articulator, is represented as a transition between two settings. For example a downward movement on the chest as in the NGT sign TIRED is represented as in Figure 3.

The settings [high] and [low] are interpreted with reference to the feature [trunk] specified in the head node. The same setting specification combined with the Location feature [cheek] would result in a much smaller path movement.

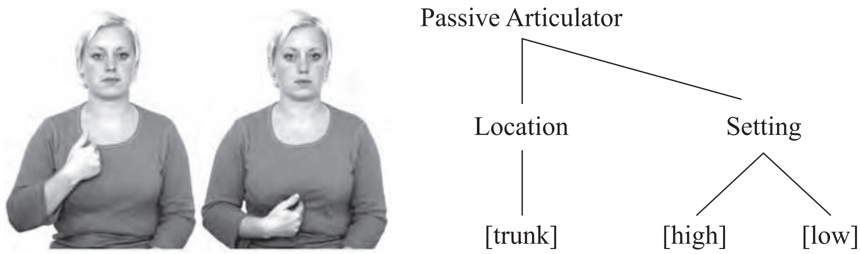


Figure 3. Downward path movement in NGT sign TIRED.

Sandler (1989) was the first to separate *Major Location* and *Setting*. In her model Setting is used to represent finer distinctions within one major area. Setting features include three pairs of features concerning the distance to the location (peripheral-distal), the height within the location (high-low), and the lateral side of the major location (ipsi-contra). Moreover, the feature [contact] is subsumed under the setting node. In Sandler's model, a combination of Major Location features and setting features establishes the set of *distinctive* locations. A distinctive location can thus consist of a Major location and one setting value, i.e., in the absence of movement. In addition, Setting in Sandler's model is used as a means to describe path movement. In that case, the setting specifications occur pairwise. A downward path movement consists of a specification of the settings 'high' and 'low'. Table 2 presents the setting specifications distinguished by Sandler.

Table 2. Setting specifications (Sandler 1989)

Setting:	
ipsilateral	setting on the same side of the body as the articulating hand
contralateral	setting on the opposite side of the body as the articulating hand
high	above the middle of the specified place (near the fingertip edge in case the place is 'hand')
low	below the middle of the specified place (near the heel edge or base in case the place is 'hand')
proximal	within a few inches of the place
distal	a comfortable arm's length from the place
contact	touching specified location by articulating hand(s)

In the current Dependency Model, the division in Location and Setting is adopted, but a clear division of labor between them is postulated. Location features are used to make underlying contrasts of distinctive areas on the head, body and arm and hand. Contrary to what is found in Sandler's model, therefore, setting is not used to make finer location distinctions. Setting features are *only* used to represent path movement within those distinctive areas. Consequently, Setting features only come in pairs, parallel to the features of other dependent nodes (*Aperture* and *Dynamic Orientation*). Path movements are thus regarded as the dynamic part of the location component. As in Sandler's model, the interpretation of the setting pairs representing the path movements is contingent on the distinctive location they depend on. In Figure 4, the passive Articulator node and the features that are subsumed under it are illustrated.

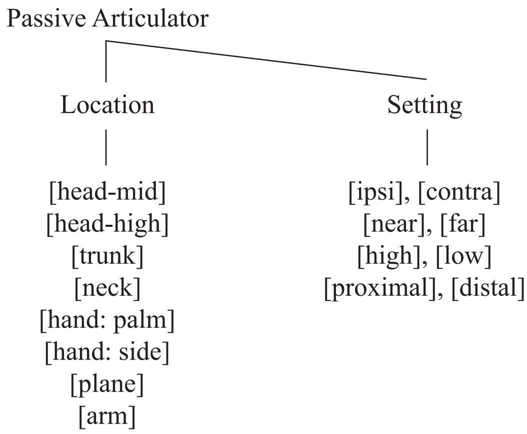


Figure 4. The Passive Articulator in the Dependency Model.

Phonological locations in the Dependency Model, then, refer to distinctive *areas* within which a setting change can take place rather than to specific *points* on the head, hand or body. This representation accounts for the observation in Battison (1978) that the movement of mono-morphemic signs stays within one major location. This observation was based on the behavior of so-called double contact signs that have contact at the beginning and at the end of the movement. For NGT it is found that not only double contact signs, but basically all signs with a simple path movement stay within their location, provided that we interpret distinctive location as an area rather than a point. In NGT there appear to be no mono-morphemic signs that, for instance, start out with contact on the chin and then touch the nose, or signs that start out near the left eye and then go to the right ear. For NGT, the generalization thus holds that *all simple path movements* take place within one location.

Further evidence for the head-dependent relation between location and setting comes from the observation about the size of path movements already made with reference to (3). The *size* of the movement seems to correlate with the major location where movement is made (Battison 1978). For instance, a linear horizontal movement on the trunk is larger than a linear horizontal movement on the chin, and a circular movement on the weak hand has similar size as a circular movement on the cheek. In the representation proposed here, this correlation is formally accounted for by a head-dependent structure (cf. van der Hulst 1993, 2000).[8] Setting is dependent on Location, thus expressing the observation that the size of path movements correlates with their location. If there were no formal account of (the size of) path movement in relation to location, one would need specifications for all sizes of movement. In the representation proposed here, the same setting pair representing a horizontal movement ([contra] to [ipsi]) will result in a larger path movement on the trunk (LATE in Figure 5b.) than for instance on the chin (MOTHER in Figure 5a.).

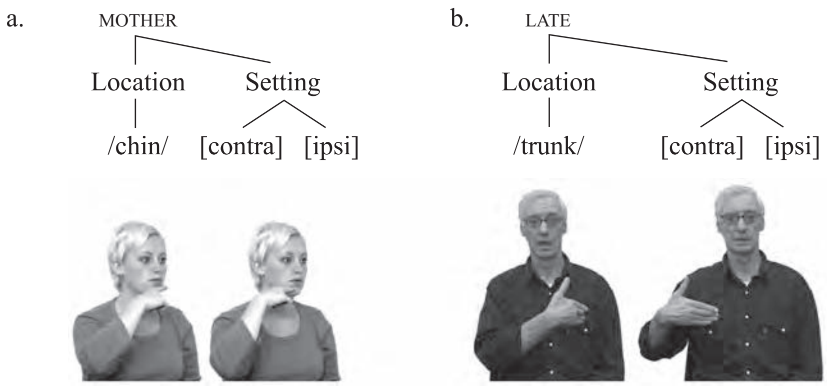


Figure 5. Partial representation of the NGT signs MOTHER and LATE

By formally linking location and movement it is implied that a distinctive location is not necessarily a specific point in space or on the body but rather an area, i.e. a set of phonetic points. This view of location as an area does more justice to the range of phonetic variation we find for certain signs. For example, the NGT sign NORMAL is made on the trunk and consists of a repeated downward brushing movement. The realization of the phonetic location of this sign ranges horizontally from the contralateral side (near the arm pit) to the center of the chest (the chest bone) and vertically from shoulder height to diaphragm-height. One even finds variants that are made in the space right in front of the diaphragm, without actually contacting the chest. In the Dependency Model all variants would be adequately described with the location feature [trunk] and the setting features [high], [low].

With this brief sketch of the Dependency Model in place, the next section focuses on one distinctive location, the trunk. The aim of this section is to illustrate the reduction strategies that were discussed in section 1. It will be shown how a small set of distinctive location features can be postulated when a proper role is given to phonetic implementation rules (to account for non-distinctive phonetic properties) and to the mechanism of semantic pre-specification (to account for meaning-driven phonetic properties).

3. A case study: the feature [trunk]

3.1. Location features

Traditionally, location (or place) is recognized as one of the major components of signs. In the different models that describe ASL phonology, we find an abundance of features for this dimension of signs. This is due, in part, to the fact that often every phonetic location that is touched by the articulator is adopted as a phonological specification in these models. In table 3 (page 277), adapted from van der Kooij (2002a), a comparison of the location features for ASL in various sign models is provided. The rightmost column lists the features resulting from an analysis of NGT that incorporates phonetic implementation rules and semantic pre-specification. Although these features are based on the description of NGT signs, the resulting proposal is put forward as a general hypothesis for sign language phonology in general.

Contrary to the models in the middle columns, there are only a few distinctive locations in the new proposal, which is actually quite close to that of Stokoe (1960). In Figure 6, a schematic summary of the needed place specification is offered.

The locations printed in italics are considered to be *marked* vis-à-vis their complementary choices. In NGT, [head] is the most frequent body-related location (61%). This makes [head] the least marked body-related location. The recurrent observation that lesser marked locations allow the greatest array of subdivisions (cf. van der Hulst 1993) is born out by the fact that both [head] and [hand] have further subdivisions. In the case of [head], it is assumed that the location features [mid] and [high] can enter into a combination. Either feature can be in syntactic head or dependent position, resulting in a four-way distinction on the head. In the case of [hand], the palmside of the hand is most frequent and is represented by the feature [broad]. The thumb side of the hand, which is represented by the feature [narrow], is much less frequent. The backside of the hand is highly marked and is represented by a further specification of the feature [broad].

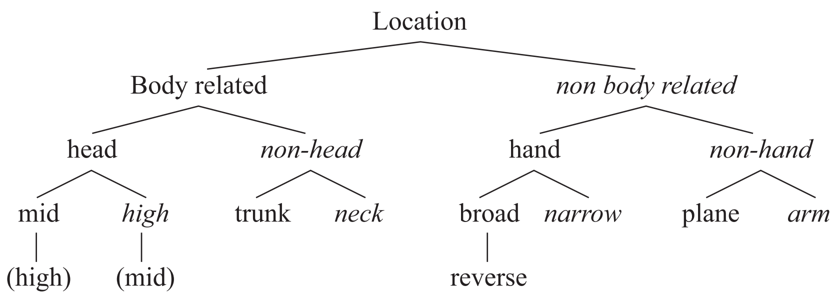


Figure 6. Schematic summary of place specifications for the Dependency Model.

The major distinction is between body-related and non-body related locations. This distinction correlates to a certain extent with linguistic function: the former are primarily relevant in the lexical domain, whereas the latter serve a more morpho-syntactic role.

In the typology of location, no feature is proposed to cover the phonetic property of 'neutral space', i.e. the space in front of the signer. Neutral space is the most frequent location in NGT lexical sings (71%). Given that, in this study, lexical frequency is considered to be an important indicator of markedness, neutral space is taken to be the *literally* unmarked location (i.e *no* location). This representation of neutral space signs does not preclude that the articulator *can* move with respect to specific locations in neutral space. These locations can be provided by other components of the grammar. The morphosyntax of agreement may provide the locations in neutral space that are associated with the arguments of a verb. Also the spatial expression of time ('time lines', see Schermer & Koolhof 1990) may demand the use of specific locations in neutral space. However, the present model recognizes *one* distinctive non-body related location in neutral space on phonological/phonetic grounds. Based on a survey of movement types, a virtual horizontal plane in space is assumed. End-hold movements in space can be analyzed in the same way as movements ending in contact with some body location, if one assumes that the end hold is caused by a (virtual) end contact with a (virtual) horizontal plane.

Table 3. Overview of location specifications of various ASL models

Stokoe (1960)	*Sandler (1989)*	*Brentari (1998)*	*Liddell&Johnson (1989)*	*Van der Kooij (2002)*
neutral space	neutral space	3 planes		plane
Head	Head	head	top of head back of head	head:high
Forehead		forehead	forehead forehead-side	
Midface		Eye		head:mid=>high
side of face		Cheek/nose	cheek ear jaw	head:high=>mid
lower face		Upper lip	Nose	head;mid
		mouth	Lip	
		Chin	Chin	
Neck	neck	under chin neck	Neck	Neck
Trunk	shoulder	Shoulder	Shoulder	Trunk
	trunk	Clavicle torso-top torso-mid	sternum chest trunk	
		torso-bottom waist hip	abdomen	
			Leg	
Upperarm	arm	upper arm elbow front elbow back	upper arm	Arm
Forearm wrist back wrist front		forearm front forearm back forearm ulnar wrist back wrist front	Forearm	
weak hand (several shapes)	weak hand	weak hand (8 locations)	weak hand (38 locations)	weak hand (palm, back, side)
13+	78[9]	32+	82[10]	11

3.2. The trunk

Stokoe (1960) proposed that no phonological distinctions are made on the trunk. The ASL signs that use the extreme upper (ANGEL, RESPONSIBILITY) and lower (RUSSIA) limits of the trunk are claimed to carry their contrast in hand configuration and movement only. However, unless all redundancies concerning the combination of handshape and location are investigated, this seems a rather *ad hoc* and unsatisfying solution.

Other researchers assume many distinctions on the trunk, mainly based on the observation that the articulator contacts the trunk in various locations. As an example of this 'phonetic' approach, Brentari (1998) assumes 28 potentially distinctive locations on the trunk, resulting from seven vertical level distinctions (shoulder, clavicle, top of the torso, torso-mid and torso-bottom, waist, hip) combined with two horizontal levels and two distinctions parallel to the body surface.

Following Stokoe, it is proposed here that there are no phonological contrasts among locations on the trunk. There is only one phonological specification: [trunk]. The actual realization of a location on the trunk can be predicted in terms of the phonetic implementation rules or semantic pre-specification. The phonetic implementation rules that account for the allophonic locations on the trunk need reference to the number of hands and to the relative orientation of the dominant hand. For some trunk locations, there is ample evidence for the role of semantic motivation. The use of extreme limits of the trunk, shoulders and waist, especially, will be shown to be controlled by semantics.

3.3. Predictability in terms of phonetic implementation rules

Based on a review of the signs in the SignPhon database, some tendencies were found in frequencies of occurrence of various phonetic distinctions. As can be seen in Table 4, most signs on the body are made in the center of the chest. For one-handed signs, this rate is even higher. The center of the chest is thus considered to be the *default* implementation, at least for one-handed signs. An articulatory reason may be that the center of the chest is the easiest position for the articulator to touch if it is lifted up while keeping the elbow in neutral (low, non-raised) position.

Two-handed signs can be divided into *balanced* and *unbalanced* signs (van der Hulst 1996). In unbalanced signs, there is one active hand; the passive hand serves as the location for the active hand. Balanced signs involve two identical articulators moving in tandem or in an alternating manner. In balanced signs, both on the body and in space, the default location of the strong hand is on the ipsilateral side of the midsagittal plane. In general, the non-dominant hand is placed in the exact mirror location, the mirror being placed at the midsagittal plane. As the location for the weak hand can be predicted in balanced signs, location in these signs is represented by specifying only the location for the strong hand.[11] However, this observation needs some refinement. If the two hands make contact with the chest at the end of the movement, the hands tend to be located in the center of the chest, one hand above the other. An example of such a sign is LONELY.

Table 4. Distribution of phonetic locations of NGT signs made on the trunk

Trunk	389	8% of all signs in SignPhon
upper chest, center	182	47%
upper chest, ipsi	77	20%
upper chest, contra	61	15%
lower chest, center	28	7%
lower chest, ipsi	16	4%
waist	14	5%
top of shoulder, ipsi/contra	7	2%
lower chest, contra	4	1%

We can summarize as follows:

a. In balanced signs, the articulators tend to stay on their own side of the midsagittal plane.
b. In balanced signs *with end contact*, the articulators tend to touch the chest in the center, one hand above the other.

The type of contact thus influences the exact location on the trunk. Another generalization concerning contact is made by Battison (1978) for ASL. He noted that contact with the contralateral side of the body is more marked than contact with the ipsilateral side of the head or body. In general, this also holds for NGT. Although one finds signs involving both contralateral and ipsilateral contact, single-setting contralateral contact on the head indeed never occurs. On the trunk, contralateral contact is marked for both one-handed and two-handed balanced signs. However, there is a group of signs that refute this tendency. These signs can be accounted for if one notes which part of the hand has contact with the location. The contacting part of the hand is subsumed under what is called the *relative orientation* of the hand. (Crasborn & van der Kooij 1997, Van der Kooij 2002) In signs that have the radial (or thumb-) side of the hand contacting the body, the *contralateral* side of the trunk appears to be unmarked location. Examples are: ALSO, BROTHER, NORMAL, TO BORROW, PRIZE. This tendency seems to be phonetically grounded. Contacting the ipsilateral side of the trunk, or even the center of the chest, with the radial side of the hand would require more articulatory effort, since it is hard to accomplish without lifting the elbow and/or bending the wrist to an extreme position. While all other points of contact in one-handed signs seem to favor the center of the trunk, thumb-side contact thus favors the contralateral side of the trunk. To re-state:

> If the relative orientation is [radial] in one-handed signs, the articulator touches the contralateral side of the trunk.

As for two-handed signs, very few have relative orientations other than palm or fingertip. According to the general tendencies for balanced signs, the hands stay on the ipsilateral side of the body. There is one two-handed balanced sign from the *Groningen* variant[12] (KITCHEN) that has contact with the ulnar (or pinky-) side of the hand and a few signs (CHIEF, DUCK) that have contact with the radial side of the hand. These two-handed signs are made lower on the trunk. Presumably this is also due to ease of articulation. In order to contact the trunk with the lateral side of the articulator, the lower arm has to rotate, but this seems less effortful in locations relatively low on the trunk.

Table 5. Summary of predictable locations on the trunk

Relative orientation	One-handed signs		Two-handed signs	
Palm	center	FRIDAY	ipsi	GLAD
radial	contra	ALSO	center	LONELY
fingertip	center	FINALLY-UNDERSTAND	ipsi	BEHAVIOR
Ulnar (few signs)	center	GUYOT[13]	lower ipsi	KITCHEN $_{\text{Groningen}}$[14]
Back/wrist (few signs)	center	POSSESSIVE-I$_{\text{VOORBURG}}$	ipsi	ACCEPT

Summarizing the noted refinements, it appears that by default, one-handed signs are located in the center of the trunk (neither ipsi- nor contralateral) whereas two-handed signs are made on the ipsilateral sides of the body. This observation can be rephrased as follows: if the sign is one-handed, the whole trunk is used, whereas if the sign is two-handed only one half of the trunk (i.e., ipsilateral of the midsagittal plane) is available.

Other relevant factors cannot, of course, be excluded at this point. First, Table 5 is not intended to be a complete account of phonetic predictability.

Moreover, for some combinations only one or two signs exist. In general, only very few signs make contact with the back or wrist part of the hand. Only one sign was found in this category, namely the *Voorburg* variant (cf. footnote 12) of the directional first person possessive MINE.

Second, these tendencies of NGT do not preclude the possibility that in other sign languages, whether in general or in specific sub-regions of the lexicon (e.g. in name signs), the ipsi-contra distinction is not predictable and would need a phonological specification. As David Perlmutter (personal communication) has pointed out, name signs in ASL do make use of the different sides of the chest; for this stratum, a distinctive use of the setting features [ipsi] and [contra] may be used. Another way of dealing with the locations of name signs could be by phonetic pre-specification. If distinctive features are typically used recurrently in the lexicon, one could propose that distinctions must be used more than once or twice in the lexicon in order to be accepted as phonological. Since name signs are typically unique, one could deal with the unique locations in these signs in a similar manner as semantically motivated locations, i.e. by phonetic pre-specification.

The next section discusses a set of one-handed signs that make contact with the palm and radial side of the hand. It will be shown how semantically motivated locations escape the phonetically motivated tendencies illustrated in this section.

3.4. Semantic pre-specification

With just one phonological distinction on the body – [trunk] – and the set of phonetic implementation rules, finer distinctions on the trunk have been accounted for. There are signs, however, that seem to escape the general tendencies. As for the height distinctions, the locations high and low on the trunk are infrequent (21 and 5 of the SignPhon signs, respectively). Almost without exception, these relatively high and low locations are semantically motivated.

Examples from SignPhon of signs that are made at shoulder height are: (UNDER-)SHIRT, COAT, SWEATER, BEAUTY-QUEEN, and PROFESSOR. These signs all refer to the shape of a piece of clothing and the way people wear it. Other signs made at shoulder height , which were collected separately, are: GYMNASTICS, BAG, TO-CARRY, FUNERAL/BURIAL, MILITARY-RANKS, TO-CARRY-RESPONSIBILITY, BROAD-SHOULDERED, SEAT-BELT, YOKE. Except for GYMNASTICS, all these signs relate to the shoulders or to carrying things (on the shoulders).

Signs from SignPhon that are made on the lower part of the body are: TROUSERS, UNDERPANTS/PANTIES, SKIRT, LIVER, ORGANS. Signs made on the lower body that were collected separately are: BELLY, FAT-ROLLS (ON THE BELLY), KIDNEYS, INTESTINES, APPENDIX/APPENDICITIS, SEX-ORGAN (M/F), ERECTION, SEXUAL-INTERCOURSE, PREGNANT, SICK, ABORTION, MISCARRIAGE, STERILIZATION (M/F), DELIVERY, CAESARIAN-SECTION, UMBILICAL-CORD, UMBILICAL-BANDAGE, CORSET, WRAP-UP-SKIRT. It is clear from the meaning of these signs that they are all related to the lower part of the body, either to the inside (the organs and functions thereof) or to the outside (clothes typically worn on the lower part of the body).

It is here proposed that the above semantically-driven locations (which do not necessarily provide a complete set) are accounted for in terms of phonetic pre-specification. If the meaning aspect 'lower part of the body' is involved, the location on the body is typically made on the lower part of the trunk. As this tendency concerns a larger set of signs it may be possible to formulate rules that predict the lower or higher part of the trunk on the basis of a meaning aspect that these signs have in common. Formulating these rules requires a semantic analysis of these signs that goes beyond the scope of this paper. Another set of signs show more isolated exceptions to the general tendencies. In these cases, the iconic phonetic location is unique to specific signs and no patterns exist.

We found two types of *unique* phonetic locations. As we showed in Table 5, the phonetic default location for one-handed signs with palm contact is the center of the chest (OF, TO FIND/TO CONSIDER, TO WANT, AFRAID). The sign SOLDIER, however, is made high on the ipsilateral side of the chest. We argue that this is due to the iconic base of the sign: the carrying of a gun. Likewise, the sign KIDNEY consists of a pointing sign toward the lower side of the trunk. The location in this sign is semantically-motivated in the sense that the pointing hand has a deictic function, pointing to the actual location of the kidneys. Similar signs are LIVER, STOMACH, APPENDIX, etc. Because in these cases the motivated relation of formal location to its referent is idiosyncratic, we propose to adopt a phonetic pre-specification of these specific locations.

Yet another set of semantically motivated locations on the trunk is the set of signs in the semantic field of feelings. These signs tend to be made in the center of the chest. Of the 299 signs made at the center of the chest, 119 (i.e., almost 40%) were related to 'feeling'. Examples are: DISAPPOINTED, PATIENT, SATISFIED, IRRITATED, SAD, ANGRY, TO LOVE, AT EASE, TO BE BOTHERED WITH SOMETHING, EGOISTIC, TO PITY, HAPPY, AFRAID, GUILTY, SORRY. It is clear, however, that there are signs not pertaining to the semantic field of feelings that are made in the center of the chest. Given that the center of the chest is already the default location, at least for one-handed signs, it may that the central location for signs associated to feeling is due simply to the phonetic default.[15] There are, however, some signs that provide evidence for the reality of the semantic tendency just noted. As was illustrated in Table 5, one-handed signs with radial contact tend to be located at the contralateral side of the trunk. Of these radial contact signs, there are at least three signs that also have the meaning component of 'feeling,' resulting in a potential conflict between the phonetic and the semantic tendency. These signs are: RELIEVED, IMPATIENT and RESTRAIN ONE'S FEELINGS. These signs all contact the center of the chest. This can be understood as evidence that there really is a semantic tendency, and that it is stronger than the phonetic tendency.

4. Conclusions

In this article, it has been argued that there is no need to 'burden' the phonology with long lists of place features. Hence, it is possible to agree with the more economical earlier proposals of Stokoe and Battison. However, the richness of phonetic locations must be accounted for and here it has been argued that this can be done in two ways: phonetic implementation rules that specify default locations and semantically-driven phonetic pre-specifications that universally override the phonetic default rules.

The lexical entry for a sign can thus contain a phonetic pre-specification alongside a phonological structure. Phonetic pre-specification can either be unique to a particular sign, or be an instance of a rule expressing a phonetic tendency that may have some degree of generality and may play a role in lexical innovation, i.e., the formation of new signs (cf. Brennan 1990). In addition to a potentially illuminating role in the study of lexical innovation, the current analysis allows the formulation of hypotheses regarding the historical development of signs. If semantic motivation reduces over time, phonetic implementation rules are predicted to take over.

Notes

1. This, of course, leaves open the possibility that the phonological form of a particular morpheme consists of one single feature, such as the feature [nasal] for

the expression of 'first person' in the South American language Terena (Akinlabi 1996).

2. A phonetic property that is predictable in one language, may, of course, later be found to be distinctive in the context of an analysis of another language. It is assumed here that a phonetic property is non-distinctive until shown otherwise.
3. Contrary to other research on iconicity in sign language, it is not assumed here that the vocabulary of a sign language can be partitioned into non-overlapping sets of iconic signs and non-iconic or arbitrary signs. Van der Kooij (2002) demonstrated for NGT that semantic motivation (including iconic motivation) may occur at all levels of description, also at the lowest level of features. This implies that signs consisting of otherwise arbitrary features may contain a semantically motivated form element, e.g. an iconically motivated location feature.
4. In this article iconic motivation is understood as a subtype of semantic motivation. In iconic motivation, the relation between the linguistic form and its referent is analogous or diagrammatic. The form may refer to its referent directly or through metaphor. Included here are also deictic signs referring to body parts (e.g. KIDNEY) and cases in which the semantic motivation of the form is mediated through some other level of symbolization. (e.g. 'loan' handshapes; handshapes representing number or letters from a surrounding spoken language). In the latter case one may argue for calling these form-meaning associations morphemes.
5. These may themselves have an infinite array of phonetic realizations.
6. It is, of course, highly questionable that the form shape of spoken languages is exclusively arbitrary. To claim that would ignore many meaning-driven phenomena that usually go under the heading of onomatopoeia and sound-symbolism. If the present proposal to pre-specify certain phonetic properties as extra-phonological is tenable, a similar proposal might well be motivated for certain aspects of the 'phonology' of spoken languages.
7. Van der Hulst (2000) offers a comparison of structures in the signed and the spoken modalities. In the present article no attention will be given to parallels, deep or superficial, between sign language and spoken language phonological structures.
8. A different sort of formal link between location and movement is proposed in Uyechi (1996). She developed a geometric model in which the size of the lexical movement is relative to (the size of) the location.
9. The 78 distinctive locations results from multiplying the six major locations in this list by the potential setting specifications (in Table 2) within these major locations.
10. The 82 distinctive locations are taken from Liddell & Johnson (1989). For some locations in the table, diacritics are added to specify the precise setting of the articulator within the location.
11. If the hands cross the midsagittal plane, this is specified by the feature [cross] in the manner node. An example of such a sign in NGT is AUSTRIA.

12. NGT, at least at a lexical level, has several regional varieties or 'dialects' that originate from Deaf educational institutions located in various sub-regions of the Netherlands.
13. GUYOT is a proper name and refers to the name of the Deaf Institute in Groningen.
14. The subscipts Groningen and Voorburg indicate the geographical region where the sign originated or is used.
15. In principle, the signs in the default location could also be semantically motivated to be produced here. Of the 299 signs made in the default location on the trunk, almost 40 % could be related to 'feeling'. This leaves 60% unaccounted for by semantic motivation. Thanks to an anonymous reviewer for bringing up this important point.

References

Akinlabi, A.
1996 Featural affixation. *Journal of Linguistics* 32, 239–289.
Battison, R.
1978 *Lexical Borrowing in American Sign Language*. Silver Spring, Md: Linstok Press.
Boyes Braem, P.
1981 Distinctive Features of the Handshape in American Sign Language. Doctoral dissertation, University of California, Berkeley.
Brennan, M.
1990 Word Formation in British Sign Language. Phd thesis. University of Stockholm.
Brentari, D.
1990 Theoretical foundations of American Sign Language phonology. Doctoral dissertation. Unversity of Chicago, Chicago.
1998 *A Prosodic Model of Sign Language Phonology*. Cambridge, MA: MIT Press
Crasborn, O.
1998 *SignPhon software* (version 2.22). Nijmegen: http://www.let.kun.nl/sign-lang/signphon2.html
2001 Phonetic Implementation of Phonological Categories in Sign Language of the Netherlands. PhD thesis, University Leiden Centre for Linguistics.
Crasborn, O., H. van der Hulst and E. van der Kooij
1998 SignPhon 2 manual. Working draft. Leiden University.
2001 SignPhon: A phonological database for sign language. *Sign Language and Linguistics* 4, 1/2, 215–228.

Crasborn, O. and E. van der Kooij
1997 Relative orientation in sign language phonology. In: J. Coerts & H. de Hoop (eds.) *Linguistics in the Netherlands 1997.* Amsterdam: John Benjamins Publishing Company, 37–48.

Hulst, H.G. van der
1993 Units in the Analysis of Signs. *Phonology* 10, 209–241.
1996 On the other hand. *Lingua* 98, 121–143
2000 Modularity and modality in phonology. In: N. Burton-Roberts, P. Carr and G. Docherty (eds.). *Phonological Knowledge: Its Nature*. Oxford: Oxford University Press, 207–244.

Kooij, Els van der
2002 Phonological categories in Sign Language of the Netherlands; Phonetic implementation and iconicity. PhD thesis. Leiden University. [Published by LOT, Netherlands Graduate School of Linguistics; http://www.lot.let.uu.nl]

Liddell, S. K. & R. Johnson
1986 American Sign Language Compound Formation Processes, Lexicalization and Phonological Remnants. *NLLT* 4, 445–513.
1989 American Sign Language: The Phonological Base. *Sign Language Studies* 64, 197–277.

Perlmutter, D. M.
1992 Sonority and Syllable Structure in American Sign Language. *Linguistic Inquiry* 23, 407–442.

Sandler, W.
1989 *Phonological Representation of the Sign: Linearity and Nonlinearity in ASL Phonology*. Dordrecht: Foris Publications.

Schermer, T. and C. Koolhof
1990 The reality of time-lines: aspects of tense in Sign Language of the Netherlands (SLN). In S. Prillwitz & T. Volhaber (eds.) *SLR'89*, 295–305. Hamburg: Signum.

Stokoe, W. C.
1960 Sign Language Structure. 2nd edition 1978. Silver Spring, MD: Linstok Press.

Uyechi, L.
1996 *The Geometry of Visual Phonology.* Stanford, CA: CSL Publications [from 1994 dissertation].

Variability in verbal agreement forms across four signed languages

Gaurav Mathur and Christian Rathmann

In signed languages, "agreeing verbs" change orientation and/or direction of movement according to targets established by an agreement rule and assigned to locations in signing space. Such verbs do not always match the targets. To determine the nature and source of the mismatches, this study investigated variability in three agreement forms across 70 to 80 verbs in four signed languages. One main finding is that two kinds of constraints determine whether the phonological form of a verb matches the target(s) imposed by the agreement rule: constraints on the degree and on the complexity of articulation. Another finding is that variability was systematic across particular targets for three agreement forms, across all verbs and signed languages. Taken together, the findings suggest that the constraints may be encoded within the phonological system of a sign language. The constraints are further discussed in terms of gestural scores within the framework of Articulatory Phonology: the constraint on degree of articulation is seen as an upper bound on a tract variable, while that on complexity is evaluated in terms of number of phasing relationships. Within the framework of Optimality Theory, which determines optimal output for a given agreement form, ranking of one markedness constraint above a faithfulness constraint determines the output according to degree of articulation, while ranking of a faithfulness constraint over another markedness constraint selects the output according to complexity of articulation.

1. Introduction

In signed languages, there is a set of verbs called "agreeing verbs" that change orientation and/or direction of movement according to "targets." Targets are established through agreement with the subject and direct object of a sentence in person and number features. Person features include first person, which refers to the signer, and non-first-person, which refers to people other than the signer (Meier 1990), while number features include singular and plural (Klima and Bellugi 1979). The targets are assigned during discourse to locations in the area in front of the signer ("signing space"), which represent the referents of the subject and the direct object. For example, to agree with a non-first-person singular subject represented by a location on the signer's right, and with a non-first-person singular object represented by a location on the signer's left, the ASL sign ASK changes its orientation so that the palm of the hand faces the location representing the direct object, as in Figure 1b. Compare with the citation form in Figure 1a. It also changes direction of movement so that the hand moves from the location representing the subject to the location representing the direct object. Similar forms of agreement have been noted in all the signed languages documented to date. See, e.g., Padden (1983), Meier (1982, 2002), Aronoff, Meir, & Sandler (2000), Lillo-Martin (2002), Rathmann & Mathur (2002), and Liddell (2003).

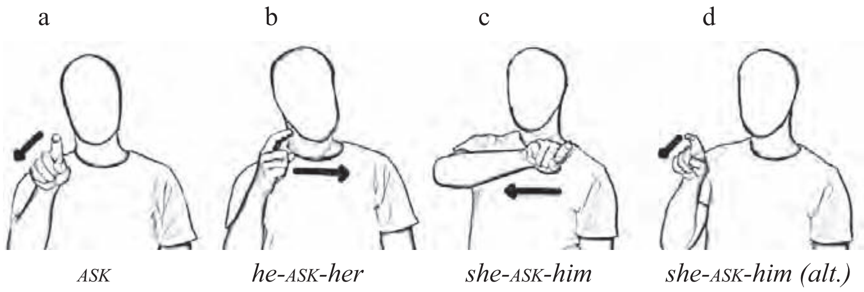


Figure 1. Forms of ASK, non-first-person singular. The hand is a fist with index finger extended, then bent, as it moves in a straight line.

Theoretically, targets may be assigned to an infinite number of locations. Since agreement is a grammatical process, an agreeing verb should match every possible target. As it turns out, this is not always the case. Figure 1c presents another example, where the locations for the targets in Figure 1b have been switched. That is, the location on the signer's left side now represents the subject and the location on the right side represents the direct object, as seen by the orientation and direction of movement of the sign. For an agreeing verb to agree with the targets in Figure 1c, a right-handed signer would have to twist the right arm so that the palm of the hand faces the right side and then move the hand from the left side to the right side. For many signers, twisting the right arm so that it faces to the right (and equivalently twisting the left arm so that it faces to the left) is articulatorily awkward. Loncke (1985) notes that this awkward configuration may be resolved if there is some compensatory movement such as supination of the arm that relaxes the arm; otherwise, if the movement is linear (moving the hand in a straight line without wrist action), the configuration remains difficult and is precisely what occurs in Figure 1c. In these cases, signers opt to leave the verb uninflected for the subject and add a pronoun (or an auxiliary-like element if available in the language) to clarify who the subject is. The palm of the hand still faces the location of the direct object but the hand no longer moves from the location of the subject; rather it moves from the chest, as in Figure 1d.

There is, then, variability in verb agreement forms in signed languages, depending both on the particular locations to which targets are assigned and on the phonological forms of the agreeing verbs, which include lexical specifications for handshape, orientation, location and movement (following Stokoe, 1960; Stokoe, Casterline & Croneberg, 1965; Battison, 1978). An instrumental study by Cormier (2002) has also identified variability in agreement forms in ASL. In particular, it revealed significantly more variability in plural forms than in singular forms. We are not so much interested in whether one agreement form, when produced over and over, is subject to greater variation than another, as we are in the issue of what allows some verbs but not others to express an agreement form in its entirety. Addressing this issue will help us to understand what makes a form articulatorily difficult, just as Mandel (1979) and Ann (1996) have done for handshape.

The present study pursues several questions regarding the nature of this variability. First, how systematic is the variability? Second, is it sufficient to derive the variability from the physiology of the articulators, or is it necessary to encode the variability in some way within the phonological system of a signed language? Next, if the variability is generated from within the phonological/phonetic system, to what extent can it be understood in terms of gestural scores in the framework of Articulatory Phonology (Browman and Goldstein 1989, 1992), and in terms of markedness constraints in the framework of Optimality Theory (Prince & Smolensky 1993, McCarthy & Prince 1993)? A related question is whether the nature of variability is the same across all signed languages. Since the forms of agreement are uniform across sign languages (Supalla 1997, Rathmann & Mathur 2002), we expect that variability in agreement forms is also uniform across sign languages.

To address these issues, a systematic study was undertaken of the variability by examining agreement forms across different sets of targets, across different verbs, and across different sign languages.

2. Methods

The data collection for this study followed the methodology of Mathur & Rathmann (2001), as detailed below.

2.1. Participants

There were eight participants in the study, two for each of the four signed languages under study: German Sign Language (Deutsche Gebärdensprache, DGS), Australian Sign Language (Auslan), Japanese Sign Language (Nihon Shuwa), and American Sign Language (ASL). These signed languages were chosen in particular for two reasons: (i) they are well-established signed languages that have been used by a stable Deaf community over the past century, and (ii) they are historically unrelated to one another and thus present a diverse sample for confirming that the variability of agreement forms is not a language-specific effect.

For each signed language, there was at least one female and one male participant, except for ASL, which was represented by two female participants. All participants ranged in age from 28 years through 50 years, with a mean age of 40 years.

Each participant was selected on the basis of three further criteria listed in Mathur & Rathmann (2001; 7). The first was "exposure to a signed language by the age of three." Mayberry (1993) and Boudreault and Mayberry (2000) have found that exposure by this age is sufficient to perform comprehension and sentential judgments on par with signers exposed to ASL from birth. All the participants met this criterion by being born deaf to deaf parents and/or by entering a residential school for the deaf by the age of three. A second criterion was the "capability to judge with ease whether or not a sentence is grammatical." This criterion ensured that participants had strong metalinguistic skills, which were required for the task. The last criterion was "daily contact with a signed language in the Deaf community for more than 10 years." This was to ensure that the participants were up-to-date on current usage of the signed language in question. Every participant was paid for his/her involvement in the test.

2.2. Stimuli

Stimulus verbs. To examine the variability of verb agreement forms in signed languages, a master list of 79 agreeing verbs with different meanings (as encoded by English glosses) was used. The master list was originally compiled by Mathur & Rathmann (2001:26) from dictionaries for Auslan (Johnston 1989) and BSL (Brien 1992). These dictionaries served as a point of departure, because Auslan was one of the languages under study and the BSL dictionary was used as well, since BSL is very closely related to Auslan (Johnston 2003, McKee and Kennedy 2000). In addition, these dictionaries were the most comprehensive sign language dictionaries available to us, with clear pictures and descriptions, and are organized according to the formational parameters of the sign, making it easier to identify signs. From these dictionaries, signs were collected that fit the criteria for an agreeing verb: transitive, with two animate arguments, generally with the roles of 'agent' and 'patient'. Verbs like 'meet', 'relate to', 'contact' or 'agree with' were excluded from the list, because they involved some inherent, mutual action between the two animate arguments which was expressed through a different form than prototypical verb agreement.

From this master list, a list of agreeing verbs was then created for ASL, DGS and Nihon Shuwa. With the assistance of a consultant who was a native, Deaf signer of the particular language, we gathered signs that approximated as closely as possible the meaning of the English glosses in the master list. This way, it was possible to compare agreement forms over verbs that had the same meaning across signed languages. It was not always possible to identify a sign that corresponded to each gloss on the master list, because a verb in one language sometimes had a different argument structure, i.e. the set of semantic roles that it assigns to subject and direct object, than the corresponding verb in another language, so that it was not an agreeing verb. As a result, the number of verbs elicited in each signed language was different, but always fell within the range of 70–80.

Stimulus targets. Next, to compare variability across agreement forms, three targets were selected, each expressing respectively the non-first-person singular subject form, the non-first-person plural object form, and the first person plural object form. This struck a balance between eliciting a feasible number of agreement forms across at least 70 verbs, and observing enough forms for each to examine the nature of variability. We selected the targets for their likelihood to induce variability based on the fact that their realization requires marked phonological forms, as will be made clear below, where each target is described.

The first target is one manifestation of the non-first-person singular subject form. Here, the subject is represented by a location on the contralateral side (i.e. the side opposite the dominant hand) and the object is represented by a location on the ipsilateral side (i.e. same side as the dominant hand). For a right-handed signer, this would mean the left and the right sides, respectively, as shown in Figure 2a.

For some verbs, this target is achieved by having the palm or the fingertips of the hand face the location representing the object. For other verbs, it is achieved by having the hand move from the subject location to the object location; for yet others, by doing both (Fischer and Gough 1978, Askins and Perlmutter 1995, Mathur 2000). Placing the dominant hand at a contralateral location and then twisting the palm of the hand so that it faces an ipsilateral location requires twisting the arm into an awkward configuration that might lead to an alternative form.

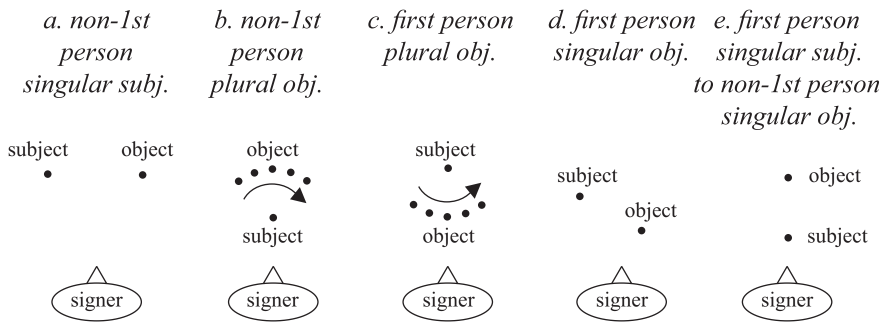


Figure 2. Targets for agreement forms.

The second target is a common way to express the non-first-person plural object form meaning ‘I *verb* you all.’ To achieve this target, the movement of the hand/arm is made in a horizontal, outward arc that proceeds from the contralateral side to the ipsilateral side, as schematized in Figure 2b. In other targets, the movement may proceed in the opposite direction, and the arc may be off to the side rather than directly in front of the signer. The choice of the particular target depends on discourse context. For the purpose of this study, we pick the most commonly used target for the non-first-person plural object form to examine how it interacts with the phonological form of the verb. Since the form requires a change not only in the direction but also in the shape of the movement, it has the potential to 'clash' with some verbs that require a specific movement shape. Thus, the form may induce variability in its expression.

The last target corresponds to the first person plural object form, which roughly means 'you (sg.) *verb* to us.' The target is similar to the target for the non-first-person plural object form, except that the movement is made near the signer's chest and is concave rather than convex, as depicted in Figure 2c. The target not only requires a specific shape of movement (arc-like), but also requires that the palm or the fingertips, depending on the verb, face the signer's body during the movement. This target then adds a layer of articulatory complexity that may lead to variability in the expression of an agreement form.

Control targets. To ensure that any variability in agreement form is due to phonological or phonetic factors, we confirmed that for each of the above stimulus targets, the verb is able to show another, similar agreement form. For the non-first-person singular subject form, the control target is a first person singular object form, as illustrated in Figure 2d. In this target, the location representing the object is different (i.e. the chest for first person singular), but the location representing the non-first-person singular subject is the same. If the verb can show the non-first-person singular subject in this form, it should be able to achieve the stimulus target as well.

The control target for the second stimulus target (the non-first-person plural object form) is the one for the non-first-person singular object form, as schematized in Figure 2e. In the control and stimulus targets, the same location represents the non-first-person object; they differ in whether the object is singular or plural. If the verb is able to show agreement with a non-first-person singular object but not with a non-first-person plural object, it can be deduced that the variability is phonological or phonetic.

The last stimulus target has two control targets. The first one is the first person singular object form, which is already seen in Figure 2d. The second one is the non-first-person plural object form, which happens to be one of the stimulus targets (see Figure 2b). If the verb can agree with a first person (singular) object and with a (non-first-person) plural object, it should agree with a first person plural object. It follows that any variability would be attributed to phonological or phonetic factors.

2.3. Procedure

For each signed language, an experimenter who was both Deaf and fluent in the language elicited all the data from the participants. For ASL, the experimenter was the first author; for DGS, it was the second author; and for Auslan and Nihon Shuwa respectively, it was the consultant who assisted in development of the list of agreeing verbs for that language.

The experimenter first explained to the participant that the goal of the task was to obtain different agreement forms (e.g., a partial conjugation) for each verb. Since agreement forms establish targets that are assigned to locations representing referents through the discourse process of nominal establishment, the experimenter instructed the participant to visualize that nominal establishment had been achieved according to visual aids, which showed an array of circles, each representing a person. One circle was marked by color as the signer, while another was marked with a different color as the addressee, and other circles representing other referents had no color. If the participant had difficulty visualizing the target(s), the experimenter added context by signing further descriptive information to clarify which particular target was desired, without showing the actual agreement form. In all signed languages, there is optional omission of subject agreement (Padden 1983). Thus, verbs that can show the target as subject do not have to. The participants were asked not to omit subject agreement and to try to produce forms that showed as much of the target as possible, as indicated on the visual aid.

After the participant practiced with an agreement form on four or five verbs that were not part of the study, the experimenter laid out the visual aid for the first agreement form. The experimenter then showed the citation form of the first agreeing verb on the list and asked the participant to produce the agreement form one time according to the targets suggested by the visual aid. After this, the experimenter showed the next verb on the list and asked the participant to show the same agreement form for that verb, and so on. After going through the entire list of verbs for the first agreement form, the experimenter repeated the procedure for each of the remaining agreement forms. The order of the agreement forms elicited varied across participants, but the order of elicited verbs was fixed to facilitate coding. During the entire elicitation session, the participant's production was video recorded for later coding.

2.4. Coding

Both authors coded the data from ASL, and split the coding of the rest of the data. As a first step, they checked for inter-participant reliability. If both participants for a given language produced the same agreement form for a verb, the form was included in the study; otherwise, it was excluded. For each form included in the study, the coders noted the handshape(s), orientation, location, manner, direction and shape of movement of the verb. Then, they noted whether the verb showed the expected agreement form. If it did not, the coders noted what form was used instead. There was high inter-coder reliability, greater than 96% agreement for the double-coded ASL data. Most coder differences occurred for handshape. Those were excluded, since they do not change under agreement. The focus, therefore, was on those parameters that change under agreement, i.e., orientation and direction of movement.

3. Results

The results for the non-first-person singular subject form and the first person plural object form are parallel to one another; these are reported first. The results for non-first-person plural objects are reported last.

3.1. Non-first-person singular subject form

Two variants were found for the non-first-person singular subject form. The first variant conforms to expectations, i.e. it achieves the targets for the subject and object. An example from ASL is seen in Figure 3a. In this example, the hands realize the subject target by starting from the contralateral side. They also realize the object target by moving toward its location ipsilaterally, with the fingertips pointing to this location. More examples of verbs from all the languages that show this variant are presented in Table 1. Appendix A provides illustrations of all signs listed in this table, as well as others.

Table 1. Variants of non-first-person singular subject form.

Language	*Verbs showing both targets*	*Verbs showing one target*
ASL	SHOW	GET-HOLD-OF
	HELP	PICTURE
DGS	BEEINFLUSSEN 'influence'	VERSPOTTEN 'tease'
	EINLADEN 'invite'	LEHREN 'teach'
Auslan	ANSWER	CONTACT
	HIRE	FLIRT
Nihon Shuwa	CHUUI SURU 'advise'	DAMASU 'deceive'
	HIHAN SURU 'criticize'	IU 'tell'

The other variant achieves the object target but fails to achieve the subject target. Figure 3b shows an example from ASL. The hands are not at the targeted location for the subject; instead, they are near the signer's chest. Otherwise, they achieve the object target by having the palms of the hands face its location. The subject target is expressed through a pronoun or, if available in the language, an auxiliary-like element. Additional examples of verbs from all the languages that display only this variant are provided in Table 1.

Table 2 shows how many verbs in each language can achieve both subject and object targets, and how many can show the object target.

Table 2. Results on variability in non-first-person singular form.

Language	*Number of verbs showing target for both subject and object (first variant)*	*Number of verbs showing target only for object (second variant)*	*Total*	χ^2	*p*
ASL	6 (10%)	56 (90%)	62	41.30	.01
DGS	8 (32%)	17 (68%)	25	03.24	.07
Auslan	21 (51%)	20 (49%)	41	00.02	*n.s.*
Nihon Shuwa	10 (32%)	21 (68%)	31	03.90	.05

The chi-square test shows that there are significantly more verbs (or just as many in the case of Auslan) that achieve only the object target than there are verbs that show both the subject and object targets.

Moreover, the coding reveals a correlation between the ability to achieve a target and the phonological properties of a verb. Specifically, all verbs that show both targets fall into one of two groups.

a. *he-*ABANDON*-her*

HND: both hands flat and spread
MVT: hands move slightly forward

b. *he-*ASK*-her*

HND: crooked index finger
MVT: hand moves in a straight line
& finger bends

c. *you-*ASK*-us*

HND: crooked index finger

MVT: hand moves in inward arc
& finger bends

d. *you-*ANALYSE*-us*

HND: crooked index &middle fingers spread
MVT: hands move in inward arc
& fingers wiggle

e. *I-*TEST*-them*

HND: both index fingers crooked
MVT: both hands move up and down repeatedly while proceeding in an arc

f. *I-*ASK*-them*

HND: crooked index finger
MVT: hand moves in arc & finger bends

Figure 3. Verb signs showing variability in agreement forms: (a) both targets in non-first-person singular, (b) one target in non-first-person singular, (c) both targets in first person plural object, (d) one target in first person plural object, (e) non-first-person plural object, without repeated movement, and (f) in non-first-person plural object, with repeated movement.

In one, the verbs change only in direction of movement, and not in orientation, to show agreement. (Recall that agreement can be manifested by a change just in orientation, or just in direction of movement, or both, depending on the verb.) In the other, verbs change in orientation to show agreement and have a 'mid' orientation: in citation form, the fingertips point away from the signer and the palms of both hands face sideways, as illustrated in Figure 3a. Verbs that show only the object target are all similar in that they require a change in orientation for agreement but do not have the 'mid' orientation, like the example in Figure 3b.

3.2. First person plural object form

Two variants were observed for the first person plural object form, depending on the verb. The first variant has the targeted arc, which marks the plural number of the object, and it also has the targeted change in orientation/direction of movement that marks the object as first person. One example from ASL is illustrated in Figure 3c, and more examples from ASL, DGS and Auslan are listed in Table 3. There are no examples from Nihon Shuwa, because the data does not have any marking for plural number that appears on the verb.

The other variant omits the target for the arc (i.e. it omits plural agreement) but preserves the target for the first person object, as seen through orientation and/or direction of movement toward the signer's chest. The target for the arc may be expressed separately through a pronoun meaning 'us'. One illustration from ASL is provided in Figure 3d, and further examples are given in Table 3.

Table 4 shows the distribution of verbs showing the plural number along with a change marking the first person object, and verbs marking only the first person object.

The most important result is that in all the signed languages under study, there are two sets of verbs, each showing a different variant of the agreement form. According to the chi-square test, ASL and Auslan have significantly more verbs showing the targets for both plural and first person object than verbs showing the targets for only the first person object, while DGS has the opposite pattern.

Coding for all of the sign languages also reveals a robust relationship between the ability to show a target and a phonological property of the verb. In all verbs that can show both targets, extension of the hand from the wrist or pronation of the radius-ulna are lacking, as exemplified by Figure 3c (see Appendix B). In contrast, in all verbs that show only the target for first person, there is extension of the hand from the wrist and/or pronation of the radius-ulna, as seen in Figure 3d.

Table 3. Variants of first person plural object form

Language	*Verbs showing both targets*	*Verbs showing one target*
ASL	SHOW	CHASE
	HELP	RESPECT
DGS	GEBEN 'give'	VERSPOTTEN 'tease'
	BEOBACHTEN 'look at'	LEHREN 'teach'
Auslan	ANSWER	ABANDON
	HIRE	REMIND

Table 4. Results on variability in first person plural object form.

Language	*Number of verbs showing target for both plural number and first person object (first variant)*	*Number of verbs showing target only for first person object (second variant)*	*Total*	χ^2	*p*
ASL	32 (84%)	6 (16%)	38	17.00	.01
DGS	7 (29%)	17 (71%)	24	04.17	.04
Auslan	15 (68%)	7 (32%)	22	02.91	.09

3.3. Non-first-person plural object form

There were also two variants for the non-first-person plural object form. One variant preserves the target for plural number in the form of an arc movement, as well as the target for the non-first-person object in the form of orientation/direction of movement toward the object's location. In addition, there is repeated movement, which refers to multiple movements of the arm/hand from a joint. This term covers both reduplicated movement and trilled movement. Reduplicated movement is countable (usually consisting of two instantiations of the base), while trilled movement refers to uncountable repetitions of the movement in reduced form (Padden and Perlmutter 1987, Brentari 1998). One example is the ASL sign in Figure 3e. Additional examples of this variant are seen in Table 5. As with the previous agreement form, Nihon Shuwa does not use the arc form of plural marking.

Table 5. Variants of non-first-person plural object form.

Language	*Verbs showing repeated movement*	*Verbs not showing repeated movement*
ASL	GET-HOLD-OF	HELP
	CHASE	RESPECT
DGS	LEHREN 'teach'	GEBEN 'give'
	VERSPOTTEN 'tease'	BEOBACHTEN 'look at'
Auslan	CONTACT	ANSWER
	REMIND	ABANDON

The second variant also preserves the target for plural number (arc movement) and that for the non-first-person object (orientation/direction of movement toward the location of the object). There is, however, no repeated movement. This variant is illustrated with an ASL sign in Figure 3f, and is further exemplified by other signs in Table 5.

Table 6 shows how many verbs show the target with repeated movement, and how many verbs show it without repeated movement.

Table 6. Results on variability in non-first-person plural object form.

Language	*Number of verbs showing target with repeated movement (first variant)*	*Number of verbs showing target without repeated movement (second variant)*	*Total*	χ^2	*p*
ASL	18 (47%)	20 (53%)	38	00.11	*n.s.*
DGS	10 (42%)	14 (58%)	24	00.67	*n.s.*
Auslan	11 (50%)	11 (50%)	22	00.00	*n.s.*

For each sign language, the chi-square test shows that there is no significant difference in frequency between each type of verb. It is at chance level whether the verb shows repeated movement or not.

Additionally, the coding reveals that all verbs that have lexically specified movement must show repeated movement along with the target, as in Figure 3e. Verbs that are not lexically specified for repeated movement but are specified for handshape change show both targets (as illustrated in Figure 3f), and verbs that are lexically specified for neither show only the target without repeated movement.

4. Discussion

The results show interactions between the phonological form of a verb and the expression of a target. If a verb requires a change in orientation for agreement and does not have a 'mid' orientation (see section 3.1), it achieves the target for the non-first-person singular form only by expressing the object target location. If a verb requires extension from the wrist or pronation of the radius-ulna, it achieves the target for the first person plural object by expressing only the object target. If a verb has no lexical specification for movement, it achieves the target for non-first-person plural object without repeated movement. Thus, interaction between the phonological form of the verb and a particular target can lead to variability in expression of verb agreement.

While details differ from one signed language to another, variability due to interaction between the verb and the target is seen in all the languages under study. For the non-first-person singular subject form, variability occurs in the same way for all the languages: some verbs express the target for both subject and object, while others express only the target for object. For all languages examined except Auslan, there are significantly fewer verbs expressing the target for both subject and object. For Auslan, the verbs are evenly distributed between expressing one or both targets. This suggests that compared to other signed languages, Auslan has relatively more verbs with a 'mid' orientation that allow the expression of both targets. Otherwise, for all the languages, if a verb does not have a 'mid' orientation, it expresses one target.

The pattern for the first person plural object form is similar. For ASL and Auslan, there are significantly more verbs that show both targets (plural and first person object), and for DGS, there are significantly more verbs that show only one (first person object). This again suggests something about the structure of the lexicon in each language, i.e. ASL and Auslan may have more verbs that lack wrist extension or radio-ulnar pronation and that therefore allow the expression of both targets, while DGS may have more verbs using wrist extension or radio-ulnar pronation that permit the expression of only one target. However, what is the same across all the languages is the interaction between the phonological form of the verb and the target: if there is wrist hyperextension in the verb, it results in partial expression of the agreement form.

Variability is also present in all the languages with regard to the non-first-person plural object form. In fact, all are similar in that there is roughly an even distribution between verbs that have a lexically specified movement and thus show the target with repeated movement, versus those that do not have specified movement and do not show repeated movement (cf. Sandler 1996, Brentari 1998, and Mathur 2000). What may vary across the languages is the proportion of verbs that can show either variant, i.e. those that involve handshape change. Otherwise, the pattern of variability is uniform across these languages.

Because the reasons for variability in agreement forms hold for all these signed languages, it is reasonable to posit that the physiology of the articulators may account for the variability. A pilot study by Rathmann and Mathur (1999) and a fuller, ongoing study, however, hint that constraints are needed in the phonological system to derive the variability. In a procedure similar to the one reported here, hearing participants who learned ASL as a second language in college consistently produced agreement forms that achieved as much of the target possible even if it was articulatorily awkward, whereas native Deaf signers would not do so. Taking the present study, which establishes the systematicity of the variability of verb agreement forms, together with the pilot study, we are led to deduce that the constraints responsible for the variability of agreement forms are systematized from the physiology of the articulators into a phonological system.

We present here two ways to understand the phonological systematization of constraints, first within the framework of Articulatory Phonology (Browman and Goldstein 1989, 1992), and then within an Optimality Theoretic approach (Prince & Smolensky 1993, McCarthy & Prince 1993). Articulatory Phonology is particularly suited to accounting for the constraints because they are rooted in the articulatory system of hands and arms, and because it suggests that two different kinds of constraints may be responsible for the variability of agreement forms. Optimality Theory (OT) is also able to handle the two types of constraints. It has the additional advantage that if a form violates a constraint, the OT ranking of constraints determines which output is optimal, given that a constraint alone is not sufficient to determine which form is ultimately used.

Articulatory Phonology takes gestures to be the primitive phonological units, in contrast to segments or features. In the case of speech, *gestures* are characterizations of discrete, physically real events that unfold during speech production and are specified through *tract variables*, which are dimensions of a vocal tract constriction, e.g. lip aperture, tongue tip constriction location and tongue tip constriction degree, among others. It is possible to re-adapt the notion of gesture for sign language by assuming a different set of tract variables, as listed in Table 7. Then a *gestural score* can be generated that shows the arrangement of the tract variables during the production of a sign, as exemplified for the ASL sign ASK in Figure 4a. The arrangement shows two things: the degree of extension/flexion or supination/pronation for each variable over time, and the phasing (temporal) relationship between each variable. It is possible to measure the activity of each variable instrumentally, as has been done by Cheek (2001) for the knuckles and by Mauk (2003) for the wrist.

Table 7. Tract variables of a gesture in sign language.

Tract variable	*Definition*
Shoulder	extension/flexion of arm from shoulder
Humerus	supination/pronation of arm via upper part of arm
Elbow	extension/flexion of arm from elbow
Radius-Ulna	supination/pronation of arm via lower part of arm
Wrist	extension/flexion of hand from the wrist
Metacarpo-phalangeal	extension/flexion of hand from knuckles connecting fingers to palm
Interphalangeal	extension/flexion of hand from knuckles in middle of fingers

Constraints implicit within the gestural score then account for the variability in verb agreement forms. For example, in the non-first-person singular form, the ASL sign ASK cannot meet the subject target on the contralateral side because it is articulatorily difficult to twist the arm so that the palm of the hand faces the location representing the object while the hand is at the location representing the subject. This configuration can be prevented by placing an upper bound on the radio-ulnar variable, i.e. the radius-ulna can be pronated only so far. Imagine an axis that runs from the elbow to the fingers. Suppose, further, that zero degree is defined to be the state in which the fingertips point away from the body and the palm of the hand faces the contralateral side. From this state, the arm may be pronated up to 90 degrees around the axis, but not more. Since the target for the non-first-person singular subject form requires pronation exceeding 90 degrees (see Figure 4b), it does not happen. This can be understood as a constraint on the degree of articulation. This constraint is similar in form to Crasborn's (2001: 277) *gesture constraints*. The analysis for the variability of the first person plural object form is parallel. There is precedent for such a constraint on degree of articulation in sign language phonology, e.g. constraints proposed by Mandel (1979) and Ann (1996) that predict the articulatory difficulty of a handshape.

a. Citation form of ASK

Shoulder	
Humerus	
Elbow	extending
Radio-Ulna	pronated
Wrist	
Meta.	
Inter.	flexing

b. Nonfirst person singular form of ASK

Shoulder	
Humerus	
Elbow	extending
Radio-Ulna	pronated > 90 degrees
Wrist	extended
Meta.	
Inter.	flexing

c. Plural form of ASK w/out repeated mvt.

Shoulder	
Humerus	supinating
Elbow	extending
Radio-Ulna	pronated
Wrist	
Meta.	
Inter.	flexing

d. Plural form of ASK with repeated mvt.

Shoulder	
Humerus	supinating
Elbow	extending
Radio-Ulna	pronated
Wrist	
Meta.	
Inter.	flexing

Figure 4. Gestural scores for various forms of ASK in ASL. (a) Citation form, (b) non-first-person singular form meeting both targets, (c) non-first-person plural object form without repeated movement, and (d) non-first-person plural object form with repeated movement.

In addition to placing constraints on the degree of articulation, the gestural score also imposes constraints on phasing relationships that explain variability in the non-first-person plural object form. For instance, to produce the ASL sign ASK in that form, supination of the humerus is added to the gestural score for the verb, as in Figure 4c, to match the target for a horizontal arc movement. The sign requires bending the index finger at the distal knuckles as part of its lexical form, as notated by 'flexing' for the interphalangeal variable. This flexing movement may be repeated throughout the arc movement, as indicated by multiple boxes for the interphalangeal variable in Figure 4d. Each of these repeated movements have to be in phase with the other variables, while there is only a single phasing relationship between the interphalangeal variable and the other variables in Figure 4c. Since fewer phasing relationships are required for the form without repeated movement, this form is generally preferred over the complex form schematized in Figure 4d. This can be seen as a constraint on the complexity of articulation. Similar kinds of constraints have been proposed in the literature, e.g. Battison's (1978) constraints on the specifications for the two hands in a two-handed sign, Mandel's (1981) constraint against more than one set of selected fingers per sign, Sandler's (1989) constraint against more than two locations for a sign, and Brentari's (1998) constraint against more than two sequential movements in a sign.

In Optimality Theory, these two constraints, on degree of articulation and on complexity of phasing relationships, can be systematized as interactions between markedness and faithfulness constraints. Moreover, the interactions can predict which output is optimal. For instance, for the first person plural object form, there are two possible outputs in case the original form cannot be fully produced: either drop the target for plural number, or drop the target for first person object. As it turns out, interaction between markedness constraints and faithfulness constraints determines that the target for plural number is dropped.

Figure 5 shows the analysis for the first person plural object form within OT framework. The first row shows the input form and the constraints that determine the optimal output. One faithfulness and one markedness constraints, when they are ranked in the order shown from left to right, are sufficient to yield the observed output. The markedness constraint ("*R-U Pron" for no radio-ulnar pronation) states that the radio-ulnar part of the arm cannot be pronated, and the faithfulness constraint (MAX) says: "Every target in the input has a correspondent in the output." A star indicates a violation of the constraint; two stars indicate multiple violations; an exclamation point denotes a "fatal" violation that rules it out as an optimal output; and gray shading marks those evaluations that follow a fatal violation.

The bottom three rows in Figure 5 show possible candidates for the optimal output. The first candidate shows both targets, but is ruled out due to a violation of the markedness constraint. The last candidate, which shows only the target for plural number, is ruled out due to two violations of the faithfulness constraint. One violation is due to the failure to achieve the object target, and the other violation is due to not achieving the plural target. It is the middle candidate that emerges as the optimal output (indicated by a pointing hand). Even though it violates the faithfulness constraint once for not expressing the target for plural number, it "wins" for not violating the two higher-ranked constraints. The same analysis applies as well to the variability of the first person plural object form.

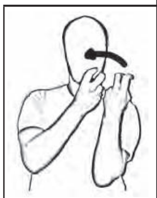
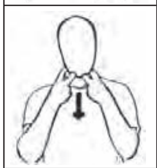
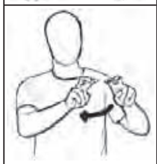
		*R-U-PRON	MAX
	Input: ANALYZE (crooked index & middle fingers) +target for first person object +target for plural number		
	a. ANALYZE (crooked index & middle fingers) +target for first person object (hands face body) +target for plural number (hands move in an arc)	*!	
	☞ b. ANALYZE (crooked index & middle fingers) +target for first person object (hands face body)		*
	a. ANALYZE (crooked index & middle fingers) +target for plural number (hands move in an arc)		**

Figure 5. OT analysis with constraint on degree of articulation

To account for the constraint on complexity of articulation, interaction between a different kind of markedness and a faithfulness constraint ensures that less complex movements are preferred if particular movements are unspecified in the lexical entry, as seen in the variability of the non-first-person plural object form. The faithfulness constraint is the same MAX constraint seen above. The markedness constraint at play, labelled as "≤2MOV", states that there cannot be more than two specifications for the movement parameter. Ranking the faithfulness constraint above the markedness constraint achieves the desired result for the non-first-person plural object form of ASL sign TEST, which has repeated movement.

Figure 6 illustrates the analysis for TEST, which has a lexical specification for repeated movement. The first candidate, which lacks repeated movement, violates the MAX constraint for deleting a specification from the input. Even though the second candidate, which contains repeated movement, violates the markedness constraint, it wins for obeying the higher-ranked MAX constraint.

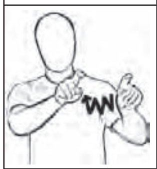
			MAX	≤2MOV
	Input: Mov:	TEST (both index fingers crooked) +target for nonfirst person object +target for plural number handshape change, arc, repeated movement		
	a. Mov:	TEST (both index fingers crooked) +target for nonfirst person object +target for plural number handshape change, arc	*!	
	☞ b. Mov:	TEST (both index fingers crooked) +target for nonfirst person object +target for plural number handshape change, arc, repeated movement		*

Figure 6. OT analysis with constraint on complexity of articulation, for verb with lexical specification for repeated movement.

5. Conclusion

In addition to establishing the systemacity of variability in three agreement forms across several signed languages, the study has identified two different kinds of constraints that are responsible for the variability. First is a constraint on the degree of articulation, which can be understood as an upper bound on a tract variable in Articulatory Phonology or as a ranking of a markedness constraint above a faithfulness constraint in Optimality Theory. The other kind of constraint is on the complexity of articulation, which can be conceived as a limit on the number of phasing relationships in Articulatory Phonology or as a ranking of a faithfulness constraint over a markedness constraint in Optimality Theory. Moreover, the Optimality Theoretic analysis shows that in order to determine the optimal output for a given target, it is necessary to have an agreement rule that specifies the target(s) and systematized constraints that determine which of the forms expressing the targets are possible. While the constraints may have different substantive content than analogous constraints in spoken languages, the existence of such constraints in signed languages suggests that a similar mechanism is at work in both modalities. That is, there is a phonological/phonetic system in place in each modality which ensures that actual forms are articulatorily easy and simple to produce.

Acknowledgments

We thank Scott Myers, Catherine Best, and two anonymous reviewers for their helpful comments. This project was supported in part by NIH HD-01994 to the Haskins Laboratories.

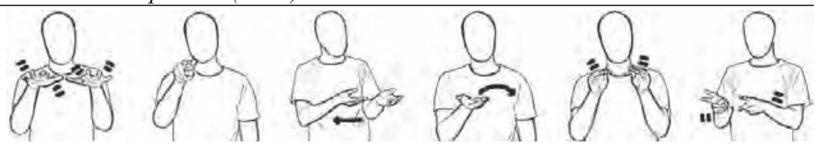
Appendix A: Illustration of signs

Key: HND = Handshape (handshape for dominant hand is listed, followed by handshape for nondominant hand)
MVT = Movement
Dom. = dominant
Nondom. = non-dominant

American Sign Language (ASL)

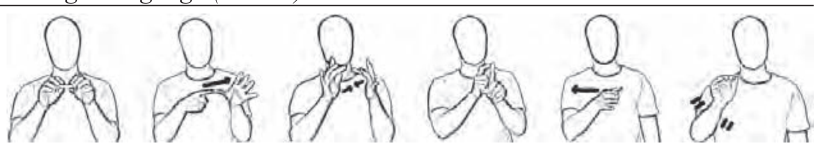
	CHASE	GET-HOLD-OF	HELP	PICTURE	RESPECT	SHOW
HND	fist; fist	index and middle fingers extended and crooked; index finger	fist with extended thumb; flat hand	lax flat hand; index finger	fists with extended and crossed index and middle fingers	index finger; flat hand
MVT	dom. arm rotates	dom. hand moves straight to contact	hands move straight	dom. hand opens and closes	hands move in downward arc	hands move away

Deutsche Gebärdensprache (DGS)



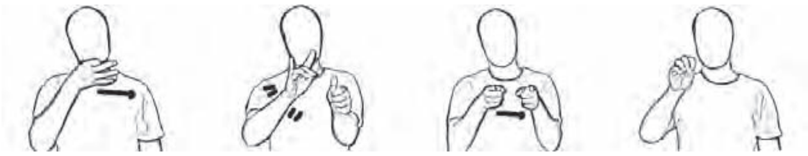
	BEEINFLUSSEN 'influence'	BEOBACHTEN 'look'	EINLADEN 'invite'	GEBEN 'give'	LEHREN 'teach'	VERSPOTTEN 'tease'
HND	both hands flat and all fingers spread	fist with index and middle fingers extended	both hands bent with thumb flexed in	flat hand	both hands bent with thumb touching fingers	both hands fists with index and middle fingers ext.
MVT	fingers wiggle and hands move back and forth	hand moves straight away from body	hands move in a straight line, and fingers close	hand moves in a slight arc	hands repeatedly move back and forth	hands repeatedly move back and forth

Australian Sign Language (Auslan)



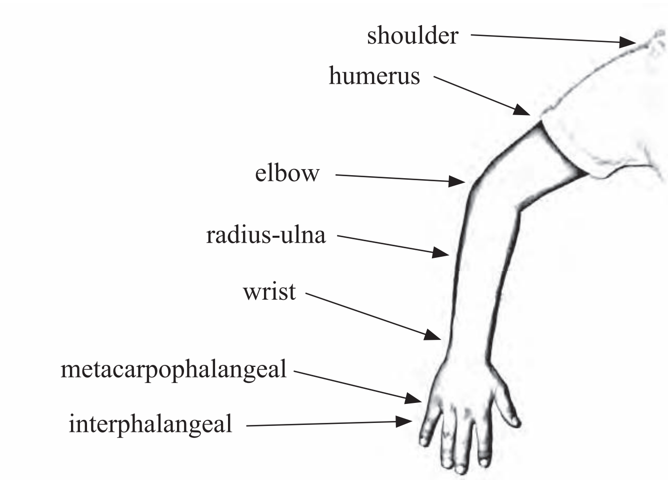
	ABANDON	ANSWER	CONTACT	FLIRT	HIRE	REMIND
HND	both hands curved with fingers touching thumb	index finger; flat hand with spread fingers	both hands flat, with fingers spread and middle finger bent	crooked index finger; index finger	curved hand	bent hand with thumb slightly extended
MVT	hands open while moving away from body	both hands move together in a straight line	hands move toward contact with each other	dom. finger wiggles up & down on nondom. finger	hand moves in a straight line toward body	hand repeatedly moves away from body

Japanese Sign Language (Nihon Shuwa)



	CHUUI SURU 'advise'	DAMASU 'deceive'	HIHAN SURU 'criticize'	IU 'tell'
HND	index finger and thumb touch, while others are extended and spread	index and pinky extended, with others touching thumb; fist	both hands extend finger pointing away from body	curved hand with fingers touching thumb
MVT	hand moves away from chin	dom. hand rotates in front of nondom. hand	hands move away from body	hand opens up and moves away from body

Appendix B: Anatomy of the arm



References

Ann, Jean
1996 On the relation between ease of articulation and frequency of occurrence of handshape in two sign languages. *Lingua* 98: 19–41.

Aronoff, Mark, Irit Meir and Wendy Sandler
2000 Universal and particular aspects of sign language morphology. In: Kleanthes Grohmann and Caro Struijke (eds.), *University of Maryland Working Papers in Linguistics* 10, 1–34.

Askins, David and David Perlmutter
1995 Allomorphy explained through phonological representation: person and number inflection of American Sign Language. Paper presented at the annual meeting of the German Linguistic Society, Göttingen.

Battison, Robbin
1978 *Lexical Borrowing in American Sign Language*. Silver Spring, MD: Linstok Press.

Boudreault, Patrick and Rachel Mayberry
2000 Grammatical processing in American Sign Language: effects of age of acquisition and syntactic complexity. Poster presented at 7th Congress on Theoretical Issues in Sign Language Research (TISLR), Amsterdam.

Brentari, Diane
1998 *A Prosodic Model of Sign Language Phonology*. Cambridge, MA: MIT Press.

Brien, David
1992 *Dictionary of British Sign Language/English*. London: Faber and Faber.

Browman, Catherine and Louis Goldstein
1989 Articulatory gestures as phonological units. *Phonology* 6: 201–251.
1992 Articulatory Phonology: an overview. *Phonetica* 49: 155–180.

Cheek, Adrianne
2001 *The phonetics and phonology of handshape in American Sign Language*. Ph.D. dissertation, The University of Texas at Austin.

Cormier, Kearsy
2002 *Grammaticization of indexic signs: how American Sign Language expresses numerosity*. Ph.D. dissertation, The University of Texas at Austin.

Crasborn, Onno
2001 *Phonetic implementation of phonological categories in Sign Language of the Netherlands*. Ph.D. dissertation, Universiteit Leiden.

Fischer, Susan and Bonnie Gough
1978 Verbs in American Sign Language. *Sign Language Studies* 18: 17–48.
Johnston, Trevor
1989 *Auslan Dictionary: A Dictionary of the Sign Language of the Australian Deaf Community*. Victoria, Australia: Australia Print Group.
2003 BSL, Auslan and NZSL: three signed languages or one? In: Ann Baker, Beppie van den Bogaerde, Beppie and Onno Crasborn (eds.), *Cross-linguistic Perspectives in Sign Language Research: Selected Papers from TISLR 2000*, 47–70. Hamburg: Signum.
Klima, Edward and Ursula Bellugi
1979 *The Signs of Language*. Cambridge, MA: Harvard University Press.
Liddell, Scott
2003 *Grammar, Gesture and Meaning in American Sign Language*. Cambridge: Cambridge University Press.
Lillo-Martin, Diane
2002 Where are all the modality effects? In: Richard P. Meier, Kearsy Cormier and David Quinto-Pozos (eds.), *Modality and Structure in Signed and Spoken Languages*, 241–262. Cambridge: Cambridge University Press.
Lombardi, Linda
1999 Positional faithfulness and voicing assimilation in Optimality Theory. *Natural Language and Linguistic Theory* 17: 267–302.
Loncke, Filip
1985 Sign phonemics and kinesiology. In: William Stokoe and Virginia Volterra (eds.), *Proceedings of the Third International Symposium on Sign Language Research*, 152–158. Silver Spring, MD: Linstok Press.
Mandel, Mark
1979 Natural constraints in sign language phonology: data from anatomy. *Sign Language Studies* 24: 215–229.
1981 *Phonotactics and morphophonology in American Sign Language*. Ph.D. dissertation, University of California at Berkeley.
Mathur, Gaurav
2000 *Verb agreement as alignment in signed languages*. Ph.D. dissertation, Massachusetts Institute of Technology.
Mathur, Gaurav and Christian Rathmann
2001 Why not *GIVE-US: an articulatory constraint in signed languages. In: Valerie Dively, Melanie Metzger, Sarah Taub and Anne-Marie Baer (eds.), *Signed Languages: Discoveries from International Research*, 1–26. Washington, D.C.: Gallaudet University Press.

Mauk, Claude
2003 *Undershoot in two modalities: evidence from fast speech and fast signing*. Ph.D. dissertation, The University of Texas at Austin.
Mayberry, Rachel
1993 First language acquisition after childhood differs from second language acquisition: the case of American Sign Language. *Journal of Speech and Hearing Research* 36: 1258–1270.
McCarthy, John and Alan Prince
1993 Prosodic Morphology I: Constraint interaction and satisfaction. Ms., University of Massachusetts at Amherst and Rutgers University, New Brunswick, NJ.
McKee, David and Graeme Kennedy
2000 Lexical comparisons of signs from American, Australian, British and New Zealand Sign Languages. In: Karen Emmorey and Harlan Lane (eds.), *The Signs of Language Revisited: An Anthology to Honor Ursula Bellugi and Edward Klima*, 49–76. Mahwah, NJ: Lawrence Erlbaum Associates.
Meier, Richard P.
1982 *Icons, analogues and morpheme: the acquisition of verb agreement in American Sign Language*. Ph.D. dissertation, University of California, San Diego.
1990 Person deixis in American Sign Language. In: Susan Fischer and Patricia Siple (eds.), *Theoretical Issues in Sign Language Research*, 175–190. Chicago: The University of Chicago Press.
2002 The acquisition of verb agreement: pointing out arguments for the linguistic status of agreement in signed languages. In: Gary Morgan and Bencie Woll (eds.), *Directions in Sign Language Acquisition*, 115–141. Amsterdam/Philadelphia: John Benjamins Publishing Company.
Padden, Carol
1983 *Interaction of morphology and syntax in American Sign Language*. Ph.D. dissertation, University of California, San Diego.
Padden, Carol and David Perlmutter
1987 American Sign Language and the architecture of phonological theory. *Natural Language and Linguistic Theory* 5: 335–375.
Prince, Alan and Paul Smolensky
1993 Optimality Theory: constraint interaction in generative grammar. Ms., Rutgers University, New Brunswick, NJ, and University of Colorado at Boulder.
Rathmann, Christian and Gaurav Mathur
1999 The linguistic status of joint-based constraints in signed languages. Paper presented at the annual meeting of the Chicago Linguistics Society (CLS), University of Chicago.

2002 Is verb agreement different cross-modally? In: Richard P. Meier, Kearsy Cormier and David Quinto-Pozos (eds.), *Modality and Structure in Signed and Spoken Languages*, 370–404. Cambridge: Cambridge University Press.

Sandler, Wendy
1989 *Phonological representation of the sign*. Dordrecht: Foris.
1996 Phonological features and feature classes: the case of movements in sign language. *Lingua* 98: 197–220.

Stokoe, William
1960 Sign Language Structure: An Outline of the Visual Communication System of the American Deaf. (Revised and reprinted in 1978 as *Sign Language Structure*.) Silver Spring, MD: Linstok Press.

Stokoe, William, Dorothy Casterline, and Carl Croneberg
1965 *A Dictionary of American Sign Language on Linguistic Principles*. Silver Spring, MD: Linstok Press.

Supalla, Ted
1997 An implicational hierarchy in verb agreement in American Sign Language. Unpublished manuscript, University of Rochester, New York.

Some current claims about sign language phonetics, phonology, and experimental results*

David M. Perlmutter

1. Introduction

This paper addresses the claims in three papers in this volume. The major issue is relevant for signed and spoken languages alike: How are we to distinguish what is phonetic from what is phonological? We pose this question for sign languages, taking as our point of departure two papers in this volume which draw the line between phonetics and phonology in ways we will challenge. The first, by van der Hulst and van der Kooij, claims that the specification of places of articulation on the trunk, previously thought to be phonological, is phonetic in Sign Language of the Netherlands. The second, by Mathur and Rathmann, proposes a phonological analysis of some gaps in morphological paradigms in four sign languages. We challenge both papers' claims, arguing that what van der Hulst and van der Kooij claim to be phonetic is phonological, and that what Mathur and Rathmann claim to require a phonological solution can be accounted for entirely in phonetic terms. We also discuss Brentari's claims based on an experiment that used signed stimuli.

Stokoe's (1960) pioneering monograph, which launched the study of sign language structure, identified three types of minimal contrasts in signs; signs could differ in handshape, in movement, or in location (place of articulation or POA), to which Battison (1978) added orientation. The potential opposition between *right* and *left* is not phonologically contrastive in any sign language yet investigated. Right-handed signers articulate 1-handed signs with the right hand, left-handed signers with the left. We call this hand the *strong* hand; the other one we call *weak*. A place of articulation on the same side of the face or body as the articulating hand is called *ipsilateral*; POA on the opposite side is called *contralateral*.

Some signs are articulated with one hand, others with two. In ASL there are minimal pairs that differ in this respect alone:[1]

(1) LIKE (1-handed) vs. INTERESTING (2-handed)

2. POA on the Trunk in NGT: Phonetics or phonology?

2.1. The Issues

Signs can be articulated at any of a number of different major places of articulation (POA): [Face/Head], [Trunk], [Arm], [(Weak) hand]. Van der Hulst and van der Kooij (this volume) address the representation of POA in the lexical entries of signs articulated on the trunk in Sign Language of the Netherlands (in Dutch, *Nederlandse Gebarentaal* or *NGT*). Earlier analyses have specified a large number of contrasting POAs on the trunk.[2] Hulst and Kooij (henceforth "HK") make these radical claims:

(2) a. Signs articulated on the trunk are phonologically marked [Trunk]; no further phonological features for POA are needed.
b. What distinguishes signs with different POA on the trunk is phonetic, not phonological.

HK's claims are important because they raise fundamental theoretical issues. How can one determine whether distinctions, the features that distinguish them, and rules that predict these features' values are phonetic or phonological? Since most research on sign language phonology has not asked what might be phonetic rather than phonological, HK's paper makes an important contribution in bringing these issues to the fore.

Among signs articulated on the trunk it is necessary to distinguish those articulated in the center of the trunk from those articulated on one or both sides; these two classes of signs can be distinguished by the feature [+/–Center]. It is also necessary to distinguish signs articulated on the same (ipsilateral) side of the body as the articulating hand from those articulated on the opposite (contralateral) side – a distinction that can be represented by the feature [+/–Ipsi].[3] Although they do not discuss these features *per se*, HK (Table V) use the same vocabulary (center, ipsilateral, contralateral) in describing POA on the trunk in NGT. If their proposals were made more precise, they would have to include these features, which would be claimed to be phonetic rather than phonological features.

HK's proposals are based on an analysis of over 3,000 signs in a data base of NGT signs in their citation forms. They offer these generalizations:[4]

(3) a. 1-handed signs tend to be articulated in the center of the chest.
b. 2-handed ("balanced") signs tend to be articulated on the sides of the trunk.

(3a–b) could be interpreted as *rules of phonetic implementation* for signs on the trunk:

(4) a. [–2-Handed], Place: [Trunk] → [+Ctr]
b. [+2-Handed], Place: [Trunk] → [–Ctr, +Ipsi]

Alternatively, instead of being rules of phonetic implementation, (4a–b) could be *redundancy rules* in the lexicon. Values of [+/–Ctr] and [+/–Ipsi] still would not need to be specified in lexical entries, for they would be predicted by (4a–b), simplifying the lexicon correspondingly.

Thus, two issues confront us. First, are (4a–b) rules of phonetic implementation or redundancy rules in the lexicon? Second, are [+/–Ctr] and [+/–Ipsi] phonetic or phonological features? Linking these issues yields two competing hypotheses about NGT grammar:

(5) *The Phonetics Hypothesis*
a. (4a–b) are rules of phonetic implementation.
b. [+/–Ctr] and [+/–Ipsi] are phonetic features.

(6) *The Phonology Hypothesis*
a. (4a–b) are lexical redundancy rules.
b. [+/–Ctr] and [+/–Ipsi] are phonological features.

Although we do not have access to NGT data, we argue that there are indications in HK's paper that the Phonology Hypothesis is preferable to the Phonetics Hypothesis for NGT, as it is for ASL.[5] In addition, HK failed to look for some of the key types of evidence that show the Phonology Hypothesis to be preferable for ASL. Thus, the case for the Phonetics Hypothesis in NGT has not been made.

2.2. Some reasons to doubt the Phonetics Hypothesis for NGT

The first argument for the Phonology Hypothesis comes from the fact that the generalizations in (3–4) are not absolute but merely tendencies; HK call these the “default implementations,” i.e. they have lexical exceptions. This violates what seems to be a valid generalization:

(7) *Phonetics without Lexical Exceptions (PWLE)*
Rules of phonetic implementation do not have lexical exceptions, i.e. lexical items that idiosyncratically fail to undergo them.

Given Lexical Phonology’s claim that post-lexical phonological rules do not have lexical exceptions, it would be bizarre if rules of phonetic implementation patterned with rules of lexical phonology in having lexical exceptions, while post-lexical phonological rules did not. Indeed, the exceptionless nature of rules of phonetic implementation seems to have been widely assumed; I am unaware of any such rules in the literature that have lexical exceptions.[6] Thus, PWLE (7) can tentatively be taken to be a principle of grammar.

Given that (4a–b) have lexical exceptions, if they were phonetic implementation rules they would violate PWLE. If, on the other hand, they are lexical redundancy rules that predict the values of phonological features in lexical entries, there is no problem. Exceptions can be marked with appropriate values of [+/–Ctr] and [+/–Ipsi] in the lexicon. Signs whose values of these features are correctly predicted by (4a–b) need not be lexically marked for these features, whose values will be supplied by the redundancy rules in (4). The net effect of (4a–b) is to simplify the lexicon by making it unnecessary to specify the values of these features in the lexical entries of many signs. PWLE thus provides an argument for the Phonology Hypothesis over the Phonetics Hypothesis in NGT.

Another argument could come from evidence for (6b) over (5b), i.e. signs whose values of [+/–Ctr] and [+/–Ipsi] are not predicted by (4a–b) and therefore must be marked in lexical entries. ASL exhibits two types of such evidence: first, the lexical exceptions to (4a–b), and second, signs that do not fall under (4a–b) but still must be marked for [+/–Ctr] and/or [+/–Ipsi]. All this is evidence that [+/–Ctr] and [+/–Ipsi] are phonological features in ASL (Perlmutter 2004).

One of the arguments that [+/–Ipsi] is phonological in ASL comes from the fact that despite the tendency for 2-handed signs to be ipsilateral, there are 2-handed signs in which each hand is contralateral, the arms crossing so the hands can contact opposite sides of the trunk. There are also both ipsilateral and contralateral 1-handed signs. Thus, a sign can be ipsilateral or contralateral independently of whether it is 1- or 2-handed; all four combinations of these two properties exist. This argues for [+/–Ipsi] as a phonological feature in ASL and hence for the distinctiveness of POA on the trunk.

From HK's paper we cannot tell whether there is parallel evidence for [+/–Ipsi] as a phonological feature in NGT. The existence of contralateral 2-handed signs might be masked by their feature [Cross] introduced for 2-handed signs in which the hands cross in violation of their generalization that each hand remains on its own side of the midsagittal plane. This does not tell us what is crucial: does each hand contact the contralateral side of the trunk? If so, the feature [Cross] does not account for this.[7]

The difficulty, in fact, goes deeper. Suppose HK claim there are no contralateral 2-handed signs in NGT; how do we know whether these are impossible signs or merely happen not to show up in HK's data base? The kinds of claims HK wish to make for NGT cannot be based solely on a data base, which cannot distinguish between accidental and systematic gaps in the data base, let alone the lexicon as a whole.

No investigation of whether a feature is phonetic or phonological can afford to ignore the question of whether there are minimal pairs that contrast in that feature alone. HK's limitation to a data base of signs, however, forces them to ignore this question. We do not know whether such minimal pairs exist but happen not to show up in HK's data base.

In fact, aspects of HK's analysis could obscure the existence of minimal pairs that would motivate [+/–Ctr] as a phonological feature marking the contrast between central and contralateral signs. HK note contrasts in NGT between central and contralateral POA on the chest. They note that in the contralateral signs the radial side of the hand contacts the chest, claiming that contralateral POA is predictable from radial contact in 1-handed signs and therefore phonetic, giving no argument against predicting radial contact from POA instead. Again we do not know whether the putative correlation is genuine or is due to the limitations of their data base.[8]

In ASL at least, there are minimal contrasts between central and contralateral POA on the upper chest – both in the general vocabulary and in arbitrary name signs (Perlmutter 2004, Supalla 1990). In a minimal pair in the general vocabulary, both contrasting signs have radial contact, so their contrasting values of [+/–Ctr] cannot be attributed to radial contact. Arbitrary name signs generate *sets* of minimal pairs with all the handshapes of the manual alphabet (including those that do not allow radial contact), so POA contrasts cannot be attributed to handshape or radial contact. One is left to wonder whether similar contrasts exist in NGT but happened not to appear in HK's data base.

HK's limitation to a data base weakens their case that POA on the trunk is phonetic in NGT in another respect as well: they failed to look for crucial evidence that would decide between the Phonetics and Phonology Hypotheses. For ASL, in addition to the evidence from minimal pairs, two additional types of evidence are adduced for the Phonology Hypothesis in Perlmutter (2004) that HK did not look for in NGT. First, outputs of the phonology must be *prevented* from undergoing (4a), which follows from the architecture of grammars without further stipulation if (4a) is a lexical redundancy rule rather than a rule of phonetic implementation. Second, the geometry of phonological representations in lexical entries makes correct predictions if [+/–Ctr] is a phonological rather than phonetic feature. HK did not ask any of the questions that could have led to such evidence because they limited themselves to a data base of signs.

2.3. Iconicity and indexicality in sign language grammars

Iconicity and indexicality are sign language phenomena that initially appear to have no counterparts in spoken languages and to require grammatical devices unlike anything needed for spoken languages, as well as an expansion of the set of POAs.

Contrary to the idea that iconicity has no counterpart in spoken languages, we propose that it is a form of borrowing – not from another language, but from the physical appearance of what the sign names:

(8) *The Loan Hypothesis*
Iconicity is a form of borrowing. Iconic signs in sign languages are loan words.

The Loan Hypothesis predicts that the phonology of iconic signs will be like loan word phonology in spoken languages in two respects. First, iconic signs may violate phonological patterns and constraints that hold in the native vocabulary. Second, they are predicted to change over time so as to conform to them.

These predictions are confirmed. Frishberg's (1975) classic study found that over time iconic ASL signs tend to become less iconic, i.e. nativized, like loan words in spoken languages. As they come to conform to ASL's phonological patterns and constraints, iconicity is lost.

If ASL is typical, most iconic signs, like many loan words in spoken languages, conform to the borrowing language's phonological patterns and constraints. For those that do not, two alternatives suggest themselves. Under one, non-conforming iconic signs would constitute a separate domain or stratum in the lexicon whose members are subject to constraints partially overlapping with and partially different from those that hold for the native vocabulary, as proposed by Itô and Mester (1995) for Japanese and Brentari and Padden (2001) for ASL. Alternatively, non-conforming iconic signs could stand outside the phonological system altogether. To decide whether to postulate a lexical domain for iconic signs in NGT, we would need to know (inter alia) roughly how many non-conforming iconic signs there are and which phonological constraints that hold for the native vocabulary they violate. Under neither alternative would iconic signs bear on phonological constraints that hold for the native vocabulary.

A class of signs that apparently has no direct counterpart in spoken languages are indexicals that indicate a body part by touching it or pointing to it so that the link between form and meaning is not arbitrary but direct. One way to deal with them in grammars would be to have a domain or stratum in the lexicon consisting solely of indexicals. This would have two advantages. First, it would eliminate the need to expand the set of phonologically relevant POAs for non-indexical signs beyond what is independently needed. Second, the meanings of indexicals need not be stated separately in each lexical entry, but can be stated once for the entire lexical domain: the meaning of each indexical sign is the body part at which it is articulated or which it points to. This would effect a considerable savings in the lexicon which would more than offset the cost of marking each indexical with a lexical feature indicating its membership in the class of indexicals.

Thus, neither iconicity nor indexicality requires an expansion of the class of POAs for the general vocabulary.

2.4. What is "phonetic prespecification?"

HK admit that POA on the trunk is not always predictable from other formational properties of NGT signs. For signs whose POA is not predictable they propose specifying POA in lexical entries – an uncontroversial move. What is controversial is their claim that such specification is phonetic ("phonetic prespecification"). On the key issue of whether such lexical specifications are phonetic or phonological HK provide no evidence whatever.

HK argue that where signs' POA on the trunk must be lexically specified this is due to their semantics. To show this, one would need to show that POA is predictable from semantic information in lexical entries, but this HK fail to do.

Consider the distinction between signs articulated high vs. low on the trunk. This can be accounted for by a feature [+/–High], as proposed for ASL by Sandler (1989), among others. To argue that this feature is "semantically driven," HK offer semantic generalizations, e.g. that signs referring to the shape of a piece of cloth and the way people wear it or which relate to carrying things (on the shoulders) are made at shoulder height. They intend the first generalization to encompass the signs BEAUTY-QUEEN and PROFESSOR, and the second to include FUNERAL, BAG, GYMNASTICS, and MILITARY-RANKS. HK fail to distinguish between vague associations with a sign's meaning and information in its semantic representation in the lexicon from which POA might be predictable. In particular, they fail to show that these signs' values of [+/–High] are predictable from their semantic representations.

Second, even if they could show that elements of signs' semantic representations could predict values of [+/–High], this would have no bearing on the key issue of whether this feature is phonetic or phonological. Nothing rules out redundancy rules predicting phonological feature values from semantic information in lexical entries. Thus, their proposal suffers from the vagueness of their notion "semantically driven," from their failure to predict POA from elements of signs' semantic representations, and from their failure to provide evidence that the features needed to specify POA on the trunk are phonetic rather than phonological.

Suppose, as seems likely, that POA in the cases HK discuss cannot be predicted from information in semantic representations. They would most likely still use phonetic prespecification for these cases, since they use it not only for signs whose POA they claim to be predictable by rule, but also for those where it is not. For example, HK use phonetic prespecification for SOLDIER, whose POA on the trunk can be accounted for under the Phonology Hypothesis with the phonological feature values [+High, +Ipsi]. HK claim this has an iconic basis (carrying a gun) but they indicate no mechanism that translates the claimed iconicity into feature values.

Since a very large number of signs have some kind of loose iconic basis, even if strict iconicity has been lost over time, the use of phonetic prespecification for such cases means that this is just a label for cases where HK are forced to admit that POA on the trunk must be specified in lexical entries.

The picture that emerges is that HK use phonetic prespecification as a catch-all for signs for which their strategy of predicting POA on the trunk from signs' other formational properties does not work.[9] This emerges even more clearly from their willingness to use phonetic prespecification for ASL name signs, which are generated productively by combining specific POAs with the handshapes of the manual alphabet (Supalla 1990, Perlmutter 2004).[10]

Name signs exhibit systematic contrasts between central and contralateral POA on the trunk, which combine with all the handshapes of the manual alphabet. This results in sets of minimal pairs for POA on the trunk, readily accounted for with the phonological features [+/–Ctr] and [+/–Ipsi] (Perlmutter 2004). There could be no stronger evidence that these features are contrastive and phonological. Using phonetic prespecification for POA in ASL name signs would mean that this device is merely a different name for specification of phonological features in the lexicon.

Further – and crucially – as throughout their discussion of phonetic prespecification, HK offer no argument that the features they admit need lexical specification are phonetic rather than phonological.

Overall, HK give two reasons for treating POA on the trunk as phonetic rather than phonological in NGT. They claim that in some signs POA on the trunk is phonetic because it is predictable, while in others POA on the trunk is phonetic because it is not predictable. That says it all. "Phonetic prespecification" is not an empirical result but a label for (a subset of) the cases where POA on the trunk must be specified in lexical entries.

While I have been highly critical of HK's proposals, their paper makes three important contributions. First, it raises the question of what is phonetic as opposed to phonological in sign language grammars. Scant attention has previously been paid to distinguishing the two. Second, HK have brought out hitherto unnoticed generalizations about the NGT lexicon, some of which may extend to other sign languages. Third, by pushing the idea that POA distinctions on the trunk are phonetic and by being forced to admit that at least some such distinctions must be specified in lexical entries, they have strengthened the case that at least some of what they claim to be phonetic is phonological.

3. Paradigm asymmetries due to phonetic gaps

3.1. The issue

In §2 we argued against a proposal that claims that specification of POA on the trunk is a matter of phonetics, not phonology. In §3, by contrast, we argue that data that Mathur and Rathmann (this volume) claim to be part of a phonological system can be accounted for solely in phonetic terms. From the standpont of Optimality Theory, the issue is whether phonetic and phonological constraints are combined into a single set of ranked constraints or separated into separate sets.[11] Even if there is evidence from other languages that phonetic and phonological constraints must be combined into a single set of ranked constraints, it is important to examine each purported case to determine whether it forces that conclusion. Whether the phonetic constraints are to be conceived of as phonetic implementation rules or as declarative constraints is not at issue, for the same issue arises under either conception: to what extent are phonetics and phonology separated or intermingled in grammars?

3.2. ASL Person Inflection: Singly and Doubly Inflected Verb Paradigms

MR present data they account for both in terms of Articulatory Phonology and through the interleaving of phonological markedness and faithfulness constraints in Optimality Theory. We discuss their first case in §3.3, showing that for ASL, a wholly phonetic solution is possible. In §3.4 we propose that sign language grammars specify which hand serves as active articulator in the phonetics, arguing that this can account for additional data MR did not consider. All the cases MR discuss involve inflectional paradigms.[12]

Key to the realization of person inflection in ASL verbs are the loci in the signing space of the verb's subject and object. Where a nominal's referent is physically present, the referent's position serves as its locus. Hence the locus for 1st person is the signer. 3rd persons whose referent is not present are assigned an arbitrary locus. The inflected forms in clauses with both subject and object come from one of two alternative person inflection paradigms.

The first paradigm I call the "doubly inflected paradigm," whose forms are inflected for person of both subject and object. ASL inflecting verbs divide into three classes, depending on how person inflection is realized. In the first class (e.g. GIVE, INFORM, HELP, SHOW), inflection is marked by direction of movement (from the subject locus toward the object locus). In the second class (e.g. PITY, OWE, BOTHER), inflection is marked by orientation (from the subject locus toward the object locus). In the third class (e.g. SEND, SAY-NO-TO, OK), inflection is marked by both direction of movement and orientation (from the subject locus toward the object locus).[13] ASL verbs INFORM and SEND are a near-minimal pair. In both, person inflection is realized as direction of movement, but in SEND it is realized as orientation in addition. Thus, SEND1:3[14] 'I send him/her and SEND3:1 's/he sends me' contrast both in direction of movement and in orientation, while INFORM1:3 'I inform him/her' and INFORM3:1 's/he informs me' contrast in direction of movement but have the same orientation.

The second paradigm is the "singly inflected paradigm," whose forms are inflected for person of the object only; the person of the subject is not expressed, the signer's position serving as an "empty" subject locus in the cases discussed here.[15] For the ASL verb SEND, for example, in the singly inflected form for a 3rd person object, the hand moves away from the signer toward the object locus while facing the object locus.

Thus, ASL grammar makes available to signers two verb paradigms:[16]

(9) a. Doubly inflected paradigm: inflected for person of both subject and object.
b. Singly inflected paradigm: inflected for person of the object only.

The availability of forms from both paradigms has a consequence:

(10) Forms from either paradigm can be used in clauses with both subject and object.

Padden (1988) described this situation in terms of subject agreement being optional.

3.3. An Apparent Gap in a Person-Inflection Paradigm

MR point out a gap in the doubly inflected paradigm of ASL verbs that realize person inflection through orientation alone (e.g. PITY) or through both orientation and direction of movement (e.g. SEND). If the subject and object are both 3rd person and if the subject locus is on the ipsilateral (strong hand's) side and the object locus is contralateral[17] (which we will call a "3I:3C" form), the doubly inflected form is well-formed. But if the subject locus is contralateral and the object locus ipsilateral (the 3C:3I form), MR note that getting the hand into the right orientation requires radio-ulnar pronation, which is articulatorily difficult. They claim such forms are ill-formed.[18] They also show that verbs that realize person inflection as direction of movement (and not orientation) have well-formed 3C:3I forms. They provide an articulatory description of the forms to be ruled out and propose the constraint:[19]

(11) *R-U-PRON: Radio-ulnar pronation is disallowed.

(11) is phonetic in that it refers to articulatory but not phonological information. MR claim, however, that it is a phonological markedness constraint, and that it must be ranked with respect to two phonological faithfulness constraints to get the right results. None of this is needed, however. *R-U-PRON accounts for the gap by itself.[20] It rules out the ill-formed 3C:3I forms because they require radio-ulnar pronation. No additional constraints or devices are needed. A form from (9b) inflected only for (3rd person) object will be available to signers as always, regardless of whether a form from the doubly inflected paradigm is available, as (10) makes clear. *R-U-PRON – a phonetic constraint – accounts for the gap in the doubly inflected paradigm. Variability in inflected forms exists independently of any articulatorily based phonetic constraints, as (9-10) make clear.

3.4. How the Gap is Filled: A Phonetic Account

Following MR, we assumed above that *R-U-PRON produces a genuine gap in the doubly inflected paradigm. We now argue that this slot in the paradigm is filled by a different doubly inflected form, based on additional data that MR did not consider.

Our argument is based on the observation that where inflection is expressed through orientation and 3C:3I forms are consequently ruled out by *R-U-PRON, a 3C:3I form can be articulated by the weak hand. This can be seen in the 3C:3I forms – articulated by the weak hand – of signs like PITY and SEND. Thus, there is no gap in the doubly inflected paradigm: the slot apparently left vacant by *R-U-PRON is filled by a sign articulated by the weak hand, which can be oriented toward the object locus and away from the subject locus without radio-ulnar pronation. It is necessary to account for the fact that the weak hand is used in the doubly inflected paradigm where use of the strong hand would violate *R-U-PRON – and only in those cases.

We posit *R-U-PRON and STRONG-ACTIVE as violable phonetic constraints:

(12) STRONG-ACTIVE: The strong hand (the right hand for right-handers, the left hand for left-handers) is used as active articulator.

*R-U-PRON must outrank STRONG-ACTIVE (or be undominated). In 3I:3C forms, the strong hand can be oriented away from the subject locus and toward the object locus without radio-ulnar pronation. Consequently, the candidate form articulated by the strong hand violates neither *R-U-PRON nor STRONG-ACTIVE, while the form articulated by the weak hand violates both constraints and is eliminated:[21]

(13)

SEND/PITY 3I:3C	*R-U-PRON	STRONG-ACTIVE
a. ☞ Articulated by strong hand		
b. Articulated by weak hand	*!	*

3C:3I forms articulated by the strong hand, however, violate *R-U-PRON. Although the weak-hand form violates STRONG-ACTIVE, the fact that the other candidate violates the higher-ranked *R-U-PRON makes it the winner:

(14)

SEND/PITY 3C:3I	*R-U-PRON	STRONG-ACTIVE
a. Articulated by strong hand	*!	
b. ☞ Articulated by weak hand		*

The constraint ranking guarantees that a weak-hand form will be well-formed if and only if the corresponding strong-hand form violates *R-U-PRON.

With a verb like GIVE, which realizes inflection through direction of movement only, orientation of the hand plays no role in inflection. Consequently, 3C:3I forms can be articulated by the strong hand without violating

*R-U-PRON. The weak-hand form, however, violates STRONG-ACTIVE and is eliminated:

(15)

GIVE 3C:3I	*R-U-PRON	STRONG-ACTIVE
a. ☞ Articulated by strong hand		
b. Articulated by weak hand		*!

Thus, this analysis accounts not only for the fact that weak-hand forms fill the gap resulting from *R-U-PRON violations, but also for the fact that this does not happen with verbs that do not use orientation to realize person inflection.

Thus, doubly inflected forms articulated by the weak hand fill what would otherwise be a gap in the doubly inflected paradigm (9a) caused by avoidance of radio-ulnar pronation. Singly inflected forms from (9b) exist alongside them. Thus, (10) still holds. The difference between weak-hand and strong-hand forms in (9a) lies in phonetic implementation, not in phonological representations, which say nothing about which hand is active articulator.

Importantly, STRONG-ACTIVE is needed independently to account for the fact that right-handers and left-handers use different hands as active articulator.

Our hypothesis potentially makes further predictions, e.g. that signs articulated by the weak hand will be well-formed in other cases where articulation by the strong hand would violate a higher-ranked constraint.

We have proposed a purely phonetic account of data MR claim to require a phonological one which accounts not only for MR's data, but also for data they did not consider.[22] Crucial to our solution is the interpretation of phonetic implementation rules as violable ranked constraints with STRONG-ACTIVE as a phonetic constraint whose violation yields signs with the weak hand as active articulator.

STRONG-ACTIVE is found in all sign language grammars, as far as I am aware. This has two consequences. First, they all should have analogs of the kind of interaction with higher-ranked constraints that has been illustrated here for ASL. The second stems from the ability of STRONG-ACTIVE to explain why sign languages' phonological feature systems do not encode the potential contrast between right and left, but rather that between ipsilateral and contralateral. If STRONG-ACTIVE is found in all sign language grammars, their feature systems should all be like ASL's in this respect.

4. Are there modality effects in word segmentation?

Brentari (this volume) makes strong claims about modality effects in word segmentation (parsing the speech or signing stream into words):

(16) a. There are strong modality effects in the word segmentation strategies used in signed and spoken languages.
b. There are strategies for segmenting visual language input that are different from those used in segmenting auditory language input.
c. These different strategies are due in large part to the visual/gestural nature of sign languages and to the auditory/vocal nature of spoken languages.

These rather sweeping claims are based on a single experiment. Subjects were shown nonsense signed stimuli and had to judge whether each stimulus consisted of one sign or two. Half the subjects were Deaf native signers, while the control group had no previous exposure to sign language. These were the basic results:

(17) Native signers used handshape cues in word segmentation, while non-signers did not. In other respects, the experiment did not detect any significant differences between native signers and non-signers. Both groups used POA and movement cues more heavily than handshape cues in making word segmentation judgments.

The first puzzle is why signers differed from non-signers in using handshape cues but not with respect to POA or movement. There are minimal pairs in ASL that contrast only in handshape, but this is also true of POA and movement. So why was handshape different?

A possible answer may be found in the degree of awareness of handshape in the signing community. Much of this is due to the large overlap between the handshapes in ordinary signs and those in the manual alphabet, which represents each English letter by a handshape. Signers' awareness of handshape can be seen in (18).

There is awareness of (18b), which is extremely common, and of (18c), since choosing a name sign for a new baby is a conscious process. There are also alphabet stories and language games in the Deaf community in which signers consciously manipulate signs' handshape as a variable. Non-signers, on the other hand, tend not to notice handshape in signs until it is pointed out to them, as anyone who has taught or attended a beginning sign language class can attest. Given all this, it is not surprising that native signers paid attention to handshape cues in word segmentation, while non-signers did not.

(18) a. Signers consciously use the manual alphabet to spell English words and names.
b. Many loan words from English are borrowed into ASL through a sign whose handshape represents the first letter of the borrowed English word.
c. Name signs in ASL use the handshape of the letter of the manual alphabet that corresponds to the first letter of the person's given name in English.

With respect to POA and movement cues, Brentari got what psychologists call a "null result:" the experiment failed to detect any difference between native signers and subjects with no previous exposure to sign language. This does not mean that there is no difference between the two groups, but only that this experiment failed to detect one. This could be due to the experimental design, the stimuli used, or any of a number of other factors. From a null result such as this, one cannot confidently conclude that there are no significant differences between the two groups. To show that would be far more difficult; it would be necessary to perform a power analysis, i.e. compute the probability that the experiment would have detected a certain (relatively small) difference (say, 10%) between the two groups if such a difference existed. This Brentari did not do.[23] Thus, Brentari's experimental results do not show that there is no difference between the two groups. Nonetheless, she assumes that they do. Let us call this the "no-difference assumption." Her interpretation of the results and her claims that there are modality effects in word segmentation are all based on this unsupported assumption.

The essence of her explanation is:

(19) *The Modality-Dependence Hypothesis*
a. Human beings use different strategies to parse signing and speech that have nothing to do with their linguistic knowledge.
b. These different strategies are due to the modality difference between signed and spoken languages.

Since the no-difference assumption is not supported by the experimental evidence, there is nothing to be explained. The Modality-Dependence Hypothesis is thus superfluous. We will now argue, however, that even under the unsupported no-difference assumption, there is at least one superior explanation.

Why doesn't signers' knowledge of ASL distinguish them from non-signers in the word segmentation task? An obvious initial hypothesis would be:

(20) *The Guessing Hypothesis*
Signers and non-signers alike were simply guessing.[24]

This would explain the supposed lack of difference between the two groups. When subjects with no previous exposure to sign language are shown signed stimuli and asked how many signs they consist of, what can they do but guess? By positing that the signers were guessing too, the Guessing Hypothesis explains why their performance was like the non-signers'. But why were they reduced to guessing?

A major finding of research on (a subset of) spoken languages is that prosodic cues are the most robust cues in word segmentation tasks. In Finnish, for example, where the major stress falls on the initial syllable of the word, stress has been shown to be the most robust cue in word segmentation (Vroomen et al. 1998). In sign languages, prosody is expressed on the face. For ASL, the face is divided into two zones, with prosodic information expressed in the upper zone – at eye level and above.[25] The eyebrows convey the kinds of information conveyed by intonation in spoken languages. The difference between a yes-no question and the corresponding declarative is that the question is signed with raised eyebrows. A furrowed brow is used in WH-questions. A constituent topicalized to initial position is signed with raised eyebrows. The prosodic information expressed in the face's upper zone is thus crucial for parsing the signing stream in ASL.

The fact that prosodic cues are most robust in word segmentation in spoken languages raises an important question:

(21) Why didn't Brentari's subjects use prosodic cues in word segmentation?

Brentari's explanation elaborates the Modality-Dependence Hypothesis, claiming that the unit of segmentation is the word in sign languages but the syllable in spoken languages. She further claims that this is related to supposed structural differences between signed and spoken languages. All this is the basis of Brentari's claims in (19). Generalizing from a single experiment with null results (except for handshape), which could be due to any of a variety of factors, she proposes the Modality-Dependence Hypothesis to answer (21).

There is, however, at least one alternative explanation, even under the no-difference assumption. Perhaps her subjects did not use prosodic cues in word segmentation because she selected stimuli that were prosodically neutral across the string; her stimuli were practiced before recording to ensure a neutral facial expression across each entire form. Consequently, there were no prosodic cues in her stimuli for her subjects to exploit. Perhaps they used other cues in the task because those were the only cues in her stimuli that subjects could base their guesses on. This explanation answers (21) without positing any modality differences whatever. Adopting it means there is no evidence supporting the Modality-Dependence Hypothesis.

Thus, if signers and speakers are alike in relying most heavily on prosodic cues in word segmentation of language input in each modality, there is no way Brentari's experiment could reveal this. The most straightforward explanation is that her subjects did not use prosodic cues in word segmentation because she gave them no opportunity to do so.

Let us briefly consider an alternative hypothesis:

(22) *The Modality-Independence Hypothesis*[26]
Signers use the same types of strategies in analyzing signing into words that speakers use in analyzing speech into words.

Since prosodic cues are the most robust in word segmentation tasks in spoken languages, the Modality-Independence Hypothesis makes a prediction:

(23) *The Prosodic Cues Hypothesis*
Both signers and speakers rely primarily on prosodic cues in word segmentation.

This can be tested with two experiments. First, test whether signers differ from non-signers in exploiting prosodic cues in signed stimuli. Second, suppress prosodic cues from the signed stimuli. Building on Brentari's result, one should also suppress cues based on handshape in order to get a clearer test of the effect of suppressing prosodic cues. The Prosodic Cues Hypothesis predicts that, with prosodic cues present, native signers will out-perform the control group with no previous exposure to sign language. However, with prosodic cues suppressed, native signers will perform no better than the control group with no previous exposure to sign language; both groups will be reduced to guessing.

Indeed, Brentari may have inadvertently performed an experiment similar to the latter. Her experimental results under the no-difference assumption are explained more parsimoniously by the Modality-Independence Hypothesis than by the Modality-Dependence Hypothesis. Her results with suppression of prosodic cues in the stimuli support the Prosodic Cues Hypothesis – a prediction of the Modality-Independence Hypothesis. This explains the similar performance of signers and non-signers as due to guessing, and it explains why native signers were reduced to guessing: because the prosodic cues on which they usually rely were absent in the stimuli. Most importantly, it does not require us to posit any modality differences whatever. There is no need to claim there are different parsing strategies for speech and signing that have nothing to do with linguistic knowledge, different units of word segmentation in signed and spoken languages, or different types of cues as inherently suited to speech and signing. These are all *ad hoc* constructs devised for the sole purpose of explaining the supposed lack of difference between signers and non-signers. The Modality-Independence Hypothesis needs no such *ad hoc* constructs. It exploits the finding (based on research on spoken languages) that the most robust cues in word segmentation tasks are prosodic cues. It need posit nothing more. It is a more parsimonious theory.

Much research needs to be done in order to learn how signers parse the signing stream into words. The Prosodic Cues Hypothesis could lead to the discovery of prosodic cues that signers use in word segmentation. The Modality-Independence Hypothesis is more parsimonious than the Modality-Dependence Hypothesis precisely because the latter needs to posit a series of *ad hoc* modality-specific constructs that the former does not need to posit. There are no doubt additional alternative explanations that could be considered. The superiority of the Modality-Independence Hypothesis to the Modality-Dependence Hypothesis, even under Brentari's unsupported no-difference assumption, is sufficient to show that the rather sweeping claims about modality effects in word segmentation embodied in her Modality-Dependence Hypothesis are not supported by the experimental evidence she cites.

Notes

* I am indebted to Eric Bakovic for valuable discussion, to Rachel Mayberry and especially John Wixted for discussion of experimental methodology, to Louis Goldstein for comments on an earlier version of this paper, and to Catherine Best for doing an extraordinary job as editor. None of these people are responsible for anything said here, for which I assume full responsibility.

1. We use the feature [+/–2-Handed] to distinguish such signs, assuming a rule of phonetic implementation that makes both hands active articulators in signs marked [+2-Handed]. These signs are called "2-handed" here and "balanced" in van der Hulst (1996) and van der Hulst and van der Kooij (this volume). The analysis of ostensibly 2-handed signs in which the weak hand is not an active articulator is controversial and is beyond the scope of this paper.
2. See Table III in van der Hulst and van der Kooij (this volume).
3. The distinctions captured by [+/–Center] and [+/–Ipsilateral] have been recognized in earlier research on ASL, but have not always been captured in the same way. For example, [+/–Ipsilateral] figures prominently in earlier work (Sandler 1989, Brentari 1998, among others), but Sandler posits [+/–Ipsilateral] and [+/–Contralateral] as separate features, interpreting [–Ipsilateral, –Contralateral] as central.
4. HK note two exceptions to these generalizations: first, signs in which the radial side of the hand(s) contacts the trunk have different POAs on the trunk (contralateral for 1-handed and central for 2-handed signs) and second, in 2-handed signs with end contact, the hands tend to touch the center of the chest, one above the other. (4) below would have to be modified to account for these exceptions.
5. Detailed arguments for rejecting the Phonetics Hypothesis in favor of the Phonology Hypothesis for ASL are given in Perlmutter (2004).
6. Louis Goldstein has pointed out to me that while no lexical item is an exception to a phonetic implementation rule, it is has been argued extensively that individual words do have their own phonetics in that they may exhibit a particular phonetic phenomenon (e.g. flapping in English) with reliably different frequencies and degrees (Pierrehumbert 2002, among others). Even if the degree of flapping must be specified in the lexicon, the class of cases that can exhibit flapping at all is still predictable and (probably) exceptionless, as PWLE claims.
   If phonetic implementation rules are interpreted as Optimality Theory constraints, as in §3 below, they can of course be overridden by higher-ranked constraints. PWLE claims only that they do not have lexical exceptions.
7. For ASL there is an additional argument against using a feature like [Cross], based on parallels between POA on the trunk and on the face (Perlmutter 2004),
8. HK cite three signs with radial contact articulated in the center of the chest which could be evidence that central vs. contralateral POA is distinctive, with radial contact predictable from contralateral POA for signs with handshapes capable

of radial contact. HK, however, draw the rather sweeping conclusion that these signs have a semantically-driven POA which universally overrides the phonetic default rules.

9. HK also use phonetic prespecification for the indexicals (e.g. KIDNEY, STOMACH, LIVER) discussed in §2.3. Under our proposal that these constitute a domain or stratum in the lexicon, the lexicon is simplified because there is no need to provide a separate semantic representation for each sign in this domain.
10. More precisely, the central name signs will be marked [+Ctr], while the contralateral ones will be marked [–Ipsi], from which [–Ctr] is predictable by a redundancy rule.
11. This bears on the larger issue of how large a class of grammars the theory characterizes. For *n* constraints, there are *n!* possible rankings, hence *n!* possible grammars. If phonetic and phonological constraints are in separate modules, where there are *j* phonetic and *k* phonological constraints, the number of possible grammars is *j!k!*. If phonetic and phonological constraints are intermingled, the number of possible grammars is *(j+k)!* As *j* and *k* increase, *(j+k)!* increases much more rapidly than *j!k!* does.
12. MR cite similar data from ASL, German Sign Language (in German, *Deutsche Gebärdensprache* or *DGS*), Australian Sign Language (*Auslan*), and Japanese Sign Language (in Japanese, *Nihon Shuwa*). Given the similarity of the data cited from the four languages, the conclusions we reach here for ASL may prove viable for the other languages as well.
13. This three-way distinction in how verbs realize person inflection plays a key role in Askins and Perlmutter (1995) and has been widely adopted in subsequent discussions of ASL verb inflection (Mathur 2000, among others). Earlier descriptions of ASL verb inflection (e.g. Padden 1988) discussed only verbs whose person inflection is realized as direction of movement (from the subject locus toward the object locus). Askins and Perlmutter propose a phonological analysis under which how a verb realizes inflection follows from other phonological properties and therefore need not be marked in the lexicon.
14. The notation "1:3" indicates 1st person subject and 3rd person object, while "3:1" indicates 3rd person subject and 1st person object. "3:3" indicates 3rd person subject and 3rd person object.
15. This can be captured by a phonetic implementation rule whose formulation is beyond the scope of this paper.
16. Number inflection is ignored here.
17. There are no phonological contrasts between ipsilateral and contralateral POA in neutral space (Perlmutter 2004). Since inflecting verbs are articulated in neutral space or use the weak hand as POA, there are no such contrasts in inflecting verbs. The *loci* of the subject and object for which they are inflected may be on the ipsilateral or contralateral side, however. Thus, the terms "ipsilateral" and

"contralateral," when applied to inflecting verbs, refer to the *loci* for which they are inflected, not to POA.

18. Throughout this discussion I assume the correctness of MR's data.
19. MR's description of what is disallowed in terms of Articulatory Phonology is somewhat different: the radius-ulna can be pronated only so far (up to 90º around the axis), i.e. it is a degree-of-articulation constraint.
20. MR's solution seems to have been motivated by the (unwarranted) assumption that the singly inflected form used in cases where the doubly inflected form is ruled out by (11) is derived from the doubly inflected form in the phonology. As (9–10) make clear, however, the singly inflected form exists independently. All that is needed is for (11) to rule out those doubly inflected forms that require radio-ulnar pronation.
21. In the upper left-hand corner of each tableau we indicate the inflected form in question and whether the subject and object *loci* are ipsilateral or contralateral.
22. MR also discuss a form inflected for a 1st person plural object which exhibits the arc movement expressing plural inflection. Another form inflected for a 1st person object has no number inflection, Thus it has no arc movement and is compatible with both singular and plural objects. In most cases, either 1st person object form can be used with a 1st person plural object.
    MR show that the form with plural inflection is impossible in cases where radio-ulnar pronation would be necessary to execute it. They claim this is due to the ranking of (11) with phonological faithfulness constraints. As with the 3C-3I forms discussed above, however, *R-U-PRON by itself is sufficient to rule out the ill-formed forms where plural inflection causes radio-ulnar pronation; the phonological constraints are not needed. The independently existing 1st person object form without plural inflection is used in these cases.
23. With only 13 signing and 13 non-signing subjects, that probability (known as the "power" of the experiment) would most likely be insufficient to inspire confidence that there is no significant difference between the two groups.
24. As pointed out above, signers' awareness of handshape would enable them to base some guesses on handshape.
25. The zone below eye level (particularly the mouth) expresses certain manner adverbials, among other things.
26. A theory need not state (22) as a principle. If modality is irrelevant, a theory of word segmentation can be developed without saying anything about modality.

References

Askins, David and David M. Perlmutter
1995 Allomorphy explained through phonological representation: Person and number inflection in American Sign Language. Manuscript, Uni-

versity of California, San Diego and University of Rochester. Presented at the annual meeting of the Deutsche Gesellschaft für Sprachwissenschaft, Göttingen.

Battison, Robbin
1978 *Lexical Borrowing in American Sign Language.* Silver Spring, MD: Linstok Press.

Brentari, Diane
1998 *A Prosodic Model of Sign Language Phonology*. Cambridge, MA/London: The MIT Press.
this vol. Effects of language modality on word segmentation: An experimental study of phonological factors in a sign language.

Brentari, Diane and Carol Padden
2001 Native and foreign vocabulary in American Sign Language: A Lexicon with multiple origins. In: Diane Brentari (ed.) *Foreign Vocabulary in Sign Languages: A Cross-Linguistic Investigation of Word Formation*, 87–119. Mahwah, NJ: Lawrence Erlbaum Associates.

Fischer, Susan and Bonnie Gough
1978 Verbs in American Sign Language. *Sign Language Studies* 18: 17–48.

Frishberg, Nancy
1975 Arbitrariness and Iconicity. *Language* 51: 696–719.

Hulst, Harry van der
1996 On the other hand. *Lingua* 98: 121–143.

Hulst, Harry van der and Els van der Kooij
this vol. Phonetic implementation and phonetic pre-specification in sign language phonology.

Itô, Junko and Armin Mester
1995 Japanese phonology. In: John Goldsmith (ed.) *Handbook of Phonological Theory*, 817–838. Oxford/Cambridge, MA: Blackwell Publishers.

Mathur, Gaurav
2000 Verb agreement as alignment in signed languages. Unpublished doctoral dissertation, Dept. of Linguistics, Massachusetts Institute of Technology.

Mathur, Gaurav and Christian Rathmann
this vol. Variability in verbal agreement forms across four signed languages.

Padden, Carol A.
1988 *Interaction of Morphology and Syntax in American Sign Language*. (Outstanding Dissertations in Linguistics) New York/London: Garland Publishing, Inc.

Padden, Carol A. and David M. Perlmutter
1987 American Sign Language and the architecture of phonological theory. *Natural Language and Linguistic Theory* 5: 335–375.

Perlmutter, David M.
2004 Separating Phonetics and Phonology in Sign Language Grammars. Manuscript, University of California, San Diego.
Pierrehumbert, Janet
2002 Word-specific phonetics. In Carlos Gussenhoven and N. Warner (eds.) *Laboratory Phonology VII,* 101–140. Berlin: Mouton de Gruyter.
Sandler, Wendy
1989 *Phonological Representation of the Sign: Linearity and Nonlinearity in American Sign Language.* (Publications in Language Sciences 32) Dordrecht, Netherlands: Foris Publications.
Stokoe, William
1960 Sign language structure: An outline of the visual communication systems of the American Deaf. *Studies in Linguistics, Occasional Papers 8*.
Supalla, Samuel J.
1990 The arbitrary name sign system in American Sign Language. *Sign Language Studies* 67: 99–126.
Vroomen, Jean, Jurki Tuomainen and Beatrice de Gelder
1998 Theories of word stress and vowel harmony in speech segmentation. *Journal of Memory and Language* 38: 133–149.

II. Sources of variation and their role in the acquisition of phonological competence

Getting the rhythm right: A cross-linguistic study of segmental duration in babbling and first words*

Marilyn May Vihman, Satsuki Nakai and Rory DePaolis

The broad goal of this study was to understand how children in the earliest stages of word use integrate their perceptual knowledge of ambient language prosody and segmental patterns with their production experience of speech motor control to begin to produce words and phrases with adult-like rhythm. Disyllabic babbling and identifiable words or phrases produced by five infants each acquiring American English, French and Welsh, at two developmental points within the single word period, were compared with elicited adult disyllables, both words and nonwords. The elicited adult forms were designed to resemble the segmental patterns produced by the infants in each group, in order to control for the effects of inherent segmental duration and phonotactic structure in making adult/child comparisons. We also considered an uncontrolled sample of prosodically isolated adult disyllabic productions of (child-directed) words and phrases derived from the same recording sessions which provided the infant vocalisation data in all three language groups.

1. Introduction

In the past 10 or 15 years of experimental studies of infant speech perception great strides have been made in our understanding of what children know about their native language by the end of the first year of life (Jusczyk 1997). Broadly speaking, we see that whereas knowledge of native language prosody is already gained in the womb in the last trimester before birth (Querleu et al. 1988), segmental patterns gradually become familiar as well over the course of the first year, with accelerated learning between 9 and 12 months. Recent work demonstrating infants' capacity for 'statistical learning' of arbitrary distributional patterns (Saffran, Aslin and Newport 1996) suggests that implicit learning is the basis for these advances. Implicit learning is also reflected in vocal production, which shows ambient language effects as early as 10 months (Boysson-Bardies et al. 1989; Boysson-Bardies and Vihman 1991). However, the relationship between the impressive prelinguistic knowledge of the ambient language and the child's deployment of motoric patterns for the production of identifiable first words remains unclear.

The study of rhythmic patterns in production in children acquiring three accentually and rhythmically distinct languages should help us to gain a purchase on this problem. More specifically, the study of segmental duration provides an opportunity to look at rhythmic factors – the infants' ability to match the durational patterning typical of (C)VCV sequences in the adult language – while at the same time taking account of differences in the segments and the targets attempted and actually produced in each language group. To our knowledge, no cross-linguistic studies of infants have previously addressed these issues, although Vihman, DePaolis, and Davis (1998) included duration along with pitch and amplitude in an investigation of the acoustic and perceptual characteristics of infant disyllabic vocalisations in English and French. Similarly, Vihman and Velleman (2000) compared medial consonant duration at two points in the single word period in three language groups, English, French, and Finnish, showing that while there is great individual variability (and relatively long medial consonants) in all three groups early in this period, by the end of the period the medial consonants of children acquiring both English and French become shorter while those of Finnish infants, exposed to contrastive consonant length in the adult language, grow longer. Neither of these studies included analyses of adult data, however.

English, French and Welsh provide a good basis for comparing the effects of exposure to different accentual systems. All three languages show final syllable lengthening (Delattre 1966; Williams 1986) but they are otherwise reported to have complementary durational patterns. The dominant stress pattern of English disyllabic words is trochaic, or strong-weak (75% of English words, according to Delattre 1965). In fact, the proportion of English disyllabic words attempted by children in the single word period can be over 90% trochaic (Vihman and McCune 1994; Vihman, DePaolis and Davis 1998). On the other hand, children acquiring English tend to produce at least as many monosyllables as disyllables in the early word period (for the same five children included in the present study, 54% of all child word forms were monosyllables, 38% disyllables, while content words in running speech in the input were 70% monosyllabic: Vihman et al. 1994). The children's disyllabic productions include attempts at apparent monosyllabic targets preceded by dummy syllables (Vihman, DePaolis and Davis 1998). Those vocalisations often give the impression of being inspired by adult phrases, which are typically iambic in the input (75%: Delattre 1965). This means that English presents the child with a relatively complex accentual learning problem, since it is necessary to match at least two distinct prosodic patterns, both of them of high incidence in the input.

In contrast, French accents the phrase-final syllable, mainly by lengthening it, so that disyllables are uniformly iambic, whether words or phrases (Fletcher 1991). Vihman, DePaolis and Davis (1998) found that toward the end of the single word period "the [first-to-second vowel] duration ratios for the French infants were relatively stable and adult-like, whereas the American infants showed only slight second syllable lengthening, on average, and a considerably higher level of variability for each syllable (especially the first) than was found in the French data" [p. 944].

Welsh disyllabic words, like English ones, are predominantly trochaic (Williams 1986); disyllabic phrases may be either iambic (*a* ˈ*fo* 'that's it!', *na* ˈ*ni* 'here we are', *yn* ˈ*dwyt* 'aren't you?') or trochaic (ˈ*da, de?* 'good, hunh?', ˈ*tyd ta* 'come-on then'), although no quantitative data as to the typical distribution are currently available. The nature of stress is quite different in the two languages, however. In English, stress is characterized by a combination of perceptually "strengthening" factors affecting the vowel nucleus: greater intensity, higher pitch, longer duration, and a qualitative difference (full as opposed to reduced vowel). In contrast, the stressed vowel in Welsh trochaic words is identifiably *short* rather than long; it is the consonant following the stressed vowel that is marked by lengthening (Williams 1986). Welsh stress is also characterised by greater intensity on the initial syllable. However, pitch prominence, although less reliable than relative duration as a cue to stress, tends to fall on the final syllable (Watkins 1993; Williams 1986). We will see below how these distinct accentual systems result in a unique durational pattern for the VCV portion of disyllables in each of these languages.

In our analyses of disyllabic units we will also distinguish onomatopoeic forms – rare in adult discourse in all three languages but more common in child-directed speech – from other words and phrases. Onomatopoeic words are more variable in terms of accentual pattern than are conventional words or phrases; they are also more likely to be produced with playful emphasis on either syllable. According to several native speakers, onomatopoeic words are characteristically iambic in Welsh, while in both French (e.g., *miam-miam* 'yum-yum') and American English (*tick-tock*) they are most often even-stressed. Thus, these unconventional lexical items tend to depart from the dominant word-accent pattern in all three languages. With regard to the homogeneity of the input the child hears, if we limit ourselves to disyllables, English and Welsh stand as mixed systems, in contrast to French, which has only a single accentual pattern for two-syllable utterances, aside from onomatopoeia.

Another potential source of variability in the median lengths of V-C-V elements in the word production of children exposed to different languages has been proposed, however. This is the difference in the inherent rhythmic variability of the adult languages, based on purely phonetic acoustic analyses rather than on phonological classification (Grabe and Low 2002; Ramus, Nespor and Mehler 1999). The durational variability captured by these rhythm class models derives from a calculation of the duration of vocalic and intervocalic intervals (exclusive of pauses), based on acoustic analysis of controlled sentences (Ramus, Nespor and Mehler 1999) or of longer passages of read speech (Grabe and Low 2002). Ramus and colleagues found the best acoustic correlate of rhythm classes to be a combination of the proportion of time allocated to vocalic intervals and the standard deviation of the duration of consonantal intervals. Grabe and Low suggested that a better indicator of rhythmicity could be obtained by separately calculating a "Pairwise Variability Index" across successive vocalic and intervocalic intervals, with normalisation for speech rate for the vocalic index. The difference between the two methods for assigning rhythm class is in the index derived for vocalic intervals: proportion of vocalic relative to intervocalic intervals in the Ramus et al. model, variability in the duration of vocalic intervals in the Grabe and Low model.

Languages are then placed within the matrix defined by the intersection of the two indices to define a (graded) rhythm class space. The incidence of vowel reduction, diphthongs and tense vowels plays a role in defining rhythm class under either model, as does the incidence of consonant clusters, including consonant length or geminates, and final consonants, all of which contribute to the variability of syllable types in a language (one of the factors thought to enter into the impressionistic classification of languages by rhythm types: Dauer 1983). Using this approach English, the classic example of the "stress-timed" language type, again stands in contrast with French, long considered to be a prototypical "syllable-timed language." Welsh falls in between. According to Grabe and Low's calculations (see their Fig. 2, p. 530), Welsh is closer to French than to British English, while the rhythm class characterization used by Ramus et al. (1999) would place Welsh about equidistant from British English and French but would classify Welsh as stress-timed, like English (see Fig. 3, Grabe and Low 2002, p. 534).[1]

The empirical goal of this study was to test the effects on child rhythmic learning of differences in input speech as regards relative homogeneity at (1) the level of word or phrase accentual pattern and (2) the level of C-V alternation, or segmental sequencing. Although prosodic patterns elicit the earliest learned (i.e., language-specific) perceptual responses in infancy, segmental patterns begin to be known perceptually and to influence infants' vocal production patterns within the first year as well, as noted above. On the other hand, the segmental sequences found in early child words and contemporaneous babble are highly similar cross-linguistically. Within the limits of children's motoric planning skills, the specific rhythmic patterns of the adult language should have some influence; the limits of those planning skills remain largely unknown for children at this stage, however.

At the level of both larger (word, phrase) and smaller units (sublexical segmental sequences) French is less variable than either English or Welsh, leading to the expectation that children acquiring French will advance more rapidly in approximating the adult pattern. At the larger unit level, there is little to choose between the variability of model structures for disyllables in English and Welsh, however. Both have mainly trochaic words in input speech and a mix of trochaic and iambic phrases, although disyllabic words are far more common in the input in Welsh than in English, as we will see. At the segmental sequence level English is more variable than either French or Welsh (based on Grabe and Low 2002) and could thus be expected to provide children with the greatest challenge; Welsh falls between English and French in this regard.

2. Method

Data from five children each were drawn from longitudinal studies conducted in three language communities: American English (California; 10 participants in the original study, five of them boys: Vihman et al. 1985), French (Paris, France: Boysson-Bardies and Vihman 1991), and Welsh (North Wales). All of the infants were normally developing. All five of the American children whose data are included here, three of the French, and one of the Welsh were first-born. Two American, three French and one Welsh child were male. Children were recorded at home on a weekly (English) or biweekly basis (French and Welsh) on audio and video, in natural interaction with their mothers and sometimes with the observer, who was always a native speaker.

Two word points were identified for sampling the data in a comparable way cross-linguistically: the 4-word point, the first month in which the child used four or more identifiable adult-based words spontaneously in a half-hour session (4wp: two sessions sampled) and the 25-word point (25wp: one session sampled). The latter corresponds to approximately a 50-word cumulative vocabulary (for word identification procedures and other methodological details used in all three studies see Vihman and McCune 1994). For English we used data only from the five children who had reached the 25wp by the age of 17 months. In both the French and the Welsh groups a sixth child dropped out of the study when word production proved to be slow in getting started. There is thus a small bias toward precocity in word production in all three groups. All analyzable disyllables, including both words and nonwords or babble vocalisations, were extracted from each child's 4- and 25-word points. In some instances supplementary disyllables were selected from the week immediately preceding the 25-word point for the English group, since these children produced less disyllables overall. Mean group ages and numbers of disyllables analysed are indicated in Table 1.

Table 1. Ages and tokens analyzed.

Language	**Mean age in months**		**Tokens analyzed**	
Word point	*4*	*25*	*4*	*25*
English	11.5	16	111	117
French	11.5	16.5	114	169
Welsh	13	17	168	167

2.1. Selection Criteria

The study was limited to disyllabic vocalisations only, including both identifiable words and babble, for two reasons: (1) the disyllable is the minimal unit needed for the investigation of intervocalic consonantal length, and (2) only monosyllables and disyllables are of high incidence in infant production at this stage, cross-linguistically (Vihman et al. 1994). Utterances selected for inclusion minimally contained two open (vocalic) phases separated by a closed (consonantal) phase. We included every disyllable which lent itself to objective analysis by the methods available. Items whose medial consonant was a glide, which poses particular problems for segmentation, were excluded. Disyllables with interfering talking or other noise were not used. Utterances which showed excessive shifts from modal register, excessive vocal effort, whisper, or creaky voice were also excluded. No more than three successive repetitions of a single word type were included in the analysis, on the grounds that a "prosodic set" could be inferred and such mechanical repetition might bias the results.

2.2. Spontaneous adult data extracted from recordings

In order to obtain a representative sample of spontaneous adult speech whose durational properties could provide an idea of the range of models the infants are exposed to we searched through the 4wp and 25wp recordings of the children included in this study, in all three languages, for all prosodically isolated adult disyllabic productions, whether words or phrases. We defined "isolated disyllables" as those that were separated from adjacent utterances by at least 300ms (following Fernald et al. 1989). In the case of Welsh, mothers as well as children wore a wireless microphone connected to a transmitter. In the case of English and French, only the infants wore a microphone in most sessions, however. Since the mothers were located at varying distances from the infant in the course of the sessions, only a relatively small subset of the disyllables extracted lent themselves to analysis for those languages. Disyllables that were whispered or overlaid with noise, or that lacked a medial consonant, were excluded.

2.3. Elicited adult data

In order to form a clear idea of the "end state" toward which the children might be supposed to be heading while abstracting away from the differences in phonotactic structure between child and adult productions, we elicited child-like disyllabic patterns from five female native speakers of each language. Since target words, or the adult models for words attempted by the children, occurred only rarely as isolated disyllables in the recorded sessions, we included in our stimuli the most common child word targets for each language group alongside nonwords modeled on frequently occurring children's disyllabic nonword patterns (Table 2). Note that one phonological pattern – /babi/ – was elicited in all three groups for direct comparison across languages (Eng. *bobby,* Fr. *babie,* Wel. *babi* 'baby').

Table 2. Elicited adult word and nonword disyllables (word spelling is in italics).

English		**French**			**Welsh**		
spelling	*phonetic*	*spelling*	*phonetic*	*gloss*	*spelling*	*phonetic*	*gloss*
apple	æpəl	*bébé*	bebe	baby	*babi*	babi	baby
baby	beɪbi	*chapeau*	ʃapo	hat	*choo choo*	tʃutʃu	choo-choo
Big Bird	bɪgbəd	*maman*	mamã	mama	*eto*	ɛtɔ	again
bottle	bɑrəl	*papa*	pɑpɑ	papa	*si-so*	sisɔ	see-saw
button	bʌʔən	*poupée*	pupe	dolly	*ta ta*	tɑtɑ̥	bye-bye
					tedi	tɛdi	teddy
bebba	bɛbə	aba	aba		ada	ada	
bebby	bɛbi	aideux	ɛdø		dede	dɛdɛ	
bobby	bɑbi	babie	babi		dwdw	dudu	
doodoo	dudu	bobeau	bobo		gaga	gaga	
edda	ɛdə	doudou	dudu		gagak	gagak	
gogga	gɑgə						

2.4. Acoustic Analysis

In the case of the spontaneous English and French adult and child data, disyllables were extracted from the audio tapes and digitized to 16 bits using an Audiomedia sound board in a PowerPC (sampling rate 22.2 kHz). The Welsh data had been recorded onto a DAT deck and so were transferred digitally for further analysis. All measurements were made using Soundscope speech analysis software. Duration measurements used concurrent information from the amplitude trace, narrow and wide band spectrograms, and intensity curve. Additional screens were used to expand the beginning and endpoints of the segments to be measured in order to obtain more detailed signal information related to each manually placed marker (see Vihman, DePaolis and Davis 1998, Fig. 2, for an illustration). Rules for segmentation of the first vowel (V_1), medial consonant (midC), and second vowel (V_2) were based on relevant transition cues, depending on the surrounding segments. Utterance-initial consonants were excluded from the measurements. Glides occurring between vowel and consonant or consonant and vowel were included in the vocalic measurement.

3. Results

3.1. Durations of V-C-V in the three languages: Adults

Figure 1 shows the proportional durations of the individual V-C-V elements of the form /babi/ as produced in isolation by five adult speakers of each of the three languages. The patterns in the three languages are clearly distinct. As expected, based on prior reports, English is marked by long V_1, French by long V_2, and Welsh by long midC. A repeated measures ANOVA (Language: Between-subject factor; Element: Within-subject factor) performed on the proportional durations of adult /babi/ indicated that the Language x Element interaction is highly significant [$F(4, 24) = 44.7$, $p < .001$]. Multiple comparisons (Bonferroni corrections applied) revealed that the following differences are significant: $\%V_1$: English > French and Welsh, %midC: Welsh > English and French, $\%V_2$: French > Welsh.

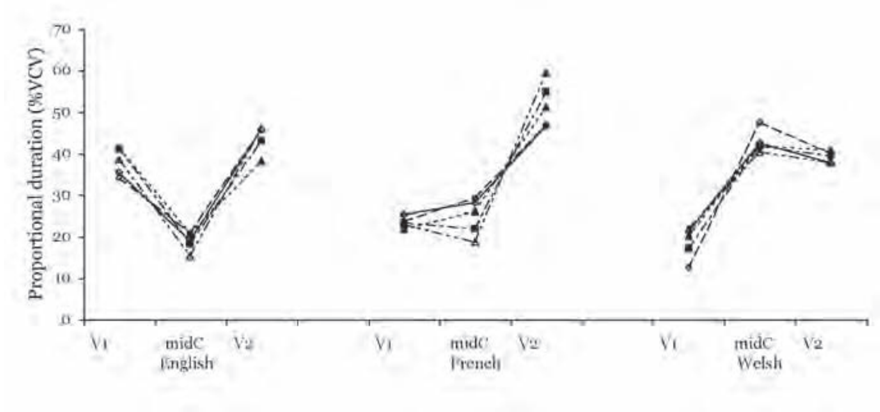


Figure 1. Proportional durations of elements of adult productions of /babi/. (Individuals are plotted in different lines.).

3.2. Durations of V-C-V in the three languages: Children

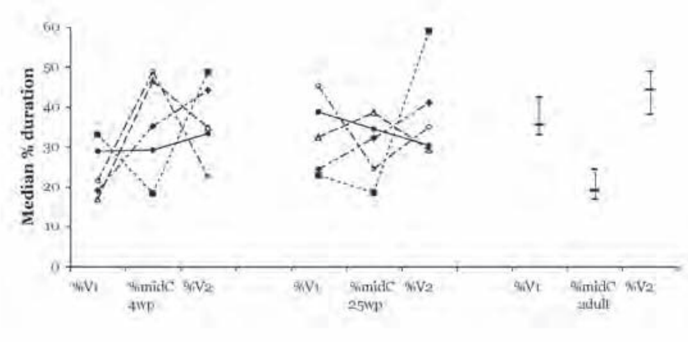
A comparison of the particular segments used in the child vocalisations measured in the three language groups revealed a number of significant differences that could be expected to affect the overall durations of VCV sequences. At both word points in all three language groups stops accounted for the highest proportion of medial consonants (close to 50% or more). The second most frequent consonant type differed across groups and ages: nasals for the American children at both age levels but fricatives for both French and Welsh at the 4wp, with a shift to more nasals by the 25wp in both those languages.[2] The duration of children's consonant types differed in the same way across all language groups, with the ranking stops > fricatives > nasals (median 173, 148 and 133ms, respectively).[3]

Since the particular consonants produced in the disyllables used for analysis could affect the results of cross-linguistic comparisons we limited our durational analyses to vocalisations with medial stops. We also limited the vocalic portions to monophthongs, excluding tokens consisting of either diphthongs or syllabic consonants, which amounted to less than 12% of all vocalic nuclei for any language at either word point but which tended to be longer than monophthongs. For example, the median duration of monophthongal V2 was ca. 210 ms while the median duration of diphthongs and syllabic consonants in that position was ca. 300ms.

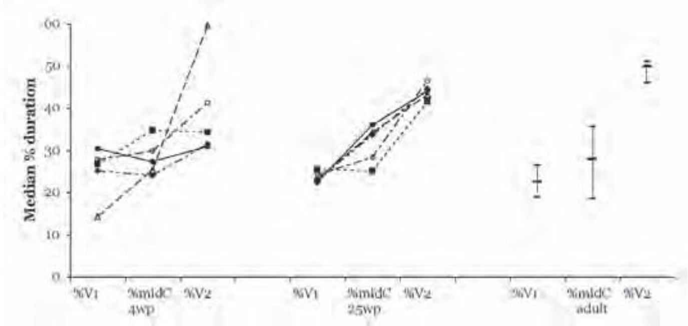
Figure 2 shows the vowel-stop-vowel (V-S-V) profile for the children's vocalisations in each language group, at the 4wp and the 25wp, with the segmental types restricted to stops and monophthongs. Median durational values for adult production of all the elicited nonword disyllables included in Table 2 are plotted on the right for comparison (recall that only stops and monophthongs were included in the elicited nonwords). Since the proportional durations of V-S-V elements differ for different consonants and vowels, even within the limits we imposed, error bars represent the range of median values of the various child-form disyllables.[4]

At the 4wp in all three groups the majority of the children, like the adults, show proportionately longer V_2 than V_1, giving the effect of final syllable lengthening. The proportional durations of the three elements taken together reveal relatively little evidence of specific ambient adult language shaping, however. Only one American child produced elements whose relative durations resembled those of the American adults' elicited forms (broadly, midC $< V_1, V_2$), while two Welsh children produced a pattern resembling that of the Welsh adults ($V_1 <$ midC, V_2) and three French children produced a pattern resembling that of the French adults (V_1, midC $< V_2$). Furthermore, considering the individual children in each group, only one American child, one French child and two Welsh children produced all three elements within the range of the adult values (±10%) at the 4wp. The French children's patterns give a more homogeneous impression than do those of the other two groups, but Cochrans C tests indicate that the cross-linguistic differences are not statistically significant ($C[4, 3] = .46$, $p = .74$ for V_1; $C[4, 3] = .61$, $p = .24$ for midC).

American children



French children



Welsh children

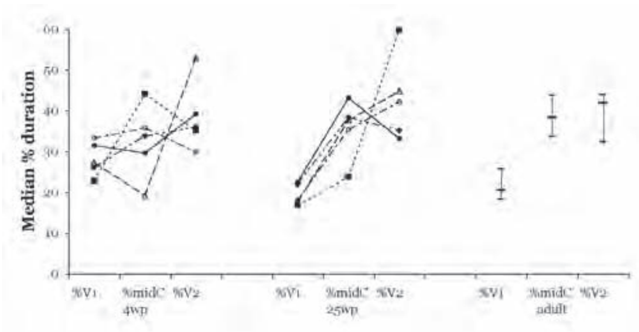


Figure 2. Proportional durations of elements of child disyllables at 4 and 25wp (leftmost panels); right panels show median and range of durations of each element in all elicited adult nonword disyllables for each language.

We see from Figure 2 that in all three language groups the children's vocalisations do show a change between the two sampling periods such that at the 25wp they more closely approximate the adult pattern of the target language. This necessarily required different patterns of change within each language group, and for the different children within each group. Overall, in proportional terms, American children's V_1 became longer and midC shorter; French children's V_1 became shorter and V_2 longer; Welsh children's V_1 became shorter while their midC and V_2 became longer. Considering individual children, however, French and Welsh children had achieved far more adult-like V-S-V proportions than had American children by the later developmental point. Of the five children in each group, five French and four Welsh children produced V-S-V proportions that approximately matched the adult shape while only one of the American children matched the ratio for all three elements. A better match between child and adult productions for French and Welsh in comparison with English at the 25wp results in greater homogeneity for the groups of French and Welsh children as a whole. A Cochrans C test indicated that at the 25wp the difference in the homogeneity of variance is significant for the proportional duration of V_1, where American children differed from the remaining two groups: $C(4, 3) = .92$ $p = .001$. The differences for the other elements are not statistically significant, however.[5]

Table 3. Number (proportion of total) and structures of mothers' isolated disyllables.

		Total (proportion)	***long /tense V, diphthong, glide + V***		***med. cluster***	***final consonant***
			V1	*V2*		
English	*words*	44 (.37)	24	31	8	16
	phrases	65 (.54)	27	37	35	59
	onomat.	11 (.09)	10	10	2	10
	Total	120	61 (.51)	78 (.65)	45 (.38)	85 (.71)
French	*words*	101 (.61)	3	5	21	40
	phrases	46 (.28)	3	7	2	7
	onomat.	19 (.11)	0	0	4	0
	Total	166	6 (.04)	12 (.07)	27 (.16)	47 (.28)
Welsh	*words*	154 (.62)	13	9	54	54
	phrases	54 (.22)	8	6	15	12
	onomat.	39 (.16)	27	9	8	17
	Total	247	48 (.19)	24 (.10)	77 (.31)	83 (. 37)

3.3. Crosslinguistic comparison of mothers' isolated disyllables

Table 3 characterises the isolated disyllables extracted from the mothers' child-directed speech and analysed for the duration of the V-C-V elements. The disyllables are categorised into the three units that might be expected to differ in accentual pattern: words, phrases, and onomatopoeic forms. As expected based on earlier studies, words make up a far smaller proportion of the disyllables produced in child-directed speech in English (37%) in comparison with French (61%) and Welsh (62%). The proportion of onomatopoeic forms is also somewhat smaller in English and French than in Welsh. Phrases make up over half of the mothers' isolated disyllables in English, as compared to well under a third in the other two languages. In particular, certain phrases are of very high incidence in the English data (over half are wh-questions, such as *who's that? what's this? where else?*). Furthermore, although trochaic words are the dominant English pattern, only 23 out of the 44 tokens analyzed were standard trochaic words (e.g., *bunny, cupcake vs. allgone, whoopee,* which have variable accent).

To better show the differences between English and the other two languages with respect to the complexity of the nucleus and other aspects of phonotactic structure that affect the rhythmic profile, we also indicate the occurrence of diphthongs and contrastively long or tense V_1 and V_2, medial clusters, and final consonants in the mothers' disyllables. From this it is clear that the phonotactic complexity of our spontaneous English data exceeds that of the other two languages on all three measures.

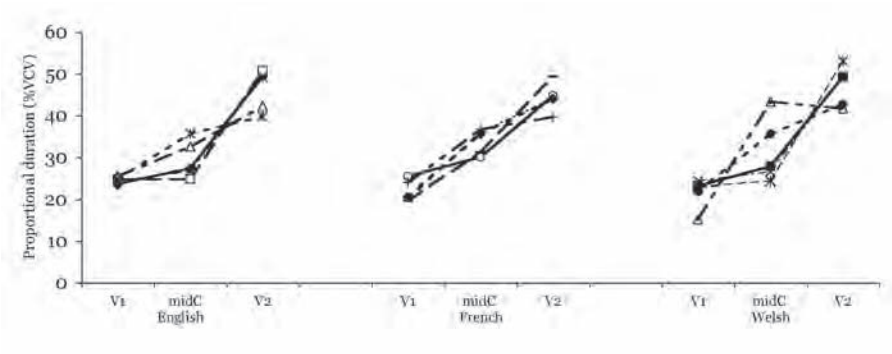


Figure 3. Proportional durations of V-C-V elements of mothers' disyllables.

Figure 3 summarizes the distribution of proportional durations of V_1, midC, and V_2 for the analyzable disyllables extracted from each mother's child-directed speech. The differences between language groups are far less evident here than in the elicited disyllabic nonwords (adult panels, Fig. 2), in which accentual pattern, segmental type, and phonotactic structure were all controlled. Furthermore, the cross-linguistic differences in variability that we expected to see are not apparent either. We see, instead, the same differences in variability across the elements of the V-C-V sequence in all three languages: V_1 covers a smaller proportion of the total duration and is the least variable, midC takes up a larger proportion of the duration and varies somewhat more widely, and V_2 is proportionately the longest and the most widely varying. Phonotactic complexity appears to interact with other factors to obscure any durational differences due to accentual patterning alone in this uncontrolled data sample.

If these findings provide an accurate picture of the V-C-V duration profiles in input speech, however, we are left with a puzzle: How can we account for the children's progress toward distinct rhythmic patterns by the 25wp if isolated disyllables taken from mothers' child-directed speech are so similar across the languages? What then are the children's models?

3.4. Children's disyllabic words in relation to targets

The children differed across languages in the types of adult disyllabic targets they attempted and produced; to some extent this reflects differences in adult language structures and in the nature of the input to children in the different groups (see Table 4). As expected, the French children produced the most words following the dominant adult word pattern (iambic words: 61%), with the remainder divided between a few onomatopoeic forms (8%) and a larger number of "other" forms including interjections (*allo, bravo)*, other words or phrases whose accentual pattern is variable (*non-non*), and monosyllabic words preceded by a likely "filler syllable" ([*à*] *voir* [aβa], [*de*] *l'eau* [dəlo]), giving the impression of an iambic phrase, although the actual target can only be guessed at (on filler syllables in French, see Veneziano and Sinclair 2000).

The Welsh children produced twice as many onomatopoeic forms as either of the other two groups (21% of all target types), with less than half of all target word types conforming to the dominant word pattern (42% trochaic words). No putative phrases consisting of filler + monosyllabic content word were identified as targets for Welsh child productions with the exception of *oh God! (*produced by a child with several older siblings).[6]

Table 4. Adult targets for children's disyllables (proportion of all target types).

	Dominant pattern		***Onomatopoeia***		***Other***		***Total***	
	4wp	*25wp*	4wp	*25wp*	4wp	*25wp*	4wp	*25wp*
English	5	23	2	4	6	19	13	46
Total	28 (.47)		6 (.10)		25 (.42)		59	
French	7	29	3	2	3	15	13	46
Total	36 (.61)		5 (.08)		18 (.31)		59	
Welsh	10	20	8	7	9	18	27	45
Total	30 (.42)		15 (.21)		27 (.37)		72	

Just under half of the targets attempted by American children were trochaic words (47%), while a sizable minority fell in the category "other", including at least five monosyllabic words preceded by a filler syllable with unverifiable target (*[a/the] bead[s]*) in addition to one iambic word (*balloon*) and a certain number of longer words or monosyllables treated in some other way.

To better understand the difference in variability of rhythmic patterns at the 25wp across the different groups of children we analyzed the elicited adult forms of five target disyllables in each language (Table 2). We selected for analysis the most frequently attempted target words in each group (taking into account the number of tokens attempted by each child as well as the number of children attempting each word type). Figure 4 presents proportional V-C-V profiles for these elicited adult productions.

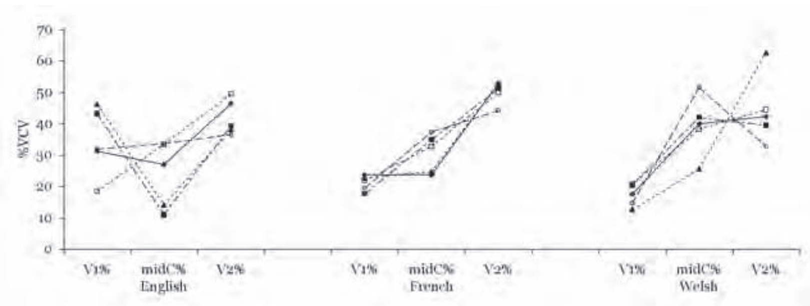


Figure 4. Proportional durations of elements of elicited adult productions of target disyllabic words.

Here we begin to gain a better idea of the source of the differential variability of children's productions in the three languages. That is, we can see that although all of the English words measured here are trochaic, the durational variability of all three elements remains considerable, reflecting the variability in the intrinsic duration of different types of segments (e.g., tense vs. lax stressed vowel, full vs. reduced unstressed vowel) as well as in phonotactic complexity (consonant singletons vs. clusters).

The similarity between the median proportional durations plotted for the different children's disyllables produced at the 25wp (Fig. 2) and those of the common target words (Fig. 4) is striking, especially in the case of French and Welsh. In order to investigate this finding more closely we undertook further analysis of these individual word targets. Figures 5–7 provide a direct comparison of the proportionate V-C-V durations of selected adult (median) and child productions.

In Figure 5 we see the proportionate durations of four French words (*chapeau* 'hat'*, maman* 'mother'*, papa, poupée* 'doll') as produced by two to four children as well as the median production pattern for the fifteen tokens elicited from adults for each of these words. The four adult V-C-V patterns are generally similar. On the whole the children are producing moderately good matches to the particular word targeted; the children's patterns are relatively similar across the four words as well.

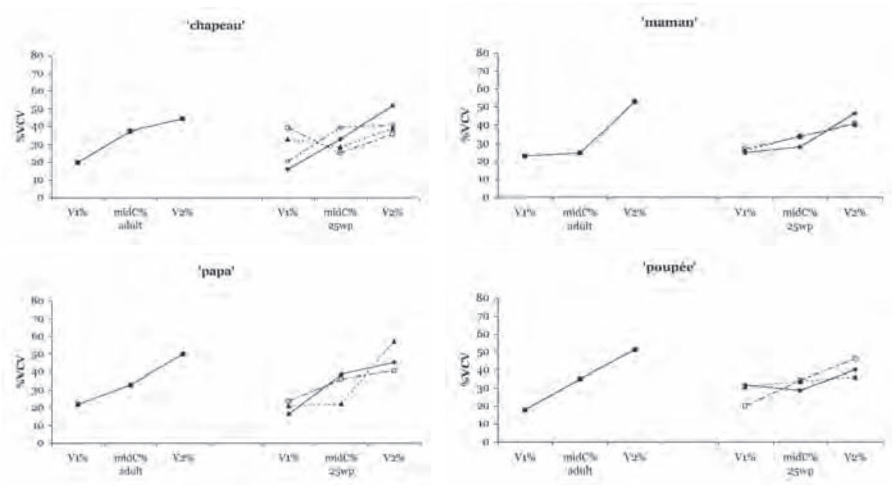


Figure 5. Individual French words; elicited adult targets and child productions.

In Figure 6, where the proportionate durations for four Welsh words are plotted (*choochoo, eto* 'again', *tata* 'bye-bye' and *tedi* 'teddy'), we can see a more striking match of child productions to individual adult word models. Each of the target words has a distinct V-C-V pattern, and the children's productions constitute good matches to the individual pattern of at least three of them. This seems to reflect item learning. The difference in the between-word similarity for adult French as compared with Welsh further suggests that the source of the child variability that we have observed may lie primarily in the variability of different adult models.

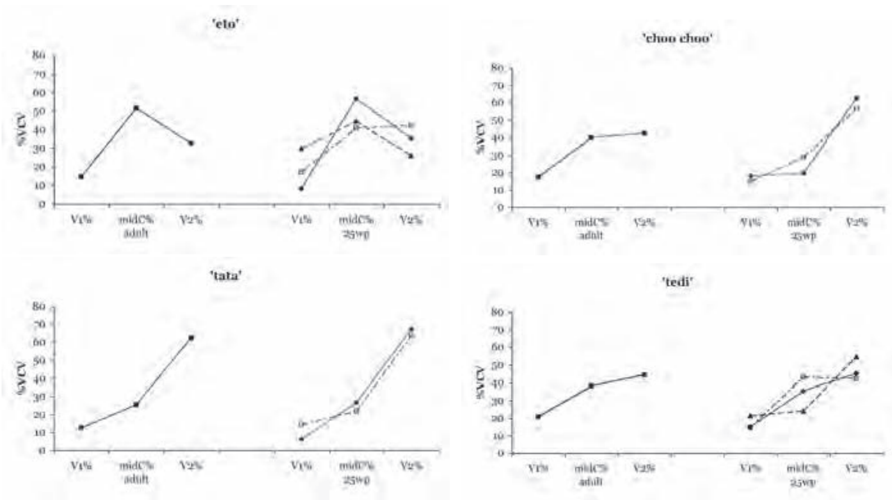


Figure 6. Individual Welsh words; elicited adult targets and child productions.

In Figure 7 we see a comparable set of comparisons for American English. Notice that although all four words are trochaic, their adult rhythmic profiles are not the same. In particular, contrast *baby*, which elicits three relatively good matches, with the other three words. In the case of *button,* in particular, the tokens produced by the two children look quite different: One of these tokens is transcribed as [bætːi], the other as [paːŋa], with the final nasal moved to medial position as in many of the child's other words (for an account of the development of this word template, see Vihman and Velleman 1989).

More generally, the differences between the fidelity of the individual children's productions as regards rhythmic match to the adult targets in the three languages can plausibly be attributed to the difference in the structural complexity of the target words. Of the five words frequently attempted in each group, all of the Welsh target words and all but one of the French target words have a simple $(C_1)VC_1V$ structure (the exception is *chapeau,* with child forms [pœpœ, bɔbo, habœ, (h)apo]: Note the variability of these tokens in Figure 5). For English, three of the target word forms include stops differing in place of articulation *(Big Bird, bottle, button*), three have a syllabic consonant as the second nucleus, *Big Bird* has a medial consonant cluster and even *baby*, which inspires the best English matches, has a diphthong for its first nucleus. None of these relatively more complex segmental types occur in any of the French or Welsh target words that were frequently attempted by the children. The relative difficulty posed by the English models has its equivalence in the phrases found in the input: Contrast the frequent English wh-questions noted above, with their two- to three-consonant clusters, with the French equivalents found in our data: *et ça?* 'and that?', *et là?* 'and there?', *qui c'est?* 'who is it?', *c'est quoi?* 'what is it?'.

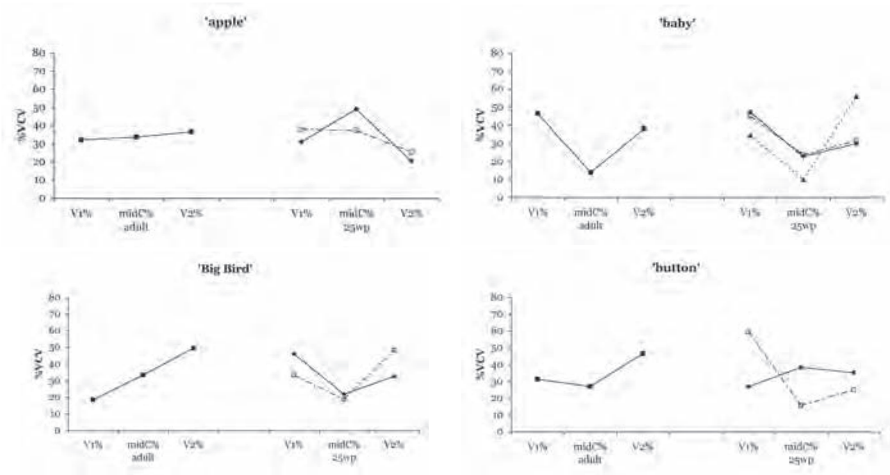


Figure 7. Individual English words; elicited adult targets and child productions.

In view of these differences it is reasonable to suppose that, in comparison with English, it is more straightforward for the child acquiring French or Welsh to succeed in making a rhythmic mapping from adult target to a form in his or her existing articulatory repertoire, at least where disyllables are concerned. (Recall that monosyllables are the dominant production pattern for most children acquiring English, but not for the other two languages.) Thus, the greater rhythmic variability that has been noted for English in comparison with French and also Welsh (Grabe and Low 2002) seems to give a reasonably good account of the children's relative success in matching the adult rhythmic standard at the 25wp.

4. Discussion and conclusion

Analysis of the elicited child-like disyllabic patterns produced by adults in the three language groups showed little within-language variability and clear between-language differences in the relative length of V_1, midC and V_2. Comparison of these results with the V-S-V productions of children at two developmental points yielded three primary findings:

(1) At the 4wp children could not be readily assigned to the appropriate language group on the basis of the way their rhythmic pattern apportioned length to V-S-V. All three groups seemed to show final vowel lengthening, also seen in the adult languages. As a group the French children appeared relatively more homogeneous and adult-like than the American and Welsh children, but the difference was not significant.

(2) By the later developmental point there was progress toward the adult model in all groups. This agrees with earlier cross-linguistic studies of segmental production over this same period (Boysson-Bardies and Vihman 1991; Vihman et al. 1994).

(3) Inter-group differences were found in the extent to which the children succeeded in matching adult rhythmic patterning. Both the French and the Welsh children generally conformed relatively closely to the adult pattern by the 25wp. In contrast, the American children remained more variable and less closely matched to the adult models.

4.1. Sources of variability

We explored two potential sources for the lower variability and greater conformity to the adult pattern seen in the French children's production, namely, greater homogeneity at the level of both larger (word, phrase, onomatopoeia) and smaller rhythmic units (sublexical segmental sequences). On the one hand, French seems to provide a more homogeneous set of target forms to model than either English or Welsh, due to the fact that French disyllabic speech forms are largely iambic, whether words or phrases. Exceptionally, forms may be evenly accented on both syllables or produced with initial-syllable stress for idiosyncratic reasons (affect, playful variation, etc.). In contrast, both English and Welsh present a mix of trochaic, iambic, and even-stressed models. However, our analyses of isolated disyllables extracted from the mothers' speech failed to provide support for the hypothesis that differences in the rhythmic patterns of different lexical types is the primary source of differences in the extent of child variability across groups.

An alternative proposal holds that the inherent rhythmic variability of a "stress-timed" language like English provides a relatively more difficult model for children to match than does French, one of the classic "syllable-timed" languages. Grabe, Post and Watson (1999), in a study of speech production in 4-year-olds acquiring (British) English vs. French, argue that the greater success shown by French children in matching the rhythmic variability level of adults is due to the fact that French is rhythmically simpler than English because French has a less variable rhythm overall, as detected in fluent adult speech.

The present study generally supports the rhythm class models (Grabe and Low 2002; Ramus, Nespor and Mehler 1999) as predictors of children's relative ease of acquisition of rhythmic patterning. On the one hand, in our elicited adult nonword productions we controlled the rhythmic complexity of the adult models by eliciting the kinds of simple patterns typically produced by children and found that French and Welsh children's vocalisations were more adult-like than those of the American children at the 25wp. On the other hand, when we compared elicited adult productions of words attempted by the children in the three languages, the difference in variability in the adult word targets was evident (Fig. 4). Furthermore, the relative variability of the five elicited target words in each language – ordered as English > Welsh > French, in agreement with the predictions of the rhythm class models – bore a close resemblance to the relative variability across the five individual children within the three language groups at the 25wp. In other words, the target words for the French children's early productions were themselves closely similar to one another, rhythmically speaking (Fig. 4), and the different children's tokens were similar as well (Fig. 5). In Welsh, in accordance with the "larger unit" variability discussed above, the words targeted by the children were more disparate in their rhythmic pattern than the French target words (Figs. 4, 6) but the children generally achieved good matches to them. In English, in contrast to both French and Welsh, the elicited target words, although all trochaic, differed considerably, apparently reflecting differences at the sub-lexical level (Figs. 4, 7). Specifically, the word targets attempted by the American children included more phonotactically complex sequences and more segments of a kind not typically found in children's earliest production repertoires. One could infer that it is a consequence of the greater phonetic challenge posed by the adult models that some of the English child word productions depart more radically from the adult forms than do any of the French or Welsh child word productions.

It may be worth noting, finally, that although the word targets attempted most often by the children learning English were more complex than those of the children learning the other languages, those targets were nevertheless less complex than the American mothers' isolated disyllabic words. Only one out of five of the children's frequent target words (20%) includes either a diphthong or a long or tense vowel (vs. 81% of the mothers' words) or a final consonant (vs. 35% of mothers' words). Without laboring the point, the children's frequent choices of early words to produce seem to reflect the kind of selectivity often reported in the literature (see Schwartz 1988).

4.2. The representation of rhythm and item learning

Returning to the broader issue of implicit prelinguistic knowledge in relation to early word production with which we began, we are now in a position to reflect again on what the child needs to know in order to "get the rhythm right". To achieve the native language pattern in production the child must eventually be able to match both the overall melody and the rhythmic pattern of individual words. In the single-word production period this might seem not to be overly challenging, since the complication of fitting content and function words, or even a succession of content words, into a single intonational contour does not yet arise. The problem is rather one of representation: In the first year, as indicated above, there is good evidence that the child is able to gain a sense of the rhythmic patterning of the language sufficient to give greater attention to words fitting the dominant pattern than to other words. Our results suggest that it is not this global representation of the rhythmic patterning in the language that underlies children's early word production, however.

We saw, first of all, that at the outset of word production children's disyllables do not yet match adult rhythmic patterning (Fig. 2). In fact, despite the extensive perceptual learning of the first year, we see here that different children acquiring the same language show distinct patterns, only a few achieving a rough approximation of the dominant rhythmic pattern of the ambient language. Thus, the construction of representations at a new level seems to be needed, involving considerable additional learning before ambient language rhythms can be successfully matched in production. Specifically, as the evidence from Welsh demonstrates the most clearly, early word learners seem to be developing adult-like rhythmic patterns for production on a word-by-word basis. A more abstract knowledge of the rhythmic patterning appropriate to the native language can be expected to emerge only gradually from the combined perceptual and production knowledge or representation of increasing numbers of individual words, with further reorganization once larger and more varied units begin to be combined and integrated into word combinations (see Snow 1994).

It has been well established that children's early words build on articulatory patterns already available in babbling. Assuming that early vocal production patterns are grounded in biomechanical constraints and are thus common to children learning different languages, the opportunities for easy matching will differ according to the particular ambient language. In addition, the findings of this study agree with earlier work in suggesting that the challenges posed by the adult language will be met differently by different children. Cross-linguistic phonological analyses of first words (4wp) in comparison with later words (25wp or beyond) suggest that each child must begin by developing a stock of individual representations of particular lexical items (Vihman 2002). These initial representations, or word production patterns, may derive from matches between the child's existing vocal patterns and (implicitly "pre-selected") adult words (Vihman and Nakai 2003).

That is, the highly selected first words reflect existing vocal patterns developed in babbling; these differ in complexity from one child to the next. Through production practice with a growing stock of adult-based words the child is then very gradually able to induce the more abstract structure of the adult language. This implicates explicit as well as implicit learning, as the child moves from the use of situationally primed first words to more intentionally targeted and more flexibly deployed later "referential" words (McCune and Vihman 2001). Our study has captured two early points in that process. At the 4wp, when babbling vocalizations make up much of the child's production, the influence of the overall rhythmic patterns of the adult language is weak, if detectible at all. By the 25wp, when the children are using 50 words or more, individual words are relatively successfully reproduced, resulting in greater homogeneity for the French group, with its less variable rhythmic patterning, than for the other groups. The achievement of adult-like rhythmic patterns will require mastery of considerably greater phonotactic complexity for all the children, but particularly for those learning English; accurate deployment of the range of accentual patterns available for words, phrases, and onomatopoeic forms can be expected to emerge in parallel with that developing mastery, as part of the process of lexical learning. Thus the implicit learning of native-language rhythms that occurs in the prelinguistic period is only a first step in the long apprenticeship that will culminate in adult-like rhythmic production.

Notes

* The authors thank the Economic and Social Research Council of the UK for its financial support. We also thank the participating families from California, Paris and North Wales and the adults who produced the elicited data. Lucy Evans collected and transcribed the Welsh infant data, Pam Martin helped with Welsh word identification, Dr. Llinos Spencer and Dr. Enlli Thomas kindly answered questions about Welsh words and phrases.

1. The classification of Welsh by the Ramus et al. method was carried out by Grabe and Low, based on their own data.
2. Note that these differences were identified only in the sample of disyllables that lent themselves to analysis in this study. For analyses based on a larger sample of English and French child vocalisations see Boysson-Bardies and Vihman (1991).
3. The greater duration of stops than of fricatives in all three groups of children is surprising. We speculate that this may be due to greater intentionality on the child's part in producing a stop than a fricative, as suggested by the greater incidence of fricatives in babbling than in words in this period (Boysson-Bardies and Vihman 1991)
4. Note that for quantitative analyses of the children's data, given skewed distribution and some extreme values, we have entered the median measurement as a summary figure for each child and, for the sake of comparability, for the adult data as well.
5. Cochrans' C tests performed on proportional durations of elements produced at the two developmental points indicate that French children's proportional durations of V1 and V2 are significantly more homogeneous at the 25wp than at the 4wp ($C[4, 2] = .96$, $p = .009$ for V1; $C[4, 2] = .98$, $p = .003$ for V2). No other groups differed in terms of homogeneity of variance.
6. The primary language of all of the Welsh children's homes was Welsh. These children rarely produced English words, but since beyond early childhood virtually all speakers of Welsh are bilingual today, it would have been impossible to avoid including in the study any Welsh children who are also exposed to English.

References

Boysson-Bardies, Bénédicte de, Pierre Hallé, Laurent Sagart and Catherine Durand
1989 A crosslinguistic investigation of vowel formants in babbling. *Journal of Child Language* 6: 1–17.

Boysson-Bardies, Bénédicte de and Marilyn M. Vihman
1991 Adaptation to language: Evidence from babbling and first words in four languages. *Language* 67: 297–319.

Dauer, Rebecca M.
1983 Stress-timing and syllable timing reanalyzed. *Journal of Phonetics* 11: 51–62.

Delattre, Pierre C.
1965 *Comparing the Phonetic Features of English, French, German and Spanish*. Heidelberg: Julius Groos Verlag.
1966 A comparison of syllable length conditioning among languages. *International Journal of Applied Linguistics* 4: 182–198.

Fernald, Anne, Traute Taeschner, Judy Dunn, Mechtilde Papousek, Bénédicte de Boysson-Bardies and Ikuko Fukui
1989 A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language* 16: 477–501.

Fletcher, Janet
1991 Rhythm and final lengthening in French. *Journal of Phonetics* 19: 193–212.

Grabe, Esther and Ee Ling Low
2002 Durational variability in speech and the rhythm class hypothesis. In: Natasha Warner and Carlos Gussenhoven (eds.), *Papers in Laboratory Phonology VII*, 515–546. Cambridge: Cambridge University Press.

Grabe, Esther, Brechtje Post and Ian Watson
1999 The acquisition of rhythmic patterns in English and French. In: *Proceedings of the XIVth International Congress of Phonetic Sciences*, 1201–1204.

Jusczyk, Peter W.
1997 *The Discovery of Spoken Language*. Cambridge, MA: MIT Press.

McCune, Lorraine and Marilyn M. Vihman
2001 Early phonetic and lexical development. *Journal of Speech, Language, and Hearing Research* 44: 670–684.

Querleu, D., X. Renard, F. Versyp, L. Paris-Delrue and G. Crépin
1988 Fetal hearing. *European Journal of Obstetrics and Reproductive Biology* 29: 191–212.

Ramus, Franck, Marina Nespor and Jacques Mehler
1999 Correlates of linguistic rhythm in the speech signal. *Cognition* 73: 265–292.
Saffran, Jenny R., Richard N. Aslin and Elissa L. Newport
1996 Statistical learning by 8-month-old infants. *Science* 274: 1926–1928.
Schwartz, Richard
1988 Phonological factors in early lexical acquisition. In: Michael D. Smith and John L. Locke (eds.), *The Emergent Lexicon*, 185–222. New York: Academic Press.
Snow, David
1994 Phrase-final syllable lengthening and intonation in early child speech. *Journal of Speech and Hearing Research* 37: 831–840.
Veneziano, Edy and Hermine Sinclair
2000 The changing status of 'filler syllables' on the way to grammatical morphemes. *Journal of Child Language* 27: 461–500.
Vihman, Marilyn M.
2002 Getting started without a system. *International Journal of Bilingualism* 6: 239–254.
Vihman, Marilyn M., Rory DePaolis and Barbara L. Davis
1998 Is there a "trochaic bias" in early word learning? *Child Development* 69: 933–947.
Vihman, Marilyn M., Edwin Kay, Bénédicte de Boysson-Bardies, Catherine Durand and Ulla Sundberg
1994 External sources of individual differences? *Developmental Psychology* 30: 651–662.
Vihman, Marilyn M., Marlys A. Macken, Ruth Miller, H. Simmons and Jim Miller
1985 From babbling to speech: A reassessment of the continuity issue. *Language* 61: 395–443.
Vihman, Marilyn M. and Lorraine McCune
1994 When is a word a word? *Journal of Child Language* 21: 517–542.
Vihman, Marilyn M. and Satsuki Nakai
2003 Experimental evidence for an effect of vocal experience on infant speech perception. In: M. J. Solé, Daniel Recasens and J. Romero (eds.), *Proceedings of the 15th International Congress of Phonetic Sciences*. Barcelona.
Vihman, Marilyn M. and Shelley L. Velleman
1989 Phonological reorganization. *Language and Speech* 32: 149–170.
2000 The construction of a first phonology. *Phonetica* 5: 255–266.
Watkins, T. Arwyn
1993 Welsh. In: Martin J. Ball (ed.), *The Celtic Languages*. London: Routledge.

Williams, Briony
1986 An acoustic study of some features of Welsh prosody. In: Catherine Johns-Lewis (ed.), *Intonation in Discourse*. London: Croom Helm.

Flexibility in the face of incompatible English VOT systems*

James M. Scobbie

The Voice Onset Time (VOT) cue to the /p/-/b/ voicing contrast in Shetland Isles English was found to demonstrate a high degree of interspeaker variation, which was non-arbitrary on two counts. First, there was an inverse relation between the amount of prevoicing for /b/ and the aspiration duration for /p/. Second, the most vernacular contrast, between typically prevoiced /b/ and relatively short lag /p/, was generally found in speakers whose parents were also native Shetlanders. These results suggest that indexical and phonological cues are simultaneously present, that phonetic targets for VOT are specified in fine detail by speakers, and that they are not restricted to using a fixed low number of VOT categories. The relevance of the results to exemplar models of the mental lexicon and the relationship of phonetics to phonology is discussed, and arguments are presented that both categorical and gradient approaches to phonology must be pursued.

1. Introduction

1.1. Language-specific and speaker-specific sound systems

Features remain central to theories of phonological categories. The idea that there is a universal set of features, however, becomes harder to defend the more it is discovered how much of learned linguistic competence is bound up in the specification of language-specific phonetic detail. In their review and preview of phonology as a laboratory science, Pierrehumbert, Beckman, and Ladd (2000: 285) state: "there are no two languages in which the implementation of analogous phonemes is exactly the same. When examined in sufficient detail, even the most common and stereotypical phonetic processes are found to differ." For example, a recent study of the fortis vs. lenis vs. aspirated stop contrast in Korean concludes: "it would take a great deal of procrustean effort to force Korean stops into the categories that have been developed for phonological descriptions of other languages" (Cho, Jun and Ladefoged, 2002: 222).

Subtle variation between cross-linguistically analogous phonemes is, however, a limited source of evidence against universal features. Each phoneme exists in the context of a wider system, and different languages have (by definition) their own syntax and lexicons and (in practice) their own phonotactic, allophonic and prosodic systems, even if they have absolutely identical phonemic inventories. Therefore, it would be advantageous to examine the phenomenon of fine-grained phonetic differences and their relevance to phonological categories not just in a cross-linguistic situation, but also in the context of linguistic systems that differ only in minimal ways. Such sister systems are, of course, dialects of the same language.

The new consensus – that significant amounts of phonetic detail are language-specific – can be dealt with theoretically in more than one way. A more conservative view holds true to the traditional modularity of phonology. It posits that the observed cross-linguistic differences in analogous phonemes and features reside outside phonological competence, say in phonetics, or in performance. Analogous phonemes may therefore be identical (i.e. have the same feature content), but are accompanied by a sophisticated language-specific phonetic grammar of phonetic realisation rules (e.g. among many, see Cho *et al*., 2002). A modular approach to phonology and phonetics does not seem to provide the conceptual framework for understanding *how* the mass of language-specific fine phonetic knowledge interacts with universal feature systems, and it predicts that a discrete interface between the relevant modules should exist.

A more radical position is that language learners create their own phonological and phonetic (gradient and categorical) systems, prompted by statistical patterning in the input and universal linguistic and cognitive predispositions. On this view there is continuity between language-specific phonetics and phonology. Many aspects of such a view in child and adult studies are decades old (see the reviews in Vihman and Velleman, 2000; Docherty, 1992:55, 75). Recently, however, further research (e.g. Bybee, 2001; Pierrehumbert, 2002; Coleman, 2002; and see the papers in Bybee and Hopper, 2002) has drawn on psycholinguistic "exemplar" or "episodic" models of the mental lexicon (e.g. Goldinger, 1997; Mullenix, 1997; Pisoni, 1997; Johnson, 1997; Hawkins and Smith, 2001), in which multiple detailed exemplars of every word are stored, making the lexicon a structured mix of abstractions and detailed memories of previous speech events and contexts.

A major motivation for exemplar models is that a variety of psycholinguistic experiments have shown listeners can be simultaneously influenced not only by *what* is being said but also by *who* is talking. The implications of this, and of the fact that listeners perceive phonetic details subtle enough to convey both lexical contrast and linguistically non-contrastive indexical factors are that "the indexical and linguistic attributes of speech are not neatly partitioned" (Pisoni, 1997: 11) into different modules. The *same* phonetic parameter can convey *both* types of information.

Exemplar models are of particular interest to phonologists, sociolinguists and others because speakers, when they store traces of individual words, are forced (by the sheer numbers of such exemplars, perhaps) to generalise across them. Such generalisations must be relatively abstract because functionally-similar exemplars are not phonetically identical: variability "causes the need for abstraction" by the learner (Pierrehumbert, Beckman and Ladd, 2000: 292). Such abstractions are presumably made on the basis of similarity in sound and articulation on the one hand and/or similarity in function and context of use on the other (where the nature and limits of "similarity" crucially must be elucidated). If these reflect the systematic nature of language, basic phonological phenomena such as phonemic contrast will tend to induce abstractions similar to those posited by traditional phonological analyses. Such a model predicts the existence of "recurrent" features: fuzzy categories with a close family resemblance will tend to emerge without being universal. Phonemic and indexical functions can be conveyed along the same acoustic parameters and either may be distributed more or less categorically.

Many phonetic (e.g. Docherty, 1992; Browman and Goldstein, 1992; Lavoie, 2001) and sociophonetic studies (Thomas, 2002; Labov, 1994; Foulkes and Docherty, in press) have shown just how fine-grained the control of phonetic and phonological variation actually is – *within* what is often considered a single dialect. So, in addition to the continuum between indexical and contrastive functions, there is a continuum of indexicality, from idiolectal variation, through micro-dialectalisms, to variation characteristic of larger and more disparate groups of speakers. The exemplar type of model predicts that *sociolinguistic* expectations should also influence the perception of contrast, for the issue of "who" is talking relates to groups as well as to individuals.

Nevertheless, and despite a large body of research arguing for continuity from phonetics to phonology,[1] many linguists have tended to hold to the conservative modular position presented above: that while speaker-specific, dialectal, stylistic or phonetic differences vary along continuous scales, phonological features are discrete and universal. Further evidence is presented here that distinctive features are recurrent, are created during acquisition as a response to complex and possibly competing environmental targets, and are intimately bound to indexical functions of language.

In exemplar models, different speakers' "sound systems"[2] end up being similar to the extent that their experiences are similar.[3] Broad generalisations are just those that have previously been – and will therefore continue to be – internalised by large numbers of speakers. Individual speakers can vary from crosslinguistic norms, within statistical limits. In this paper the focus is on the systematicity and flexibility of such normal variation in speakers whose target language offers more than one possible feature analysis of a contrast which they all share.

1.2. Voice Onset Time from a cross-linguistic perspective

On the basis of a major cross-linguistic study, Lisker and Abramson (1964; 1967) proposed the universally-available parameter of Voice Onset Time (VOT), comprising three acoustic categories to cue VOICING contrasts for stops in word-initial position.[4] The categories are voicing lead, short lag (voiceless unaspirated) and long lag (voiceless aspirated). More recently, Cho and Ladefoged explored a wider range of languages under a consistent methodology and concluded (exemplifying the point with specific values for velars): "it is not at all clear that there are just two phonetic categories [in addition to voicing lead] from which languages can choose… it would certainly be plausible to say that there are four phonetic categories, one around 30 ms representing unaspirated stops, another around 50 ms for slightly aspirated stops, a third for aspirated stops at around 90 ms, and a fourth for … highly aspirated stops" (Cho and Ladefoged, 1999: 223).

Lack of any universal boundary between aspirated and unaspirated categories is independent of the questions of whether there *are* recurrent categories of Voice Onset Time, and if so, *how many* there are. If four modal values of positive VOT are observed crosslinguistically, Cho and Ladefoged take a rather agnostic stance on the implications for feature theory: "we consider what might appear to be phonetic categories as *at best* modal values within the continua formed by the physical scales – the parameters – that define each feature" (Cho and Ladefoged, 1999: 225 [emphasis mine]). In any case, further research is required to identify such general crosslinguistic tendencies for categorisation.

Cho and Ladefoged's survey compares only those phonological categories *distinguished by VOT "alone."* They adopt the standard phonological assumption that when languages have more than two homorganic stops with measurably distinct voicing lag, in all but two cases the VOT differences must be a "secondary" cue which "enhances" some other contrast, in voicing, place, constriction, or airstream mechanism. Cho and Ladefoged (1999) and Docherty (1992) are both quite clear there are not just three universal *phonetic* targets for VOT. Consequently, they seem to disagree with the position that there is a "fixed and universally specified set" (Keating, 1984: 289) of VOT targets relevant to phonetics and phonology alike. For example, Cho and Ladefoged show that VOT for EJECTIVE /k/ tends to be intermediate between UNASPIRATED /k/ and ASPIRATED /k/, but conclude that the intermediateness of EJECTIVE /k/ is not proof that there is a phonological category of "medium lag". Rather, VOT must be a secondary cue for one of the three types of /k/ (EJECTIVE in this case). Clearly, there is a danger of such argumentation being circular. A variation on this argument was proposed in a recent very detailed analysis of the three-way voiceless stop contrast in Korean (a system recognised as a challenge for the VOT system by Lisker and Abramson, 1964). Cho, Jun, and Ladefoged (2002) proposed that one of the three stops (the lenis one) is unspecified, and that its VOT target is specified *phonetically*.

In the exemplar model there is no a priori demarcation between primary and secondary: instead, functional reasons are expected for any patterns observed. For example, there is likely to be a diminishing reliance on VOT as a cue to the identity of all members of a set of homorganic stops if the VOT targets get to be too close together or overly numerous. This avoids the claim that languages specify a precise maximum of two positive VOT targets, and that a greater number of distinct VOT targets must be due to a categorical distinction between VOT as a primary vs. enhancing cue (or that at most two positive VOT targets can be specified by phonology and all others are specified by a different phonetic module).

It seems that Docherty (1992) was right to question the value of rigid phonetic classifications such as voiced, unaspirated and aspirated. It is theoretically more useful to present VOT values numerically in tandem with the relevant phonological contrasts. This may be particularly valuable for those languages that have no VOICING contrast for stops at all, including those in which phonetic voicing is contextually conditioned. One major outcome of Cho and Ladefoged's paper is that there are no clear demarcations in VOT to classify such stops in a universal or deterministic way.

1.3. VOT variation in standard and non-standard English

Previous studies of VOT within a single language have mainly investigated the influence on VOT of linguistic factors such as prosodic context, place of articulation, segmental context, organic factors such as sex, age, pathology, and intermediate factors such as speech rate and dialect in a very broad sense. But "arbitrary" interspeaker variation has, of course, also been observed. A recent example from English is a study of eight General American speakers which found VOT for post-pausal /p/ varied from about 60 ms to about 110 ms (Allen, Miller and DeSteno, 2003). Normalising for speech rate reduced the indexical interspeaker variation, but Allen et al.'s conclusion was nevertheless that VOT varies and may function indexically, in line with exemplar models. The source of the variation could not be identified, because the subjects were a homogenous group. One speaker's normalised mean VOT for /ptk/ was, at 74 ms, distinct from the other more closely grouped seven speakers, but all the VOICELESS stops were what is normally called long lag.

Other studies have found variation in word-initial, post-pausal position in the occurrence of prevoiced and short lag variants of /bdg/ (Westbury, 1979; Docherty, 1992; Smith, 1978; Lisker and Abramson, 1964; Flege, 1982). Prevoicing is *normal* in this position in English as has pointed out by, for example, Westbury (1979) who found 62% of /b/ tokens were prevoiced; Smith (1978), 56%; and Flege (1982), 59%. As Lisker and Abramson (1967: 5,7) observe, in American English "it appears that while /ptk/ have distributions that are essentially unimodal, /bdg/ show values that fall into two discontinuous ranges, with modes at about –100 msec and near zero." Their interpretation is that English speakers select from one or both of "two phonetic categories of /bdg/... In fact the relation between the two distribution modes for each member of the /bdg/ set is not *overtly* different from the relation between phonemically distinct categories in certain other languages" (Lisker and Abramson, 1967: footnote 13 [emphasis mine]). The cautious nature of this statement about a possible equivalence between English allophonic categories and contrastive categories in other languages has perhaps been overlooked in an enthusiastic search for a universal model of distinctive and redundant features.

The existence of positive and negative VOT values is taken as evidence that English uses two categories for VOICED stops. Further support comes from interspeaker variation in the use of prevoicing. Some speakers appear to use prevoicing, others to avoid it, and others to vary freely. Lisker and Abramson (1964; 1967) discuss such speaker-specific use of prevoicing in some detail. Docherty (1992) also found such variation in his comprehensive and detailed study of obstruent voicing in five speakers of Southern Standard British English. Recently, age, race/ethnicity and sex have been shown to be conditioning factors for VOT, such that "there appear to be many more instances of prevoicing" than is normally acknowledged (Ryalls, Zipprer and Baldauff, 1997: 644).

Prevoicing is also a common characteristic of the speech of many English L2 speakers and natively bilingual English speakers, and their (grown-up) children.[5] This is to be expected if sound systems are able to differ in very subtle ways. For example, in a detailed longitudinal study of English/Arabic bilingual pre-teenagers, Khattab found very fine differences in VOT between the analogous phonemes of each language. The bilinguals had very slightly different language-specific targets when using the "same" phonetic category. Her conclusion was that "there are important phonetic differences between English VOICED and Arabic VOICELESS stops involving divisions that are finer than the boundaries suggested by the three supposedly universal categories, and suggesting that neither of the two broad categories 'short lag' or 'long lag' can adequately describe Lebanese VOICELESS stops" (Khattab, 2003: 290).[6]

In order to combine these findings on very fine within-language variation with Cho and Ladefoged's more dramatic cross-linguistic results, this paper examines the VOT parameter in a single speech community (the Shetland Isles of Scotland) in which the phonology and lexicon can be held fairly constant while VOT can be expected to vary across categories. The traditional Shetlandic system has a voicing lead target for /bdg/ and a short lag voice onset target for /ptk/, whereas the majority of standard English varieties have voicing lead and more often short lag voice onset for /bdg/ and a long lag voice onset target (i.e. aspiration) for /ptk/. On standard assumptions, these are *incompatible* VOT systems because short lag VOT is ambiguous as to whether it cues VOICED or VOICELESS stops. The implicit decisions taken by different individuals in such a situation should reveal something about the phonetic options that are available to all speakers.

2. Method: interspeaker variation as an experimental resource

The language reported in this paper is English, specifically those varieties of Scots and Scottish English spoken in the Shetland Isles.[7] The Shetlands are an isolated archipelago of about 15 inhabited and more than 80 uninhabited islands, about 300km (180 miles) north of the Scottish city of Aberdeen, a major transportation link, and 350km (220 miles) west of Bergen in Norway. The three biggest settlements are Lerwick (population circa 7,500), Scalloway (c. 1,000) and Sandwick (c. 900), all on the same island.

In recent decades, historic depopulation has been reversed in the Northern Isles (Orkney and Shetland). This has been particularly dramatic in Shetland, where there has been significant immigration due to the oil industry: the population swelled by about a third in the decade leading up to 1981. Though the population had fallen back from this peak by 1991, it has remained relatively steady (21,988 in the census of 2001). The net increase from 1971 is made up in part by incomers who settled in the islands and subsequently had children. Thus, at the time the data was collected (1999), a significant minority of young adult native Shetlanders had parents who themselves were incomers from England and elsewhere in Scotland.

Contemporary vernacular Shetlandic preserves many relic forms of older Scots, perhaps including the Scots instantiation of the VOICING contrast: "early authorities [from the 1930s and 1940s] are united as to the unaspirated nature of Scots voiceless stops in syllable onsets" (Johnston, 1997: 505). Scottish-accented Standard English on the other hand has long-lag VOT (aspiration). Traditional Scots VOT patterns can indeed still be heard in vernacular varieties throughout Scotland, though "aspirated ones equal to the equivalent /ptk/ allophones in other English dialects" are spreading into Scots (Johnston, 1980: 78). VOT in Southern British English /p/ VOT in post-pausal, word-initial position is around 47 ms (Hawkins, 1979), 40 ms (Suomi, 1980), or 46 ms (Docherty, 1992). There are no comparable published figures for Scottish Standard English, though Docherty's review of VOT studies suggests that, in initial position, long lag /p/ is unlikely to be under 40 ms. However, consider the very small study of VOT in vernacular southern Scots (Masuya, 1997). Unexpectedly, only one Scottish English speaker (Sp19) out of 20 had unquestionably short lag /p/. However, since there are only two tokens per subject, this cannot be given much credence. Of greater reliability and relevance are the pooled results, which show that the Scottish subjects have a *slightly* lower VOT than the English controls: phrase-medially, lowland Scottish /p/ was 32 ms (n=29, s.d. 16 ms), while English /p/ was 48 ms (n=10, s.d. 17 ms). This Scottish mean is not easily classifiable as being either "short" or "long".

Children growing up in the Shetlands in recent decades are likely to have been exposed to short lag and long lag variants of /p/, and short lag and prevoiced variants of /b/. Children growing up in families where the parents are non-Shetlandic incomers would have had to face incompatible functions for the same VOT value. The meaning of short lag stops would vary depending on the speaker. As a result, wide interspeaker variation in VOT is to be expected in young Shetlanders. Yet all speakers would typically be regarded by linguists as having the same contrast at an abstract level: there is no *phonological* conflict to resolve in acquisition or in adult communication. The incompatibility of different systems is likely to lie in the location of the VOT targets and perhaps in whether VOT is a robust cue to the contrast at all. If human language provides only three universal categories of VOT, then we should expect to see categorical variation.

Exploration of such a situation demands a fundamental departure from normal practice: it requires subjects who might be *expected* to vary in their systemisation of the phonetic or phonological phenomena under consideration along partially predictable non-linguistic lines. Augmenting standard experimental phonetic methodology with *socially structured* groups of subjects offers a new basis for advancing phonological theory. Two linked strands of previous research have greatly influenced the adoption of such a methodology here. One has been the work of Docherty and colleagues (Docherty, Foulkes, Milroy, Milroy and Walshaw, 1997; Docherty and Foulkes, 2000; Foulkes and Docherty, 1999). The other has been collaborative work with Stuart-Smith (Scobbie, Timmins, Stuart-Smith, Tweedie, Hewlett and Turk, 2000). However, while many instrumental phonetic studies of VOT have addressed a wide range of sources of variation (see above), social stratification has tended to be seen as the preserve of sociolinguistics. This is despite the fact that social variation can be a more delicate tool than cross-linguistic variation for probing subtle differences in sound systems.

The analysis is based on wordlist recordings made by Marie Cluness, a native Shetlander (then aged 20). The wordlist comprised a random ordering of 270 single words, intended to enable study of a number of dialect variables (Cluness, 2000; Scobbie, 2005). On a number of grounds, the subjects might be expected to have *similar* accents: all had been born in the Shetlands and lived there all their lives; all lived in the same geographical area (Westside); six were final year students at the same high school; eleven were 16 to 22 years old; many were acquainted with each other (and Cluness). (Indeed, S9 and S11 are brothers.) However, one major structuring factor was introduced to provide a basis for possible interspeaker differences (Table 1). The subjects form three groups based on the geographical origin of their parents (and, therefore, by hypothesis, the parental accents). Crucially, one group (S1–S4) had parents who themselves were native Shetlanders and so were likely to have broader, more vernacular accents than members of the other two groups.

Table 1. Subject identifiers and age details coded by sex and group.

Parents born and raised in	Shetland	Scotland	England
Male Shetlanders	S1: age 20 S3: age 21	S5: age 16 S7: age 17	S9: age 20 S11: age 19
Female Shetlanders	S2: age 30 S4: age 22	S6: age 17 S8: age 17	S10: age 17 S12: age 17

VOT is an acoustic measure which can be defined in a number of ways (cf. Fischer-Jorgensen and Hutters, 1981; Docherty, 1992: 24). Here it is the difference in time from the burst at the end of stop closure to the onset of the periodicity, which continues into the following vowel.

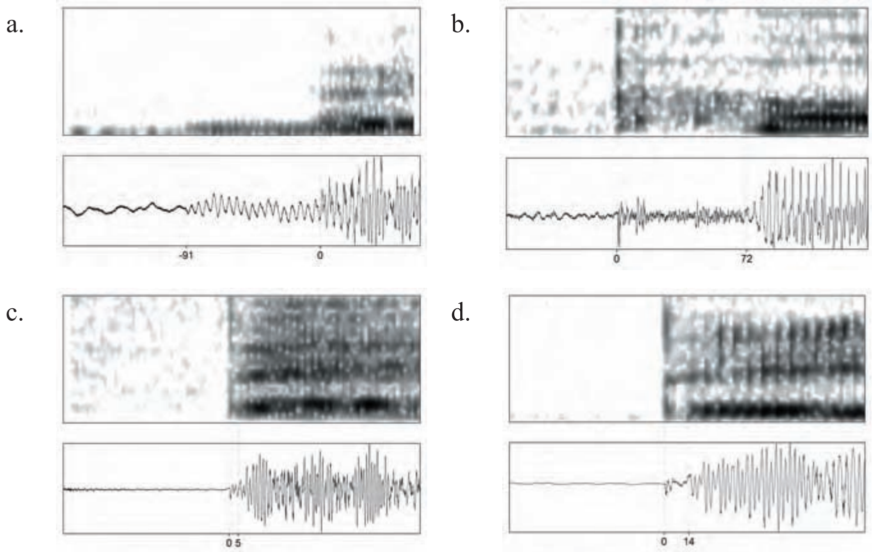


Figure 1. Annotation of VOT ranging from pre-voiced to long-lag aspirated. The x-axis in each panel is 200 ms, with 0 ms at the burst. The timing (in ms) and direction of VOT is indicated. The y-axis of each spectrogram is 0–5kHz. a. S4 brood [brʉd̥], b. S2 pod [pʰɔd̥], c. S10 both [peθ], d. S1 piece [pis].

If periodicity was detected following the release only, it was recorded as *positive* VOT (Figure 1b,c,d). Typically, VOT was easy to identify (Figure

1d). If the stop had a particularly energetic release, 5–10 ms of frication energy were present, perhaps obscuring the initiation of voicing. In common with other acoustic studies, this effect presumably boosts mean VOT in unaspirated stops by a few milliseconds. In 18 cases (1.5% of the data) a weak burst was simultaneous with the onset of periodicity, so a VOT of zero was assigned. These tokens were classified as examples of voicing lag. Of these tokens, 17 were /b/ and one was /p/, and two subjects (S7 and S6) provided 11 of the examples.

If periodicity was first detected at a point clearly before the stop release, it was recorded as *negative* VOT (Figure 1a). There was no requirement, however, for voicing to be maintained from this VOT point until the burst and beyond, though this occurred in most cases. In some tokens, periodicity died away again before the burst, and only began again at or just after the release.

3. Results

3.1. Pooled results

Consider mean VOT for /p/ and /b/ for all subjects pooled (Table 2). The VOICELESS stop /p/ has a long voicing lag. The VOICED stop /b/ shows voicing lead, i.e. is prevoiced. The mean difference in onset time cueing the VOICING contrast is 85 ms. Note the very large ranges and the large standard deviation for /b/ in particular, which will be explored in detail below.[8] About half (52%) of the VOT measures for the VOICED stop /b/ were greater than or equal to zero. This proportion is broadly comparable to previous findings for /b/ discussed above (§1.3).[9]

Table 2. Voice Onset Time (ms) based on pooled data from all subjects.

	mean	s.d.	count	min	max
/p/	56	24	280	0	112
/b/	–29	51	334	–190	41

As discussed above in §1.3, the presence of positive and negative VOT in English has been taken as evidence that there are two different phonetic (acoustic) categories: voicing lead and short voicing lag. Stronger evidence for this conclusion is the apparently *bimodal* distribution of VOT in Figure

2, rather merely than a unimodal distribution spanning both negative and positive territory.

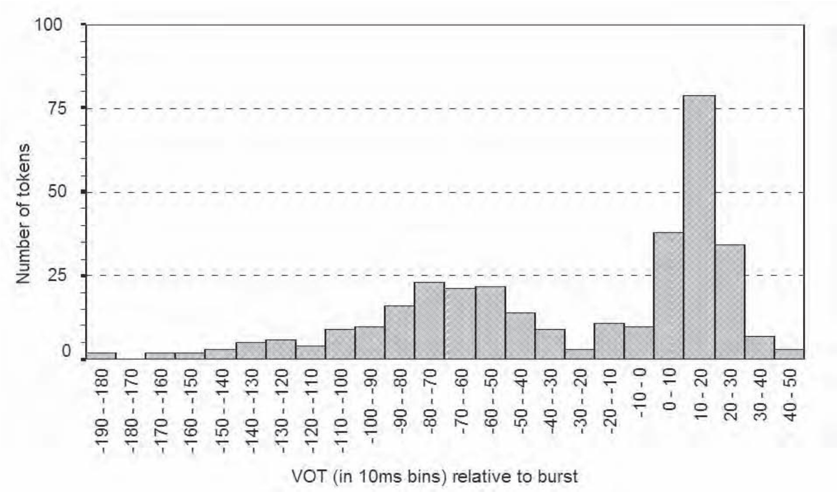


Figure 2. Histogram of Voice Onset Time (ms) for /b/, all subjects pooled.

The standard solution to this distribution is to separate VOT into positive and negative ranges. Such an approach provides two means for VOT: positive and negative. These are usually assumed to meaningfully reflect the VOT targets of the two modes of articulation (Docherty, 1992). The stop release (which defines this positive/negative split) initiates the emission of trapped intra-oral air, hence the *beginnings* of the increase in trans-glottal pressure differential which is necessary for phonation, so 0 ms VOT as the boundary point is articulatorily meaningful, but its use is based at least in part on convenience. An empirical alternative – selecting a boundary point based on the *actual* distribution of the data – is clearly less convenient, because it would be impossible with small sample sizes, might be nondeterministic, and would be specific to the speaker, task, style, place of articulation, language, and so on. On the basis of the 10 ms bins in Figure 2, it would be somewhere between –30 ms and –20 ms. Even if the adoption of 0 ms as a unique boundary point between voicing lead and voicing lag modes is less than ideal, using it has certain practical advantages: it lets us compare results with previous studies of English and can also be applied deterministically in individual cases or small samples.

Table 3 presents revised pooled results for /b/ on this basis. Standard deviations are, unsurprisingly, much lower than those in Table 2. Negative

VOT has a wider range and greater standard distribution than positive VOT, a characteristic visible in Figure 2. To the extent that the contrast between VOICED and VOICELESS stops in Shetlandic is cued by two degrees of positive VOT (short vs. long lag), a tight distribution of short lag VOT for VOICED stops would likely occur for perceptual reasons. Such factors are relevant even if half the VOICED stops have voicing lead, and even though the voicing lag /p/ and /b/ ranges overlap.

Table 3. Voice Onset Time (ms) for /b/ calculated from pooled negative VOT data for voicing lead /b/ and pooled positive VOT data for voicing lag /b/.

	mean	s.d.	count	%	min	max
Voicing lead /b/	–71	38	172	52	–190	–2
Voicing lag /b/	15	9	162	48	0	41

It is well-known that that aspirated stops show large amounts of glottal abduction, which delays voice onset and makes the closure voiceless, but Flege (1982) is unusual in examining post-pausal word initial VOICED stops. He showed that such stops in American English may or may not be adducted. One subject in that study (his S8) produced tokens of /b/ which were nearly always acoustically voiceless yet with vocal cords adducted suitable for phonation, while some produced voiceless unaspirated /b/ without adduction. Three subjects varied between prevoicing and short lag, but there was no simple laryngeal/acoustic relationship: "much of the variation in the presence/absence of prevoicing observed in utterance-initial voiced stops is due to factors other than variation in laryngeal timing" (Flege, 1982: 189).

It is also well known that phonation is unlikely to begin towards the *end* of a stop closure: as seen in Figure 2 there are generally few tokens with small negative VOT. Flege is skeptical that this is simply due to aerodynamic conditions within the vocal tract tending to make phonation harder to initiate as the supraglottal pressure builds up. Rather, the discontinuous acoustic pattern may be due to a combination of factors such as the strength and timing of glottal gestures, allied to the extent to which a speaker suppresses "passive" devoicing. In the Shetland data, there are a relatively small number of cases in which prevoicing is initiated (at about –100 ms or so), only to become attenuated before the stop release. The lack of such cases argues against high intraoral pressure being solely responsible for the lack of voicing initiation just before the burst.

It is possible, however, that the bimodal distribution in Figure 2 may come in part from unimodal variation in the underlying articulatory strategies (cf. Pierrehumbert and Talkin, 1992). The bimodal nature of Figure 2 does not tell us whether the acoustic distribution reflects "intentional" targets or is the "accidental" output of other types of underlying variation that may even be unimodal. Either way, the acoustic distribution itself is an aspect of grammar – it is not *necessary* for speakers to produce prevoiced tokens, but some do. The observed VOT patterns are therefore linguistically relevant and must reflect the internalised grammar of speakers. This conclusion is in keeping with exemplar theory, which tolerates such ambiguity and non-determinism.

3.2. Individual results

The main benefits of considering individuals are: first, that we can determine how many categories each speaker uses, which lets us probe small systematic differences; and second, that categories *in use* can be compared within and across speakers in order to reveal patterns to replace the unstructured variation of the previous section. The analysis developed below will be based on preliminary individual results for three straightforwardly definable units, namely /p/, positive VOT /b/ and negative VOT /b/.

Table 4 shows that the great variability in the pooled results in Table 2 is due to individuals having their own VOT targets, including distributions above and below 0 ms for /b/. Note that *all* these individual categories have a lower standard deviation (not shown) than the pooled results. Two speakers stand out: S1 has a strict categorical opposition of prevoiced /b/ and short lag /p/, exemplifying the traditional Shetlandic VOT system. S12, on the other hand, has short lag /b/ vs. long lag /p/.

It might seem that in addition to these two, the ten other speakers use two categories for /b/, and that no more needs to be said. However, though all have positive and negative VOT, some speakers have only three or four tokens of one category but 20 or 25 in the other. Consequently, a threshold will be used to allow us to discover which categories are *most clearly* used. If five or fewer tokens are positive or negative (always <20%), they will be treated as noise and not evidence of a category. This threshold is justified by an obvious natural discontinuity in the data itself: the least populated category beyond the threshold is S10's prevoiced /b/, with nine tokens (32%).[10] Its location is arbitrary, but the use of a threshold reduces the arbitrariness of the final analysis, because S1 and S12 are not treated as having a completely different sound system to the other subjects merely on the grounds of the presence or absence of a single example of positive or negative VOT. This analysis of the raw data means that S3 and S9 join S1 as speakers who prevoice /b/ and do not use short lag VOT much.[11] S5, S7 and S6 join S12 as speakers who do not have a clear prevoicing category. The remaining five present the strongest evidence that the use of both positive and negative VOT for /b/ is typical and systematic.

Table 4. Mean Voice Onset Time (ms), all subjects, for /p/, and for both positive and negative /b/. Values in italics are revised in the text. Subjects S1–S12 are presented in rank order of increasing mean VOT for /p/.

	S1	S4	S9	S3	S2	S11	S7	S6	S12	S8	S10	S5
/p/	22	33	36	37	47	55	61	69	73	78	81	83
/b/		9	*18*	*17*	14	16	*15*	*15*	17	23	5	17
/b/	–64	–60	–64	–60	–62	–87	*–9*	*–11*		–96	–121	*–76*

One final adjustment to the results for /b/, however, can be made. For two of the speakers with only a few negative VOT tokens (S7 and S6), all eight tokens were clustered between –4 ms and –15 ms. As noted above, these speakers were already unusual in that a number of tokens were excluded on the grounds that no release could be detected. It thus seems likely that they cue /b/ in part through lenition, or with a very low intra-oral pressure which enables voicing to begin just before the point identified as the release during annotation. Consequently, all their measured tokens are analysed as examples of short lag voicing, even though some are below 0 ms. The revised means (and standard deviations) for /b/ are therefore amended (slightly) to S7: 12 ms and S6: 11 ms. Their mean VOT is now a bit lower and a bit more variable, in keeping with their actual behaviour. Figure 3 presents the results of this analysis.

Each speaker has either one or two categories for /b/, as represented by positive and negative means. All speakers are assumed to have a single category for /p/, because speaker-specific histograms of VOT provide no strong evidence to the contrary; but such a technique really requires much more data.[12] Note, however, that the histograms seem to show that even the three subjects in the middle of the distribution of /p/ (S11, S7 and S6) are not varying between short and long lag targets. Despite the fact that these subjects have the relatively high standard deviation of 17 ms compared to the other subjects' range of 10–14 ms, the minimum values for S7 and S6 are both above 35 ms, suggesting within-category variation. S11 has a single token at 14 ms, one at 27 ms, and the rest above 34 ms.

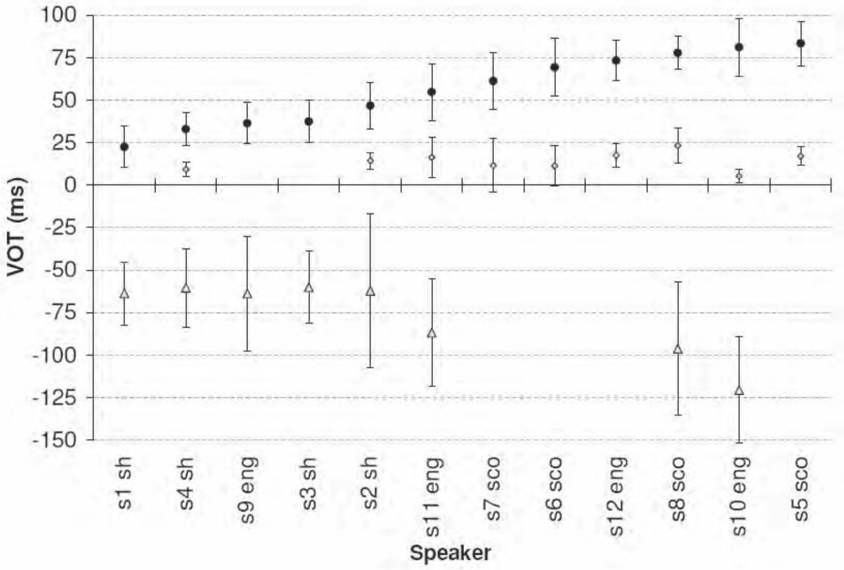


Figure 3. Mean Voice Onset Time (ms) relative to burst, with whiskers indicating one standard deviation for each speaker for the categories identified in the text. Circles indicate /p/, diamonds and triangles indicate positive and negative means for /b/.

What, then, of each speaker's /p/? The gradient increase in mean VOT for /p/ across the subjects means that there is no basis for allocating /p/ into short or long lag categories on the basis of data on /p/ alone. This conclusion is similar to that of Cho and Ladefoged (1999), but arises from within a language, so avoids the problems of interpretation attaching to crosslinguistic studies. It is also far more radical than the intercategory differences revealed by Allen et al. (2003). It is possible, however, to make some judgments about the categorical affiliation of /p/ on the basis of the distribution of VOT means for /b/, an analysis which neither was able to undertake. In particular, the hypothesis that there are only three universal categories of VOT predicts that the subjects with short lag /b/ *must have*, by definition, long lag /p/. This is not problematic for the contiguous eight-strong group of subjects between S2 and S5 in Figure 3 and Table 4. But S4 clearly has short lag /b/, so her /p/ on this basis should be long lag, though it is only 33 ms.[13] This creates problems for the analysis of S9 and S3. Are VOT means of 36 ms and 37 ms for /p/ (in opposition with heavily prevoiced /b/) to be categorised as short or long lag?

Consider now the claim that short lag stops are ambiguous. In Shetlandic a short lag stop can represent either a standard VOICED stop (Figure 1c) or a vernacular VOICELESS one (Figure 1d). However, VOT is slightly greater on average when it cues /p/ and less when it cues /b/. The lowest VOT for /p/ (S1, 22 ms) is right at the top end of the scale of the nine speakers with short lag /b/, whose means range downwards from 23 ms to 5 ms. This effect can neither be put down to cross-linguistic differences nor to experimental artifact. The small VOT differences between short lag /b/ and /p/ are observed in speakers of the *same language*, from the *same speech community*, using the *same phonemic contrast*, and reading the *same materials*. Koenig (personal communication) points out that, since phonetic models must be able to deal with the combination of phonetic cues, the VOICELESS specification of /p/ may make a short lag target for VOT a little longer (i.e. relatively more voiceless) than the "same" VOT target for /b/. The idea that the "same" value of short lag VOT could be either VOICED or VOICELESS may express no more than the idea that a VOICING contrast will never be cued *solely* by a tiny VOT difference.

VOT may well be predictable from the phonological specification of the phonemes concerned, but it is likely that it is a robust indicator of the contrast – it just happens to be in different parts of the continuum for different speakers. In Shetlandic English, the "/p/ vs. /b/" contrast seems to be shared at some abstract level, while individual speakers encode their own VOT targets. This offers a challenge to feature-based modular theories, in which VOT must be either distinctive or redundant, if phonological, or be indexical. The exemplar model, however, denies that such strict delineations are necessary, and assumes each part of the system is linked more or less directly to the other relevant cues that might pertain.

Furthermore, if short lag VOT is different in /b/ and /p/, then what is to be gained by positing at most three oppositions in VOT? For example, in the case of subject S4 there are two voicing lag targets very close together in the short lag range. *Both* /b/ and /p/ may be "short" lag, or /p/ may be a phonetically short "long" lag, if we are forced to assign categorical identities. But what surely matters is that short positive VOT may be less robust (not simply either distinctive or redundant) as a cue to the contrast for speaker S4. It is ultimately unhelpful to assume that a category *label* such as "short lag" in the absence of (or even in addition to) the actual quantitative details captures all the important parts of the system, a point made strongly by Docherty (1992).

Shetlandic short-lag labial stops do not exist in isolation, but function as one pole of a binary opposition. We can hypothesise that the more prevoicing for /b/ that a speaker uses, the shorter their VOT for /p/ can be without causing perceptual confusion. Conversely, if a speaker uses very little prevoicing for /b/, their VOT target for /p/ is more phonetically distinctive if it is long lag. In Figure 4 we may be able to see some evidence to support these hypotheses. There is generally an inverse relationship between the rate of prevoicing and the mean VOT for /p/, although there are two subjects (S8 and S10) who contrast aspirated /p/ with a /b/ which is prevoiced in about 60% of cases.

This is not to rule out a simultaneous categorical analysis, so long as it can be approached flexibly. With reference both to the raw data plotted in Figure 4 and the analysis in Figure 3, it may be that S10 and S8 could be classified as having a mix of VOICED and VOICELESS /b/ plus ASPIRATED /p/. Subjects S5, S6, S7 and S12 could have VOICELESS /b/ and ASPIRATED /p/. Subjects S3, S9 and S1 could have UNASPIRATED /p/ with VOICED /b/. Subject S4 could have UNASPIRATED /p/ and vary in VOICING for /b/. This still leaves some doubt about S2 and S11.

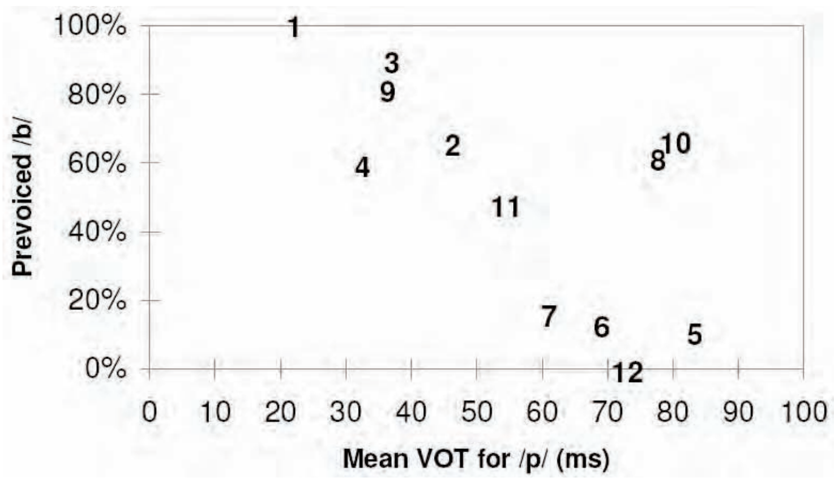


Figure 4. Relationship of rate of prevoicing of /b/ (%) to mean Voice Onset Time (ms) of /p/ for all speakers (identified by code number), based on the raw data in Table 4.

Even if a categorical interpretation of Figure 4 were possible, so that all variation were either categorical feature switching or phonetically gradient, as the more conservative approaches to fine phonetic differences demand, three problems remain. First, the categorisation of some subjects such as S11 in particular remains highly ambiguous. Consider also S4. Since 40% of her /b/ are voiceless, perhaps her /p/ is VOICELESS ASPIRATED even though it is only 33 ms and even though S9 and S3 have a longer VOT for their /p/ which could be VOICELESS UNASPIRATED. Second, more than one binary feature would be required to encode /p/ vs. /b/ for the group of speakers in order that short lag /b/ is distinguished from short lag /p/. Third, there appears to be a gradient inverse relationship between prevoicing for /b/ and aspiration for /p/ that exists *within-category*, independently of any particular categorisation that might be imposed.

Within a modular framework, a categorical analysis ought to reveal the simpler structure underlying variable data, not replicate in an impoverished way the detailed linguistic distinctions and patterns of variation that the speakers make. An exemplar approach, on the other hand, predicts flexibility among speakers and permits flexibility in our linguistic analysis between gradient and categorical aspects of sound systems.

4. Summary and conclusions

As expected, the distinction between prevoiced /b/ and unaspirated /p/ is present as a phonemic contrast in Shetlandic English, for at least one speaker. Generally there is an inverse relationship between the rate of prevoicing for /b/ and the VOT target for /p/, though two speakers seem basically to oppose long lag /p/ with substantial amounts of prevoicing for /b/. It is unclear whether this inverse relationship reflects a gradient tendency within a single system of /p/ vs. /b/, or two inverse relationships within two categorically distinct systems (which we might call [b] vs. [p] and [p] vs. [p^h]). In a modular approach to phonology and language-specific phonetics a decision would have to be made on this issue, but we do not seem to have reliable methods for doing so, a difficulty reflected in previous discussions by Cho and colleagues. This may be a general methodological failing, but it may instead be seen as support for the view (e.g. Browman and Goldstein, 1991; Ohala, 1990) that there is no strict interface between phonetics and phonology – as support for "meta-gradience." This would mean that the distinction between gradient and categorical phenomena is itself gradient (Scobbie, in press).

VOT also seems to function indexically. Mean VOT for /p/ falls on a continuum, such that the four speakers with Shetlandic parents reflect the local vernacular (i.e., shorter VOT values for /p/, and pre-voiced /b/) more than most of the other subjects in this study. Interestingly, those whose parents are from elsewhere in Scotland (S5–S8) are at the other extreme on the continuum. English-parented subjects, including the two brothers, appear throughout the continuum. Speakers' systems therefore reflect to an extent the parental and community target systems in addition to arbitrary individual differences. VOT for /p/ and the rate of prevoicing for /b/ are therefore likely to be sociolinguistic variables.

In Shetlandic English, VOT seems to be a robust indicator to the VOICING contrast for each individual, despite the fact that the VOT values for /p/ and /b/ may be scattered through a large and apparently ambiguous region of the relevant phonetic space, when thinking of the community as a whole. The range of positive VOT values suggests that there is little to be gained from the theoretical supposition that phonetically meaningful features exist universally. Rather, speakers share two things: a comparable phonological VOICING contrast in the same lexical items, and the same abilities to respond to the functional demands of production, perception and acquisition of lexical contrast and other aspects of linguistic competence.

The VOICING contrasts observed range along a continuum reflecting traditional Scots and English, or draw on the extreme aspects of both. Such solutions are presumably typical of English-speaking communities the world over in both multidialectal and multilingual situations. These results and those of Khattab (2003) suggest that situations in which learners are faced with distinct – even incompatible – systems can lead to the acquisition of arbitrary targets for VOT, as exemplar theory allows. However, such stochastic models must not limit learners to acquiring a slavish recapitulation of the distribution of raw tokens to which they have been exposed. Higher-level abstraction and analysis by learners is also necessary.

Labov (1994: 25) expresses his desire "to reinforce the natural alliance of dialect geography, sociolinguistics, phonetics and historical linguistics – fields that share a common interest in objective [speech production] data." We can add at least child language acquisition and speech pathology to this list. To date, most of the impetus for a rapprochement between sociolinguistics and laboratory-based phonetics and phonology has arisen due to the demands for quantitative phonetic data in sociolinguistic research. By definition, the field of laboratory phonology also focuses on the descriptive and theoretical importance of subtle phonetic and phonological distinctions.

What has not yet been fully appreciated by the experimental community is the potential value of the more vernacular varieties of English in their own right, or, moreover, the varied systematic relationships which are known to exist between different socially-distributed varieties of a language.

Unwanted variation is a problem for all experimental studies. Variation due to differences between subjects themselves is typically minimised rather than being controlled as an experimental factor. This is explicit in the design when subjects are only selected if they have *a priori* similar accents of a language, typically English. Often, experimental subjects are university colleagues or students: well-educated adults who almost always speak a standard variety. Thus an implicit homogenisation arises because study after study draws on a limited number of English accents. This paper rejects the assumption that the use of homogenous pools of subjects simplifies phonetic and phonological analysis. Instead, the approach to sound system analysis advocated here draws on a structured pool of subjects who can a priori be *expected* to vary linguistically. This structured heterogeneity reveals more about the underlying uniformities of the speakers' linguistic systems, as well as providing new challenges for the theorist.

Notes

* Without Marie Cluness, Jane Stuart-Smith and the Economic and Social Research Council (Personal Research Fellowship R000271195), this research would never have happened. Many thanks to Cathi Best, Laura Koenig and an anonymous reviewer for very careful and helpful comments on an earlier draft, and additionally to Bob Ladd, Gerry Docherty, Paul Foulkes, Ben Matthews, Klaske van Leyden, Lisa Lavoie, and Susanne Fuchs for further comments, discussion, help and inspiration.

1. Consider the enormous influence of Ohala, Articulatory Phonology and research reported in this series. Of particular relevance are the instrumental sociophonetic studies of Docherty and colleagues (Docherty, Foulkes, Milroy, Milroy and Walshaw, 1997; Docherty and Foulkes, 2000), and acquisition research by Vihman and colleagues (Vihman and Velleman, 2000).
2. A sound system comprises all the learned aspects of phonetics and phonology.
3. Such experiences include internal psychological states and innate predispositions as well as real world interactions. Acquisition itself is an experience: the very first abstractions drawn by children influence those that come later, perhaps explaining why the developmental paths which children take are apparently even more varied than the resulting adult systems.

4. Capitalised terms (other than the acronym VOT) indicate possible phonological features but do not assume any specific set.
5. The implications of this are profound for research in acquisition, perception, bilingualism and second-language learning which often assumes, contrary to fact, that English speakers never produce or experience prevoicing. In the USA, the census of 2000 reports that more than 50 million people (18%) live in a household where a language other than English is spoken (in addition to or instead of English). More than half of these people live in a household in which Spanish (a language with prevoicing) is used. English is often acquired as a native language in a multilingual or multidialectal context, and the adult English used may be very different from what is assumed in the literature.
6 . The rate of occurrence of prevoicing in the English speech of older bilingual children (English/Panjabi) has also been found to be intermediate relative to two control sets of monolinguals (Heselwood and McChrystal, 2000).
7. Scots, like English, is West Germanic, and is sometimes said, rather whimsically, to be the language most closely related to English.
8. In all 614 labial stops were measured. The minimum number of tokens analysed per subject for /p/ and /b/ were 17 and 21. All /p/ were singleton consonants, but 107 /b/ were in the clusters /bl/ and /br/. The means for singleton /b/ were –71 ms (s.d. 15 ms), for cluster /b/ –71 ms (s.d. 16 ms).
9. Lisker and Abramson (1964) and Docherty (1992), however, found a smaller proportion of negative VOT, due largely to a single speaker. Docherty argues that this suggests first that individual speakers can choose a particular VOT target and second that interspeaker variation is an important but neglected aspect of the description of a language.
10. Only 11 tokens are excluded (of 334), for S3, S9 and S5. Eight other tokens (from S7 and S6) fall below the threshold but will be reassigned, see the text.
11. Under a modular interpretation, brothers S9 and S11 would now be classed as having different systems, a rather arbitrary conclusion.
12. No speaker has a bimodal distribution for all their VOICELESS stops.
13. S4 does not have a bimodal distribution for /p/, nor do her distributions of /b/ and /p/ overlap, so her low mean for "long lag" /p/ is not due to intrasubject variation between short and long lag targets.

References

Allen, J. S., Joanne L. Miller and David DeSteno
2003 Individual talker differences in voice-onset-time. *Journal of the Acoustical Society of America* 113: 544–52.

Browman, Catherine P. and Louis Goldstein
1991 Gestural structures: distinctiveness, phonological processes, and historical change. In: I. G. Mattingly and Michael Studdert-Kennedy (eds.), *Modularity and the Motor Theory of Speech Perception.* Hillsdale, NJ: Lawrence Erlbaum Associates, 331–338.
1992 Articulatory Phonology: an overview. *Phonetica* 9: 155–80.
Bybee, Joan
2001 *Phonology and Language Use.* Cambridge: Cambridge University Press.
Bybee, Joan and Paul Hopper (eds.)
2002 *Frequency and the Emergence of Linguistic Structure*. Amsterdam: John Benjamins.
Cho, Taehong, Sun A. Jun and Peter Ladefoged
2002 Acoustic and aerodynamic correlates of Korean stops and fricatives. *Journal of Phonetics* 30: 193–228.
Cho, Taehong and Peter Ladefoged
1999 Variation and universals in VOT: evidence from 18 languages. *Journal of Phonetics* 27: 207–29.
Cluness, Marie
2000 An investigation into the parental influence on dialect acquisition. Queen Margaret University College, Edinburgh.
Coleman, John
2002 Phonetic representations in the mental lexicon. In: Jacques Durand and Bernard Lax (eds.), *Phonetics, Phonology, and Cognition.* Oxford: Oxford University Press, 96–130.
Docherty, Gerard J.
1992 *The timing of voicing in British English obstruents.* Berlin: Foris Publications. Netherlands Phonetic Archives IX.
Docherty, Gerard J. and Paul Foulkes
2000 Speaker, speech and knowledge of sounds. In: Noel Burton-Roberts, Philip Carr and Gerard J. Docherty (eds.), *Phonological Knowledge: Conceptual and Empirical Issues.* Oxford: Oxford University Press, 105–129.
Docherty, Gerard J., Paul Foulkes, James Milroy, Lesley Milroy and David Walshaw
1997 Descriptive adequacy in phonology: a variationist perspective. *Journal of Linguistics* 33: 275–310.
Fischer-Jorgensen, E. and B. Hutters
1981 Aspirated stop consonants before low vowels: a problem of delimitation – its causes and consequences. *Annual Report, Institute of Phonetics, University of Copenhagen* : 77–102.

Flege, James E.
1982 Laryngeal timing and phonation onset in utterance-initial English stops. *Journal of Phonetics* 10: 177–92.

Foulkes, Paul and Gerard J. Docherty (eds.)
1999 *Urban Voices: Accent Studies in the British Isles*. London: Arnold.

Foulkes, Paul and Gerard J. Docherty
in press The social life of phonetics and phonology. *Journal of Phonetics.*

Goldinger, Stephen D.
1997 Words and voices: Perception and production in an episodic lexicon. In: Keith Johnson and John W. Mullenix (eds.), *Talker Variability in Speech Processing.* London: Academic Press, 33–66.

Hawkins, Sarah
1979 Temporal coordination of consonants in the speech of children: further data. *Journal of Phonetics* 7: 235–267.

Hawkins, Sarah and Rachel Smith
2001 Polysp: A polysystemic, phonetically-rich approach to speech understanding. *Italian journal of Linguistics / Rivista di Linguistica* 13: 99–188.

Heselwood, Barry and Louise McChrystal
2000 Gender, accent features and voicing in Panjabi-English bilingual children. In: Nelson, Diane and Foulkes, Paul (Eds.) *Leeds Working Papers in Linguistics and Phonetics* 8: 45–70.

Johnson, Keith
1997 Speech perception without talker normalization. In: Keith Johnson and John W. Mullenix (eds.), *Talker Variability in Speech Processing.* London: Academic Press, 146–165.

Johnston, Paul A.
1980 A synchronic and historical view of border area bimoraic vowel systems. University of Edinburgh.
1997 Regional variation. In: Charles Jones (ed.), *The Edinburgh History of the Scots Language.* Edinburgh: Edinburgh University Press, 433–513.

Keating, Patricia A.
1984 Phonetic and phonological representation of stop consonant voicing. *Language* 60: 286–319.

Khattab, Ghada
2003 Sociolinguistic competence and the bilingual's adoption of phonetic variants: Auditory and instrumental data from English-Arabic bilinguals. University of Leeds.

Labov, William
1994 *Principles of Linguistic Change: Internal Factors.* Oxford: Blackwell.

Lavoie, Lisa M.
2001 *Consonant Strength: Phonological Patterns and Phonetic Manifestations.* New York and London: Garland Publishing Inc. Outstanding Dissertations in Linguistics.

Lisker, Leigh and Arthur A. Abramson
1964 A cross-language study of voicing in initial stops: acoustic measurements. *Word* 20: 384–422.
1967 Some effects of context on voice onset time in English stops. *Language and Speech* 10: 1–28.

Masuya, Yosihiro
1997 Voice Onset Time of the syllable-initial /p/, /t/ and /k/ followed by an accented vowel in Lowland Scottish English. In: Yosihiro Masuya (ed.), *Phonetics and Phonology.* Tokyo: Kobian-Shobo, 139–172.

Mullenix, John W.
1997 On the nature of perceptual adjustments to voice. In: Keith Johnson and John W. Mullenix (eds.), *Talker Variability in Speech Processing.* London: Academic Press, 67–84.

Ohala, John J.
1990 There is no interface between phonology and phonetics: a personal view. *Journal of Phonetics* 18: 153–171.

Pierrehumbert, Janet B.
2002 Exemplar dynamics: Word frequency, lenition and contrast. In: Joan Bybee and Paul Hopper (eds.), *Frequency and the Emergence of Linguistic Structure.* Amsterdam: John Benjamins, 137–157.

Pierrehumbert, Janet B., Mary E. Beckman and D. R. Ladd
2000 Conceptual foundations of phonology as a laboratory science. In: Noel Burton-Roberts, Philip Carr and Gerard J. Docherty (eds.), *Phonological Knowledge: Conceptual and Empirical Issues.* Oxford: Oxford University Press, 273–303.

Pierrehumbert, Janet B. and Deborah Talkin
1992 Lenition of /h/ and glottal stop. In: Gerard J. Docherty and D. R. Ladd (eds.), *Papers in Laboratory Phonology II.* Cambridge: Cambridge University Press, 90–117.

Pisoni, David B.
1997 Some thoughts on “normalisation” in speech perception. In: Keith Johnson and John W. Mullenix (eds.), *Talker Variability in Speech Processing.* London: Academic Press, 9–32.

Ryalls, Jack, Allison Zipprer and Penelope Baldauff
1997 A preliminary investigation of the effects of gender and race on Voice Onset Time. *Journal of Speech Language Hearing Research* 40: 642–645.

Scobbie, James M.
2005 Interspeaker variation among Shetland Islanders as the long term outcome of dialectally varied input: speech production evidence for fine-grained linguistic plasticity. *QMUC Speech Science Research Centre Working Papers WP-2*.

Scobbie, James M.
in press Interface and overlap in phonetics and phonology. In: Charles Reiss and Gillian Ramchand (eds.): *Handbook on Interface Research in Linguistics.* Oxford: Oxford University Press.

Scobbie, James M., Claire Timmins, Jane Stuart-Smith, Fiona Tweedie, Nigel Hewlett and Alice E. Turk
2000 Fieldwork in the urban jungle: an empirical phonological study of Glasgow English. *The VIIth Conference on Laboratory Phonology*. Nijmegen.

Smith, Bruce L.
1978 Temporal aspects of English speech production: a developmental perspective. *Journal of Phonetics* 6: 37–67.

Suomi, K.
1980 Voicing in English and Finnish stops. *Publications of the Department of Finnish and General Linguistics of the University of Turku* 10.

Thomas, Erik R.
2002 Instrumental phonetics. In: Jack K. Cambers, Peter Trudgill and Natalie Schilling-Estes (eds.), *The Handbook of Language Variation and Change.* Oxford: Blackwell Publishers Ltd, 168–200.

Vihman, Marilyn M. and Shelley L. Velleman
2000 Phonetics and the origin of phonology. In: Noel Burton-Roberts, Philip Carr and Gerard J. Docherty (eds.), *Phonological Knowledge: Conceptual and Empirical Issues.* Oxford: Oxford University Press, 305–339.

Westbury, John R.
1979 Aspects of the temporal control of voicing in consonant clusters in English. *Texas Linguistic Forum* 14: 1–304.

On the scope of phonological learning: Issues arising from socially-structured variation

Gerard Docherty, Paul Foulkes, Jenny Tillotson and Dominic Watt

This study set out to track the path taken by children in learning variable phonetic forms in the accent to which they are exposed, and as part of this to make a three-way comparison between the children's performance, that of their immediate caregiver in child-directed speech (CDS), and that of the broader speech community. We suggest that theories of acquisition (and by extension theories of representation in mature speakers) may gain by adopting a broader focus which acknowledges that phonological learning is not simply focused on a static or uniform system structured only in relation to lexical/prosodic contrast. We highlight the compatibility of this approach with episodic models of lexical representation in which token-to-token variation is encoded as part of the constellation of memory traces which make up the mental representation of lexical items.

1. Introduction

Research in phonological acquisition has yielded a wealth of information on how children, once presented with linguistic input, learn the structures of words and the phonological processes that those structures may undergo. A parallel tradition of research in sociolinguistics has revealed that language use is highly variable, but also that many facets of variability are structured in line with non-linguistic factors such as gender, age and speaking style. Our aim in this chapter is to draw these two research traditions together, in order to explore the implications for accounts of phonological acquisition of socially-structured variability. This approach leads us to advocate an account of acquisition in which 'phonological knowledge' embraces all of the systematic relationships between the sound patterns of spoken language and the external environment rather than a subset of these as is typically found in most accounts.

We begin with a general discussion of the task of phonological acquisition. We argue that aspects of variable performance must be learned alongside reflexes of the system of lexical contrast, and indeed may be indivisible from them. In section 3 we outline two projects that we have carried out to investigate aspects of structured variability in the variety of English spoken in Newcastle (in the north-east of England). In section 4 we describe some findings from these projects which bear on the issue of how variability may impact upon the acquisition process. Finally, section 5 positions the findings in a broader theoretical context.

2. Background: socially-structured variability

It would be fair to say that there remains a great deal of controversy, even fundamental incompatibility, within the various methodologies and theoretical approaches adopted in work on phonological acquisition. Nevertheless, we can identify a substantial consensus on what the key dimensions of the debate currently are, no matter what methodology or theoretical stance one adopts. We would like to take one of these as our point of departure, as follows: *to what extent do the primes/processes/constraints proposed within phonological theory adequately capture the phonological knowledge acquired by a child?*

That phonological theory has been an enormously productive source of hypotheses regarding the nature of phonological learning is beyond doubt, as evidenced by the way in which accounts of acquisition have been heavily influenced by the swings within theory from segmental accounts, to distinctive features, to phonological processes, to feature geometries, and most recently to ranked constraints. Nevertheless, it is precisely this question which is at the heart of many of the long-standing sources of controversy within the field of acquisition. For example, a key strand of work relates to whether specifically linguistic predispositions are activated in phonological bootstrapping (as hypothesised by UG) or whether those predispositions are largely or maybe entirely related to general properties of learning, audition, motor control and memory (e.g. Sabbagh and Gelman 2000, Vihman and Velleman 2000). Likewise this central question bears on the issue of abstractness of representation; on the one hand phonological theory tends to assume relatively abstract and invariant lexical representations implying a complex mapping to surface forms, while on the other hand certain theories of memory and representation applied to phonological learning and knowledge (e.g.

Pisoni 1997, Johnson 1997, Pierrehumbert 2001, Coleman 2003) argue for strongly surface-oriented representations with a relatively simple mapping to and from surface forms.

One of the principal consequences of the influence of phonological theory on acquisition is the fact that most models of acquisition presuppose that the object of phonological learning is fundamentally a *system for expressing lexical contrast*. This follows from the tradition established in almost all mainstream linguistic theories of the 19th and 20th centuries, where rigid distinctions are drawn between form and substance, langue and parole, competence and performance, phonology and phonetics. Hume and Johnson (2001: 12) provide a succinct description of their view of the system which seems to capture the general spirit:

> Phonological systems…are symbolic in nature, dissociated from any particular physical event in the world. Indeed, such is the independence of phonology from the physical world that it can be said that two people share the same symbolic phonological system, speak the same language, even though their experience of physical events in the world does not overlap at all.

We do not want to question this view of the system *per se*, nor do we want to suggest that the learning of such a system is *not* a consequence of acquisition. However, the issue of variability in the spoken medium forces us to be more circumspect when it comes to assessing the task of the child in acquiring phonology.

It is obvious that spoken language – the vital input to a child without which normal acquisition cannot fully succeed – is variable. The relationship between the linguistic system and the physical world, or between phonology and phonetics, is a highly complex one. Phonological units are realised in the physical world in a potentially infinite (if constrained) set of physical forms. Variability may furthermore have many (often simultaneous) sources; e.g. within the same speaker, systematic effects of coarticulation, rate, and timing effects correlated with prosodic constituency can all coincide within the same time slot. These effects can be further confounded by cross-speaker variability arising from differences in voice quality, pitch range, and vocal tract length.

The phonetic properties of words are also subject to socially-structured variation (henceforth SSV). They may vary, for instance, as a reflex of gender, age, socioeconomic background, geography, speaking style, attention, affect, conversational structure and addressee (Labov 1994, 2001; Local 2003). SSV forms the focus of our present work, since it has received little attention in work on phonological acquisition, and thus its effects on the acquisition process are not well understood. However, the facts of socially-structured variability raise various questions for our understanding of the acquisition of phonology. First, SSV clearly forms part of what a language learner needs to learn. Learning about speech sounds involves more than acquiring knowledge about lexical contrast and the associated motor skills that enable the speaker to reproduce and understand contrastive information within a phonetic medium. In addition to this, language learners must learn to control their speech production mechanisms in order to sound like other members of their geographically- and socially-defined speech community. That is, they learn a sociolect and accent. They must also learn to adapt their phonetic output to achieve particular goals in communication (cf. the 'H and H' (hyper-hypo) model, Lindblom 1990). Typically this involves, for example, learning to speak carefully in certain circumstances to provide clear information to a listener, speaking quickly in other circumstances to execute speedy exchange of information, and speaking in more or less standard forms according to addressee and speaking situation. Learners furthermore learn to interpret and understand the phonetic variation they hear in other speakers' talk.

The point of interest for acquisition is that the auditory and vocal mechanisms that are used to encode and decode lexically-contrastive information are precisely the same as those which serve to encode and decode SSV. If we take it as axiomatic that acquisition is at least in part the result of exposure to spoken language input, then we can see that a child learner is faced with a specific problem: (s)he must extract different types of knowledge from the same phonetic input but initially without any basis on which to interpret the different sources of the phonetic variability. The problem is illustrated in Figure 1.

The data in Figure 1 represent a tiny fraction of the task facing a child learning a variety of British English such as that of Newcastle, which we will say more about in due course. The ambient language input is shown to the left, while the ellipses to the right illustrate the knowledge that will be derived via correct analysis of the input. Type A data include minimal pairs such as *foot* and *put*.[1] The phonetic variation differentiating the pairs is likely to be consistent across speakers the child hears, and correct interpretation of this acoustic/auditory variability will lead to two semantically contrasting words being learned. These form part of what we have labeled knowledge

Base 1, which corresponds to what might traditionally be called 'phonological competence.'

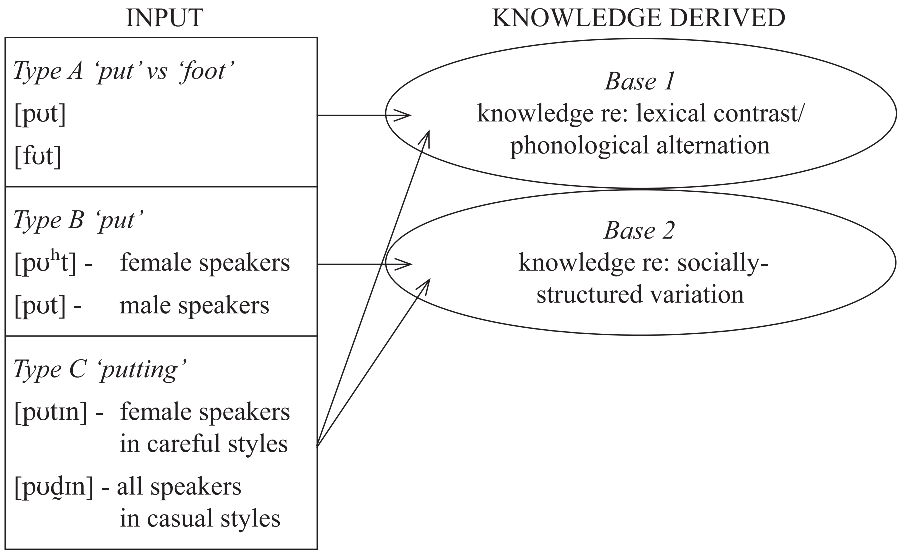


Figure 1. Schematic view of tasks in learning about sound structure.

Type B data illustrate a typical pattern of SSV for Newcastle English. Exemplars of words such as *put, foot* etc may vary phonetically according to the gender of the speaker. The alternation shown in Figure 1 between pre-aspirated [ʰt] and plain [t] must be learned, although it is not relevant for lexical contrast and therefore has a different status from knowledge derived from Type A data. A child must also learn to reproduce such alternatives appropriately to conform to the social conventions of the community he/she lives in. Put simply for the sake of exposition, girls should learn to produce more realisations of [ʰt] than boys to conform to their vernacular norms (see further below). We might call this sort of knowledge, again derived from analysis of phonetic variation, SSV knowledge (Base 2, which equates to 'sociolinguistic competence' in traditional terms).

The 'problem' for the child is shown via Type C data. The example of *putting* shows that individual exemplars in the input may encode both types of information simultaneously. The morphological derivation of *putting* demands phonetic alternation from the base form. Such phonologically-governed alternation is clearly a part of Base 1: in traditional terms it can be described as a morphophonemic rule. However, in Newcastle the choice of variants is very strongly influenced by gender and speech style. The variant [d̰] is the default form for all young adults in casual interaction, and is also used by most males in formal speech styles. An alternative variant, [t], may also be used, but this is mainly restricted to use by women in formal styles (Docherty et al. 1997, and see further below). To understand the socially-correlated use of such forms, and to learn to reproduce such forms appropriately, must form part of knowledge Base 2.

How does the child perform this dual task? Is it, for example, easier to learn forms that are relatively consistent phonetically (Type A), but more difficult to learn forms that are more variable and which serve to encode social and stylistic as well as lexical information (Type C)? More generally we might ask how the learning of lexically-contrastive information relates to, or differs from, the learning of other parallel channels of phonetic control which are necessary in learning to become a fully functional member of a speech community.

As we have already commented, such issues have received relatively little attention. There is some evidence from sociolinguistic work that children show signs of mastering local variable forms as early as 3;0 (Roberts and Labov 1995; Roberts 1997a, 1997b, 2001), but in general early child speech has received little attention from sociolinguists. Within the literature on developmental phonology, statements can be found to the effect that by age 2;0 children "speak with an "accent" typical of a native speaker" (Kuhl 1994: 812). It is not clear, however, on what evidence claims such as this are based. Clearly children of this age do not produce fully mature speech patterns; it is therefore an empirical question of some importance to ascertain precisely *which* components of 'accent' have been mastered by a child, particularly with regard to the SSV which pervades mature speech. It is also striking that few links have been made between work in the sociolinguistic and developmental arenas (exceptions being Payne 1980, Chambers 1992, Kerswill 1996).

In order to advance the discussion of these issues, the three questions we want to address in this paper are as follows:

(1) What is the nature of the SSV in linguistic input to children?
(2) How do children acquire the SSV that is characteristic of their community?
(3) What are the implications of the answers to (1) and (2) for our understanding of how phonological knowledge is represented in the child's, and ultimately the adult's, mind?

We have sought to address these issues through two projects. The first of these centered on adult subjects, while the second shifted the focus to children. In section 3 we outline the overall aims and methodologies of each project, before moving in section 4 to a discussion of some of our findings and their implications for the above questions.

3. Phonological variation in Newcastle English: aims and methods

3.1. Variability in adult speech: the PVC project

In the adult project (*Phonological Variation and Change in Contemporary Spoken British English,* henceforth referred to as PVC; Milroy et al. 1997), recordings were made of 32 speakers, equally divided by age (young group: 16–25, older group: 45–65) and gender. Half the sample at each age were drawn from a social network in a broadly defined 'middle class' neighborhood, the other half from a network in a 'working class' sector of the city.[2] The corpus comprises around 13 hours of unscripted conversations from the speakers in self-selected pairs, as well as word-list readings designed to elicit a more self-conscious style of speech. Analysis of several vocalic and consonantal variables was undertaken, using a combination of auditory and acoustic analysis.

The study revealed systematic patterns of variation correlating with gender, age, class and/or style (for extensive discussion see Docherty et al. 1997, Docherty and Foulkes 1999, Watt and Milroy 1999). The use of acoustic analysis in particular revealed aspects of fine-grained – but nevertheless socially-structured – variability in the phonetic forms used by the informants. Examples of our findings are discussed in Section 4.

3.2. Variability in child speech: the ESV project

The second project, focusing on child subjects, was entitled *The Emergence of Structured Variation in the Speech of Tyneside Infants* (henceforth ESV; Docherty et al. 2002). Two parallel studies were undertaken, one cross-sectional and one longitudinal. In total 96 recordings were produced from 53 children drawn from the same broadly 'working-class' community investigated in PVC. We present only a sample of the findings in this paper, drawing mainly from the cross-sectional study. The cross-sectional study was based on 40 children, with four boys and four girls recorded in each of five age groups: 2;0, 2;6, 3;0, 3;6 and 4;0 (± 1 month). This age-range was chosen on the basis that the level of phonological development typically achieved at 2;0 would probably be the earliest feasible starting point for addressing the research questions. Beyond 4;0 it was reasoned that it would be difficult to control for other sources of input, e.g. from the peer group and younger siblings. All children were born to monolingual parents, had normal speech and hearing, and were first-born in order to minimise the impact of communication with siblings. The parents were also the main care-givers.

The principal material collected was a sample of interaction between mother and child, mediated by the fieldworker. Typical interactional situations were playing with a toy or reading a book. Materials were chosen to elicit words containing variables of interest. We were not prescriptive about the amount of material to be obtained from the children. The goal of each session was to obtain as much data as possible, with the focus of the interaction geared by the fieldworker towards eliciting tokens of relevant variables. No attempt was made to manipulate children's speech style.

Recordings were made in as quiet surroundings as could be achieved at the subjects' homes, using radio lapel microphones and digital recording facilities. Fieldwork was performed by a group of speech and language therapists, all with extensive experience in obtaining language samples from infants and their parents. Analysis of the resultant data made use of a combination of acoustic and auditory methods. Auditory analysis was used to record transcriptions using IPA symbols. This was supplemented by acoustic analysis, both to register measurements of key parameters (e.g. voice onset time) and also to compile a detailed profile of the acoustic properties found in each token. In the case of (t), for example, the profiling recorded the presence or absence of periodicity during the stop, release burst, pre-aspiration and creaky phonation (see further Docherty and Foulkes 1999). All tokens were analysed where possible. For (t) this amounted to over 3,000 tokens from mothers and over 7,500 from children.

4. Results: variability and acquisition

In this section we turn to the findings of the two Newcastle projects, and discuss their implications for our understanding of phonological learning.

Question (1) above asks: what is the nature of the SSV in linguistic input children receive? Our first task in providing an answer to this question is to assess the parameters of phonological variability in the language which forms the input for the acquisition process. This enables us to identify the phonetic targets the child must aim for as well as any social constraints upon those targets. On the one hand this is aided by our detailed description of phonological variation in the Newcastle adult community derived from the PVC project. However, in assessing the task of a child in acquisition it is imperative to take full account of the input the child actually receives. It is well known that speech directed to children may differ markedly from speech between adults (Snow 1995).

We therefore present data relevant to question (1) in two parts. In section 4.1 we discuss aspects of variability in the adult community at large, while in section 4.2 we present data from an analysis of child-directed speech (henceforth CDS). In 4.3 we then assess the children's performance in acquiring variable phonological targets.

4.1. Variability in the ambient language

Although several variables were analysed we restrict ourselves here to a discussion of (t), as this is the variable which yields the most extensive sociolinguistic and phonetic variability. Table 1 summarises the distribution of five variants of (t) in word-final pre-vowel context (e.g. *what̠ if*), with data drawn from the working class speakers in the PVC study.

variant	ɹ	t̬	t	d̰	ʔ	*n*
older females	40	18	27	12	2	*404*
older males	15	35	7	42	2	*178*
young females	21	39	5	20	13	*402*
young males	3	59	4	23	12	*230*

Table 1. Percentage realisations of (t) in word-final pre-vocalic context, PVC working class adults (Docherty et al. 1997)

Table 1 testifies to a striking degree of socially-correlated heterogeneity in the data. [ɹ] is very strongly a marker of female speech, while [t] is only found to a significant degree for older women.[3] Males display a higher use of [t̬] and [d̰] than the corresponding females. The glottal stop is strongly associated with younger speakers, and as such reflects a change in progress affecting the dialect. Note also that the five variants collectively represent a fairly wide degree of variation in phonetic form. Variation is found with respect to voicing, plosivity, anteriority, and use of laryngeal constriction.

Variants of (t) found in other word positions differ somewhat from those used in word-final pre-vocalic contexts. In word-medial intersonorant position (for example *pu<u>t</u>ting, bo<u>t</u>tle*) all five variants shown in Table 1 can be used, but their distribution is very different. In casual style all speakers except the older women were extremely consistent in using [d̠]. For example, the young working class women produced this form between 80% and 100% of the time (cf. the 20% in Table 1). Use of [t] averaged 10%, with the other variants barely in evidence at all. A gender-correlated stylistic difference was also found for medial (t). In word-list readings, which were designed to yield a more formal, self-conscious speech style, most men continued to use the local variant [d̠] on the whole. Most women, however, displayed a stark style-shifting pattern and preferred near-categorical use of the more standard-like [t]. In pre-pausal positions (e.g. *who's tha<u>t</u>?*) only two of the five variants shown in Table 1 ([t] and [ʔ]) are found. The latter is rare and largely restricted to younger speakers in turn-endings (Docherty et al. 1997). Acoustic analysis revealed that pre-aspirated stops were very common in pre-pausal data, but were barely found at all in the pre-vocalic context shown in Table 1 (see further Docherty and Foulkes 1999). Pre-aspiration was strongly correlated with age and gender, as shown in Figure 2. Note in particular that for young women the pre-aspirated form is the dominant variant, and also that the dramatic age differences suggest this is an innovative form in the dialect.

From the perspective of a child learning the Newcastle dialect, the data outlined in this section probably present a rather daunting picture. First, of the five variants available in word-final pre-vowel position, only two ([t] and [ʔ]) are also permissible as realisations of word-final (t) in pre-pausal position. The other three variants therefore reflect dialect-specific morphophonological processes, and must be learned as part of 'Base 1' phonological knowledge. All five variants are furthermore subject to socially-determined variability according to a speaker's age, class and/or gender. A child growing up in our 'working class' neighborhood is therefore likely to hear all five variants being used by adults, but to different degrees according to the speaker. The child must learn to interpret all five forms as linguistically equivalent, and must also learn to use the variants appropriately in his/her own speech in order to conform to the sociolinguistic norms of the community. Such information constitutes part of 'Base 2' knowledge in our schema shown in Figure 1. Similarly, an interlocking complex of information must be learned with respect to articulating and understanding medial and pre-pausal (t), some of which relates to phonological context and some of which is socially-governed.

The extent of variability apparent with respect to (t) presents by no means the only difficult task for the child learner. Similar variation must be handled for a child to master the auditory and articulatory information characterising many other phonological units, including (p), (k), (r), (th), and several vowels (Watt and Milroy 1999).

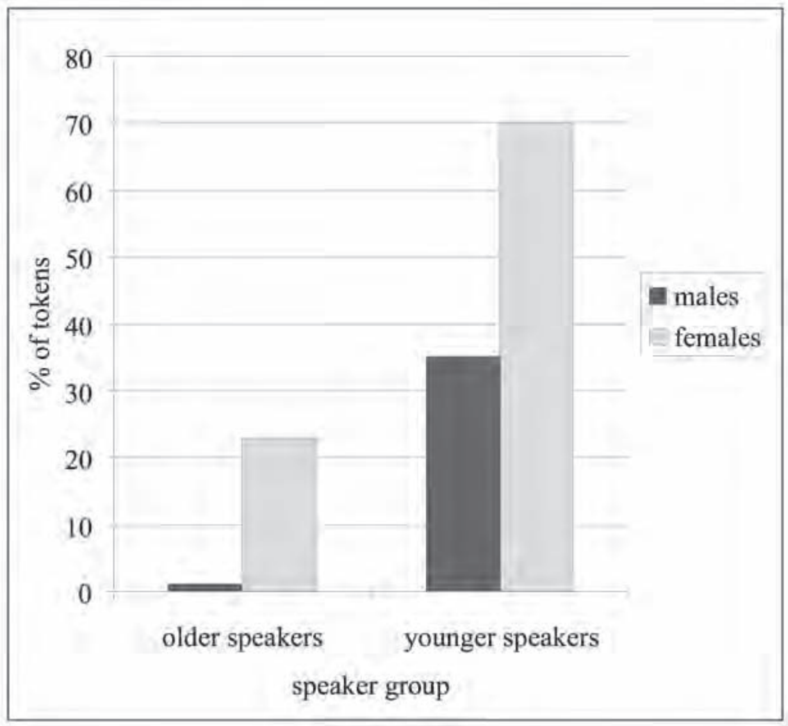


Figure 2. PVC study: distribution of pre-aspirated variants of pre-pausal (t) in Newcastle English (mean N = 90 tokens per group, data drawn from word-list items, social classes combined).

The dialect we have been investigating is not unusual in this regard. All the evidence available suggests that SSV is ubiquitous within speech communities. This renders all the more striking the fact that we do not have an account, or even very many detailed descriptions, of how this aspect of sound patterning emerges in phonological acquisition.

4.2. Assessing variability in input to children

We turn now to our analysis of phonological variation in CDS. In part this investigation sought to examine the extent to which the children in our study did indeed experience the range of SSV apparent in adult-to-adult speech, as described in section 4.1. It was also undertaken with another aim in mind, namely to assess whether variation in CDS differs from that found in inter-adult speech. It is well known, for example, that features of intonation, syntax, speech rate and vocabulary are modified in speech addressed to children (Snow 1995). However, relatively little attention has been paid to the segmental properties of CDS, and the work that has been done paints a rather inconsistent picture. Some research suggests that CDS is characterised by more careful articulation than inter-adult speech (e.g. Bernstein Ratner 1984), leading to the conclusion that, at least with respect to its formal properties, CDS is "a simpler, cleaner corpus from which to learn language" (Snow 1995: 180). Other work, however, has found CDS to be no less variable that inter-adult speech (e.g. Shockey and Bond 1980). Foulkes, Docherty, and Watt (2005) provide a detailed overview of work on the phonological properties of CDS.

As part of the ESV project we therefore also analysed the segmental properties of speech addressed to children. Here we highlight some of the findings for word-medial (t) (Foulkes et al. 2005).

In section 4.1 we noted that in inter-adult speech there was extremely consistent use of [d̰] for all speakers in casual style, averaging 90% for the young working class women. The form was also used by most men in the more formal word-list style, but women in word-list readings tended to prefer [t]. Analysis of the mothers' speech to the children in the ESV corpus revealed sharp differences compared with the findings of the adult-to-adult speech of PVC (see Figure 3). In speech to children the proportion of [t] increased to an average of 59%, with [d̰] averaging 36%. The mothers therefore used far fewer of the characteristic local variants overall, preferring forms that were more like those used in the standard accent and in formal speech styles. There was incidental evidence from our data that men engaging in CDS did not behave in the same way as women. A small amount of data from three men who were fortuitously present for part of a number of recording sessions showed that [d̰] was favoured. This aspect remains a topic for future research, although the fact men appeared to make less marked linguistic adaptations in CDS has been noted for other linguistic phenomena (Snow 1995).

The patterns apparent in Figure 3 raise two interesting issues with regard to variability in input heard by children. First, if we take account of the overall distribution of variants in proportional terms, and concentrate solely on (t) in this context, CDS is actually more variable than adult-to-adult speech. The high degree of consistency shown in adult-to-adult speech (90% usage of [d̰]) is not apparent in CDS. Instead, CDS contains two alternative variants in relatively high proportions.

Secondly, there may also be increased variability in CDS emanating from social factors, as shown by the differential phonetic behaviour of men and women. Recall that in adult-to-adult conversation a consistent use of [d̰] was found for all speakers in unscripted speech.

The data in Figure 3 are pooled from 39 mothers. A more detailed analysis of the performance of the individual mothers revealed further aspects of socially-influenced variability. Differences in variant use across the sample of mothers was related to the gender of their child. Mothers of boys used significantly more [d̰] than did mothers of girls. The latter produced a much higher proportion of [t] than mothers of boys (Table 2).

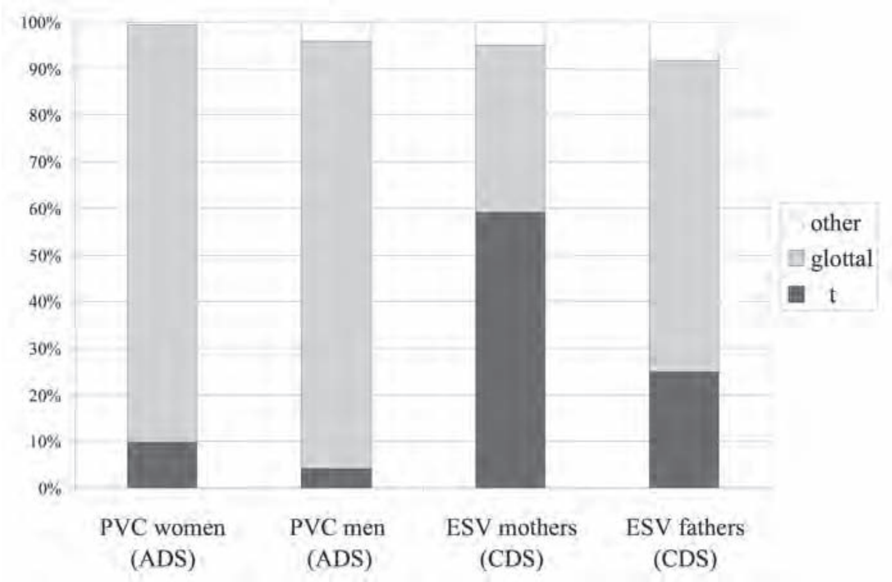


Figure 3. % variant use for word-medial intersonorant (t) (N tokens: PVC women 163; PVC men 120; ESV mothers 570; ESV fathers 36) (Foulkes et al. 2005).

Table 2. Percentage occurrence of variants in word-medial position, CDS (Foulkes et al. 2005)

	t	d̰	*n*
mothers of boys	48.4	44.8	293
mothers of girls	69.3	27.7	277

4.3. Assessing variability in children's performance

Having outlined some aspects of the variability which children face in the input they hear, we turn now to a discussion of the children's performance. As expected, the performance of children in the study was itself highly variable. However, we do not focus here on individual variability but on the overall differences yielded by our analysis. We discuss results relating to three aspects of (t) production. In each case the central issue is to investigate to what extent there is evidence that a particular aspect of SSV is emerging in children's performance.

4.3.1. Use of glottal variants across contexts

We drew attention in sections 4.1 and 4.2 to significant differences across contexts in the use of variants containing laryngealisation, [d̰] and [ʔ]; in word-medial position (*pu<u>tt</u>ing*) [d̰] is almost categorical in adult-to-adult speech (but significantly less frequent in speech to children). In word-final pre-vowel context (*pu<u>t</u> in*) both laryngealised forms are possible, but they are variable across speaker groups and overall occur significantly less frequently than in word-medial contexts. In pre-pausal positions (*pu<u>t</u>* #) laryngealised tokens are rare.

In light of these findings from adults, our interest in the children's productions is twofold. First, do the children show signs of differentiating their use of glottal forms across the various contexts? Secondly, what is the children's reaction to the stark differences in word-medial contexts, where patterns in CDS are different from inter-adult speech?

The findings from the children's data are shown in Figure 4, with both types of adult data shown for comparison. Children's data were classified on the basis of the acoustic profiling analysis (Docherty and Foulkes 1999).

With respect to word-medial tokens, Figure 4 provides evidence that the children have adopted production patterns closely matching those of mothers in CDS. The children's overall use of tokens with laryngealisation amounted to 32% compared with the mothers' 36%, and is thus distinct from the pattern typical of adult-to-adult speech where [d̰] is so dominant.

The relationship between contexts is less clear, but again offers some suggestions that the children are differentiating variant use appropriately. In both modes of adult speech there are more glottal tokens in word-medial position than the other two contexts. The children do not display the same differentiation between word-medial and word-final pre-vowel positions, using slightly more glottals in the latter context, but with this difference failing to reach significance. However, (t) in word-final positions *is* differentiated, depending on whether the word is followed by a vowel or a pause. The children produce significantly fewer pre-pausal glottals than they do in pre-vowel contexts, thus paralleling the patterns for adults.

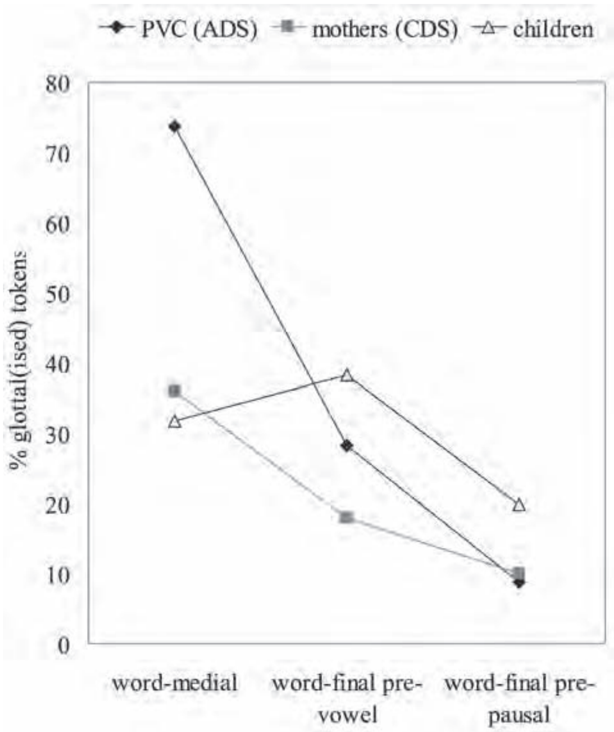


Figure 4. Use of glottal tokens across contexts (PVC – all WC speakers, ESV – mothers' CDS, ESV children).

4.3.2. Pre-aspiration

As we showed in section 4.1, pre-aspirated variants of pre-pausal (t) are very strongly associated with young women in Newcastle. Pre-aspiration was also found for many of the mothers recorded for the ESV project, although the larger body of speakers (39 compared with 8 in PVC) not surprisingly revealed a degree of variability across individuals with usage ranging from 0 to 89%. Nevertheless, the adult data from both projects show that [^{h}t] is a form which children acquiring the dialect typically must learn to understand in the input they hear. Children might also potentially learn to reproduce it. In particular, we would predict that at some point in development girls should favour the form more than boys if they are to conform to the apparent gender-patterning extant in the community.

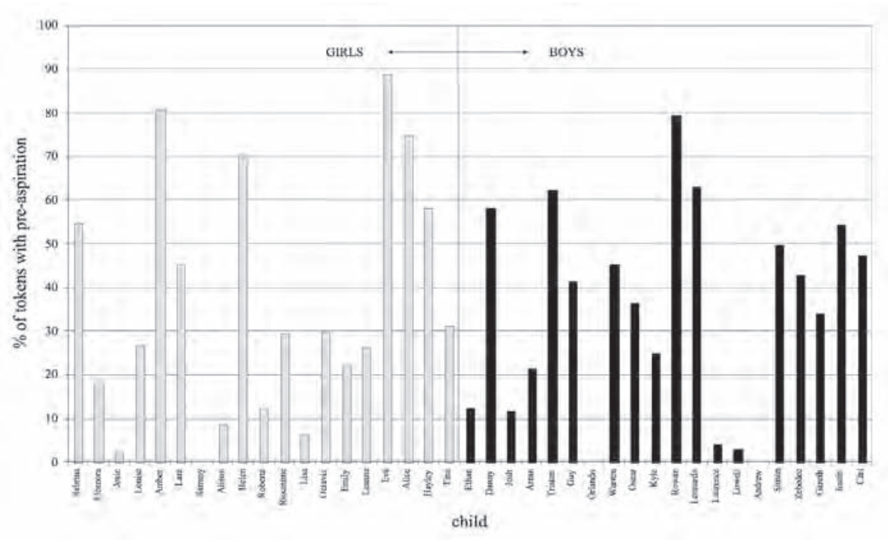


Figure 5. Percent occurrence of preaspirated tokens in word-final pre-pausal context (39 children ordered by age; girls to the left, boys to the right).

Figure 5 displays data for children in the cross-sectional corpus, arranged from left to right in order of increasing age.[4] The bars of the histogram show the proportion of pre-pausal tokens containing pre-aspiration produced by each child. There is clearly a wide range of variability in the children's performance, coincidentally matching the mothers' range of 0 to 89%. However, it is also apparent that pre-aspiration is well established across the sample as a whole, showing that the majority of the children have acquired the capacity to reproduce the pattern they experience in their ambient input.

Figure 6 illustrates the relationship between pre-aspiration use by mothers and their children.

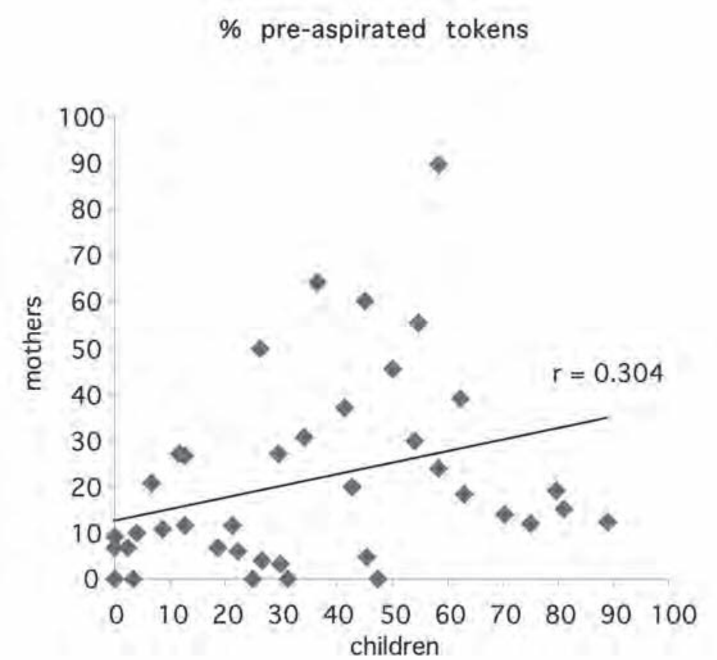


Figure 6. Scatter plot showing % preaspiration used by 39 mother/child pairs.

The children's proportional use of pre-aspiration is scaled to the *x* axis, with the mothers' use scaled to the *y* axis. Each data point in the scatter plot therefore indicates the proportion of pre-aspirated tokens used by each mother-child pairing. The variability for both children and adults already noted is again apparent. What is striking, however, is that for the group as a whole there is evidence of a relationship between the performance of mother and child. This is not an overwhelming effect for the pooled data ($r = 0.304$, $p < 0.06$), although it does strengthen considerably if pairings with low Ns are removed from the analysis (for example, restricting the analysis to the 30 mother-child pairs where $N > 10$ for both mother and child, the significance of the correlation improves to $p < 0.04$). Hence, despite all the other factors mediating between ambient language and a child's speech performance, our results suggest that mothers who use relatively high rates of pre-aspiration tend to have children who produce relatively high rates of pre-aspiration. It is important to acknowledge that children receive input from many sources, not just their mothers, of course. However, the ESV children were selected according to the criterion that the parents were the dominant caregivers, which in most cases meant that the bulk of this role was undertaken by the mother. As such, the mothers' recorded data presumably represent a characteristic sample of the type of phonetic input to which the children have extensive exposure. We might therefore interpret this finding as indicative of learning of phonetic detail on the part of the children in line with patterns in the input they receive.

Figure 7 shows more clearly the effects of age and gender on the use of pre-aspiration, suggesting that there may be incipient gender differentiation among the older children in the corpus.

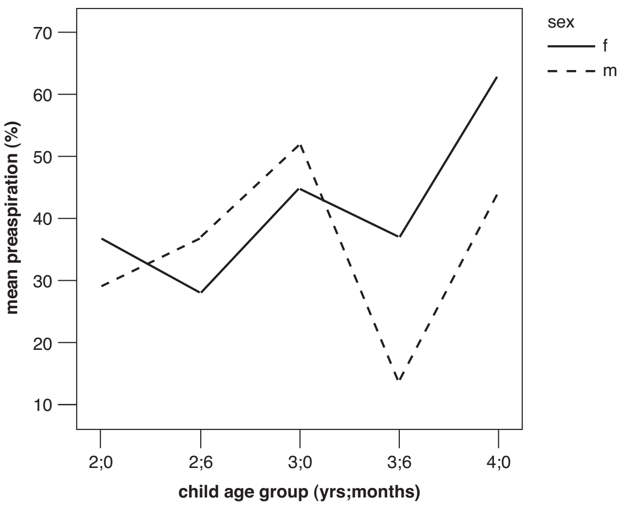


Figure 7. Percent occurrence of pre-aspiration in pre-pausal context by age and sex of child.

In the first three age groups no differences were found in girls' versus boys' use of pre-aspiration. At age 3;6, however, the girls use significantly more pre-aspiration than the boys. A similar difference is found for the oldest children, narrowly missing significance at the 5% level. We appear to have evidence, then, that appropriate sociolinguistic marking may be emerging with respect to this variant, following the model we have seen in the adult community (cf. similar findings by Roberts (1997a, 1997b) from studies of children at similar ages). Note that this is the only case of significant gender-differentiation yielded by our children's performance data (a point we return to in Section 5).

5. Discussion

In light of our data from children and adults, we now return to the questions we posed at the end of section 2.

5.1. Question (1)

Question (1) asked "what is the nature of the SSV in the linguistic input children receive?" It can be seen from the PVC data that the distribution of (t) variants in this variety is sensitive to both phonological and lexical context, as well as being dependent on several non-linguistic factors. Each variant encodes information about linguistic structure, the social attributes of the speaker, or both simultaneously. Because acquisition of such complex patterns might be expected to pose a serious challenge to children, one might also anticipate (especially in light of studies such as Bernstein Ratner 1984) that adults would modify their speech when addressing children in the direction of a simpler, less subtly conditioned form, in which the sets of phonetic realisations are kept as small as possible. By so doing they would assist children in their acquisition of the phonetic contrasts they need to master if they are to learn the phonological system of the ambient language. Perhaps, then, the characteristics of CDS with respect to (t) among the ESV mothers would more closely resemble the less variable patterns observed in word list readings by the women in the PVC study (e.g. in the greater use of [t] in word-medial position).

The findings presented in section 4.2 to some extent bear out this assumption, but in other ways cast doubt on it. The mothers do appear to modify their linguistic behaviour when addressing their children, and in the case of word-medial (t) there is indeed a significant increase in [t] usage. However, in comparison with the sample of (t) in inter-adult speech from the PVC women (see Figure 4) there is actually *less* consistency in CDS as far as the choice of (t) variant in this context is concerned. It is therefore far from clear that the mothers are 'simplifying' the set of alternants they use. The phonetic variability in CDS is different in key respects from that found in ADS, but it is not noticeably simpler (cf. Shockey and Bond 1980).

5.2. Question (2)

This last finding leads us on to question (2), "how do children acquire the SSV that is characteristic of their community?" The first response to make is that children clearly are acquiring some of the fine-grained patterns of realisation which are strongly socially marked in their community. There is evidence in our data that both boys and girls are latching on to patterns which are known to be predominantly female markers within the adult community (e.g. use of pre-aspiration). This is perhaps not surprising given the central role of the mothers in caring for their children, and is in accord with the claims of Labov (1990) that variable features dominant in women's speech have the best chance of being transmitted to the next generation. The fact that we only find a single case of gender differentiation within the performance of the child subjects reflects the likelihood that sociolinguistic identity emerges at a later stage of development, probably heavily influenced by peer group interactions (Kerswill 1996).

The fact that our subjects are beginning by age 3;0–3;6 to reproduce fine-grained aspects of the socially-structured variation which they are exposed to suggests that they are attuned to these features from an early stage of development. It does not appear to be the case that the learning of a system of phonological contrasts is independent of the process of learning the way in which those contrasts are typically realised within the immediate community. Likewise, if we accept the view that completion of a mature phonological system is not attained until well beyond the age-range of our child subjects (Vihman 1996, pp. 228, 237), it seems that acquisition of detailed phonetic patterns is not dependent on the child having previously 'worked out' the properties of the phonological system. In this respect, our findings support models of acquisition which emphasise the role of "phonetic grounding" as an "entry point into linguistic structure" (Vihman and Velleman 2000:305, see also MacNeilage 1997, Jusczyk 1993).

In relation to (2) we might also focus on the role of CDS. If we consider the sorts of knowledge that children must acquire in their phonological development (as discussed in section 2), it is vital that children become aware of the sorts of phonetic variability that are present in the community, and that aspects of this variability are associated with particular types of speaker (men vs. women, old vs. young, etc.). Thus, CDS should ideally contain features that provide the child with information that will allow him/her to interpret utterances made by any speaker, but which will also provide indications of what sorts of linguistic behaviour are considered appropriate for the individual child. If mothers of girls, for example, produce higher levels of [t] in word-medial contexts than mothers of boys do, we can hypothesise that mothers of children of both sexes are demonstrating that more than one pronunciation of medial (t) is possible, while at the same time attempting to bias the child towards using one variant over another, as is deemed appropriate to his or her gender. This would be in accord with abundant evidence showing differential parental behaviour to male versus female children (e.g. Wells 1986), but we are unaware of any other findings demonstrating that such differentiation is present in the fine phonetic detail of CDS.

5.3. Question (3)

What are the implications of the answers to (1) and (2) for question (3), relating to how phonological knowledge is represented in the child's, and ultimately the adult's, mind? In Section 2 we contrasted the predominant view within phonological theory by which SSV is, at best, present on the margins of theory, with a stance suggesting that the intertwined strands of lexical and other types of meaning encoded within sound patterning may in reality be more difficult to pull apart. The results of our study provide some support for this latter position. We find evidence of SSV emerging in tandem with other aspects of phonological learning. This is not to say that children have acquired a sociolinguistic identity at age 2–4 or that they are systematically using speech production for social marking purposes at that age. But our results do suggest that, in approaching the task of acquisition, children may not be predisposed to analyse *some* of the patterning they are exposed to as relevant to a system of lexical contrast while classifying other aspects as being irrelevant to this. Since it is unlikely that they have *a priori* awareness of this distinction, it might not be unreasonable to hypothesise that children are sensitive to sound-meaning associations of all sorts within the ambient sound patterns without excessive privilege being assigned to lexical meaning. For example, a child may learn to associate the sequences [wɔtə] and [wɔd̰ə] with a drink. But the same process may also lead to an association of this word produced with [t] to 'mother' and with a glottal form to 'father', thus simultaneously encoding another strand of meaning. As investigators of language we tend to treat the former as more central to the task of learning; the empirical question is whether children do the same.

The suggestion that lexical and non-lexical sound-meaning associations may not be (immediately) channelled into separate strands may not be easy to accommodate within the architecture of most approaches to theoretical phonology. However, it is compatible with the view of representation espoused by exemplar models of representation (see Pisoni 1997, Johnson 1997 for overviews). One of the features of exemplar models is that they obviate the 'non-invariance' problem by encoding variability within representation. That is, exemplars with varying phonetic shapes may be encoded in respect of many or all of the dimensions with which they are associated (e.g. lexical meaning, speaker identity, situational context, etc.). Unlike conventional models of representation, it is not necessary to posit algorithms for separating out these overlaid strands nor to strip out the 'noise' in the stimulus which masks the underlying invariance. Thus exemplar models may offer a plausible means of accounting for the learning and emergence of features of SSV alongside other systematic aspects of sound patterning. It is fair to say, though, that proponents of exemplar models have not yet addressed in any depth the areas of acquisition or speech production (but see Pierrehumbert 2001). Nevertheless, the fact that exemplar models offer an 'in-principle' account of the sort of phonological learning we have explored in this chapter, together with the fact that they appear to be compatible with findings on phonetic perception in infants (Jusczyk 1993, Jusczyk and Luce 2002), does offer a channel for future investigation. In pursuing this, one specific question is whether, as acquisition/learning becomes more sophisticated, children eventually separate out the different strands of meaning as many models would predict, or whether the diverse associations between sound and meaning remain embedded within one another even after segmental awareness has emerged. This brings us back to the quote from Hume and Johnson (2001) in Section 2 regarding the traditional view of phonology as a symbolic system for lexical contrast. We cannot rule out that increasing segmental awareness (enhanced in alphabetic cultures by coaching in literacy) leads an individual to 'work out' that there is a system to the sound patterning of language which is akin to that conceived of by most phonologists. But it is equally possible that knowledge of this sort could be present alongside the other sound-meaning associations which are present within the listener's experience. Indeed, the evidence which has been adduced in support of exemplar representations (largely based on effects of multiple speakers/different voices on memory-related tasks) suggests that if a system of lexical phonological contrast is indeed acquired eventually by an individual, this does not entail that individual jettisoning other key dimensions of their experience as listeners.

6. Conclusions

Our study has attempted to bridge the wide gap between, on the one hand, work on phonological acquisition and the phonetics-phonology interface, and on the other, the insights of sociolinguistics. Using a combination of acoustic and auditory techniques and appealing to the results of two studies we have found evidence of socially-structured patterns of speech production emerging alongside other systematic phonetic behaviour, all at a stage generally considered to be prior to the full acquisition of a complete 'phonological system'. Our findings are compatible with an account of acquisition in which children attempt to tune into systematic variability in the ambient language irrespective of whether it is lexically-derived or a reflex of one of the many other sound-meaning associations inherent within the ambient linguistic input. This in turn is compatible with models of representation and learning in which some or all of the relevant information can be retained in memory. It is less compatible with models which require speakers to direct attention away from many aspects of the sensory stimulus in the search for underlying lexical invariants.

Of course, a degree of caution needs to be applied to our work to date. While we have investigated a number of variables, the largest dataset only relates to a sample of one small part of one dialect. However, the outcomes of our analysis persuade us that this is an area which merits further investigation and where combining insights from lines of research that only rarely come into contact may enhance our understanding. It is also important to exercise caution in respect of the sociolinguistic assumptions built in to work of this sort. In particular, the issue of what the relevant social dimensions of variation might be is far from resolved (Mufwene 1994, Eckert 1999), and there is an increasing awareness that assumptions typically made about homogeneity of variance within the socially-defined groups of speakers may be underplaying the significance of the individual as a key factor in linguistic variation (Mufwene 1994, Johnstone and Bean 1997, Scobbie this volume). A further point is that while we have focused our discussion on the juxtaposition of lexically- and socially-structured sound patterning, we need to extend our account to take into consideration the role of factors such as discourse type and prosodic structure, to give but two examples of other features which are likely to correlate with the use of (t) variants described above.

Socially-structured phonetic variability of the kind we have discussed here is presumably found in all speech communities. In light of this, the emphasis on stable and invariant forms within the acquisition literature is perhaps all the more striking. Edwards (2000: 245) makes a similar point, calling for a more sophisticated assessment of targets in phonological acquisition studies. Edwards points out:

> we cannot consider acquisition of a phoneme category independently of the segmental and prosodic contexts in which that phoneme is produced. We can't simply ask "Can the child produce /k/?" Rather, we must ask whether a child can produce /k/ word-initially and word-finally, in clusters, before vowels, after back vowels, in stressed onsets, in unstressed codas, and so on.

To this list we would also add the rider that children must learn to understand and reproduce sociolinguistically appropriate variants. The availability of data from a detailed study of the adult community, such as that afforded by PVC, is an obvious source of help in assessing the targets children must aim for in phonological acquisition, as well as identifying the parameters of variability in those targets.

Failure to take account of such factors may lead to the view that variability in children's performance is the result of a high degree of inconsistency. Variability itself is widely recognised as one of the most obvious characteristics of children's speech. However, some investigations of children's speech have tended to try to reduce this variability, concentrating on characterising the patterns of *simplification* between child and adult forms (see further MacNeilage 1980, 1997). The abundant variability found in child speech has often been attributed to imperfect learning, physical production constraints, random factors, and/or the operation of universal phonological rules. For instance, glottalisation and deletion of /t/ in coda position have been analysed as the result of children operating a phonological rule of simplification (Locke 1983: 230 ff.). While such factors may indeed play a part in explaining variability, our data offer evidence that variability in child speech is (at least to some extent) linked to variability in adult speech. That is, children's variability is in part structured, and moreover it is structured in line with the input they receive. Our findings therefore offer support to other studies which show evidence for detailed learning, based on specific aspects of input, of features in other areas of the grammar (e.g. Snow 1995). The child's task in acquiring phonology cannot therefore be adequately described as the learning of a language-specific set of contrasts. The task is in fact much more complicated. Assessing that task demands an understanding of the parameters and constraints on variability in the ambient dialect.

Returning to the over-arching question set out in Section 2, our findings highlight the social embedding of phonological learning; acquisition is partly about learning the sound patterns for lexical contrast, but it is also about learning to be a member of one's family and broader community and to be able to communicate effectively within those groups (i.e., learning the social meaning embedded within the ambient sound patterns). Our understanding of this latter part of the task remains scanty to say the least, and is not facilitated by theoretical frameworks which as a matter of principle rule this out of scope.

Notes

* This investigation has been supported by the UK Economic & Social Research Council – awards R000237417/R000234892. Our thanks to Barbara Dodd, Ghada Khattab, John Local, Lesley Milroy, Jim Scobbie, and Jane Stuart-Smith for helpful discussions, and to Alice Faber and Bill Labov for their comments on the chapter. We are also grateful to the fieldworkers from the Newcastle Paediatrics Speech & Language Therapy service, and to the participating families who gave freely of their time.

1. Segmental transcription is used here for the sake of illustration. It is not intended to signal an *a priori* assumption that either adults or children necessarily operate with a purely segmental phonology.
2. The terms 'middle class' and 'working class' are used here as convenient labels to describe two communities within the city that are clearly differentiated along a number of socio-economic parameters, including housing, employment rates, and type of occupation.
3. In PVC, variation was analysed using log-linear models (for further details, see Docherty et al. 1997). Within *ESV* a range of statistical procedures were used, including linear models and the identification of disjoint 95% confidence intervals. All analyses were designed to provide a conservative means of identifying differences and relationships between groups of speakers in the context of varying sample sizes from individual speakers and the between-speaker variability characteristic of child-speech. Further details are given in Foulkes et al. (2005). Differences referred to as significant were found to be so at an alpha level of at least $p<0.05$. We are grateful to Dr Tom Chadwick of the University of New-

castle School of Population and Health Sciences for his support with this aspect of our study.

4. Data are shown for 39 children. One of the recordings of the mothers in the cross-sectional study proved unsuitable for analysis.

References

Bernstein Ratner, Nan
1984 Patterns of vowel modification in mother-child speech. *Journal of Child Language* 11: 557–578.

Chambers, Jack
1992 Dialect acquisition. *Language* 68: 673–705.

Coleman, John
2003 Discovering the acoustic correlates of phonological contrasts. *Journal of Phonetics* 31, 351–371.

Docherty, Gerard and Paul Foulkes
1999 Derby and Newcastle: instrumental phonetics and variationist studies. In: Paul Foulkes and Gerard Docherty (eds.), *Urban Voices,* 47–71. London: Arnold.

Docherty, Gerard, Paul Foulkes, Barbara Dodd and Lesley Milroy
2002 The emergence of structured variation in the speech of Tyneside infants. Final report to ESRC for award #R000237417.

Docherty, Gerard, Paul Foulkes, James Milroy, Lesley Milroy and David Walshaw
1997 Descriptive adequacy in phonology: a variationist perspective. *Journal of Linguistics* 33: 275–310.

Eckert, Penelope
1999 *Linguistic Variation as Social Practice.* Oxford: Blackwell.

Edwards, Jan
2000 Commentary. Lexical representations in acquisition. In: Michael Broe and Janet Pierrehumbert (eds.), *Acquisition and the Lexicon. Papers in Laboratory Phonology 5,* 240–249. Cambridge: Cambridge University Press.

Foulkes, Paul, Gerard Docherty, and Dominic Watt
2005 Phonological variation in child-directed speech. *Language* 81: 177–206

Hume, Elizabeth and Keith Johnson (eds.)
2001 *The Role of Speech Perception in Phonology.* New York: Academic Press.

Johnson, Keith
1997 Speech perception without speaker normalization. In: Keith Johnson and John Mullennix (eds.), *Talker variability in speech processing,* 145–165. San Diego: Academic Press.

Johnstone, Barbara and Judith Bean
1997 Self-expression and linguistic variation. *Language in Society* 26: 221–246.

Jusczyk, Peter
1993 From general to language-specific capacities: the WRAPSA model of how speech perception develops. *Journal of Phonetics* 21: 3–28.

Jusczyk, Peter and Paul Luce
2002 Speech perception and spoken word recognition: past and present. *Ear and Hearing*, 23: 2–40.

Kerswill, Paul
1996 Children, adolescents and language change. *Language Variation and Change* 8: 177–202.

Kuhl, Patricia
1994 Learning and representation in speech and language. *Current Opinion in Neurobiology* 4: 812–822.

Labov, William
1990 The intersection of sex and social class in the course of linguistic change. *Language Variation and Change* 2: 205–254.
1994 *Principles of Linguistic Change:Volume 1 – Internal Factors*. Oxford: Blackwell.
2001 *Principles of Linguistic Change: Volume 2 – External Factors*. Oxford: Blackwell.

Lindblom, Björn
1990 Explaining phonetic variation: a sketch of the H and H theory. In: William Hardcastle and Alain Marchal (eds.), *Speech Production and Speech Modelling,* 403–439. Dordrecht: Kluwer.

Local, John
2003 Variable domains and variable relevance: interpreting phonetic exponents. *Journal of Phonetics* 31: 321–339.

Locke, John
1983 *Phonological Acquisition and Change*. New York: Academic Press.

MacNeilage, Peter
1980 The control of speech production. In: Grace Yeni-Komshian, James Kavanagh and Charles Ferguson (eds.), *Child Phonology. Volume 1: Production*, 9–21. New York: Academic Press.
1997 Acquisition of speech. In: William Hardcastle and John Laver (eds.), *The Handbook of Phonetic Sciences*, 301–332. Oxford: Blackwell.

Milroy, Lesley, James Milroy and Gerard Docherty
1997 *Phonological Variation and Change in Contemporary Spoken British English*. Final report to the ESRC, research grant R000234892.
Mufwene, Salikoko
1994 Theoretical linguistics and variationist analysis: strange bedfellows? *Papers from the 30th Meeting of the Chicago Linguistics Society (Parasession on Variation),* 202–217.
Payne, Arvilla
1980 Factors controlling the acquisition of the Philadelphia dialect by out-of-state children. In: William Labov (ed.), *Locating Language in Time and Space,* 143–178. New York: Academic Press.
Pierrehumbert, Janet
2001 Words and wordoids. Paper presented at Workshop on Early Phonological Acquisition. Carry-le-Rouet, October 2001. http://www.ling.northwestern.edu/~jbp/carry/carry.html (accessed 18/4/06).
Pisoni, David
1997 Some thoughts on 'normalization' in speech perception. In: Keith Johnson and John Mullennix (eds.), *Talker Variability in Speech Processing,* 9–32. San Diego: Academic Press.
Roberts, Julie
1997a Acquisition of variable rules: A study of (-t, d) deletion in preschool children. *Journal of Child Language* 24: 351–372.
1997b Hitting a moving target: Acquisition of sound change in progess by Philadelphia children. *Language Variation and Change* 9: 249–266.
2001 Child language variation. In: Jack Chambers, Peter Trudgill, and Natalie Schilling-Estes (eds.), *The Handbook of Language Variation and Change*, 333–348. Oxford: Blackwell.
Roberts, Julie and William Labov
1995 Learning to talk Philadelphian. *Language Variation and Change*, 7:101–112.
Sabbagh, Mark and Susan Gelman
2000 Buzzsaws and blueprints: what children need (or don't need) to learn language. *Journal of Child Language* 27: 715–726.
Scobbie, James
this vol. Flexibility in the face of incompatible English VOT systems.
Shockey, Linda and Zinny Bond
1980 Phonological processes in speech addressed to children. *Phonetica* 37: 267–274.
Snow, Catherine
1995 Issues in the study of input: finetuning, universality, individual and developmental differences, and necessary causes. In: Paul Fletcher

and Brian MacWhinney (eds.), *The Handbook of Child Language,* 180–193. Oxford: Blackwell.

Vihman, Marilyn
1996 *Phonological Development*. Oxford: Blackwell.

Vihman, Marilyn and Shelley Velleman
2000 Phonetics and the origin of phonology. In: Noel Burton-Roberts, Philip Carr and Gerard Docherty (eds.), *Phonological Knowledge: Conceptual and Empirical Issues,* 305–339. Oxford: Oxford University Press.

Watt, Dominic and Lesley Milroy
1999 Patterns of variation and change in three Newcastle vowels: is this dialect levelling? In: Paul Foulkes and Gerard Docherty (eds.), *Urban Voices,* 25–46. London: Arnold.

Wells, Gordon
1986 *The Meaning Makers*. Portsmouth: Heinemann.

Variation in developing phonologies: Comments on Vihman and colleagues, Docherty and colleagues, and Scobbie

Shelley Velleman

Variation has often been ignored in favor of “ideal” speaking-hearing conditions or considered to be an obstacle to be overcome over the course of development of a mature phonological system. In many models, such an adult phonological system is presumed to include a single static underlying representation of each word and a strictly-ordered set of principles (e.g., constraints) that determines an invariant output for each. Most studies of early child phonological development have ignored variation in the ambient language to which the child is exposed as well as in the child’s productions, choosing instead to assume an accurate invariant target for each word and to focus on one “typical” child output. The basic assumption has been that variant information about speaker physiology (e.g., age, gender, size), dialect, and other factors is somehow filtered out, yielding a single abstract percept (e.g., a phoneme-based underlying representation) against which incoming stimuli are compared. One major task for the child, in this view, is to learn what to ignore. The implication of such reductionist theories is that the learner will master the more abstract invariant properties of the language, those that are key to lexical contrasts, before more variable aspects of the input are reflected in his or her speech (Docherty et al. this volume).

More recently, other authors (e.g., Johnson 1997) have proposed models in which new instances are compared for categorization to an accumulation of memory traces, which retain details about the speaker, the context, etc. These models are a response to the fact that such factors have been shown experimentally to influence the efficiency of listeners’ recognition processes. Thus, perception is a “resonant state between signals and memories” (Goldinger 1998: 267). Attentional weighting determines the influence of different aspects of variation, as well as the roles of word frequency, recency, semantic priming, and syntactic priming, in speech recognition. In this view, retaining variability within one’s internal representation of a word does not impair efficient speech recognition, but in fact facilitates it. Abstraction is not eliminated; rather it "occurs during retrieval as countless partially redundant traces respond to an input" (Goldinger 1998: 251).

The majority of the studies that have addressed the question of the impact of variation on speech processing have focused on adult speech perception. A largely untapped source of data regarding the processing and storage of variability in the speech signal exists in learners' responses to such variation, as exemplified by their attempted reproductions of the variable input to which they have been exposed. Furthermore, linguistic variation (variously attributed to exceptions to rules, constraint interactions, interactions among linguistic units at different levels, etc.) has typically been studied independently of paralinguistic sources of variation. The set of papers discussed here raises intriguing questions about the types of variation that may occur in a learner's linguistic environment, and the impact of such variation on the resulting linguistic system at different points in development. In addition to ambient conditions and phonetic contextual effects (Pisoni 1997), the types of variation to which the learner is exposed may include:

1. the consistency of given patterns within a language (applicability across the lexicon; applicability at different levels e.g., word versus phrase);
2. the transparency of given patterns within a language (e.g., number and nature of cues);
3. linguistic cues (i.e., cues embedded in the language) that may simultaneously signal both social or physiological differences (e.g., age, gender, or social status) and linguistic differences (i.e., lexical contrasts); and
4. exposure to more than one system:
   a. where the two systems differ phonologically in substantial ways (i.e., different languages)
   b. where the two systems are the same phonologically, but differ with respect to the phonetic instantiations of that phonology (i.e., dialects of the same language).

Type 4a represents the situation of the emerging bilingual, and has been discussed at length elsewhere. The other types of variation listed above have been largely ignored until recently, and they are the focus of the following discussion.

Vihman, Nakai, and DePaolis (this volume), Docherty, Foulkes, Watt, & Tillotson (this volume), and Scobbie (this volume) also allow us to view the learner's response to these types of variation at various stages in the developmental process: at the onset of word usage (the *4 word point*), at the emergence of phonological systemization (the *25 word point*), during a slightly later period of rapid linguistic expansion (ages 24–48 months), and in young adulthood (late teens–early 20s). As a group, these papers validate theories that incorporate phonetic details into the listener's representations. I will argue that these papers, together with additional recent results to be presented below, also validate Pierrehumbert, Beckman, and Ladd's (2000) claim that "variability causes the need for abstraction" (p. 292), and is therefore critical to the development of a far more complex, far more flexible, adult phonological system than has been proposed before.

Vihman and co-authors (this volume) compare the very early acquisition of three languages whose phonologies differ with respect to the amount of linguistic ambiguity present in their metrical structure. According to these authors, French, with iambic lexical and phrasal stress (marked primarily by lengthening of the stressed final syllable) and syllable-timing, is the most consistent of the three. Its metrical system is phonetically grounded in phrase-final lengthening and is homogeneous across phonological levels. It also applies across the board to all items within the lexicon, except for onomatopoeia which are equally-stressed but very rare. In fact, it has been claimed that stress is not lexically encoded by French speakers at all, based upon evidence that such speakers are actually *stress deaf* – unable to distinguish stress minimal pairs in other languages (Peperkamp, Dupoux and Sebastian-Galles 1999).

In contrast to French, Vihman and colleagues (this volume) tell us, English has a predominance of trochaic words, although about 25% of English lexical items are iambic. The English metrical waters are further muddied by the facts that English is stress-timed (which adds further variability to the ambient input) and that about 70% of the words addressed to children are monosyllabic. Therefore, about half of the disyllables that American children hear are, in fact, bisyllabic article+noun phrases, which are iambic in English. Another 4% are equally-stressed onomatopoeia. There is also more phonotactic variability in English children's word targets, including consonant clusters, the inclusion of two different consonants within the same word, and the use of syllabic sonorants. However, stress is quite clearly marked in English, with the vowel nucleus of the stressed syllable being produced with more intensity, a longer duration, and a higher pitch.

Vihman and colleagues argue that Welsh falls somewhere in between these two systems, with predominantly trochaic words and predominantly iambic phrases (as in English), but a higher percentage of disyllables represented by single words. Thus, disyllables are more likely to be trochaic in Welsh than in English (because a given disyllable addressed to a child is more likely to be a word rather than a phrase). This effect is mitigated by the fact that iambic onomatopoeia are more frequent, at 12% of words in child-directed speech in Welsh. However, the onomatopoeia have the advantage of being phonotactically simple; child-directed speech words are also phonotactically simpler than in English. Welsh rhythm is intermediate between stress-timing and syllable-timing. In these senses, then, Welsh is somewhat less variable than English. However, the cues to stress are less clear-cut in Welsh than in English: stressed syllables are of higher intensity, but they are shorter in duration rather than longer. Furthermore, the following consonant, not the vowel nucleus, is lengthened, and the vowel nucleus in the following syllable is of higher pitch. In short, the increase in prominence on stressed syllables in Welsh is less transparent than in English because the cues are far less localized.

What do children on the brink of meaningful speech make of these systems? At the 4-word point (see Vihman, Satsuki and DePaolis this volume for a definition), three out of five French children, two out of five Welsh children, and only one out of five American children produce stressed/unstressed syllable duration ratios that are roughly similar to those of the adults in the same language group. In all three groups, individuals demonstrate very distinct patterns, although the French infants' productions appear to be more homogeneous. This French>Welsh>American pattern is even more evident at the 25-word point, at which all of the French and four out of five of the Welsh but none of the American children have mastered the predominant ambient metrical pattern for disyllables. As a result, the French and Welsh groups are also more homogeneous than the American group. The Welsh children attempt, and match, more varied stress patterns than the French, but not quite as consistently. Apparently, the less-localized nature of stress in Welsh is less of a challenge for learners than the overall higher levels of variability in English, perhaps due to the consistency of the association among the cues.

Vihman and colleagues emphasize the individual nature of each child's learning path, and conclude that "each child can be seen to be developing a stock of individual representations of particular words from which the more abstract structure of the adult language will eventually be induced" (p. 27). The representations that the American children must develop are more varied overall (including phonotactic factors), and the children are therefore slower to pick out and match the predominant patterns. The French children, on the other hand, as argued by Peperkamp and colleagues (Peperkamp, Dupoux and Sebastian-Galles 1999), need not represent stress patterns lexically at all. They "master" the adult system not by abstracting from a set of patterns in the ambient language but by ceasing to attend to stress linguistically.

The comparison among these three languages suggests that the language-learning process is indeed sensitive to variability in the languages to which the children are exposed. At this point in development, variability does indeed slow down the learning process, as predicted by reductionist theories, but it is noteworthy that the type of variability that has delayed the American children's metrical learning is neither social nor physiological, but linguistic. They do not appear to have been able to sort out the different types of linguistic variability present in their ambient language: grammatical (word versus phrase), lexical (iambic versus trochaic words), or phonotactic (duration differences due to clusters versus singletons, etc.).

Docherty and co-authors (this volume) investigate the development of phonological variation in slightly older children, 2–4 years of age, and their development of both linguistic and sociolinguistic sources of variation. These authors explore the systematic variability of adult speakers of Tyneside English, both in adult-adult speech and in child-directed speech. They demonstrate that, while child-directed speech does differ from adult-adult speech, it is actually no less variable, and that it reflects the same social factors as adult-adult speech. One of the key points made by the authors is that the same linguistic cue (i.e., a phonetic cue embedded within a linguistic context) can be used by speakers and listeners to transmit sociolinguistic information, such as the gender, age, and social group of the speaker, as well as to transmit linguistic information, such as a lexical contrast. The absolute frequencies of vowel formants, for example, reveal important physiological facts about the speaker, while their relative frequencies differentiate different vowel phonemes (Pisoni 1997). Clearly, to be a fluent member of the dialect community, a learner must eventually master both. The question these authors pose is whether children learn the linguistic value of such cues first, before demonstrating awareness of their social value. The answer provided by Docherty and colleagues is "no". Their results provide support for item learning in that the children produce one very restricted variant of the adult phoneme /t/ only in the appropriate lexical items. The use of this variant is not generalized across /t/ contexts by any of the children, as a reductionist theory might predict. Furthermore, children's rates of production of various sociolinguistic variants are related to frequencies of use in the mother's speech, indicating that the learners are sensitive to probabilities as well as contexts of occurrence. By 3;6 even gender differences, paralleling those in the adult community, begin to emerge. Docherty and colleagues conclude that their subjects demonstrate a "...high degree of sophistication in their ability to make correct analyses of phonological patterns based on both linguistic and non-linguistic factors" (p. 22) despite, or perhaps due to, the lack of invariance in the ambient language.

Scobbie (this volume) presents data from far older subjects, young adults with a mean age of 19 years (range 16–30 years) from the Shetland Islands in northern Scotland. As a result of a variety of demographic factors, these subjects have been exposed to dialects which differ phonetically in a critical way: the voice onset time boundaries in the different dialects are contradictory with respect to the phonological voicing categories of certain exemplars. There is no one consistent manner in which every subject in any dialect group responds to this linguistic variability. Some subjects' phonetic patterns parallel those of one of the dialects to which they were exposed; others have adopted intermediate values; and still others use the extreme voice onset time values offered by both (i.e., one extreme from each). Yet, these speakers are mutually intelligible. Scobbie's results thus strongly support the co-existence of both

– abstract categories (such as *voiced* and *unvoiced*) that are critical to lexical decision making, and
– detailed exemplar sets from which those traces retained from previous exposure to similar speakers are most highly activated to support the efficient perception of such speakers' current utterances.

Goldinger (1998) proposes such a hybrid model in which detailed memory traces co-exist and co-function alongside more abstract representations, and the memory traces themselves may be influenced by cognitive processes as well as by perceptual experiences.

How could such a model be applied to the developmental process? The potential organizing principles derived thus far from the studies above include:

1. Linguistic as well as paralinguistic sources of variability impact children's ability to abstract general patterns (Vihman, Satsuki and DePaolis this volume);
2. Children appear to respond to both types of variability, by the age of 3;6 if not earlier (Docherty et al. this volume);

3. At the same time, children appear to be sensitive to variants with restricted ranges; that is, there are at least some allophonic patterns that they do not overgeneralize, suggesting that item learning is occurring (Docherty et al. this volume); but
4. It's not just item learning; some abstraction must occur given that speakers on Shetland are able to process the voiced/voiceless distinction in others' speech even when the voice onset time boundaries are apparently phonologically contradictory (Scobbie this volume).

The studies reviewed above are largely atheoretical, though they have important implications for theories of phonology. Clearly, an appropriate model must allow for the influence of a variety of factors and for variable outcomes across production tokens of the same or similar word types by the same child. The relatively recent phonological model of Optimality Theory is more flexible than previous generative or parameter-based accounts, in that it posits a ranked set of constraints whose interactions determine output forms. Within this theory, phonology is taken to consist of universal sets of *markedness constraints*, which specify the features, patterns, or structures that are typically preferred by human languages, and *faithfulness constraints*, which specify the features, patterns, or structures of the *input* (presumed underlying representations of words) that are preferentially preserved in production universally and in specific languages. Faithfulness corresponds approximately to linguistic conventionality among the speakers of a language, with contrast maintenance as its motivation. That is, words must be different enough from each other to be distinguishable. In the terms of the discussion above, item learning corresponds roughly to faithfulness in the sense that both require preservation of the characteristics (structure and elements) of a particular word.

Markedness constraints, on the other hand, reflect preferred output forms – which are typically, but not always, phonetically grounded. Thus, many markedness constraints may be based upon human physiological limitations, experienced by the learner through his/her prelinguistic and early linguistic vocal activities. Those constraints are *inductively grounded* to begin with (Boersma 1999; Hayes 1999), and are then *phonologized* at some point into categorical phonological constraints. However, markedness constraints can be language-specific as well as universal. In the terms used above, language-specific markedness corresponds to consistency of patterning within a language. Those patterns that are preferred, and therefore most common, in the language are unmarked; those that are rare or nonexistent are highly marked. Either way, the markedness of a form can be determined independently of the target, while faithfulness can only be determined by comparison of the output to the target.

Within Optimality Theory, each language has an established *ranking* for the set of faithfulness and markedness constraints. The constraints are presumed, by a principle known as Richness of the Base, to be universal. The ranking of the constraints determines which sounds and combinations of sounds are present in any given language. The ranking further determines which characteristics of a given underlying form are actually preserved and which are sacrificed in a given output (production). Thus, whether or not a given constraint is respected in a given production depends upon which other constraints are operative in that word. The irrelevance of some higher-ranked constraints to particular word forms allows the influence of lower-ranked constraints to be evident in some cases. In this sense, some variability is inherent to the theory.

There are two problems with previous applications of Optimality Theory to child phonological data, however. The first is inherent to the theory as originally proposed and the second is a hypothesis that has inappropriately been received as an axiom:

1. Optimality Theory was designed to account for the fact that different phonological principles (constraints) appear to bear primary responsibility for production patterns under different linguistic circumstances. However, every production of the same target word is expected to have the same phonetic form, because the only factor that influences the role played by each constraint is its applicability to that particular phonological target. Furthermore, each word of the lexicon is presumed to have one underlying target form (the *input*, in Optimality Theory terms). Thus, phonetic, paralinguistic and sociolinguistic factors are not considered to be part of the lexical representation nor of the phonological process of generating the output form.
   In order to reconcile this theory with the data presented above, it is necessary to introduce two additional sources of variation:
   a. variable target forms, presumably stored in the lexicon, and
   b. gradient constraint-ranking, in which constraints are ranked probabilistically rather than strictly (Boersma and Hayes 2001; Boersma and Levelt 2000; Hayes 2000). That is, each constraint has a normally distributed range of ranking values. Therefore, constraint A that normally dominates constraint B may occasionally be violated in favor of B if their ranges of ranking values overlap. The constraints remain categorical in nature (i.e., their descriptions refer to abstract elements and structures, not to fine-grained articulatory or acoustic distinctions),

but their application patterns are, to some extent, stochastic. Furthermore, the constraints are continuously rather than discretely ranked. Davidson, Smolensky and Jusczyk (2004) have recently shown the usefulness of such a model in accounting for the variable productions of adults under different task conditions.

2. It has been assumed, based upon Gnanadesikan's (1995) findings and claims about a child already well beyond the stage of first-word production, that all markedness constraints are initially ranked above all faithfulness constraints (e.g., Demuth 1995; Gnanadesikan 1995; Levelt and van de Vijver 1998; Smolensky 1996; Smolensky, Davidson, and Jusczyk 2004). This proposal (MARK>>FAITH) implies that language-specific effects should not be evident at the *initial state* (onset of word use). However, studies such as the current one by Vihman and colleagues (and many others including Foulkes, Docherty and Watt 1999; Johnson and Jusczyk 2001; Levelt and van de Vijver 1998; Roark and Demuth 2000; Smolensky, Davidson and Jusczyk 2004; Stoel-Gammon, Buder and Kehoe 1995; Tesar 2000; Tesar and Smolensky 1998) clearly demonstrate ambient language effects in the very earliest stages of word production.

Both of these issues were explored in a recent study by Velleman and Vihman (2002). In this work, the authors examined early constraint rankings using data from twenty children, five each learning one of four languages: English, French, Japanese, and Welsh. The participants' pseudonyms, languages, and ages at the two data points (where applicable) are given in Table 1.

Data from each of the twenty children were analyzed at the 25-word point. Each token of each word produced by each child in each session was considered with respect to the core set of phonotactic markedness and faithfulness constraints relevant to early child phonologies (as suggested by Bernhardt and Stemberger 1998; Demuth 1996; Fee 1996; Gnanadesikan 1995; Kehoe and Stoel-Grammon 1997; Levelt and van de Vijver 1998; and others). To facilitate mark-faith comparisons, markedness and faithfulness constraints were paired with respect to their effects on the output. For instance, the frequency of violations of the markedness constraint that requires a consonantal onset for every syllable – ONSET – was compared to the frequency of violations of a faithfulness constraint against omitting or adding an onset to the heard form – CORRESPONDENCE(ONSET), a.k.a. CORR(ONSET). The relative ranking of ONSET versus CORR(ONSET) was compared within children and also across children from different language backgrounds. Frequencies of violation of these pairs of related markedness and faithfulness constraints were compared to derive a probabilistic constraint ranking for each child.

The purpose was to determine whether or not markedness constraints consistently outranked faithfulness constraints (either generally or specifically), whether strict or gradient rankings of these constraint pairs could be identified, and further whether children's errors reflected articulatory difficulty only or other more abstract factors as well.

Table 1. List of subjects by language, name, and age.
*Last session @ 19-word point.

Name	Age @ 25 Words	Age @ 4 Words
ENGLISH		
Deborah	1;3.24	0;11.4 & 0;11.11
Emily	1;3.29	
Molly	1;2.20	
Sean	1;3.23	
Timmy	1;4.22	
FRENCH		
Carole	1;2.5	0;10.26 & 0;11.10
Charles	1;3.19	
Laurent	1;5.15	
Marie	1;7.24	
Noel	1;5.23	
JAPANESE		
Taro	1;11.2	1;03.09 & 1;03.14
Emi	1;4.7*	
Haruo	1;7.17	
Kazuko	1;3.28	
Kenji	1;6.17	
WELSH		
Fflur	1;5.2	1;1.6 & 1;1.20
Carys	1;5.29	
Gwyn	1;2.24	
Elen	1;6.6	
Nona	1;6.18	

The initial plan for this research was to evaluate the children's constraint violations at their 25-word points, as the earliest point at which a substantial number of word types and tokens is available in a child's speech. However, given the somewhat surprising results found at the 25-word points (to be discussed below), four children (one per language) were considered in greater depth. These children's constraint rankings were studied at the 4-word-point as well. Two 4-word point sessions were used for each child to increase the number of word tokens available for analysis.

The source of information for each child's phonology at each point was the same: transcripts of the child's word productions during unstructured 30-minute audio- and video-recorded parent-child play sessions. The two developmental points, the 4-word point and the 25-word point, were defined as the sessions in which the child's expressive lexicon reached a level of approximately 10 or 50 words, respectively, as evidenced by:

– parental diary report of 5–10 or 40–60 words, respectively, and
– the production of at least four or 20–30 words, respectively, during the 30-minute recording session (see Vihman and McCune 1994 for the protocol for distinguishing words from babble vocalizations).

Data for English and Japanese were collected in California, USA; for French in Paris, France; and for Welsh in Bangor, Wales (UK). All children were being raised as monolinguals in their respective languages at the time of the recordings (see Boysson-Bardies and Vihman 1991; Vihman and McCune 1994 for further details).

Native transcribers prepared transcripts of each child using the International Phonetic Alphabet. Reliability was first tested within each language. Although many prelinguistic vocalizations were included in these measures, agreement as to the specific identity of the consonant ranged from .75 (Japanese) to .80 (French and English). Crosslinguistic reliability was also checked for some pairs of languages, with percentage of agreement ranging from .81 to .86 (see Boysson-Bardies and Vihman 1991; Vihman et al. 1985 for further details). Transcription differences were resolved by consensus or by a third transcriber.

Every token of every word type produced during each session was analyzed in two ways:

– Markedness: Each token was categorized as either satisfying or violating each markedness constraint. For example, if a syllable of the word as produced did not begin with a consonant, that syllable was coded as violating the constraint ONSET, regardless of the target word.
– Faithfulness: Each token was categorized as either matching or violating the target word's status with respect to each constraint. For example, if the overt form lacked a consonant onset, but the child produced a

> consonant onset, this would be considered to be a violation of the constraint CORR(ONSET), even though the child's production was respectful of markedness. An inherent limitation of this procedure is that it assumes that just one target/overt form is stored by the child, and that it fully matches one invariant form produced by adults. This simplifying assumption is made quite consciously here. Its impact on the results will be discussed later in this paper.

Further details of the constraints used and of the ranking procedures are available in Velleman and Vihman (2002).

The results of this study disconfirm both of the common Optimality Theory assumptions about early phonology described above:

- that constraints are strictly ranked, yielding a consistent output form for each target word, and
- that all markedness constraints outrank all faithfulness constraints at the onset of word production.

The children's violations of the phonotactic constraints studied range from 0% to 100%. Certain universal markedness constraints, especially PEAK (the requirement that every syllable include a vowel peak), appear to be very highly ranked in all of the children's systems, regardless of the language of exposure. Similarly, the markedness constraint NOCODA (final consonants are disallowed) dominates the related faithfulness constraint CORR(Coda) (the produced form must match the target form with respect to the presence or absence of a coda) for all languages except Welsh; the children frequently omit codas when they are called for by the adult language. However, the position of NOCODA in the overall ranking differs from language to language. For example, in Japanese NOCODA is ranked very highly while CORR(Coda) is quite frequently violated – the children frequently omit the few target codas, especially in word-final position, and when they do produce a coda (typically in syllable-final word-medial position), it may not have been present in the target. For example, Emi produces *wan-wan* 'doggy' as [wɑwɑ] (occasionally with a word final glottal stop), yet one of her productions of *mama* is [mɑmmɑʔ]. In the other three languages, both NOCODA and CORR(Coda) fall towards the middle of the overall ranking. Welsh children may over-ride the tendency for NOCODA to dominate CORR(Coda) due to a combination of factors. First, Welsh includes many final consonants (as does English, but not French or Japanese). Furthermore, despite a trochaic stress pattern, these final consonants are typically released (unlike final consonants in English).

Thus, certain markedness constraints do tend to dominate the corresponding faithfulness constraints in the majority of the languages studied. However, the children's markedness constraints do not consistently outrank their faithfulness constraints. Several faithfulness constraints, such as CORR(PEAK) (produce a peak if there is one in the target, but don't if not), CORR(CC) (produce an intrasyllabic consonant cluster if there is one in the target, but don't if not), and CORR(ONSET) (produce an onset consonant if there is one in the target syllable, but don't if not) appear near the top of the rankings in all four languages. (Note: These CORR(ONSET) results, and also the ONSET results, were the same when calculated by treating glottal stop as an onset as when calculated treating glottal stops as non-consonantal onsets.) Furthermore, as exemplified by NOCODA and CORR(Coda) in Welsh, the lack of dominance of markedness constraints applies at the level of specific constraints as well as generally. That is, not only do all markedness constraints *not* dominate all faithfulness constraints, but specific markedness constraints also fail to dominate the related faithfulness constraints. For example, CORR(ONSET) dominates ONSET in all languages except Japanese, in which the two constraints are very closely ranked but in the opposite direction. Clearly, the frequency or salience of vowel-initial words in English, French, and Welsh have already made an impact on these children's phonologies by early in the word production process, such that they produce vowel-initial words as such despite the relative articulatory difficulty of doing so.

Similarly, CORR(SYLBIN) (produce an even or odd number of syllables, depending on the number in the target word) dominates SYLBIN (produce an even number of syllables) for all languages. As is known from other types of psycholinguistic studies (Aitchison and Chiat 1981; Smith, Macaluso and Brown-Sweeney 1991; Vihman 1981), the number of syllables in a word – or perhaps the rhythmic pattern derived from even versus odd numbers of syllables – is highly recognizable and this apparently has an impact on the child's early phonological system. In fact, binarity markedness constraints in general are very low-ranked in all of the children's phonologies, across all languages. Thus, binarity may not play as strong a role as has sometimes been assumed (Demuth 1996; Fee 1996).

The finding that all markedness constraints dominate all faithfulness constraints neither generally nor specifically holds true for each individual language as well. In English, for instance, particular markedness constraints do outrank the corresponding faithfulness constraints. For example, *CC (do not produce intrasyllabic consonant clusters) dominates CORR(CC), reflecting the fact that intrasyllabic consonant clusters are often simplified to singletons. Similarly, NoCoda (produce no final consonants) is violated less often than Corr(Coda) (produce a final consonant if there is one in the target but don't do so if not); the children tend to omit codas even when they appear in the target (e.g., Emily's [bɪ] for *bib*). NoCoda and Corr(Coda) appear fairly low in the overall ranking, however, indicating that both are frequently violated. Furthermore, there is a wide range in the children's rankings of NoCoda – from near the top of the hierarchy (rarely violated) to near the bottom (rarely satisfied). In fact, two of the children (Molly and Sean) actually rank Corr(Coda) somewhat higher than NoCoda; they produce codas whenever they are called for. In addition, certain faithfulness constraints, such as Corr(Onset), are very rarely violated in English regardless of the markedness of the resulting form (e.g., a word with a vowel onset). Furthermore, even some individual faithfulness constraints that are sometimes violated (such as Corr(Onset), Corr(SylBin), and Corr(Geminate: if the target includes a geminate, then the production must also; if not, it must not) nonetheless do dominate the corresponding markedness constraints (i.e., Onset, SylBin, and Geminate: sequences of two consonants must be made up of two identical consonants i.e., a geminate).

The French children, like the English-learners, show clear evidence of faithfulness to phonotactic features of target words at this early stage of word production. Not surprisingly, given that French has fewer codas than English and Welsh, the universal markedness constraint against codas does dominate faithfulness to syllable-final consonants in their speech. However, Corr(Onset), Corr(SylBin) and Corr(MoraicBin: if the target syllable includes two moras, then the produced syllable must also) all outrank the corresponding markedness constraints for all of the French children.

For Japanese, too, we see several high-ranked faithfulness constraints, both generally and also specifically. Particular faithfulness constraints such as Corr(SylBin) and Corr(Gem) outrank the corresponding markedness constraints. Kazuko, for instance, always produces a geminate when one is called for by the adult form, but never produces any other medial consonant sequence as a geminate (though she does sometimes geminate singleton medial consonants).

In Welsh, faithfulness to syllable binarity, to the presence/absence of an onset, and to the presence/absence of a coda are more highly ranked than the related universal markedness principles. The Onset and NoCoda markedness constraints are particularly low-ranked, far below the corresponding faithfulness constraints. The fact that final consonants are generally released in Welsh may be a factor in the low ranking of NoCoda. Interestingly, the

Welsh children not only preserved codas as targeted but also added codas even where these were not required by the target. For example, Elen produced *baby* as [babap] as well as [baba]. In contrast, many of the words that these children attempted began with vowels or glottals ([h] or [ʔ]) in the target; these non-salient onsets may have influenced them to downgrade Onset. In summary, each child speaker of each language studied provided evidence against the Mark>>Faith hypothesis. General trends also were seen across children by language.

Our findings also cast doubt upon the assumption of strict constraint domination, at least in child phonology. Every child in the study exhibited at least one pair of mark-faith related constraints for which the resulting word productions varied in their outcome, sometimes violating the markedness constraint and sometimes the faithfulness constraint; most exhibited several such pairs. These variable outcomes could not be attributed to within-word coarticulatory effects, as they occurred across multiple productions of the same word. Given that these children were at or just barely emerging from the one-word stage of production, this variability cannot be attributed to phrasal effects, either.

It is true that only phonotactic constraints were considered in this study. It is quite possible that some of the violations of these constraints could be due to the influences of segmental constraints. For example, Deborah's productions of *kitty* with a [tl] or [kl] medial consonant cluster (e.g., as [kɛkli]) could reflect a faithfulness constraint requiring that the lateral feature of the medial flap be preserved. Even if so, the fact remains that Deborah also produced this same word without a medial cluster upon occasion (e.g., as [k^hiwe]). Therefore, the purported faithfulness constraint, whatever it may have been, applied variably (i.e., gradiently) to this word.

Evidence in these results for gradient constraint ranking also include evidence against articulatory limitations as an adequate explanation for discrepancies between target forms and child forms. In many cases, the children's errors constituted more difficult productions than required by the target. Something other than motor difficulty is compelling the child to produce a different, more articulatorily challenging form. Phonologization, i.e., generalization of grounded constraints beyond their motivating word forms, seems the most likely explanation. Another possible explanation is that the child, having been exposed to systematically variable forms of the targets, has stored underlying forms that include such gradient information, and that her/his outputs reflect those variable targets. It has been assumed (e.g., by Dinnsen et al. 2000) that the child's underlying representations of words match one canonical surface adult form. Thus, all children learning the same language are presumed to have the same, invariant, word targets. This assumption ignores the child's perceptual experience of variable adult productions (Docherty et al. this volume; Foulkes, Docherty and Watt 1999; Matthews 2002; Scobbie this volume). If constraints, underlying forms, or both are *induced* based upon statistical learning from the ambient overt forms, or based upon the child's own motoric experience, or both, then variability should be a hallmark of the child's output, as it will have pervaded his/her experience of both types. The results of this study, and those of Vihman and colleagues (this volume), Docherty and co-authors (this volume) and Scobbie (this volume), confirm this assumption. Gradient constraint rankings applied to invariant targets and strictly ordered constraint rankings applied to systematically variant targets (reflecting ambient-language patterns) are alternative means of modeling this reality; they may be empirically undifferentiable.

Thus, Velleman and Vihman have shown that children in the very early stages of word production have already moved on from phonological universals in two respects:

– they already show a high degree of faithfulness to individual word forms (item learning), and
– their sets of markedness constraints have been altered (as well as reranked) to match the tendencies of the languages to which they have been exposed.

In summary, variation is already a hallmark of the very earliest word productions, and this variation is systematic and not merely the result of poor articulatory control. Both abstraction away from examplars and item learning play roles in individuals' phonological systems. Both linguistic and sociological variables influence the child's production in systematic ways. Optimality Theory as originally conceived with both invariant targets (inputs) and strictly ranked constraints cannot be used to model this variation. However, a model of Optimality Theory that allows for either variant targets or gradiently-ranked constraints and for learned, child-specific ambient-dialect-based markedness constraints would be appropriate for this purpose.

References

Aitchison, Jean and Shula Chiat
1981 Natural phonology or natural memory? The interaction between phonological processes and recall mechanisms. *Language and Speech* 24: 311–326.

Bernhardt, Barbara H. and Joseph P. Stemberger
1998 *Handbook of Phonological Development from the Perspective of Constraint-Based Nonlinear Phonology*. San Diego: Academic Press.

Boersma, Paul
1999 On the need for a separate perception grammar. M.S. thesis, University of Amsterdam.

Boersma, Paul and Bruce P. Hayes
2001 Empirical tests of the gradual learning algorithm. *Linguistic Inquiry* 32: 45–86.

Boersma, Paul and Clara Levelt
2000 Gradual constraint-ranking learning algorithm predicts acquisition order. In: Eve V. Clark (ed.), *Proceedings of the Thirtieth Child Language Research Forum*, Volume 30, 229–237. Stanford: CSLI Publications.

Boysson-Bardies, Benedicte de and Marilyn M. Vihman
1991 Adaptation to language: Evidence from babbling and first words in four languages. *Language* 67: 297–319.

Davidson, Lisa, Paul Smolensky, and Peter W. Jusczyk
2004 The initial and final states: Theoretical implications and experimental explorations of richness of the base. In René Kager, Joe Pater and Wim Zonneveld, eds. *Constraints in Phonological Acquisition.* Cambridge University Press.

Demuth, Katherine
1995 Markedness and the development of prosodic structure. In: Jill Beckman (ed.), *Proceedings of the North Eastern Linguistic Society*, Volume 25, 13–25. Amherst, MA: GLSA, University of Massachusetts.
1996 The prosodic structure of early words. In: James L. Morgan and Katherine Demuth (eds.), *Signal to Syntax: Bootstrapping From Speech to Grammar in Early Acquisition*, 171–184. Mahwah, NJ: Lawrence Erlbaum Associates.

Dinnsen, Daniel A., Laura W. McGarrity, Kathleen M. O'Connor and Kimberly A. B. Swanson
2000 On the role of sympathy in acquisition. *Language Acquisition* 8: 321–361.

Docherty, Gerard J., Paul Foulkes, Jenny Tillotson and Dominic Watt
this vol. On the scope of phonological learning: Issues arising from socially-structured variation.

Fee, E. Jane
1996 Syllable structure and minimal words. In: Barbara H. Bernhardt, John Gilbert and David Ingram (eds.), *Proceedings of the UBC International Conference on Phonological Acquisition*, 85–98. Somerville, MA: Cascadilla Press.

Foulkes, Paul, Gerard J. Docherty and Dominic Watt
1999 Tracking the emergence of sociophonetic variation. *Leeds Working Papers in Linguistics and Phonetics* 7: 1–25.

Gnanadesikan, Amalia E.
1995 Markedness and faithfulness constraints in child phonology. M.S. thesis, University of Massachusetts-Amherst.

Goldinger, Stephen D.
1998 Echoes of echoes? An episodic theory of lexical acess. *Psychological Review* 105: 251–279.

Hayes, Bruce P.
1999 Phonetically driven phonology: The role of optimality theory and inductive ground. In: Michael Darnell, Edith Moravscik, Michael Noonan, Frederick Newmeyer and Kathleen Wheatly (eds.), *Formalism and Functionalism in Linguistics: General Papers*, 243–285. Amsterdam: John Benjamins.
2000 Gradient well-formedness in optimality theory. In: Jost Dekkers, Frank van der Leeuw and Jeroen van de Weijer (eds.), *Optimality Theory: Phonology, Syntax, and Acquisition*, 88–120. Oxford: Oxford University Press.

Johnson, Elizabeth K. and Peter W. Jusczyk
2001 Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory and Language* 44: 548–567.

Johnson, Keith
1997 Speech perception without speaker normalization. In: Keith Johnson and John W. Mullennix (eds.), *Talker Variability in Speech Processing*, 145–165. San Diego: Academic Press.

Kehoe, Margaret and Carol Stoel-Grammon
1997 The acquisition of prosodic structure: An investigation of current accounts of children's prosodic development. *Language* 73: 113–144.

Levelt, Clara and Ruben van de Vijver
1998 Syllable types in cross-linguistic and developmental grammars. Paper presented at: *Third Biannual Utrecht Phonology Workshop*. Utrecht, the Netherlands.

Matthews, Benjamin
2002 On variability and the acquisition of vowels in normally developing Scottish children (18–36 months). Paper presented at: *Eighth Conference on Laboratory Phonology*. Yale University: New Haven, CT.
Peperkamp, Sharon, Emmanuelle Dupoux and Nuria Sebastian-Galles
1999 Perception of stress by French, Spanish, and bilingual subjects. *Proceedings of Eurospeech* 6: 2683–2686.
Pierrehumbert, Janet, Mary E. Beckman and D. R. Ladd
2000 Conceptual foundation of phonology as a laboratory science. In: Noel Burton-Roberts, Philip Carr and Gerard J. Docherty (eds.), *Phonological Knowledge: Conceptual and Empirical Issues*, 273–304. New York: Oxford University Press.
Pisoni, David B.
1997 Some thoughts on "normalization" in speech perception. In: Keith Johnson and John W. Mullennix (eds.), *Talker Variability in Speech Processing*, 9–32. San Diego: Academic Press.
Roark, Brian and Katherine Demuth
2000 Prosodic constraints and the learner's environment: A corpus study. In: S. Catherine Howell, Sarah A. Fish and Thea Keith-Lucas (eds.), *Proceedings of the 24th Annual Boston University Conference on Child Language Development*, 597–608. Somerville, MA: Cascadilla Press.
Scobbie, James M.
this vol. Flexibility in the face of incompatible English VOT systems.
Smith, Bruce L., C. Macaluso and Sharon G. Brown-Sweeney
1991 Phonological effects shown by normal adult speakers learning new words: Implications for phonological development. *Applied Psycholinguistics* 12: 281–298.
Smolensky, Paul
1996 On the comprehension/production dilemma in child language. *Linguistic Inquiry* 27: 720–731.
Stoel-Gammon, Carol, Eugene H. Buder and Margaret Kehoe
1995 Acquisition of vowel duration: A comparison of Swedish and English. *Proceedings of the 13th International Congress of Phonetic Sciences* 4: 30–37.
Tesar, Bruce
2000 Using inconsistency detection to overcome structural ambiguity in language learning. *Rutgers University CCS Technical Report TR-58*.
Tesar, Bruce and Paul Smolensky
1998 Learnability in Optimality Theory. *Linguistic Inquiry* 29: 229–268.

Velleman, Shelley L. and Marilyn M. Vihman
2002 The emergence of the marked unfaithful. In: Angela Carpenter, Andries W. Coetzee and Paul De Lacy (eds.), *University of Massachusetts Occasional Papers in Linguistics 26: Papers in Optimality Theory II*, 397–419. Amherst, MA: GLSA, University of Massachusetts.

Vihman, Marilyn M.
1981 Phonology and the development of the lexicon: Evidence from children's errors. *Journal of Child Language* 8: 239–264.

Vihman, Marilyn M., Marcy A. Macken, Ruth Miller, H. Simmons and Jim Miller
1985 From babbling to speech: A reassessment of the continuity issue. *Language* 61: 395–443.

Vihman, Marilyn M. and Lorraine McCune
1994 When is a word a word? *Journal of Child Language* 21: 517–542.

Vihman, Marilyn M., Nakai Satsuki and Rory DePaolis
this vol. Getting the rhythm right: A cross-linguistic study of segmental duration in babbling and first words.

III. Knowledge of language-specific organization of speech gestures

Interaction of prosody and gestures

Local gesture interaction and perception

Prosody first or prosody last? Evidence from the phonetics of word-final /t/ in American English

Stefanie Shattuck-Hufnagel

The apparent syllabification of verb-final /t/ into the onset position of a following direct object pronoun, which has been described for Southern British English in phrases like escort us*, suggests that a) the process of phonological encoding involves the use of metrical frames that correspond to word-level prosodic constituents, and b) under some circumstances these word-level prosodic constituents can incorporate two lexical words, so that syllabification within them can sometimes obliterate allophonic cues to morphosyntactic boundaries. Cross-lexical-word syllabification in this dialect of English is inferred from the observation that in these contexts /t/ is produced with audible release noise, suggesting full oral closure with enough pressure buildup behind the constriction to create turbulence noise at its release. This implementation is viewed as characteristic of syllable onset position. Our results suggest that American English speakers do not produce verb-final /t/ before a direct object pronoun as an onset, but instead produce a glottalized or flapped /t/ that is more consistent with non-onset position, and thereby provides potential cues to the lexical boundary between the verb and the pronoun. This highlights the need to develop models of phonological encoding that can be adjusted to account for differences across language communities. It also underlines the desireability of closer study of the phonetics of both Southern British English and American English final /t/ in connected speech, to determine whether a noisy release is necessarily a correlate of syllable-initial position.*

1. Introduction

One of the greatest mysteries of human behavior is how a plan becomes an action. In the domain of speech, this problem can be expressed as the question of how a speaker turns a sentence into an utterance. It is reasonable to assume that at some point in the planning process the speaker's intended sentence is represented in terms of symbolic categories that carry contrastive meaning, and that at a later point the intended utterance is represented in terms of quantitative neuromuscular specifications which control articulatory movements. Although most theorists agree that the representation is symbolic up to the point of accessing the lexicon, just how and at what point in the planning process the symbolic representation of a sentence becomes a quantitative specification of the utterance is a point of disagreement. Views range from the fully articulatory lexical representations postulated by Articulatory Phonology to the fully symbolic phonetic representations captured by traditional IPA transcriptions, and from a fresh computation of quantitative parameter values entirely from scratch for each new utterance (as in Generative Phonology) to a fully selectional process in which the parameter values appropriate to each context are stored and retrieved as needed (as in an Episodic production lexicon).

A critical issue for all of these approaches is context-governed surface phonetic variation. Because phonetic variation is pervasive, extensive and systematic, it is imperative for a production planning model to provide an adequate account of it. Recent work suggests that a great deal of surface phonetic variation can be understood by relating it to the prosody, which can vary significantly from one utterance of a given sentence to another. Since the extensive development of theories of the prosodic hierarchy in the 1970s and '80s, an increasing amount of empirical evidence has been gathered in support of the hypothesis that the presence vs. absence of prosodic boundaries and prominences of various levels can have a systematic effect on the surface phonetic forms of words and sound segments. This concept has been embodied in a view of production planning called Prosody First, sketched out in Keating and Shattuck Hufnagel (2002). In this view, the speaker generates a morphosyntactic structure for the underlying sentence, and then, critically, uses this morphosyntactic structure to derive a prosodic structure for the specific utterance being planned; that prosodic structure includes a number of levels of constituent structure (e.g. Utterance, Intonational Phrase, Phonological Phrase, Prosodic Word) and of prominence (e.g. nuclear pitch accent, pre-nuclear pitch accent, unaccented full vowel, reduced vowel). It is this multi-level prosodic structure, along with other, non-grammatical aspects of prosody such as speaking rate, which then govern the surface phonetic form of the utterance. On this view, utterances of a given sentence which are organized into different constituent groupings and have different prominence patterns can show systematically different phonetic forms, and these differences involve not only variation in traditional prosodic parameters (e.g. f0, duration, amplitude) but also variation in the characteristics that are more traditionally associated with distinctive feature contrasts among segments (such as choice of articulator) as well as characteristics that have not historically been the focus of linguistic study (such as strength of articulation).

This Prosody First view can be extended to deal explicitly with the mechanism by which prosody can govern phrase-level-related variation in the phonetics of word form. For example, one might postulate that the abstract contrastive categories that characterize the lexical representations of words (Stevens 2002) are translated into abstract gestures, which can be thought of as articulatory configurations (Zsiga 2002), and the utterance-specific parameter values for these abstract configurational representations are computed on the basis of their location in the prosodic structure. This provides a mechanism by which the boundaries of constituents at different levels of the prosodic hierarchy can affect the phonetics differently, i.e. by specifying different values for timing and duration (Sproat and Fujimura 1993), and perhaps even in some cases different gestures altogether, as in glottal closure for syllable-final /t/ in some contexts and glottal spreading in others. On this view, variation in both the articulatory parameters and the resulting acoustic parameters can often be gradient-valued, i.e. the timing and amplitude of gestures will vary continuously (Browman and Goldstein 1992, Byrd and Saltzman 2003), although this may not occur when articulatory changes result in quantal shifts in the acoustics (Pierrehumbert 2001, 2002), or when different gestures are selected in different contexts to enhance the perceptual contrasts among distinctive features (Keyser and Stevens 2006). If this view is correct, many aspects of phonetic variation will be best described in terms of changes in articulatory parameters (such as overlap) governed by acoustic goals, rather than in terms of the categorical changes in feature values which are appropriate for phonological-level variation (such as vowel harmony, Zsiga 2002).

A different set of predictions emerges from the most thoroughly-developed model of production planning in the current literature, described in Levelt (1989) and extended in Levelt, Roelofs and Meyer (1999). In the word-form-encoding component of this model, only a single level of prosodic constituent, the Prosodic Word (PWd), governs segmental variation, and it does so via selection of the gestural scores for each of the syllables within the PWd. The LRM99 view, the PWd encompasses sequences such as <verb+affix>, as in *escorting*, as well as sequences such as <verb+pronoun>, as in *escort us*. (In other models of prosodic constituent structure, the PWd is defined differently; we adopt the LRM99 definition here for simplicity in the discussion.) Higher levels of prosodic structure, such as Intonational Phrases, are constructed later, and thus are not available to influence the selection of a gestural score for each syllable during Phonological Encoding. Thus, the model reflects a Prosody Last point of view; it predicts that while other levels of the prosodic hierarchy may affect traditional prosodic parameters (such as syllable F0, syllable duration and syllable amplitude), and possibly the timing of gestures within the syllable (e.g. with changes in duration), they will not affect the set of gestures that is chosen to implement each individual phonological contrast within the syllable, because this set is fixed during the earlier phonological encoding process.

Developed largely on the basis of data from British English and Dutch speakers, this Prosody Last model is inspired in part by an interesting observation about the behavior of word-final consonants before vowel-initial words in certain contexts in these languages. The prosodic constituent PWd is accorded a central role because of evidence that syllabification can occur across lexical word boundaries within this constituent. Such cross-lexical-boundary syllabification is taken as evidence that, during the planning process, the two lexical words have been combined into a single PWd, and that syllabification within this PWd structure governs the phonetics of the lexical-word-final consonant. For example, the verb-final /t/ in *escort us* is released directly into the following vowel-initial direct object pronoun, just as the /t/ in *escorting* is released directly into the following vowel-initial affix, with a noticeable amount of release noise which gives the perceptual impression of a syllable-initial stop. This behavior is consistent with the structural intuitions reflected in comments such as

> "For example, in *Peter doesn't understand it* the syllabification of the phrase *understand it* does not respect lexical boundaries, that is, it is not *un-der-stand-it*. Rather it becomes *un-der-stan-dit*, where the last syllable, *-dit-*, straddles the lexical word boundary between *understand* and *it*." (LRM99 p. 20)

and

> "…*understander*, which the speaker will unhesitantly syllabify as *un-der-stan-der*." (LMR99 p. 20)

Booij and Lieber (1993) report a similar intuition for Dutch; they note that for Dutch examples like *Komt hij* (*Komt ie*, Engl transliteration *Comes he*?),

> "That *ie* forms one prosodic word with the preceding word is clear from the syllabification patterns (kom) (tie)…" (p. 37)

The description of the phonetic behavior of final /t/ in these contexts in at least some variants of Southern British English and Dutch initially appears to support a Prosody Last view. In this view, only one prosodic constituent, the PWd, is required to account for constituent-position-driven phonetic variation, so that computation of the rest of the prosodic constituent hierarchy can be postponed to a later point in the planning process. Syllable organization within the PWd accounts for aspects of phonetic variation that are governed by syllable structure and also for variation that is governed by adjacent segments within the syllable, and in this model these two kinds of variation are the primary phonetic concern.

There are of course a number of other factors that govern surface phonetic variation – how do they fare in this model? One such source of systematic variation is prosody, e.g. the patterns of f0, relative timing etc. that are governed by prosodic specifications of prominence and constituent structure, and another source is gestural overlap between successive syllables and words. In earlier proposals, Levelt (1989) emphasized the role of a later component called the Prosody Generator, which computes higher-level prosodic structure after Phonological Encoding has generated the gestural scores of syllables. This module was motivated by the need to compute intonational and durational patterns for the utterance, but it could also be invoked to control gestural overlap via the specification of relative timing of gestures. This might account for many cross-PWd-boundary context effects, such as the acoustic labialization of final /t/ in *cut baloney*, or the palatalization of final /s/ in *gas shortage*. It might also be able to account for the hierarchical aspect of constituent-initial strengthing and constituent-final lengthening (described by Keating et al. 2003, among others), by building a hierarchy of higher-level prosodic structures to control the timing of events within and between the already-selected syllabic gestural scores. Thus, on many fronts, the Prosody Last approach to phonological and phonetic encoding provides, or could be expanded to provide, a reasonable account of phonetic variation.

Where this approach may run into some difficulty is in accounting for phonological adjustments to context that on the one hand require look-ahead to higher-level structures, and on the other hand have phonetic consequences that are not amenable to description in terms of greater or lesser gestural overlap, or other changes in the relative timing of articulatory events between PWds. For example, Keating and Shattuck-Hufnagel (2002) point out that if higher-level prosodic structure above the level of the PWd is computed late in the utterance planning process, i.e. after Phonological Encoding via selection of syllable-sized gestural scores is complete, there might be some difficulty accounting for certain types of interactions *between* PWds. These include the sensitivity of accent placement within the word to intonational phrasing (e.g. early accent in *JApanese FOOD is important* but not in *For the JapanESE, FOOD is important*, Shattuck-Hufnagel et al. 1996), or the occurrence of cross-PWd sublexical errors (e.g. *Vile Nalley* for *Nile Valley*, *sale and pickly* for *pale and sickly*, and *Rount Mushmore* for *Mount Rushmore*).

Such arguments highlight the fact that there is clearly more to be learned about the nature of the representations that are formed during Phonological Encoding, as well as about the mechanisms by which surface phonetic variation occurs. One aspect that needs to be addressed is the possibility of differences in encoding mechanisms across languages, or even across different varieties of the same language. This is particularly important when the phonetic phenomena that provide the strongest support for a particular model differ from one language community to another. For example, the description of onset syllabification of /t/ in both *escorting* and *escort us* in Southern British English, which provides the key evidence for LRM99's proposal for phonological encoding into PWd constituents (and PWd constituents only), does not accord with the behavior or intuitions of speakers of American English. It is true that many speakers of this variety of English are described as producing similar phonetic implementations in these two contexts, but it is a flapped /t/ (Hayes 1989), rather than a /t/ produced with closure silence followed by release noise.

In accord with this behavior, and in contrast to speakers of Dutch, American English speakers do not have strong intuitions about the syllable affiliation of e.g. the /t/ in *escorting* or the /d/ in *understanding*. However, there is evidence that, whatever the direction of affiliation of these verb-final stops before an inflectional affix, speakers of American English may treat them differently from the way they treat a final stop before a pronoun, as in *escort us*. Hayes (1989) notes that, although both the affix and the pronoun context are appropriate for flapping in American English, they are not treated identically. Instead, a small amount of release noise is acceptable in the flap before an affix (e.g. in *visiting*) but not before a pronoun (e.g. in *visit it*). Hayes proposes that this phonetic difference reflects a difference in prosodic constituent structure: the <verb+affix> sequence is grouped into a single PWd, [*visiting*]PWd, while the <verb+pronoun> sequence is separated into two PWds, [*visit*]PWd [*it*]PWd, which in turn form a single higher level constituent, the Clitic Group. Because syllabification occurs within the PWd, only the /t/ in *visiting* can exhibit any of the release noise characteristic of syllable-initial voiceless stops; the same /t/ in *visit it* must remain syllable-final, because *visit it* is structured prosodically into two separate PWds combined into a Clitic Group.

If this intuition about the difference in phonetic behavior of final /t/ before an affix vs. a pronoun in American English is correct, it is consistent with the hypothesis that more than one level of prosodic structure is available when the surface phonetic shape of the /t/ is determined, and thus with the Prosody First approach to the planning of spoken utterances, at least for American English. More immediately, if speakers show evidence of combining <verb + affix> sequences into one kind of prosodic constituent and <verb + direct object pronoun> sequences into another, then models of phonological encoding must provide a way of accounting for this difference between varieties of English. The experiments described below were designed to begin the exploration of this issue by providing a quantitative measure of the phonetic behavior of verb-final /t/ in these two contexts for American English. A preliminary study, with 9 /t/-final verbs (such as *edit, limit, remit* and *create,* produced before the affixes -*ed* and -*ing* and the direct object pronouns *us* and *it*), and 5 speakers, generated somewhat equivocal results. Three of the speakers treated the two contexts differently: they produced a glottalized /t/ (consistent with final position) more often before the pronoun, and produced an oral closure released directly into the vowel (consistent with initial position) more often for the affix. The other two speakers usually flapped the /t/ in both contexts. Two further experiments were designed to examine this phenomenon in more detail.

2. Experiment 1

The preliminary study described above provided some evidence for a phonetic difference between pre-affix and pre-pronoun final /t/, and where there was a difference, it was consistent with the hypothesis that /t/ is produced with more syllable-initial characteristics before an affix than before a pronoun. However, many tokens were produced as flaps, so that they fit the description of neither a fully initial nor a fully final /t/. Moreover, there appeared to be differences both within and between speakers in the phonetic treatment of this segment. To explore this phenomenon more thoroughly, a larger set of verbs was used to elicit utterances in both the affix and pronoun contexts.

2.1. Method

Stimuli: A set of 30 /t/-final verbs were embedded in six frame sentences: two frames included inflectional affixes (*Please say editing again, Please say edited again*), two had direct object pronouns (*Please say edit us again, please say edit it again)* and two had an adverb directly following the verb (*Please say edit again*). Ten of the verbs were trochees (e.g. *edit, credit, limit*), ten were iambs (*complete, remit, create*) and ten were monosyllables (*seat, meet, grate*). The combination of 30 verbs and 6 frame sentences resulted in a set of 180 stimuli.

Subjects: Three female undergraduate students served as speakers. They reported no history of hearing or speech problems, and were paid $10 for their participation.

Elicitation: The stimuli were presented in the form of typed scripts. Speakers were recorded individually while seated in a sound-attenuated room. The speech was recorded on audio cassettes and later digitized at a 10K sampling rate.

Analysis: Waveform and spectrogram displays provided by xwaves and xkl software were used to analyse the digitized utterances. Tokens were sorted into three categories, separating flaps from both phonetically syllable-initial and phonetically syllable-final tokens. Figures 1–3 illustrate typical tokens from each class. For illustrative purposes, Figure 1 shows a verb-final /t/ released as a syllable-initial /t/ in the word *planted* (not by one of the speakers in Experiment 1); the /t/ occurs at approximately 19.15 to 19.25 seconds. Note the noisy release, consisting of frication (with energy present across a wide portion of the higher frequencies in the spectrum) just after 19.2 seconds, followed by aspiration noise showing formant structure similar to that of the following vowel which begins at about 19.25 seconds. Such tokens, with a closure period resulting in near-silence and a noisy release directly into the vowel and without an intervening post-release silence, glottalization or glottalized vocoid plus silence before the vowel onset, are consistent with buildup of interoral pressure behind an oral closure and release directly into the vowel. They were categorized as initial /t/s, and taken as evidence for the integration of *plant-* and *-ed* into a single word-like constituent.

/ p l æ n t ə d /

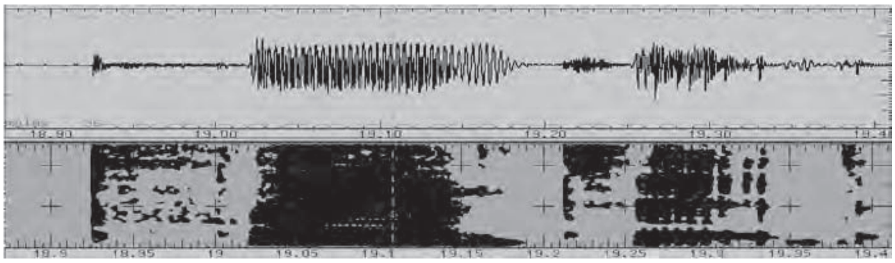


Figure 1. A verb-final /t/ produced in the word *planted*, with a closure silence and release noise consistent with full oral closure and release directly into the following vowel.

In contrast, tokens with a clear acoustic separation between the two morphemes, such as the silence between glottal closure for the /t/ followed by silence and glottal release into the vowel shown in Figure 2, were classified as final /t/s. The glottal closure for /t/ in this token occurs at approximately 11.45 seconds; oral closure may have also occurred but there is no evidence in the acoustic signal for release of pressure buildup behind such a closure, in the form of turbulence noise at its release. The spectrum of the glottal release into the vowel does not have the higher-frequency concentration that would be expected for a tongue-tip release, and the two kinds of release events are easy to distinguish perceptually.

/ b l a t # ɪ t /

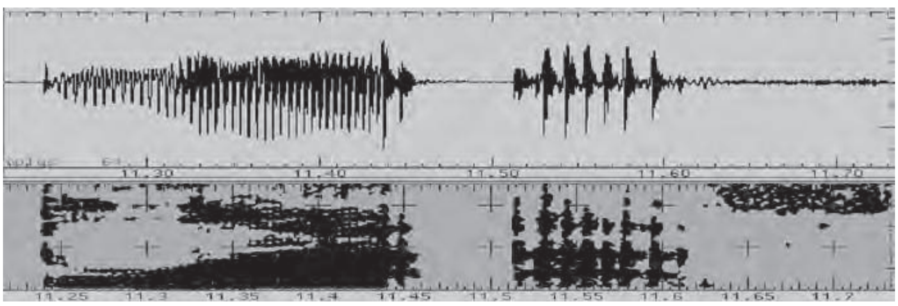


Figure 2. A verb-final /t/ produced in the phrase *blot it,* with the irregular pitch periods that are characteristic of a phonetically syllable-final /t/, followed by a silence, followed by a glottal onset of the vowel.

Such tokens were categorized as final /t/s, and taken as evidence that the verb was not combined with the following morpheme into a single word-sized prosodic constituent.

In addition to tokens with a noisy release of the /t/ directly into the vowel (classed as initial) and tokens with an acoustic separation between the /t/ and the vowel (classed as final), there was a third type of token, showing the characteristics of a flapped /t/. The criteria for a flapped final /t/ included a) perceptual judgment of a flap rather than an /t/, and b) continuation of voicing throughout the low-amplitude region (if there was one); in some cases there was also an extremely short release noise riding on a single pitch period, but there was no discernable period of closure silence. (The range of acoustic shapes for the tokens categorized as flaps is illustrated in the Appendix. The direction of affiliation of these tokens was not determined.) Judgments for each token were made by two experienced phoneticians, one of whom was the author, and disagreements were resolved via discussion.

2.2. Results

For two of the three subjects there was a clear distinction between the phonetic treatment of final /t/ before an inflectional affix vs. before a direct object pronoun, as can be seen in Tables 1 and 2.

Table 1. Phonetic treatment of final /t/ by Speaker 1

Type of /t/ produced	V_again	V_pronoun_again	V_affix_again
Fric/asp	0	0	15
Other /t/	63	58	0
Flap /t/	1	3	45

Table 2. Phonetic treatment of final /t/ by Speaker 2

Type of /t/ produced	V_again	V_pronoun_again	V_affix_again
Fric/asp	5	0	35
Other /t/	55	60	0
Flap /t/	0	0	25

Before pronouns (*us, it*), these speakers produced almost no /t/s with the earmarks of syllable-initial position. Instead they often produced a glottalized final /t/, or released the /t/ into an intervening silence (presumably with a glottal closure after the release) or a glottal squeak (i.e. a lens-shaped region of very high frequency in the wave form), separating the verb from the pronoun acoustically. This suggests that these speakers did not integrate the <verb+pronoun> sequence into a constituent which permits syllabification across the morphosyntactic boundary. However, before affixes (*-ed, -ing*) they produced a noticeable number of /t/s with the kind of fricated/aspirated release noise characteristic of prevocalic position. This suggests that they could integrate the <verb+affix> sequence into a single constituent which permits syllabification across the morpheme boundary. For both speakers, the phonetic behavior of verb-final /t/ in <verb+pronoun> sequences was similar to that for <verb+*again*> sequences. Since most theories would ascribe separate PWd status to *again*, the similarity of phonetic treatment of /t/ before *again* and before a pronoun is consistent with a model in which the direct object pronoun in <verb+pronoun> sequences is a separate PWd as well.

The third speaker showed a different pattern, seen in Table 3.

Table 3. Phonetic treatment of verb-final /t/ by Speaker 3

Type of /t/ produced	V_again	V_pronoun_again	V_affix_again
Fric/asp	0	17	14
Other /t/	60	2	0
Flap /t/	0	41	46

This speaker treated the <verb+pronoun> and the <verb+affix> sequences differently from the <verb+*again*> sequence. All 60 of the <verb+*again*> sequences were produced in a way that was consistent with syllable-final position, i.e. with a glottal closure and/or an acoustic separation between the two target words, but the <verb+pronoun> and <verb+affix> sequences showed almost no evidence for final position for the /t/. Moreover, the pronouns and affixes showed approximately the same number of tokens with the syllable-initial signature of full closure and noisy release (17 and 14). This pattern is consistent with the claim that both of these morphosyntactic sequences form a prosodic structure which allows syllabification of the verb-final /t/ into syllable-initial position at least some of the time.

A similar picture across emerges from analysis of the flapped tokens, also shown in Tables 1, 2 and 3. Speakers 1 and 2 again treat the <verb+adverb>

and <verb+pronoun> sequences similarly, flapping almost none of them, while flapping a substantial number of the <verb+affix> sequences. In contrast, as before, Speaker 3 treats the <verb+pronoun> and <verb+affix> sequences similarly, flapping both at similar rates (41 and 46 tokens), and treats the <verb+adverb> sequences differently, flapping none of them. Thus the distribution of flaps, like the distribution of initial-like and final-like /t/s, is consistent with the claim that Speaker 3 grouped <verb+pronoun> and <verb+affix> sequences into a similar type of prosodic constituent, while speakers 1 and 2 grouped <verb+pronoun> and <verb+adverb> sequences similarly.

2.3. Discussion

The phonetic shapes produced for verb-final /t/ in this experiment clearly show that speakers of American English do not invariably syllabify a verb-final /t/ into the onset of a following vowel-initial pronoun, as has been described for some dialects of British English. Instead, for at least some speakers, the pronoun forms a separate word-like constituent, just as the adverb does, and thus resists syllabification of the /t/ into its onset. In contrast, a verb and its inflectional affix *can* be integrated into a single word-like constituent, facilitating syllabification of the verb-final consonant into the syllable-onset position of the affix. While one speaker failed to distinguish affixes from pronouns in this way, all three speakers produced tokens with flapped /t/ (rather than with full closure followed by a noisy release into the vowel) either for the affix or for both the affix and the pronoun condition. What are the implications of flapped /t/ for the underlying prosodic constituent structure?

In an attempt to determine whether speakers produced flaps differently before affixes vs. before pronouns, we classified the flaps into two groups: those with a discernable noise burst riding on one pulse of the continuous voicing, and those without. Results showed a familiar pattern for the three speakers. Speakers 1 and 2 produced some flaps with noise before affixes, but essentially no flaps with noise before pronouns or adverbs. In other words, they behaved as if they could at least partially syllabify the verb-final /t/ into the following onset position in an affix, but not in a pronoun or adverb. In contrast, as for the other measures, Speaker 3 treated pronouns and affixes similarly (but differently from the adverb). That is, she produced some flaps with a noise burst for both pronouns and affixes (i.e. 27 tokens for each of these conditions), but none for the adverb condition.

Like the results of the pilot study, Experiment 1 suggests that, although it is possible to release a verb-final /t/ directly into the vowel of a following direct-object pronoun in American English, this /t/ does not usually show the full closure, pressure buildup and release noise characteristic of initial voiceless stops. Thus, there is only limited support for the notion that final stops can be syllabified into the onset of a following reduced pronoun like *us* and *it*, as has been reported more generally for British English. Hayes' (1989) suggestion that a degree of release noise is possible for the inflected forms but not for the pronoun forms (e.g. for *visited* but not for *visit it*) finds some support in these data: one speaker produced a /t/ with syllable-onset characteristics before more than half of her affixes (but did not do this before pronouns), and another speaker did this before a quarter of her affixes (but no pronouns). (Hayes was speaking of a variation in the amount of release noise; in this study we report a different parameter: its likelihood.) However, we note that the results for these three speakers do not provide evidence that <verb+affix>, <verb+pronoun> and <verb+adverb> sequences form prosodic constituents at three different levels in the prosodic hierarchy. Speakers treated the <verb+pronoun> sequence either like the <verb+adverb> sequence, i.e. presumably as two separate PWds (Speakers 1 and 2), or like the <verb+affix> sequence, presumably like a single PWd (Speaker 3). This provides evidence for only two different mid-level constituent organizations, which could correspond to a) a single PWd, and b) two separate PWds. It may be that in other speaking conditions (such as conversational speech) or for other speakers, evidence will be found for a difference in the phonetics of final /t/ that implicates an intervening level of prosodic constituent like the Clitic Group.

One aspect of the data from Experiment 1 does support an intervening level of prosodic organization for <verb+pronoun sequences>. In a post hoc analysis we compared the three kinds of /t/ configurations that are consistent with a syllable-final position: (1) /t/ release into an intervening silence, (2) /t/ release into a glottalized region (i.e. into either the glottalized onset of the following vowel or a small glottalized pseudo-vowel followed by silence before the true vowel onset) and (3) no /t/ release after glottal closure. (All of these tokens are included in the 'other /t/' row in Tables 1, 2 and 3.) For Speakers 1 and 3, these variants were produced at similar rates among the three types of morphosyntactic sequences, but Speaker 2 showed a different pattern. For the adverb case, she produced about the same number of /t/ releases into silence and into a glottalized region: 25 and 24. For the pronoun condition, the results were very different: she produced only half as many releases into silence (12), but twice as many releases into glottalization (47). And, she produced no releases of either type for the affix condition, where she produced only flaps and initial stops. Thus for this speaker there is some evidence for a 3-way phonetic distinction, which treats verb-final /t/ differently before adverbs, pronouns and affixes. This is consistent with Hayes (1989) proposal that pronouns form a constituent with a preceding verb (a Clitic Group), and that this constituent is different both from the constituent formed with an inflectional affix (a PWd) and from the structure that is formed with an adverb (i.e. two separate PWds). However, the phonetic distinction between affixes and pronouns for this speaker does not involve the presence vs. absence of a noisy release, which would distinguish initial from final position for the /t/. Instead, it involves release of the /t/ into silence vs. release into a glottalized vocalic region, both of which are consistent with final position.

In Experiment 1, speakers did not reliably produce the verb-final /t/ as an initial stop (i.e. with closure silence and a noisy release directly into the vowel) in any condition, although they sometimes did this in tokens with following affixes (such as *editing* and *edited*). Thus it would be of interest to find a context in which the noisily-released form of these verb-final consonants occurs more freely, in order to test further whether inflectional affixes are grouped differently with their preceding verb than pronouns are. One such context, i.e. where the element that follows the verb base is contrastively pitch accented, formed the basis for Experiment 2.

3. Experiment 2

The purpose of this experiment was to test further the hypothesis that in American English a verb-final consonant is reliably syllabified into the empty onset position of a following affix, but not into the onset of a following pronoun. As in Experiment 1, we are looking for clues to the way these morphosyntactic structures are grouped into the prosodic structures that govern their surface phonetic form. In this experiment we asked the speakers to produce the post-verbal affix or pronoun with contrastive pitch accent, as in *Please say editING, don't say edit US.* This prominence pattern was selected because informal observation suggested that it permitted a strongly-released /t/ in at least some cases.

3.1. Method

Stimuli: Twelve sentences were constructed using six /t/-final verbs similar to the ones in Experiment 1: *edit* (2x), *limit*, *comfort*, *meet*, *grate*. The sentences took the contrastive form shown in the examples *Please say editED, don't say edit US*, and *Please say edit IT, don't say editING*. Each verb was included in two stimuli, one with an inflected form in the first phrase and a direct object pronoun form in the second phrase, and one with the reverse order, for a total of 12 sentences (6 verbs, two orders) and 24 target word utterances (2 per sentence).

Subjects: The subjects were the same 5 undergraduate speakers recorded for the pilot experiment.

Elicitation: Elicitation methods were the same as for Experiment 1, with the addition that for each stimulus sentence, the syllables which were to receive contrastive pitch accents were typed in capital letters. Care was taken not to capitalize the verb-final /t/ (e.g. subjects saw *editING,* not *ediTING,* just as they saw *edit IT,* but not *ediT IT*), to avoid biasing the speakers toward treating this consonant differently in the two conditions.

Analysis: Analyses were as in Experiment 1.

3.2. Results

The five speakers used different strategies, but all five treated the inflections and the direct object pronouns in phonetically different ways.

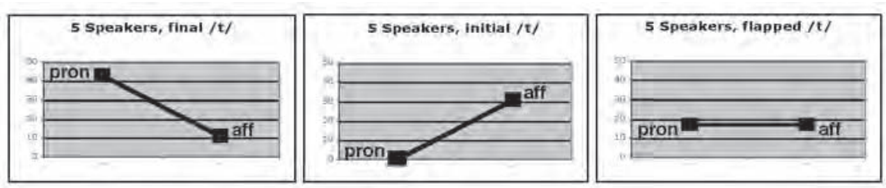


Figure 3. Comparison between types of verb-final /t/ releases in pronoun vs. affix contexts for the 5 speakers combined. The left graph shows the number of phonetically final /t/s (glottalized closure or oral closure with release into silence or a short pseudo-vowel followed by silence); the middle graph shows the number of phonetically initial /t/s (full closure with noisy release directly into the vowel); and the right graph shows the number of flaps. Pronouns are distinguished from affixes in the phonetic treatment of /t/.

Overall, final /t/s were produced more often for pronouns, initial /t/s more often for affixes, and flaps equally often for both morphosyntactic conditions, as shown in Figure 3.

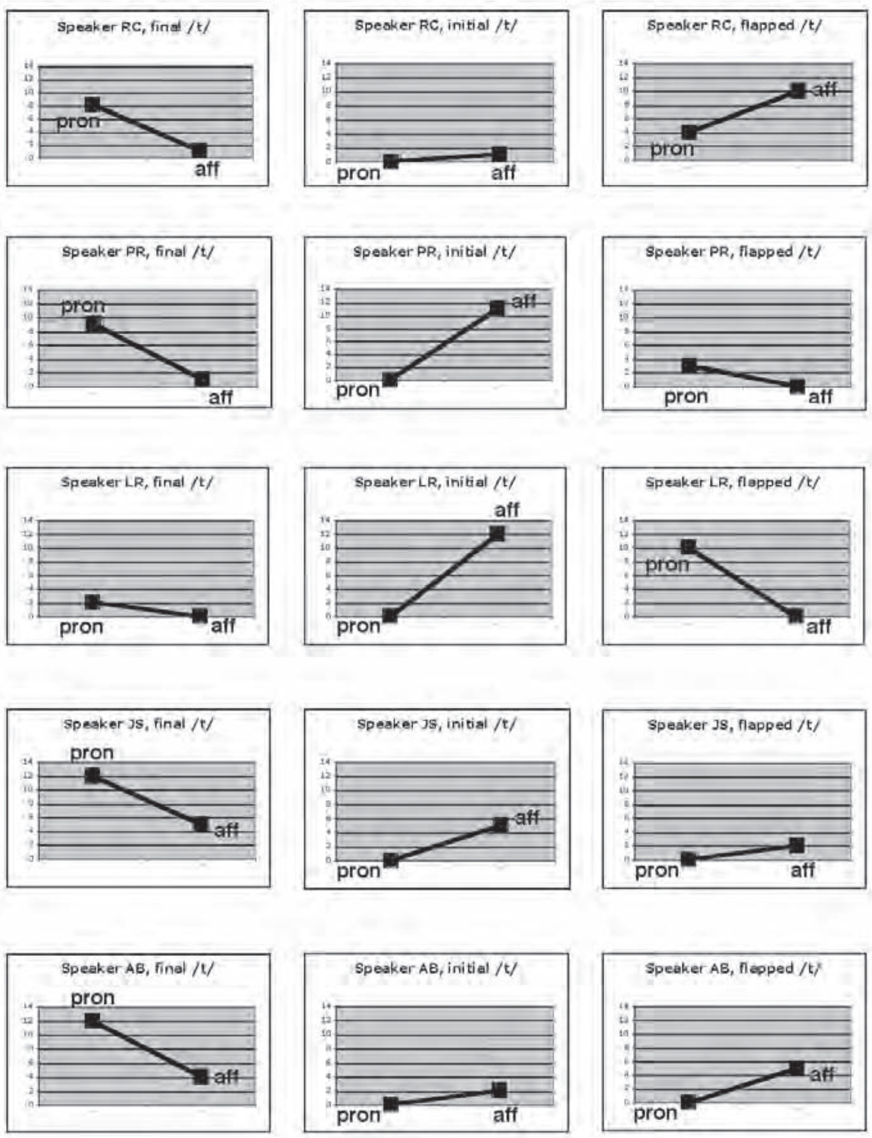


Figure 4. Comparison between types of verb-final /t/ releases in pronoun vs. affix contexts for 5 speakers (see Figure 3 for details.)

Comparison of the individual speakers (Figure 4) reveals that different speakers accomplished this in different ways. For example, Speaker RC usually treated the verb-final /t/ before a pronoun as a final /t/, i.e. released it into a silence or a glottalization, or showed only a glottal closure. But before an affix she was more likely to produce a flap. Speaker AB showed a similar pattern. Speaker PR showed a similar pattern for pronouns, i.e. treated the /t/ as final, but before affixes she did not flap; instead, she tended to produce an initial /t/ in the affix context, with full closure, pressure buildup and a noisy release directly into the vowel. Speaker LR showed the same pattern as PR for affixes, i.e. an initial-like release, but showed a different pattern for pronouns, where she preferred to flap. Finally, Speaker JS treated the /t/ as final before a pronoun, and sometimes before an affix, but she produced initial /t/s and flaps only before affixes. Another way of describing these individual results is that all speakers produced final /t/ more often before pronouns, and initial /t/ more often before affixes, but this distinction was not absolute. For example, Speaker LR showed little preference for final /t/ in either context, and Speakers AB and RC showed little preference for initial /t/ in either context. Moreover, flapping behavior was quite different from speaker to speaker. For example, Speaker LR preferred to flap before pronouns and speaker RC before affixes, while Speakers JS and PR showed little interest in flapping anywhere.

3.3. Discussion

While these data are sparse (only 12 productions in each context per speaker), they illustrate the point that individual speakers can distinguish verb-final /t/ before affixes from /t/ before pronouns, but they can do so in idiosyncratic ways. While the data summaries for the five speakers show an equal number of flaps for affixes and pronouns, data from individual speakers hint at an asymmetry in flapping between the two contexts, with a different pattern of preference from one speaker to another. More comprehensive analysis of the speaking habits of individual speakers will show whether such differences persist, and whether speakers additionally vary in the way they treat final /t/ before a vowel-initial morpheme from one occasion to another, and in different contexts.

As in the earlier experiment, however, these speakers of American English show a striking difference from the behavior reported for speakers of Southern British English. In general these five speakers treat final /t/ differently before affixes vs. pronouns, and they do so in a way that is consistent with a closer structural grouping between a verb and its affix than between a verb and its direct object pronoun. That is, they produce more /t/s with initial-position phonetic characteristics before affixes, and more /t/s with final-position characteristics before pronouns. This is consistent with syllabification of verb-final /t/ into the onset position of a following affix but not into the onset position of a following pronoun (at least when the post-verbal element is contrastively accented, as in this experiment).

4. General Discussion

The questions of interest here are whether a final stop can be integrated with a following vowel-initial morpheme in American English, and if so, whether such structural integration is similar for inflectional affixes and pronouns. Such a finding would be consistent with the Prosody Last view that a PWd-sized planning unit can account for aspects of phonetic variation that occur within the syllable. As we have seen, on this view the phonetic syllables selected on the basis of syllabification within the PWd can be adjusted at a later stage in the planning process by varying the parameter settings for the f0, duration and amplitude of each syllable, on the basis of higher-level prosodic structure which is built later. This later adjustment mechanism could perhaps be expanded to account for aspects of variation that depend on higher-level prosodic contexts, potentially including processes like cross-PWd gestural overlap and hierarchical strengthening and lengthening of articulatory gestures at the edges of prosodic constituents.

A crucial piece of evidence for the Prosody Last view is the report of syllable-initial productions of verb-final /t/ across a verb-pronoun boundary in Southern British English. If this occurs, it strongly suggests that syllabification occurs within a potentially multi-word prosodic constituent that can be made up of e.g. a verb its direct object pronoun, and implicates this level of constituent as the prosodic unit of Phonological Encoding. However, the results from these experiments are not consistent with this view for speakers of American English, because these speakers do not uniformly treat final /t/ in the same manner in the verb-pronoun and verb-affix contexts. Moreover, where phonetic differences between the two contexts are observed, they suggest a stronger boundary before the pronoun than before the affix. What are the implications of these findings for models of phonological encoding during the speech production planning process?

One obvious implication is that if Southern British English does syllabify verb-final /t/ into an empty onset position for both following affixes and following direct object pronouns, and American English does not, then an adequate model of encoding must permit different mappings from the morphosyntactic structure to the prosodic structure for different language varieties. On this view, Southern British English groups both <verb+affix> and <verb+pronoun> sequences into the same type of word-level prosodic constituent, and syllabifies the verb-final /t/ into an empty onset slot of the following syllable in both contexts; American English does the same thing for <verb+affix> sequences but not for <verb+pronoun> sequences. Such a cross-language difference is not implausible, nor does it seem particularly difficult to implement in the Prosody Last model. Moreover, this possibility encourages us to separate out several aspects of the Prosody Last view of Phonological Encoding, i.e. the claims that 1) prosodic frames provide the constituent structure for the encoding process, 2) these prosodic frames consist solely of PWds, with higher levels of prosody structure computed (and having their phonetic effects) during a later, post-Phonological-Encoding stage of processing, and 3) the two kinds of morphosyntactic sequences (<verb+affix> and <verb+pronoun>) form the same kind of prosodic constituent. Allowing different language varieties to map a given morphosyntactic structure onto different mid-level prosodic constituents would give the claim in (3) the status of a dialect- or language-specific characteristic, which could accommodate the difference between the results reported here and observations reported for Southern British English. The claim in (2) is not adequately addressed by the results of these experiments, although there are hints that additional constituent levels will be required. The claim in (1) is not challenged by the present results.

However, there is another possible view of the difference between Southern British English and American English in their treatment of verb-final /t/ in different following contexts, and that is that the noisy /t/ releases observed in Southern British English in phrases such as *escort us* are not necessarily indicative of syllable-initial position. Several observations support this possibility. First, it appears anecdotally that speakers of this dialect can produce a noisy released final /t/ in contexts where there is no opportunity to combine the /t/-final word with the following word into a single PWd. An example of such an utterance produced by a native speaker of this dialect is shown in Figure 5a. The speaker has released the final /t/ of *it* into the initial vowel of the adverb *again* with a noticeable burst of frication, despite the presence of a glottalized onset in the following vowel which is not consistent with an onset /t/. An enlarged view of this fricated release is shown in Figure 5b, where the release noise is clearly visible just before the irregular pitch periods of the initial vowel of *again*.

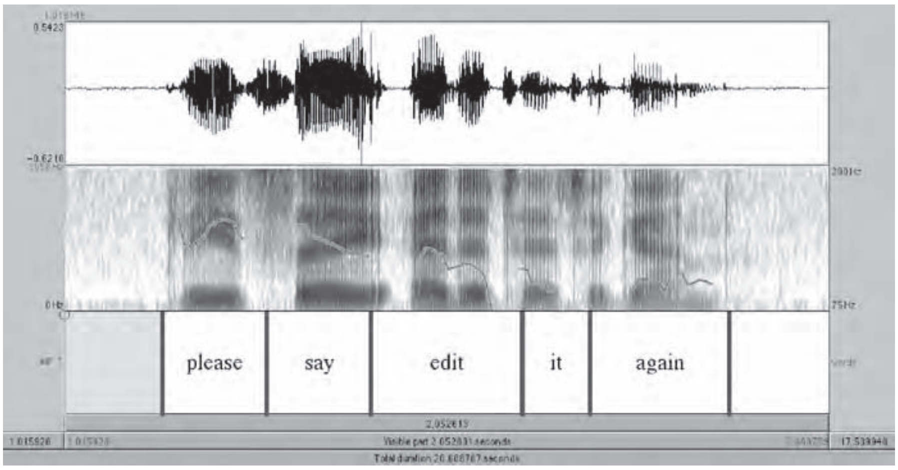


Figure 5a. Example of /t/ in “it again”. Southern British speaker.

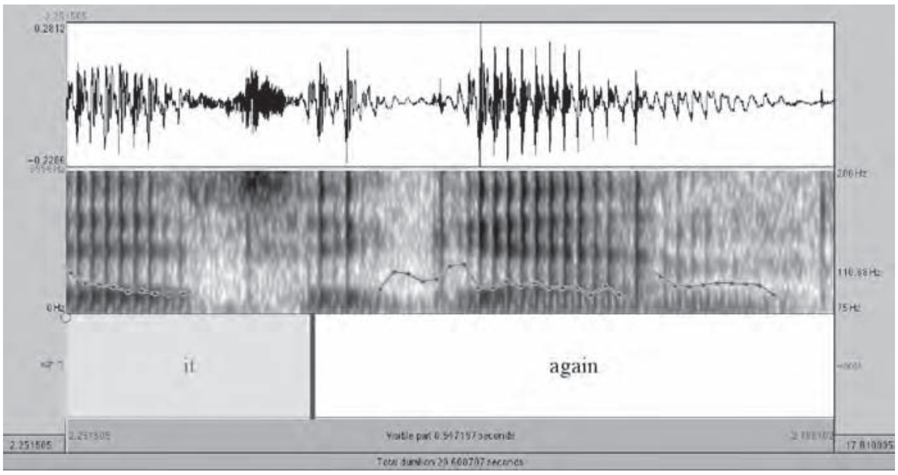


Figure 5b. Example of /t/ in “it again” (zoomed). Southern British speaker.

Another example is shown in Figure 5c, where the final /t/ of *escort* is released into the following PWd *Ed* with a noticeable release noise. Interestingly, at least to native American English ears, this does not sound like *Ted*, despite the presence of the noisy release. Presumably this is because of the lack of the aspiration which would be expected in an initial /t/, as the vocal folds move toward a more approximated position to enable periodic vibration during the vowel. The Prosody Last model does not predict syllabification of the verb-final /t/ across a PWd boundary, so it could not account for this noisy release. But if the noisy release is simply typical of final /t/ in any context in this dialect, it weakens the argument that noisy release indicates a syllable-initial position. This in turn weakens the argument that the final /t/ is syllabified into a following vowel-initial syllable within the PWd, removing an important supporting argument for the Prosody Last view.

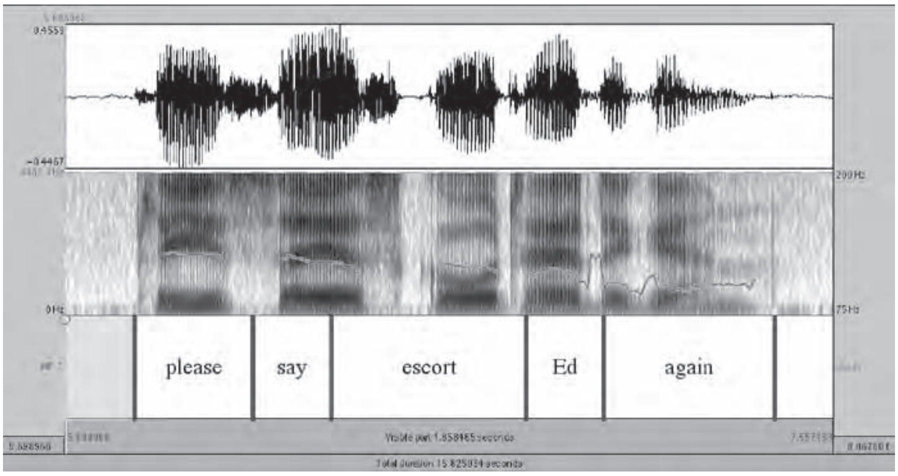


Figure 5c. Example of /t/ in "escort Ed". Southern British speaker.

A second line of argument that calls into question whether a noisy /t/ release is in itself evidence of syllable onset position in Southern British English is that final /t/ can also be released with a burst of frication when there is no following vowel at all, e.g. in utterance-final position. Figure 6 shows an utterance by the same speaker in which the utterance-final /t/ of *edit* shows a recognizably noisy release.

In view of these examples, it is not clear that a noisy release of final /t/ in phrases like *escort us* in Southern British English is a reliable cue to syllabification of the /t/ into onset position. To establish this, we would need to show that it is indistinguishable from a true onset /t/ before similar vowels;

further studies of the acoustics of such stop consonants in continuous speech will be needed to resolve this issue. It would be particularly interesting to determine whether a /t/ with noisy release has both frication and aspiration before a reduced vowel (as in *escort us, escorting*) and before an accented vowel (as in *escort US, escortING*) in the two varieties of English. Thus the questions of whether Phonological Encoding models need to accommodate syllabification across lexical word boundaries in English, and if so, whether this occurs in different morphosyntactic contexts in different varieties of the language (as the result of different morphosyntax-to-prosody mappings) remain open.

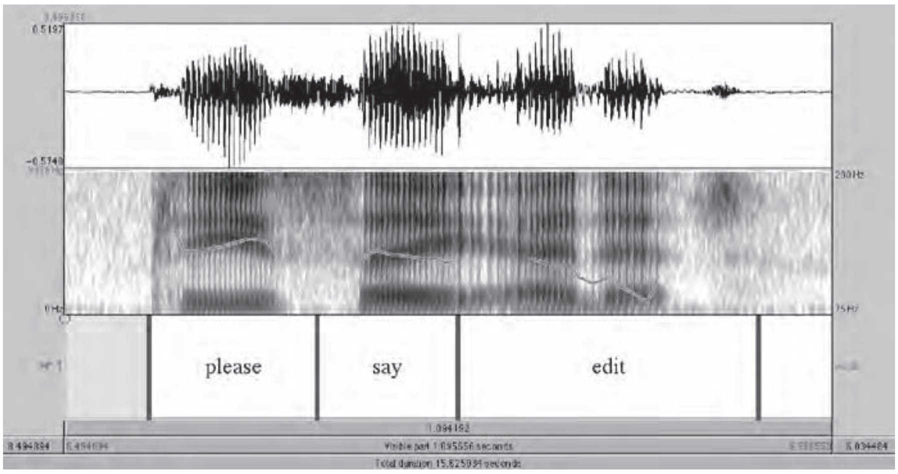


Figure 6. Example of final /t/. Southern British speaker.

Flapping as evidence for syllable position in American English. The study of contexts for flapping has a long and distinguished history, revealing at least two contexts for the effect: lexical-word-medial /t/ can flap before a weak vowel (*butter, city, data*) and lexical-word-final /t/ can flap before both strong and weak vowels (*but I can, but Alicia can*). Anecdotal evidence suggests that even word-onset /t/ can be produced as a flap in some contexts (*I'd like to go tomorrow*), although not easily in others (*buy tomatos),* and never in others (*pay taxes*). It is unclear what role is played by frequency of use and/or rhythmic constituency in these distinctions, or how consistently they can be observed across speakers. Another observation of interest here is that, for some speakers, flapping appears to be acceptable across rather deep prosodic constituent boundaries, such as Intonational Phrase boundaries (as in *It's not.] I already told him that.]).* Thus the positional status of the flapped /t/s produced by the subjects in these experiments remains unclear.

Individual speaker differences. These experiments provide evidence that different speakers can compute the surface phonetic form of the same sentence in different ways. It is possible that some of the differences observed here are due to the way each individual approached the demands of these specific tasks, and do not reflect what they would do in connected speech in a communicative situation. However, like earlier reports (Byrd and Saltzman 2003, Ellis and Hardcastle 2002 and their references), these results suggest the importance of surveying a number of speakers before reaching conclusions about the range of possible surface phonetic variation in a given dialect of a language, as well as the importance of providing for individual speaker behavior in models of speech production planning. A model in which speakers have the option of constructing different surface prosodic structures from the same underlying morphosyntactic structure, and in which those prosodic structures have an important role in determining the surface phonetic shape of an utterance, would provide an account of this behavior.

5. Conclusion

The kinds of systematic phonetic variation that have been observed in these experiments, and in a growing body of earlier and ongoing investigations in many laboratories, suggest that we will see a number of interesting developments over the coming years. The first is more detailed and comprehensive study of the range and nature of this variation, particularly in conversational speech, which will help to distinguish among competing models of the mechanisms by which variation arises in speech production. Another is the incorporation of algorithms for the synthesis of acoustic output into computer-implemented models that are now restricted to earlier parts of the phonological and phonetic encoding stages. The inclusion of this final output stage will permit the evaluation of such models by the most sensitive of instruments for detecting phonetic naturalness and well-formedness: the human perceptual system. Another anticipated development is more widespread consideration of the implications of phonetic variation for production planning models in general, not only for the final stages of such models (where phonetic parameters themselves are computed), but also for earlier stages, which must generate representations of the factors that will govern those phonetic computations. The influence of frequency/predictability on phonetic reduction, which is increasingly well-documented (e.g. Jurafsky et al. 2001) must be incorporated into models of phonological and phonetic encoding as well. As these expected developments unfold, we can expect to see considerable progress in understanding how speakers turn a speech plan into a completed utterance.

Appendix A: Variation in flapped /t/

A striking characteristic of the set of flapped /t/s produced in this experiment was the wide range in their acoustic shapes. They ranged from a stop with abrupt closure, voiced throughout, with pressure buildup and a short release noise resembling that of a /d/, to a nearly indiscernible lessening of amplitude during an otherwise smooth and continuously voiced transition between the two vowels. In between were a variety of tokens in which one could discern a larger or smaller burst-like noise at the end of the voicing in the flap region, riding on one or two pitch periods. Figures A1a and A1b show flaps with a small burst (at 4.530 seconds and 6.125 seconds), and Figure A2 shows one without such a burst.

/ g r e i t ə d /

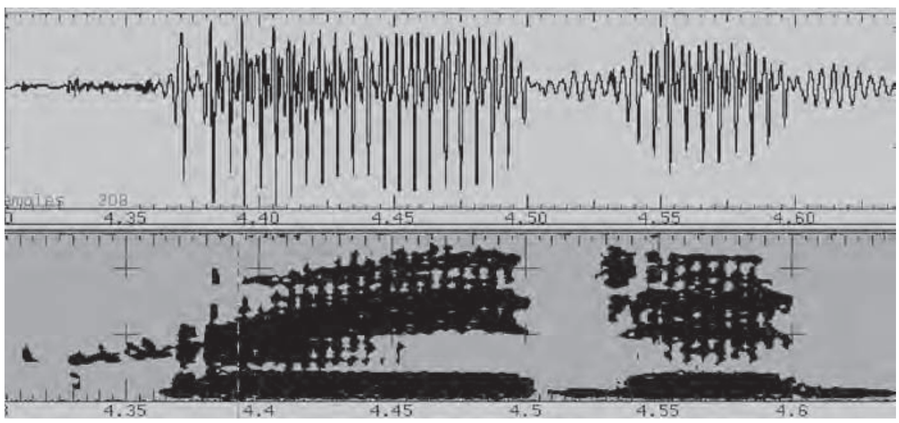


Figure A1a. Example of a flapped /t/ with low-amplitude release noise indicating some buildup of pressure behind the oral constriction.

/ n ɛ t ə d /

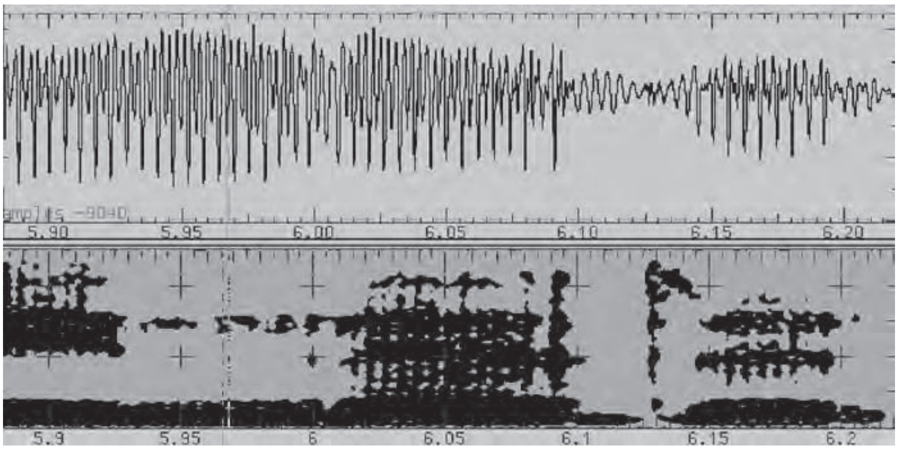


Figure A1b. Another example with a small release noise, more easily seen in the spectrogram than in the wave form.

/ n ɛ t ə d /

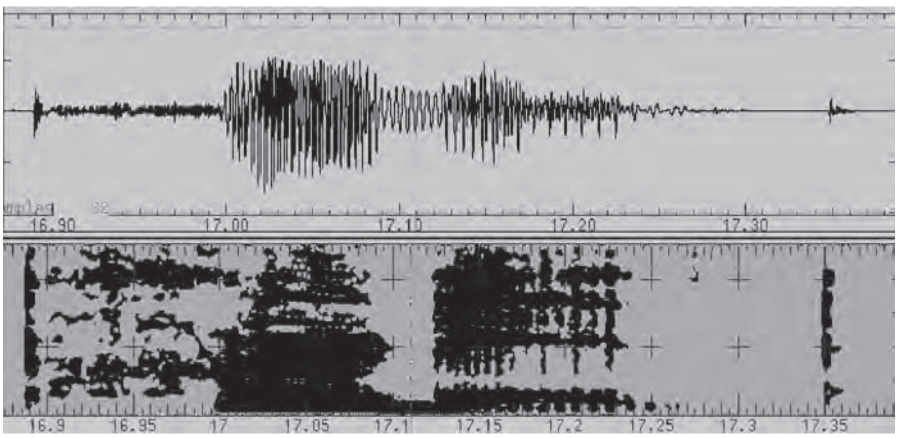


Figure A2. Example of a flap with no acoustic evidence for a release noise burst

There were other tokens which showed a waveform profile remarkably like that of a glide, i.e. the amplitude diminished smoothly to a minimum and then smoothly increased (at about 2.65 seconds in Fig. A3a).

/g r i t ə d ə g ɛ n /

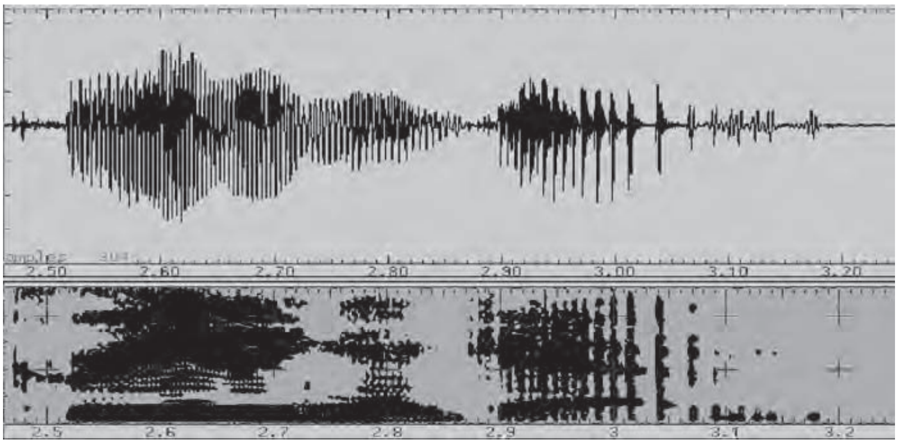


Figure A3a. Example of a ‘flapped’ /t/ realized in a glide-like manner.

An expanded view of this example is shown in Figure A3b.

/(gr) i t ə (d)/

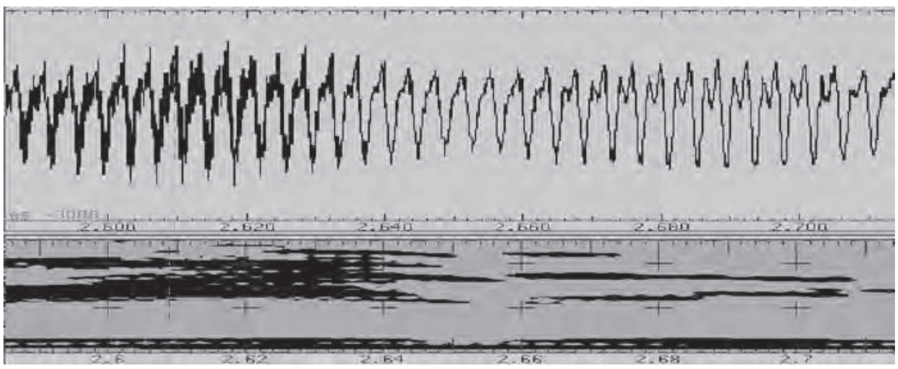


Figure A3b. Expanded Example of a ‘flapped’ /t/ realized in a glide-like manner.

Still other tokens showed an abrupt stop-like closure (in the form of a rapid fall in the amplitude of voicing, albeit to a level that was still audible to the listener and visible in the waveform) with a gradual glide-like release, or the reverse: a gradual glide-like closure with an abrupt increase in amplitude at the following vowel. It was difficult to sort these different types of tokens reliably into separate categories, because the nature of the variation appeared to be continuous, from a fully-voiced stoplike closure, to a less constricted closure which nevertheless permitted some pressure buildup and a visible release noise, to a glidelike constriction without enough pressure buildup to result in a release noise, to a minimal constriction whose acoustic results were sometimes difficult to detect at all. This kind of acoustic variation lends itself particularly well to a description in terms of continuously-varying articulatory parameters, such as constriction degree.

Acknowledgements

* This work was supported by NIH grants DC00075, DC02978 and DC02125, and by the Undergraduate Research Opportunities Program and the Keith North Fund at MIT. Special thanks to Margaret Renwick, who carried out many of the phonetic analyses, to Pat Keating and Alice Turk for extensive background discussions, to Alejna Brugos for timely technical assistance, to Louis Goldstein for invaluable advice, and to Ken Stevens for providing the intellectual environment which made this work possible. Any errors are the responsibility of the author alone.

References

Bell, A., D. Jurafsky, E. Fosler-Lussier, C. Girand, M. Gregory and D. Gildea
2003 Effects of disfluencies, predictability and utterance position on word form variation in English conversation. *Journal of the Acoustical Society of America* 113: 1001–1102.

Browman, C. and L. Goldstein
1992 Articulatory phonology: an overview. *Phonetica* 49: 155–180.

Booij, G. and R. Lieber
1993 On the simultaneity of morphological and prosodic structure. In: S. Hargus and E.M. Kaisse (eds.), *Phonetics and Phonology 4: Studies in Lexical Phonology*, 23–44. Boston: Academic Press.

Byrd, D. and E. Saltzman
2003 The elastic phrase: modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics* 31: 139–278.

Ellis, L. and W. J. Hardcastle
2002 Categorical and gradient properties of assimilation in alveolar to velar sequences: evidence from EPG and EMA data. *Journal of Phonetics* 30: 373–396

Hayes, B.
1989 The prosodic hierarchy in meter. In P. Kiparsky and G. Youmans (eds.), *Phonetics and Phonology 1: Rhythm and Meter*, 201–260. Boston: Academic Press.

Jurafsky, Daniel, Alan Bell, Michelle Gregory and William D. Raymond
2001 Probabilistic Relations between Words: Evidence from Reduction in Lexical Production. In: Joan Bybee and Paul Hopper (eds.), *Frequency and the emergence of linguistic structure.* Amsterdam: John Benjamins.

Keating, P.A. and S. Shattuck-Hufnagel
2002 A prosodic view of word form endocing for speech production. *UCLA Working Papers in Phonetics* 101: 112–156

Keating, P., T. Cho, C. Fougeron and C. Hsu
2003 Domain-initial articulatory strengthening in four languages. In: J. Local, R. Ogden, R. Temple (eds.), *Phonetic Interpretation (Papers in Laboratory Phonology 6)*, 143–161. Cambridge University Press.

Keyser, J. and K. N. Stevens
2006 Enhancement and overlap in the speech chain. *Language.* 82(1): 33–63

Levelt, W. J. M.
1989 *Speaking: From intention to articulation.* Cambridge, MA: MIT Press

Levelt, W. J. M., A. Roelofs and A. S. Meyer
1999 A theory of lexical access in speech production. *Behavioral and Brain Sciences* 22: 1–75

Pierrehumbert, J.
2001 Exemplar dynamics: Word frequency, lenition, and contrast. In Joan Bybee and P. Hopper (eds.), *Frequency effects and the emergence of linguistic structure*, 137–157. Amsterdam: John Benjamins.
2002 Word-specific phonetics. In: C. Gussenhoven and N. Warner (eds.), *Laboratory Phonology VII*, 101–140. Berlin: Mouton de Gruyter.

Stevens, K. N.
2002 Toward a model for lexical access based on acoustic landmarks and distinctive features. *Journal of the Acoustical Society of America* 111: 1872–1891.

Zsiga, E.
1997 Features, gestures and Igbo vowels: An approach to the phonology/phonetics interface. *Language* 73: 227–274.

Focusing, prosodic phrasing, and hiatus resolution in Greek*

Mary Baltazani

I investigate whether hiatus resolution in Greek (a) is a categorical rule, (b) applies within a certain prosodic domain, and (c) what effect focus exerts on this rule. The results suggest that the process is gradient, not categorical. Moreover, deletion of one of the two vowels creating hiatus occurs only for 25% of the data and there is no assimilation between the two vowels at all for 30% of the data. For the remaining 45% of the data there is assimilation in varying degrees. Vowels around strong boundaries and vowels in focused environments show greater resistance to assimilation than those around weak boundaries and non-focused environments. A strong positive correlation was found between duration and the degree of assimilation: Vowels with long duration resist assimilation more than vowels with short duration. There is also a significant influence of the quality of the input vowels both on segmental duration and the degree of assimilation.

1. Introduction

Recent studies have shown the effects of prosodic phrasing on segments. For example, final lengthening before prosodic boundaries is a well known process, and so is the strengthening of segments in phrase-initial positions (Jun 1993; Keating et al. in press; Pierrehumbert and Talkin 1992). Prosodic phrasing has also been shown to affect coarticulation between segments: Cho (2001; 2002; in press) found greater resistance to coarticulation across stronger boundaries than across weaker ones. In addition, focus has been found to affect segments: A focused word shows expanded pitch range and lengthening (Frota 2002 for a typology of focus realization; Jun and Fougeron 2000 and references therein; Jun and Lee 1998).

This is an experimental study which concentrates on the influence of prosodic contexts on *hiatus* resolution in Greek. Hiatus, two immediately adjacent vowels, is often avoided in casual speech. In this paper only *external* (cross-word) hiatus cases will be examined. In (1) [ao] creates hiatus: *megala onomata* surfaces with [a] deleted.

(1) megàla onòmata → megàl onòmata
'big names'

Cross-linguistically hiatus is avoided through different strategies (Casali 1997). In the Greek phonological literature, hiatus resolution has been treated as a categorical rule (Condoravdi 1990; Fallon 1994; Hadzidakis 1905; Kaisse 1977; Nespor and Vogel 1986). Greek researchers agree that external hiatus is optionally resolved in fast speech, usually by deletion of the first vowel (V1). According to these analyses, which are impressionistic, hiatus resolution depends on the quality of the hiatus vowels and on the syntactic relation between the carrier words. However, there is disagreement on how extensive the hiatus resolution process is, what the domain of each proposed rule is, and which vowels are affected in each domain.

The experimental evidence presented in this paper suggests that the output of hiatus is variable and gradient. Cross-linguistically, several sandhi processes have been shown to be gradient. For example, vowel assimilation in Igbo, as analyzed in Zsiga (1997), is a process very similar to the Greek phonological description of vowel deletion in hiatus environments. Zsiga convincingly shows that the phenomenon, described in phonological studies as categorical, is in fact gradient. Gradient assimilation is accounted for within Articulatory Phonology (Browman and Goldstein 1986; Browman and Goldstein 1989; Browman and Goldstein 1992): articulatory gestures for two segments overlap to a greater or lesser extent. In Igbo, this partly depends on prosodic phrasing. Korean Lenis Stop Voicing has also been shown to have gradient output (Jun 1995). The gradient nature of this rule was also accounted for within Articulatory Phonology, by showing the correlation between the duration of lenis stop closure and the degree of voicing; that is, it was shown that the shorter the closure duration of the stop, the more likely it was to be voiced.

Turning to Greek, Pelekanou and Arvaniti (2001) in an acoustic study of several sandhi processes in Greek also examined external hiatus. The deletion of V1 predicted in the phonological literature occurred only in 35% of their data, in 26% there was no deletion at all, and for the remaining 39%, the outputs were gradient, including reduction, coalescence, and diphthongization. In many cases more than one of these outputs were observed for the same pair of words. Pelekanou and Arvaniti conclude that vowel deletion is best viewed as gradient overlap of articulatory gestures which can be perceived as 'deletion' when overlap is complete, and as different degrees of partial overlap when gradient outputs range anywhere between total deletion and no deletion at all. The domain of vowel deletion reported in Pelekanou and Arvaniti is a small prosodic phrase, called the *intermediate phrase* in Greek ToBI (Arvaniti and Baltazani 2000; Arvaniti and Baltazani 2004). The evidence I present supports the conclusion in Pelekanou and Arvaniti and sheds more light on the effects of prosodic structure on hiatus by exploring the importance of prosodic phrasing as well as the effect of focus on hiatus.

The results suggest that, first, vowel deletion in hiatus environments in Greek is less common than reported in the phonological literature, confirming the Pelekanou and Arvaniti results. The most common hiatus output is V1V2 variable assimilation. Second, the extent of V1V2 assimilation correlates with their duration, which in turn is affected mainly by prosodic phrasing and occasionally by focusing: vowels that have lengthened are less prone to assimilate than vowels that have not. That is, the hiatus resolution process is gradient, not categorical. The degree of V1V2 assimilation in hiatus outputs covers the continuum between deletion on one side and total lack of assimilation on the other.

The prosodic organization of Greek laid out in GRToBI is assumed here, that is, a hierarchical system with three prosodic levels, in descending order, the Intonational Phrase (IP), the intermediate phrase (ip), and the Prosodic word (PrWd). A strong prosodic boundary is present across IPs and boundary strength diminishes for the lower levels. V1V2 were placed in sentences with these three levels of prosodic phrasing separating them.

1.1. Prosodic strength hypothesis

Cross-linguistically phrase final lengthening is stronger at higher prosodic boundaries (Cooper and Paccia-Cooper 1980; Edwards, Beckman and Fletcher 1991; Klatt 1975; Wightman et al. 1992). More recently, Cho (2001; in press) in an articulatory study found more resistance to V-to-V coarticulation across strong prosodic boundaries. In accordance with these results, the first hypothesis of the experiment is that in Greek the duration and the quality of the hiatus vowels are affected by the strength of the boundary between them: Stronger boundaries induce greater resistance of vowels to shorten or assimilate.

It is unclear, however, whether assimilation is blocked by boundaries as such or by lengthening regardless of the presence of boundaries. In a study of the acoustic characteristics of Greek vowels, Fourakis, Botinis and Katsaiti (1999) found that focused vowels are longer and the vowel space is more expanded in focus conditions than non-focused ones. In addition, Baltazani and Jun (1999) found that words after a focused item in Greek are de-accented and prosodic boundaries are deleted. These facts combined allow the use of focus-induced lengthening to test whether less assimilation occurs in the absence of strong boundaries. If resistance to assimilation occurs whenever segments are lengthened, then we should find correlation between less assimilation and lengthening due to focus alone without the confounding factor of strong prosodic boundaries. The second hypothesis of the experiment, then, is that V1V2 duration and quality are affected by focus: focused vowels resist shortening or assimilation more than non-focused ones.

2. Method

2.1. Material and speakers

In a production experiment subjects read sentences containing eight different hiatus vowel pairs: [ae], [ao], [oa], [ia], [ua], [eo], [oe], and [ou]. Greek has a five vowel system [a, e, i, o, u] (Joseph and Philippaki-Warburton 1987; Koutsoudas and Koutsoudas 1962; Mackridge 1985). The particular pairs in this experiment were chosen because the words they occurred in sounded natural across all the prosodic contexts. One pair of words for each vowel pair was created (henceforth carrier words), with V1 at the end of the first word and V2 at the beginning of the second. Table 1 shows the eight carrier word pairs with hiatus vowels underlined.

Table 1. The eight carrier word pairs used in the experiment. Each one, containing one of the eight vowel pairs, was embedded in seven different prosodic environments.

1	[ae]	*diávas<u>a é</u>gera* read-pst-1s in time 'I read in time'
2	[ao]	*megál<u>a o</u>rámata* big-pl vision-pl 'A grand vision'

3	[oa]	*diavázo astinomiká* read-pres-1s detective-novels 'I read detective novels'
4	[ia]	*ápiri amerikána* inexperienced American 'Inexperienced American'
5	[ua]	*(I mamá tu) Andrónikou anevéni* (the mother of) Andronikos-gen go-up-3s 'Androniko's mother is going up'
6	[eo]	*katedafísoume olókliro* demolish-3p all 'We will demolish (it) whole'
7	[oe]	*aftokínito érhete* car come-3s '(The) car is coming'
8	[ou]	*nóstimo uzáki* tasty ouzo 'Tasty ouzo'

The carrier words were embedded in seven prosodic contexts. Three tested different strengths of prosodic boundary: Prosodic Word (PrWd), intermediate phrase (ip), and Intonational Phrase (IP). The remaining four contexts tested the effect of focus. Carrier words relative to focus were post-focal (postF), pre-focal (preF), or focused (w1f and w2f). Table 2 gives an example of the contexts for [ou]. Sentence A illustrates PrWd boundary between the words *nostimo uzaki* 'tasty ouzo'. The same two words are separated by ip and IP boundary in B and C respectively. The carrier words are post focal in D and pre-focal in E. Sentence F shows word 1 and G word 2 in focus.

Table 2. The seven prosodic contexts for [ou]. Brackets show prosodic boundary. V1V2 are underlined and focus is shown by small caps.

A	*Tha sou etoimáso éna nóstimo*]$_{Wd}$ *uzáki me mezé.* will for-you prepare-1s one tasty ouzo-dim with tidbits 'I'll prepare tasty ouzo with tidbits'
B	*An éxei mezé nóstimo,*]$_{ip}$ *uzáki pínoun óloi.* If has tidbits tasty ouzo-dim drink-3p all 'If the tidbits are tasty, everyone drinks ouzo'
C	*Tha fáo mezedáki an ínai nóstimo.*]$_{IP}$ *Uzáki den píno.* Will eat-1s tidbits-dim if is tasty ouzo-dim not drink-1s 'I'll have tidbits if they're tasty. I won't drink ouzo'

D *Tha sou etoimaso éna nóstimo uzáki, den boró na su arnithó*
will for-you prepare-1s one tasty ouzo-dim not can to refuse-1s
'I WILL prepare tasty ouzo for you, I can't refuse you'
E *Tha sou etoimáso éna nóstimo uzáki me meze.*
will for-you prepare-1s one tasty ouzo-dim with tidbits
'I'll prepare tasty ouzo WITH TIDBITS for you'
F *Tha sou etoimáso éna nostimo uzáki, óxi san tis Annas*
will for-you prepare-1s one tasty ouzo-dim not like Anna's
'I'll prepare TASTY ouzo for you, not like those that Anna makes'
G *Tha sou etoimáso éna nóstimo uzaki, óxi bíra*
will for-you prepare-1s one tasty ouzo-dim not beer
'I'll prepare tasty OUZO for you, not beer'

Four Athenian Greek speakers, one male and three females ranging in age between 18 and 40, repeated each sentence three times, creating a corpus of 652 tokens (8 V-pairs X 7 Prosodic conditions X 4 speakers X 3 repetitions). Each speaker was recorded separately in a quiet room.

2.2. Measurements

The duration and formant frequencies (F1 and F2) of the vowels in hiatus were measured. *Pitchworks* and *PCquirer* (Scicon) speech analysis programs were used for the measurements.

All duration measurements reported are those of the vowel *pair*: for pairs with a pause between them, the duration of V1 and V2 were added and the silence excluded. F1 and F2 were measured at a stable part near the middle of each vowel, when the two vowels were separate. When there was no break between V1 and V2 two measurements were taken, the first one at a stable part around 1/4 to 1/3 of the way into the vowel pair and the second at a stable part around 1/3 to 1/4 before the end.

The outputs are classified into four categories, shown in Table 3. One-vowel outputs are called *deletion* if the output is in a position (in the F1xF2 vowel space) typical for either vowel and *merger* otherwise. Two-vowel outputs are called *no assimilation* if they occur in their typical positions and *partial assimilation* if either vowel occurs in a non-typical position.

I use the Euclidean distance between the vowels in the F1F2 space[1] to quantify V1V2 assimilation. Smaller distance means more assimilation. These results are reported in section 3.

Table 3. Classification of hiatus outputs according to the hiatus resolution process.

OUTPUT	TYPICAL FORMANTS	NON-TYPICAL FORMANTS
1 vowel	Deletion	Merger
2 vowels	No assimilation	Partial assimilation

3. Results

3.1. Duration

As a reference for the duration of Greek vowels, data from Fourakis, Botinis and Katsaiti (1999) are shown in Table 4. Duration was measured in that study in stressed, unstressed, focused, and unfocused conditions, in fast and slow tempos. Table 4 presents single unstressed vowel durations, because the vowels in my study were unstressed (except V2 in pairs [ae] and [oe]). Moreover, the vowels in the present study were produced in normal tempo, so the average duration between fast and slow will be used for comparison here.

Table 4. Durations (in ms) of unstressed Greek vowels produced in slow and fast tempo from an acoustic study in Fourakis, Botinis and Katsaiti (1999).

Vowel	Slow	Fast	Average
[i]	44	36	40
[e]	57	51	54
[a]	78	69	73.5
[o]	67	58	62.5
[u]	54	40	47

Figure 1 shows durations from the present study in striped bars and in Fourakis, Botinis and Katsaiti in solid bars (vowel pair duration was calculated by adding the averages for single vowels from Table 4). Mean durations are provided under the graph. It is not clear how similar prosodic contexts in the two studies are because no intonational realization details were provided in that paper. Because of this, averages across all prosodic conditions were taken.

The average V1V2 duration in the present study is almost the same as the sum of two single "typical" vowels. Of course, as the error bars show, durations from different prosodic positions within each vowel pair vary considerably: in some prosodic contexts V1V2 duration is as short as a single vowel and in some it is much longer than two single vowels. Table 5 shows the durations for the eight vowel pairs in each of the seven prosodic positions examined. Pooled durations are given in each cell and standard deviations are included in brackets.

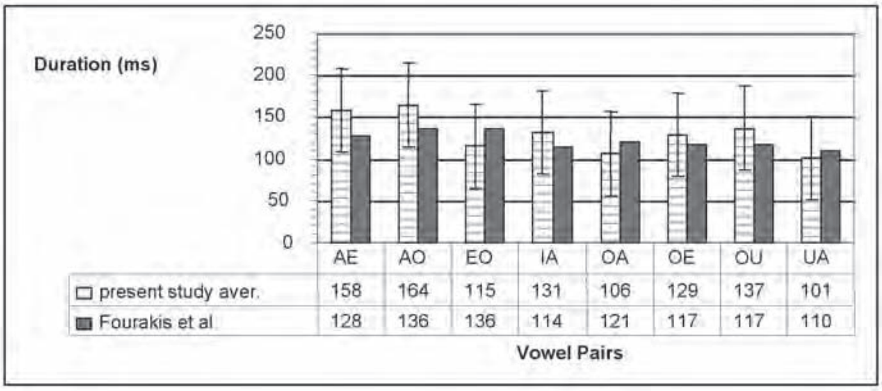


Figure 1. Durations in the present study and in Fourakis, Botinis and Katsaiti (1999).

An analysis of variance tested the effect of prosodic condition (PrWd, ip, IP, postF, preF, w1F, w2F) as well as the type of vowel pair ([ae], [ao], [eo], [oe], [ua], [oa], [ia], [ou]) on duration. Both factors were significant: Prosodic position ($F[1, 6] = 56.553$, $p < .0001$) and Vowel Pair ($F[1, 7] = 14.591$, $p < .0001$). Post-hoc tests show that for all vowel pairs, the longest duration was found for vowels straddling a strong prosodic boundary, either that of an intermediate phrase (ip) or that of an Intonational Phrase (IP): these two prosodic positions, IP and ip, are significantly longer than all the rest. No other significant differences were found for the remaining prosodic positions.

Vowel pairs were significantly longer around strong prosodic boundaries, supporting the first hypothesis. No significant effect of focus on duration was found, refuting the second hypothesis. Vowel pairs in focus positions were longer than those in non-focus positions, but the difference was not significant. Table 5 shows three clusters in the duration values: one with the strong boundary positions (the only one showing significant difference), one comprising w1F and w2F, and a third cluster comprising the rest, all phrase medial positions. This clustering more or less corresponds to the initial experimental setup. It is not clear whether the failure of the lengthening in focus positions to reach significance is due to the small number of participants in this study.

Table 5. Duration for the eight vowel pairs in the seven prosodic positions: Prosodic word boundary (PrWd), intermediate phrase boundary (ip), Intonational Phrase boundary (IP), post focal position (postF), pre-focal position (preF), focus on the first word (w1F), focus on the second word (w2F).

Prosodic Condition	[ae]	[ao]	[eo]	[ia]	[oa]	[oe]	[ou]	[ua]
PrWd	159 (25)	123 (10)	78 (15)	102 (21)	72 (21)	94 (16)	135 (13)	83 (24)
Ip	217 (50)	246 (22)	179 (11)	186 (32)	240 (96)	93 (16)	178 (13)	165 (24)
IP	249 (60)	233 (25)	199 (56)	193 (53)	163 (46)	205 (37)	152 (13)	131 (44)
postF	101 (15)	127 (3)	61 (14)	112 (25)	62 (17)	87 (19)	112 (12)	67 (9)
preF	108 (25)	123 (5)	62 (17)	108 (25)	62 (19)	94 (17)	116 (11)	90 (8)
W1F	124 (43)	143 (33)	138 (61)	102 (23)	69 (2)	107 (27)	122 (4)	72 (8)
W2F	150 (39)	157 (11)	87 (27)	117 (23)	77 (14)	220 (38)	142 (40)	98 (43)

The formant frequency measurements reported in section 3.2 show the influence both of the prosodic context and of the input vowels on V1V2 assimilation.

3.2. Formant frequencies

Section 3.2.1 shows which hiatus resolving strategy (deletion, merger, partial assimilation, and no assimilation) was used for the different vowel pairs and prosodic positions tested; 3.2.2 shows the influence of prosodic context and input vowel quality on the output; finally, 3.2.3 gives more evidence for gradience.

3.2.1. Output classification

As a reference for the formant values of Greek vowels, data from Fourakis, Botinis, and Katsaiti (1999) were used. These vowels were produced within sentences, just like in the present study, so the formant values between the two studies are fairly comparable. An important difference between the two studies is that whereas the Fourakis, Botinis and Katsaiti experiment had only male participants, the current experiment included three females and one male. In Table 6 the male speaker formant values in the present experiment in the IP condition are quite similar to those in the Fourakis, Botinis and Katsaiti study. Those produced by the female speakers in the IP condition are lower and more front than those produced by males, schematically shown in Figure 2.

Table 6. Top row: Average F1, F2 formant values in the Fourakis, Botinis and Katsaiti (1999) study (all male speakers); middle row: the male speaker of the present study; bottom row, the female speakers of the present study.

[i]		[e]		[a]		[o]		[u]	
F1	F2	F1	F2	F1	F2	F1	F2	F1	F2
305	2025	472	1680	680	1315	468	1097	327	1069
366	2098	522	1700	635	1325	487	1077	389	1000
366	2430	620	2070	780	1650	562	1235	487	1100

V1V2 tended to remain in their "typical" positions the most in the strong boundary conditions; formant frequencies in the remaining prosodic positions tended to shift towards one another. Details about assimilation in the different prosodic contexts are given in the following section. For now, the main point is that the formant frequency values in the IP condition will be considered the norm and values in all other prosodic positions will be compared to them. Due to the difference in formant frequencies between males and females, the comparison between the typical values and the centralized ones was made separately for males and females.

Table 7 shows the percentage of each hiatus resolving strategy defined in section 2.2, broken down by vowel pair. Almost half of the outputs show neither deletion nor total lack of assimilation, but intermediate degrees of assimilation.

The quality of input vowels affects the hiatus resolution process used. More specifically, for vowel pair [ia] there are no tokens in the deletion category and very few in the merger category. For [oa] deletion is rare, with merger the most frequent output. For the remaining vowel pairs, deletion occurs around 30% of the times. In the no assimilation column, the average shown in the TOTAL row, 30%, is very near the percentage for each vowel pair. In the partial assimilation column, there are three values that differ from the average: [ia] and [ua] have more outputs than average in this category, while [oa] has extremely few.

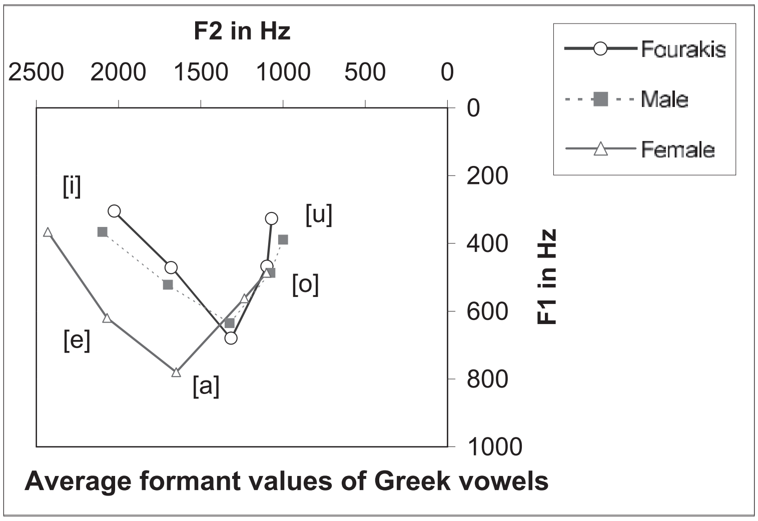


Figure 2. The formant values found in the Fourakis, Botinis and Katsaiti (1999) (all males) and those found for the male and female speakers in the present study.

Table 7. Percentage of the different types of output across input vowel pairs.

Vowel Pair	Deletion	Merger	Partial Assimilation	No Assimilation
[ae]	32	4	29	31
[ao]	39	14	18	29
[oa]	11	57	3	29
[ia]	0	7	61	31
[ua]	25	0	36	39
[eo]	29	21	14	36
[oe]	32	25	21	21
[ou]	39	11	25	25
All 8 pairs	26	17	26	30

Table 8 shows the percentage of outputs in each hiatus resolving category, broken down by prosodic positions.

Table 8. Percentage of the different types of hiatus output in each prosodic position.

Prosodic condition	Deletion	Merger	Partial Assimilation	No Assimilation
PrWd	31	22	41	6
Ip	17	3	6	75
IP	0	0	3	97
postF	41	25	28	3
preF	34	31	31	3
w1F	34	19	34	12
w2F	25	22	37	17
All Conditions	26	17	26	30

The hiatus resolution process is not uniform across prosodic positions, that is, the prosodic position a specific vowel pair is found in influences the hiatus resolution process. More specifically:

- In the no assimilation column, values are widely dispersed from the average: no assimilation is the most common output around strong prosodic boundaries, and a rare output around weak prosodic boundaries, and in unfocused positions. Furthermore, no assimilation occurs in approximately one fifth of the outputs with focus, w1F and w2F, which is considerably more often than in the non-focus, phrase medial positions.
- In the deletion column, the distribution is again uneven. Deletion never occurs in the IP condition and relatively seldom in the ip condition. For the remaining prosodic positions, the distribution of deletion outputs is between 30–40%, with the exception of positions with focus on the second carrier word, w2F.
- As for the resolution processes in the two middle columns, there are low numbers in the ip and IP conditions and for the remaining prosodic positions, the distribution of merger and partial assimilation is fairly even.

Two major findings emerge from the results in this section. First, the claim made in the phonological literature that the most frequent hiatus output in Greek is V1 deletion is not supported. Instead the results here support the findings presented in Pelekanou and Arvaniti (2001). Almost a quarter of the outputs belong in the deletion category (V1 deletion mostly), and a quarter each belong in the merger, partial assimilation, and no assimilation categories.

Second, the degree of V1V2 assimilation is influenced by the quality of these vowels as well as the prosodic position the pair is embedded in.

No unifying generalization can be made regarding which hiatus resolution strategies apply for which vowel pairs. The percentage of no assimilation is fairly constant around 30% for all vowel pairs, but for the remaining three hiatus resolution processes, each vowel pair displays its own pattern of hiatus resolution.

In the phonological literature several vowel quality properties have been claimed to affect deletion as hiatus resolution strategy[2]. The report that high vowels do not delete is verified by the data in the present study only for [i], not for [u], which deletes 25% of the time. The report that a round V1 does not delete in adjective-noun pairs, is not verified by the data, because [o] deletes 39% of the times in the pair [ou], although embedded in such an environment. The report that a V1 higher in the sonority hierarchy [o > a > u > i > e] than V2 does not delete in verb-adverb pairs, is not verified either, because [a] deletes before [e] 32% of the time.

As far as merger (coalescence) is concerned, Casali (1997), in an influential cross-linguistic typological study of hiatus resolution, notes that this strategy is common when V1 is lower than V2 and rare for the reverse. No such tendency was found in the present study. The merger rate for [oa] is high, 57%, although [o] is higher than [a].

The reasons behind the difference in the process of hiatus resolution for each vowel pair remain elusive. A larger scale study may reveal some generalization.

The effect of prosodic contexts on hiatus resolution has not been systematically examined in a quantitative study before. Results show that assimilation is blocked across strong prosodic boundaries, like those at the edge of intermediate phrases (ip) and Intonational Phrases (IP). Moreover, in phrase-medial positions some amount of assimilation is almost always the case. The assimilation-blocking effect of focus on the carrier words is much weaker than that observed for strong boundaries (not statistically significant as we will see in section 3.2.3) but stronger than that for phrase-medial positions. Finally, the rates of deletion, merger, and partial assimilation are approximately equal in every prosodic context except for the strong (IP, ip) boundaries.

In summary, the most robust effect on hiatus resolution is exerted by strong boundaries. A weak effect of focus is only seen for the no assimilation strategy: In contexts with focus on one of the carrier words (i.e., w1f, w2f), more outputs showed total lack of assimilation than in the phrase-medial contexts (i.e., post-F, pre-F, PrWd). In all other hiatus resolution categories (i.e., deletion, merger, and partial assimilation) prosodic positions related to focus are not much different from the phrase-medial prosodic positions.

3.2.2. Assimilation across prosodic positions

In this section I present the V1V2 Euclidean distance in each prosodic position and the correlation between this distance and duration. First, Table 9 shows the V1V2 distance. The numbers on the left column under each vowel pair give the absolute V1V2 distance (pooled across speakers). The numbers on the right column present the 'normalized' distance, which is the distance of the vowel in the given prosodic position expressed as a proportion of the distance in the IP condition in the column (thus 1 in the IP condition). This was done to facilitate comparison across vowel pairs.

Table 9. Euclidean V1V2 distance across vowel pairs in each prosodic condition. The first column provides the distance in absolute values and the second column provides the distance in each prosodic position as a proportion of the distance in condition IP.

	[ae]		[ao]		[oa]		[ia]		[ua]		[eo]		[oe]		[ou]	
PrWd	240	0.47	152	0.3	110	0.3	587	0.59	401	0.52	183	0.5	133	0.25	184	0.74
Ip	498	0.97	477	1	406	1.1	863	0.87	613	0.79	583	1.5	82	0.15	220	0.89
IP	512	1	466	1	355	1	996	1	778	1	389	1	534	1	248	1
postF	291	0.57	123	0.3	76	0.2	600	0.6	267	0.34	178	0.5	107	0.2	158	0.64
preF	217	0.42	131	0.3	117	0.3	511	0.51	432	0.56	171	0.4	191	0.36	166	0.67
w1F	231	0.45	138	0.3	88	0.3	637	0.64	294	0.38	419	1.1	170	0.32	160	0.65
w2F	257	0.5	201	0.4	114	0.3	584	0.59	330	0.42	165	0.4	274	0.51	168	0.68

An analysis of variance tested the effect of prosodic position and the types of vowel pair on Euclidean distance. The within-subject factors considered were Prosodic Position (PrWd, ip, IP, postF, preF, w1F, w2F) and Vowel Pair (ae, ao, eo, oe, ua, oa, ia, ou). Both showed significant effects, Prosodic Position ($F[1, 7] = 25.1214$, $p < .0001$) and Vowel Pair ($F[1, 7] = 40.4224$, $p < .0001$). Post hoc comparisons show that distances across an IP and an ip boundary were indeed significantly greater than those in the remaining prosodic positions. No other prosodic position differences reached significance.

As for the vowel pairs, post hoc comparisons show that distances for the [ia] pair were significantly greater than those in the remaining vowel pairs, those for [ua] the next greatest, and no other significant differences were found.

Table 10 shows the correlation between distance and duration for each vowel pair for each speaker. The values compared were averages across prosodic positions.

Table 10. Correlation between duration and the Euclidean distance for each vowel pair across prosodic positions for the four speakers.

Vowels	Speaker 1	Speaker 2	Speaker 3	Speaker 4
[ae]	0.80	0.89	0.90	0.87
[ao]	0.99	0.80	0.96	0.98
[oa]	0.81	0.94	0.59	0.71
[ia]	0.94	0.86	0.92	0.79
[ua]	0.86	0.76	0.99	0.66
[eo]	0.98	0.89	0.83	0.93
[oe]	0.83	0.47	0.98	0.67
[ou]	–0.13	–0.19	0.80	0.81

For most of the vowel pairs, there is high correlation, over 0.8, between duration and distance. Except for two cases, the correlation is positive, meaning that as duration increases, so does the resistance of a vowel pair to assimilation.

The findings in this section show once more the robust inhibiting effect of strong prosodic boundaries on assimilation. The distance between the two vowels in the strong boundary positions is significantly greater than that in all other prosodic positions. A less robust assimilation inhibiting effect, which didn't reach statistical significance, was found for focus on a word carrying a hiatus vowel. In other words, the hypothesis that focus will block assimilation was not supported. Finally, phrase medial positions, where prosodic boundaries are weak, allowed for the greatest degree of assimilation.

Combining the results of this section with the previous one, it is necessary to stress that although these outputs were classified into four different hiatus resolution categories, these categories do not have sharply defined borders, since the statistical results showed no significant difference for most of the distances. The outputs in the no assimilation category – which in the majority are the outputs in the IP and ip conditions – are the only exception, being significantly different from the rest. Of course there are some 'prototypical' outputs of deletion, no assimilation, and so on, but these constitute the minority. If the four hiatus resolution processes defined in this study are seen as a continuum from no assimilation to deletion, then the measurements showed that outputs are spread all over this continuum and the borders between one category and its neighbor were never clear but had to be artificially imposed upon the data. In other words, this is a gradient process, not a categorical one. Gradience is evident not only across prosodic positions but also within them, as will be shown in the next section, 3.2.3.

The degree of assimilation between the two vowels in a pair also varied as a function of the input vowels. The statistical results show that vowel pairs [ia] and [ua] are different from the remaining pairs, showing the least amount of assimilation and also different from one another, [ia] assimilating less than [ua]. No significant effect of the remaining vowel pairs was found.

Finally, a high positive correlation was found between duration and the Euclidean distance between V1 and V2. It seems that in addition to the effect of boundaries, duration affects the assimilation between the two vowels. Although no statistical support was given to the initial hypothesis that focusing affects the degree of assimilation, the high correlation between segmental duration and assimilation suggests that lengthening in the absence of boundaries does influence the assimilation process.

3.2.3. Assimilation within prosodic positions

Speakers produced three repetitions per prosodic position and vowel pair. For part of the data the degree of assimilation varies across the three repetitions within the same prosodic position.

For example, Table 11 shows formant values for [ia] in two prosodic conditions, IP and postF, produced by speaker 1. The values for each repetition are shown separately. F1 values for [i] and [a] are shown in columns two and three, respectively; F2 values are shown in columns four and five. The last column shows V1V2 Euclidean distance.

For the IP condition, formants in all repetitions are close to typical. For the postF condition, the output in repetition 3 is different from the remaining two: the distance between the two vowels is smaller because [a] has moved to a much higher and more front position than it is typical for it and [i] has moved to a much more back position. Both vowels have moved towards the center but still are distinct from each other: a partial assimilation output. In repetitions 1 and 2, the formant frequencies show that the vowels occupy slightly more centralized positions (especially [a] is around 100 Hz more front) than is typical for them but, according to the classification in this paper, these are no assimilation outputs, qualitatively and quantitatively different from the output in repetition 3.

Table 11. Formant values of [i] and [a] in the vowel pair [ia] in prosodic conditions IP and postF produced by speaker 1. Values in each repetition are shown separately.

Speaker 1 [ia] Condition IP	F1 [i]	F1 [a]	F2 [i]	F2 [a]	Euclidean Distance
Repetition 1	342	799	2328	1656	813
Repetition 2	338	828	2393	1585	945
Repetition 3	417	809	2512	1508	1078
Speaker 1 [ia] Condition postF	F1 [i]	F1 [a]	F2 [i]	F2 [a]	Euclidean Distance
Repetition 1	416	687	2328	1785	607
Repetition 2	401	714	2522	1784	802
Repetition 3	410	563	2112	1807	341

Table 12 shows how much variability of the type shown in the example above occurred within prosodic conditions. In general, 13% of the data display such variability. Table 12A shows variability in prosodic conditions and 12B variability in vowel pairs.

Table 12. The percentage of variability among the repetitions is shown in each prosodic position (A) and in each vowel pair (B).

A.

Prosodic Position	Variability
PrWd	13%
ip	13%
IP	6%
postF	21%
preF	18%
W1F	8%
W2F	11%
AVERAGE across prosodic positions	13%

B.

Vowel Pair	Variability
[ae]	11%
[ao]	11%
[oa]	14%
[ia]	8%
[ua]	12%
[eo]	14%
[oe]	15%
[ou]	14%
AVERAGE across V pairs	13%

The greatest amount of variability occurs in phrase-medial positions, PrWd, postF, and preF. There is less variability in the two focus positions, w1F and w2F, and the least amount of variability is found in strong boundary position IP. Contrary to the general pattern observed so far of ip and IP boundaries influencing segments in a similar way, in this case there is considerably more variability found around intermediate phrase boundaries than around IP boundaries.

As for the variability across the vowel pairs, in Table 12B, only [ia] is obviously different from the rest: it has outputs with the least amount of variability. The remaining vowel pairs cluster around the average of 13%.

For one eighth of the data, the amount of assimilation between the vowels varies across the three repetitions. The greatest amount of variability occurs in phrase-medial positions and the least in strong boundary positions. One noteworthy exception to this generalization is the high degree of variability around intermediate phrase boundaries. In all other respects, as was repeatedly shown throughout this study, the positions around both intermediate and Intonational phrase boundaries induce the same behavior for vowels: they impede assimilation and facilitate lengthening. This is one instance where we see a difference between the two kinds of boundary, and this nicely illustrates the difference in strength between the two otherwise very similar prosodic environments – there is more plasticity around intermediate phrase boundaries than around Intonational Phrase ones.

As for variability across vowels, only for vowel pair [ia] is there a link between variability and the amount of assimilation found: less assimilation correlates with less variability, just as noted for IP boundaries. However, no such pattern was observed for the pair [ua] which showed the second lowest degree of assimilation.

In summary, the results suggest once more that the hiatus resolution process is gradient. I would like to stress that stating that the hiatus resolution process is not categorical does not contradict the fact that prosodic context influences this process: the prosodic position effect is a tendency, not an absolute.

4. General discussion and conclusion

Several conclusions can be drawn from the results of this experiment. First, it is evident that hiatus resolution in Greek cannot be described as a categorical rule: the same speaker, in the same prosodic position, produced different outputs, with varying degrees of assimilation between the two vowels. Even though there is a clear cut difference between total lack of assimilation and the remaining hiatus resolution processes, we cannot speak of categorical rules, as whenever assimilation happens, it does so in varying degrees.

Second, several factors influence the amount of V1V2 assimilation. Vowels around strong boundaries show significantly greater resistance to assimilation than those around weak boundaries. Moreover, vowels show a non-significant trend to assimilate less around focused environments than around non-focused ones. Connected to these findings, I showed that segments have shorter duration around weaker boundaries than around stronger ones and in non-focus environments than in focused ones. Although we cannot argue for a causal link between the assimilation results and the duration results, there is strong positive correlation between segmental duration and the degree of assimilation: Vowels with long duration display more resistance to assimilation than vowels with short duration. The combination of these results makes clear the need to investigate further the effects that lengthening due to focus has on assimilation, to reveal whether lengthening alone, without the confounding factor of prosodic boundaries, is what inhibits assimilation. This study failed to establish statistically significant effects of focus on the hiatus resolution process. Further investigation is necessary to confirm this result.

The acoustic effects of varying degrees of assimilation cross-linguistically have been described in the literature as results of variable overlap between the articulatory gestures for the two hiatus vowels. It has been claimed that more time allows articulators to reach their target positions, but when time is compressed articulators do not have time to reach the target and only reach intermediate positions approximating the target. As a result, gestures are compressed by overlapping with each other more, that is, articulators do not reach their target, and each gesture lasts a shorter time. Although the results of this study could certainly accommodate such an explanation, it should be borne in mind that this was an acoustic study, not an articulatory one, and no hard evidence can be provided for or against an explanation within Articulatory Phonology.

Finally, one of the questions left unanswered in is how the quality of the input vowels influences hiatus resolution. The results show a significant influence of the input vowels both on segmental duration and the degree of assimilation. Several vowel properties were examined in section 3.3 in connection to this question, among them vowel height, sonority hierarchy, vowel roundness, and relative height between V1 and V2, but no single general property can account for these results. A larger scale study may reveal generalizations that the present study failed to detect.

In summary, we have shown that the process of hiatus resolution in Greek is gradient, not categorical. Moreover, the degree of assimilation between hiatus vowels in Greek and their duration are affected by their prosodic position, with less assimilation and more lengthening found around strong boundaries and focus.

Notes

* I am indebted to Sun-Ah Jun for her advice, encouragement, and invaluable help through all the stages of this work. I would also like to thank Pat Keating for her advice and improvement suggestions, audiences at UCLA and LabPhon 8 for helpful comments, as well as Marios Fourakis and an anonymous reviewer for many helpful comments on the first version of the manuscript.

1. This distance is the square root of the sum of the squares of the difference between the vowel formant frequencies ($ED= \sqrt{((F1V1–F1V2)^2 + (F2V1–F2V2)^2)}$). I thank one of the reviewers for suggesting this measure.
2. All three phonological rules for deletion mentioned here were employed in Kaisse (1977).

References

Arvaniti, Amalia and Mary Baltazani
2000 Greek ToBI: A system for the annotation of Greek speech corpora. In: *Proceedings of the Second International Conference on Language Resources and Evaluation*, Volume 2, 555–562.
2004 Intonational analysis and prosodic annotation of Greek spoken corpora. In: Sun-Ah Jun (ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press.

Baltazani, Mary and Sun-Ah Jun
1999 Focus and topic intonation in Greek. In: *Proceedings of the XIVth International Congress of Phonetic Sciences*, Volume 2, 1305–1308. San Francisco, CA.

Browman, Catherine P. and Louis M. Goldstein
1986 Towards an articulatory phonology. *Phonology Yearbook* 3: 219–252.
1989 Articulatory gestures as phonological units. *Phonology* 6: 201–251.
1992 Articulatory phonology: An overview. *Phonetica* 49: 55–180.

Casali, Roderic F.
1997 Vowel elision in hiatus contexts: Which vowel goes? *Language* 73: 493–533.

Cho, Taehong
2001 Effects of prosody on articulation in English. Ph.D. dissertation, UCLA.
2002 *The Effects of Prosody on Articulation in English*. New York: Routledge.
in press Prosodically-conditioned strengthening and vowel-to-vowel coarticulation in English. *Journal of Phonetics*.

Condoravdi, Cleo
1990 Sandhi rules of Greek and prosodic theory. In: Sharon Inkelas and Draga Zec (eds.), *The Phonology-Syntax Connection*. University of Chicago Press.

Cooper, William E. and Jeanne Paccia-Cooper
1980 *Syntax and Speech*. Cambridge, MA: Harvard University Press.

Edwards, Jan E., Mary E. Beckman and Janet Fletcher
1991 The articulatory kinematics of final lengthening. *Journal of the Acoustical Society of America* 89: 369–382.

Fallon, Paul
1994 Naturally occurring hiatus in Modern Greek. In: Irene Philippaki-Warburton, Katerina Nicolaidis and Maria Sifianou (eds.), *Themes in Greek Linguistics*, 217–224. London: John Benjamins Publishing Co.

Fourakis, Marios, Antonis Botinis and Maria Katsaiti
1999 Acoustic characteristics of Greek vowels. *Phonetica* 56: 28–43.

Frota, Sónia
2002 The prosody of focus: a case-study with cross-linguistic implications. In: *Proceedings of Speech Prosody*. Aix-en-Provence.

Hadzidakis, Georgios
1905 *Meseonika ke Nea Ellinika (Medieval and Modern Greek)*. Athens.

Joseph, Brian D. and Irene Philippaki-Warburton
1987 *Modern Greek*. London: Croom Helm.

Jun, Sun-Ah
1993 The phonetics and phonology of Korean prosody. Ph.D. dissertation, Ohio State University.
1995 Asymmetrical prosodic effects on the laryngeal gesture in Korean. In: B. Connell and Amalia Arvaniti (eds.), *Phonology and phonetic evidence: Papers in Laboratory Phonology IV*, 235–253. Cambridge: Cambridge University Press.

Jun, Sun-Ah and Cecile Fougeron
2000 A phonological model of French intonation. In: Antonis Botinis (ed.), *Intonation: Analysis, Modeling and Technology*, 209–242. Kluwer Academic Publishers.
Jun, Sun-Ah and Hyuck-Joon Lee
1998 Phonetic and phonological markers of contrastive focus in Korean. In: *Proceedings of the 5th International Conference on Spoken Language Processing*, Volume 4, 1295–1298. Sydney.
Kaisse, Ellen
1977 Hiatus in Modern Greek. Ph.D. dissertation, Harvard University.
Keating, Patricia, Taehong Cho, Cecile Fougeron and Chai-Shune Hsu
in press Domain-initial articulatory strengthening in four languages. *Journal of Phonetics*.
Klatt, Dennis H.
1975 Vowel lengthening is syntactically determined in connected discourse. *Journal of Phonetics* 3: 129–140.
Koutsoudas, Andreas and O. Koutsoudas
1962 A contrastive analysis of the segmental phonemes of Greek and English. *Language and Learning* 12: 211–230.
Mackridge, Peter
1985 *The Modern Greek Language: A Descriptive Analysis of Standard Modern Greek*. Oxford: Oxford University Press.
Nespor, Marina and Irene Vogel
1986 *Prosodic Phonology*. Dordrecht: Foris.
Pelekanou, Theodora and Amalia Arvaniti
2001 Post-lexical rules and gestural overlap in a Greek spoken corpus. In: *Proceedings of the 5th International Conference on Greek Linguistics*.
Pierrehumbert, Janet and D. Talkin
1992 Lenition of /h/ and glottal stop. In: Gerard J. Docherty and D. R. Ladd (eds.), *Papers in Laboratory Phonology II*, 90–117. Cambridge: Cambridge University Press.
Wightman, Colin W., Stefanie Shattuck-Hufnagel, Mari Ostendorf and Philip J. Price
1992 Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America* 91: 1707–1717.
Zsiga, Elizabeth
1997 Features, gestures, and Igbo vowels: An approach to the phonology-phonetics interface. *Language* 73: 227–274.

Early vs. late focus: Pitch-peak alignment in two dialects of Serbian and Croatian*

Rajka Smiljanić

The overall goal of this study was to investigate the extent to which the existence of a lexical tonal contrast affects variation in the acoustic properties that cue the contrast. More specifically, the lexical, pragmatic and prosodic effects on F_0 peak alignment in two dialects of Serbian and Croatian with distinct phonological properties were examined: the Belgrade dialect with a lexical contrast between 'rising' and 'falling' accents and the Zagreb dialect without such lexical pitch-accent contrast. In the dialect with a lexical pitch-accent contrast, the effect of narrow focus on peak alignment is to enhance the contrast. In the dialect without such contrast, the effect of narrow focus is to enhance the prominence of the focused word/syllable by locating the salient pitch peak more strictly on the accented syllable. It was argued that in Zagreb, alignment differences are due to pragmatically defined accent categories. Tonal crowding in the phrase-final positions additionally affects tonal alignment regardless of the presence or absence of the lexical pitch-accent contrast. These results suggest that the variation in the acoustic properties that cue the contrast is restricted by the existence of a phonemic contrast. Furthermore, the phonetic implementation of F_0 targets is determined by an interaction of lexical, pragmatic and prosodic factors. Combined, all of these results have implications for the distinction between tonal alignment and association.

1. Introduction

Tune-text association has been one of the central questions in Intonational Phonology (Pierrehumbert 1980, Horne (ed.) 2000). Categorical tonal distinctions defined in the phonology of a language provide tonal targets in a tune that associate with segmental anchors. Work by Bruce (1977) on Swedish and by Pierrehumbert (1980) on English intonation has shown early on that tonal alignment with the segmental string is precisely controlled. However, alignment of tonal targets varies across and within languages. Alignment differences can arise from factors such as: categorical tone contrasts, pragmatic context (e.g., focus), tonal context (crowding), and position in prosodic phrase. In Swedish, tonal alignment is used to distinguish between two lexically contrastive accents: Accent I and Accent II (HL* vs. H*L, Bruce 1990). In English, tonal alignment makes a similar distinction between contrastive *pragmatically*-defined pitch-accents (L*+H vs. L+H*). Pragmatic information such as narrow focus can be expressed through changes in pitch contours and changes in alignment. Narrow focus can be expressed through introduction of an additional sentential H tone as in Swedish (Bruce 1977), or by pitch-peak retraction as in Spanish and Greek (Face 2002, Botinis 1998), or through pitch-peak protraction as in European Portuguese and Palermo Italian (Frota 2000, Grice 1995). For instance, in Spanish, a prenuclear pitch-accent has a 'late' peak, aligned on the posttonic syllable, in a broad focus sentence and an 'early' peak, aligned within the accented syllable, when the same word is narrowly focused (Face, 2002). Some of these pragmatically induced alignment changes have been analyzed as consisting of categorical distinctions between accent types: H+L* vs. H*+L for European Portuguese (Frota 2000) and Palermo Italian (Grice 1995), L*+H vs. L+H* for Spanish (Face 2002), etc. In these analyses, the difference in the tonal alignment in narrow focus is placed in the realm of phonology through the inventory of contrastive pitch-accents. However, in other analyses, the same phenomena of timing differences in the expression of pragmatic information are explained through gradient alignment effects rather than categorical distinctions (Nibert 2000, Hualde 2002a, Ladd 1996: 96–98).

Alignment patterns are further finely tuned by proximity to other tones. Prieto et al. (1995) have shown that in Spanish pitch-peak location crucially depends, among other things, on the adjacency to intonational boundaries and the proximity to other tones, which causes earlier alignment. Silverman and Pierrehumbert (1990) show that in English pitch peaks on words in nuclear position overall have earlier alignment than in prenuclear position, presumably due to the upcoming boundary tones in the final position. Furthermore, they showed that pitch-peaks are aligned later in vowels which are elongated due to a slowing down of speech. The opposite is true when vowels are lengthened due to final lengthening effects (e.g. in phrase final positions), i.e., prosodic lengthening induces earlier peak alignment. All of these studies suggest that tonal alignment is sensitive to a combination of the factors mentioned above and that several mechanisms, such as those underlying pragmatic focus and prosodic position, define tonal realization. In the recent years, other factors that additionally affect peak alignment, such as speaking rate (Xu 1998, Ladd et. al. 1999) and phonological vowel length (Ladd et. al. 2000) have been identified in numerous, mostly acoustic, studies of various languages.

In this paper, I explore the interplay of lexical, pragmatic and prosodic factors and how some of these potentially conflicting demands on F_0 alignment are resolved, through a comparative study of two dialects of Serbian and Croatian (S/C): Belgrade Serbian and Zagreb Croatian (henceforth Belgrade and Zagreb respectively). Three factors and their interactions that might determine pitch peak alignment in these two dialects are considered: 1) lexical i.e., presence or absence of a lexical pitch-accent; 2) pragmatic, i.e., broad vs. narrow focus in utterance-initial position; and 3) prosodic, i.e., adjacent vs. non-adjacent to an intonation boundary/boundary tones (early vs. late focus). The two dialects investigated here have distinct phonological properties: Belgrade has a lexical contrast between 'rising' and 'falling' accents with 'late' and 'early' pitch-peak alignment respectively while Zagreb does not have such a lexical pitch-accent contrast (Lehiste and Ivić 1986, Skarić 1991, Smiljanić and Hualde 2000, Smiljanić 2004).

The hypothesis for the present study is that the peak alignment patterns in these two dialects will vary as a function of the presence vs. absence of a lexical pitch-accent contrast (similar to different coarticulation effects in languages with different phonemic inventories, Manuel 1990, Cohn 1993). Zagreb without the lexical pitch contrast will freely use phonetic space through alignment adjustment to express narrow focus. In Zagreb, then, peak alignment patters would be *pragmatically* rather than lexically driven as in other intonation languages. However, the expression of narrow focus will be different from Belgrade with the lexical contrast. Belgrade will exhibit less freedom in utilizing the phonetic space since the location of peaks crucially differentiates categorical distinctions at the lexical level. The maintenance of the lexical pitch-accent contrast will, therefore, limit the amount of alignment change available for expressing pragmatic narrow focus. Furthermore, the prosodic effect (proximity to an intonational phrase boundary/boundary tones) is expected to exert influence on F_0 peak alignment in both dialects. It is expected that the upcoming boundary will cause earlier alignment for both accent types (Silverman and Pierrehumbert 1990). This effect might potentially jeopardize the lexical contrast in Belgrade where the 'falling' peaks are already aligned 'early', i.e., with the word-initial tonic syllable. 'Falling' peaks might, therefore, have less room for retraction than the word-initial 'rising' accents which are aligned with the posttonic syllable. In this situation, the peaks for the two accentual categories might overlap to some extent (see below for the description of the experimental materials).

Through this investigation several questions are addressed: First, what role do lexical, pragmatic and prosodic factors play in shaping pitch contours in S/C? Second, how does the presence vs. absence of a lexical pitch contrast limit pragmatic and prosodic effects on pitch peak alignment? Third, are there differences between pragmatic and prosodic effects on F_0 peak alignment and, if so, how are these potentially conflicting effects resolved? Finally, how are phonological tonal contrasts implemented phonetically? The results show that pitch peak alignment in Belgrade and Zagreb is greatly sensitive to a combination of lexical, pragmatic and prosodic factors. As hypothesized, there is a difference between the two dialects in patterns of pitch-peak adjustments that corresponds to the presence vs. absence of a lexical pitch-accent contrast. In Zagreb, all peaks are retracted in narrow focus while in Belgrade, the 'falling' peaks are retracted and 'rising' peaks are unchanged or slightly protracted. Thus, the contrast between the two lexical accent categories is enlarged in Belgrade through the asymmetric manipulation of the 'rising' and 'falling' peaks rather than just maintained through the possible uniform retraction of both 'falling' and 'rising' peaks. Also as hypothesized, the upcoming prosodic boundary/boundary tones cause earlier alignment of peaks in utterance-final position when compared with the utterance-initial position in both dialects. However, it was not found that the lexical contrast in narrow-focused words in Belgrade is ever endangered. Finally, the results show that phonetic alignment is not a direct mapping of phonological categories, supporting a crucial distinction between tonal association and alignment (Ladd 1996).

2. Serbian and Croatian

S/C is a stress-language with a lexical pitch alignment contrast only in words with initial stress (Smiljanić and Hualde 2000). This contrast is manifest in 'late' vs. 'early' alignment of accent peaks. Traditionally these accent types are designated as 'rising' and 'falling' respectively (Lehiste and Ivić 1963, 1986, Godjevac 1999, 2000, among others). In the 'rising' accents F_0 rises through most of the accented syllable and the peak is aligned with the posttonic while in the 'falling' accents F_0 rises and falls within the accented syllable with the peak being aligned with the accented vowel. In addition, there is a lexical contrast in vowel length in S/C (Lehiste and Ivić 1986, among others). According to Lehiste and Ivić (1986) similar F_0 patterns hold for both long and short series. This description corresponds to most varieties of standard S/C and is reflected in the speech of Belgrade speakers (one of the dialects under investigation here). Although standard Croatian is said to have a contrast in accent type and vowel length similar to those of Standard Serbian, Magner and Matejka (1971) show that some of the accentual distinctions are not found in Zagreb. Furthermore, Zagreb Kajkavian, a non-standard dialect of the area, has been described as lacking both the length and the pitch accent contrasts (Magner 1966). Although the existence of these contrasts is taught in the schools, it is at least doubtful that some Zagreb speakers with the Kajkavian background make either of the two contrasts, in any style (their local dialect or the more formal standard dialect). Therefore, it is expected that some of the Zagreb speakers will exhibit different prosodic characteristics even in their standard speech which is used in the Experiment in this paper. The validity of this hypothesis was established in a pilot study by Smiljanić and Hualde (2000) who show that some Zagreb speakers lack the pitch-accent contrast when speaking the standard variety. For a more detailed description of the dialectal situation see Magner (1966), Inkelas and Zec (1988) and Smiljanić (2004) among others. It is important for this paper to emphasize that standard Croatian and standard Serbian are mutually comprehensible to a very high degree. Impressionistically they do not differ from each other more than, say, British RP and General American English. In the experiments conducted in this investigation, all words are segmentally identical in both varieties. The only expected differences are prosodic.

3. Experiment

Six subjects were recorded: three Belgrade speakers (two female and one male), two Zagreb speakers (one male and one female), and one Karlovac male speaker (some 30 kilometers south of Zagreb). Both Zagreb speakers and the Karlovac speaker are of the same Kajkavian language background and are expected to exhibit the same behavior with regard to the pitch-accent contrast, i.e., they would transfer the prosodic patterns of their local dialect, in which pitch-accent contrast is absent, into their standard dialect productions, which is the style expected in the formal reading situations.[1] All speakers produced the same disyllabic target words in carrier sentences. The traget words have either a 'falling' (F) or a 'rising' (R) accent on the first syllable. All target syllables have the vowel /a/ and only long vowels were chosen, since accent distinctions are most clearly seen with long vowels (for the behavior of short vowels, see Smiljanić 2004). Mostly sonorant sounds were chosen to avoid segmental pitch perturbations. Example sentences are given in (1) where target words are bolded and the accented syllable within the target word is capitalized:

(1)(R)	a.	***MAra***	*je*	*jela*	*bananu*.
		Mara	3rd person Sg-to be	PAST-eat	ACC-banana.
		'Mara ate a banana.'			
	b.	*Nada*	*je*	*vidjela*	***MAru***.
		Nada	3rd person Sg-to be	PAST-sea	ACC-Mara.
		'Nada saw Mara.'			
(F)	c.	***MLAda***	*je*	*jela*	*bananu.*
		Bride	3rd person Sg-to be	PAST-eat	ACC-banana.
		'The bride ate a banana.'			
	d.	*Mama*	*je*	*vidjela*	***MLAdu***.
		Mom	3rd person Sg-to be	PAST-sea	ACC-bride.
		'Mom saw the bride.'			

Target words were placed in either sentence- initial or final position. The test sentences were read in three blocks: first, with the broad focus reading over the entire test sentence; second, with the narrow focus reading on the sentence-initial target word, and third, with the narrow focus reading on the sentence-final target word. The broad focus reading over the entire test sentence was elicited by the triggering question '*What happened yesterday?'* where the answer contains all new information. The narrow focus reading for the target word in the initial position was prompted by a question such as '*Did the bridegroom eat a banana?* for the answer sentence, '*The bride ate a banana*' which requires contrastive focus on the target word, in this example 'bride'. The narrow focus reading for the same target word in the final position was elicited with '*Did mom see the bridegroom?'* for the answer sentence '*Mom saw the bride*'. The prompting questions and answers were written on index cards. The experimental text was written in the Roman alphabet, with which all speakers were familiar and comfortable. Accent types are not indicated in the orthography. The speakers were instructed to read out loud only the answers to the prompting questions.

In all carrier sentences, the number of syllables in sentences was kept at 8. The number of syllables between targeted accent and the preceding and following accents was kept at two so that accent clash would be avoided. Arvaniti et al. (2000) show for Greek that three syllables are needed between accents for a canonical alignment of accents. It is expected, however, that the two-syllable inter-accent interval will allow for accentual gestures to be executed close to their full form. Furthermore, the syllable interval between accents is kept constant across accent types and pragmatic conditions. If there is any tonal crowding present it is of the same amount for both 'rising' and 'falling' accents and in both broad and narrow focus conditions.

The sentences within each block were randomized and repeated (5 times in the initial position and 10 times in the final position).[2] 120 sentences per speaker were obtained in this way (30 broad focus initial + 30 narrow focus initial + 60 narrow focus final). The total number of target words obtained was 720. Speakers were recorded under quiet conditions using the Kay Computerized Speech Laboratory (CSL) 4300B at the sampling rate of 16 kHz. The sentences were analyzed using PRAAT software for speech analysis (Boersma 1996).

3.1. Measurements

The sentences were transcribed and segmented manually by a combination of listening and by inspection of the spectrogram, F_0 track and sound wave. Measurements were made of the accented vowel duration and the F_0 maximum location with respect to the end of the accented syllable. F_0 peaks were fairly easy to determine. In cases of high plateau (rather than a clear peak) the beginning of the plateau (elbow) was chosen as the pitch peak. Some of the sentences were discarded due to either bad recording, or the wrong accent type put on the word, or the wrong syllable accented, or other factors, such as extensive coarticulation, that obscured segment boundaries (as determined by the experimenter). This yielded a total of 696 sentences used in the analyses.

The choice to measure the peak alignment data with respect to the ˈVC-boundary was made because it best captures the phonological 'targets' for the 'falling' and 'rising' accents, as previously described, i.e., the peaks for the 'rising' accents target the posttonic syllable and the peaks for the 'falling' accents target the accented syllable in Belgrade. The end of the accented vowel/syllable, therefore, seems to be a landmark relevant for peak alignment. The end of the stressed vowel is in most cases the syllable boundary as well, i.e. most stressed syllables are open syllables (with the exception of the word *marva* 'cattle' which has a word-initial closed stressed syllable). Where necessary, peak alignment data relative to C'V-boundary are included as well.

4. Results I: expression of focus in the initial position

F_0 peak alignment data from the utterance-initial position are given in scatter plots in Figure 1 separately for each speaker. The left three panels show peak alignment for the Belgrade speakers and the right three panels for the Zagreb speakers. In all scatter plots, the X-axis shows peak alignment with respect to the end of the accented vowel/syllable (*0* on X-axis). Peak alignment is plotted against vowel duration (given on Y-axis). It can be observed that patterns of peak alignment differ in two dialects. In Belgrade, the largest distinction seems to be between circles and squares, i.e., 'rising' peaks cluster together and are later than the 'falling' peaks. The biggest difference in Zagreb, on the other hand, is between empty and filled symbols, i.e., between broad and narrow focus. The 'falling' and 'rising' peaks within a pragmatic condition (e.g. filled circles vs. filled squares) are greatly overlapped in Zagreb. Peak alignment data were submitted to a two-way ANOVA for each speaker separately. The analyses involved two fixed factors: accent type ('rising' vs. 'falling') and pragmatics (broad vs. narrow focus). Since numerous ANOVAs were performed in all statistical analyses alpha level was adjusted to .001 to ensure against committing a type I error. For all three Belgrade speakers there was a significant effect of accent on peak alignment: B1: $F(1,55) = 981.858$, $p < .0001$; B2: $F(1,56) = 306.409$, $p < .0001$; B3: $F(1,54) = 289.507$, $p < .0001$. For B2 and B3 there was a significant effect of pragmatics: B1: $F(1,55) = 4.642$, $p = .036$ ns; B2: $F(1,56) = 70.577$, $p < .0001$; B3: $F(1,54) = 60.050$, $p < .0001$. Additionally, for all three speakers there was a significant two-way interaction: B1: $F(1,55) = 23.292$, $p < .0001$; B2: $F(1,56) = 30.534$, $p < .0001$; B3: $F(1,54) = 21.592$, $p < .0001$. Pairwise comparisons reveal that accent is a significant factor for all three speakers for broad focus condition: B1: $t(13) = 13.892$, $p < .0001$; B2: $t(14) = 7.746$, $p < .0001$; B3: $t(12) = 7.083$, $p < .0001$, and for narrow focus condition: B1: $t(14) = 23.892$, $p < .0001$; B2: $t(14) = 19.660$, $p < .0001$; B3: $t(14) = 7.083$, $p < .0001$. In other words, in both pragmatic conditions, 'falling' accents are aligned differently than 'rising' accents (they are earlier as seen in Figure 1). Furthermore, pairwise comparisons of accent type across two pragmatic conditions show a significant effect of focus for 'falling' accents: B1: $t(13) =$

6.640, p < .0001; B2: t(14) = 9.725, p < .0001; B3: t(12) = 6.965, p < .0001, but not for ‘rising’ accents: B1: t(14) = 2.790, p = .014 ns; B2: t(14) = 2.693, p = .018 ns; B3: t(14) = 3.455, p = .004 ns.

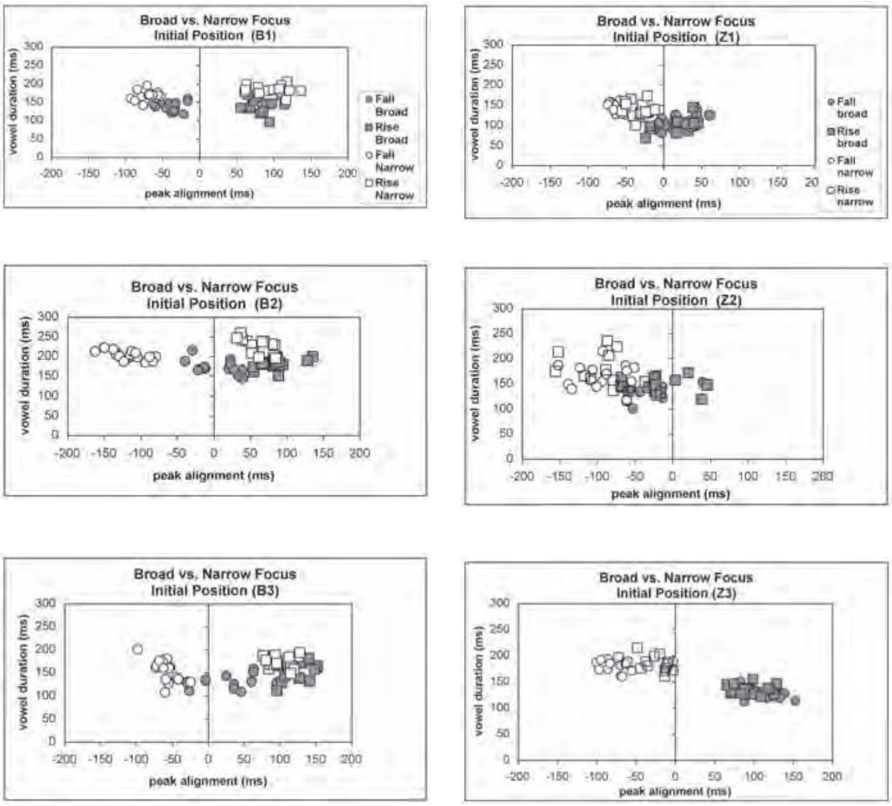


Figure 1. F_0 peak alignment in broad (filled symbols) and narrow (empty symbols) focus conditions in utterance-initial position. The panels on the left show the results for the Belgrade speakers and on the right for the Zagreb speakers. ‘Rising’ accents are shown as squares and ‘falling’ accents as circles. X-axis shows peak alignment with respect to the end of the accented vowel/syllable (0). Y-axis shows vowel duration.

This shows that the ‘falling’ accents change their alignment with respect to the end of the accented vowel significantly (empty circles vs. filled circles), i.e., they are aligned earlier with respect to the end of the accented vowel/syllable when compared to ‘falling’ peaks in broad focus. The ‘rising’ accents remain fairly stable in their alignment with respect to the end of the accented vowel (empty squares vs. filled squares).

For none of Zagreb speakers was there a significant effect of accent on F_0 alignment: Z1: $F(1,55) = 4.965$, $p = .031$, ns; Z2: $F(1,48) = 1.630$, $p = .208$ ns.; Z3: $F(1,54) = 1.132$, $p = .293$ ns. For all three speakers there was a significant main effect of pragmatics on F_0 peak alignment: Z1: $F(1,55) = 99.594$, $p < .0001$; Z2: $F(1, 48) = 46.878$, $p < .0001$; Z3: $F(1,54) = 431.079$, $p < .0001$. There were no significant two-way interactions: Z1: $F(1,55) = 1.658$, $p = .204$ ns; Z2: $F(1,48) = .116$, $p = .735$ ns; Z3: $F(1,54) = .742$, $p = .039$ ns. These results show that, in Zagreb, F_0 alignment is not different for 'rising' and 'falling' accents (circles and squares within a pragmatic condition are largely overlapped in Figure 1, right panels). F_0 peaks do, however, change their alignment in narrow focus, i.e., they are aligned significantly earlier (empty symbols are earlier than filled symbols in Figure 1).

4.1. Discussion

The results of the broad focus condition confirm earlier experimental findings that Belgrade speakers differentiate between 'early' and 'late' peak alignment corresponding to the lexical pitch accent categories. (Lehiste and Ivić 1986, Smiljanić and Hualde 2000). However, it can be seen from the plots in Figure 1 that the 'falling' peaks are not necessarily on the tonic syllable as traditionally described. Speakers B2 and B3 align 'falling' peaks in broad focus on both tonic and posttonic syllables (filled circles). Despite this 'non-canonical' alignment of 'falling' peaks they are distinguished from the 'rising' peaks which are placed significantly later onto the posttonic syllable. This 'non-canonical' alignment is not due to tonal crowding, since F_0 peaks are aligned 'later' than their traditional descriptions suggest, i.e., closer to the following tonal target. Such alignment patterns could possibly be due to hypo-articulation and a less-precise coordination of segmental and tonal targets characteristic of more informal speaking styles to which broad focus sentences might belong (Lindblom 1990, de Jong 2000). The inherent variability observed in larger distribution of the 'falling' accents data can partly be explained by the variable syllable structure of the words involved. One of the words, *marva* 'cattle', has a closed accented syllable and one of the words, *mlada* 'bride', has an onset cluster. Both of these factors could affect peak alignment and could account for the variability in the results.[3]

The results of the broad focus condition further confirm earlier experimental findings that the pitch accent contrast is absent for some Zagreb speakers (Smiljanić and Hualde 2000). For Zagreb speakers, F_0 peaks in broad focus can be characterized as 'late' peaks with somewhat variable alignment for different Zagreb speakers (filled symbols, right panels in Figure 1). This peak alignment does not correspond clearly with the late/'rising' peak alignment in Belgrade for all speakers. The variability in peak alignment observed among Zagreb speakers could partly be accounted for by the fact that one speaker (Z2) is from Karlovac. It could be argued that this speaker does not clearly belong to the Zagreb dialect but to a distinct dialectal group. However, the fact that this speaker lacks the pitch accent contrast makes him more similar to Zagreb speakers than to Belgrade speakers. In any case, the variability between/among dialects suggests that the prenuclear 'rising' tonal categories are phonetically implemented in different ways. The variability in the alignment of 'rising' peaks cannot be accounted for by the varying structure of the test words since all 'rising' words have the same ˈCVCV structure. It is more likely that there are dialect differences in the instantiation of this accentual category (similar to Atterer and Ladd, 2004). Finally, differences in peak alignment among the speakers within a dialect (even excluding Z2) suggest that the alignment of tonal targets in this case is not as invariable as that of tonal targets in Greek pre-nuclear accents (Arvaniti et al. 1998).

The results for narrow focus show that F_0 peak alignment is changed in both dialects. However, the alignment change is different in the two dialects. The alignment adjustment crucially depends on the 'availability' of the phonetic space to realize the change in the alignment. For Belgrade speakers the phonetic space in broad focus is already carved in two parts for the lexical distinction between 'rising' and 'falling' accents. Such division of the phonetic space limits the available alignment change for Belgrade speakers. In narrow focus the lexical distinction is maintained and, despite the alignment changes that peaks undergo, they do not cross over these spaces ('rising' peaks are not realized on the tonic syllable although potentially the retraction of both 'falling' and 'rising' peaks could be a signal of narrow focus). In Zagreb, there is no such division of phonetic space in broad focus. The availability of phonetic space for 'early' alignment is utilized completely for the expression of narrow focus. For all Zagreb speakers all F_0 peaks in narrow focus are aligned with the tonic syllable (where the lexical category of 'falling' accents is realized for Belgrade).

An important acoustic correlate of narrow focus, which enhances the salience of the accented syllable in both dialects equally, is vowel lengthening.

For all speakers white symbols are above the filled symbols in Figure 1. It is important to look closer at vowel duration data in order to establish whether vowel lengthening is the cause of the change in the peak alignment rather than the independent peak retraction. It is possible that the peaks have the same alignment with respect to the beginning of the accented vowel in both pragmatic conditions. The observed change in the alignment could then entirely be due to the lengthening of vowels in narrow focus. Two-way ANOVAs for each Belgrade speaker separately were performed for the vowel duration data. The analyses involved two fixed factors: accent type ('rising' vs. 'falling') and pragmatics (broad vs. narrow focus). For all three Belgrade speakers there is a significant main effect of pragmatics on vowel duration: B1: $F(1,55) = 75.463$, $p < .0001$; B2: $F(1,56) = 85.597$, $p < .0001$; B3: $F(1,54) = 14.057$, $p < .0001$. For B2 and B3 there is a significant effect of accent on vowel duration: B1: $F(1,55) = 7.629$, $p = .008$ ns; B2: $F(1,56) = 12.540$, $p = .001$; B3: $F(1,54) = 15.620$, $p < .0001$. No two-way interactions were significant: B1: $F(1,55) = 1.345$, $p < .251$ ns; B2: $F(1,56) = 6.288$, $p < .015$ ns; B3: $F(1,54) = .010$, $p = .919$ ns. Vowels are lengthened in narrow focus on average 35 ms for B1, 37 ms for B2, and 18 ms for B3. Vowel length also differs between 'rising' and 'falling' accents. ANOVAs with only pragmatics as a fixed factor were performed on the vowel duration data for Zagreb speakers (accent is no longer considered to be a factor). For all three speakers there is a significant main effect of pragmatics on vowel duration: Z1: $F(1,55) = 43.259$, $p < .0001$; Z2: $F(1,48) = 15.744$, $p < .0001$; Z3: $F(1,54) = 256.396$, $p < .0001$. Vowels are lengthened in narrow focus on average 33 ms for Z1, 28 ms for Z2 and 50 ms for Z3.

In order to check whether the difference in the alignment is due to vowel lengthening or to the active peak alignment manipulation, mean peak alignment change (ΔH) between broad and narrow focus was calculated with respect to the beginning of the accented vowel. The results are given in Table 1. The results in Table 1 show that for B2 and B3 'falling' peaks are retracted independently of vowel lengthening ('falling' peaks for these speakers were in the 'non-canonical' posttonic position in broad focus). This is not the case for B1 whose 'falling' peaks were already on the tonic in the broad focus condition. For this speaker vowels are significantly lengthened in narrow focus which indicates that the peak alignment results are not a consequence of 'misreading' of the pragmatic context. This suggests that the primary function of peak retraction is enhancing the contrast (already large enough for B1 in broad focus). Additionally, 'falling' peaks are treated differently from the 'rising' peaks, which are either slightly protracted or not changed (for B3) but are never retracted. This strategy allows the Belgrade speakers to additionally exaggerate the contrast between 'falling' and 'rising' accents. All three Zagreb speakers retract all the peaks independently of vowel lengthening. These results suggest that pitch manipulation due to narrow focus is somewhat different in the two dialects, i.e., it reflects their different phonological properties. In Belgrade lexical contrast is enlarged through the asymmetrical manipulation of 'rising' and 'falling' accents (similar to the lexical vowel length contrast enhancement reported by, e.g., de Jong and Zawaydeh 2001, Smiljanić 2004). In Zagreb, on the other hand, the F_0 peak alignment change is completely a function of expression of pragmatic narrow focus (confirming Smiljanić and Hualde, 2000). Therefore, peak alignment is not utilized in the same way in expressing pragamtic narrow focus in the two dialects.

Table 1. Mean amount of F_0 peak alignment change between broad and narrow focus conditions in the initial position. Negative numbers indicate retraction.

ΔH (ms)	B1	B2	B3	Z1	Z2	Z3
F	–1.36	–80.7	–72.2	–32.6	–40.8	–103.1
R	13.2	25.1	–4.5			

5. Results II: early vs. late focus

Here, the results of the comparison between early and late focus (utterance-initial vs. final position) are given (e.g., ***Mara*** in the initial narrow focus condition (3a) vs. ***Maru*** in the final narrow focus condition (3b)). The data in Figure 2 are plotted in the same way as in Figure 1. In all plots in Figure 2 the empty symbols are data from the narrow focus initial position condition (the same points were seen in Figure 1). The filled symbols are data from the narrow focus final position. For Zagreb speakers, 'falling' and 'rising' points are collapsed since it was shown in Results I that accentual distinctions are absent in this dialect. Peak alignment in the final position is additionally modified when compared with the initial position (filled symbols vs. empty symbols). The most striking adjustment is in the alignment of the 'rising' accents in Belgrade (filled squares). Recall that for this accent type in both broad and narrow conditions in the initial position peaks are consistently 'late' (on the posttonic syllable). In the final position they are clustered fairly tightly around the end of the tonic syllable. For Zagreb speakers Z1 and Z3,

F_0 peaks in the final position (filled triangles) are earlier than the peaks in the initial position. For Z2 the peaks in both conditions are largely overlapped.

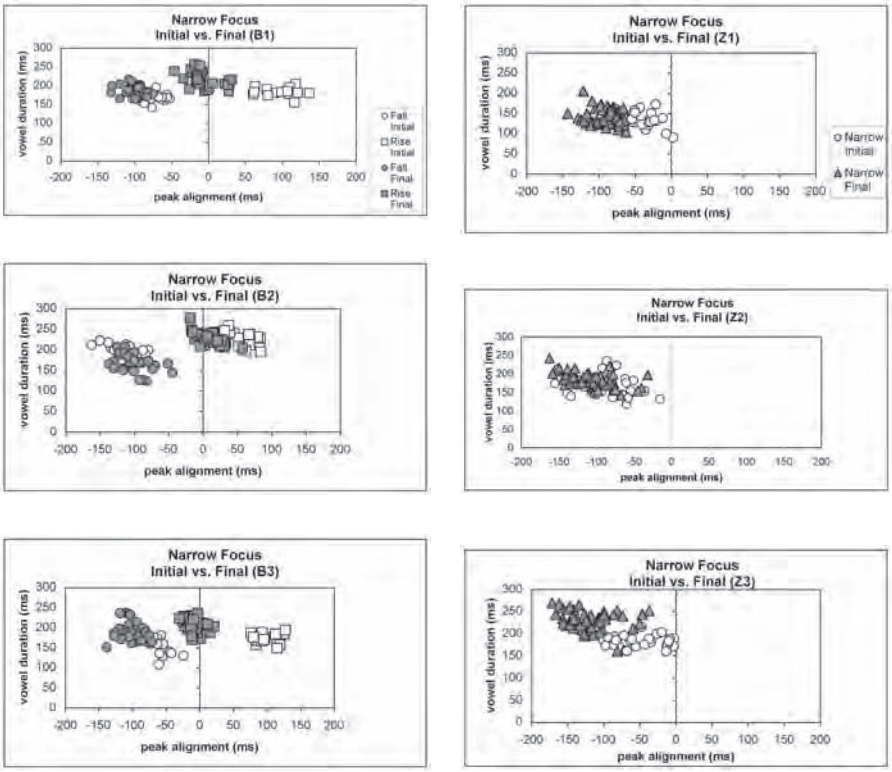


Figure 2. F_0 peak alignment for the Belgrade speakers (left panels) and the Zagreb speakers (right panels) in narrow focus in initial and final positions. The data are plotted in the same way as in Figure 1.

Two-way ANOVAs with position (initial vs. final)[4] and accent ('falling' vs. 'rising') as fixed factors show that for all Belgrade speakers there is a main effect of position on peak alignment: B1: $F(1,56) = 179.541$, $p < .0001$; B2: $F(1,56) = 9.231$, $p < .001$; B3: $F(1,56) = 256.289$, $p < .0001$. There is also a significant main effect of accent on peak alignment: B1: $F(1,56) = 692.805$, $p < .0001$; B2: $F(1,56) = 636.459$, $p < .0001$; B3: $F(1,56) = 764.281$, $p < .0001$. Additionally, there is a significant two-way interaction for all three speakers: B1: $F(1,56) = 45.922$, $p < .0001$; B2: $F(1,56) = 47.137$, $p < .0001$; B3: $F(1,56) = 48.684$, $p < .0001$. It was established earlier that 'falling' and ‘rising’ accents are significantly different in narrow focus in the initial position for all three speakers. Pairwise comparisons reveal that accent is a significant factor in the final position as well: B1: $t(14) = 15.890$, $p < .0001$; B2: $t(14) = 12.649$ $p < .0001$; B3: $t(14) = 15.305$, $p < .0001$. Furthermore, for all three speakers ‘rising’ accents change their alignment significantly in the final position when compared with the initial position: B1: $t(14) = 11.402$, $p < .0001$; B2: $t(14) = 11.073$, $p < .000$; B3: $t(14) = 17.038$, $p < .0001$. Finally, the alignment of ‘falling’ accents is significantly different in the final position for B1 and B3 only: B1: $t(14) = 6.693$, $p < .0001$; B2: $t(14) = 1.895$, $p = .097$ ns; B3: $t(14) = 5.319$, $p < .0001$. These results show that in both positions, ‘falling’ and ‘rising’ peaks are distinct with ‘falling’ accents being aligned earlier. Additionally, F_0 peaks are aligned earlier in the final position than in the initial position (filled vs. empty symbols). For two Zagreb speakers ANOVAs with the fixed factor of position (initial vs. final) showed a significant effect of position on peak alignment: Z1: $F(1,56) = 107.058$, $p < .0001$; Z2: $F(1,55) = 3.773$, $p = .057$ ns; Z3: $F(1,51) = 103.669$, $p < .0001$. For Z1 and Z3, F_0 peaks are aligned earlier in final position. For Z2 there is no difference in the alignment between two conditions.

5.1. Discussion

The results show that ‘falling’ and ‘rising’ peaks are distinct for Belgrade speakers in narrow focus final position. The results also show that peaks are additionally modified in the final position in both dialects. The ‘falling’ peaks for Belgrade speakers are just slightly earlier in the vowel (not for B3) but the ‘rising’ peaks are much earlier than in the sentence-intial position. For Zagreb speakers, peaks are earlier in the final position except for Z2.

Vowel lengthening was examined in the final position as well. Two-way ANOVAs with position (initial vs. final) and accent (‘rising’ vs. ‘falling’) as fixed factors for each Belgrade speaker show that there is a main effect of position on vowel duration: B1: $F(1,56) = 62.805$, $p < .0001$; B2: $F(1,56) = 16.792$, $p < .0001$; B3: $F(1,56) = 30.819$, $p < .0001$. Accented vowels are on average longer in the final position for B1 by 29 ms and for B3 by 28 ms but are shorter for B2 by 19 ms. There is also a main effect of accent for all speakers: B1: $F(1,56) = 34.041$, $p < .0001$; B2: $F(1,56) = 96.706$, $p < .0001$; B3: $F(1,56) = 12.844$, $p < .001$. For all speakers, there is an overall significant durational distinction between ‘rising’ and ‘falling’ accents. Finally, there is a significant two-way interaction for B2: B1: $F(1,56) = 1.948$, $p <$

.168 ns; B2: F (1,56) = 20.455, p < .0001; B3: F(1,56) = .080, p < .778 ns. Pairwise comparisons for B2 show that in both positions vowels with 'rising' accents are different from vowels with 'falling' accents: $t_{init}(14) = 4.489$, p = .001; $t_{fin}(14) = 10.827$, p < .0001. It can be seen in Figure 2 that for B2, vowels for 'rising' accents are longer than vowels for 'falling' accents (squares vs. circles). The effect of position is significant for vowels with 'falling' accents: t(14) = 5.934, p < .0001 but not for vowels with 'rising' accents: t(14) = .288, p = .778. Vowels for 'falling' accents are shortened while there is no major change in the duration of 'rising' accents in the final position. ANOVAs with position (initial vs. final) as a fixed factor show that there is a main effect of position on vowel duration for Z3: Z1: F(1,56) = 1.022, p = .316 ns; Z2: F(1, 55) = 8.098, p = .006 ns; Z3: F(1,51) = 120.898, p < .0001. Vowels are lengthened on average 4 ms for Z1, 18 ms for Z2 and 42 ms for Z3. The overall effect of final position on vowel duration (filled symbols vs. empty symbols in Figure 2) seems to be of smaller magnitude than was the effect of focal lengthening in the initial position for all speakers.

The mean change in peak alignment (ΔH) in the final position with respect to the beginning of the accented vowel is given in Table 2. This was done in order to see whether the observed peak alignment change is due to vowel lengthening alone or to an independent retraction:

Table 2. The change in the peak alignment between the initial and final positions with respect to the beginning of the accented vowel.

<table>
<tr><th>ΔH (ms)</th><th>B1</th><th>B2</th><th>B3</th><th>Z1</th><th>Z2</th><th>Z3</th></tr>
<tr><td>F</td><td>–9.8</td><td>+18.2</td><td>–12.32</td><td rowspan="2">–41.7</td><td rowspan="2">–0.23</td><td rowspan="2">–28.6</td></tr>
<tr><td>R</td><td>–68.3</td><td>–53.1</td><td>–79.9</td></tr>
</table>

Peak alignment change in the final position has the effect of 'repelling' the peaks away from the boundary in both dialects (similar to Silverman and Pierrehumbert 1990) rather than exaggerating the phonemic accentual contrast. For most speakers peaks are retracted independently of vowel lengthening. For Z2, however, peaks are not retracted in the final position. It can be noted that for Z2 already in the narrow focus initial position F_0 peaks were placed earlier than for the other two speakers (a consequence of earlier peak alignment in broad focus for this speaker, i.e., peaks are on the tonic in broad focus). For Belgrade speakers, the retraction is large for the 'rising' peaks and smaller or absent for the 'falling' peaks. Table 3 shows the amount of accentual difference between ‘falling’ and ‘rising’ peaks in the two positions for Belgrade speakers:

Table 3. The mean amount of difference between ‘falling’ and ‘rising’ peak alignment in the initial and final narrow focus conditions.

‘falling’ vs. ‘rising’ distinction	B1	B2	B3
F-R/Initial	167.6 ms	178.3 ms	160.4 ms
F-R/Final	125.3 ms	167.8 ms	112.42 ms

The change in the peak alignment induced by the final position decreases the difference in the peak alignment between ‘rising’ and ‘falling’ accents. The proximity to the IP boundary affects peak alignment in the opposite direction from the pragmatic effect of narrow focus which increases the lexical pitch distinction (cf. intial broad focus vs. intial narrow focus). Furthermore, the effect of the final position goes against the claim that phonetic space is carved into two areas and that peak adjustments never cause ‘falling’ and ‘rising’ peaks to cross over these spaces. The ‘rising’ alignment in the final position is the same as the ‘falling’ alignment in broad focus in the initial position by some Belgrade speakers (see Figure 1). This suggests that in a situation when ‘late’ alignment is unavailable due to tonal crowding and/or unavailability of segmental material, ‘rising’ peaks can ‘encroach’ on the ‘falling’ ‘territory’. However, for all Belgrade speakers, the difference in the alignment between ‘falling’ and ‘rising’ in the final position is still significant (as shown in Results II). Asymmetric vowel shortening/lenghthening of ‘rising’ and ‘falling’ vowels could function as a repair mechanism allowing for the distinction between the two accent types to be maintained (see B2 results). It remains to be investigated further whether the changes in the alignment of peaks in the final position, such that the ‘rising’ peaks are ‘early’ compared to their alignment in the intial position, affect listeners’ perception of the these lexical categories.

6. General discussion

The results show that lexical, pragmatic and prosodic factors determine pitch peak alignment in both dialects of S/C. In postulating phonological tonal categories, therefore, all of the discussed factors have to be taken into account. The following analyses are proposed for the two dialects: in Belgrade there is a lexical contrast between L*+H and L+H* for the 'rising' and 'falling' accents respectively (for additional arguments for the proposed analysis, see Smiljanić 2003, 2004; for an alternative account see Godjevac 2000, 2001 and Godjevac and Arvaniti, 2003). The timing difference between the late-aligned 'rising' peaks and the early-aligned 'falling' peaks is indicated by the starred notation.[4] In Zagreb, the phonological tonal category is L*+H. The actual alignment data show inter- and intra-speaker variation in preferred peak position. The prenuclear 'rising' category, L*+H, is not implemented in the same way in the two dialects. The variation observed shows that these categories can be implemented in different ways on an alignment continuum (similar argument is given in Atterer and Ladd, 2004). The proposed pitch-accent categories, thus, present idealized associations with segmental anchors rather than the actual alignment facts.

In narrow focus, the pattern of F_0 peak alignment change in Zagreb could be explained by postulating a *pragmatically*-defined pitch-accent category L+H* (similar to the analyses for Spanish and Portuguese, Face 2002, Frota 2000, and possibly for Greek, based on data described in Botinis 1998). The Belgrade narrow focus data, on the other hand, do not lend themselves to the same analysis. Despite peak alignment modifications, we wouldn't want to postulate two new pitch-accent categories ('earlier' early peak vs. 'later' late peak). It is not clear what the anchors for these tonal targets would be (they would have to refer to the pitch-accents in broad focus and specify the alignment to be 'earlier' or 'later' in narrow focus although still on the tonic syllable for 'falling' and on the posttonic syllable for 'rising'). It was shown above that the peak alignment is modified in such a way that the pitch-accent contrast is enlarged. Therefore, it is proposed that F_0 peak alignment changes in Belgrade are due to phonetic level adjustments in the expression of narrow focus without changing the accent category. This analysis would capture the distinction in the mechanisms underlying narrow focus effects in the two dialects, whereby in Belgrade the lexical contrast is enlarged while in Zagreb the focus mechanism is entirely driven by pragmatic considerations.

We need to, briefly, consider the alignment patterns in the final position as well. For Belgrade speakers, 'rising' peaks in the final position are aligned earlier (on the tonic) when compared with the initial narrow focus condition. By looking at the alignment facts in this condition alone one could argue that in fact H is a starred tone (aligning with the end of the accented vowel). However, through the comparison of all three conditions investigated here we can conclude that 'rising' peak alignment in the final position is not a new pitch-accent category (somewhat intermediate to 'falling' and 'rising' alignment in the initial position). Such alignment is a result of the interaction of pitch accents and boundary tones. Furthermore, the same analysis accounts for the Zagreb data and reflects the fact that the effect of the prosodic position is the same for both dialects. All of these analyses support the difference between tonal alignment and association as given in Ladd (1996). These results show that tonal alignment is controlled by various mechanisms and that in order to arrive at the right phonological analyses of pitch patterns we have to take into account lexical, pragmatic, and prosodic components which can become obvious only through experimental investigation.

Finally, it is important to note that the results in the final position are taken from the narrow focused words, which may be characterized as hyperarticulated or 'listener oriented' (Lindblom 1990, de Jong 2000). A different pattern is observed for target words in the final position with broad focus reading (not discussed here). There, lexical contrast seems to be jeopardized due to deaccenting, i.e., there are no discernable peaks in this condition. It seems that the overall contrast, in both broad and narrow focus, is diminished in the final position. This is in agreement with the finality effects discussed by Hock (1999) whereby the pitch contrasts are diminished or obliterated in prepausal contexts. Furthermore, the displacement of H to the left in the final narrow focus could lend support to the claim made by Becker (1979, as cited in Hock 1999) that the leftward accentual shift, which occurred in S/C in the fifteenth century and reorganized the accentual paradigm, originated in such prepausal contexts.

7. Conclusions

This paper has examined lexical, pragmatic and prosodic effects on F_0 peak alignment in two dialects of S/C. It finds that phonetic implementation of F_0 targets is determined by all of these factors. The effects on peak alignment vary for these two closely related language varieties. We have compared segmentally identical texts produced by S/C speakers with and without a lexical pitch-accent contrast. In the dialect with a lexical pitch-accent contrast, the effect of narrow focus on peak alignment is to enlarge the contrast. In the dialect without such contrast, the effect of narrow focus is to enhance the prominence of the focused word/syllable by locating the salient pitch peak more strictly on the accented syllable similar to Spanish (Face 2002), Italian (Grice 1995) and Portuguese (Frota 2000). It was argued that in Zagreb, alignment differences are due to pragmatically defined accent categories. The effect of the final position in the prosodic phrase (proximity to a boundary and boundary tones) is similar in both dialects. There seems to be a constraint against F_0 peak alignment on the posttonic syllable which is the last syllable in the utterance and a host for the realization of boundary tones. Therefore, in both dialects F_0 peaks are further away from the boundary in final position. This causes a conflict in the Belgrade dialect between the demands to maintain the lexical pitch-accent contrast in a hyperarticulated narrow focus context and to accommodate the upcoming boundary tones with little segmental material available. However, the contrast is maintained albeit of smaller magnitude.

This investigation by no means exhausts all the factors that determine phonetic implementation of tonal targets in S/C. A number of other effects need to be calculated for deriving the right model of phonetic tonal implementation. Peak alignment might change as a function of vowel duration differences due to changes in speech rate (Xu 1998, Ladd et al. 1999), of phonological vowel length (Xu 1998, Ladd et al. 2000), of segmental syllable composition (Arvaniti et al. 1998, Ladd et al. 2000), etc. Furthermore, it is possible that it is not just the F_0 peak that changes its alignment. The preceding L target could change as well (Face 2002). In that case the investigated factors would affect the entire rising gesture rather than just the F_0 peaks. Finally, articulatory constraints, such as the maximum speed of pitch change, and perceptual constraints may impose additional constraints that determine tonal alignment (Xu 2002, D'Imperio 2000). Research is needed to determine what these factors are and how they interact, potentially shedding light on the interplay between universal and language-specific constraints.

Notes

* I am grateful to Jennifer Cole, Jose Hualde, Zsuzsanna Fagyal, Carlos Gussenhoven and Ken de Jong for help at various stages of this project. I am also grateful to two anonymous reviewers and Louis Goldstein for helpful suggestions and comments on an earlier draft. All errors are mine.

1. The label 'Zagreb' dialect is used for the speakers of standard Croatian with Kajkavian background as opposed to the speakers of standard Serbian. The speakers included in this group might differ among themselves, in addition to being different from the Belgrade speakers.
2. The data in the initial position are a subset of a larger set collected in another experiment (Smiljanić 2004). The subset used here is limited to the words with the long vowel /a/ to make the comparison with the same words in the final position

more felicitous. This is the reason for the asymmetry in the number of repetitions between initial and final positions.

3. I want to thank an anonymous reviewer for bringing up this point.
4. For statistical analyses of peak alignment data only first five repetitions of each word obtained for the final position were included. This was done in order to balance the number of data points per cell for initial and final position.
5. It would be possible to argue that neither tone is starred as was done by Arvaniti et al. (2000) for Greek. However, this analysis would not capture the regularities in tonal alignment observed in these data.

References

Arvaniti, A., D. R. Ladd, and I. Mennen

1998 Stability of tonal alignment: the case of Greek pre-nuclear accents. *Journal of Phonetics* 26: 3–25.

2000 What is a starred tone? Evidence from Greek. In: M. Broe and J. Pierrehumbert (eds.), *Papers in Laboratory Phonology V: Acquisition and the Lexicon*, 119–131. Cambridge University Press.

Atterer, M. and D. R. Ladd

2004 On the phonetics and phonology of 'segmental anchoring' of F0: evidence from German. *Journal of Phonetics* 32(2): 177–197.

Becker, L.

1979 A contribution to an explanation of the Neo-Stokavian accent retraction. *Zbornik za filologiju i lingvistiku.* 22:1: 1–23.

Boersma, P.

1996 Praat: A system for doing phonetics by computer, University of Amsterdam. Available online from http://www.praat.org.

Botinis, A.

1998 Intonation in Greek. In: D. Hirst and A. Di Cristo (eds.), *Intonation systems: A survey of twenty languages*, 288–310. Cambridge: Cambridge University Press.

Bruce, G.

1977 *Swedish word accents in sentence perspective*. Lund: CWK Glerup.

1990 Alignment and composition of tonal accents: comments on Silverman and Pierrehumbert's paper. In: J. Kingston and M. E. Beckman (eds.) *Papers in Laboratory Phonology I: Between the grammar and the physics of speech*, 72–106. Cambridge: CUP.

Cohn, A.

1993 Nasalization in English: Phonology or Phonetics? *Phonology* 10: 43–81.

D'Imperio, M.
2000 The role of Perception in Defining Tonal Targets and their Alignment. Ph.D. Thesis. The Ohio State University.

de Jong, K.
2000 Attention modulation and the formal properties of stress systems. In: A. Okrent and J. P: Boyle (eds.): *CLS 36, vol. 1, The Main Session*, 71–92. Chicago: Chicago Linguistic Society.

de Jong, K. and B. Zawaydeh
2001 Comparing stress, lexical focus, and segmental focus: patterns of variation in Arabic vowel duration. *Journal of Phonetics* 30:1, 53–75.

Face, T.
2002 Contrastive Focus and Local Intonation Patterns in Spanish. *Probus* 14.1: 71–92.

Frota, S.
2000 *Prosody and Focus in European Portuguese: Phonological Phrasing and Intonation.* New York: Garland Publishing.

Godjevac, S.
1999 *An autosegmental/metrical analysis of Serbo-Croatian intonation*. In: C. Roberts, J. Muller and T. Huang (eds.): *Ohio state working papers in linguistics: Varia,* volume 54. Columbus, Ohio: Ohio State University.
2000 Intonation, Word Order, and Focus Projection in Serbo-Croatian. Unpublished Ph.D. dissertation, Ohio State, Columbus.
2001 Serbo-Croatian ToBI (SC_ToBI). http://www.ling.ohio-state.edu/~godjevac/SCTobi5.pdf

Godjevac, S. and A. Arvaniti
2003 Phonological structure and phonetic realization of Serbian pitch accents. Presented at the *Experimental Approaches to Tone and Intonation* Workshop of the European Science Foundation Network „Tone and Intonation in Europe.“ Oxford.

Grice, M.
1995 *The intonation of interrogation in Palermo Italian: implications for intonation theory.* Tübingen: Niemeyer.

Hock, H. H.
1999 Finality, prosody, and change. In: O. Fujimura, B. D. Joseph and B. Palek: *Proceedings of LP 98*, 15–30. Prague: Karolinum Press.

Horne, M. (ed.)
2000 *Prosody: Theory and Experiment.* Dordrecht: Kluwer.

Hualde, J. I.
2002 Intonation in Spanish and other Ibero-Romance languages: overview and status quaestionis. In: C. R. Wiltshire and J. Camps (eds.): *Romance Phonology and Variation*, 101–115. Amsterdam: Benjamins.

Inkelas, S., and D. Zec
1988 Serbo-Croatian Pitch Accent: The Interaction of Tone, Stress, and Intonation. *Language: Journal of the Linguistic Society of America,* 64(2): 227–248.

Ladd, R.
1996 *Intonational Phonology*. Cambridge: Cambridge University Press.

Ladd, R., D. Faulkner, H. Faulkner and A. Schepman
1999 Constant ‚segmental anchoring' of F_0 movements under changes in speech rate. *Journal of the Acoustical Society of America*, 106(3): 1543–1554.

Ladd, R., I. Mennen and A. Schepman
2000 Phonological conditioning of peak alignment in rising pitch accents in Dutch. *Journal of the Acoustical Society of America,* 107(5): 2685–2695.

Lehiste, I. and P. Ivić
1963 Accent in Serbo-Croatian: An experimental study. *Michigan Slavic Materials, 4*.
1986 *Word and sentence prosody in Serbocroatian*. Cambridge, MA: MIT Press.

Lindbloom, B. E. F.
1990 Explaining phonetic variation: a sketch of the HandH theory. In: H. J. Hardcastle and A. Marchal (eds.) *Speech Production and Speech Modeling, NATO ASI Series D: Behavioral and Social Sciences,* Vol. 55, 403–439. Dordrecht: Kluwer.

Magner, T.
1966 *Zagreb Kajkavian Dialect* (Vol. 18). University Park, PA: Pennsylvania State University.

Magner, T., and L. Matejka
1971 *Word Accent in Modern Serbo-Croatian*. Pennsylvania UP: University Park, Pa.

Manuel, S.
1990 The role of contrast in limiting vowel-to-vowel coarticulation in different languages. *Journal of the Acoustical Society of America,* 88: 1286–1298.

Nibert, H.
2000 Phonetic and phonological evidence for intermediate phrasing in Spanish intonation. Ph.D. Dissertation. University of Illinois at Urbana-Champaign.

Pierrehumbert, J.
1980 The phonology and phonetics of English intonation. Unpublished doctoral dissertation, Cambridge, MA: MIT.

Prieto, P., J. v. Santen and J. Hirschberg
1995 Tonal Alignment Patterns in Spanish. *Journal of Phonetics,* 23(4): 429–451.
Silverman, K. and J. Pierrehumbert
1990 The timing of prenuclear high accents in English. In: J. Kingston and M. E. Beckman (eds.) *Papers in Laboratory Phonology I: Between the grammar and the physics of speech*, 72–106. Cambridge: CUP.
Skarić, I.
1991 Fonetika Hrvatskog Knjizevnog Jezika. In: R. Katicic (ed.), *Povijesni pregled, glasovi i oblici hrvatskoga knjizevnog jezika*, 61–379. Zagreb: Globus, Nakladni zavod.
Smiljanić, R.
2004 *Lexical, Pragmatic and Positional Effects on Prosody in two dialects of Croatian and Serbian: An Acoustic Study.* Outstanding Dissertations in Linguistics. New York: Routledge.
to appear Lexical and Pragmatic Effects on Pitch Range and Low Tone Alignment in Two Dialects of Serbian and Croatian. In: *CLS 39, vol. 1*. Chicago: Chicago Linguistic Society.
Smiljanić, R. and J. I. Hualde
2000 *Lexical and Pragmatic Functions of Tonal Alignment in Two Serbo-Croatian Dialects.* In: A. Okrent and J. P: Boyle (eds.): *CLS 36, vol. 1, The Main Session*, 469–482. Chicago: Chicago Linguistic Society.
Xu, Y.
1998 Consistency of tone-syllable alignment across different syllable structures and speaking rates. *Phonetica* 55: 179–203.
2002 Articulatory constraints on tonal alignment. In: *Proceedings of SP2002*, Aix-en-Provence, France.

Manifestation of prosodic structure in articulatory variation: Evidence from lip kinematics in English

Taehong Cho

This study investigates effects of prosodic structure on kinematic variations that may illuminate how prosody is manifested in articulatory variation. Kinematic characteristics of lip aperture (e.g., articulatory displacement, duration and velocity) were examined with respect to three prosodically strong locations: domain-initial, and domain-final syllables. The results (obtained with the Electromagnetic Articulograph, EMA) showed that each of these prosodic locations are associated with distinctive kinematic patterns that can distinguish itself from others: All of the three prosodically important locations showed strengthening effects with generally longer and larger movements in common, but they differred primarily in velocity: faster for accented gestures, no change for domain-initial/final lip opening gestures, and slower for cross-boundary lip closing gestures. This suggests that a hierarchically-nested prosodic structure is marked by systematic kinematic variation. The results were further evaluated in terms of the dynamical parameter settings (e.g., stiffness, intergestural timing, target amplitude, re-scaling) in a mass-spring gestural model. Close examination of relationships between various kinematic parameters (e.g., displacement, velocity, duration, time-to-peak velocity) suggested that at the very least one should look for a combination of settings for multiple dynamical parameters, in order to account for the prosodically-driven articulatory systematicity in a way that is both descriptively and explanatorily adequate.

1. Introduction

The term *prosody* has traditionally been used as the cover term for suprasegmental features such as pitch, duration and intensity, but more recently, it is often used to refer to more abstract linguistic notion “a hierarchically organized structure of phonologically defined constituents and heads" (Beckman 1996), in which lower domains (e.g., syllables) are typically grouped into immediately higher levels (e.g., words), eventually forming the Intonational Phrase (IP) (see Shattuck-Hufnagel and Turk 1996 for a review). This structural view of prosody assumes that prosody is a grammatical entity in its own right, which is a crucial element of speech production and speech comprehension processes.

One line of research, taking the structural view of prosody, has vigorously shown that the prosodic structure of an utterance is phonetically manifested on the surface by distinctive pitch patterns and temporal structure at the right edge of prosodic constituent (e.g., Beckman and Pierrehumbert 1986; Pierrehumbert and Beckman 1988; Edwards, Beckman and Fletcher 1991; Jun 1993; 1998, and Beckman 1996). Further, a recent articulatory study has shown that domain-final vocalic articulation in English also undergoes a substantial increase in spatial magnitude (Cho 2002, 2005), contrary to a common assumption that domain-final position is subject primarily to temporal expansion (cf. Beckman, et al. 1992).

While these domain-final phonetic events have widely been considered as major phonetic correlates of prosodic structure, researchers have recently started to look at domain-initial positions for other potential phonetic events correlated with prosodic structure (Pierrehumbert and Talkin 1992; Jun 1993; Fougeron and Keating 1997; Cho and Keating 2001; Fougeron 2001; Cho 2002, 2004, 2005; Keating, Cho, Fougeron and Hsu 2003). For example, a phrase-initial stop /t/ is likely realized with a longer VOT and a larger linguopalatal contact than the same /t/ occurring in the middle of a phrase, a phenomenon know as *domain-initial strengthening*.

Yet other phonetic correlates of prosodic structure other than domain-final phenomena come from cross-boundary phenomena such as cross-boundary vowel-to-vowel coarticulatory resistance (Cho 2004) and the relative timing of consonant and vowel gestures (Byrd 2000). For example, in an articulatory study, Cho (2004) showed that vowels in prosodically stronger locations are coarticulated less with neighboring vowels, but do not exert a stronger influence on the articulation of neighboring vowels.

Such robust phonetic phenomena in the vicinity of prosodic boundaries have led to a growing awareness that it is no longer fruitful to describe the sound properties of a language without adequately taking into account the interface between prosodic structure and phonetics. Accordingly, the focus of recent laboratory work has been on more diverse prosodic locations, including domain-initial and -final positions, as well as stressed (pitch-accented) syllables (de Jong 1991, 1995, Cho 2002). These three positions have been shown to give rise to some type of strengthening of articulatory properties of features or gestures (also known as *prosodic strengthening*), which is taken to be an articulatory signature of prosodic structure.

The majority of phonetic research has, however, generally been limited to one prosodic effect (the stress or the edge effect), and has thus failed to provide a comprehensive account of articulatory characteristics of prosodic structure. The present study examines the three prosodic locations concurrently, in order to understand the prosody-phonetics interface in a coherent way. In particular, it aims to understand the effects of prosody in English on kinematic variations and considers dynamical accounts that may illuminate how prosody is manifested in articulatory variation. To this end, it examines lip movement kinematics of accent- and boundary-induced articulatory strengthening and how accent-induced kinematic patterns differ from boundary-induced ones. Further, given that prosodically-conditioned articulatory variation may be controlled by a particular dynamical parameter setting in a mass-spring gestural model (Beckman, et al. 1992; Harrington, et al. 1995; Byrd and Saltzman 1998; Byrd, et al. 2000), it is further evaluated whether and how prosodically-driven strengthening may be accounted for by a particular dynamical parameter setting, and whether different dynamical mechanisms govern the articulatory characteristics that arise from different prosodic locations.

Task dynamic model and dynamical parameters. In the task dynamic model, the articulatory gesture is described in terms of the behavior of an abstract 'mass' (an idealized articulator such as the tongue) which is connected to a 'spring' and a 'damper' in a critically damped mass-spring system (Saltzman and Munhall 1989). As Hawkins (1992) describes, it is as if one end of the spring is attached to the mass, and the other end is held at the target location. Then, as the target location changes, the spring is stretched, and the mass is pulled towards the target location. In a critically damped mass-spring gestural model, however, the mass does not oscillate, but asymptotes towards the equilibrium position, such that the gesture is generally realized as a one-directional movement towards the target.

In the model, the gesture is defined as a dynamical system specified with a set of parameter values. Relevant dynamical parameters include target (underlying amplitude), stiffness (or natural frequency), damping ratio, intergestural timing, and activation time. Characteristics of the articulatory movements that result from executing gestures depend on the values of the parameters specified for a given gesture. Crucially, any systematic kinematic variation is interpreted as the consequence of dynamical parameter settings. Thus, in theory, systematic kinematic variations arising from prosodic conditions should be accounted for by either a particular dynamical mechanism or an interaction of more than one mechanisms.

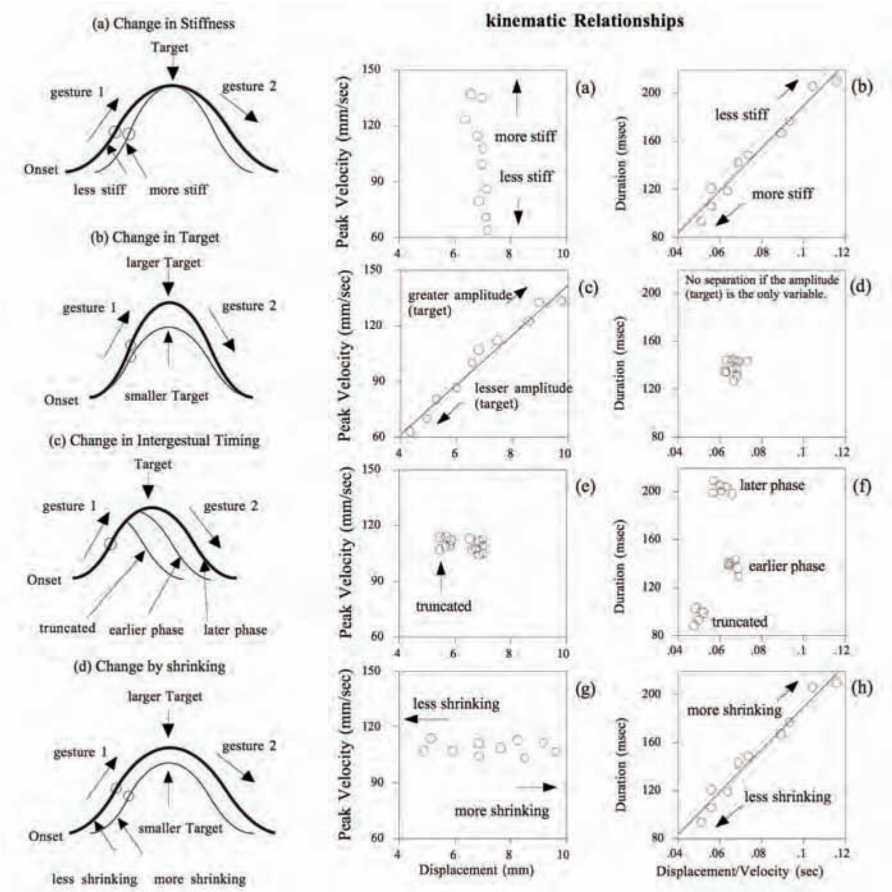


Figure 1. Hypothetical trajectories that correspond to a change in each parameter (left) and relationships among kinematic variables that manifest dynamical parameter settings (right). In the left panel, empty circles indicate the timepoint of the peak velocity attainment. In the right panels, (a–b) show change in stiffness; (c–d), change in amplitude (target); (e–f), change in intergestural timing; and (g–h), change by shrinking. Figures (a)–(f) are adapted from Beckman, et al. (1992).

Some researchers (Beckman, et al. 1992; Byrd, et al. 2000) provide useful summaries of the kinematic consequences of various mass-spring parameter manipulations. The left panel of Fig. 1 shows schematized movement trajectories that correspond to changes in four dynamical parameters. (Although shrinking is a scaling of two parameters, it will be called 'parameter' for the sake of simplicity.) The right panels of Fig. 1 visualize idealized kinematic manifestations of different dynamical specifications by relating some kinematic measures to each other.

(1) *Stiffness.* Variation in movement duration is thought to be controlled by the stiffness parameter: the stiffer the spring (the articulator), the faster the movement (see the left panel of Fig. 1). If stiffness is the only parameter underlying kinematic differences, there should be a change in peak velocity (the maximum velocity during the movement), but not in displacement (spatial distance that the articulator travels), therefore showing vertical distribution of the datapoints (Fig. 1a). In addition, there should be a proportional change in both duration and displacement/velocity ratio, with a diagonal distribution of the datapoints (Fig. 1b), i.e., as duration increases, peak velocity would decrease, making the displacement/velocity ratio increase (because of constant displacement). Further, the time-to-peak-velocity (the interval from the onset to the attainment of peak velocity) will vary as stiffness changes (the less stiff, the longer) (cf. Byrd and Saltzman 1998; Byrd, et al. 2000; and Byrd 2000).

(2) *Target (articulatory amplitude).* A change in target induces a change in displacement. In a pure target change, peak velocity and displacement changes proportionally without a change in duration: With stiffness being held constant, articulators have to travel farther with no extra time. The only way to reach the increased target is by increased velocity in proportion to the change in the target value, with a diagonal distribution of the datapoints (Fig. 1c). Further, since displacement and velocity change proportionally, there should be no change in the displacement/velocity ratios, nor should there be a change in duration (due to no stiffness change) (Fig. 1d). Time-to-peak-velocity also remains constant (the left panel of Fig. 1).

(3) *Intergestural timing or truncation.* The articulatory movement towards the target can be 'truncated' by an early activation of the following gesture, which keeps the movement from reaching its assumed target. Thus, under a pure change in intergestural timing, there should be no change in peak velocity because the effect of a substantially earlier following gesture is mainly to prevent the preceding gesture from reaching its target. As Byrd, et al. (2000) noted, if the gesture has a plateau-like shape at its peak displace-

ment and the truncation applies primarily to this region, the change in displacement will be small or zero, but relatively larger if the truncation applies to the region beyond the plateau-like shape (See Fig. 1e). Further, there will be a substantial change in duration as the following gesture is phased earlier or later, whereas the ratio of the displacement to the peak velocity remains relatively unchanged because of no change in displacement and velocity, except when enough truncation brings about a decrease in displacement (Fig. 1f). Finally, the durational change comes from a change in the interval from the timepoint of peak-velocity to the target (deceleration duration) with no change in time-to-peak-velocity.

(4) *Shrinking.* Shrinking can be defined as a change in both target and stiffness which are scaled proportionally. Shrinking can be thought of as a unique dynamical parameter that may underlie prosodically conditioned kinematic variation (see Harrington, et al. (1995), and Byrd, et al. (2000)). As can be seen in Fig. 1 (left), in a pure proportional change in target and stiffness, there would be a proportional increase in both duration and displacement, which results in no change in peak velocity, giving a horizontal distribution of the datapoints (Fig. 1g). Further, the displacement/velocity ratio will increase as duration increases, giving a diagonal distribution of the datapoints (Fig. 1h). Note that the pattern in Fig. 1h is similar to that in Fig. 1b under a change in stiffness. However, the difference between these two is that in a change in stiffness (Fig. 1b), the displacement/velocity ratio increases as duration increases not because displacement increases, but because velocity decreases with displacement being held constant.

2. Experiment

In order to examine effects of various prosodic conditions on speech production, lip movement data in American English were collected, using Electromagnetic Midsagittal Articulography (Carstens Articulograph). An important criterion for building the corpus was to include both prosodic and segmental variables. Each item in the corpus included two test syllables (pre- and post-boundary), yielding a $C_1V_1\#C_2V_2$ sequence (# = a prosodic boundary) across words. C_1 and C_2 were always /b/, whose articulation is known to minimally interfere with the vocalic lingual articulation. V_1 and V_2 were either /i/ or /a/, resulting in four pairs: /bi#bi/, /ba#ba/, /bi#ba/, and /ba#bi/, but in this study only the data for /bi#bi/ and /ba#ba/ are examined to control for the vowel type.

The boundary between the test syllables was varied from the Intonational Phrase boundary (IP), to the Intermediate Phrase boundary (ip), to the Word boundary (Wd). At the same time, accentuation was manipulated in preboundary and postboundary syllables, resulting in four patterns: ACC#ACC, ACC#UNACC, UNACC#ACC, UNACC#UNACC. Such a manipulation yields three prosodic variables: (a) prosodic boundary; (b) accentuation of syllables (accented, unaccented); (c) position of test syllables (initial, final). Thus, the corpus contained every combination of the prosodic and segmental factors, yielding a total of 24 different sequences (3 boundaries 2 accentual patterns in the preboundary syllable 2 accentual patterns in the postboundary syllable 2 vowel type (/bibi/ vs. /baba/). (Sample sentences are given below in Table 1.)

Six American English speakers participated. In order to control for rounding in the low vowel, speakers whose dialect lacked /ɔ/ were chosen. They were all trained in producing English sentences in the ToBI framework (Beckman and Elam 1997) prior to the experiment. Before the actual recording date, each speaker participated in an approximately two-hour long practice session in order to be able to produce the intended renditions as naturally as possible. During the experiment, speakers were instructed to produce two different versions (ip vs. IP) of a sentence in order to obtain balanced ip and IP tokens.

The target sequences were obtained from real sentences in a mini discourse situation intended to induce the desired variety of accent-placement patterns and prosodic groupings. A sample sentence set with /ba#ba/ tokens in *Little Bah bopped the girl* is given in Table 1. (/bi#bi/ sequence tokens were produced in similar discourse frames as in *Donna B. beeped at him*) In each target sentence, the words in bold received pitch accent. The prompt was read silently by the speaker to cue the intended accent patterns, which were provided using partial ToBI transcriptions in the script. Six American English speakers were recorded reading the target sentences. Each sentence was read twice in succession, and the entire list was read twice, for a total of four repetitions per sentence. This yielded a total of 576 sentence tokens (24 sentence types x 6 speakers x 4 repetitions).

In the EMA experiment, seven transducer coils were used. Two reference transducers were placed on the nose and upper gumline, or maxilla, in order to correct for head movement inside the helmet. Two transducers were mounted on the upper and lower lips at the vermilion borders (L1, L2) to monitor lip closing and opening movements. (The remaining three transducers were located on the tongue; the data from these transducers were analyzed in Cho (2002, 2004, 2005)).

Table 1. A subset of the corpus containing /ba#ba/ sequences with different prosodic boundaries (IP, ip, Wd) and accentual patterns.

= Word boundary:

(a) *ACC.-UNACC.*		
	Prompt:	Did you just say "Little **Boo** bopped the girl last night"?
	Target:	No, "Little **Bah** # bopped the girl"
	rendition:	H* L– L%
(b) *UNACC.-ACC.*		
	Prompt:	Did you just say "Little Bah **popped** the girl last night"?
	Target:	No, "Little Bah # **bopped** the girl"
	rendition:	H* L– L%
(c) *ACC.-ACC.*		
	Target:	You know what? Little **Bah** # **bopped** the girl.
	rendition:	H* H* L– L%
(d) *UNACC.-UNACC.*		
	Prompt:	Did you just say "**Big** Bah bopped the girl last night"?
	Target:	No, "**Little** Bah # bopped the girl"
	rendition:	H* L– L%

= Intermediate or Intonational Phrase boundaries (ip or IP):

(e) *ACC.-UNACC.*		
	Prompt:	Did you say "Little **Boo** bopped the **boy** last night"?
	Target:	No, " Little **Bah** # bopped the **girl**."
	rendition 1:	H*L– H* L– L%
	rendition 2:	H*L–L% H* L– L%
(f) *UNACC.-ACC.*		
	Prompt:	Did you say "**Big** Bah **popped** the girl last night"?
	Target:	No, " **Little** Bah # **bopped** the girl.
	rendition 1:	H* L– H* L– L%
	rendition 2:	H* L–L% H* L– L%
(g) *ACC.- ACC.*		
	Prompt:	Did you say "Little **Boo popped** the girl"?
	Target:	No, " Little **Bah** # **bopped** the girl.
	rendition 1:	H*L– H* L– L%
	rendition 2:	H*L–L% H* L– L%
(h) *UNACC.-UNACC.*		
	Prompt:	Did you say "**Big** Bah bopped the **boy** last night"?
	Target:	No, " **Little** Bah # bopped the **girl**"
	rendition 1:	H* L– H* L– L%
	rendition 2:	H* L–L% H* L– L%

Next, the articulatory space was rotated so that the x-axis was the occlusal plane using a bite-plate with two additional transducers on it. The EMA data were sampled at 500 Hz and were then submitted to low-pass filtering with a filter cutoff of 50 Hz.

The relevant $C_1V_1\#C_2V_2$ portion of the audio recording was transcribed, with the aid of an acoustic display, by two trained ToBI transcribers (one the author) following the criteria in the ToBI transcription system (Beckman and Elam 1997). The two transcribers identified the same locations of pitch accent in every token of the entire dataset. The only difference between the transcribers came from a choice between IP and ip. Because the difference is an important experimental variable in this study, only tokens whose renditions were agreed on by the two transcribers were used. (There was a 96.3% agreement between the two transcribers in distinguishing ip and IP boundaries.)

2.1. Measurements

To obtain lip opening and closing movement data, horizontal and vertical position signals for two lip sensor coils are combined into one dimension, Lip Aperture. The Euclidean distance between these two sensor coils is used as an index of Lip Aperture (cf. Byrd and Saltzman, 1998). The derived signal serves as the basis for all the lip measurements. The onset and target timepoints of the lip closing and opening movements were determined from the zero-crossings in the velocity signal with a velocity noise window, defined as 5% of the highest peak velocities of each lip opening and closing movements across the entire dataset. This procedure was done separately for each of the six speakers.

Various dependent variables are calculated at/between the moments of movement onset, target, and peak velocity. The measured variables that are examined in this paper are schematized in Fig. 2. There are three movement events (opening-closing-opening) and five different measures are made for each lip opening and closing movement: (a) *displacement (mm):* the spatial difference between the onset and the target (C_1-to-V_1 lip opening displacement; V_1-to-C_2 lip closing displacement; and C_2-to-V_2 lip opening displacement); (b) *total movement duration*: the temporal interval from the onset to the target (C_1ONS-TO-V_1TARG; V_1ONS-TO-C_2TARG; and C_2ONS-TO-V_2TARG); (c) *time-to-peak-velocity* (acceleration duration): the temporal interval from the onset to the timepoint of peak velocity (C_1ONS-TO-V_2PKVEL; V_1ONS-TO-

C_2PKVEL; and C_2ONS-TO-V_2PKVEL); (d) *deceleration duration*: the interval from the timepoint of peak velocity to the target (C_1PKVEL-TO-V_2TARG; V_2PKVEL-TO-C_2TARG; and C_2PKVEL-TO-V_2TARG); (e) *peak velocity*: the actual peak velocity value for each opening and closing lip movement.

Based on these measured variables, the relationships between some of them were examined further, in order to investigate detailed dynamical aspect of prosodic effects as discussed above (see Fig. 1).

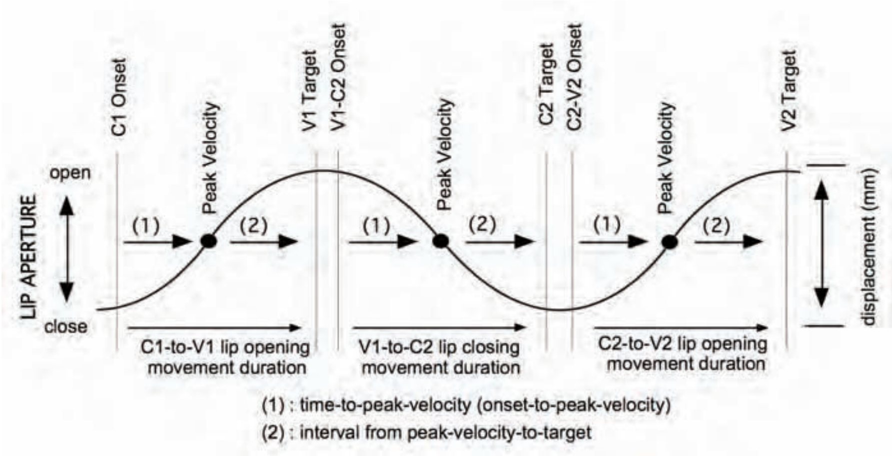


Figure 2. Schema of the lip opening and closing movement trajectory with an indication of the measured kinematic variables.

The systematic influence of prosodic factors on lip opening and closing gestures were evaluated, based on repeated measures Analyses of Variance (ANOVAs). The within-subject factors are V_1 Accent (ACC, UNACC), V_2 Accent (ACC, UNACC), and Boundary Type (IP, ip, Wd). The results are reported based on three-way ANOVAs performed separately for /a/ and /i/. To avoid violating the sphericity assumption (for Boundary Type with more than two levels), a Huynh-Feldt corrected degree of freedom was used in generating F ratio and p values. Next, for relationships between kinematic variables, simple regression analyses were performed. Since we are interested in overall patterns across speakers, and since each speaker had different magnitude of absolute measurements, data were normalized across speakers by transforming measured kinematic values into percentages. This returns for each datapoint that datapoint's percentage contribution to the sum of the entire dataset, which makes the variables more comparable across speakers.

3. Results

This section reports on the effects of Accent and Boundary Type on the kinematics of (1) a preboundary (domain-final) C_1-to-V_1 lip opening gesture, (2) a postboundary (domain-initial) C_2-to-V_2 lip opening gesture, and (3) a cross-boundary V_1-to-C_2 lip closing gesture. It should be noted that in this paper, only overall effects across speakers are reported. In general, for statistically significant findings reported in this study, speakers showed similar patterns. Due to the space limit, this paper will focus on the main effects. (For further details, readers are invited to refer to Cho (2002)).

3.1. Accent effects on kinematics

Let's first look at the results of lip *opening* movement. The basic finding is that lip opening movements in accented CV sequences are characterized by an increase, relative to unaccented sequences, in all measured variables (i.e., displacement, total movement duration, time-to-peak-velocity, deceleration duration, peak velocity), as summarized in Table 2. That is, when accented, lip opening movements are simply bigger in all ways regardless of position-in-domain (final vs. initial) and vowel type (/bi/ vs. /ba/).

Table 2. Summary of effects of accent on kinematics for domain-final C_1-to-V_1 and domain-initial C_2-to-V_2 lip opening gestures. The description in each cell (e.g., larger, longer, higher) is based on main effects, showing a pattern for the accented CV vs. the unaccented counterpart.

Kinematic measures	domain-final (C_1-to-V_1#)		domain-initial (#C_2-to-V_2)	
	/ba#/	/bi#/	/#ba/	/#bi/
Displacement	larger F=19.98**	larger F=88.18**	larger F=29.541**	larger F=41.82**
Total Movement Duration	longer F=37.41**	longer F=74.75**	longer F=79.40**	longer F=60.77**
C_1ONS-To-V_1PKVEL (TIME-TO-PEAK-VEL)	longer F=44.18**	longer F=103.8**	longer F=51.61**	longer F=61.27**
V_1PKVEL-To-V_1TARG (DECELERATION)	longer F=16.32**	longer F=38.07**	longer F=54.86**	longer F=43.98**
Peak Velocity	higher F=15.92**	higher F=62.58**	higher F=17.81**	higher F=37.69**

(** p<0.01, degrees of freedom = F[1,5])

Turning to lip *closing* movement, as shown in Table 3, V_1-to-C_2 lip closing gestures are influenced by both V_1 and V_2 Accent factors. With respect to the V_1 Accent effect, the results show patterns similar to those of lip opening gestures: V_1-to-C_2 lip closing gestures are associated with an increase in displacement, duration, time-to-peak-velocity, and peak velocity (showing a larger, longer, and faster movement).

As for the V_2 Accent effect, it also influences V_1-to-C_2 lip closing gestures with respect to several kinematic measures, but in a way that is somewhat different from V_1 Accent. First, V_1-to-C_2 lip closing gestures before accented V_2 are larger, but not faster, showing an increase in displacement for both /a#b/ and /i#b/, but this time with no change in time-to-peak-velocity. Second, there is a significant change in deceleration duration as a function of V_2 Accent, which was not the case for the effect of V_1 Accent. Finally, the total movement duration is not consistently longer when V_2 is accented: Only /a#b/ (not /i#b/) shows an increase in duration together with an increase in peak velocity.

Table 3. Summary of effects of V_1 and V_2 accent on kinematics for V_1-to-C_2 lip closing gesture. The description in each cell (e.g., larger, longer, higher) is based on significant main effects, showing a pattern for the accented CV as compared with the unaccented counterpart.

	When V_1 accented		When V_2 accented	
	/a#b/	/i#b/	/a#b/	/i#b/
Displacement	larger F=8.944*	larger F=37.882**	larger F=28.251**	larger F=8.691*
Total Movement Duration	longer F=6.354*	longer F=20.384**	longer F=6.414*	*n.s.*
V_1ONS-To-C_2PKVEL	longer F=6.567*	longer F=19.172**	*n.s.*	*n.s.*
VC_2PKVEL-To-C_2TARG	*n.s.*	*n.s.*	longer F=32.477**	longer F=11.833*
Peak Velocity	higher F=4.924*	higher F=8.923*	higher F=11.431*	*n.s.*

(** $p<0.01$, * $p<0.05$, ^{tr} $p<0.07$; degrees of freedom = F[1,5])

Dynamical aspects of Accent. One of the underlying assumptions in a task dynamics model is that distinct kinematic patterns that might arise from linguistic factors can be characterized by different settings of a specific dynamical parameter. Under this assumption, there arises a question as to what dynamical parameter can best characterize accent-induced kinematic variation. Some investigators (Edwards, et al. 1991; Beckman, et al. 1992; Harrington, et al. 1995) have already suggested that accent-induced kinematic variation in jaw opening movements is best captured by a single dynamical parameter, *intergestural timing*. However, we found no evidence that this intergestural timing account or any other parametric account can be extended to the accent-induced kinematic patterns.

Let's first consider lip opening movements. Regarding ACC/UNACC differences in lip opening gestures, we found that an accented lip opening gesture is associated with an increase in all kinematic parameter values (the longer, larger, faster pattern). When these results are compared to kinematic consequences of various mass-spring parameter manipulations, there seems to be no single specific mass-spring parameter that can account for ACC/UNACC differences (compare with the predictions in Fig. 1): (a) If intergestural timing were the only dynamical parameter, we would have observed an increase in both displacement and duration but no change in time-to-peak-velocity and peak velocity; (b) If gestural target (or underlying amplitude) were the only dynamical parameter, there would have been no change in total movement duration and time-to-peak-velocity; (c) In a pure change in stiffness, there would have been no change in displacement but a decrease in peak velocity for accented gestures; (d) Finally, in a pure change by shrinking, there would have been no change in peak velocity. However, none of these idealized descriptions matches the results presented here. Furthermore, relationships between various kinematic variables revealed that no particular dynamical parameter setting can be singled out as an absolute dynamical mechanism underling ACC/UNACC kinematic differences. (Due to the space limit, regression plots are not shown here. For a full description of the data, please see Cho, 2002.)

Next, for the cross-boundary V_1-to-C_2 *lip closing* gesture, kinematic patterns are different depending on the source of Accent (preboundary vs. postboundary) and Vowel Type. On the one hand, the effect of preboundary (V_1) accent shows a pattern similar to the effect of accent on the lip opening gesture, favoring no dynamical account. On the other hand, the fact that only the second component of the total duration (i.e., deceleration duration) is influenced by V_2 Accent appears to support the intergestural timing account, which is especially true for /i#b/ with no change in peak velocity and time-to-peak-velocity. (Recall that the patterning of no change in peak velocity along with an increase in displacement fits the descriptions of a delayed intergestural timing (see Fig. 1)). However, this intergestural timing account is critically weakened for /a#b/ which shows a change in peak velocity (the larger, the faster). The kinematic relationships also show that there is substantial overlapping between ACC and UNACC datapoints, not matching any idealized pictures for a pure change in any particular dynamical parameter. (Again, figures are not provided.)

3.2. Boundary effects on kinematics

C_1-to-V_1# (domain-final) lip opening gesture. The pattern of kinematics in common to both /ba#/ and /bi#/ is that a C_1-to-V_1 lip opening gesture before a higher boundary is associated with an increase in total movement duration, time-to-peak-velocity and deceleration duration, with no increase in peak velocity, i.e., showing a longer, but neither faster nor slower movement. Statistical results are summarized in Table 4. Furthermore, although there is no main effect of Boundary on displacement for /ba#/ (Table 4), there is a significant Accent x Boundary interaction ($F[1.3,6.4]=5.99$, $p<0.05$) because of a pattern of IP>(ip=Wd) only when /ba#/ is *unaccented* (Bonferroni/Dunn *posthoc* test). That is, the C_1-to-V_1 lip opening gesture is generally larger (increased displacement) before a higher boundary for both /ba#/ and /bi#/ except when /ba#/ is accented.

#C_2-to-V_2 (domain-initial) lip opening gesture. As in the case of C_1-to-V_1 (domain-final) lip opening gesture, a C_2-to-V_2 lip opening gesture after a higher boundary is associated with an increase in total movement duration, time-to-peak-velocity (C_1ONS-To-V_1PKVEL) with no increase in peak velocity, again showing a longer, but neither faster nor slower movement. This time, however, there is no effect of Boundary on deceleration duration, suggesting that the temporal effect lies primarily in the first component of the duration (i.e., time-to-peak-velocity).

With respect to displacement, there is no main effect of Boundary. However, there is a significant Boundary x Accent interaction for both /#ba/ and /#bi/ ($F[1.6,8.1]=6.345$, $p<0.025$; $F[2,10]=4.65$, $p<0.04$, respectively). One noteworthy point drawn from Bonferroni/Dunn *posthoc* tests is that unaccented /#ba/ shows an increase in displacement before a higher boundary when /#ba/ is *unaccented* (IP>Wd, $p<0.01$).

V_1-to-#C_2 (transboundary) lip closing movement. As summarized in Table 5, there is systematic boundary-induced kinematic variation in all measured kinematic variables. V_1-to-C_2 lip closing gestures at a higher prosodic boundary show a progressive increase in displacement, total movement duration,

time-to-peak-velocity and deceleration duration, but a progressive decrease in peak velocity. This pattern holds for both vowels.

Table 4. Summary of boundary effects on C-to-V lip opening movements. The results of posthoc tests (p<0.01) is provided when there is a main effect.

Kinematic measures	domain-final (C_1-to-V_1#)		domain-initial (#C_2-to-V_2)	
	/ba#/	/bi#/	/#ba/	/#bi/
Displacement	$F_{[1.2,6.0]}=0.71^{n.s.}$ (IP=ip)>Wd (when unaccented)	$F_{[1.1,5.8]}=6.40$* (IP=ip)>Wd	$F_{[1.3,6.9]}=1.71^{n.s.}$ —	$F_{[1.1,5.6]}=0.99^{n.s.}$ —
Total Duration	$F_{[1.1,5.6]}=23.85$** IP>ip>Wd	$F_{[1.6,8.0]}=66.14$** IP>ip>Wd	$F_{[1.8,9.3]}=16.53$** IP>(ip=Wd)	$F_{[2,10]}=9.91$** IP>(ip=Wd)
C_1ONS-To-V_1PKVEL	$F_{[1.2,5.9]}=8.63$* (IP=ip)>Wd	$F_{[1.8,9.1]}=49.29$** IP>ip>Wd	$F_{[1.9,9.4]}=25.85$** (IP=ip)>Wd	$F_{[2,10]}=17.02$** IP>ip, IP>Wd
V_1PKVEL-To-V_1TARG	$F_{[2,10]}=8.05$** IP>ip>Wd	$F_{[1.3,6.6]}=38.79$** IP>ip>Wd	$F_{[1.1,5.8]}=2.02^{n.s.}$ —	$F_{[1.1,5.9]}=1.01^{n.s.}$ —
Peak Velocity	$F_{[1.1,5.9]}=2.10^{n.s.}$ —	$F_{[1.2,6.0]}=3.02^{n.s.}$ —	$F_{[1.3,6.5]}=1.97^{n.s.}$ —	$F_{[1.2,6.1]}=1.02^{n.s.}$ —

Table 5. Summary of boundary effects on V_1-to-#C_2 lip closing movements. The results of posthoc tests (p<0.01) is provided when there is a main effect.

	/ba#/	/bi#/
Displacement	$F_{[1.3,6.6]}=9.05$* (IP=ip)>Wd	$F_{[1.4,7.1]}=15.777$* IP>ip>Wd
Total Duration	$F_{[1.4,7.2]}=20.018$** IP>ip>Wd	$F_{[1.2,6.6]}=66.654$** IP>ip>Wd
V_1ONS-To-C_2PKVEL	$F_{[1.4,6.7]}=15.289$** IP>ip>Wd	$F_{[1.2,6.1]}=55.448$** IP>ip>Wd
C_2PKVEL-To-C_2TARG	$F_{[2,10]}=35.001$** IP>ip>Wd	$F_{[2,10]}=37.049$** IP>ip>Wd
Peak Velocity	$F_{[1.7,8.7]}=32.754$** IP<ip<Wd	$F_{[2,10]}=5.978$** IP<ip<Wd

Dynamical aspects of boundary effects. As was the case for Accent effects, the boundary-induced kinematic variations are not fully accounted for by any single dynamical parameter setting. First let's consider lip opening movements. There are some close cases in which the kinematic patterns suggested by ANOVA match the shrinking account, showing the requisite larger and longer movement with no change in peak velocity, especially for domain-final kinematic patterns. The shrinking account for *domain-final* cases appears to be further supported by kinematic relationships: (a) a close relationship between duration and displacement/velocity ratio with a remarkable separation among boundary types (R^2 = 0.82 to 0.89 for /ba#/; R^2 = 0.85 to 0.97 for /bi#/), which matches the idealized pattern of a change in shrinking (Fig. 1h); (b) a close relationship between total movement duration and time-to-peak-velocity (R^2 = 0.73 to 0.78 for /ba#/; R^2 = 0.86 to 0.96 for /bi#/). (Note that although Byrd and Saltzman (1998) used this temporal relationship as an index of the degree of stiffness, the close relationship between total movement duration and time-to-peak-velocity also supports the shrinking account because the re-scaling involves a proportional change between the two measures.) However, a close examination of the relationship between peak velocity and displacement, as shown in the left panel of Fig. 3 reveals that the shrinking account is not an absolute fit to the observed pattern, not even domain-finally. If the kinematic pattern were due to a pure change in shrinking, datapoints would be horizontally scattered (see the idealized picture in Fig. 1g), showing a distinct separation among boundary types, which is not what we observe in the figure.

Domain-initially, the evidence for the shrinking account becomes even less clear because of quite a substantial overlap among points belonging to different boundary types (figures not shown). Again, the relationship between peak velocity and displacement (the right panel of Fig. 3) reinforces this by showing substantial overlap among datapoints, rather than the idealized horizontal distribution of datapoints.

Instead, an interesting pattern emerges from the plots in Fig. 3, especially for domain-final cases: the data points are generally scattered diagonally with the datapoints for IP clustering beneath the regression line to the right, and the datapoints for Wd clustering above the regression line to the left. This pattern appears to indicate that some kind of complicated, yet, systematic kinematic mechanism is involved in marking prosodic boundaries, though not accounted for by any single dynamical parameter setting.

Now, let us move on to lip closing (V_1-to-C_2) movements. As seen earlier, V_1-to-C_2 closing gestures at a higher boundary show progressive increase in displacement, total movement duration, time-to-peak-velocity and deceleration duration, but a progressive decrease in peak velocity. This holds for both vowels. The pattern of a larger, longer, and *slower* movement does not single out any particular dynamical parameter as an underlying mechanism. For example, while the patterning of the longer duration with a lowered peak velocity favors the stiffness account, the systematic variation in displacement requires a further dynamical mechanism which cannot be pinpointed here.

Relationships between kinematic measures show evidence that might favor the stiffness account to some extent. The longer and slower movement for V_1-to-C_2 lip closing gestures at a higher prosodic boundary might be accounted for by a decrease in stiffness, as evident in: (a) a close relationship between total movement duration and time-to-peak-velocity (R^2=0.88–0.97); (b) a close relationship between duration and displacement/velocity ratio (R^2=0.88–0.97) with datapoints for a higher prosodic boundary gathering towards the upper right corner of the regression space (bearing resemblance to the idealized picture in Fig. 1b); and (c) datapoints for a higher prosodic boundary being scattered in the lower side of the regression space that relates peak velocity and displacement (again bearing some resemblance to the idealized picture in Fig. 1a, but for the actual plots, see Cho, 2002). While these results indicate apparent temporal aspects that support the stiffness account to some extent, however, the systematic change in displacement (the higher the prosodic boundary, the larger the movement) adds a great deal of dynamic complexity, which again makes it difficult to pinpoint a unified dynamical account.

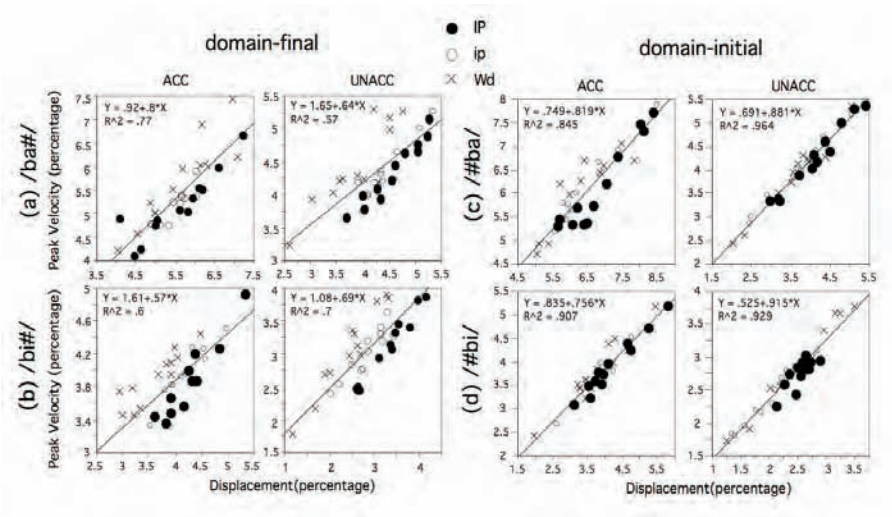


Figure 3. Effect of Boundary Type on relationship between peak velocity and displacement for lip opening gestures, by vowel Type and Accent.

4. General Discussion

Accent-driven kinematic characteristics. In this study, we found that the *lip opening* gesture under accent is associated with an increase in almost all measured kinematic variables including displacement, total movement duration, time-to-peak-velocity, deceleration duration, and peak velocity, regardless of whether it is domain-initial or domain-final. This indicates that the accent-induced articulatory strengthening can be further characterized with a larger, longer, and faster lip opening movement. This result is consistent with findings for jaw opening gestures under accent reported in the literature (as in English *put* reported in de Jong 1991; and in *Pope* and *pipe* in Fowler 1995), but not with those reported in Beckman, et al. (1992) who found that under accent, the jaw opening gesture is associated with an increase in duration and displacement without a substantial increase in peak velocity. The acc/uacc differences are also in line with those coming from lexical stress (Kelso, et al. 1985 for jaw and lower lip movements for reiterant /ba/).

With respect to the *lip closing* (V_1-to-C_2) gesture, one of the significant findings is that the V_1-to-C_2 lip closing gesture is influenced not only by postboundary (V_2) accent but also by preboundary (V_1) accent. Some measured variables are affected primarily by preboundary accent and some, by postboundary accent, while yet others are affected by both. For example, spatial displacement is significantly affected by both V_1 Accent and V_2 Accent, such that lip closing displacement is larger for ACC than for UNACC, regardless of whether accent comes from the preboundary or the postboundary syllables. As for durational variation, the preboundary accent affects primarily the first durational component (time-to-peak-velocity) of the movement duration, whereas the postboundary accent affects only the second component (deceleration duration) of the movement duration. Finally, while peak velocity is consistently influenced by preboundary (V_1) accent such that it is higher for ACC than for UNACC, there is no consistent effect of postboundary (V_2) accent on peak velocity. In short, although there are some compounding effects of Accent arising from both sides of the boundary, the effects of both V_1 and V_2 Accents converge on a larger and longer lip closing movement, while a faster movement comes primarily from V_1 Accent. This is generally consistent with accent-induced kinematic characteristics for the lip opening gesture.

As an aside, a noteworthy point concerns whether the V_1-to-C_2 lip closing gesture should be considered solely as a postboundary phenomenon. Byrd and Saltzman (1998) define the V_1-to-C_2 lip closing gesture as being postboundary because it is activated in order to form a lip constriction for $\#C_2$

which belongs to the postboundary syllable. According to this account, the V_1-to-C_2 lip closing gesture should perhaps be influenced only by postboundary (V_2) accent. However, the fact that some measured kinematic variables are affected only by preboundary accent and some only by postboundary accent suggests that kinematic variation for the V_1-to-C_2 lip closing gesture may be better defined as a *transboundary* phenomenon rather than as a postboundary phenomenon. From the results, we can make a generalization that the V_1-to-C_2 lip closing gesture can be thought of as consisting of two components, with the timepoint of peak velocity as a landmark. First, the articulation during the time course from V_1 onset to the peak velocity landmark may be characterized as a preboundary phenomenon which is governed by the preboundary accent. Second, the articulation during the time course from the peak velocity landmark to the C_2 target attainment may be thought of as a postboundary phenomenon which is governed by the postboundary accent. (Note also that if we apply this transboundary nature of the C_2 lip closing gesture to the framework of syllable structure, we can further posit that the C_2 lip closing gesture is "ambisyllabic" in that the first half of it belongs to the preceding syllable and the second half of it to the following syllable.)

Boundary-driven kinematic characteristics. The one obvious kinematic characteristic for both lip opening and closing gestures at edges of prosodic domains is that they are consistently longer, but this time, not necessarily faster. However, there is an inconsistent boundary effect on spatial displacement in lip opening and closing gestures. The larger displacement was found consistently for the cross-boundary V_1-to-C_2 lip closing gesture, showing a progressive increase in displacement as the prosodic boundary moves up in the prosodic hierarchy. This is consistent with results reported in Byrd and Saltzman (1998). On the other hand, domain-edge lip opening gestures show some interaction between Accent and Vowel Type. (As pointed out by a reviewer, the kinematic difference between lip opening and closing gestures may be in part due to some physiological reasons: there is lip compression at the end of a closing gesture which is likely to change articulatory patterns.)

As for the *domain-final* lip opening gestures, /bi#/ shows an increase in displacement before a higher prosodic boundary, but /ba#/ shows such an effect only when it is *unaccented*. Similarly, for the *domain-initial* lip opening gesture, there is an increased displacement after a higher boundary only for *unaccented* /ba#/. There is thus an increased displacement at a higher boundary at least when the target gestures are unaccented. This is presumably because of some sort of ceiling effect due to accent, that is, when gestures are accented, articulation is already expanded such that an expanded articulation would not leave much room for an additional articulatory expansion from boundary type.

At first glance, this result appears to be consistent with previous findings (e.g., Edwards, et al., 1991) whereby an expanded jaw opening displacement was found domain-finally, only when the gestures being compared are *unaccented*. Recall, however, that some of our results show an expanded lip opening displacement when *accented* (e.g., final /bi#/ and transboundary /a#b/ and /i#b/), suggesting that boundary-induced spatial expansion is not limited to the unaccented gestures only. (However, it should be noted that the difference between the present study and previous studies may be due to the articulators that have been examined. See below for discussion on limitations of lip kinematics.)

All in all, the results suggest that there is some sort of articulatory strengthening as evident by the longer and sometimes larger lip opening and closing gestures at a higher prosodic boundary. However, the boundary-induced strengthening pattern is somewhat different from that arising from accent in that the latter is associated with a faster movement whereas the former is not. Further, this pattern, especially the longer opening movement duration, is found not only in domain-final but also in domain-initial positions. As discussed above, while Byrd and Saltzman (1998) did not consider domain-initial lip opening gestures (thus, for example, it may not undergo lengthening), the present study suggests that the domain-initial lip opening gesture has temporal characteristics much the same as the domain-final lip opening gesture.

4.1. Can prosodically-driven kinematic variations be modeled by a particular dynamical parameter setting?

Accent-driven kinematic variations. Some researchers (e.g., Edwards, et al. 1991; Harrington, et al. 1995) have proposed that an intergestural timing mechanism underlies accent-induced kinematic variation in jaw opening and closing gestures. However, when the present kinematic findings regarding lip opening gestures were considered, no single dynamical parameter setting could be singled out as the underlying mechanism. For the lip opening gesture, the longer and larger movement pattern suggested that, if anything, a change in both stiffness and target was the more probable account for accent-induced differences, with a change in intergestural timing as the least likely mechanism. (Note that findings reported by de Jong (1991) and Fowler

(1995) also suggested that if anything an increase in target is the most likely source for an increased displacement.) For the cross-boundary V_1-to-C_2 lip closing gesture, the effect of preboundary (V_1) accent shows a pattern similar to the effect of accent on the opening gesture, favoring no dynamical account. The longer and larger articulation (with no change in peak velocity and time-to-peak-velocity) due to postboundary accent for /i#b/ seems to be ascribable to a change in intergestural timing. Again, however, relationships between various kinematic variables did not support this, weakening the intergestural timing account.

What emerges from the data is then that no single dynamical mechanism can account for accent-induced kinematic variations, contrary to what has previously been assumed among researchers who have attempted to characterize prosodically-conditioned kinematic variations in terms of a mass-spring dynamical parameter setting.

Boundary-driven kinematic variations. As was the case for Accent effect, the boundary-induced kinematic variations were not fully accounted for by any single dynamical parameter setting. If we consider only temporal kinematic measures, namely, the total movement duration and time-to-peak-velocity, as Byrd & Saltzman (1998) did, the boundary-induced durational difference is likely ascribable to a change in stiffness, given the proportional change in the total movement duration and time-to-peak-velocity as a function of prosodic boundary. However, when we consider additional kinematic measures, the stiffness hypothesis is seriously undermined. For instance, when peak velocity (which was not included in Byrd & Saltzman) is considered, only the lip *closing* gesture shows a slower movement (with lowered peak velocity) at a higher prosodic boundary, favoring the stiffness account, whereas no change in peak velocity in the case of the lip *opening* gestures weakens the stiffness account. Moreover, when the variation in displacement is figured in, it becomes even more obvious that a change in stiffness is not the only dynamical mechanism underlying the boundary-induced longer, larger, and sometimes slower movement.

Here, it is worthwhile noting two possible sources of variation in displacement. Both Byrd & Saltzman and the present study have measured the displacement of the lip closing gesture by differentiating the lip opening maxima and minima. However, as pointed out by Goldstein (p.c.), the displacement in lip opening in V_1#C_2 may vary not only due to a change in the target of the lip closing gesture but also due to a change in the value of the Lip Aperture at the onset of the gesture associated with the preceding vocalic gesture. This becomes clearer with the results of the present study regarding accent-induced variation in displacement. It was found that the V_1-to-C_2 lip closing gesture was associated with an increase in displacement when either V_1 or V_2 was accented. It is therefore possible that when V_1 was accented, the increased displacement was mainly due to the more extreme opening value at the onset of lip closing while the increased displacement due to accented V_2 was primarily attributable to the more extreme target value obtained during C_2 constriction. In other words, as Cho (2002) discussed, the reason for effects of both V_1 Accent and V_2 Accent is presumably because the maximum Lip Aperture for V_1 is significantly larger for V_{1ACC} than V_{1UNACC} ($p<0.01$), and the minimum Lip Aperture for C_2 is significantly smaller for V_{2ACC} than for V_{2UNACC} ($p<0.01$).

One might then question whether measuring the spatial difference (displacement) between the onset and the offset of the V_1-to-C_2 movement adequately reflects the target (gestural amplitude) of the relevant dynamical system (here, the lip closing gesture). For example, the increased displacement at a higher prosodic boundary found in this study may not exclusively reflect the change in target in the dynamical system. Cho (2002) indeed reported that Lip Aperture maxima for preboundary V_1 were generally larger at a higher prosodic boundary than at a smaller prosodic boundary. Thus, it requires caution to interpret boundary-induced kinematic variation in displacement in terms of a dynamical parameter setting, for both the present study and Byrd & Saltzman 1998.

Some discussion on the relationship between kinematics and dynamics. With all these in mind, let us return to the issue of how the kinematic results can be accounted for in the framework of dynamics. One might raise a rather fundamental question about the validity of the current mass-spring dynamical model. If the dynamical model were assumed to predict that modification to a single dynamical parameter is the only way to control kinematics, then the failure to single out any particular dynamical parameter setting would suggest that the current mass-spring dynamical model is not adequate to account for the prosodically-induced kinematic patterns. With respect to accent-induced kinematic pattern, Fowler (1995) indeed proposed that gestural behaviors under accent may not be best described in terms of dynamical parameter settings, but rather they are most consistent with the "global effect" hypothesis that stress consists of a global increase in *production effort* in order to maximize prominence in the stressed syllable. Such prominence maximization can then be obtainable simply by the larger, longer, and faster lip opening and closing movements, as found in this study.

Alternatively, speech mechanisms may not be as simple as has been assumed by researchers who adopt the mass-spring dynamical model in explaining certain speech phenomena. The observed data could be explained under a mass-spring dynamical model, if we further explore the possibility that more than one dynamical parameter governs the accent-induced kinematic patterning. For example, from the present study, one might infer that both stiffness and target changes govern the lip opening movements under accent, and the lip closing movement under V_1 accent, whereas changes in both stiffness and intergestural timing likely underlie the lip closing movement under V_2 accent. Likewise, the consistently larger displacement in the lip closing movement at a higher prosodic boundary can be dealt with by either the target or the intergestural timing parameters in combination with the stiffness parameter that accounts for the observed temporal aspects. Further, we cannot entirely reject the possibility that all the dynamical parameters are interactively influential on kinematic realizations with different degrees of effect, such that breaking down such compounding effects into individual dynamical parameter settings would be extremely difficult without fine-grained computational modeling on ample empirical data.

Finally, there is another caveat interpreting the kinematic data of the present study in dynamical terms. In this study, following Byrd & Saltzman (1998), the lip opening and closing movements (Lip Aperture) have been assumed to be regulated by a single dynamical regime (gesture). While it is reasonable to assume that lip closing is controlled by a single dynamical system (i.e., a consonantal lip closing gesture), it is less clear whether the lip opening (e.g.,. C_1-to-V_1 and C_2-to-V_2 movements) is indeed modulated by a single dynamical gesture (Goldstein, p.c.). Lip opening movements are usually associated with a vocalic gesture which may regulate tongue task variables primarily, and Lip Aperture may be influenced by not only the tongue movement but also the action of the jaw which accompanies it. Therefore, the failure to interpret lip opening kinematics in terms of dynamics may be attributable in part to such articulatory complexity associated with the lip opening during the vocalic movement.

At the very least, however, the findings in the present study motivate future studies to look for not only the complexity of dynamical parameter settings but also articulatory complexity associated with a single gesture, rather than seeking what particular dynamical parameter setting 'best' matches speech phenomena. However, even if practicing linguists can develop such a complicated model (building on the currently available dynamical model) which can adequately describe all the complex kinematic patterns as presented in the present paper, it will still be interesting to see how such a complex dynamical system is learned in the course of the language acquisition.

The π-gesture. Another way of characterizing boundary-adjacent kinematic variation is suggested by Byrd and Saltzman (Saltzman 1995, Byrd, et al. 2000; Byrd 2000; Saltzman & Byrd 2000; Byrd & Saltzman 2003): there might be abstract, non-tract variable prosodic boundary gestures that are governed by prosodic constituency in a mass-spring dynamical model. The so-called 'π-gesture' was hypothesized initially to affect stiffness in tract variable articulatory gestures over its activation period, roughly in proportion to the strengths of the boundary: the larger the prosodic boundary, the less stiff the articulatory gestures in the vicinity of the boundary. In Byrd & Saltzman (2003), the stiffness modulation approach was replaced with the clock-slowing modulation approach: the π-gesture locally slows the clock that controls the timecourse of gestural activation. In this framework, boundary-induced temporal variation can be interpreted as a change in clock-slowing under the influence of the π-gesture. The temporal activation interval of the π-gesture overlaps with the activation interval of articulatory gestures adjacent to prosodic boundaries, such that the boundary-adjacent articulation lengthens in proportion to degree of the π-gesture's strength, which is again roughly proportional to level of prosodic boundary.

Degree of lengthening is also influenced by the temporal extent of the π-gesture. In an earlier model of π-gesture, Byrd (2000) suggested that the π-gesture's domain of influence is local at edges of prosodic domains – i.e., "only the constriction gestures within the π-gesture's temporal field of activation are directly affected, not gestures remote from the phrasal boundary (p. 14)." Thus, Byrd hypothesized that for the sequence $C_1V_1\#C_2V_2$, articulations that are closest to the prosodic boundary are most influenced by the π-gesture, resulting in the maximal elongation. In the present study, however, it was found that not only V_1-to-C_2 movement (which is the closest to the prosodic juncture) but also C_1-to-V_1 and C_2-to-V_2 movements were all significantly affected by boundary type. Of course, it is likely that the lengthening of C_1-to-$V_1\#$ comes from the effect of the π-gesture on $V_1\#$ and the lengthening of $\#C_2$-to-V_2, from the effect of the π-gesture on C_2. It is also possible that articulations for the rather remote C_1 and V_2 are still within the activation field of the π-gesture but presumably with somewhat reduced degree of the π-gesture's influence, under the assumption that the π-gesture's strength tapers out towards edges of its temporal activation interval. However, it is not entirely clear what are the exact mechanisms that underlie lengthening of gestures that are not immediately next to a prosodic boundary. Byrd and

Saltzman (2003) explain that "its [π-gesture's] effect will be felt on any of the gestures with which it is coarticulated; under the assumption that the π-gesture is anchored to the prosodic juncture, these will be gestures closest to the phrase edges." Specifically, it is hoped that future studies provide us with more information not only about the precise temporal extent of the π-gesture, but also about its relationship with the declining nature of the π-gesture's strength towards the edges of the activation interval.

Another issue regarding the π-gesture model is whether the π-gesture influences degree of spatial magnitude directly or not. The available information with respect to variation in spatial magnitude comes from simulations (Saltzman and Byrd 2000, Byrd and Saltzman 2003) which demonstrate that a clock-slowing implementation of π-gestures may entail variation in displacement associated with domain-initial consonant constriction. For example, for the domain-initial consonant-vowel constriction sequence, the π-gesture initiates the CV constriction sequence (rather than intervening it) such that under the influence of π-gesture, the consonantal constriction will get not only longer but also overlap less with the following vowel gesture. In the current model, the decreased overlap (or less truncation) between the consonantal gesture and the following vocalic gesture accounts for the domain-initial strengthening phenomenon – i.e., the increase in gestural amplitude associated with domain-initial consonants (e.g., Fougeron and Keating 1997; Cho and Keating 2003). It remains to be seen whether this model would be able to account for the full range of results presented here and elsewhere, and again whether it could provide a simple and unified theory about the prosodically-driven systematicity in speech production.

4.2. Enhanced consonant-vowel contrasts at domain-edges

One of the central issues with respect to boundary-induced kinematic variation is whether it is a linguistically significant phenomenon. It has been suggested in the literature (Fougeron and Keating 1997; Hsu and Jun 1998; Fougeron 2001) that expanded #CV or V#C displacement adjacent to a prosodic boundary would serve as an articulatory signature for marking that prosodic boundary. The present results are generally supportive of this proposal. In particular, the V#C lip closing gesture shows the most robust boundary effect on displacement with a pattern of IP>ip>Wd. A similar result was found for domain-final (CV#) lip opening gesture, whereas the domain-initial (#CV) lip opening gesture did not show a consistent effect. This is compatible with Fougeron and Keating's observation that domain-initial consonantal strengthening, as measured by linguopalatal contact, induces a greater V#C displacement at edges of prosodic domains, while such an effect is less evident in degree of #CV displacement. This observation is reinforced by kinematic data reported in this study. Further, the results presented in this study show that even the domain-final CV# displacement is expanded at higher prosodic boundaries, which Fougeron and Keating did not find in their EPG data. Overall, we can infer that contrasts between consonants and vowels are enhanced at edges of prosodic domains (syntagmatic contrast enhancement) via an increase in displacement adjacent to a prosodic boundary, which can be seen as the articulatory manifestation of prosodic structure. (See Cho and Jun (2000), Cho and McQueen (2005) for discussion on domain-initial strengthening in terms of enhancement of distinctive features.) Recent research has begun to investigate the role of domain-initial consonantal strengthening in lexical segmentation in English (McQueen and Cho 2003; Cho, McQueen and Cox, in press), showing that the acoustic consequences of initial strengthening facilitate word recognition.

4.3. Conclusion

The present study has investigated how segmental phonetic realizations are conditioned by various prosodic factors by examining kinematic variations in accented syllables, domain-initial, and domain-final syllables. While previous studies have looked at these locations separately, the present study differs from them in that it examined all these locations concurrently. Crucially, each of the three prosodically important locations showed distinctive kinematic patterns that can distinguish itself from others. Several major points have emerged. First, accent-induced articulatory strengthening can be characterized by larger, longer, and faster lip opening and closing movements. That is, when accented, vowel movements are simply bigger in all ways – in distance, time, and speed. Second, unlike accent-induced strengthening, boundary-induced strengthening effects are evident in longer, but this time not necessarily faster, articulation in both domain-initial/final positions. The spatial expansion is found quite consistently at the domain edge when the gestures are *unaccented,* and more consistently for cross-boundary lip closing gestures regardless of accent. Further, temporal characteristics are similar for the domain-final and initial lip opening gestures (i.e., longer duration with no change in velocity), though the domain-final gesture is longer than domain-initial one. In short, all of the three prosodically important locations show strengthening effects with generally longer and larger movements in common, but they differ primarily in velocity: faster for accented gestures, no change for domain-initial/final lip opening gestures, and slower for cross-boundary lip closing gestures.

Finally, the results regarding movement kinematics suggest that speech mechanisms are more complex than has generally been assumed. These results challenge the theories previously advanced in the framework of a mass-spring gestural model. It was proposed that in order to account for prosodically-conditioned kinematic patterns in the framework of a mass-spring gestural model, at the very least one should look for a combination of settings for multiple dynamical parameters, rather than seeking one particular dynamical mechanism governing kinematic patterns arising from each prosodic condition. Alternatively, the best solution to the problem might be to find a simple and unified dynamical theory (not necessarily in the framework of a mass-spring gestural model) which can model the prosodically-driven systematicity in a way that is both descriptively and explanatorily adequate.

This study suggests that phonetic realization is governed by high level prosodic conditions, and that prosodically-conditioned kinematic patterns, in turn, manifest high level prosodic structures. Furthermore, the systematic phonetic variation conditioned by prosodic structure should be taken more seriously into account in developing linguistic theories, especially in modeling speech production and speech perception. It is ultimately hoped that this study will contribute to theories of the phonetics-prosody interface, making progress towards gaining better insight into prosodically-driven speech phenomena.

Acknowledgments

I thank Pat Keating, Sun-Ah Jun, Peter Ladefoged, Bruce Hayes, Jody Kreiman, Adam Albright, Elizabeth Johnson and James McQueen for their comments and suggestions. Special thanks goes to Louis Goldstein and two other anonymous reviewers for critical and constructive comments. This work was supported in part by the NSF doctoral improvement grant (BCS-0001716) to the author and Pat Keating.

References

Beckman, M. E.
1996 The parsing of prosody. *Language and Cognitive Processes* 11: 17–67.

Beckman, M. E., J. Edwards and J. Fletcher
1992 Prosodic structure and tempo in a sonority model of articulatory dynamics. In: G. Docherty and D.R. Ladd (eds.), *Papers in Laboratory Phonology II: Gesture, Segment, Prosody,* 68–86. Cambridge: Cambridge University Press.

Beckman, M. E. and G.A. Elam
1997 Guidelines for ToBI labelling, version 3.0. Unpublished Manuscript. Ohio State University.

Beckman, M. E. and J. Pierrehumbert
1986 Intonational structure in Japanese and English. *Phonology Yearbook* 3: 255–309.

Browman, C. and L. Goldstein
1990 Tiers in Articulatory Phonology, with Some Implications for Casual Speech. In: J. Kingston and M. E. Beckman (eds.), *Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech*, 341–376. Cambridge: Cambridge University Press.

Byrd, D.
2000 Articulatory vowel lengthening and coordination at phrasal junctures. *Phonetica* 57: 3–16.

Byrd, D. and E. Saltzman
1998 Intragestural dynamics of multiple prosodic boundaries. *Journal of Phonetics* 26: 173–200.
2003 The elastic phrase: modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics* 31: 149–180.

Byrd, D., A. Kaun, S. Narayanan and E. Saltzman
2000 Phrasal signatures in articulation. In: M. Broe and J. Pierrehumbert (eds.), *Acquisition and the Lexicon: Papers in Laboratory Phonology V*, 70–88. Cambridge, UK: Cambridge University Press.

Cho, T.
2002 *The Effects of Prosody on Articulation in English*. New York and London: Routledge.
2004 Prosodically-conditioned strengthening and vowel-to-vowel coarticulation in English. *Journal of Phonetics* 32: 141–176.
2005 Prosodic strengthening and featural enhancement: Evidence from acoustic and articulatory realizations of /a,i/ in English. *Journal of the Acoustical Society of America* 117(6): 3867–3878.

Cho, T. and J. McQueen
2005 Prosodic influences on consonant production in Dutch: Effects of prosodic boundaries, phrasal accent and lexical stress. *Journal of Phonetics* 33(2): 121–157.
Cho, T., J. McQueen and E. Cox
in press Prosodically driven phonetic detail in speech processing: The case of domain-initial strengthening in English. *Journal of Phonetics.*
Cho, T. and P. Keating
2001 Articulatory and acoustic studies of domain-initial strengthening in Korean. *Journal of Phonetics* 29: 155–190.
Cho, T. and S.-A. Jun
2000 Domain-initial strengthening as featural enhancement: Aerodynamic evidence from Korean. *Chicago Linguistics Society 36* (CLS 36): 31–44.
De Jong, K.
1991 The Oral Articulation of English Stress Accent. Ph.D. Dissertation, Ohio State University.
1995 The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *Journal of the Acoustical Society of America* 97: 491–504.
Edwards, J. E., M. E. Beckman and J. Fletcher
1991 The articulatory kinematics of final lengthening. *Journal of the Acoustical Society of America* 89: 369–382.
Fougeron, C.
2001 Articulatory properties of initial segments in several prosodic constituents in French. *Journal of Phonetics* 29: 109–135.
Fougeron, C. and P. A. Keating
1997 Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America* 101: 3728–3740.
Fowler, C. A.
1995 Acoustic and kinematic correlates of contrastive stress accent in spoken English. In: F. Bell-Berti and J. J. Raphael (eds.), *Producing Speech: Contemporary Issues: For Katherine Safford Harris*, 355–373. Woodbury: American Institute of Physics.
Harrington, J., J. Fletcher and C. Roberts
1995 Coarticulation and the accented/unaccented distinction: Evidence from jaw movement data. *Journal of Phonetics* 23: 305–322.
Hawkins, S.
1992 An introduction to task dynamics. In G. Docherty and D. R. Ladd (eds.), *Papers in Laboratory Phonology II: Gesture, Segment, Prosody,* 9–25. Cambridge: Cambridge University Press.

Hsu, C.-S. and S.-A. Jun
1998 Prosodic strengthening in Taiwanese: Syntagmatic or paradigmatic? *UCLA Working Papers in Phonetics* 96: 69–89.

Jun, S.-A.
1993 The Phonetics and Phonology of Korean Prosody. Ph.D. Dissertation, Ohio State University.
1998 The accentual phrase in the Korean prosodic hierarchy. *Phonology* 15: 189–226.

Keating, P. A., T. Cho, C. Fougeron and C. Hsu
2003 Domain-initial strengthening in four languages. In: J. Local, R. Ogden, and R. Temple (eds.), *Papers in Laboratory Phonology IV: Phonetic Interpretation,* 145–163. Cambridge, UK: Cambridge University Press.

Kelso, J. A. S., E. Vatikiotis-Bateson, E. Saltzman and B. Kay
1985 A qualitative dynamic analysis of reiterant speech production: phase portraits, kinematics, and dynamic modeling. *Journal of the Acoustical Society of America* 77: 266–280.

McQueen, J. and T. Cho
2003 The use of domain-initial strengthening in segmentation of continuous English speech. *Proceedings of the 15th International Congress of Phonetic Sciences (ICPhS)*. Barcelon, Spain 2003.

Pierrehumbert, J. and M. E. Beckman
1988 *Japanese tone structure*. Cambridge, Mass.: MIT Press.

Pierrehumbert, J. and D. Talkin
1992 Lenition of /h/ and glottal stop. G. Docherty and D. R. Ladd (eds.), *Papers in Laboratory Phonology II: Gesture, Segment, Prosody,* 90–117. Cambridge: Cambridge University Press.

Saltzman, E.
1995 Intergestural timing in speech production: Data and modeling. In *Proceedings of the XIII International Congress of Phonetic Sciences 2*: 84–91.

Saltzman, E. and D. Byrd
2000 Demonstrating effects of parameter dynamics on gestural timing. Presented at the May 2000 meeting of the Acoustical Society of America, Atlanta, Georgia. *Journal of the Acoustical Society of America* 107(5,2): 2904.

Shattuck-Hufnagel, S. and A. E. Turk
1996 A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research* 25: 193–247.

Relating prosody and dynamic events: Comments on the papers by Cho and Smiljanić

Dani Byrd

> From the Proceedings of LabPhon I: "We have recently begun a research program with the goal of providing explicit, formal representations of articulatory organization appropriate for use as phonological representations." (Browman and Goldstein 1990: 341)

1. Introduction

At the first meeting of the Conference on Laboratory Phonology in 1987, Browman and Goldstein outlined a model of speech production (Browman and Goldstein 1990) that made three significant claims. First, they proposed that linguistically significant vocal tract constrictions – *gestures* – were the fundamental cognitive unit of speech production. Second, they specifically presented these units as phonological, thereby rejecting a grammatical structure having a mediated relationship between phonology and phonetics in favor of one in which informational and motor task units are isomorphic. Thirdly, they characterized these gestural units as dynamical systems, thereby endowing them with spatiotemporal characteristics, which, they argued, provide the raw materials of phonological patterning. Specifically they addressed the importance of temporal relations between gestural units in understanding variability in surface realization of underlyingly invariant phonological units.

In the 15 years since that time, the field of Laboratory Phonology has achieved a reasonably sophisticated understanding of how articulatory gestures as dynamical systems interact to produce context-conditioned variability. This has been explored by the LabPhon community and others with regard to coarticulation, speech rate variation, casual speech phenomena, and syllable-structure related allophony. However, despite remarkable progress in this past decade, or perhaps because of it, we are now seeing a dichotomy in the treatment of "segmental" and "suprasegmental" phenomena

– that is (in more theoretically appropriate terms) between the treatment of word-level and prosodic phenomena. Prosodic phenomena are still largely characterized non-dynamically as symbolic, categorical, and atemporal. This has brought us to the point of an acute mismatch with well-established views in the LabPhon community as to the nature of segmental articulation.

Under the aegis of Commentary, I submit the following question to the community involved in the next decade of LabPhon: Will an exploration of prosody (i.e., information constituency & prominence) as having dynamical characteristics that interact with gestural dynamics help illuminate variability of words in meaningful contexts? If such an exploration proves fruitful, a greater theoretical consistency in conceptualizing prosodic structure and word-level structure within Articulatory Phonology will be made possible, and we will advance our understanding of the prosody-phonetics interface. The various papers in this session of LabPhon8 provide an opportunity to glimpse this nascent area of LabPhon research, and their intriguing research results should propel the community further in the direction of more sophisticated accounts of prosodically conditioned gestural variability.

2. Gestures as action units

As a foundational basis for considering the dynamics of prosody, it is worthwhile to review briefly the dynamical characterization of gestural units within Articulatory Phonology. Gestures are described as bifunctional; they act as units of information (and as such, the basis of linguistic contrast) and units of action, i.e. vocal tract constriction formation (see Browman and Goldstein 1992, 1995). Because these informational and action units are one and the same, there is no translation between units of representation and their execution as speech tasks. At an organizational level, constriction *tasks* are controlled, rather than the movement of individual articulators (Saltzman and Munhall 1989). This control is achieved by turning on and off point attractors in the task space defined within the vocal tract. Point attractor dynamics are an appealing system type for characterizing the control of discrete movements because of several characteristics: namely, because it is possible to associate a particular point in task space with the target of the discrete movement and because the target is approached regardless of initial conditions or perturbations along the way. This type of perturbation-resilient behavior has been observed for speech gestures in both the spatial and temporal domains (Kelso, Tuller and Fowler 1984, Saltzman, Löfqvist, Kay, Kinsella-Shaw, and Rubin 1998), and the temporal stability has also been modeled with point attractors (Saltzman and Byrd 2000a). The point attractor system typically used to model gestural dynamics is that of a critically damped mass-spring system. Recall the familiar mass-spring equation: $m\ddot{x} + b\dot{x} + k\,(\,x - x_0\,) = 0$, where m is assumed to equal 1; x_0 is the rest position (i.e., constriction target), b is set to critical damping, and k indicates the stiffness term governing gestural speed to target. A gesture's dynamical system controls the vocal tract articulators for a discrete interval of time in an utterance. Further, the *activation* of this control regime waxes and wanes gradually (see e.g., Byrd and Saltzman 1998).

The hypothesis of gestural action units in speech production makes several predictions regarding the ways an utterance can be malformed during production – that is, regarding speech errors – that differ from a symbolic account of errors. A symbolic account views errors as exchanges or substitutions of symbolic units in a plan that is then, after the symbol manipulation, executed normally. The predictions of a gestural account of speech errors are reviewed in Goldstein, Pouplier, Chen, Saltzman, and Byrd (subm). Briefly, if the atomic phonological units are gestures with time-varying activation strengths, errors may result from (partial or complete) activation of a gestural unit at an inappropriate time during production. Such gradient errors may fail to be represented in transcriptional corpora of errors. Suggestions that gradient errors do indeed occur can be found in Mowrey and McKay's (1990) EMG study and Frisch and Wright's acoustic study (2002).

Patterns of speech errors – their systematicity and tendencies – have long been thought to provide important information about the nature of cognitive units in speech production. The data presented by Goldstein et al. (subm.), Mowrey and MacKay (1990), and Frisch and Wright (2002) join the legacy of empirical investigation demonstrating that vocal tract constriction gestures are fundamental units of speech production. Depending on the granularity of analysis, other units such as segments, syllable/word onsets, syllables, and words are also evidenced as playing active roles in speech production. Goldstein and colleague's recent work on bonding relations (Browman and Goldstein 2000, Goldstein and Fowler, in press; see also Byrd 1996 and Saltzman and Byrd 2000a) pursues an account of how the presence and strength of temporal relations among atomic gestural units might be important in the accretion of larger 'molecular' structures.

3. Underlying gestures and surface prosodic variability

This brings us full circle to a foundational LabPhon topic. If words are composed of a small set of stable combinatorial units, how do we account for surface variability that seems to indicate that the physical characteristics of these units are not fixed? A dynamical systems account of phonological gestures offers the advantage of assuming an underlying invariance at the level of control but still accounting for surface variability in performance via gestural interaction (e.g. blending and competition). The output acoustics results from the combined influence of all active gestures (plus the initial conditions, usually the result of previously active gestures.). (For further explication see for example Saltzman and Munhall [1989].) In the first LabPhon proceedings, Browman and Goldstein state, "[T]he gestures are invariant across different contexts…[O]verlapping activation…results in context-varying articulatory trajectories" (1990: 342). Now, a decade later, is has become clear that *other kinds* of surface variability occur in addition to that driven by neighboring segmental/gestural context. Phrasal constituent structure and prominence structure *also* result in articulatory variability. How is this to be captured in a dynamical gestural account of speech production while maintaining the linguistic insight of stable underlying units that are called upon in a variety of prosodic contexts?

A stable lexical representation requires defining a set of gestures, their internal and relative activation patterns, and their underlying parameter values (e.g. target, stiffness). Yet recent laboratory work (e.g. Edwards, Beckman and Fletcher 1991, Beckman and Edwards 1992, Keating, Cho, Fougeron, and Hsu in press, Byrd and Saltzman 1998, Byrd, Kaun, Narayanan, and Saltzman 2000, Byrd 2000) has demonstrated that intra- and inter-gestural timing and magnitude vary as a function of phrasal position. Specifically these studies have shown that gestures exhibit boundary adjacent lengthening, reduced temporal overlap at boundaries, and spatially larger gestures in phrase-initial position. These effects are generally more robust at larger boundaries. Keating and Shattuck-Hufnagel (2002) view prosodic structure and prominence as available to the processes that compute the timing and amplitude of articulatory gestures.

Cho's contributes to this volume a meticulous articulatory investigation into the effects of both prominence and phrasal position on intra-gestural characteristics. In a sequence of the type [[*….opening*][*closing-opening…*]], he finds that under focal accent, gestures become larger, longer, and faster (in peak velocity); at phrasal boundaries, gestures get longer, and, for phrase-initial closings, slower and larger. He comments that temporal characteristics are similar for domain-final and domain-initial openings in the sequence he investigated. When Beckman, Edwards, and colleagues presented the first data on articulatory kinematics at boundaries, lowered gestural stiffness provided a sensible dynamical account capturing the data (e.g., Beckman and Edwards 1992). While it is clear that their original insights were correct in asserting that gestural slowing was a vital aspect of this phenomena, further empirical work such as that of Cho's indicate that the picture is not yet complete. Gestural stiffness lowering alone will not account for the complexity of spatiotemporal patterns observed at phrasal boundaries.

At this point, it is worth noting that these phrasal phenomena (e.g. longer, less overlapped, larger gestures) have received a variety of labels in the Laboratory Phonology literature that attribute various functional roles to these phenomena: consonant-vowel differentiation, lexical-contrast enhancement, initial strengthening, boundary-adjacent slowing, prominence maximization, juncture marking for listeners, planning consolidation on the part of speakers. However, orthogonal to the question of functional motivation is the question of *how* this prosodic action is effected. What is the articulatory foundation(s) of these varied effects observed at boundaries?

One effort, building on the earlier work of Beckman, Edwards, and colleagues, to model gestural variation at phrase edges is the introduction of the π-gesture (prosodic-gesture) into the Articulatory Phonology framework by Byrd, Saltzman, and colleagues (Byrd et al. 2000, Saltzman and Byrd 2000b, Byrd and Saltzman 2003). Under this account, prosodic variability can emerge from lexically-specified dynamics of constriction gestures *interacting* with π-gestures that represent boundaries as opposed to constrictions. This approach is motivated in part by a desire for theoretical consistency with how contextual variability has been successfully handled in the Articulatory Phonology framework. Namely, it preserves stable underlying constriction gestures but allows surface variability to arise through the interaction of co-active gestures, in this case, co-active constriction and prosodic gestures. These π-gestures span a discrete temporal interval during which they act vicariously on all concurrently activated constriction gestures. π-gesture activation is determined by boundary strength and its effect is in proportion to activation level. Since activation waxes and wanes gradually, it is *proximity* to a juncture (yielding overlap and co-activation of π- and constriction gestures), not adjacency to the boundary (i.e., not status as domain final or initial *per se*) that is the source of effect on articulation.

Smiljanić's study in this volume is of interest in a discussion of prosodic events as having temporal (activation) intervals, as suggested in the π-gesture approach. It provides an examination of how prosody affects the coordination of segmental material and lexical pitch accent. Recall from Smiljanić's description that in Belgrade Serbian, a falling lexical accent is indicated by a pitch peak early in the vowel and a rising accent by a pitch peak after (or late in) the vowel. Smiljanić demonstrates that for narrow focus in phrase initial position, the pitch accent peaks earlier. But this is especially the case for the falling (early peak) accent (see her Fig. 1, left panels, note circles vs. squares). She proposes a mechanism of contrast enhancement. In phrase final position under narrow focus, pitch peaks are early. But an examination of her Figure 2 shows that this is mostly just found for the rising (late peak) accent (see her Fig. 2, left panels, note squares vs. circles). She proposes that this is the result of tonal crowding.

Can this apparent asymmetry be understood via a single mechanism if prosodic events are *intrinsically* temporal and overlapping gestural events? Consider the Serbian stimuli as presented schematically in Figure 1, shown with overlapping π-gestures coordinated leftward in the vowel for the phrase-initial situation (left column) and rightward in the vowel for phrase-final position (right column). When viewed in such a schematic that captures temporal relations among laryngeal, oral, and prosodic events, it seems sensible that the early lexical accent (bottom row) should be most affected in phrase initial position (left bottom) and the late accent in final position (right top). This is precisely because the π-gesture has a temporal activation interval coordinated with the articulatory events.

Smiljanić's findings have a further interesting aspect with regard to prosodically-driven dynamical changes at the gestural level. Accent alignment differences between initial and final positions are greater (i.e., dark more to the left of light in her Fig 2, both panels) when final vowel lengthening is greater compared to initial (i.e., dark more above light in her Fig 2 both panels). This suggests a relation between *inter*gestural and *intra*gestural effects in prosodic dynamics. Recall that Cho's study suggested that the intragestural change of stiffness lowering at boundaries was not sufficient to account for his data (see also Saltzman and Byrd 2000b, Byrd and Saltzman 2003). Cho posits that multiple parameter changes with complex interactions might be at work. Byrd and Saltzman (2003 and Saltzman and Byrd 2000b) demonstrate that stiffness scaling cannot account for changes in relative timing or phrase-initial magnitude, two other important effects at boundaries in addition to lengthening. This is because such stiffness adjustment modulates a single point-attractor parameter value but does not specifically influence the domain of gestural activation, which controls the gating in of multiple parameter values. One approach that we've taken is endowing junctures with dynamics at the *activation* level, rather than the parameter level (Byrd and Saltzman 2003). This approach slows the timecourse of gestural activations and has both intra- and inter-gestural timing consequences. The juncture related changes in gestural overlap, in turn, have spatial consequences, though spatial parameters are not manipulated directly.

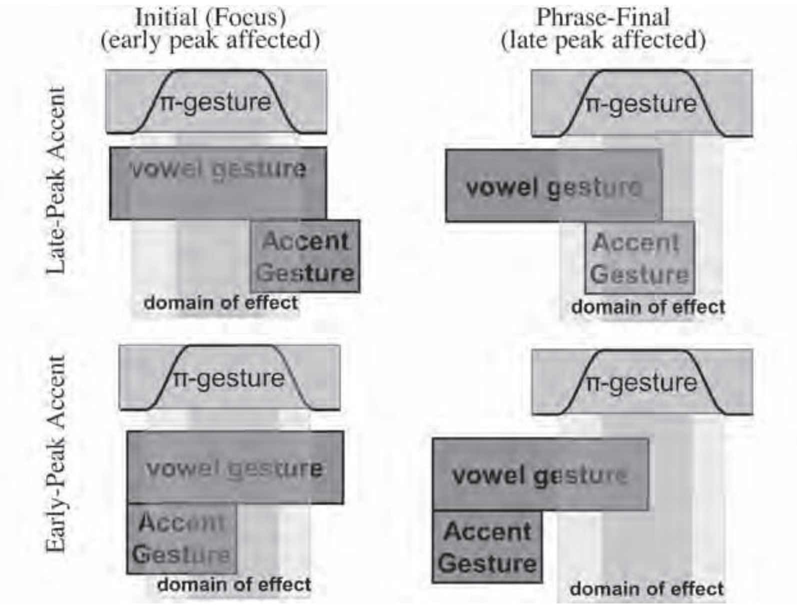


Figure 1. Schematic representation of Smiljanić's Serbian stimuli, shown with overlapping π-gestures coordinated leftward in the vowel for the phrase-initial situation (left column) and rightward in the vowel for phrase-final position (right column). The maximal effect of the π-gestures would be observed for the early-peak accent in the initial condition and the late-peak accent in the final condition.

4. Challenges for a dynamical approach to prosody

Further challenges for dynamical accounts of prosody abound, and one of these revisits the traditional phonology-phonetics tension between categorical and gradient characterization of phonological units. Many discussions of prosody assume a small fixed set of hierarchically-related prosodic domains with recursion within a level disallowed (or at least frowned upon) (e.g., Selkirk 1984, Nespor and Vogel 1986). Categorical differences in boundary type are generally implicit in this approach. It is worth carefully evaluating both assumptions – namely, the existence of a few qualitatively different prosodic boundaries in a language and the strict-layering[1] of these domains. Strict Layering implies that any difference in boundary strength is by definition a difference in boundary type. Ladd (1996, Section 6.3), to which we refer the interested reader, discusses some difficulties with this, focusing on the plausibility of recursion. In the context of the present Commentary, let's here take a brief look at the postulation of a small set of categorically different boundary types and the accommodation of this within a dynamical approach. Within Articulatory Phonology, the dynamical description of phonological units allows, in principle, gradiently varying specification of gestural parameter values and their relative timing (see e.g. Byrd 1996). However in many, though not all, instances, only certain attractor basins in this potential continuum are observed to be used in languages.[2] So if prosodic events have a dynamic reality, one needs to explore whether the strength (activation) of these prominence and junctural events is continuously variable (e.g. gradient), or whether prosodic events demonstrate clear and specific differences in this regard, perhaps behaving as scalar or even further constrained.

In Cho's study and many others, no distinct separation among data from different boundary conditions is observed within subjects (*excluding* those cases with pausing). Contrast such behavior, for example, with the distribution of the acoustic (or articulatory) correlates of other contrastive phonetic categories, such as the timing-based VOT contrast in English unaspirated and aspirated stops. In such contrasts, the two categories are definitely separated in their observed values, i.e. coordination, into two clusters (see e.g., Blumstein 1980 Fig. 1 top panel). This is quite different than the timing behavior we observe for contrasting prosodic categories such as PP and IP (in the absence of pausing, arguably a qualitatively different phenomenon), which doesn't seem to definitively separate into categorically non-overlapping groups. Of course, this is sensible in light of the fact that phonemic contrast plays a critical role in lexical access while prosodic structure plays a different role, namely one related to information processing by speakers and listeners, which is influenced by nongrammatic as well as grammatic variables.

Further, in production experiments having boundaries classified according to a small categorical set by the experimenter, there is often little between-subject consistency in which posited boundary type is apparently chosen by a subject for a given stimulus sentence (exclusive of when pausing [or no boundary] occurs). Such inconsistency between subjects in a category choice is not typical of segmental categories for speakers (sharing a dialect). Again, this suggests that assigning a symbolic boundary type to such productions could be the act of partitioning a continuum of boundary strength. The degree of agreement and consistency among transcribers or among naïve listeners in identifying boundary types is another knotty issue – see Swerts (1997) who finds that textual cues rather than prosodic variables predominate for labelers making prosodic judgments (but see de Pijper and Sanderman 1994 for contrasting results), though a consideration of phonetic information does increase inter-listener consistency for the zero and strongest boundary ratings (presumably the strongest boundaries are the ones with substantial pauses). The experience of many linguists suggests, for example, that Accentual, Intermediate/Phonological, and Intonational Phrases are notoriously difficult to define and distinguish consistently[3]; as Ladd (1996) describes, some Intonational Phrases seem bigger than others. De Pijper and Sanderman (1994) evaluate a measure termed Perceptual Boundary Strength and find that their results argue against the Strict Layer Hypothesis in "that there are no apparent peaks in the distribution of PBS values: "The 175 observed PBS values do not appear to cluster around a limited number of target values, which could then be said to reflect prosodic constituent structure" (p. 2046). In contrast, listeners seem relatively reliable at determining relative strength (or ordering) of boundaries (e.g., de Pijper and Sanderman 1994).

If a scalar description were to capture important aspects of articulation at phrase edges, laboratory phonologists would want to consider how a dynamic description of domain edges – for example, in terms of variable strength activation of prosodic gestures at junctures – might be applied to other prosody-related linguistic phenomena, such as domain-limited phonological rules (Hyman 1990) and the distribution of boundary tones. For example, the permissible distribution of boundary tones (i.e., their phonotactics) would presumably be strongly influenced or constrained by the strength of the juncture with which they are associated. Likewise, phonological processes restricted to particular prosodic domains or junctures would be expected to be sensitive to scalar boundary strength. It is less clear whether a dynamic account could accommodate a situation in which a phonological phenomenon selected an *intermediate* (symbolically designated) boundary type *and only* *that boundary (none higher/stronger and none lower/weaker)*. Interestingly, empirically confirmed evidence of such phenomenon appears to be rare or non-existent. Finally, one aspect of prosody that our current knowledge suggests *is qualitatively different,* is the difference between boundary effects and accent effects (see Cho, this volume, and Beckman and Edwards 1992), as these are not on a continuum (Goldstein, p.c.) and serve different communicative functions.

A future challenge to a dynamical description of prosody will be to understand to what extent prosodic events exhibit or fail to exhibit categorical characteristics, since a dynamic description will be inherently quantitative and potentially continuous. The data patterns described above might suggest that boundary strengths (as captured for example in the activation of π-gestures) may differ gradiently. Alternatively, it is possible that just as constriction gesture parameter values and coordination relations have been considered to cluster into distinctive regions (see e.g. Goldstein 1989), despite what is in principle a continuum of values, perhaps the defining attributes of prosodic events will be determined to behave similarly. However, since phrasal and prominence information do not serve the role of encoding lexical contrast, it might be unsurprising for each of these to be scalar or gradient in nature, reflecting, for example, degree of embedding. A dynamical systems approach to speech production allows for both categorical (though quantitative rather than purely symbolic) differences among constriction gestures forming the basis of contrast and for gradient differences among junctural and prominence events organizing information above the single-word level. In contrast, it is difficult to accommodate both types of linguistic phenomenon – categorical and gradient – seamlessly within a symbolic approach.

5. Prosody within the dynamical approach

This collection from LabPhon8 encourages an exploration of prosody from a dynamical perspective. If a dynamical approach can be successfully developed for phrasal and prominence structure, this will allow the articulatory variability engendered by the informational structure of utterances to be treated in a way conceptually parallel to the treatment of variability due to segmental context. The future may show us that phrasal junctures and prominence events, just like constriction gestures, have inherent temporal properties, are temporally coordinated, and have a dynamical instantiation.

Notes

* This work was supported by NIH grant DC03172. The author thanks the contributors to LabPhon8, the organizers of LabPhon8, and Edward Flemming, Bob Ladd, Mark Liberman, Louis Goldstein, Elliot Saltzman, and Doug Whalen.

1. The Strict Layer Hypothesis states that a category of level i in the hierarchy immediately dominates a (sequence of) categories of level i–1 (Selkirk 1984). Ladd (1996) notes that the Strict Layer Hypothesis was originally formulated by Selkirk (1984) as a 'useful working hypothesis' (p. 238). For a more flexible approach to layering see Ito and Mester (1992) on Weak Layering. See also Ladd and Campbell (1991) on compound prosodic domains.
2. In some instances this is due to nonlinearities in articulatory-to-acoustic mapping and requirements for successful signal transmission (see e.g. Goldstein 1989, Browman and Goldstein 2000).
3. The acoustic cues (or the articulatory maneuvers they result from) that are the dependent variables for many experiments looking at effects of these phrasal categories can unfortunately be the very same cues used to distinguish these independent variables for listeners coding the data utterances in the first place – a clear circularity problem.

References

Beckman, Mary E. and Jan Edwards
1992 Intonational categories and the articulatory control of duration. In: Y. Tohkura, E. Vatikiotis-Bateson, Y. Sagisaka (eds.), *Speech Perception, Production, and Linguistic Structure*, 359–375. Tokyo: Ohmsha, Ltd.

Blumstein, S. E., W. E. Cooper, H. Goodglass, S. Statlender and J. Gottlieb
1980 Production deficits in aphasia: A voice-onset time analysis. *Brain and Language* 9: 153–170.

Browman, Catherine P. and Louis Goldstein
1990 Tiers in articulatory phonology, with some implications for casual speech. In John Kingston and Mary E. Beckman (eds.), *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*, 341–376. Cambridge: Cambridge University Press.
1992 Articulatory phonology: An overview. *Phonetica* 49: 155–180.
1995 Dynamics and articulatory phonology. In R. Port and T. van Gelder (eds.) *Mind as Motion: Dynamics, Behavior, and Cognition*: 175–193. Cambridge, MA: MIT Press.
2000 Competing constraints on intergestural coordination and self-organization of phonological structures. *Bulletin de la Communication Parlée* 5: 25–34.

Byrd, Dani
1996 A phase window framework for articulatory timing. *Phonology* 13(2): 139–169.
2000 Articulatory vowel lengthening and coordination at phrasal junctures. *Phonetica*, 57(1): 3–16.

Byrd, Dani, Abigail Kaun, Shrikanth Narayanan and Elliot Saltzman
2000 Phrasal signatures in articulation. In Michael B. Broe and Janet B. Pierrehumbert (eds.) *Papers in Laboratory Phonology V*, 70–87. Cambridge: Cambridge University Press.

Byrd, Dani and Elliot Saltzman
1998 Intragestural dynamics of multiple phrasal boundaries. *Journal of Phonetics* 26: 173–199.
2003 The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics* 31: 149–180.

De Pijper, J. R. and A. A. Sanderman
1994 On the perceptual strength of prosodic boundaries and its relation to suprasegmental cues. *Journal of the Acoustical Society of America* 96(4): 2037–2047.

Edwards, Jan, Mary E. Beckman and J. Fletcher
1991 The articulatory kinematics of final lengthening. *Journal of the Acoustical Society of America*, 89(1): 369–382.

Frisch, Stefan and Richard Wright
2002 The phonetics of phonological speech errors: An acoustic analysis of slips of the tongue. *Journal of Phonetics*, 30(2): 139–162.

Goldstein, Louis
1989 On the domain of the quantal theory. *Journal of Phonetics* 17: 91–97.

Goldstein, Louis and Carol Fowler
2003 Articulatory Phonology: A phonology for public language use. In Antje Meyer and Niels Schiller (eds.) *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities*, 159–208.

Goldstein, Louis, Marianne Pouplier, Larissa Chen, Elliot Saltzman and Dani Byrd
subm. Gestural action units slip in speech production errors.

Hyman, Larry
1990 Boundary tonology and the prosodic hierarchy. In Sharon Inkelas and Draga Zec (eds.) *The Phonology-Syntax Connection*, 109–126. Chicago: The University of Chicago Press.

Itô, J. and A. Mester
1992 Weak layering and word binarity. *Linguistics Research Center-92-09*.

Keating, Patricia, Taehong Cho, Cécile Fougeron and Chai-Shune Hsu
2004 Domain-initial strengthening in four languages. In John Local, Richard Ogden, and Rosalind Temple (eds.) *Papers in Laboratory Phonology VI*.
Keating, Patricia and Stefanie Shattuck-Hufnagel
2002 A prosodic view of word rorm encoding for speech production *UCLA Working Papers in Phonetics* 101: 112–156
Kelso, J. A. S., B. Tuller, E. Vatikiotis-Bateson and C. A. Fowler
1984 Functionally specific articulatory cooperation following jaw perturbations during speech: Evidence for coordinative structures. *Journal of Experimental Psychology: Human Perception and Performance* 10: 812–832.
Ladd, D. R.
1996 *Intonational Phonology*. Cambridge: Cambridge University Press.
Ladd, D. R. and N. Campbell
1991 Theories of prosodic structure: evidence from syllable duration. In *Proceedings of the 12th International Congress of Phonetic Sciences.* Aix-en-Provence: Université de Provence.
Mowrey, R. A. and I. R. MacKay
1990 Phonological primitives: Electromyographic speech error evidence. *Journal of the Acoustical Society of America*, 88(3): 1299–1312.
Nespor, M. and I. Vogel
1986 *Prosodic Phonology.* Dordrecht: Fortis.
Saltzman, Elliot and Dani Byrd
2000a Task-dynamics of gestural timing: Phase windows and multifrequency rhythms. *Human Movement Science*, 19(4): 499–526.
2000b Demonstrating effects of parameter dynamics on gestural timing. May 2000 meeting of the Acoustical Society of America, *Journal of the Acoustical Society of America* 107(5,2): 2904.
Saltzman, Elliot, Anders Löfqvist, Bruce Kay, J. Kinsella-Shaw and P. Rubin
1998 Dynamics of intergestural timing: A perturbation study of lip-larynx coordination. *Experimental Brain Research* 123: 412–424.
Saltzman, Elliot and Kevin G. Munhall
1989 A dynamical approach to gestural patterning in speech production. *Ecological Psychology* 1: 333–382.
Selkirk, Elizabeth
1984 *Phonology and Syntax: The Relation between Sound and Structure*. Cambridge, MA: MIT Press.
Swerts, Marc
1997 Prosodic features at discourse boundaries of different strength. *Journal of the Acoustical Society of America* 101(1): 415–521.

III. Knowledge of language-specific organization of speech gestures

Interaction of prosody and gestures

Local gesture interaction and perception

Syllable position effects and gestural organization: Articulatory evidence from Russian

Alexei Kochetov

Previous articulatory studies have shown that English syllable-initial and syllable-final consonants exhibit different patterns of gestural organization. These differences – syllable position effects – *are manifested primarily in the relative timing and magnitude of gestures. In general, syllable-initial consonants show more stable patterns of coordination and "tighter" articulatory constrictions than the same consonants in syllable-final position. This paper addresses the question of whether syllable position effects hold for other languages by examining the articulatory properties of some Russian syllable-initial and syllable final consonants: the palatal glide /j/ and labial stops /p^j/ and /p/. In general, the articulometer (EMMA) results confirm the hypothesis that the same consonants in these two positions differ with respect to their inter-gestural timing and gestural magnitude. At the same time, some predictions made based on patterns observed in English are not supported. The results thus provide evidence that, although syllable positions are characterized by different patterns of gestural organization, actual manifestations of this organization are not always the same; they may vary between gestures within a language and between similar gestures in different languages. It is further suggested that while the task-dynamics of gestural coordination is crucial to explaining syllable position effects, some non-contrastive and variable information still has to be specified – possibly by being lexically encoded.*

1. Introduction: Background and predictions

While the role of the syllable as a phonological unit has been long established, attempts to find the basis of the syllable in phonetics – articulation and/or acoustics – have not been entirely successful until recently. A number of articulatory studies (Krakow 1989, Sproat and Fujimura 1993, Browman and Goldstein 1995, Byrd 1996, Fougeron and Keating 1997, Gick 2003, among others) have provided evidence for the gestural organization of the syllable in English. Specifically, it has been established that syllable-initial and syllable-final consonants tend to exhibit different patterns of gestural organization. These differences, referred to in this paper as *syllable position effects*, are manifested primarily in the relative timing of gestures, and in their magnitude, that is, in the degree of constriction.

A classic example of these contextual differences involves positional allophones of the American English /l/, namely the syllable-initial *clear* [l] as in 'leap' and the syllable-final *dark* [ɫ] as in 'peel'. Investigations of these allophones using a variety of methods (Giles and Moll 1975, Sproat and Fujimura 1993, Browman and Goldstein 1995, and Gick 2003) have shown that both are produced with two articulators, a tongue tip (TT) closure at the alveolar ridge and a tongue body (TB) retraction. The difference between the dark and clear /l/ is in the timing and magnitude of these gestures. First, while the TT gesture is near-synchronous with the TB gesture syllable-initially, TT follows TB syllable-finally. Second, the relatively consistent achievement of the TT target syllable-initially contrasts with the frequent lack of complete closure syllable-finally, particularly in casual and fast speech (Giles and Moll 1975). In sum, distinct timing and magnitude patterns of the same gestures result in the two different kinds of acoustic and perceptual consequences commonly associated with the allophones [l] and [ɫ]. Similar positional effects on timing have been observed for other English consonants that involve multiple oral gestures, such as the nasal /m/ and the glide /w/. Thus, the lip aperture and velum opening gestures of syllable-initial /m/ achieve their goals synchronously. In contrast, in the production of the syllable-final /m/, the more closed articulator, the lips, lags substantially behind the more open one, the velum (Krakow 1989).

Magnitude effects, similar to the final reduction of the tongue tip of /l/, have also been documented for a number of English consonants: oral and nasal stops (Browman and Goldstein 1995, Byrd 1996, Turk 1994, Fougeron and Keating 1997), the glides /w/ and /j/, and the rhotic /r/ (Gick 2003). While exhibiting the same general pattern, individual gestures tend to vary in the degree of reduction and overall stability. For instance, it was found that the oral gestures of velar and labial stops (tongue dorsum and the lips) show a moderate reduction in magnitude in coda position, while the articulation of alveolar stops in the same position is highly variable, with TT being considerably reduced in magnitude (Browman and Goldstein 1995, Byrd

1996, among others). Although the magnitude reduction effect is commonly displayed by syllable-final consonants in English, the opposite process has been found to occur in some cases. Thus the reduction of TT in the articulation of the *dark* [l] is often accompanied by some increase in magnitude of the TB retraction gesture (Giles and Moll 1975). Similarly, the lowering of the velum of the syllable-final /m/ was found to be more extreme in final than in initial position (Krakow 1989). Yet some other consonants, for example the sibilant fricatives /s/ and /z/, do not seem to be affected by syllable position (Byrd 1996).

One question raised by the findings for English is whether they can be extended cross-linguistically, that is whether other languages also show differences between syllable-initial and syllable-final consonants in inter-gestural timing and magnitude. The few articulatory investigations of syllable effects that have been conducted on other languages have provided mixed results: while some found positional differences in gestural organization similar to English (see Krakow 1999 for a review), others reported either no consistent syllable position effect or effects somewhat different from the findings for English (see, for example, Wang 1995 on the timing and magnitude of the velum and lips in Cantonese; Gick, Campbell, Oh, and Tamburri-Watt 2006, on the magnitude and timing of tongue tip and tongue dorsum for /l/ in a number of languages).

The current work aims to contribute to the cross-linguistic study of the gestural properties of the syllable by examining positional effects in three Russian consonants: the palatalized and non-palatalized voiceless labial stops /p^j/ and /p/, and the palatal glide /j/. The first consonant, /p^j/, presents an interesting case for an articulatory study of inter-gestural timing. The consonant is a complex gestural constellation consisting of two coordinated oral gestures, Lips [bilabial, closed] and Tongue Body [palatal, narrow] (Kochetov 2002; see also Recasens and Romero 1997, Zsiga 2000 on the gestural organization of palatalized consonants). An examination of this consonant will allow us to investigate the timing of the two gestures in syllable-initial and syllable-final positions, and to compare this timing to the timing of the same gestures in combinations of the consonants /p/ and /j/, sequences /pj/ and /jp/. In addition, I will examine the magnitude and duration properties of the two gestures in the production of the syllable-initial and syllable-final consonants /p^j/, /p/, and /j/. (An examination of the velarization of the non-palatalized /p/ is beyond the scope of this paper; but see Kochetov 2002.)

The general hypothesis tested in this work is that the Russian consonants in question differ syllable-initially and syllable-finally with respect to inter-gestural timing and their internal gestural properties. That is, the hypothesis predicts that the timing of the gestures of Lip Aperture (LA) and Tongue Body (TB) of /p^j/, as well as the magnitude (and possibly duration) of these gestures in /p^j/, /p/, and /j/ differ depending on the position. Further, more specific predictions can be made about the types of differences to be observed. The timing patterns found in English suggest that we may expect the two gestures to be synchronous syllable-initially, and the more closed gesture, the lips, to follow the more open gesture, TB, syllable-finally. In addition, in all positions, the timing of the two gestures of /p^j/ in [p^{j}a] and [ap^j] may differ from the timing of the same gestures in the sequences of /p/ and /j/ in [pja] and [ajp], since these are lexically distinct in Russian (Avanesov 1984: 139–140, Jones and Ward 1969: 93–96). These specific predictions will be referred to as timing hypotheses A and B.

Based on the findings for the LA and TB [palatal] gestures of the English consonants /p/ and /j/, we may expect the corresponding gestures of the Russian /p/ and /j/ to be reduced syllable-finally, with more reduction for /j/ than for /p/ (the magnitude hypothesis). For the palatalized labial /p^j/, we may find the same effect for both LA and TB or, we may find a reduction of the more closed constriction, LA, and an increase in the magnitude of the more open constriction, TB (as in the English syllable-final /l/). Overall, the syllable-final consonants are expected to show more variation in timing and magnitude (the variability hypothesis). No specific predictions are made about temporal differences between initial and final gestures, or about differences between the same gestures of different consonants (LA of /p/ and /p^j/; TB of /p^j/ and /j/).

2. Experiment

Data were collected using the EMMA (Electromagnetic Midsagittal Articulometer: Perkell et al. 1992) magnetometer system at Haskins Laboratories. Four speakers of standard Russian participated in the experiment: three females (subjects AS, NT, and DK) and one male (the author, subject AK). Subjects AS and DK were originally from Moscow and subjects NT and AK were from Perm', Russia. (Some of the data, nonwords with /p/ and /p^j/ for subjects AS, NT, and AK, were used in Kochetov (2002), a study investigating the relation between articulation, perception, and phonotactics of palatalized stops.)

2.1. Materials

The stimuli included nonwords and real Russian words with the consonants /pʲ/, /p/, and /j/. Since the data for subject AK were collected earlier as part of a larger-scale exploratory study, not all of the stimuli for this subject were identical to those used later for the remaining subjects. Therefore, the stimuli used for subjects AS, NT, and DK are described here separately from the stimuli used for subject AK. The utterances used for subjects AS, NT, and DK are shown in Table 1. The utterances consisting of a sequence of nonwords and real words had the same target consonants in the same immediate environments. The stress pattern in both types of utterances was also controlled for: both test words of the utterance carried primary stress, for example, [ˈtapʲ ˈpapɨ] or [ˈgrapʲ ˈkaz̯dəvə].

Table 1. The nonword and real word stimuli presented to the subjects AS, NT, and DK and the analysis of these stimuli in the studies of inter-gestural timing, TB and LA magnitude and duration.

Cons.	Position	Nonword stimuli		Real word stimuli		Analysis		
		IPA	Translit.	IPA	Translit.	Timing	TB	LA
/pʲ/	onset	ta **pʲ**apɨ	*ta pjapy*	ṣla **pʲ**atəjə	*shla pjataja*	✓	✓	✓
		tap **pʲ**apɨ	*tap pjapy*	grap **pʲ**atəvə	*grab pjatogo*		✓	
		tat **pʲ**apɨ	*tat pjapy*	brat **pʲ**atəvə	*brat pjatogo*			✓
	coda	ta**pʲ** apɨ	*tap' apy*	gra**pʲ** angʲɪlə	*grab' angela*	✓	✓	✓
		ta**pʲ** papɨ	*tap' papy*	gra**pʲ** padəjə	*grab' padaja*		✓	
		ta**pʲ** tapɨ	*tap' tapy*	gra**pʲ** tomnəvə	*grab' tomnogo*			✓
		ta**pʲ** kapɨ	*tap' kapy*	gra**pʲ** kaz̯dəvə	*grab' každogo*			✓
/p/	onset	ta **p**apɨ	*ta papy*	ṣla **p**adəjə	*šla padaja*		✓	
		tat **p**apɨ	*tat papy*	brat **p**adəjə	*brat padaja*		✓	
		tak **p**apɨ	*tak papy*	brak **p**adəjə	*brak padaja*			✓
	coda	ta**p** apɨ	*tap apy*	gra**p** angʲɪlə	*grab angela*			✓
		ta**p** tapɨ	*tap tapy*	gra**p** tomnəvə	*grab tomnogo*			✓
		ta**p** kapɨ	*tap kapy*	gra**p** kaz̯dəvə	*grab každogo*			✓
/j/	onset	ta **j**apɨ	*ta japy*				✓	
		tap **j**apɨ	*tap japy*			✓	✓	
	coda	ta**j** apɨ	*taj apy*				✓	
		ta**j** papɨ	*taj papy*			✓	✓	

The need to control precisely for target consonants and environments took precedence over the semantic plausibility of these word combinations. (The real word utterances were verbs + noun/adjective combinations of the following Russian words: *šla* (she) 'walked,' *grab'* 'rob' (imp.), *grab* 'hornbeam,' *brat* 'brother,' *brak* 'flaw,' *pjataja* 'the fifth' (fem.), *angela* 'angel' (acc. sg.), *pjatogo* 'the fifth' (acc. sg.), *padaja* 'falling,' *tomnogo* 'fat' (acc. sg.), *každogo* 'every.') All stimuli were embedded in a carrier phrase [ˈɛtʌ ______ ʌˈpʲatʲ] 'This is ____ again' and presented in Cyrillic in alternating blocks of nonword and real word utterances. Five tokens for each nonword and each real word utterance were collected, yielding a total of 70 tokens per subject for /pʲ/ (30 onset and 40 coda tokens), 60 tokens for /p/ (30 onset and 30 coda tokens), and 20 tokens for /j/ (10 onset and 10 coda tokens). The last three columns of the table indicate which stimuli were used in which particular study (described in Section 2.2), the studies of: inter-gestural timing (30 tokens per subject), TB magnitude and duration (60 tokens), and LA magnitude and duration (110 tokens). Note that due to errors such as false starts or certain technical problems, the number of analyzed tokens was sometimes fewer than collected; in other cases, some additional tokens were collected and analyzed. The number of tokens per subject analyzed in each particular study is given in the presentation of the results in Section 3.

Table 2. Real word stimuli presented to subject AK. (Nonword stimuli for /pʲ/ and /p/ were the same as for the other subjects).

Cons.	Position	Real word stimuli
/pʲ/	onset	**pʲ**at *pjat,* **pʲ**atʲ *pjat',* krap **pʲ**atəvə *krab pjatogo,* brat **pʲ**atəvə *brat pjatogo*
	coda	gra**pʲ** adʊ *grab' Adu,* gra**pʲ** padəjə *grab' padaja,* gra**pʲ** tantsɨ *grab' tancy,* gra**pʲ** kadətʃnʲɪka *grab' kadočnika*
/p/	onset	**p**at *pat,* brat **p**adəjə *brat padaja,* brak **p**adəjə *brak padaja*
	coda	kra**p** adɨ *krab Ady,* kra**p** tantsə *krab tanca,* kra**p** kadətʃnʲɪka *krab kadočnika*
/j/	onset	p**j**anɨj *pjjanyj*
	coda	ba**j**t *bajt,* vo**j**n *vojn*

2.2. Procedure and analysis

Receivers for the articulometer were placed at the following midsagittal points: *upper lip* and *lower lip* (UL and LL; placed at the border of vermillion), lower incisors (as an estimate of jaw movement), and four points on the tongue. One of the tongue receivers was positioned about 5 mm from the *tongue tip* (TT); another receiver was attached as far back on the tongue as was possible, roughly on the *tongue dorsum* (TD). The other two receivers, *tongue body 1* (TB1, more anterior) and *tongue body 2* (TB2, more posterior), were placed between these first two receivers at approximately equal distances from TT and TD respectively, and from each other. The movement data were collected at a sampling rate of 500 Hz and the acoustic data at 20 kHz. The kinematic data were converted from voltage to distance, calibrated, and corrected for head movement and for any possible shifts of the receivers relative to the transmitters. Some malfunctions of the receiver coil for UL for subject AS resulted in a defective set of data for the corresponding articulator. For this subject the analysis of lip movements was based on the LL trajectories only.

The analysis of the articulatory data collected involved a number of measured variables: The distance between the lower and upper lip trajectories – the *lip aperture* (LA) – was taken as an indicator of the formation and release of the lip gesture of /p/ and /p^j/. Since the UL data were missing for one of the speakers (subject AS), the LL raising movement was also compared across the subjects. The position of the receiver on the tongue that showed the maximum amplitude of raising and fronting movement during the articulation of /j/ and /p^j/ was taken as an indicator of the palatal gesture of these consonants. For subjects NT and AK, this receiver was TB2, while for subjects AS and DK it was TD. For consistency of presentation, I will refer to the variables analyzed as the "lip aperture" (LA) and the "tongue body (position)" (TB).

Tangential velocity minima for all receivers were automatically calculated, based on the velocity in the x- and y-coordinates of these receivers (see Chitoran, Goldstein, and Byrd 2002). The velocity minima obtained were used to determine temporal articulatory landmarks – the beginning of the closing movement, the achievement of the target, and the release from the target. The beginning of the closing movement and the release from the target were located when the velocity for a given receiver exceeded a threshold of 20% of the mean peak velocity for that receiver over all utterances. The achievement of the target was located when the velocity for a given receiver fell below the 20% threshold. Positional maxima and minima (*peaks* and *valleys*) for LA, LL, and TB were also obtained automatically. These were used in determining the magnitude of gestures.

Two measurements of inter-gestural timing were used in the study: *achievement lag* and *release lag*. The first measurement was defined as a period of time between the achievement of two constrictions, the lips (LA or LL) and TB; the second measurement was defined as a period of time between the release of these constrictions. The lag was considered positive if the achievement or release of the TB followed the corresponding landmarks of the lips; it was considered negative if the achievement or release of the TB preceded the corresponding landmarks of the lips. These measurements were limited to the minimally contrastive nonword and real word utterances with a single /p^j/ (in syllable onset and coda) and the word-boundary sequences of /p/ and /j/ (/pj/ and /jp/; see Table 1).

The gestural magnitude measurements were based on the vertical and horizontal peak values of the TB constriction and the vertical peak values of LL, as well as on the minimum values of LA. Gestural duration was obtained based on the same articulatory landmarks as the analysis of inter-gestural timing. It involved the measurements of the duration of the *closing movement* (the time from the onset of movement towards a constriction to the achievement of this constriction) and *plateau* (the time between the achievement of the constriction and the beginning of the movement away from this constriction). The TB measurements were done for /p^j/ and /j/; the LL raising and LA measurements were done for /p^j/ and /p/ (see Table 1).

The results of all measurements were further tested in a number of Analyses of Variance (ANOVAs), separately for each subject. Each test involved three between-item factors: Consonant (/p/, /p^j/, and /j/), Position (syllable onset vs. syllable coda), and Type of stimuli (nonwords vs. real words). In each case, Tukey HSD post-hoc tests were performed to investigate significant interactions.

3. Results

3.1. Inter-gestural timing

The results for achievement lag are presented in Figure 1. Here the values for the lag in sequences /pj/ and /jp/ are given at the bottom and the values for /p^j/ at the top. Significant effects and interactions for both lag measurements are shown in Table 3. Overall, the values for the lag measurements were lower in coda position than in onset position, indicating that, syllable-finally, the TB constriction was both achieved and released earlier with respect to the LA constriction than was the case in initial position. There was, however, a significant Consonant X Position interaction for all subjects, pointing to considerable differences in the effect of position between the sequences of /p/

and /j/ on the one hand and the consonant /p^j/ on the other. In addition, there was a main effect of Consonant for some of the subjects.

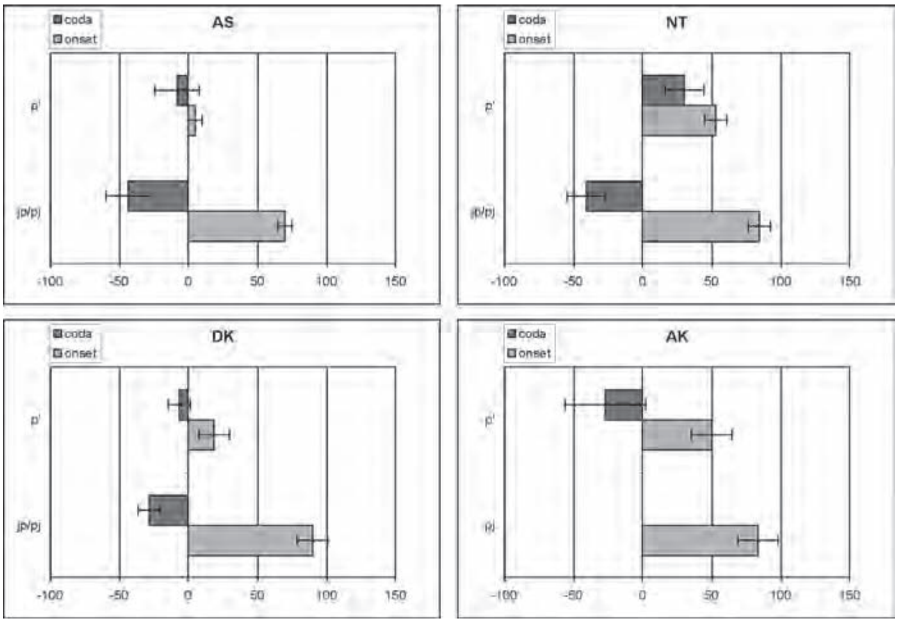


Figure 1. Mean values for achievement lag for /p^j/ in a#_a (onset) and a_#a (coda) in nonwords and real words and for the sequences /p#j/ (compared with onset) and /j#p/ (compared with coda) in nonwords (in ms) for four subjects.

Table 3. A summary of significant main effects and interactions for achievement lag (a) and release lag (b) for four subjects.

		AS	NT	DK	AK
a.	Position	F(1,26) = 70.153, p < 0.001	F(1,28) = 233.683, p < 0.001	F(1,28) = 203.432, p < 0.001	F(1,25) = 101.314, p < 0.001
	Consonant	not significant	F(1,28) = 18.419, p < 0.001	F(1,28) = 12.874, p < 0.01	F(1,25) = 10.135, p < 0.01
	Position X Consonant	F(1,26) = 46.809, p < 0.001	F(1,28) = 121.579, p < 0.001	F(1,28) = 92.228, p < 0.001	not available
b.	Position	F(1,26) = 32.469, p < 0.001	F(1,28) = 159.748, p < 0.001	F(1,28) = 475.753, p < 0.001	F(1,25) = 43.818, p < 0.001
	Consonant	not significant	F(1,28) = 5.867, p < 0.05	F(1,28) = 10.785, p < 0.01	not significant
	Position X Consonant	F(1,26) = 15.498, p < 0.01	F(1,28) = 205.975, p < 0.001	F(1,28) = 309.290, p < 0.001	not available

3.1.1. Sequences of /p/ and /j/

The results for the sequences for the first three subjects (AS, NT, and DK) are considered separately from those for subject AK, because of certain differences in the stimuli. We can see from Figure 1 that the positional differences in achievement lag between the sequences /pj/ and /jp/ for subjects AS, NT, and DK were substantial: For /pj/, the achievement of TB followed the achievement of LA by about 70–90 ms on average – a positive lag. For /jp/, the achievement of TB preceded the achievement of LA by about 30–45 ms – a negative lag. The positional differences in release lag were similar: for /pj/, TB followed LA by about 65–100 ms; for /jp/, TB preceded LA by about 50–75 ms. Overall, the lag values indicate that the constrictions of the two gestures were timed sequentially, as one might expect for sequences of two consonants; yet there was less lag in the articulation of /jp/ than for /pj/.

The timing observed for subject AK in the word-internal sequence /pj/ was similar to the timing for the other subjects (where /pj/ was a word-boundary sequence): there was a substantial positive achievement and release lag (about 85 and 75 ms respectively). An additional examination of the timing of TB and TT in the word-internal clusters /jt/ and /jn/ available for this subject (see Table 2) showed negative achievement and release lag (about 25 and 30 ms), but of a somewhat lesser degree than the TB-LA lag found for the other subjects.

3.1.2. The consonant /p^j/

We can see from Figure 1 (the upper bars) that the positional differences in achievement lag between the onset and coda /p^j/ were much smaller than between the two sequences: For the syllable-initial /p^j/, the achievement of TB followed the achievement of LA by about 5–55 ms (mean values). For the syllable-final /p^j/, the lag values were more variable: the achievement of TB preceded the achievement of LA for subjects AS, DK, and AK (by about 5–25 ms), while the achievement of TB followed the achievement of LA for subject NT (by about 30 ms). All subjects, however, showed a significant difference in lag values between the onset and coda /p^j/, which varied from about 15 to 75 ms ($p < 0.001$–0.01). The positional differences in release lag were similar: a positive lag of about 5–55 ms for the syllable-initial /p^j/ and either negative or positive lag ranging from –30 to 40 ms for the syllable-final /p^j/. Three out of four subjects showed significant differences in lag values between the onset and coda /p^j/ (AS, DK, AK: $p < 0.001$–0.05); these difference, however, varied substantially, from about 20 ms for subject AS to about 70 ms for subject AK.

Overall, the two constrictions in the production of the consonant /p^j/ were timed relatively simultaneously, yet the positional differences in timing were in the same direction as the differences between sequences of consonants: a positive lag for the onset /p^j/ comparable to the substantial positive lag for the sequence /pj/, and a zero-to-negative lag for the coda /p^j/ in some cases comparable to the negative lag for the sequence /jp/. The difference between /p^j/ and the sequences of /p/ and /j/ in timing was always maintained: the onset /p^j/ and the sequence /pj/ differed in about 30–70 ms in achievement lag and in about 10–85 ms in release lag; the coda /p^j/ and the sequence /jp/ differed in about 20–70 ms in achievement lag and in about 60–90 ms in release lag. An informal examination of standard deviations suggested slightly higher variability for the coda /p^j/ for at least some subjects (AS, NT, AK; see error bars in Figure 1).

3.2. Magnitude and duration of Tongue Body

3.2.1. Magnitude

The magnitude results for maximum TB height, with respect to the upper teeth (a) and TB fronting (b) are presented in Table 4. The values for syllable-initial and syllable-final /p^j/ are presented on the left and those for /j/ in the same positions are given on the right. The results for TB height (see Table 4a) indicate that there was a main effect of Position for all four subjects [AS: $F(1, 58) = 12.806$, $p < 0.01$; NT: $F(1, 58) = 68.244$, $p < 0.001$; DK: $F(1, 59) = 22.902$, $p < 0.001$; AK: $F(1, 50) = 44.931$, $p < 0.001$]: TB was significantly lower in coda position than in onset position. There was a main effect of Consonant for all subjects [AS: $F(1, 58) = 170.784$ $p < 0.001$; NT: $F(1, 58) = 37.860$, $p < 0.001$; DK: $F(1, 59) = 26.035$, $p < 0.001$; AK: $F(1, 50) = 27.133$, $p < 0.001$]: TB during /j/ was substantially higher than during /p^j/. For one subject there was a significant Consonant X Position interaction [DK: $F(1, 59) = 16.216$, $p < 0.001$]: the factor Position was significant for /j/ only. Overall, syllable-final /j/ showed more reduction (on average by about 1.5 to 4 mm) than syllable-final /p^j/ (about 0.5 to 2 mm). The reduction of the TB for /j/ was also more consistent across subjects than the reduction of the same gesture for /p^j/. As a result, the TB raising difference between the two consonants – rather substantial in onset position (3 to 6.5 mm) – was less apparent in coda position (0.5 to 6 mm).

Table 4. Mean values of (a) tongue body height (mm above the upper incisors) and (b) tongue body fronting (mm posterior to upper incisors – smaller numbers are more fronted) for /p^j/ and /j/ in syllable onset (a#_a, p#_a) and syllable coda (a_#a, a_#p) in nonwords and real words (for /p^j/ only) for four subjects.

		AS (N=59)	NT (N=59)	DK (N=60)	AK (N=51)
a.	onset /p^j/	4.18 (1.06)	10.07 (0.82)	3.22 (1.10)	4.27 (1.24)
	coda /p^j/	3.27 (0.89)	8.04 (1.54)	2.92 (1.42)	2.12 (1.81)
	onset /j/	10.52 (1.02)	13.07 (0.73)	7.43 (0.69)	7.81 (1.44)
	coda /j/	9.19 (2.13)	9.63 (1.12)	3.27 (1.78)	3.45 (1.27)
b.	onset /p^j/	37.77 (2.16)	33.91 (1.00)	27.58 (0.83)	30.16 (1.19)
	coda /p^j/	37.93 (1.89)	33.91 (1.55)	27.49 (1.35)	35.13 (1.95)
	onset /j/	37.63 (1.13)	33.56 (0.61)	25.53 (1.02)	28.54 (1.04)
	coda /j/	37.48 (1.63)	35.61 (0.89)	28.60 (2.33)	31.61 (1.12)

For TB fronting (see Table 4b), there was a main effect of Position for three subjects [NT: $F(1, 58) = 9.777$, $p < 0.01$; DK: $F(1, 59) = 8.439$, $p < 0.01$; AK: $F(1, 50) = 106.478$, $p < 0.001$]: For all of them, TB in coda position was less front (i.e., reduced) than in onset position. There was a main effect of Consonant for three subjects [AS: $F(1, 58) = 5.116$, $p < 0.05$; NT: $F(1, 58) = 15.131$, $p < 0.001$; AK: $F(1, 50) = 50.928$, $p < 0.001$]: For two of these subjects the TB gesture for /j/ was slightly more front than for /p^j/ (AS and AK); for the other subject (NT), the TB position for /j/ was more to the back than /p^j/. There was a significant Consonant X Position interaction for three subjects [NT: $F(1, 58) = 14.007$, $p < 0.001$; DK: $F(1, 59) = 10.615$, $p < 0.01$; AK: $F(1, 50) = 7.929$, $p < 0.01$]: For the first two speakers /j/ was more susceptible to reduction than /p^j/; for the third speaker the reverse was observed. Overall, the results of TB fronting showed high inter-speaker variability. The positional differences were limited primarily to the TB for /j/ (three subjects), which was moderately reduced in coda.

There were no consistent differences in TB raising or fronting between more specific contexts: the post-vocalic vs. post-consonantal onset (a#_a vs. p#_a) and the prevocalic vs. preconsonantal coda (a_#a vs. a_#p). An examination of standard deviations (see Table 4) suggested that the TB raising and fronting values for the coda consonants, and particularly for the coda /j/, were more variable than those for the same consonants in onset position, at least for some speakers (AS, NT, and DK). For example, standard deviations for the syllable-initial /j/ of subject DK were 0.69 mm for TB height and 1.02 mm for TB fronting, while standard deviations for the syllable-final /j/ were substantially higher: 1.78 mm and 2.33 mm respectively (See Table 4). The TB gesture in the real words was less front (AS, NT, and AK) than in the nonwords.

3.2.2. Duration

The results for the closing movement (a) and the plateau (b) duration of the TB raising gesture are presented in Table 5.

Table 5. Mean duration (in ms) of the closing movement (a) and plateau (b) of the tongue body raising gesture for /pʲ/ and /j/ in syllable onset (a#_a, p#_a) and syllable coda (a_#a, a_#p) in nonwords and real words (for /pʲ/ only) for four subjects.

		AS (N=54)	NT (N=59)	DK (N=60)	AK (N=51)
a.	onset /pʲ/	65.70 (26.02)	142.35 (37.19)	107.64 (19.51)	11.19 (42.50)
	coda /pʲ/	62.00 (30.76)	122.26 (58.23)	77.65 (29.82)	88.94 (33.75)
	onset /j/	111.00 (40.37)	105.40 (42.12)	120.50 (17.30)	148.00 (20.59)
	coda /j/	110.00 (29.00)	150.20 (29.94)	93.50 (20.30)	100.25 (25.76)
b.	onset /pʲ/	89.70 (37.03)	44.80 (16.89)	36.36 (9.65)	54.14 (26.58)
	coda /pʲ/	111.05 (44.06)	49.95 (25.09)	46.75 (20.62)	54.67 (26.03)
	onset /j/	63.33 (16.28)	40.40 (19.16)	22.75 (13.52)	59.00 (17.63)
	coda /j/	41.20 (25.63)	30.20 (10.78)	17.80 (5.45)	18.00 (3.55)

The ANOVA results for the closing movement indicate that two subjects showed a main effect of Position [DK: $F(1, 59) = 28.378$, $p < 0.001$; AK: $F(1, 49) = 18.993$, $p < 0.001$]: this movement was significantly shorter in coda position (e.g., about 20% shorter for subject AK). There was a main effect of Consonant for three subjects [AS: $F(1, 53) = 18.567$, $p < 0.001$; NT: $F(1, 58) = 12.412$, $p < 0.001$; AK: $F(1, 49) = 22.928$, $p < 0.001$], although, in different directions: the closing movement was longer either for /j/ (AS, AK) or for /pʲ/ (NT). There was no significant Consonant X Position interaction for any of the speakers.

For the plateau duration, there was no main effect of Position for any of the subjects. The factor Consonant was significant for two subjects [AS: $F(1, 54) = 26.351$, $p < 0.001$; DK: $F(1, 59) = 13.090$, $p < 0.01$; AK: $F(1, 50)$

= 14.204, p < 0.001], showing longer plateaus for /p^j/ than for /j/. It should be noted, however, that the longer plateaus of /p^j/ for these speakers were often accompanied by relatively short closing periods; there were no significant differences between the two consonants in the total duration of the gesture (closing movement + plateau). There was a significant Consonant X Position interaction for one subject only [AK: F(1, 50) = 6.023, p < 0.05], indicating that the reduction in duration in coda position affected /j/ but not /p^j/. Note that this may be due to the difference in the stimuli used for this subject.

A comparison of standard deviations for the duration measurements in syllable-initial and syllable-final positions revealed somewhat higher variation in syllable-final /p^j/, at least for some speakers (AS, NT, DK). Overall, the duration for /p^j/ and /j/ varied substantially in both positions. In terms of the type of stimuli, there were some significant differences between nonword and real word utterances: For real words, the closing movement of the TB gesture was shorter (NT, DK, and AK), while the TB plateau was longer for all subjects.

3.3. Magnitude and duration of Lip Aperture

3.3.1. Magnitude

The results for the lip gesture, based on the measurements of maximum lower lip height (a) and lip aperture (b) are summarized in Table 6.

Table 6. Mean value of (a) lower lip height (mm above the upper incisors) and (b) lip aperture (in mm) for /p/ and /p^j/ in syllable onset (a#_a, t#_a, k#_a) and syllable coda (a_#a, a_#t, a_#k) in nonwords and real words for four subjects.

		AS (N=113)	NT (N=108)	DK (N=112)	AK (N=88)
a.	onset /p/	–15.93 (1.40)	–10.56 (1.56)	–12.97 (0.60)	–21.30 (1.88)
	coda /p/	–15.86 (1.02)	–10.50 (1.10)	–12.81 (0.42)	–20.40 (1.50)
	onset /p^j/	–15.83 (1.22)	–10.03 (1.25)	–12.69 (0.48)	–21.09 (1.60)
	coda /p^j/	–16.03 (0.86)	–10.24 (1.20)	–12.73 (0.53)	–20.69 (1.55)
b.	onset /p/	not available	3.83 (1.88)	9.23 (0.40)	18.64 (1.58)
	coda /p/	not available	3.76 (1.88)	9.00 (0.38)	17.32 (1.14)
	onset /p^j/	not available	3.58 (1.65)	9.11 (0.45)	18.18 (1.55)
	coda /p^j/	not available	3.71 (1.80)	8.84 (0.43)	17.49 (1.07)

In terms of the magnitude of LL, there were neither significant effects of Position and Consonant, nor significant Consonant X Position interaction for any of the subjects. This means that the magnitude of this gesture did not differ significantly between the consonants, /p^j/ vs. /p/, or between positions (onset vs. coda). For LA, there was a main effect of Position for two of the three subjects [DK: $F(1, 111) = 11.265$, $p < 0.01$; AK: $F(1, 87) = 14.039$, $p < 0.001$; no data for AS]: These speakers showed smaller LA values (i.e., a narrower constriction) in coda position than in onset position. There was no main effect of Consonant for any of the subjects and there was no significant Consonant X Position interaction. An informal examination of the standard deviations of the results (see Table 6) suggests that the LL magnitude was somewhat more variable in onset position than in coda position.

3.3.2. Duration

The closing movement and plateau duration results for the four subjects are shown in Figure 2. The productions for syllable-initial and syllable-final /p^j/ (at the top) and /p/ (at the bottom) are aligned in time by the achievement of the target, the onset of the plateau.

For the closing movement, there was a main effect of Position for three out of four subjects [AS: $F(1, 111) = 41.940$, $p < 0.001$; NT: $F(1, 106) = 7.658$, $p < 0.001$; DK: $F(1, 111) = 53.170$, $p < 0.001$], indicating that the closing movement was substantially shorter for coda consonants: This reduction in duration amounted to as much as 30% of the duration of the movement in onset position (see Figure 2). There was a main effect of Consonant for one subject only [NT: $F(1, 106) = 9.978$, $p < 0.05$], pointing to a slightly longer movement to the constriction of /p/ than to that of /p^j/.

For the duration of the plateau, there was a main effect of Position for all subjects [AS: $F(1, 111) = 17.133$, $p < 0.001$; NT: $F(1, 106) = 9.788$, $p < 0.01$; DK: $F(1, 111) = 96.304$, $p < 0.001$; AK: $F(1, 85) = 14.192$, $p < 0.001$]: The plateaus were consistently shorter in coda position than in onset position, with the reduction amounting to as much as 40% of the duration of the syllable-initial plateau. There was no main effect of Consonant for any of the subjects. There was a significant Consonant X Position interaction for two subjects [NT: $F(1, 106) = 4.146$, $p < 0.05$; AK: $F(1, 85) = 20.586$, $p < 0.001$]: one of the subjects showed the effect for the syllable-final /p^j/, while the other one for the syllable-final /p/. In terms of variability (based on standard deviations; see error bars in Figure 2), the duration of the closing movement in onset position was more variable than in coda position. In terms of the type of stimuli, none of the significant differences in magnitude and duration were consistent across all four subjects.

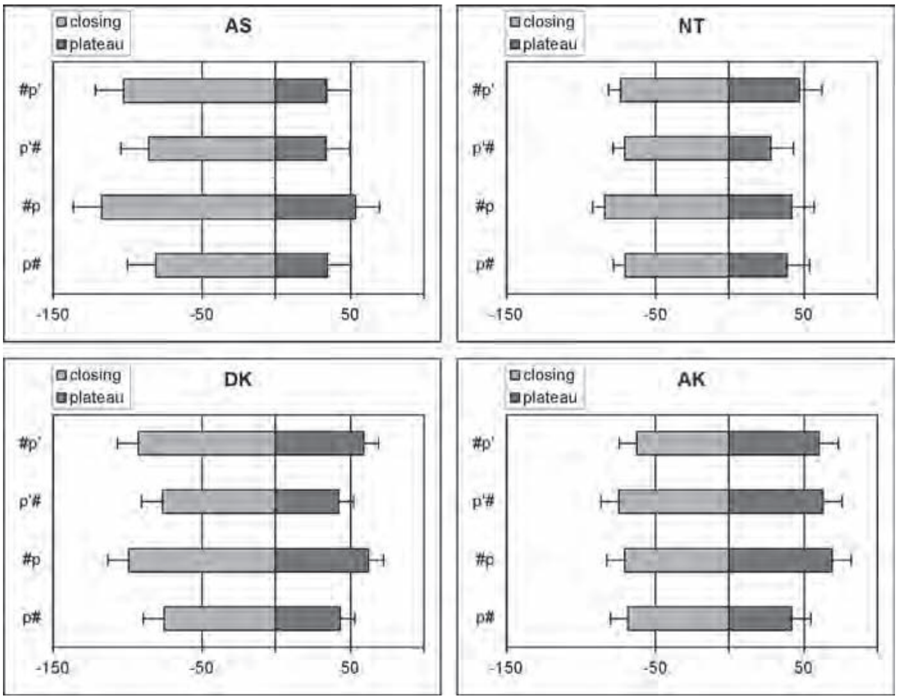


Figure 2. Mean duration of the closing movement and plateau of the lower lip gesture during the articulation of for /p/ and /p^j/ in syllable onset (a#_a, t#_a, k#_a) and syllable coda (a_#a, a_#t, a_#k) in nonwords and real words (in ms) for 4 subjects.

4. Discussion

In this section I discuss the results of the EMMA experiment with respect to the general and specific hypotheses made in Section 1. The results of our experiment largely confirm the general hypothesis: we do find substantial differences between the same consonants in different syllable positions. The syllable effects are manifested primarily in: (a) the timing of the lips and tongue body for the onset and coda /p^j/, which follows the general pattern of timing in the sequences /pj/ and /jp/; (b) the magnitude of the TB raising gesture for /p^j/ and /j/; and (c) the duration of the LA gesture for /p^j/ and /p/. The differences in the magnitude of TB fronting, in the duration of TB raising, and in the magnitude of LA are less consistent, being found only for certain speakers. All these differences are discussed in detail below.

With respect to the timing hypothesis, our findings (see Section 3.1) indicate that the gestures of TB and LA for the Russian /p^j/ show more lag (are more sequential) in onset position and less lag (are more synchronous) in coda position. The lag for the onset /p^j/ is always positive; that is, the more open gesture, TB, is delayed with respect to the more closed gesture, LA. The lag for the coda /p^j/ is either positive or negative: in the latter pattern, the more open gesture (TB) slightly precedes the more closed gesture (LA). The observed timing pattern is not fully consistent with the first prediction made with regard to timing (timing hypothesis A): no lag in onset and a substantial negative lag in coda. Recall that this prediction was based on the pattern observed for the positional variants of the English /l/. While it differs from the timing pattern for the English /l/ (especially in onset), the pattern for the Russian /p^j/ is similar to the one observed for the English /w/ (Gick 2003): For this consonant, LA for the syllable-initial /w/ were found to precede the equally open TB backing gesture in the syllable onset, while the two were near-synchronous in the syllable coda.

The timing of gestures found for the initial /p^j/ is likely influenced by perceptual factors (see Mattingly 1981 and Silverman 1997 on gestural timing and recoverability). Specifically, a somewhat later achievement of the TB target compared to the LA target results in the relatively long and perceptually salient CV transitions characteristic of palatalized consonants in general (mainly high F2; see, for example, Ladefoged and Maddieson 1996: 364; see Kavitskaya, this volume; Kochetov 2002). In addition, this timing results in a high frequency noise burst of /p^j/, which is perceptually different from the weak burst of the non-palatalized /p/. In addition, the timing pattern for the onset /p^j/ (and to some degree for the coda /p^j/) is consistent with sonority generalizations about inter-gestural timing (Sproat and Fujimura 1993, Gick 2003): the more open, more sonorous, TB gesture is timed closer to the syllable peak, than the more closed, less sonorous, LA gesture.

Another important finding of the current study is that the timing of gestures for the onset /p^j/ is similar to those for the word-boundary or within-word sequence /pj/, showing a substantial delay of the TB gesture (a positive lag). Also, the timing of gestures for the coda /p^j/ tends to be in the same direction as the timing in the sequence /jp/. Yet, the timing patterns for the single consonants are distinct from those for the sequences, thus confirming our second prediction about inter-gestural timing (timing hypothesis B). Perceptual factors are likely at play here again: A delay of the TB for the onset /p^j/ by more than 50–60 ms would cause a confusion of the palatalized consonant with the sequence /pj/; similarly, a lead of the TB for the coda /p^j/ by more than 20–30 ms may cause a confusion of the palatalized consonant with the sequence /jp/. Note also that a completely simultaneous production of the two gestures may obscure the acoustic transitions characteristic for a lip closure leading to the percepts [t^j] or [t] (see Kochetov 2002). Thus, the timing patterns found for tongue body and lips in Russian seem to be optimal in a sense that they allow to maintain multiple lexical distinctions that rely on timing of these gestures, and ensure that these gestures (and the corresponding phonological categories) are reliably recovered by listeners.

With respect to the magnitude hypothesis, the current results (see Section 3.2) confirm our prediction about the coda reduction of the TB gesture of /j/ (single and in sequences) in magnitude, primarily in TB raising. The reduction in magnitude of the TB gesture was in some cases accompanied by a reduction in duration. The results are consistent overall with the findings for the English /j/, which shows consistent reduction in TB magnitude (both raising and fronting: Gick 2003). The reduction of the TB gesture in coda was also observed for the palatalized stop /p^j/, thus confirming the first version of the magnitude hypothesis for this consonant (See Section 1). This finding for Russian, however, contrasts with the lack of reduction and some expansion of the secondary velar articulation of the English /l/. Thus, our alternative prediction about the TB magnitude of effects for /p^j/ is not supported. The robust TB magnitude differences between syllable-initial and syllable-final /p^j/ (as well as the timing differences) appear to correlate with significant differences in the perception of these positional variants. In a perceptual study in Kochetov (2004) both native and non-native listeners showed substantially longer reaction time and lower rate of correct identification of /p^j/ syllable-finally than syllable-initially.

Another interesting finding is that the TB gestures for /p^j/ and for /j/ (including for /j/ in the sequences /pj/ and /jp/) were significantly different in the degree of TB raising and, for some speakers, in TB fronting and duration. The TB of /j/ was higher than that of /p^j/ and was often characterized by a shorter plateau. The two consonants also differed in the degree of gestural reduction and variation, with /j/ being more affected by the position. This suggests that the TB gesture of /p^j/ is a distinct articulatory structure, with different target parameters than those of /j/ (either as a single consonant or in a sequence with /p/). Thus, the consonant /p^j/ and the sequences of /p/ and

/j/ in Russian differ not only in the timing of the gestures LA and TB, but also in the magnitude of the TB gesture (especially for the onset /p^j/ and the sequence /pj/).

While our predictions about the reduction of the TB gesture are fully confirmed, the predictions about the magnitude reduction of the lip gesture are not supported (see Section 3.3). None of the subjects showed any reduction of this gesture in coda; in fact, two speakers showed a tighter lip constriction (i.e., expansion in LA) in coda. In addition, our results show no syllable-final LA reduction for the palatalized /p^j/, disconfirming both versions of the magnitude hypothesis for this consonant. Interestingly, however, the lip gesture for both /p/ and /p^j/ in coda was substantially reduced in duration. These results are not consistent with certain findings for English, where labial stops in coda have been reported to be reduced in magnitude, although to a lesser degree than coronals and dorsals (Browman and Goldstein 1995). At the same time, our results for Russian labials are fully consistent with the behavior of Russian stops in general, which do not show reduction in magnitude, but show consistent reduction in duration (Kochetov and Goldstein 2006; cf. the results of an EPG study of the Russian /n/ in Barry 1991). It is possible that the differences between Russian and English in reduction patterns of the LA gesture (and other oral gestures in stops) are related to the considerable differences in the degree of overlap in stop sequences in the two languages, as well as to differences in the proportion of coda stops that are audibly released in clusters and before a pause (Zsiga 2000, Kochetov 2002).

With respect to the variability hypothesis, informal observations of the results suggested that the consonants in coda position were somewhat more variable in terms of inter-gestural timing, TB magnitude, and TB duration (for /p^j/ only). The opposite effect – higher variability in onset – seemed to hold for some lip magnitude and duration measurements. Based on these preliminary observations, the variability hypothesis is only partly confirmed.

5. Summary and conclusions

While providing evidence for the general hypothesis and some of our specific predictions, the results of our experiment challenge some other predictions and, placed in the context of other studies, raise a number of questions: First, why do we observe syllable position effects in the first place? What are the underlying reasons for the phenomenon? Second, how can we explain the observations that only some gestures show syllable position effects and that the actual manifestations of these effects differ from gesture to gesture? Finally, and most importantly for the current study, how can we account for both the similarities and the differences in the degrees and kinds of reduction of similar gestures in different languages?

The discussion of various manifestations of syllable position effects and of their possible underlying causes is well beyond the scope of this paper; however, it is worth outlining a possible direction to follow in explaining the attested patterns. One approach within the framework of Articulatory Phonology (Browman and Goldstein 1989, et seq.) has been seeking to account for syllable position effects attested in English by reference to the dynamics of coordinated articulatory gestures modeled as coupled oscillators (Nam and Saltzman 2003; Browman and Goldstein, in prep.; see Saltzman and Byrd 2000 on the task-dynamics of phasing using oscillator coupling). In particular, the *dynamic approach* assumes two basic types of coordination relations between gestures: in-phase coordination of gestures in syllable-initial position and out-of-phase coordination of gestures in syllable-final position. The first coupling mode is presumed to result in dynamically stable patterns that exhibit high amplitude oscillations of the corresponding gestures; the second mode renders the gestures dynamically unstable, with lower amplitude oscillations. The actual consequences for individual gestures are expected to vary according to differences in their dynamic properties.

While the dynamic account represents a promising direction to follow in identifying and explaining the general mechanism behind the syllable position effects, it is not yet clear whether application of general, language-independent task-dynamics is sufficient to account for the differences in the kinds and degrees of gestural reduction in different languages. It appears that a more plausible (although less parsimonious) alternative is to combine the dynamic approach within Articulatory Phonology with phonetically richer and intrinsically variable exemplar-based lexical representations (Pierrehumbert 2001). In this view, not all manifestations of syllable position effects are direct consequences of gestural dynamics: Some of them may reflect listener-based interpretation of the variable input, stored as detailed gestural structures (see Browman and Goldstein 1995). These structures would encode phasing of gestures and their magnitude parameters for each lexical item, with the values potentially differing in syllable onset and coda. Moreover, a single semantic representation for a lexical item may be associated with a number of alternative gestural structures, or exemplar memories. Weights for the competing structures, as well as values for phasing and other gestural parameters for each particular representation are constantly updated based on an individual's perception/production experience. Crucially, the differences in reduction patterns between various gestures within a language result not only from the dynamic properties of these gestures, but also from how well these units can be recovered by a listener given the variable output of production in various contexts (see, e.g., Surprenant and Goldstein 1998, Chitoran, Goldstein, and Byrd 2002). Differences between languages in types and degrees of reduction of otherwise similar gestures are likely to arise from language-particular gestural coordination patterns (e.g., degree of gestural overlap), more specifically, from essentially random differences the way these patterns are recovered and stored by speakers/listeners. Thus, no one-to-one correspondence between dynamic properties of gestures and realizations of syllable-final reduction is expected; nor should we expect similar gestures in different languages to pattern identically with respect to the reduction process. This approach could potentially capture the general relation between dynamic properties of gestural coordination and syllable position effects, while at the same time accounting for variation in effects within and across languages.

To conclude, more empirical research is needed to identify the range of possible syllable-position effects cross-linguistically. In addition, more theoretical modeling work is necessary to provide new insights into the nature of these effects, and ultimately, into the underlying relation between the microscopic properties of gestural organization and the macroscopic category of the syllable.

Acknowledgments

The author is grateful to Louis Goldstein for assistance with the design of the experiment, data collection, and discussion of the results. Many thanks to Keren Rice, Ron Smyth, and two anonymous reviewers for many helpful comments and suggestions, to Marianne Pouplier, Larissa Chen, and Mark Tiede for invaluable assistance with data collection and analysis. The research was supported by Post-doctoral Fellowship # 756-01-0145 from the Social Sciences and Humanities Research Council of Canada (SSHRC) held at Haskins Laboratories, New Haven, CT. All errors are my own.

References

Avanesov, R. I.
1984 *Russkoe Literaturnoe Proiznoshenie* [Russian literary pronunciation]. Moscow: Prosveshchenie.

Barry, Martin C.
1991 Temporal modeling of gestures in articulatory assimilation. In: *Proceedings of the 12th International Congress of Phonetic Sciences* 4: 1417.

Browman, Catherine P. and Louis Goldstein
1989 Articulatory gestures as phonological units. *Phonology* 6: 201–252.
1995 Gestural syllable position effects in American English. In: Fredericka Bell-Berti and Lawrence J. Raphael (eds.), *Producing Speech: Contemporary Issues* (for Katherine Safford Harris), 19–33. Woodbury, N.Y.: AIP Press.
in prep. Articulatory Phonology. Yale University ms.

Byrd, Dani
1996 Influences on articulatory timing in consonant sequences. *Journal of Phonetics* 24(2): 209–244.

Chitoran, Ioana L., Louis Goldstein and Dani Byrd
2002 Gestural overlap and recoverability: Articulatory evidence from Georgian. In: Carlos Gussenhoven, Toni Rietveld, and Natasha Warner (eds.), *Papers in Laboratory Phonology VII*, 419–448. Berlin: Mouton de Gruyter.

Fougeron, Cecile and Patricia A. Keating
1997 Articulatory strengthening at edges of prosodic domains, *Journal of Acoustical Society of America* 101: 3728–3740.

Gick, Bryan
2003 Articulatory correlates of ambisyllabicity in English glides and liquids. In: John Local, Richard Ogden and Rosalind Temple (eds.), *Papers in Laboratory Phonology VI: Constraints on Phonetic Interpretation*. Cambridge, UK: Cambridge University Press.

Gick, Bryan, Fiona Campbell, Sunyoung Oh and Linda Tamburri-Watt
2006 Toward universals in the gestural organization of syllables: A cross-linguistic study of liquids. *Journal of Phonetics* 34(1): 49–72.

Giles, Stephen B. and Kenneth L. Moll
1975 Cinefluorographic study of selected allophones of English /l/. *Phonetica* 31: 206–227.

Jones, Daniel and Dennis Ward
1969 *The Phonetics of Russian*. Cambridge, UK: Cambridge University Press.

Kavitskaya, Darya
this vol. Perceptual salience and palatalization in Russian.
Kochetov, Alexei
2002 *Production, Perception, and Emergent Phonotactic Patterns: A Case of Contrastive Palatalization.* New York: Routledge.
2004 Language-specific and language-independent asymmetries in the perception of place and secondary articulation contrasts. *Language and Speech* 47(4):351–382.
Kochetov, Alexei and Louis Goldstein
2005 Position and place effects in Russian word-initial and word-medial clusters. *Journal of the Acoustical Society of America* 117(4), Pt. 2: 2571.
Krakow, Rena A.
1989 The articulatory organization of syllables: A kinematic analysis of labial and velar gestures. Ph.D. dissertation, Department of Linguistics, Yale University.
1999 Physiological organization of syllables: A review. *Journal of Phonetics* 27: 23–54.
Ladefoged, Peter and Ian Maddieson
1996 *The Sounds of the World's Languages*. Cambridge, M.A.: Blackwell.
Mattingly, Ignatius G.
1981 Phonetic representation and speech synthesis by rule. In: T. Myers, J. Laver and J. Anderson (eds), *The Cognitive Representation of Speech*. Amsterdam: North Holland: 415–420.
Nam, Hosung and Elliot Saltzman
2003 A competitive, coupled oscillator model of syllable structure. In: *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona: 2253–2256.
Perkell, J. S., M. Cohen, M. Svirsky, M. Matthies, I. Garabieta and M. Jackson
1992 Electro-magnetic midsagittal articulometer (EMMA) systems for transducing speech articulatory movements. *Journal of the Acoustical Society of America* 92: 3078–3096.
Pierrehumbert, Janet
2001 Exemplar dynamics: Word frequency, lenition, and contrast. In: Joan Bybee and Paul Hopper (eds.). *Frequency Effects and Emergent Grammar*, 137–157. Amsterdam: John Benjamins.
Recasens, Daniel and Joaquin Romero
1997 An EMMA study of segmental complexity in alveopalatals and palatalized alveolars. *Phonetica* 54: 43–58.
Saltzman, Elliot and Dani Byrd
2000 Task-dynamics of gestural timing: Phase windows and mulitfrequency rhythms. *Human Movement Science* 19: 499–526.

Silverman, Daniel
1997 *Phasing and Recoverability*. New York, N.Y.: Garland Publishing.
Sproat, Richard and Osamu Fujimura
1993 Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics* 21: 291–311.
Surprenant, Aimée and Louis Goldstein
1998 The perception of speech gestures. *Journal of the Acoustical Society of America* 104: 518–529.
Turk, Alice
1994 Articulator phonetic cues to syllable affiliation: Gestural characteristic of bilabial stops. In: Patricia A. Keating (ed.), *Papers in Laboratory Phonology III: Phonological Structure and Phonetic Form*. Cambridge, UK: Cambridge University Press: 107–135.
Wang, Qi Emily
1995 Are syllables units of speech motor organization? – A kinematic analysis of labial and velar gestures in Cantonese. Ph.D. dissertation, Department of Linguistics, University of Connecticut.
Zsiga, Elizabeth C.
2000 Phonetic alignment constraints: Consonant overlap and palatalization in English and Russian. *Journal of Phonetics* 28 (1): 69–102.

Perceptual salience and palatalization in Russian*

Darya Kavitskaya

This paper reports on a gating experiment testing the relative salience of sonority, place of articulation, voicing and palatalization. The experiment showed that while sonorant segments are identified by the listeners in a higher percentage of cases than non-sonorant ones, the hypothesis that secondary features will be perceptually less salient than primary features is not supported: The cues for palatalization are at least as perceptually salient for the speakers of Russian as cues for voicing and place of articulation.

1. Introduction

Since the advent of the idea that phonological segments can be further analyzed into distinctive features (Jakobson 1939; building on Trubetzkoy 1939), there had been a great deal of research which developed various aspects of feature theory, from the early proposals, e.g. by Clements (1985), Sagey (1986), Ohala (1975) to later developments, as, for example, Padgett (1995), Avery and Idsardi (2001), Clements (2001). The studies in feature theory cover a wide range of topics, such as the nature of featural representations, the issue of binary vs. privative nature of certain features, and various proposals with respect to the organization of features into hierarchical classes, that is, feature geometry; the acoustic properties of features have been invoked in the discussion of all of these topics. While it has been claimed from the early days of the feature theory that distinctive features are necessarily correlated with phonetic properties of sounds, the exact correlation is still a subject of debate. Defining precise acoustic and articulatory properties of features has been problematic, since phonological features are by definition abstract and often have more than one acoustic correlate.[1]

This paper focuses on one aspect of a proposal which posits a classification of distinctive features on the basis of their acoustic and articulatory properties. This proposal is advanced in a series of works by Kenneth Stevens; it is based on Halle and Stevens (1991), continued in Stevens (1994), and further developed in more recent work, such as Stevens (2000; 2001; 2002). The research program developed in Stevens' work provides a phonetically-grounded approach to the hierarchical organization of features and makes a number of predictions with respect to the notion of salience of distinctive features. The subsequent sections will show that while some of these predictions are borne out, the others appear not to hold at least for the data tested. Even though the present experiment cannot fully settle the issue of the relative salience of distinctive features, it nevertheless advances our understanding of the matter.

2. Background and predictions

Both acoustic studies (e.g., Stevens 1989) and auditory studies (e.g., Chistovich and Lublinskaya 1979; Delgutte and Kiang 1984) present evidence in favor of the claim of standard feature theory that segmental units are represented through sets of distinctive features. Building on this evidence, Halle and Stevens (1991) develop a classification of binary features on the basis of the articulatory properties of a given feature. Based on Halle and Stevens (1991), Stevens (1994:242) divides distinctive features into the following three classes:

> (1) articulator-free features that indicate whether a narrow constriction is made in the vocal tract and, if so, whether or not a complete closure is formed and whether pressure is built up behind the constriction; (2) articulator-bound features indicating the primary articulator that is active in forming the constriction (whether it be a narrow consonantal constriction or a less severe vocalic constriction; and (3) articulator-bound features indicating active adjustments of secondary articulators (i.e. articulators other than the primary ones), such as larynx, the soft palate, and the tongue body (for cases in which the tongue body is not the primary articulator).

The classification above corresponds to the standard feature theory in the following way. Articulator-free features in (1) refer to major class features (such as [±consonantal], [±sonorant], [±continuant]), articulator-bound features in (2) cover place of articulation and voicing features, and different type of articulator-bound features in (3) encompasses secondary articulations, such as labialization, pharyngealization, laryngealization, and palatalization.[2]

It has been proposed by Stevens, Keyser and Kawasaki (1986) and Stevens and Keyser (1989) that frequently occurring feature combinations in languages of the world arise since such combinations maximize perceptual distinctiveness through the mechanisms of feature enhancement. The enhancement theory holds that certain features in a sound inventory are perceptually the most salient, or *primary*, as opposed to less salient, or *secondary*, features which are selected to enhance the strength of primary features. This notion of salience is further developed in Stevens (1994) to reflect the proposed featural organization. Stevens (1994:242) states that "[t]his hierarchical organization can serve as a basis for ordering the identification of features from the acoustic signal, with the more context-independent features being identified first and the more context-dependent features identified later."[3]

One of the predictions which follows from the above interpretation of the hierarchical organization of features is that articulator-free features should be identified by listeners before articulator-bound features, and articulator-bound primary features should in turn be more quickly perceived by listeners than articulator-bound secondary features.[4] In other words, articulator-free features are expected to be maximally salient, articulator-bound features produced with the primary articulator are predicted to be less salient, and secondary features produced with the secondary articulators are predicted to be the least salient. This paper is a report on the results and on the theoretical and methodological implications of an experiment designed to test this prediction.[5]

3. Experiment

A perception experiment was designed in order to test the connection of Stevens' feature classification to perceptual salience. This theory generates the following predictions with respect to the timing of the listener's disambiguation of a certain feature in the signal: a) articulator-free features will be disambiguated first; b) the confusions in articulator-bound primary features will be disambiguated second; and c) the confusion in articulator-bound secondary features will clear last. The experiment described below uses several types of confusions made by native Russian listeners to test these predictions. In particular, the confusions between oral and nasal segments are expected to be disambiguated first since they differ with respect to the articulator-free feature as [±sonorant]. The confusions with respect to the feature [±voice] and the place of articulation features [LABIAL] and [CORONAL] are predicted to be cleared second. Finally, the confusion in secondary features (which in the case of Russian is palatalization) is expected to persist for the longest time. The following sections show that while the first prediction is borne out, the other two do not appear to hold for the Russian data tested.

3.1. Materials

The main reason that the study was based on Russian is that the presence of phonemic palatalization in Russian is non-controversial, and almost all non-palatalized segments in the consonant inventory of Russian have palatalized counterparts. The following partial sound inventory was used for the purposes of the experiment:

(1) a. Consonants:

		LABIAL		CORONAL[6]	
STOP	*voiceless*	p	pʲ	t	tʲ
	voiced	b	bʲ	d	dʲ
NASAL		m	mʲ	n	nʲ

b. Vowels:

[u]

[e]

[a]

A list of 36 words was constructed on the basis of this inventory. All words began with a single consonant, followed by a stressed vowel. Even though all phonemically contrastive Russian vowels [i e a o u ɨ] can bear stress, only three vowels were recorded for the experiment. A high front vowel [i] and a high mid vowel [ɨ] were not considered since they are in complementary distribution in Russian: [i] occurs only after the palatalized consonants, while [ɨ] surfaces only after the plain ones, which makes it impossible to have minimal pairs with respect to palatalization.[7] Only one back vowel [u] was considered, in order to keep the size of the experiment manageable.

The recorded words were no longer than two syllables: whenever monosyllabic words were used, they were of CV or CVC syllable structure, and disyllabic words used consisted of open syllables only, with the stress on the first syllable. Most tokens were real words, but it was impossible to balance the list for the frequency of occurrence for the following reasons. First, Russian native vocabulary lacks sequences of non-palatalized consonants followed by [e] almost entirely.[8] To solve this problem, the names of the letters [pe], [be], [de] and [te] were used. These are less frequent than most words in the list, but nonetheless letter names are often pronounced when the alphabet is recited. For the nasal followed by [e] sequences, a borrowed word for 'major' and an acronym were used which are also quite infrequent. Second, Russian lacks stressed syllables like [p^ju] and [m^ju] so nonsense words were used in these cases.[9]

(2) shows the full word list used in the experiment:[10]

(2)	pe	'letter p'	p^jena	'foam'
	papa	'father'	p^jat^j	'five'
	putɨ	'chains'	p^ju	nonce-word
	be	'letter b'	b^jedɨ	'troubles'
	baba	'woman' (colloq)	b^jaka	'unpleasantness'
	budu	'I will'	b^ju	nonce-word
	te	'letter t'	t^jema	'topic'
	tak	'thus'	t^japka	'hoe'
	tuk	'knock'	t^juk	'bundle'
	de	'letter d'	d^jedɨ	'grandfathers'
	data	'date'	d^jad^ja	'uncle'
	duma	'thought'	d^juna	'dune'
	mer	'major'	m^jed^j	'copper'
	mama	'mother'	m^jata	'mint'
	muka	'torment'	m^ju	word-word
	nep	acronym	n^jet	'no'
	nado	'necessary'	n^jan^ja	'nanny'
	nu	'so'	n^jura	name

The word list was recorded as read by a male speaker of Russian from Moscow. The recording was made on a DAT recorder in a double-walled sound booth. After the recording, individual tokens were saved as audio files (.wav, the sampling rate 20 kHz) and subsequently transferred to the *Waves* program. A gating program (part of the *Waves* program) was used to truncate words from the releases of consonants at the first gate of 30 ms (and then 60, 90, 120 ms) of any given word. Stop releases were marked at the first indication of the burst, and nasal releases were labeled at the place of amplitude and/or spectral discontinuity in nasals. The remaining part of a word was replaced by gaussian noise[11]. After truncation, only the first consonant and the first part of the following vowel remained. No information about transitions to postvocalic consonants which could provide additional undesired cues for word recognition was included in the signal.[12]

3.2. Procedure

The *SoundEdit* program was used to design the part of the experiment responsible for the program-listener interaction. The stimuli were organized into a 144-token list (12 consonants x 3 vowels x 4 gates) in random order. 10 native Russian listeners participated in the experiment. For each token played, a card with 12 fixed response options was offered to a listener on a computer screen. For example, if a listener heard a truncated word such as [papa] 'father', a card with twelve possible choices of a CV sequence labeled in Cyrillic was shown. (3) shows a sample card with fixed responses for all CV sequences with the vowel [a]: [pa], [pʲa], [ba], [bʲa], [ta], [tʲa], [da], [dʲa], [ma], [mʲa], [na], and [nʲa].[13]

(3) A sample card for *Ca* tokens

па	пя	ба	бя
та	тя	да	дя
ма	мя	на	ня

The listeners were instructed to press a button corresponding to the sequence they heard. Each token was played only once, and a listener had unlimited time to click the button of their choice. As soon as the response was selected, the next token was played.

3.3. Analysis

After the data collection, confusion matrices[14] based on listeners' responses were constructed for each listener at four gates, for all listeners together at four gates, and for all responses as a whole (see Appendix 1). The results were analyzed by running repeated-measures ANOVA. (Statistical Package for the Social Sciences (SPSS) was used for the analysis.) Statistical analysis performed on the listeners' responses (dependent variable: total number of errors) showed that the effect of the time (gate) as an independent factor was significant at $p < 0.01$ and the effect of the vowel and of the vowel-gate interaction was not significant. The further analyses used the following independent factors:

(1) Sonority with two levels (nasal and oral);
(2) Obstruent voicing with two levels (voiced and voiceless);
(3) Place of Articulation with two levels (labial and dental);
(4) Palatalization with two levels (plain and palatalized).

The percentage of confusions was calculated as the percentage of erroneous responses given the total possible number of errors for a given feature. The analysis reported below is based on the results including the three vowels used in the study and all ten subjects who participated in the listening experiment.

4. Results and discussion

The following sections present the results of the experiment in the form of graphs showing the percentage of confusion with respect to the following parameters:

I. Articulator-free "primary" feature [±sonorant]
   – sonorants (nasals) heard as obstruents (stops) and vice versa
II. Articulator-bound "primary" features
   – labials heard as dentals and vice versa [Place of Articulation]
   – voiced segments heard as voiceless and vice versa [±voice]
III. Articulator-bound "secondary" features
   – palatalized segments heard as non-palatalized and vice versa

The percentage of confusion is shown separately for each of the four gates. In the statistical analysis, the dependent variable is always defined as the percentage of errors, and time (gate) is treated as an independent factor along with other parameters.

4.1. Primary features

4.1.1. Sonority

First, we consider the direction of confusion in primary features, such as sonority and voicing. Figure 1 illustrates the direction of confusion for nasals and stops. The three graphs show nasals misperceived as oral sounds (N-O), and the confusion of obstruents as nasals is broken down further to voiced stops misheard as nasals (V-N) and voiceless stops misheard as nasals (Vl-N).

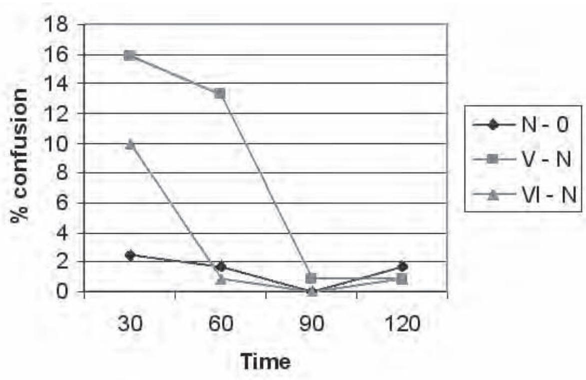


Figure 1. Sonority confusions: oral vs. nasal.

Statistical analysis was performed with gate and sonority as dependent factors. The main effects of Gate was significant (F(3,27) = 5.03, p < 0.01, as was Sonority (F(2,18) = 3.25, p < 0.1). The interaction was also significant (F(6,54) = 2.42, p < 0.05).

There are two noticeable asymmetries in these data. First, as predicted by Stevens (1994), the confusion of sonorants with obstruents is the smallest at all four gates. Statistical analysis shows that the effect of gate is significant, but sonority is not (marginally significant at p < 0.1). This result holds independent of the direction of confusion and provides support to the prediction that the feature [±sonorant] (or arguably [nasal] in this case) should be identified first in the sound signal as the most salient.

It can be hypothesized that the asymmetry in the nasal/stop confusions could be connected with the privative nature of the feature [nasal] which corresponds to velum lowering. Privativity of [nasal] was argued for by Steriade

(1993a; 1993b), Trigo (1993), and in subsequent work. Steriade (1995:149) states that "[t]here is virtually no evidence left suggesting that orality is represented phonologically, in any language."[15] The facts in Figure 1 follow naturally from the privativity of [nasal]. Once the privative feature is present in the signal, it is unambiguously heard as such. Once it is absent, there is no information with respect to the nasalization present, so the listeners can interpret the signal as either nasal or oral.[16]

Second, we can see that at the first two gates voiced obstruents are confused with nasals significantly more often than voiceless obstruents. Bonferroni tests indicate that the percent of confusions for voiced stops differed both from voiceless stops and nasals which did not differ from each other. I suggest that this asymmetry is to be expected. In Russian, phonologically voiced obstruents (both plain and palatalized) are fully voiced phonetically (Bolla 1981; Matusevich 1976), so the vocal cord vibration continues throughout the closure interval (Stevens 1998). Such "prevoicing" in stops can be easily interpreted as nasalization, especially syllable-initially, as in the case of our data.[17]

4.1.2. Voicing

The direction of confusion with respect to voicing is asymmetrical as well, as shown in Figure 2: voiceless segments are heard as voiced consistently more often than voiced segments are heard as voiceless (nasals were not included in this comparison).

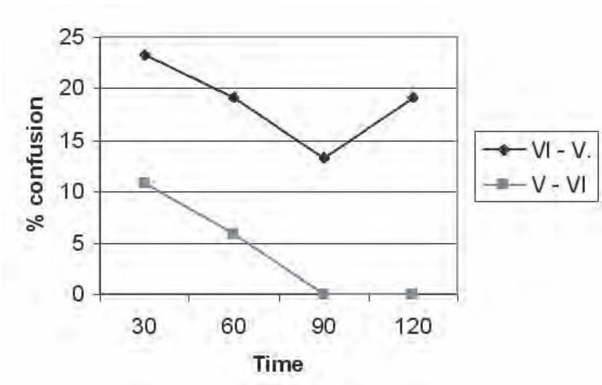


Figure 2. Voicing confusion.

The dependent variables for ANOVA were gate and voicing of the initial consonant. Confusions varied by Gate ($F(3,27) = 5.24$, $p < 0.01$). Voiceless segments are heard as voiced significantly more often than voiced segments are confused with voiceless ($F(1,9) = 10.99$, $p < 0.01$). This occurred at all four gates, as seen in the lack of a significant interaction ($F(3,27) = 1.17$, n.s.).

It is tempting to analyze voicing confusion asymmetry in the same way we analyzed the directionality of confusion in nasals vs. orals. If we assume that [voice] is privative, as we did for [nasal] in discussing the graphs in Figure 1, we can hypothesize that this feature is disambiguated by the listeners as soon as it is present in the signal; if the feature [voice] is missing, as in the case of voiceless stops, the listeners could be guessing either its presence or its absence. Although this is an appealing solution, the privative nature of voicing is considerably more controversial than the privativity of [nasal]. Researchers such as Mester and Ito (1989) and Lombardi (1991; 1995) argued for the privativity of [voice] (and some phonologists, for example, van Rooy and Wissing (2001), assume voicing to be privative without further discussion), but Rubach (1996) provides a convincing argument against this view.[18]

4.2. Secondary features: palatalization

We now turn to the main topic of this paper, palatalization as a factor in consonant confusion. The patterns of confusion with respect to place of articulation and palatalization appear to go against the predictions of Stevens (1994). Recall that it is predicted that place of articulation features would be more robust and more readily (quickly) heard than secondary articulation features, such as palatalization.

The graphs in Figure 3 show that that the percentage of confusion of palatalized segments with non-palatalized is slightly higher than of non-palatalized segments with palatalized (where X stands for all non-palatalized tokens while X' signifies all palatalized tokens). However, the statistical analysis has shown that this tendency is not statistically significant ($F(1,9) = 1.19$, n.s.).

Even though the difference in the direction of confusion with respect to palatalization is not significant, patterns emerge with the further analysis of the palatalization data, as soon as such factors as sonority, voicing, and place of articulation of the palatalized vs. plain consonants in question are taken into consideration. These patterns are presented in the following sections.

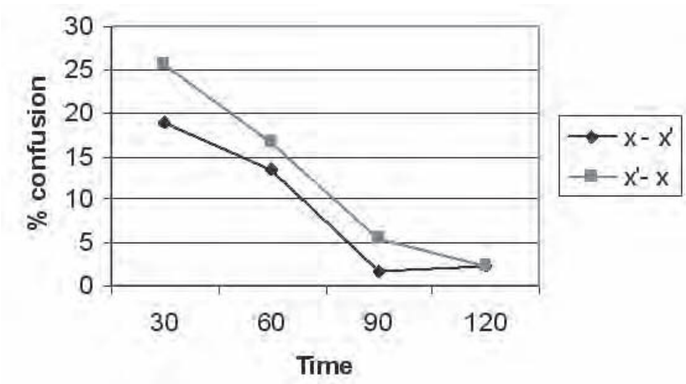


Figure 3. Palatalization confusion.

4.2.1. Palatalization and sonority

The graph in Figure 4 shows confusions of palatalization in orals (any plain oral segment heard as palatalized and vice versa) and in nasals (any plain nasal segment heard as palatalized and vice versa). It is evident that palatalization is hardly ever confused in nasals, as opposed to orals, especially in the first two gates.[19]

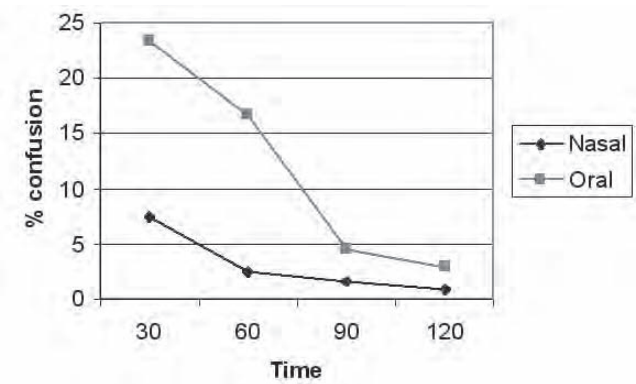


Figure 4. Palatalization confusion in orals vs. nasals.

Both Gate ($F(3,27) = 27.57$, $p < 0.001$) and Sonority ($F(2,18) = 71.91$, $p < 0.001$) were statistically significant. There was also a significant interaction ($F(6,54) = 11.9$, $p < 0.001$).

The asymmetry in Figure 4 appears to have an explanation in acoustics. Even though in nasals and orals the phasing of the palatalization gesture with respect to other gestures is similar (Louis Goldstein, *p.c.*), in the case of stops no acoustic information is available during the first two gates (30 and 60 ms) to disambiguate palatalization. The cue for palatalization is mainly the transition to the following vowel which is lacking in the first two gates.[20] In the case of palatalized nasals, however, the difference in acoustics is present in nasal closure (Bolla 1981), which prevents confusion of palatalized vs. non-palatalized nasals.[21]

4.2.2. Palatalization and place of articulation

The next factor to be considered in the palatalization confusion is place of articulation. Figure 5 shows that palatalized labials are confused with their plain counterparts significantly more often than palatalized dentals are confused with plain dentals.

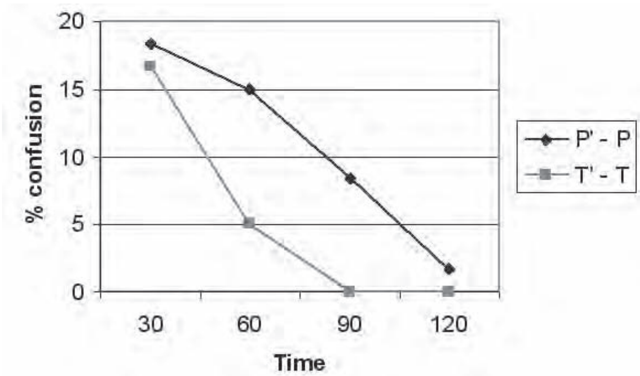


Figure 5. Palatalized as plain confusion, by place of articulation.

Both Gate ($F(3,27) = 15.93$, $p < 0.001$) and POA ($F(1,9) = 33.75$, $p < 0.001$) were statistically significant while the interaction was not ($F(3,27) = 1.82$, n.s.).

The asymmetry present in Figure 5 is expected. Kochetov (2002:116) predicts that "the palatalized labial /p^j/ should be more likely to be perceived as plain than the palatalized coronal /tʲ/...". The reason for this prediction lies in the articulatory properties of palatalized labials and coronals. In both palatalized labials and coronals two articulators are involved, but there is nevertheless the difference in production. According to Ladefoged and Maddieson (1996:364), "bilabial sounds do not require any specific position of the tongue for their articulation, the tongue body can assume an **i**-like position during their production without any conflict with the demands of the primary articulation." However, in the case of the palatalized coronal the primary articulator is shifting its position. Ladefoged and Maddieson (1996:365), note that palatalization of a coronal "consists of a displacement of the surface of the tongue front from the position that it would assume in the non-palatalized counterpart, when its role is to support the movement of the tongue tip or blade," so that palatalized articulation of coronals "can be viewed as the summation of two movements, with the displacement of the tongue front often producing a slightly different primary constriction location." These differences between the interaction of labiality and coronality of a segment with the palatalization gesture can explain the asymmetry in Figure 5. Acoustically, the palatalization gesture affects the coronal burst more than the labial burst, so more information about palatalization is present earlier in the signal. In the case of a labial, palatalization is more an offglide effect detected later.

Since the palatalization gesture is similar to the coronal gesture we predict that palatalized labials which have coronal cues will be confused with coronals [t] and [d] more often than non-palatalized labials, which do not have such cues. Figure 6 shows graphs for the confusions of plain and palatalized labials with dentals.

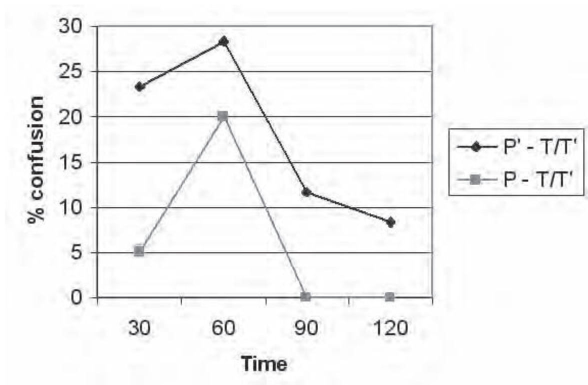


Figure 6. Palatalized and plain labials confused with dentals.

There is clear change across by Gate ($F(3,27) = 20.97$, $p < 0.001$). There is also a statistically significant tendency for palatalized labials to be confused with dentals more often than for the plain ones ($F(1,9) = 7.92$, $p < 0.02$). There was no interaction ($F(3,27) < 1$, n.s.).

4.2.3. Palatalization and voicing

Finally, Figure 7 addresses the direction of palatalization confusion in voiced vs. voiceless stops. The graphs in Figure 7 illustrate four types of confusions:

(1) voiced plain stops heard as voiced palatalized stops (V – V');
(2) voiced palatalized stops heard as voiced plain stops (V' – V);
(3) voiceless plain stops heard as voiceless palatalized stops (VL – VL');
(4) voiceless palatalized stops heard as voiceless plain stops (VL' – VL).

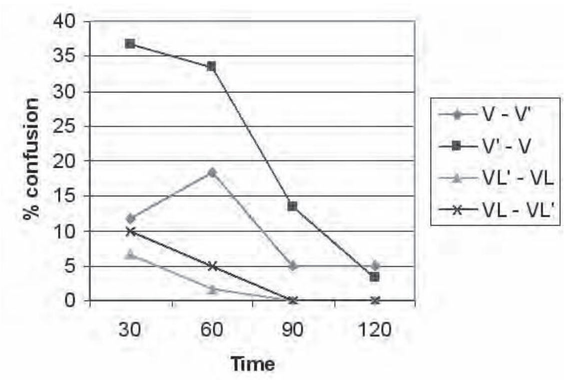


Figure 7. Palatalization confusion in voiced vs. voiceless (by direction, both labial and dental).

Gate was highly significant ($F(3,27) = 13.27$, $p < 0.001$). Plain voiced stops are significantly more likely to be misperceived ($F(1,9) = 11.25$, $p < 0.001$), and the interaction with Gate was also significant ($F(3,27) = 4.94$, $p < 0.007$). Palatalization has no effect ($F(1,9) = 2.35$, n.s.).

I suggest that the explanation of this asymmetric pattern is crucially connected with aerodynamic constraints that govern the generation of turbulence in air flow, specifically, with the effect of a following high vowel, glide, or palatalization *per se* on the burst of a stop. Due to the high turbulence noise caused by the changes of aperture, palatalized voiceless stops frequently have longer and more fricated releases which sometimes results in their phonological spirantization (Kavitskaya 1997; Ohala 1989). The presence of high frequency noise and a longer duration of noise upon the release can help listeners to distinguish voiceless palatalized stops from their non-palatalized counterparts early in the signal.

5. Summary and conclusions

The listening experiment described above provided an opportunity to test the relative salience of sonority, place of articulation, voicing and palatalization. The results of the experiment partially supported the predictions of Stevens (1994). Indeed, nasals were identified by the listeners more readily and quickly than orals. However, the direction of confusion was asymmetrical: orals were identified as nasals more often than vice versa. We proposed that the privative nature of the feature [nasal] was responsible for this asymmetry.

We have also seen an asymmetry in confusions with respect to voicing. First, voiced stops were confused with nasals more often than voiceless stops. We have proposed that this asymmetry was connected with acoustic properties of stops and nasals, and that prevoicing in voiced stops could be mistaken for nasalization and responsible for the higher rate of confusion. Second, voiceless stops were confused with voiced stops more often than vice versa. We tentatively suggested that the privativity of [voice] might be responsible for this asymmetry.

Contrary to the expectations, the difference in confusion between place of articulation and palatalization was not statistically significant. However, a more subcategorized analysis revealed further asymmetries in the direction of confusion. First, palatalized nasals were identified better than palatalized orals. We proposed that this is due to the fact that palatalization can be detected very early in the nasal closure. Second, palatalization cues were missed in the case of palatalized labials significantly more often than in palatalized coronals. Since acoustically the palatalization gesture affects the coronals more than labials, this asymmetry received a phonetic explanation as well.

The results of this experiment have shown that cues for palatalization are at least as perceptually salient for the speakers of Russian as cues for voicing or place of articulation. This suggests that the distinction between the primary and secondary features is not correlated with the notion of perceptual salience. This study provides evidence that features are not extracted from a sound signal in some particular order. Rather, the listeners disambiguate the featural information whenever it is present, and if the interaction of acoustic and articulatory properties makes it possible to detect certain features early in the acoustic signal, it is done so by the listeners.

Appendix

Perceptual confusions among consonants.

Total perception; Gate = 30 ms													
	Produced												
		p	p^j	t	t^j	b	b^j	d	d^j	m	m^j	n	n^j
Perceived	p	21	4	2				3				1	
	p^j	4	12		4	3	3		1				
	t			14									
	t^j		1	2	16	1		2					1
	b	3	1			9	6	8	4		1		
	b^j	2	5			1	3	3	1				
	d			7	6	8	8	9	4				
	d^j		2		2	4	3	1	16				
	m								2	27	1	4	2
	m^j		2	2	1	1	2	1	1	2	25		5
	n			1		3	3	1	1	1	1	24	2
	n^j		3	2	1		2	2			2	1	20

Total perception; Gate = 60 ms													
	Produced												
		p	p^j	t	t^j	b	b^j	d	d^j	m	m^j	n	n^j
Perceived	p	26											
	p^j	1	18	1	2	1	1	2					
	t			22	1								
	t^j		3	1	21		1	2			1		
	b	2				11	9	5	2				1
	b^j		6			3	3	2					
	d	1		6		7	7	14	2				
	d^j		2		6	4	4	2	22				
	m					1			1	30	1		1
	m^j		1				1		1		25		2
	n					1	4	1	1			29	
	n^j					2		2	1		3	1	26

Total perception; Gate = 90 ms													
		Produced											
		p	p^j	t	t^j	b	b^j	d	d^j	m	m^j	n	n^j
Perceived	p	30											
	p^j		23		1								
	t			22									
	t^j		2		26								
	b					30	5	2					
	b^j		5				19						
	d			8			3	25					
	d^j				3		2	3	30				
	m									30		1	1
	m^j						1				29		3
	n										1	29	
	n^j												26

Total perception; Gate = 120 ms													
		Produced											
		p	p^j	t	t^j	b	b^j	d	d^j	m	m^j	n	n^j
Perceived	p	30											
	p^j		15		1								
	t			22									
	t^j		1		27								
	b		1			27							
	b^j		11			2	27	1					
	d			7			2	29		1		1	
	d^j		1	1	2		1		30				
	m					1				29			
	m^j		1								28		2
	n										1	29	
	n^j										1		28

Notes

* I would like to thank Louis Goldstein and John Ohala for their help and encouragement and Doug Whalen and Matthew Richardson for help with statistics. I owe much gratitude to two reviewers of the manuscript. I am also indebted to Jonathan Barnes for insightful comments and helpful discussions.

1. For example, Lisker (1986) lists the following perceptual cues which can be relevant for the listener in categorizing a sound as voiced or voiceless: pre-closure cues (vowel duration, duration of F_1 transition, F_1 offset frequency, F_1 transition offset time, time of voice offset, F_0 contour, decay time signal); closure cues (closure duration, duration of glottal pulsing, intensity of glottal signal), and

post-closure cues (release burst intensity, VOT, onset of F_1 transition, F_1 onset frequency, F_1 transition duration, F_0 contour).

2. Voicing belongs to group (2); even though larynx is mentioned in (3), it is only to refer to laryngealization as a secondary articulation.
3. However, see Whalen (1984) study for an argument that certain secondary cues (e.g. transitions of a fricative vowel) can be taken into account even when the "primary" cues (fricative noise) are unambiguous.
4. In this paper, I will treat the prediction that certain features should be "identified first" as roughly equivalent to "identified best". I thank an anonymous reviewer for pointing this out.
5. Note that we are not questioning other premises of the enhancement theory, such as the claim that secondary features will be included in languages' inventories only given the selections of primary ones.
6. Note that [t] and [d] in Russian have dental, rather than alveolar, articulation.
7. Padgett (2001) argues that the contrast between consonants before /i/ and /ɨ/ is that of palatalized vs. velarized ones.
8. Such sequences occur in recent unassimilated borrowings, e.g. [**te**mp] 'tempo', [po**te**ntsija] 'potency', etc.
9. Nonsense words did seem not present any problems to the test group. According to the listeners, these tokens with did not sound any more foreign or unnatural than the others.
10. Unstressed vowels undergo reduction in Russian which is not shown since it is irrelevant for the purposes of this work.
11. The noise portions were uniform in amplitude and duration, not modulated to mimic the missing part of the word.
12. It was also important to control for the third consonant in a word used as a token. The words were chosen so this consonant would not be a fricative, in order not to provide listeners with additional cues which would be present in a vowel.
13. In Cyrillic alphabet, the distinction between plain and palatalized consonants is marked on the following vowel. Thus, [pa] is written as 'па', while [p^ja] is written with the same sign for the consonant but a different one for the vowel,'пя'.
14. See Miller and Nicely (1955) on the procedure of interpreting confusion matrices.
15. Steriade (1995: 170) mentions that according to Cohn (1990) there is evidence that "oral segments may possess specified articulatory targets for a raised velum position." However, since the presence of a phonetic target does not necessarily imply the presence of a corresponding phonological feature, we do not consider it to be counterevidence to the privativity of [nasal].
16. This analysis constitutes a possible counterexample to John Ohala's theory of asymmetrical confusion. According to Ohala (2001), it is easier to mishear a feature which is present than to "hallucinate" an absent feature.

17. I suspect that different results would follow if syllable-final postvocalic consonants were tested in the same set-up. For syllable position effects see Kochetov (this volume).
18. As was mentioned in Footnote 1, voicing has many acoustic cues. It is also unclear if the feature [voice] can be associated with one gesture. This feature can be interpreted as larynx lowering, as well as vocal fold vibration. Jessen (2001) notes that while the feature [voice] is generally interpreted as vocal fold vibration, such vibration is not a single phenomenon but several gestures that need to be coordinated.
19. Note that the direction of confusion of palatalization in nasals (that is, palatalized nasal as plain vs. plain nasal as palatalized) is not statistically significant.
20. But see discussion below for the differences between palatalized labials and coronals.
21. The same is true of the formant structure of laterals which were not tested here. I predict that the same effects will be observed for liquids tested in a similar experiment.

References

Avery, Peter and William J. Idsardi
2001 Laryngeal dimensions, completion and enhancement. In: T. Alan Hall (ed.), *Distinctive Feature Theory*, 41–70. Berlin: Mouton de Gruyter.

Bolla, Kálmán
1981 *A conspectus of Russian speech sounds*. Köln: Böhlau.

Chistovich, L. A. and V. V. Lublinskaya
1979 The 'center of gravity' effect in vowel spectra and critical distance between the formants: Psychoacoustical study of the perception of vowel-like stimuli. *Hearing Research* 1: 185–195.

Clements, G. N.
1985 The geometry of phonological features. *Phonology Yearbook* 2: 223–250.
2001 Representational economy in constraint-based phonology. In: T. Allan Hall (ed.), *Distinctive Feature Theory*, 71–146. Berlin: Mouton de Gruyter.

Cohn, Abigail
1990 Phonetic and phonological rules of nasalization. Doctoral dissertation, UCLA.

Delgutte, Bertrand and Nelson Y. S. Kiang
1984 Speech coding in the auditory nerve: IV. Sounds with consonant-like dynamic characteristics. *Journal of the Acoustical Society of America* 75: 897–907.

Halle, Morris and Kenneth N. Stevens
1991 Knowledge of language and the sounds of speech. In: Johan Sundberg, Lennart Nord and Rolf Carlson (eds.), *Music, Language, Speech, and Brain: Proceedings of an International Symposium at the Wenner-Gren Center, Stockholm, 5–8 September 1990*, 1–19. London: Palgrave Macmillan.
Jakobson, Roman
1939 Observations sur le classment phonologique des consonnes. In: *Proceedings of the 3rd International Congress of Phonetic Sciences*, 34–41.
Jessen, Michael
2001 Phonetic implementation of the distinctive auditory feature [voice] and [tense] in stop consonants. In: T. Allan Hall (ed.), *Distinctive Feature Theory*, 237–294. Berlin: Mouton de Gruyter.
Kavitskaya, Darya
1997 Aerodynamic constraints on the production of palatalized trills: the case of the Slavic trilled [r]. In: *Proceedings of the 5th European Conference of Speech Communication and Technology*, Volume 2, 751–754. Rhodes, Greece.
Kochetov, Alexei
2002 *Production, Perception, and Emergent Phonotactic Patterns: A Case of Contrastive Palatalization*. New York/London: Routledge.
Ladefoged, Peter and Ian Maddieson
1996 *The Sounds of the World's Languages*. Oxford: Blackwell Publishing.
Lisker, Leigh
1986 "Voicing" in English: a catalogue of acoustic features signaling /b/ versus /p/ in trochees. *Language and Speech* 10: 1–28.
Lombardi, Linda
1991 Laryngeal features and laryngeal neutralization. Ph.D. dissertation, University of Massachusetts-Amherst.
1995 Laryngeal features and privativity. *Linguistic Review* 1: 35–59.
Matusevich, Margarita I.
1976 *Sovremennyi Russkii Iazyk: Fonetika [Contemporary Russian Language: Phonetics]*. Moscow: Prosveshchenie.
McCarthy, J. J.
1988 Phonetic explanation for nasal sound patterns. In: C. A. Ferguson, L. M. Hyman, and J. J. Ohala (eds.), *Nasalfest: Papers from a symposium on nasals and nasalization*, 289–316. Stanford University, Language Universals Project.
Mester, R. Armin and Junko Itô
1989 Feature predictability and underspecification: Palatal prosody in Japanese mimetics. *Language* 65: 258–293.

Miller, George A. and Patricia E. Nicely
1955 An analysis of perceptual confusions among some English consonants. *Journal of the Acoustic Society of America* 27: 338–352.

Ohala, John J.
1975 Phonetic explanation for nasal sound patterns. In: Charles A. Ferguson, Larry M. Hyman and John J. Ohala (eds.), *Nasalfest: Papers From A Symposium On Nasals and Nasalization*, 289–316. Stanford: Stanford University Language Universals Project.
1989 Sound change is drawn from a pool of synchronic variation. In: Leiv E. Breivik and Ernst H. Jah (eds.), *Language Change: Contributions to the Study of Its Causes*, 173–198. Berlin: Mouton de Gruyter.
2001 Markedness and consonant confusion asymmetries. In: Elizabeth Hume and Keith Johnson (eds.), *The Role of Speech Perception in Phonology*, 79–101. Orlando: Academic Press.

Padgett, Jaye
1995 *Stricture in feature geometry*. Stanford: CSLI Publications.
2001 Contrast dispersion and Russian palatalization. In: Elizabeth Hume and Keith Johnson (eds.), *The Role of Speech Perception in Phonology*, 187–218. Orlando: Academic Press.

Rubach, Jerzy
1996 Nonsyllabic analysis of voice assimilation in Polish. *Linguistic Inquiry* 27: 69–110.

Sagey, Elizabeth
1986 The representation of features and relations in nonlinear phonology. Ph.D. dissertation, MIT.

Steriade, Donca
1993a Closure, release and nasal contours. In: M. Huffman and R. Krakow (eds.), *Nasality*. San Diego: Academic Press.
1993b Orality and markedness. In: B. Keyser and J. Guenther (eds.), *Papers from BLS 19*. Berkeley, CA: BLS.
1995 Underspecification and markedness. In: John A. Goldsmith (ed.), *The Handbook of Phonological Theory*, 114–174. Blackwell Publishers.

Stevens, Kenneth N.
1989 On the quantal nature of speech. *Journal of Phonetics* 17: 3–46.
1994 Phonetic evidence for hierarchies of features. In: Patricia Keating (ed.), *Phonological Structure and Phonetic Form: Papers in Laboratory Phonology III*, 242–258. Cambridge: Cambridge University Press.
1998 *Acoustic Phonetics*. Cambridge, MA: MIT Press.
2000 Diverse acoustic cues at consonantal landmarks. *Phonetica* 57: 139–151.

2001 The properties of the vocal-tract walls help to shape several phonetic distinctions in language. *Travaux du Cercle Linguistique de Copenhague* 31: 285–297.

2002 Toward a model for lexical access based on acoustic landmarks and distinctive features. *Journal of the Acoustical Society of America* 111: 1872–1891.

Stevens, Kenneth N. and Samuel J. Keyser

1989 Primary features and their enhancement in consonants. *Language* 65: 81–106.

Stevens, Kenneth N., Samuel J. Keyser and Haruko Kawasaki

1986 Toward a phonetic and phonological theory of redundant features. In: Joseph S. Perkell and Dennis H. Klatt (eds.), *Invariance and Variability in Speech Processes*, 241–254. Dordrecht: Foris Publications.

Trigo, Loren

1993 The inherent structure of nasal segments. In: M. Huffman and R. Krakow (eds.), *Nasality*. San Diego: Academic Press.

Trubetzkoy, N. S.

1939 *Grundzüge der Phonologie. [Principles of phonology. English translation by C. A. M. Baltaxe, University of California Press, 1969.]*. Göttingen: Vandenhoeck & Ruprecht.

van Rooy, Bertus and Daan Wissing

2001 Distinctive [voice] implies regressive voicing assimilation. In: T. Allan Hall (ed.), *Distinctive Feature Theory*, 295–334. Berlin: Mouton de Gruyter.

Whalen, D. H.

1984 Subcategorial phonetic mismatches slow phonetic judgements. *Perception and Psychophysics* 35: 49–64.

Integrating coarticulation, assimilation, and blending into a model of articulatory constraints

Daniel Recasens

An electropalatographic investigation of Catalan consonant clusters composed of dentals, alveolars and alveolopalatals may be accounted for on the basis of consonant-dependent differences in degree of articulatory constraint (where articulatory constraint is conditioned by the degree of involvement of the speech articulators in the formation of a closure or constriction). Thus, more unconstrained stops and nasals undergo complete regressive assimilation to a following more highly constrained fricative or trill, and clusters composed exclusively of unconstrained consonants result in blending through contact addition rather than through the formation of an intermediate articulation. Manner requirements on laterals explain why these consonants are usually produced with a very front closure and avoid assimilation and blending. Progressive place assimilation occurs rarely in constrained - unconstrained consonant clusters, which is in line with carryover coarticulation effects being less systematic and less temporally fixed than anticipatory effects. Sound changes involving place and manner assimilation and elision in clusters appear to be in support of this degree of articulatory constraint model.

1. Introduction

The nature of the adaptation processes in place of articulation between consecutive lingual consonants in clusters is by no means clear judging both from theoretical predictions and from production data in the literature. A comprehensive account of those processes is provided by Articulatory Phonology. According to Articulatory Phonology, the consonant-to-consonant adaptation outcome depends on whether the two consonants in the cluster are placed on different tiers or on the same tier, i.e., a hierarchical phonological structure organized by the degree of independence of articulatory organs

(Browman and Goldstein 1989, 1990, 1992). Consonantal gestures which are relatively independent and thus, located on different tiers (e.g., the tongue tip gesture for final /d/ and the tongue dorsum gesture for initial /g/ in "good girl") may overlap to different degrees, and yield reduction and even hiding of the first consonant in the cluster (C1) rather than elision or assimilation of this consonant to the following consonant (C2). On the other hand, two contiguous consonantal gestures on the same tier (e.g., the tongue dorsum gestures for /g/ and for /j/ in "beg your") are expected to undergo blending at an intermediate constriction location.

Articulatory data for several languages suggest however that regressive assimilation processes may occur in sequences composed of consonants on different tiers, e.g., a dental or an alveolar followed by a dorsal consonant such as /tk/ > [kk] and /sʃ/ > [ʃʃ] in English (Ellis and Hardcastle 1999; Holst and Nolan 1995) or /nk/ > [ŋk] in Italian (Farnetani and Busà 1994). Moreover, differently from clusters with a dorsal consonant, CC sequences composed exclusively of apical and laminal consonants exhibiting contiguous places of articulation and thus located on the same tier may be implemented through C1-to-C2 sliding instead of through blending, as for example /sd/ and /rd/ in Catalan for the production of which the tongue tip travels from the alveolar position for /s, r/ to the dental target for /d/ (Recasens 1995). In summary, the articulatory mechanisms of clusters composed of consonants articulated with the same or adjacent tongue regions appear to be complex and closely dependent on the manner and place requirements at work.

In the present paper, the production mechanisms of consonant clusters will be explored via a reanalysis of electropalatographic (EPG) data for Catalan CC sequences composed of dental, alveolar and alveolopalatal consonants on the same tongue front tier reported in Recasens and Pallarès (2001). A careful investigation of those mechanisms reveals that C-to-C adaptation strategies conform to the severity of the production requirements for the consonantal gestures involved, and may be interpreted in the light of the DAC (degree of articulatory constraint) model of coarticulation (Recasens, Pallarès and Fontdevila 1997).

The DAC model is based on the assumption that phonetic segments are characterized for different degrees of articulatory constraint depending on the degree of involvement of the speech articulators in the formation of a closure or constriction. Thus, alveolopalatal and palatal consonants are believed to be more constrained than labials (and /i/ more constrained than /a/) at the tongue dorsum since this lingual articulator is required to achieve an articulatory target for the implementation of the former class of segments vs the latter. Dentoalveolars would generally fall in between since the tongue dorsum is not directly involved in their production but subject to coupling effects with the primary articulator.

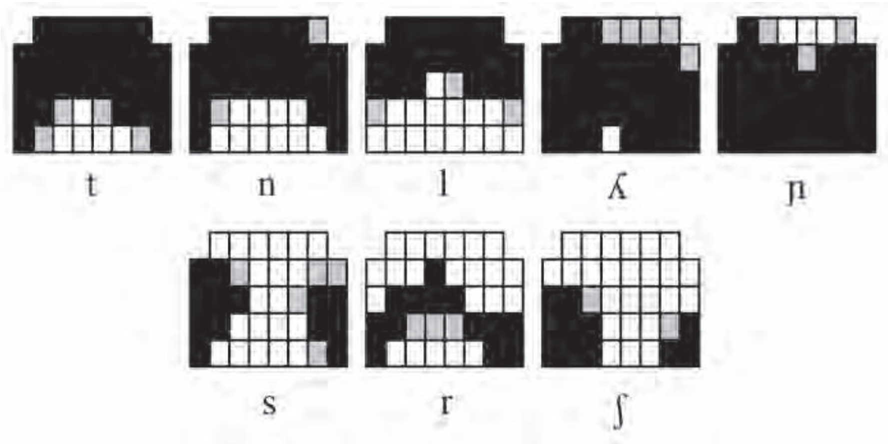


Figure 1. Linguopalatal contact patterns at consonant midpoint for unconstrained /t, n, l, ʎ, ɲ/ (top) and for constrained /s, r, ʃ/ (bottom) in symmetrical /aCa/ sequences (speaker DR). Contact is shown on five rows of electrodes proceeding from the front alveolar zone (at the top of the contact configurations) to the postalveolar and prepalatal zone (at the bottom of the contact patterns). Electrodes may appear in black (80–100% of activation across repetitions), grey (40–80% of activation across repetitions) and white (less than 40%).

The aerodynamic requirements for specific manners of articulation (e.g., trilling, frication) also contribute to an increase in degree of articulatory constraint. Crucially for the purpose of this study, manner of articulation requirements allow classifying Catalan dental, alveolar and alveolopalatal consonants as either highly constrained fricatives and rhotics (i.e., /s, r, ʃ/ where /r/ is a trill) and more unconstrained stops, nasals and laterals (i.e., /t, n, l, ʎ, ɲ/ where /t/ is dental, /l/ is dark and /ʎ, ɲ/ are alveolopalatal). Moreover, there appears to be a relationship between degree of articulatory constraint and place of articulation such that consonants of the former group are essentially more retracted than consonants of the latter, i.e., /s, r, ʃ/ cannot be articulated at the front alveolar zone while /t, n, l/ are typically front alveolar and /ʎ, ɲ/ may exhibit a central closure which includes the front alveolar zone as well. Those consonants may also differ regarding contact extension at the place of articulation, i.e., a larger contact area is available for apicolaminal

/t/ and laminopredorsal /ʃ, ʎ, ɲ/ than for more apical /n, l, s, r/. Figure 1 provides linguopalatal contact configurations for all seven Catalan consonants for illustration of the places of articulation and of the closure or constriction degrees just referred to.

C-to-C place adaptation processes in consonant clusters should conform to the principle that degree of articulatory constraint is inversely related to coarticulatory sensitivity (i.e., the extent to which a segment allows the coarticulatory influence of another segment) and directly related to coarticulatory aggression (i.e., the extent to which a segment exerts coarticulatory effects on another segment) (Bladon and Nolan 1977; Fowler and Saltzman 1993). Given the differences in degree of articulatory constraint and in place of articulation between the Catalan consonants of interest, a goal of this study is to test the following predictions on C-to-C adaptation in consonant clusters:

(a) Clusters with an unconstrained C1 (/t, n, l, ʎ, ɲ/) and a constrained C2 (/s, r, ʃ/) ought to show maximal C1-to-C2 adaptation and thus, undergo regressive place assimilation. More specifically, front C1 is expected to retract to the C2 place of articulation in these circumstances. Moreover, the assumption that assimilations are categorical processes occurring at the segmental level implies that they should be implemented throughout C1 regardless of contact degree, speaker and speech rate.

(b) Clusters exhibiting the reversed DAC composition, i.e., a constrained C1 (/s, r, ʃ/) and an unconstrained C2 (/t, n, l, ʎ, ɲ/), should be realized through two successive articulatory targets or else exhibit prominent coarticulatory effects from C1 onto C2 approaching progressive assimilation. The rationale for why regressive assimilations are more prone to occur than progressive assimilations (see *(a)*) follows from two findings: on the one hand, anticipatory effects tend to be more temporally fixed than carryover effects presumably since the former reflect planning and the latter are due to inertia (Whalen 1990); on the other hand, C-to-V coarticulatory effects associated with tongue dorsum lowering for highly constrained trills and lingual fricatives favor the anticipatory over the carryover component (Recasens 1999; Recasens 2002; Recasens, Pallarès and Fontdevila 1997).

(c) Clusters composed of unconstrained front consonants /t, n, l, ʎ, ɲ/ ought to exhibit blending, namely, a compromise articulatory configuration between C1 and C2. A goal of our research is to characterize the spatiotemporal implementation of blending, namely, whether blending begins at C1 onset or later on during C1, and whether it involves the formation of an intermediate location between the two adjacent consonants in the cluster or else the addition of their contact characteristics at the place of articulation. The

latter possibility appears to be consistent with Munhall and Löfqvist (1985) who found that the two separate glottal opening gestures for English /s##t/ were aggregated into a single smooth glottal opening at fast rates. Blending through closure addition ought to be facilitated by consonants produced with an extended closure (e.g., /t/ and alveolopalatals).

An important issue of our research is whether a C1-to-C2 adaptation lasting throughout the entire C1 period should be associated with an assimilatory process or reflects a long coarticulatory effect extending back to C1 onset. In principle, data on the temporal extent of C-to-V effects in VCV sequences could serve as reference for estimating the temporal span of coarticulation in consonant clusters. The fact is, however, that both measures are not strictly comparable since vowels do not oppose coarticulatory effects exerted by consonants to the extent that other consonants do. Consequently, we take an assimilation to occur when C1 adapts completely to the C2-dependent articulatory target already at its onset, and this adaptation process applies to comparable CC sequences and to other speakers.

In addition to testing the validity of a model of C-to-C adaptation based on articulatory constraints, the present paper will try to show the extent to which the model can contribute to the interpretation of specific sound changes and phonological issues (see section 4).

2. Method

2.1. Recording and analysis procedure

Electropalatographic (EPG) and acoustic data were recorded for the crucial CC combinations with the consonants /t, n, l, s, r, ʃ, ʎ, ɲ/ in /aC##Ca/ sequences with two meaningful words. Sentences were uttered at a confortable rate by the Catalan speakers DR, JP and JS 5, 3 and 3 times, respectively. A small and variable number of repetitions is due to time limitations at the University of Reading where those recordings were made back in 1992. Linguopalatal contact and acoustic data were gathered every 5 ms with 62 electrode artificial palates using electropalatography (Reading EPG system). The acoustic signal was recorded at a 20 kHz sampling rate and processed with a Kay CSL analysis system using the same temporal resolution as the EPG data.

It may be claimed that the absence of speech rate manipulation in the present study renders impossible to ascertain whether the C-to-C effects were gradient (coarticulatory) rather than categorical (assimilatory). The rationale underlying this objection is that overlap between the two consecutive consonants would gradually diminish at slower rates and increase at faster rates if the C-to-C effects were gradient, while staying unaffected if those effects were categorical. As discussed in the *Results* section, we feel that there was enough spatiotemporal overlap in the CC sequences under study that we can be pretty confident that discrete or categorical changes were available in most crucial cases.

In order to match the other voiced consonant clusters, consonant sequences starting with /t/ and /s/ were rendered voiced by placing C1 before a voiced consonant (e.g., [dz] was recorded instead of [ts]). Moreover, according to other phonotactic constraints of Catalan, the palatal fricative had to be underlyingly voiceless when acting as C1 and underlyingly voiced when acting as C2. In spite of these voicing differences, stops and fricatives will always be represented with phonetic symbols for the voiceless correlates in the present paper. No data will be reported for a few clusters which underwent full assimilation rather than other expected processes (i.e., /tn, tl, tʎ/ for all speakers, and /nl, lʎ/ for speaker JP) or exhibited an epenthetic stop between the two consecutive consonants (i.e., /ʎs/ for speaker DR).

Linguopalatal contact patterns for the same consonants were also recorded and analyzed at consonant midpoint in symmetrical /aCa/ sequences, and will be used for evaluation of assimilation, blending and C1-to-C2 sliding in CC sequences (see Method section). Intervocalic consonants had to be voiceless since voiced stops become approximant in this context in Catalan. The same vowel (i.e., /a/) was used in VCV sequences as in CC sequences in order to avoid different V-to-C contextual effects in both sequence types.

EPG contact patterns for the intervocalic consonants were averaged across repetitions at PMC (point of maximum constriction), i.e., the temporal frame exhibiting the maximum number of on-electrodes over the entire palate surface. Those for clusters were averaged across repetitions at four temporal points, i.e., at three points during C1 (at C1 onset, midpoint and offset) and at one point during C2 (i.e., at 15 ms after C1 offset). Segmentation procedures have been described in Recasens and Pallarès (2001); they were based on articulatory events (i.e., presence of a complete closure for stops, of a narrow constriction for fricatives and approximants, and of a closure, constriction or central contact for laterals and rhotics) or, in the absence of reliable articulatory evidence, on events available in the acoustic waveform and spectrographic displays.

As shown in Figure 1, linguopalatal contact data were computed from mean palatographic configurations for the five front rows of electrodes, and proceeds from the front alveolar zone (at rows 1 and 2 at the top of each EPG pattern) to the back alveolar zone (at rows 3 and 4) and prepalate (at row 5). The number of electrodes amounts to 6 on the first row and to 8 on rows 2–5.

2.2. Determining place of articulation, assimilation and blending

In order to determine the existence of assimilation in clusters with consonants differing in degree of constraint, we had to come up with an objective method for estimating the place of consonant articulation. This was so since, as pointed out in the Introduction, assimilation is taken to occur whenever the place of articulation for target C1 or C2 coincides with that for the triggering contextual consonant. The evaluation of blending in clusters with two unconstrained consonants was not based, however, on place of articulation but on closure or constriction degree (see section 3.3).

It is not always obvious where closure or constriction location occurs in linguopalatal contact patterns such as those in Figure 1. This applies not only to lingual fricatives but to other consonants as well. Thus, should row 2 be considered part of the closure for /ɲ/ or row 4 assigned to the place of articulation for /r/ given that not all electrodes on those rows are activated more than 80%?. Also, does the intended place of articulation for /t/ occur at rows 1 through 3 or at a more restricted location?

In the present study, place of articulation was identified with the row(s) showing the maximum degree of electrode activation of all rows over the surface of the artificial palate. Percentages of electrode activation were computed for each row of electrodes, and differences between contact averages for pairs of adjacent rows were submitted to paired t-tests. Place of articulation was taken to occur at the row(s) yielding the significantly highest contact average. The level of significance was set at 0.15 in view of the small number of repetitions for each cluster.

This procedure was applied to the linguopalatal contact configurations at consonant midpoint in all VCV sequences and at the four temporal points subject to measurement in all CC clusters, and will be exemplified for intervocalic /t, s, r, ʃ/ and for the clusters /ts, tr, tʃ/. Figure 2 allows tracking changes in linguopalatal contact during C1 (/t/) as a function of C2 (/s, r, ʃ/) for speaker DR. It displays mean EPG contact configurations across repetitions of the consonants /t, s, r, ʃ/ in symmetrical /aCa/ sequences (to the left of the vertical line), and of the clusters /ts, tr, tʃ/ at C1 onset, midpoint and offset and at C2 onset (to the right of the line). Figure 3 displays row contact averages for the EPG configurations in Figure 2, i.e., at PMC for intervocalic /t, s, r, ʃ/ (two left columns), and at C1 onset, midpoint and offset and at C2 onset for /ts/, /tr/ and /tʃ/ (right column).

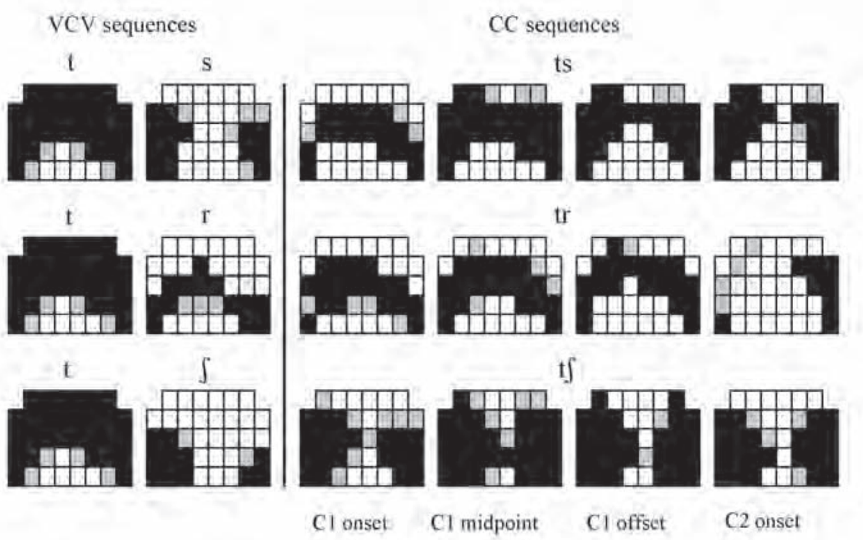


Figure 2. Linguopalatal contact patterns at consonant midpoint for /t, s, r, ʃ/ in symmetrical /aCa/ sequences (left), and at C1 onset, midpoint and offset and at C2 onset for /ts, tr, tʃ/ (right) (speaker DR). See Figure 1 for details.

Maximum electrode activation for /t, s, r, ʃ/ in intervocalic position was found to occur at row 3 for /s/, at row 4 for /r/, and at a region exhibiting a contact plateau for /t/ (i.e., at rows 1–3) and for /ʃ/ (i.e., at rows 4–5). These statistical results are consistent with row contact averages in Figure 3. Whenever contact maximum extended over two or three consecutive rows, place of articulation was fixed at the mean across rows. Thus, e.g., the place of articulation value for intervocalic /ʃ/ was 4.5 since statistical analyses yielded a contact maximum at rows 4 and 5. In cases where closure location extended over four or five rows, place of articulation was taken to occur at the row showing a contact maximum from visual inspection of the contact percentage displays. The stop /t/ was assigned a 0 value since this consonant is dental in Catalan irrespective of whether it involves central alveolar contact or not.

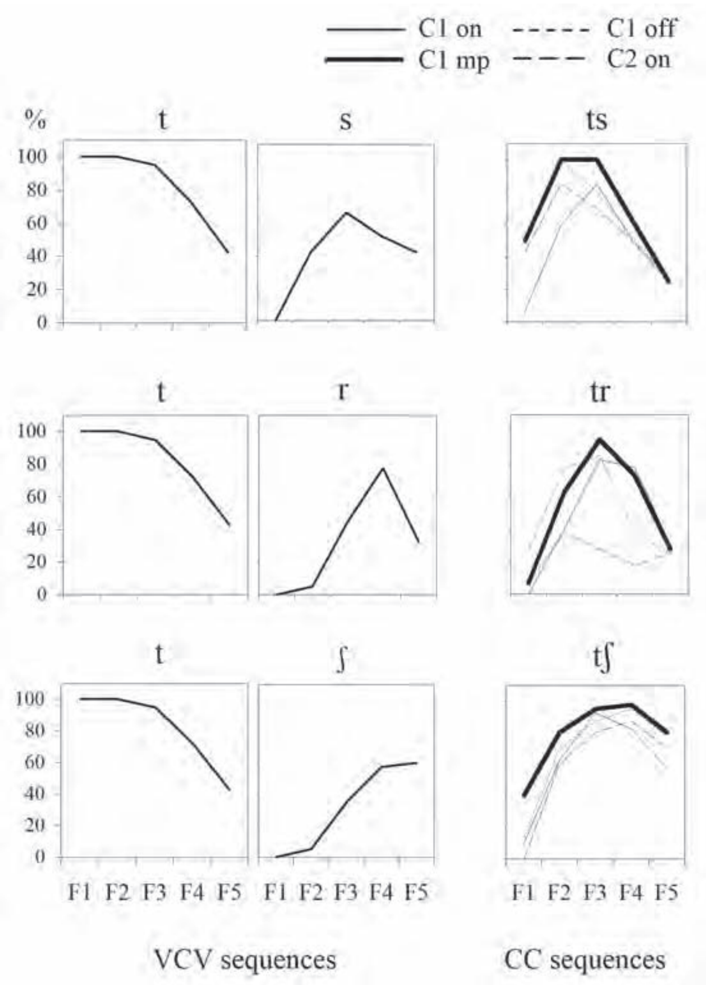


Figure 3. Percentages of electrode activation at each of the five frontmost rows of the artificial palate (F1 through F5). Data are presented for intervocalic /t, s, r, ʃ/ (left) and the clusters /ts, tr, tʃ/ (right). Data have been sampled at closure or constriction midpoint for the intervocalic sequences, and at four points in time (C1 onset, C1 midpoint, C1 offset and C2 onset) for consonants in clusters.

A decision about whether C-to-C adaptation resulted in assimilation or not was based on a joint consideration of the place of articulation for the two consecutive consonants in intervocalic position. Thus, according to Figure 3, place of articulation values at C1 onset (3), C1 midpoint (2.5) and C1 offset (2) for /t/ in the cluster /ts/ differ from the place of articulation value for intervocalic /t/ (0) while equalling or approaching that for intervocalic /s/ (3). Based on the similarity between the places of articulation for intervocalic /s/ and during the stop closure period of /ts/, it looks as if regressive assimilation could be available in this case.

3. Results

3.1. Unconstrained C1 + constrained C2

Place of articulation values for combinations of an unconstrained C1 (/t, n, l, ʎ, ɲ/) and a constrained C2 (/s, r, ʃ/) in Table 1 allow studying cases of regressive assimilation. From left to right, data in the columns correspond to C1 in intervocalic position ((V)C1(V)), to C1 and C2 in clusters (at C1 onset, midpoint and offset, and at C2 onset) and to C2 in intervocalic position ((V)C2(V)). As pointed out in the Method section, place of articulation values for the target and contextual consonants in intervocalic position were used as reference for determining whether the former assimilates to the latter in clusters. Clusters for speaker JS could only be subject to analysis if including dental /t/; this was so since non dental consonants for this speaker were found to exhibit roughly the same postalveolar or prepalatal closure or constriction location in clusters as in intervocalic position, i.e., at about rows 3–4 for /n, l/ and about rows 4–5 for /ʎ, ɲ, s, r, ʃ/.

Let us, first of all, pay attention to data for speaker DR at the top of the table. Phonetic symbols on the leftmost column indicate the presence of large differences in place of articulation between unconstrained C1=/t, n, l, ʎ, ɲ/ in intervocalic position varying in the progression /t/ (0) > /l/ (1.5) > /n/ (2.5) > /ʎ/ (3) > /ɲ/ (4). On the other hand, place of articulation values for constrained C2 in intervocalic position on the rightmost column amount to 3 for /s/, 4 for /r/ and 4.5 for /ʃ/. A look at the place of articulation values for C2= /s, r, ʃ/ in intervocalic position and at C2 onset shows that the former cannot be used alone for determining whether C1 assimilates to C2 or not. This happens to be so since the two values are not strictly the same but less front in intervocalic position than in clusters, i.e., 3 for intervocalic /s/ and 2–2.5 for

/s/ at C2 onset in the clusters /ts, ns, ls, ɲs/, 4 for intervocalic /r/ and 2.5–3 for /r/ at C2 onset, and 4.5 for intervocalic /ʃ/ and 3.5–4.5 for /ʃ/ at C2 onset. In order to cope with these differences, the characteristic place of articulation for a given consonant was considered to encompass the intervocalic and CC values, i.e., 2–3 for /s/ (which includes 3 in intervocalic position and 2–2.5 at C2 onset), 2.5–4 for /r/ (which includes 4 in intervocalic position and 2.5–3 at C2 onset), and 3.5–4.5 for /ʃ/ (which includes 4.5 in intervocalic position and 3.5–4.5 at C2 onset).

Place of articulation values in the four central columns reveal small changes during the C1 period for all clusters under analysis. Moreover, in comparison to the place of articulation values in intervocalic position, those in clusters exhibit a much more restricted range and approach the place of articulation for C2. Thus, with a few exceptions, values for all C1 oscillate between 2 and 3 before /s/ (which is precisely the characteristic range for /s/), between 2.5 and 4 before /r/ (which is exactly the characteristic range for /r/), and between 3.5 and 4 before /ʃ/ (which is close to the characteristic range for /ʃ/). These data are consistent with the notion that regressive assimilation is available in clusters with an unconstrained C1 followed by a constrained C2. The table also shows that the clusters /lr/, /ʎr/ and /lʃ/ are exceptional in that the place of articulation for the lateral occurs at a more front location than expected in line with the manner requirements involved.

According to data for speaker JP (Table 1, middle), unconstrained /n, l, ʎ, ɲ/ show place of articulation values ranging between 3 and 4.5 in intervocalic position, i.e., 3 for /l/, 3.5 for /n/, 4 for /ʎ/ and 4.5 for /ɲ/. Analogously to the data for speaker DR, values for /s, r, ʃ/ are generally less front in intervocalic position than at C2 onset in clusters, i.e., 3 (VCV) and 2–3 (CC) for /s/, 4 (VCV) and 3.5–4.5 (CC) for /r/, and 5 (VCV) and 3.5–4.5 (CC) for /ʃ/. Characteristic ranges encompassing both sets of values are 2–3 for /s/, 3.5–4.5 for /r/, and 3.5–5 for /ʃ/.

The place of articulation values for C1 in clusters resemble those for speaker DR in remaining highly stable and approaching those for C2. Thus, except for a few clusters with a lateral C1 (/ls, lʃ/), they range generally between 3 and 3.5 before /s/ (which is slightly higher than the characteristic range for /s/), between 3.5 and 4.5 before /r/ (which is precisely the prototypical range for /r/), and between 3.5 and 4.5 before /ʃ/ (which is comparable to the characteristic range for /ʃ/). A straightforward C1-to-C2 adaptation occurs for /t/ which changes from dental in intervocalic position to postalveolar before /s, r, ʃ/ (see also data for speaker JS at the bottom of Table 1). Again, consonant clusters with a lateral C1 may be produced with a more anterior closure than expected.

Table 1. Place of articulation values for clusters composed of an unconstrained C1 and a constrained C2 for speakers DR, JP and JS. Data correspond to intervocalic C1 and C2 (leftmost and rightmost columns), and to C1 onset, C1 midpoint, C1 offset and C2 onset in clusters (central columns).

		(V)C1(V)	C1on	C1mp	C1off	C2on	(V)C2(V)
Speaker DR	t s	0	3	2.5	2	2	3
	n s	2.5	3	3	3	2	3
	l s	1.5	3	2.5	3	2.5	3
	ɲ s	4	3	3	3	2.5	3
	t r	0	3.5	3	2.5	2.5	4
	n r	2.5	4	4	4	3	4
	l r	1.5	2	3	3	3	4
	ʎ r	3	1.5	1.5	2	2.5	4
	ɲ r	4	3	3	3	2.5	4
	t ʃ	0	3.5	3.5	4	4	4.5
	n ʃ	2.5	3.5	3.5	4	4	4.5
	l ʃ	1.5	2	3	3.5	3.5	4.5
	ʎ ʃ	3	3.5	3.5	3.5	3.5	4.5
	ɲ ʃ	4	3.5	4	3.5	4.5	4.5
Speaker JP	t s	0	3	3	3	3	3
	n s	3.5	3.5	3.5	3	3	3
	l s	3	1	2	2.5	2.5	3
	ʎ s	4	4	3	3	2	3
	ɲ s	4.5	4	3.5	3.5	3	3
	t r	0	4.5	4.5	4	3.5	4
	n r	3.5	4.5	4.5	4.5	4.5	4
	l r	3	3.5	4	4.5	4	4
	ʎ r	4	4	4	3.5	3.5	4
	ɲ r	4.5	4	3.5	4	3.5	4
	t ʃ	0	4	3.5	4	4.5	5
	n ʃ	3.5	3.5	4	4	4	5
	l ʃ	3	1.5	1.5	3.5	3.5	5
	ʎ ʃ	4	4.5	4	4	3.5	5
	ɲ ʃ	4.5	4.5	4.5	4.5	4.5	5
Speaker JS	t s	0	3.5	3.5	3.5	4.5	3
	t r	0	3.5	3.5	3.5	3	4
	t ʃ	0	4.5	4.5	4.5	4.5	5

3.2. Constrained C1 and unconstrained C2

Place of articulation values for clusters composed of a constrained C1 and an unconstrained C2 for studying the C1-to-C2 effects are given in Table 2. From left to right, they correspond to C1 in intervocalic position, to CC sequences (at C1 onset, C1 midpoint, C1 offset and C2 onset), and to C2 in intervocalic position.

Table 2. Place of articulation values for clusters composed of a constrained C1 and an unconstrained C2 for speakers DR, JP and JS. Data correspond to intervocalic C1 and C2 (leftmost and rightmost columns), and to C1 onset, C1 midpoint, C1 offset and C2 onset in clusters (central columns).

		(V)C1(V)	C1on	C1mp	C1off	C2on	(V)C2(V)
Speaker DR	s t	3	4	2.5	2	2	0
	s n	3	4.5	3	2.5	2.5	2.5
	s l	3	4	2.5	2	2	1.5
	s ʎ	3	4	3	2	2	3
	r t	4	4.5	3.5	3	1.5	0
	r n	4	4.5	4	4	3.5	2.5
	r l	4	4.5	4	3.5	3	1.5
	r ʎ	4	4.5	3.5	3.5	1.5	3
	ʃ t	4.5	5	4.5	3.5	3	0
	ʃ n	4.5	4.5	3.5	3.5	2.5	2.5
	ʃ l	4.5	4.5	3	2.5	1.5	1.5
	ʃ ʎ	4.5	4.5	3.5	2.5	3	3
Speaker JP	s t	3	4	2	1.5	1	0
	s n	3	4	3.5	3	3	3.5
	s l	3	4	3	2	1	3
	s ʎ	3	4	3	3	3	4
	r t	4	4	4	3.5	0	0
	r n	4	4	4	4	4	3.5
	r l	4	4	4.5	4	2.5	3
	r ʎ	4	4.5	4	3.5	3	4
	ʃ t	5	5	4	3	3	0
	ʃ n	5	4.5	4	3.5	3	3.5
	ʃ l	5	4.5	3.5	2.5	1.5	3
	ʃ ʎ	5	4.5	4	3.5	3.5	4
Speaker JS	s t	3	5	5	4.5	4	0
	r t	4	5	4.5	4.5	2	0
	ʃ t	5	5	5	4.5	4	0

Data for speaker DR generally show a gradual closure or constriction fronting from C1 onset to C2 onset rather than a fixed place of articulation. It is thus clear that the production of these clusters involves two consonant targets. The C1 target is apparent at C1 onset where the consonant exhibits a similar place of articulation to that occurring in intervocalic position (more so for /r/ and /ʃ/ than for /s/). The C2 target, on the other hand, is already available at C1 offset in clusters with C1=/s/ and to a large extent in those with C1=/ʃ/, but not in clusters with C1=/r/ where changes in place of articulation still occur during the transition from C1 offset to C2 onset.

Those data also reveal that place of articulation values at C2 onset may approach those for C1. Thus, C2= /t/ changes from dental in intervocalic position to alveolar after a lingual fricative, i.e., 2 for /st/ and 3 for /ʃt/ which lie close to the characteristic values for /s/ (2–3) and /ʃ/ (3.5–4.5) given in section 3.1. The dental place of articulation is approached however after /r/ in the cluster /rt/. As a general rule, the remaining clusters may exhibit carryover place effects (except for /rʎ, ʃl/ where the lateral is more front than expected) and even possible instances of progressive place assimilation (/rn/). In summary, CC sequences with a constrained C1 and an unconstrained C2 may undergo different place adaptation processes, i.e., carryover coarticulation, progressive assimilation or no C2-to-C1 adaptation at all.

The scenario for speaker JP is similar to that for speaker DR with the exception of some clusters with C1=/s/, and progressive assimilation appears to be available for /st, ʃt/ in the case of speaker JS.

3.3. Unconstrained C1 and C2

Results for clusters with two unconstrained consonants (i.e., /t, n, l, ʎ, ɲ/) are presented in Table 3. In the table, the three columns on the left report the row(s) of electrodes exhibiting maximal activation at closure or constriction location for C1 and C2 in intervocalic position, and the predicted outcome in case that blending involves the addition of the closure areas for the two consonants in the cluster. Thus, e.g., it is predicted that the closure area for /nt/ for speaker DR should encompass rows 1, 2 and 3 since intervocalic /n/ and /t/ for this speaker are articulated at rows 2–3 and 1–3, respectively. The four right columns provide the closure area at each of the four points in time subject to measurement, i.e., at C1 onset, midpoint and offset and at C2 onset.

Table 3. Closure areas for clusters composed of unconstrained C1 and C2 for speakers DR, JP and JS. (Left columns) Data for intervocalic C1 and C2 and for predicted blending by closure addition. (Right columns) Data at C1 onset, C1 midpoint, C1 offset and C2 onset in clusters. Instances of blending are represented in grey and contact areas approaching blending in boldface.

		(V)C1(V)	(V)C2(V)	Predicted	C1on	C1mp	C1off	C2on
Speaker DR	n t	2--3	1--3	1--3	1--2	1--3	1--4	1--4
	n l	2--3	1--2	1--3	2--3	3	2--3	1--3
	n ʎ	2--3	2--4	2--4	**1--3**	1--4	1--4	1--5
	l t	1--2	1--3	1--3	1	1--3	1--3	1--3
	l n	1--2	2--3	1--3	1	1	1--2	1--2
	l ʎ	1--2	2--4	1--4	1	1	1--4	1--4
	ʎ t	2--4	1--3	1--4	2--3	1--4	1--4	1--4
	ʎ n	2--4	2--3	2--4	**3--5**	3--4	1--4	3--4
	ʎ l	2--4	1--2	1--4	**1--3**	1--2	1--2	1--2
	ɲ t	3--5	1--3	1--5	3	1--5	1--5	1--5
	ɲ n	3--5	2--3	2--5	2--3	**1--4**	**1--4**	1--3
	ɲ l	3--5	1--2	1--5	**1--4**	1--3	1--2	1--2
Speaker JP	n t	3--4	1	1--4	**1--3**	1--2	1--2	1--2
	n ʎ	3--4	4	3--4	3--5	2--5	2--5	1--5
	l t	3	1	1--3	1	2	1--3	1--3
	l n	3	3--4	3--4	1	1--2	1--2	1--2
	ʎ t	4	1	1--4	4--5	**3--5**	1--5	1--4
	ʎ n	4	3--4	3--4	4--5	1--5	1--5	1--5
	ʎ l	4	3	3--4	4	1--4	**1--3**	**1--3**
	ɲ t	4--5	1	1--5	3--4	1--5	1--5	1--5
	ɲ n	4--5	3--4	3--5	4--5	1--5	1--5	**1--4**
	ɲ l	4--5	3	3--5	3--4	**1--4**	1--2	1--2
Speaker JS	n t	4--5	1--4	1--5	3	**1--4**	**1--4**	**1--4**
	n l	4--5	3--4	3--5	4	3--4	**1--4**	**1--4**
	n ʎ	4--5	5	4--5	3--4	1--5	1--5	1--5
	l t	3--4	1--4	1--4	1--5	1--4	1--4	1--4
	l n	3--4	4--5	3--5	2	**1--3**	**1--3**	**1--3**
	l ʎ	3--4	5	3--5	**1--3**	**1--4**	1--5	1--5
	ʎ t	5	1--4	1--5	5	5	1--5	1--5
	ʎ n	5	4--5	4--5	5	4--5	2--5	1--5
	ʎ l	5	3--4	3--5	1--5	1--5	**1--4**	**2--4**
	ɲ t	4--5	1--4	1--5	5	1--5	1--5	1--5
	ɲ n	4--5	4--5	4--5	4--5	1--5	2--5	2--5
	ɲ l	4--5	3--4	3--5	5	1--5	1--5	**2--4**

Observed closure areas which may be accounted for through blending are represented in grey whether they coincide with the predicted ones or exceed them by one or more rows. The table also highlights in boldface less robust instances of blending resulting into closure areas extending over 3 or 4 rows, and including partially the predicted closure area or being located immediately adjacent to it.

Inspection of the cells filled in grey reveals that most consonant sequences under analysis undergo blending involving the summation of the closure areas for C1 and C2. Moreover, blending usually begins after C1 onset (as suggested by the presence of a small number of grey cells at this temporal point), and goes on all the way into C2 (as shown by the presence of a large number of grey cells at C2 onset). Inspection of the linguopalatal configurations after C2 onset confirms that blending usually affects the entire C2 period.

A closer look at the individual CC sequences reveals that blending is favoured in clusters composed of an alveolopalatal consonant and a dental or an alveolar consonant produced with some tongue dorsum raising (i.e., /nʎ, ʎt, ʎn, ɲt, ɲn/). Data in the table show indeed a large closure area extending over rows 1–4, 1–5 or 2–5 for these clusters, which amounts to the addition of a large closure area for intervocalic /ʎ, ɲ/ and a smaller and more front closure area for intervocalic /t, n/. This additive outcome is illustrated in Figure 4 for /ɲt/ and /nʎ/ (speaker DR). Indeed, linguopalatal configurations at C1 midpoint and successive frames show a large closure extending over rows 1 through 4 or 5 which results from adding up the closure areas for C1 and C2 in intervocalic position.

The articulatory implementation of the sequences /nt, lt/ has been associated traditionally with an assimilatory process by which the alveolar consonant becomes dental before dental C2. Data in Table 3 suggest however that blending rather than assimilation occurs in this case. This is so since central contact may increase gradually from C1 onset to C1 midpoint such that the closure area during the second half of C1 approaches the predicted outcome for a blending scenario. Data for /nt/ in Figure 4 (speaker DR) are in agreement with this general picture.

The presence of C1=/l/ may delay the implementation of blending (e.g., see data for /lʎ/ for speakers DR and JS in the table) or prevents blending from applying (e.g., see data for /ln/ for speakers DR and JP). Indeed, linguopalatal configurations for /lʎ/ for speaker DR in Figure 4 show a progressive increase in closure extent over rows 2, 3 and 4 as we proceed from C1 onset to C2 onset. On the other hand, the presence of C2=/l/ may cause some postalveolar contact evacuation during the preceding consonant and thus, a premature termination of blending (see data for /ʎl, ɲl/ in the table). Linguopalatal contact configurations for /ɲl/ in Figure 4 show indeed a decrease in central contact at rows 3 and 4 as C2 is approached. This behavior is consistent with the requirement for laterals to exhibit a front realization so as to allow airflow through the sides of the mouth.

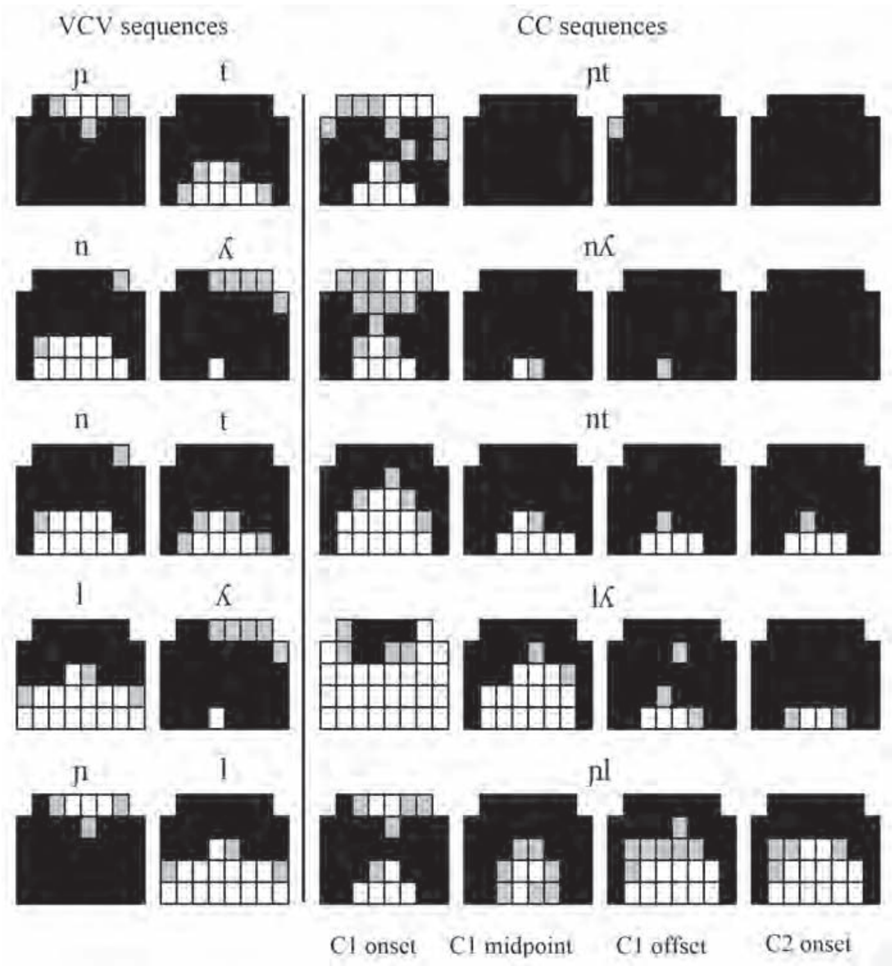


Figure 4. Linguopalatal contact configurations in symmetrical /aCa/ sequences (left) and at C1 onset, midpoint and offset and at C2 onset (right) for the clusters /ɲt, nʎ, nt, lʎ, ɲl/ (speaker DR). See Figure 1 for details.

4. Summary and discussion

Results presented in the preceding section show that a model of articulatory constraints, i.e., the DAC model, may account to a large extent for assimilatory and blending processes in Catalan clusters composed of dentals, alveolars and alveolopalatals. Our model differs from other theoretical accounts in allowing for complete assimilations (vis-à-vis Articulatory Phonology) as well as for assimilations occurring at different locations within the same articulatory zone (vis-à-vis articulatory phonetics and generative phonology). According to generative phonology (Chomsky and Halle 1968), dental and alveolar consonants differ from palatoalveolars in that the former are [+ant], [–high] and the latter [–ant], [+high]; within our framework, however, dentals, alveolars and palatoalveolars or alveolopalatals may be front or back depending on their production requirements. Assimilations may be considered the phonological consequence of prominent coarticulatory effects; this view is consistent with the finding that regressive assimilations occur more systematically than progressive ones while the temporal extent of anticipatory coarticulation is also more fixed than that of carryover effects. Evidence has been presented for blending through contact addition starting after C1 onset rather than through the formation of an intermediate articulation. Laterality requirements cause /l/ to exhibit a very front closure and to avoid assimilation and blending which is in support of the alveolar lateral being subject to higher manner demands than stops and nasals.

EPG data are highly consistent with the formulation of hypotheses *(a)*, *(b)* and *(c)* in the Introduction. C-to-C adaptation mechanisms in place of articulation will be now summarized, and their implication for the interpretation of specific sound changes and phonological issues will be underlined.

(a) As a general rule, regressive assimilation in clusters with unconstrained front /t, n, ɲ/ followed by highly constrained back /s, r, ʃ/ involves closure retraction at different alveolar locations throughout C1, e.g., /n/ is more front before /s/ than before /r/, and before /r/ than before /ʃ/, in line with small but consistent differences in place of articulation between the two lingual fricatives and the rhotic. Laterality requirements explain why clusters with unconstrained C1=/l, ʎ/ may undergo changes in place of articulation rather than regressive assimilation.

A shift in place of articulation towards the back alveolar zone affects the tautosyllabic cluster /tr/ in Chilean Spanish and in northern areas of Peninsular Spanish (Alonso 1967; Malmberg 1965), as well as in English (Jones 1956).

Retroflex consonants behave like the back alveolars of our study in attracting a preceding dentoalveolar to their back place of articulation. This behavior conforms to coarticulation data showing that these consonants exert prominent anticipatory effects, e.g., an early F2 and F3 lowering for American English retroflex /r/ during a preceding front vowel (Boyce and Espy-Wilson 1997; Lehiste 1964; Monnot and Freeman 1972; Stevens 2000). Therefore, retroflexes could be specified for a high degree of articulatory constraint independently of their manner of articulation. Indeed, /tɽ, dɽ/ are realized [tɽ, dɽ] in S. Italian dialects ([tɽi] "three", [ˈtɽumma] "trumpet") (Millardet 1925: 720; Millardet 1933: 358; Rohlfs 1966: 264), and /ds̢, ts̢, ns̢/ become [ɖs̢, ʈs̢, ɳs̢] in Norwegian (*budsjett* "budget", *matskje* "spoon for food", *kanskje* "maybe"). Norwegian exemplifies the special status of laterals since /s/ becomes retroflex before /ɭ/ while /l/ stays non retroflex before /s̢/ in this language ([s̢ɭ] in *slå* "beat", [ls̢] in *stålski* "steel ski") (Simonsen, personal communication).

Instances of C1 deletion or regressive manner assimilation in clusters with an unconstrained C1 and a constrained C2 may also be attributed to regressive place assimilation. The assumption here is that place assimilation is a precondition for elision or manner assimilation to occur and thus, that these processes apply provided that C1 and C2 are homorganic. Indeed, a shift in place of articulation towards the back alveolar zone could account for /nr, lr/> [r] in phonetic variants taken from dialectal Italian ([oˈrare] *on(o)rare* "to honor", [ˈtore] *togliere* from TOLL(E)RE "to remove"; Rohlfs, 1966) and Old Picard (<u>*terront*</u> *tiendront* from TEN(E)RE "(they) will have", <u>*mourre*</u> from MOL(E)RE "to grind") (Gossen 1970).

(b) Clusters with a highly constrained back C1 and an unconstrained front C2 may exhibit contact fronting proceeding from C1 to C2, carryover coarticulatory effects at the C2 place of articulation, or C1-to-C2 place assimilation. A factor causing C2 to reach or fall short of its target is the flexibility of the primary articulator; thus, it appears that a dental C2 is articulated at the teeth after /r/ but not after /s, ʃ/ since the syllable final rhotic involves a short closure which leaves the tongue tip enough time to travel from the alveolar zone to the teeth.

Progressive postalveolarization of dentals and front alveolars as a function of /r/ may be exemplified with /rt/ in Spanish from Paraguay (Malmberg 1971: 436), /rt, rd, rn, rl, rs/ in Chilean Spanish (Alonso 1967: 228; Eliasson 1986: 293) and /rt, rd, rn, rs/ in Norwegian and Swedish where postalveolar /r/ is lost after triggering assimilation. Again, clusters with laterals appear to show a special behavior; thus, in Swedish, the sequence /rl/ may be realized [rl], [ɭ] and even [l] in a few words or expressions (Eliasson 1986: 279).

Place substitutions may also be induced by a retroflex C1 on an unconstrained C2. Thus, dentals become retroflex in the Sicilian clusters /ɽt, ɳd/ ([ˈpaɽʈi] "part", [ˈuɳɳi] UNDE "where") (Millardet 1925: 732, 738) and after /ɭ/ in Andalusian Spanish ([ˈgoɭɖo] "fat", [ˈaɭʈo] "tall") (Llorente-Maldonado 1962: 239). Also, in Swedish and Norwegian, progressive place assimilation is triggered by consonants rendered retroflex after /r/, e.g., [ɳːʂ] (Norw. [æɳːʂʈ] *Ernst*), [ʂʈ] (Sw. [vɑʂʈo] *var står* "where are"), [ʈɳ] (Norw. *partner*), [ʈʂ] (Norw. *fortsette* "continue"), [ʂɳ] (Norw. *korsnebb* "cross beak"), [ɖɽ] (Sw. [fu:ɖɽa] *fordra* "demand") (Eliasson 1986; Simonsen, personal communication). In support of the special status of laterals, progressive retroflexion does not occur after /ɭ/ in Swedish ([pæ:ɭˈtro:d] *pärltråd* "string of pearls") (Eliasson 1986: 280).

Progressive assimilatory retroflexion is also frequent in Asian and Australian languages where it applies to clusters such as /ʈ,ɖ/ + /t, l/ (Tamil [ʈʈ], Tulu [ʈɭ, ɖɭ]), /ɭ/ + /t, d, n, l/ (E. Arrernte [ɭʈ], Kuvi [ɭɖ], Murinbata [ɭɳ], Kannada [ɭɭ]), /ɳ/ + /t, d, n, l/ (Malayalam [ɳʈ], Tulu [ɳɖ], Urali [ɳɳ], Tulu [ɳɭ]) and /ʂ + t, n/ (Sanskrit [ʂʈ, ʂɳ]) (Steriade 2001: 245–6). In Steriade's view these place assimilations are progressive rather than regressive for perceptual reasons: the fact that the tongue tip for a retroflex slides forward during the closure period renders the VC transitions more perceptually effective than the CV transitions in signalling the non retroflex-retroflex distinction (e.g., F2 and F3 at the VC boundary are about 1900 Hz and 2750 Hz for /ʈ/ and 2700 Hz and 3500 Hz for /t/). There are however conflicting data which may be accounted for within the DAC framework. In the first place, Steriade's hypothesis predicts that progressive place assimilation should apply not only to non retroflex dentoalveolar apicals but also to retroflex ones. A look at the Asian and Australian data reveals however that instances of progressive assimilation in retroflex + non retroflex clusters outnumber very clearly those taking place in non retroflex + retroflex sequences, i.e., retroflexes are prone to assimilate non retroflexes rather than the other way around. Secondly, manner of articulation requirements appear to play a role since highly constrained fricatives and rhotics may not assimilate to more unconstrained stops, nasals and laterals, e.g., stops and nasals acquire the retroflexion feature from /ʂ/ (e.g., /ʂt/> [ʂʈ]) while /s/ only assimilates to preceding /ʂ/ and /ɽ/ (/ʂs/> [ʂʂ], /ɽs/> [ɽʂ], /ʈs/> [ʈʂ]). Finally, the trend for nasals to undergo regressive and progressive place assimilation in front consonant clusters (e.g., /ʂn/> [ʂɳ], /ɳt/ > [nt]) could be associated with the fact that these consonants are specificed for a low degree of articulatory constraint rather than with the low perceptibility of nasal murmurs. Indeed, in contrast with stops, nasals involve a lower intraoral pressure level and may be realized through an incomplete closure.

Progressive place assimilation in clusters with a constrained C1 and an unconstrained C2 is presumably at the origin of regressive and progressive changes in manner of articulation. This could be so for the replacement of C2 by C1 in sequences such as /rn, rl/ in Sardinian ([ˈfuru] FURNU, [ˈfewra] FERULA) (Contini 1987: 398), and /ɽl/ in Sicilian ([ˈpaɽɽu] *parlu* "I speak", [ˈkaɽɽu] *Carlo* "Charles") (Millardet 1925: 737–8). It may also account for the replacement of C1 by C2 in analogous clusters, i.e., /rn, rt, rl/ > [nn, tt, ll] (Sicilian [ˈfunnu] *forno* "oven", [ˈmɛlla] *merla* "blackbird", Spanish from Panamá [ˈpwetta] *puerta* "door") (Alonso 1967: 234; Rohlfs 1966: 339), /ɖl/> [ɭɭ] and /ɖn/> [ɳɳ] in Telegu, and /ɭt/> [ʈʈ] in Malayalam (Steriade 2001: 245).

C1 elision may also apply after place assimilation. This is so in Norwegian and Swedish where postalveolar /r/ is lost after triggering assimilation in the sequences /rt, rd, rn, rs/ (e.g., Swedish /fort/> [fɔʈː] "fort", /fors/ > [fɔʂː] "rapids") (Eliasson 1986: 278). Also, in Swedish, C1 may drop in the cluster /rl/ after exerting place assimilation on the following lateral thus yielding [ɭ] (see above).

(c) As a general rule, clusters composed of two unconstrained front consonants yield a blended realization involving the summation of the closure areas for C1 and C2 at C1 midpoint and offset. A requirement for laterals to leave enough room behind the place of articulation may delay the initiation of blending or cause blending to terminate prematurely.

The implementation of blending may help to explain why /l, n/ may assimilate in manner of articulation to a dental consonant (and perhaps why manner assimilation may occur in the case of the clusters /ln, nl/ as well). Illustrative examples are found in Tuscan [ˈspilla] SPINULA "spinule", [ˈalla] ALINA "arm", Salentino [kutˈtente] *contento* "happy", Old Romanesco *callo* for *caldo* "hot" (Rohlfs 1966: 339–40, 355), Sardinian [sa nˈnuɛzɛ] /sal ˈnuɛzɛ/ "the clouds", [laθˈθare] LANCEARE "to throw" (Contini, 1987: 139, 490), and Spanish from Panamá [ˈfadda] *falda* "skirt", [ˈatto] *alto* "high" (Alonso 1967: 234).

In addition to predicting place adaptation, the sound change data just referred to indicate that the DAC model appears to be a good predictor of place and manner of assimilation processes and of consonant elision in clusters. The fact that manner assimilations and elisions in *(a)* and *(b)* often occur in clusters with a rhotic suggests that both process types have been implemented after regressive or progressive assimilation has applied. Progressive place assimilations may also be at the origin of instances of manner assimilation and simplification in clusters with a retroflex consonant. Clusters composed of unconstrained consonants and thus, subject to blending, exhibit a different behaviour: place identification between the two consecutive consonants may give rise to manner assimilation but not to consonant elision.

Acknowledgments

This research was funded by project BFF2003-09453-C02-01 from the Spanish Ministry of Education and FEDER, by projects 1999SGR00290 and 2001SGR00425 from the Catalan Government, and by a project on Catalan phonetics from the Institut d'Estudis Catalans. I would like to thank Doug Whalen and two anonymous reviewers for comments, and Aina Espinosa and Maria Dolors Pallarès for help with data processing and statistical treatment.

References

Alonso, A.
1967 *Estudios lingüísticos (Temas hispanoamericanos)*. Madrid: Gredos.
Bladon, R. A. W. and F. J. Nolan
1977 A video-fluorographic investigation of tip and blade alveolars in English. *Journal of Phonetics* 5: 185–193.
Boyce, S. and C. Y. Espy-Wilson
1997 Coarticulatory stability in American English /r/. *Journal of the Acoustical Society of America* 101: 3741–3753.
Browman, C. P. and L. M. Goldstein
1989 Articulatory gestures as phonological units. *Phonology* 6: 201–251.
1990 Tiers in articulatory phonology, with some implications for casual speech. In: J. Kingston and M. E. Beckman (eds.), *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*, 341–376. Cambridge: Cambridge University Press.
1992 Articulatory phonology: An overview. *Phonetica* 49: 55–180.
Chomsky, N. and M. Halle
1968 *The Sound Pattern of English*. New York: Harper and Row.

Contini, M.
1987 *Étude de géographie phonétique et de phonétique instrumentale du sarde*. Alessandria: Edizioni dell'Orso.

Eliasson, S.
1986 Sandhi in Peninsular Scandinavian. In: H. Andersen (ed.), *Sandhi Phenomena in the Languages of Europe*, 271–300. The Hague: Mouton.

Ellis, L. and W. J. Hardcastle
1999 An instrumental study of alveolar to velar assimilation in careful and fast speech. In: J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville and A. C. Bailey (eds.), *Proceedings of the XIVth International Congress of Phonetic Sciences*, Volume 3, 2425–2428. San Francisco: American Institute of Physics.

Farnetani, E. and G. Busà
1994 Italian clusters in continuous speech. In: *Proceedings of the 3rd International Conference on Spoken Language Processing*, 359–362. Yokohama, Japan.

Fowler, C. A. and E. Saltzman
1993 Coordination and coarticulation in speech production. *Language and Speech* 36: 171–195.

Gossen, C. T.
1970 *Grammaire de l'ancien picard*. Paris: Klincksieck.

Holst, T. and F. Nolan
1995 The influence of syntactic structure on [s] to [ʃ] assimilation. In: B. Connell and A. Arvaniti (eds.), *Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV*, 315–333. Cambridge: Cambridge University Press.

Jones, D.
1956 *An Outline of English Phonetics*. Cambridge: Cambridge University Press.

Lehiste, I.
1964 *Acoustical Characteristics of Selected English Consonants*. Bloomington, IN: Indiana University Press.

Llorente-Maldonado, A.
1962 Fonética y fonología andaluzas. *Revista de Filología Española* 45: 227–240.

Malmberg, B.
1965 *Estudios de fonética hispánica*. Madrid: CSIC.
1971 *Phonétique générale et romane*. The Hague: Mouton.

Millardet, G.
1925 Études siciliennes. In: *Homenaje ofrecido a Menéndez Pidal*, 713–757. Madrid.

1933 Sur an ancien substrat commun à la Sicile, la Corse et la Sardaigne. *Revue de Linguistique Romane* 9: 346–369.

Monnot, M. and M. Freeman
1972 A comparison to Spanish single-tap /r/ with American /t/ and /d/ in post-stress intervocalic position. In: A. Valdman (ed.), *Papers on Linguistics and Phonetics in Memory of Pierre Delattre*, 409–416. The Hague: Mouton.

Munhall, K. G. and A. Löfqvist
1985 Gestural aggregation in speech: Laryngeal gestures. *Journal of Phonetics* 20: 111–126.

Recasens, D.
1995 Coarticulació i assimilació en fonologia. Dades de moviment lingual sobre els grups consonàntics amb C2=/d/ en català. *Caplletra* 19: 11–26.
1999 Lingual coarticulation. In: W. J. Hardcastle and N. Hewlett (eds.), *Coarticulation*, 80–104. Cambridge: Cambridge University Press.
2002 An EMA study of VCV coarticulatory direction. *Journal of the Acoustical Society of America* 111: 2828–2841.

Recasens, D. and M. D. Pallarès
2001 Coarticulation, blending and assimilation in Catalan consonant clusters. *Journal of Phonetics* 29: 273–301.

Recasens, D., M. D. Pallarès and J. Fontdevila
1997 A model of lingual coarticulation based on articulatory constraints. *Journal of the Acoustical Society of America* 102: 544–561.

Rohlfs, G.
1966 *Grammatica storica della lingua italiana e dei suoi dialetti*. Torino: Einaudi.

Steriade, D.
2001 Directional asymmetries in place assimilation. In: E. Hume and K. Johnson (eds.), *The Role of Speech Perception in Phonology*, 219–250. San Diego: Academic Press.

Stevens, K. N.
2000 *Acoustic Phonetics*. Cambridge, MA: MIT Press.

Whalen, D. H.
1990 Coarticulation is largely planned. *Journal of Phonetics* 18: 3–35.

Excrescent schwa and vowel laxing: Cross-linguistic responses to conflicting articulatory targets*

Bryan Gick and Ian Wilson

Physical conflicts between articulatory goals are inevitable in speech. One such conflict is investigated in this paper, namely that between posterior tongue targets of consonants and vowels. Multiple strategies are identified across dialects and languages for dealing with these conflicts, and described in terms of language-specific rankings of phonological constraints on the organization of articulatory gestures. In English, sequences of high tense vowel + liquid have been observed to result in the percept of an intervening schwa in some dialects (as in, e.g., heel*,* hail*,* hire*, etc.), giving many listeners the impression of an "extra" syllable or half-syllable (cf.* higher*). While previous studies have given this phenomenon special phonological status, the present paper argues that this schwa is the phonetic result of one possible strategy for reconciling an intrinsic conflict between articulatory (tongue dorsum/root) targets. This same conflict, however, results in vowel laxing/lowering in other dialects, implying a different strategy. Preliminary laboratory data are presented, first supporting the notion that the schwa percept is the incidental result of the tongue passing through a schwa-like configuration or "schwa space" during the transition between opposing tongue root targets. Second, data are presented for several additional languages supporting the notion that similar cases of articulatory conflict occur in other languages, often resulting in similar effects. The constraint-based proposal is evaluated in light of these cross-language findings.*

1. Introduction

It has been observed that sequences of high tense vowel + liquid in many dialects of English may elicit the percept of an intervening schwa (as in, e.g., *hee*[ə]*l*, *hai*[ə]*l*, *hi*[ə]*re; cf. higher*), giving many listeners the impression of an "extra" syllable or half-syllable. Previous studies have attributed special phonological status to this phenomenon, either by stipulating a process of phonological epenthesis (McCarthy 1991), proposing segment-specific constraints (Halle and Idsardi 1997; Orgun 2001), or licensing trimoraic syllables in the phonology (Lavoie and Cohn 1999; Cohn and Lavoie 2000).

It is argued in this paper that this impression of an epenthesized schwa is not an isolated phonological process restricted to English, nor is a schwa percept itself systematically regulated by the grammar (thus this paper will not include a perception study, nor a complete explanation of why a certain string may or may not be perceived as a vowel in general). Rather, this vowel is argued to be an example of "excrescent" schwa (as described by Levin 1987) – in the present case, a phonetic by-product of one strategy for reconciling an intrinsic conflict between articulatory targets. The specific conflict under investigation appears, for example, in words like *feel* and *file* in many English dialects between an advanced tongue root/dorsum target for the palatal vowel or glide, and a retracted target for the following uvular/upper pharyngeal constriction for /l/ [see Narayanan, Alwan, and Haker 1997 and Gick, Kang, and Whalen 2002 for discussion of the dorsal constriction for /l/]). A similar conflict obtains for the retracted tongue root position of /r/ in many dialects (Delattre and Freeman 1968). The strategy invoked here is thus simply to allow both of the conflicting goals to be achieved, forcing a serial ordering with an audible transition.

Another logically possible strategy for resolving a conflict between two targets, however, is to reduce or eliminate one of the conflicting targets (Wood 1996). Indeed, this same sequence of high tense vowel + liquid appears elsewhere in the literature on English dialectal phonology, but with a different result: In Utah English (Di Paolo and Faber 1991) and Pittsburgh English[1] (Walsh Dickey 1997), the vowels in words like *feel*, for example, have laxed or lowered to merge with *fill*, *pool* with *pull*, etc.

The presence of these two distinct strategies is taken to indicate that responses to this physical conflict are not automatic, but are controlled language-specifically. Thus, a proposal is made in the present paper expressing these strategies in terms of rankings of constraints on gestural organization. These rankings are argued to result in excrescent schwa in some languages, vowel laxing in others, and a context-dependent combination of excrescent schwa and laxing in still others. Using this English case as a stepping-off point, this paper also investigates similar cases of conflict across a number of languages to test for the occurrence of similar strategies.

The present paper is organized as follows: Section 2 focuses on the English case, evaluating previous analyses of the English excrescent schwa, developing support for the proposal that these phenomena are in fact the result of articulatory conflict, and proposing a constraint-based account for gestural coordination that can accommodate these observations. In Section 3, previous descriptive data and new pilot laboratory data are presented from several languages beyond English to test for similar effects, and to find whether the additional patterns predicted by the constraint-based account exist in the languages of the world; languages include Beijing Chinese, Nuu-chah-nulth (Wakashan), Chilcotin (Athapaskan) and Korean. Section 4 summarizes the cross-linguistic results and evaluates the constraint-based analysis in light of these data.

2. Excrescent schwa and vowel laxing in English

2.1. Previous analyses of excrescent schwa

As discussed above, three recent approaches have been used to account for the excrescent schwa phenomenon in English. McCarthy (1991: 198) claims that “a glide + liquid sequence presents too small a sonority cline” (see Steriade 1982). Consequently, the liquid “cannot be syllabified with the preceding diphthong and schwa epenthesis applies instead.” This account predicts that other forms with equally small sonority clines should elicit epenthesis. However, words such as *barn* and *bust* do not surface as [báɹən] or [bʌ́sət], nor does epenthesis occur even in codas with no sonority cline (e.g., *act*) or a negative cline (e.g., *adze*). This apparent epenthesis thus appears to result from qualities more specific to the liquids and tense vowels.

Orgun (2001), following an approach similar to that proposed by Halle and Idsardi (1997), uses Optimality Theoretic constraints to deal with schwa epenthesis. This account is the same in spirit as that of McCarthy (1991) above, however, Orgun uses phoneme-specific constraints (e.g., *Coda-*r*) to derive the context for the schwa. Thus, while able to get the schwa effect, this approach is ultimately driven by specifying a “special” relationship between /r/ and a particular set of preceding vowels. In both Orgun’s and Halle and Idsardi’s cases, this special status is motivated by the vocalization of postvocalic /r/ in the dialect in question (E. Massachusetts). However, it is not clear why /l/ should also elicit schwa epenthesis in this dialect, or why both /r/ and /l/ should elicit the same effect in other dialects without /r/ vocalization.

Lavoie and Cohn (1999) propose an alternative analysis, whereby tense vowel + liquid sequences constitute trimoraic syllables (which they refer to as “sesquisyllables”). While this approach holds for the high vowels, it does not explain why the low vowel /a/ fails to elicit the schwa (e.g., *hall*). Their analysis of the low vowels, proposed in Cohn and Lavoie (2000), offers the constraint that "r/l can bear a mora after [–low] vowels but not [+low] vowels." However, this constraint is essentially descriptive. Another problematic case for the Lavoie and Cohn analyses is that they find that the vowel /o/ groups with the low/lax vowels in not eliciting the excrescent schwa effect (e.g., *hole*; see Lavoie and Cohn 1999: 111). In his study of tongue root positions in English vowels, MacKay (1977) identifies another case where /o/ patterns with the low/lax vowels: Here, /o/ was the only non-low "tense" vowel with a non-advanced tongue root position. Both of these findings regarding non-high back vowels, though seemingly exceptional, converge to support the analysis we propose in this paper: It is only where there is an opposition in tongue root/dorsum position that we expect a transition to move through "schwa space". Thus, in the case of a vowel such as /o/ or /a/ followed by a liquid, there is no conflict in tongue root/dorsum target (as all of these have a retracted tongue position); therefore no transition occurs, and the tongue does not pass through a schwa-like configuration. We will pursue this connection between tongue root position and excrescent schwa throughout the rest of this paper.

2.1.1. Previous experiments: Syllable rime durations

Some laboratory tests such as measures of duration may be applied in these cases to clarify the status of the schwa. If this apparent schwa in English is in fact an inserted vowel, or if this context corresponds with an additional mora, we should expect the schwa to contribute to the duration of the syllable rime. Previous data on rime durations, however, show that these syllables in English are no longer than other syllables ending in voiced consonants.

Lavoie and Cohn (1999) give acoustic duration measurements for various three-segment sequences, including low vowel + liquid + stop (e.g., /-ald/), low vowel + glide + stop (e.g., /-ajd/) and low vowel + glide + liquid (e.g., /-ajl/). The presence of an additional timing unit or syllable in the glide + liquid cases should cause a greater duration in /-ajl/ relative to /-ald/ and /-ajd/. We are unable to make statistical comparisons without the original data, however, the durations cited for the /-ald/ and /-ajd/ cases are 213 ms and 185 ms, respectively (averaging 199 ms), while that of /-ajl/ is 201 ms. Thus, these data do not appear to indicate a substantial additional duration contributed by the excrescent schwa in glide-liquid combinations.

As the proposal in the present paper rests upon a conflict between articulatory gestures, verification of the acoustic results presented in 2.1.1 in terms of articulatory timing is perhaps more relevant. Gick and Wilson (2001) report an ultrasound experiment conducted to test the hypothesis that durations from vowel onset to final consonant closure are stable regardless of the number of intervening (vowel) gestures and regardless of segment type. Their results show no significant differences in duration between rimes ending in the three different final consonants measured (pVn vs. pVl vs. pVd; e.g., pine vs. pile vs. pied, etc.), lending further support to the prediction that the presence of the excrescent schwa does not contribute to the duration of the syllable.

2.2. Gestural modeling of the problem

Articulatory Phonology (Browman and Goldstein 1992) treats articulatory events as basic units of phonology, thus lending phonological meaning to the notions of individual gestural targets and thereby the conflicts between them. As units of phonological representation, phonological constraints may of course be applied to them. Thus, the present paper assumes a representation of articulatory gestures as phonological units in the Articulatory Phonology sense. A further advantage of Articulatory Phonology for the present study is that it allows the component gestures of compound (multi-gesture) segments such as English /l/ to be treated independently. This is crucial as presumably it is normally only individual gestures and not whole segments that come into physical conflict. In the case of /l/, the various manifestations of its component gestures in different allophones, etc., have been reasonably well studied (e.g., Sproat and Fujimura 1993; Browman and Goldstein 1992, 1995; Gick 2003), allowing us to draw upon this previous data to further understand the English excrescent schwa phenomenon, at least with regard to interactions with /l/. The following section proposes a model of speech production to help illustrate the implications of the above duration data and to link these findings with the data to follow in the remainder of the paper.

2.2.1. A production model for gestural timing in syllable codas

Sproat and Fujimura (1993) propose that English /l/ is composed of two distinct types of gestures: A “consonantal” gesture (tongue tip [TT] raising) and a “vocalic” gesture (tongue rear/root [TR] retraction). They base this categorization in part on the timing relationship between the two gestures: While both of these occur more or less simultaneously in prevocalic allophones, TR retraction occurs significantly earlier than TT closure in postvocalic allophones. This observation is very likely to be relevant to the present discussion, as it is the transition from the preceding vowel toward the TR gesture for /l/ that is proposed here to be the locus for the schwa percept.

Consider a set of three syllable rimes parallel to the example types cited above in 2.1.1, but where all three rimes involve different numbers of retraction/advancement movements of the tongue posterior, e.g., /-id/, /-ajd/, and /-ajl/. Thus, /-id/ involves only a single vocalic gesture (TR advancement for /i/); /-ajd/ involves two contrary vocalic movements (a retraction for /a/ followed by an advancement for /j/); and /-ajl/ involves three conflicting movements (retraction for /a/, advancement for /j/, and retraction again for /l/). If the overall duration is indeed the same across all of these rimes, as was established in Section 2.1.1, then as more conflicting gestures occur within that duration, they must be more tightly packed. Thus, it must be the case that the duration of the temporal lag between vocalic and consonantal gestures decreases as the number of conflicting movements increases. A schematic illustration of this hypothesis is shown in Figure 1.

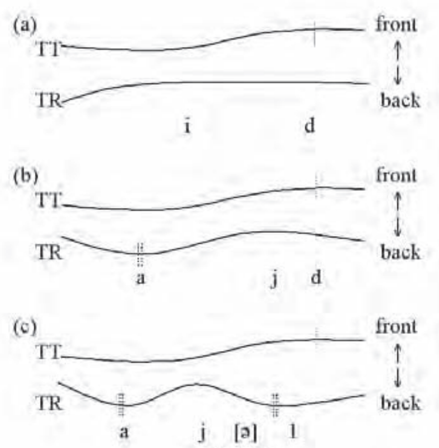


Figure 1. Schematic diagram of tongue tip (TT) vs tongue rear/root (TR) movement in (a) /-id/, (b) /-ajd/, and (c) /-ajl/ syllable rimes. Single dotted lines show estimated time of achievement of the tongue tip closure gesture for the final consonant; double dotted lines show time of achievement of TR retraction. Note multiple retraction gestures in (c).

This model suggests that stable timing is not maintained within segments postvocalically as clusters increase, at least in cases that do not allow for gestural overlap (i.e., cases of direct articulatory conflict). Rather, in this model, syllable coda timing is a relationship between the syllable peak and the first consonantal gesture, with vocalic gestures compressed into the available time window. Gick (1999) provides magnetometer data consistent with this model, showing that the temporal "lag" between the TR and TT gestures of /l/ are shortest in /-ojl/ and /-ajl/ rimes (those with the largest number of opposing TR targets). Data from Gick (1999) are shown in Figure 2. These results again support the notion that /j/ and /l/ have conflicting TR targets, and thus that the excrescent schwa appears during the resulting vocalic transition.

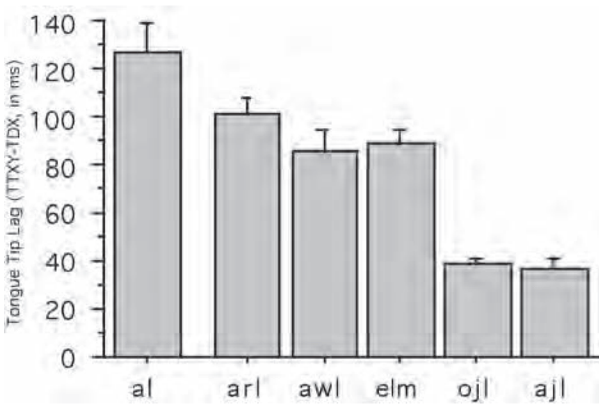


Figure 2. Tongue tip lag in postvocalic /l/ allophones (figure reproduced from Gick 1999). The three sets (al, arl/awl/elm, and ojl/ajl) are all statistically distinct.

2.3. A pilot experiment testing the "schwa space" hypothesis for English

The above sections presented evidence showing that the schwa percept in English syllable rimes does not contribute to rime duration and presented a simple production model wherein such an effect might be feasible. However, while this is consistent with the contention that this schwa is a by-product of articulatory conflict, none of the data thus far offers any explanation as to why this particular conflict should elicit the percept of a schwa, *per se* – nor has the predicted transition actually been observed to occur at all. Thus, before leaving English behind, the results of a pilot experiment clarifying these issues will be presented. An experiment was conducted using direct (ultrasound) and indirect (acoustic) measurement to illustrate: 1) whether the proposed transition from an advanced to a retracted tongue root/dorsum position visibly occurs in English high vowel/glide+/l/ clusters; and 2) whether the tongue and acoustic signal smoothly pass through a schwa-like configuration along their trajectories in the excrescent schwa cases, as proposed above.

2.3.1. Methods

Subjects: Two native speakers of North American English participated in this study, both in their late 20's, and both unaware of the nature of the experiment. W1 (female) was raised in Southwestern Ontario; M1, (male) was raised in Southwestern Manitoba.

Stimuli: Stimuli were presented in the carrier phrase "Pop is a __ ." Tokens included real and nonsense words of the form pV(G)C, where V(G) consisted of the set /a, i, aw, ej, aj, ɔj/, and where C consisted of the set /d, n, l/, giving 18 combinations. In addition to these, the form [pǽpə] was also collected to provide "canonical" schwas for comparison with excrescent schwas.

Data collection and analysis: Six repetitions of each token were collected as follows: All tokens were presented in writing to the subjects, who read them aloud; the entire list was repeated six times. The first reading of the list was discarded to ensure that subjects were accustomed to the procedure. Tokens within the list were presented in blocks of six, of which the sixth member was discarded. Given the small number of tokens per condition, statistical comparisons were not made in this pilot study.

Articulatory data were recorded to VHS from an Aloka SSD-900 portable ultrasound machine using a 3.5MHz electronic convex intercostal probe UST-9102 with a 90-degree field of view. The probe was held by the subject against his or her own neck, just above the larynx, so that a midsagittal section of the tongue was visible from the tongue root to the tongue tip. A constant probe position was maintained using a laser pointer attached to the probe. Subjects were seated facing a wall at a distance of 2 meters. The laser pointer projected an image of crosshairs onto the wall, where a 10cm square was drawn. Subjects were instructed to keep the crosshairs upright and within the square; their accuracy was monitored by the investigators during the experiment. See Gick (2002) for a more in-depth treatment of the ultrasound data collection methods used in this study.

The acoustic signal was simultaneously recorded to VHS to ensure synchronization with the video signal, using a Pro-Sound YU-34 unidirectional dynamic microphone amplified through a built-in amplifier in a Tascam cassette recorder.

After collection, videos were digitized to a Macintosh G4 from the VHS tape using an XLR8 video card with Final Cut Pro v.1.2 video editing software. Images were edited and analyzed using Final Cut Pro. Mean tongue posterior positions were calculated using Adobe Photoshop v.6 to measure the distance moved along the trajectory of greatest displacement, starting at the spatial maximum for the vowel immediately preceding /l/ and ending at the maximum point of retraction for /l/. Acoustic signals were analyzed using the freeware Praat v.3.9.13 (http://www.fon.hum.uva.nl/praat/). Formants were calculated at the midpoint of canonical schwas, and compared with the crossover points in words with excrescent schwas. A linear interpolation Burg LPC was used to automatically extract formant trajectories, with a time step of 10 ms, window length of 25 ms, and pre-emphasis from 50 Hz.

2.3.2. Results

First, it is apparent from the ultrasound data, as suggested by the acoustic transition, that the successive tongue root/dorsum targets in the excrescent schwa cases are indeed in the predicted conflict relationship. Thus, the tongue moves from a relatively advanced position in the mid pharyngeal region for all of the vowels (except /a/) to a relatively retracted position for the /l/, but not for the other postvocalic consonants. Mean distances of tongue posterior displacement for this transition movement were: (for M1) /il/=9.3mm, /awl/=3.2mm, /ejl/=12.6mm, /ajl/=5.6mm, /ɔjl/=8.5mm; and (for W1) /il/ =17.1mm, /awl/=4.5mm, /ejl/=12.7mm, /ajl/=8.5mm, /ɔjl/=10.0mm.

Second, the mean formant values for the canonical schwas were: (for M1) F1=658 Hz, F2=1217 Hz, F3=2539 Hz; and (for W1) F1=710 Hz, F2=1550 Hz, F3=2980 Hz. These formants were used to identify the crossover point in the transition from /j/ to /l/ in the word *pile*. Our findings show that in the region at approximately the midpoint of this transition both F1 and F2 cross from above to below the formant values recorded for canonical schwa. Example ultrasound images of the tongue shapes during both excrescent and final schwa are shown in Figure 3, and formant trajectories are compared in Figure 4.

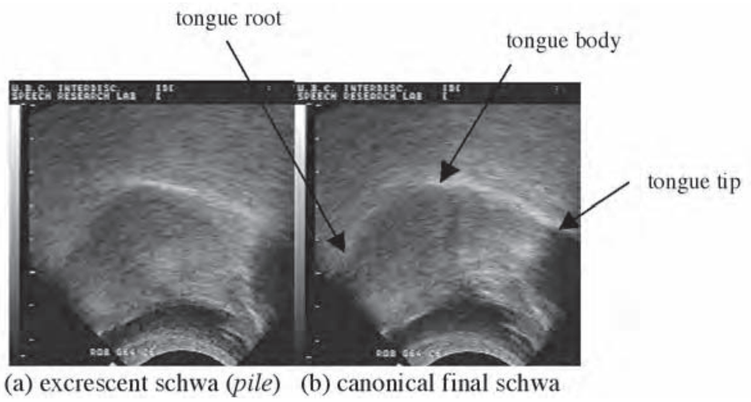


Figure 3. Midsagittal ultrasound images of tongue shape for (a) excrescent and (b) canonical schwa (M1).

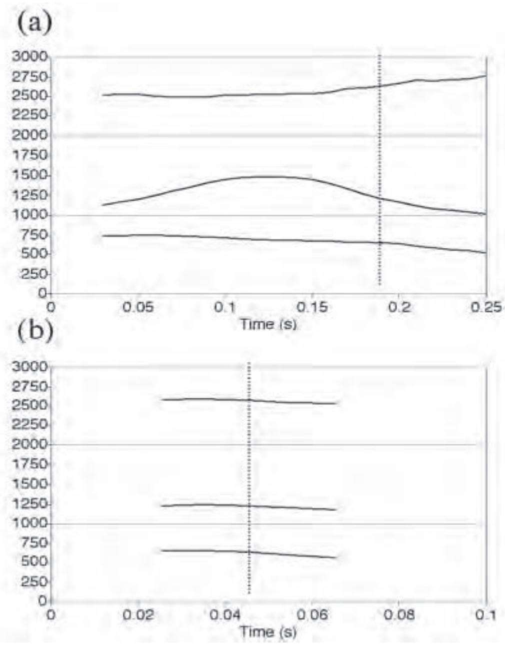


Figure 4. Example formant trajectories for (a) *pile* and (b) canonical final schwa. The vertical dotted line in (a) shows the point of intersection with canonical schwa F1 and F2; and in (b) shows the midpoint, where measures were made.

2.3.3. Discussion

The results of this experiment illustrate, as predicted, that: 1) the tongue root/dorsum targets for the tense vowels/glides and /l/ appear to be in conflict, forcing a fast transitional movement; and 2) the resulting transition moves the tongue through an articulatory and acoustic space almost identical to that of canonical schwa. Thus, it may be presumed that the percept of schwa in these cases emerges phonetically as a result of more general properties of syllable rime timing, although the actual *percept* of an additional vowel here may be influenced by other factors as well, such as rate of transition or presence of other gestures. This latter point is further suggested by the wide range in the magnitude of the effect, with /il/ and /ejl/ having the largest transitions, and /awl/ having the smallest. There is additionally a good deal of dialectal variation in the strength (impressionistically) of the glide components of certain diphthongs.

2.4. A constraint-based proposal

When one talks of different phonologies employing different strategies for resolving conflicts, it is natural to think of Optimality Theory (McCarthy and Prince 1993; Prince and Smolensky 1993) because, as Kager (1999: xi) states, "the central idea of Optimality Theory (OT) is that surface forms of language reflect resolutions of conflicts between competing demands or constraints." If the resolution of gestural conflicts is phonological, then we expect it to be possible for different languages to have different winning candidates. From the English examples of excrescent schwa and vowel laxing, it is apparent that different strategies are available for resolving cases of articulatory conflict. The two strategies that we will focus on most in the present paper are: 1) achieving all specified targets, but not achieving them simultaneously, thus allowing excrescent schwa to occur in some cases, and 2) compromising one of the component targets. As there are different strategies available for different languages, or different contexts within a language, it is a reasonable next step to propose a formalization that will allow us to identify the full set of logically possible responses to gestural conflict.

There appear in these data to be two fundamental tendencies coming into opposition. The first of these is that all gestures want to reach their targets; the second is that gestures want to occur as close as possible to the syllable peak. In cases of physical conflict, both of these constraints cannot be satisfied. In OT there are two general types of constraints, faithfulness constraints and markedness constraints (Kager 1999), each corresponding to one of these two tendencies. Thus, as the excrescent schwa scenario discussed above is the result of two segments both having their specified gestural targets achieved, this may be easily expressed in terms of a general faithfulness constraint on gestures. The constraint we propose is IDENT(target), defined as follows:

IDENT(target): The gestural target of an input segment must be preserved in its output correspondent.

As for the markedness constraints, we use two Align constraints that are independently motivated in Gafos (2002: 281). These constraints are CV-COORD(INATION) and VC-COORD(INATION) and they are defined as in Gafos (p.281) as follows:

CV-COORD: the c-centre of the C gesture must be synchronous with the onset of the V gesture.

VC-COORD: the target of the C gesture must be synchronous with the release of the V gesture.

As we are only concerned with the TR targets, only these gestures will be discussed in the present paper. However, it should be noted that previous studies, as well as the production model presented in Section 2.2 above, have found that the component gestures of postvocalic compound segments are asynchronous, suggesting that VC-COORD does not apply equally to all C gestures in English. In fact, the posterior gestures in question are specifically the ones termed "vocalic" by Sproat and Fujimura (1993), and "V-gestures" by Gick (2003). It is perhaps natural that they should be more closely coordinated with the syllable peak.

The three constraints listed above can be ranked with respect to each other in numerous ways. Table 1 outlines the possible rankings. The following section investigates similar cases of conflict in several additional languages that are included in Table 1 to test whether these constraints adequately account for cross-linguistic patterns in gestural conflict resolution. The proposed constraint-based analysis is evaluated with respect to these data in Section 4.

Table 1.

	CV (onset) result	VC (coda) result	Constraint Ranking	Languages
A	—	ə	IDENT >> VC-COORD (CV-COORD doesn't apply because no conflict in onset)	Many dialects of English; Beijing Chinese
B	laxing/ lowering	ə	CV-COORD >> IDENT >> VC-COORD	Nuu-chah-nulth
C	ə	ə	IDENT >> CV-COORD, VC-COORD	?
D	ə	laxing/ lowering	VC-COORD >> IDENT >> CV-COORD	Chilcotin
E	laxing/ lowering	laxing/ lowering	CV-COORD, VC-COORD >> IDENT	Skye Scots Gaelic
F	—	laxing/ lowering	VC-COORD >> IDENT (CV-COORD doesn't apply because no conflict in onset)	Pittsburgh/Utah English; Korean
G	ə	—	IDENT >> CV-COORD (VC-COORD doesn't apply because no conflict in coda)	No known potential cases (i.e., where a back gesture is present in onset but not coda allophone)
H	laxing/ lowering	—	CV-COORD >> IDENT (VC-COORD doesn't apply because no conflict in coda)	No known potential cases (i.e., where a back gesture is present in onset but not coda allophone)
I	—	—	Not ranked because there is no conflict	All languages lacking or avoiding conflicting sequences, e.g., Yoruba, Cook Island Maori, Bernera Scots Gaelic

3. Cross-linguistic cases of gestural conflict

This section draws on field description and pilot ultrasound imaging and acoustic data to observe cross-linguistic responses to tongue posterior conflicts similar to that discussed above for English. A number of languages have been identified where similar articulatory conflicts are expected to arise, and conflict resolution strategies used in different languages are surveyed.

3.1. Excrescent schwa

In this section we identify several cross-linguistic examples of one strategy languages use to resolve conflicting TR targets, namely that all targets are achieved in sequence, and that this strategy may at least sometimes result in the percept of an intervening excrescent schwa. This shows first that English is not an isolated case, and second, allows for a wider range of manifestations of this strategy to be observed.

3.1.1. Beijing Chinese

A similar pattern to that of English excrescent schwa has been reported in Beijing Chinese. Beijing Chinese does not normally have non-nasal postvocalic consonants at all, with the one exception of the diminutivizing suffix *-r*. This postvocalic *-r* impressionistically has a pharyngeal retroflex quality surprisingly similar to that of the rather unusual postvocalic /r/ of English.

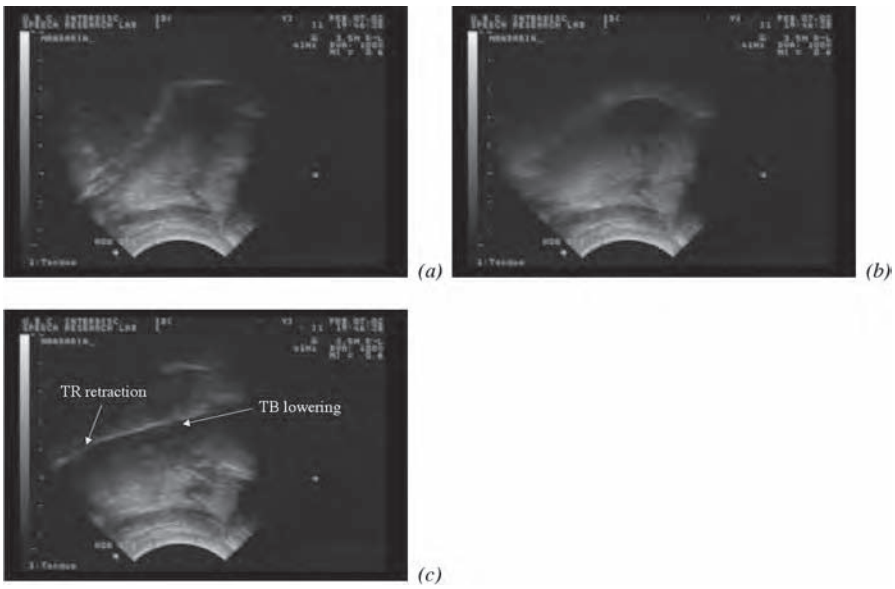


Figure 5. Beijing Mandarin /pi-r/ 'skin-DIM'. (a) 0 ms: 1st part of /i/. (b) 66 ms: 2nd part of /i/ ([ə]); note the tongue root retraction. (c) 210 ms: end of retroflex approximant; note the tongue root retraction and tongue body lowering.

According to Pulleyblank (2003), adding this suffix to a word ending in /i/, such as *jī* [tɕi] 'chicken' results not in *jīr* [tɕiɻ], but rather it "requires the insertion of a schwa vowel before the suffix, giving [tɕjər][2]."

Pilot ultrasound data was collected to test whether this phenomenon is parallel to the English excrescent schwa, by verifying that the Beijing Chinese postvocalic /r/ involves a tongue root or dorsum retraction opposing the advanced position for /i/.

The speech of three adult speakers of Beijing Chinese was measured (see 2.3.1 above for data collection and processing methods). All speakers exhibited tongue root advancement for /i/, a tongue root retraction for /r/, and the predicted transition through an apparent schwa-like configuration. As there is no tongue root retraction in the prevocalic liquid (see Gick, Campbell, Oh and Tamburri-Watt in press), and hence no conflict, this appears to be a language of type A in Table 1. Results for one subject are shown in Figure 5 for illustration.

3.1.2. Nuu-chah-nulth

Nuu-chah-nulth (formerly known as Nootka), a Wakashan language spoken on the west coast of Vancouver Island, British Columbia, is a language that includes a number of consonants articulated in the uvular and pharyngeal areas of the vocal tract. In order to produce constrictions at these areas of the vocal tract, it is necessary for the tongue dorsum/root to retract. There are primarily three vowels in the language, /i/, /u/, and /a/ (/e/ is also used, but only occurs in very specific styles of speech examined in the Ahousaht dialect). Given that the language has uvular and pharyngeal consonants as well as a high front vowel /i/, an articulatory conflict is predicted (assuming that /i/ specifies tongue root advancement in this language). It will be shown that Nuu-chah-nulth employs two different methods for resolving this conflict. When /i/ precedes a retracted consonant such as /q/, tongue root targets are not compromised, and an intervening excrescent schwa offglide results: [iᵊq]. However, when /i/ follows /q/, the tongue root target of /i/ is compromised (becomes retracted) and the tongue body (TB) commensurately lowers, resulting in a lowered and/or laxed vowel [qe] ~ [qɛ]. The vowel laxing/lowering case will be discussed in Section 3.2.2 below.

In words that have an /i/ followed by the uvular stop /q/, a very distinct schwa is apparent in the transition from /i/ to /q/. For example, in the word /ci:qci:qa/ 'talking' two perceived excrescent schwas are reminiscent of the

English and Beijing Chinese examples discussed above. On the other hand, no excrescent schwas are observed in a word with velar, as opposed to uvular stops: /ci:kci:ka/ '(a vessel) listing/tilting back and forth'. If the excrescent schwa in Nuu-chah-nulth is, as in English and Beijing Chinese, a purely phonetic by-product of two opposing tongue root targets (i.e. advanced for /i/ and retracted for /q/), then this conflict will be evident from the tongue images. Furthermore, as the tongue retracts from /i/ to /q/ position, the tongue is expected to pass through "schwa space" in the course of its transition. The TB lowering suggested in the transcription to occur simultaneously with TR retraction is likewise not surprising, given the volume-preserving nature of the hydrostatic tongue.

Using the same methods as described in 2.3.1, ultrasound data was collected from two native speakers of Nuu-chah-nulth to test this hypothesis. The movement from advancement to retraction can been seen in Figure 6, with TR retraction and TB lowering occurring simultaneously in the transition from the /i/ in (a) to the /q/ in (d).

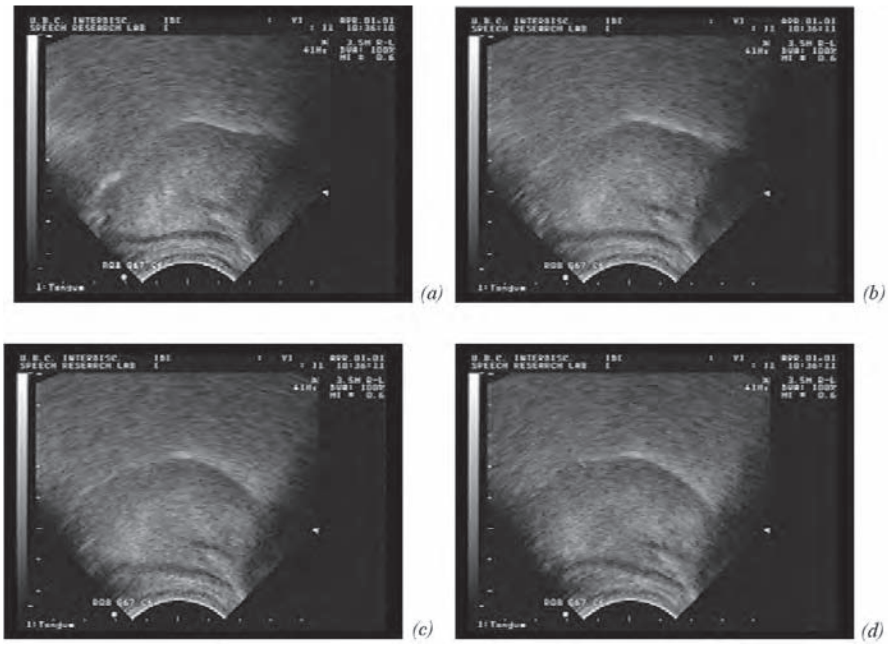


Figure 6. Nuu-chah-nulth /ci:qci:qa/ 'talking'. (a) 0 ms: 1st part of 1st /i:/. (b) 33 ms: 2nd part of 1st /i:/; note the tongue root retraction and body lowering. (c) 66 ms: end of [ə]; lead-in to /q/. (d) 99 ms: /q/ (uvular stop).

3.1.3. Chilcotin

In Chilcotin, an Athapaskan language spoken in northern British Columbia, there are a number of so-called “flat” consonants produced with a retracted tongue root (Cook 1983, 1993). These consonants include pharyngealized fricatives and affricates, as well as a set of plain and labialized uvulars (stops, fricatives, and approximants). In addition, there are three pairs of vowels described by Cook (1993: 152) as being “tense (long)” versus “lax (short)”: /i/ & /ɪ/, /æ/ & /ɛ/, and /u/ & /ʊ/ (no articulatory data exists as yet with which to verify the tongue root/dorsum position for the tense/lax distinction in these vowels). Thus, the stage is set again for a conflict in TR targets between the retracted consonants and the (presumably) advanced vowels.

Like Nuu-chah-nulth, Chilcotin exhibits the same two strategies as discussed above for resolving this TR conflict. Interestingly, however, the two strategies are employed in exactly the opposite contexts from Nuu-chah-nulth (i.e., vowel laxing/lowering occurs *preceding* back consonants, and excrescent schwa *after* back consonants). Vowel lowering will be described in Section 3.2.3 below.

According to Cook’s transcriptions, sequences of these flat consonants and the high front vowels do result in the percept of an excrescent schwa. However, unlike Nuu-chah-nulth, where the excrescent schwa appears in the transition from /i/ to a following retracted consonant, in Chilcotin the excrescent schwa occurs in the transition from a preceding retracted consonant to a tense vowel. A minimal pair from Cook (1983: 128) is given below:

(1)	/sid/	[sit]	‘I’
(2)	/ŝid/	[sˤᵊit]	‘kingfisher’

Example (1) shows that when /i/ precedes/follows a non-retracted consonant, it is realized as [i]. Example (2) shows that an excrescent schwa results when /i/ is preceded by a pharyngealized consonant.

3.2. Vowel laxing

In this section, several cases are presented where the same combination of segments (high front tense vowel + retracted consonant) consistently yields a different result: vowel lowering/laxing. It appears that, unlike the languages in the previous section, this set of languages prefers to compromise the achievement of articulatory targets. That is, rather than let the TR go to its fully advanced position for /i/, these languages reduce or eliminate that vowel gesture and produce /i/ without advanced TR, resulting in a laxed or lowered vowel (usually something impressionistically in the range of [ɪ-e-ɛ]). The implications of this pattern are discussed in more detail in Section 4 below.

3.2.1. Korean

Korean has not been previously described as a laxing language. However, Korean was chosen for this study for two reasons: first, because it gave the auditory impression of some laxing of the vowel /i/ before /l/; and second, our earlier pilot ultrasound work on Korean vowel articulation showed that /i/ appeared to be the only vowel in Korean with tongue root advancement. These two factors conspired to give the impression that this again may be a potential case of articulatory conflict. A small acoustic study was conducted, therefore, to test whether this auditory impression of laxing is accurate.

A single female native speaker of Korean in her thirties participated in this experiment. The subject produced repetitions of the words p^h*it*, p^h*in*, and p^h*il* in a fixed carrier phrase context. Values for F1 and F2 were extracted from the mid-point of each vowel using Praat software, according to the methods described in 2.3.1 above.

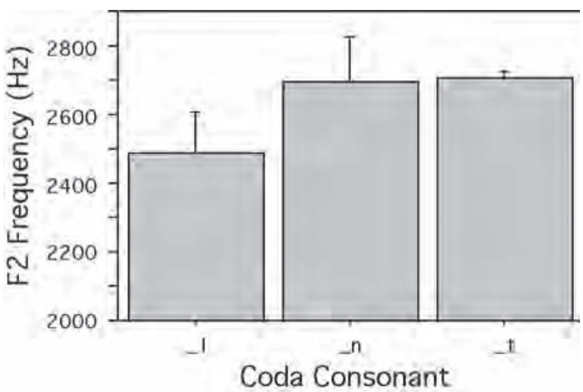


Figure 7. Effect of coda consonant on F2 of preceding /i/.

ANOVA results for this speaker indicate that F1 of /i/ is not significantly affected by coda consonant, but that F2 is ($p < .05$). Post-hoc tests (Fisher's

PLSD) indicate a significant difference (p < .05) between the effects on F2 both between final /l/ and /t/ and between /l/ and /n/, but no significant difference between the effect of final /t/ and /n/. These results are shown in Figure 7. This F2 lowering effect supports the auditory impression of laxing of Korean /i/ before /l/. As there is no such evidence of conflict in the initial allophone (see Gick, Campbell, Oh and Tamburri-Watt [in press] for relevant articulatory data), Korean must be classed with Pittsburgh and Utah English as a type F language in Table 1.

3.2.2. Nuu-chah-nulth

In Section 3.1.2, it was mentioned that in addition to the excrescent schwa cases, Nuu-chah-nulth also shows vowel lowering, i.e. a compromise of the [ATR] target of /i/ resulting in lowering of the TB. Two examples from the Ahousaht dialect of Nuu-chah-nulth are given in (3) and (4) below:

(3)	/siqiːɬ/	[sɪˤqeːɬ]	‘to cook’
(4)	/qitʃin/	[qetʃɪn]	‘louse’

The difference between /i/ in a non-retracted environment and /i/ following a retracted consonant is clearly seen in Figure 8.

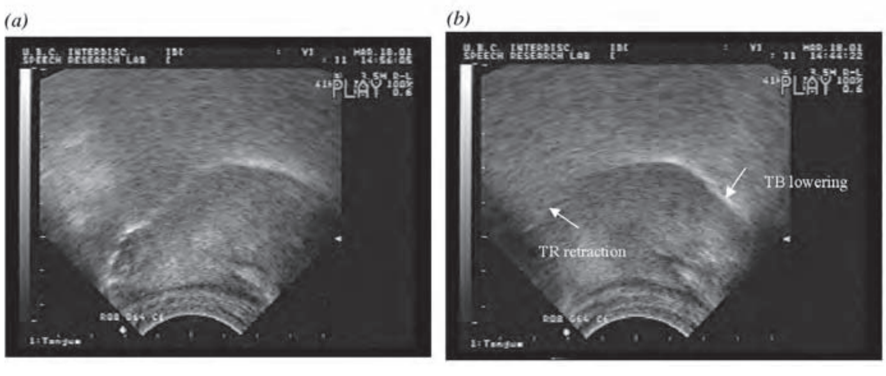


Figure 8. (a) Final /i/ in the word /-a-itʃi/ (eyebrow) – phonetically [i]. (b) First /i/ in the word /qitʃin/ (louse) – phonetically [e].

In Figure 8 (a), the advanced TR and raised TB are key to producing this [i]. However in Figure 8 (b), /i/ is seen to have a retracted TR and a lowered TB, resulting in a vowel transcribed as being close to [e] or [ɛ]. Given that Nuu-chah-nulth shows laxing/lowering in onset sequences, but serial production with an accompanying schwa percept in coda sequences, this language must be classed as type B in our Table 1.

3.2.3. Chilcotin

As discussed above, while vowel lowering takes place after /q/ in Nuu-chah-nulth, in Chilcotin it happens *before* /q/. This mirror image effect was also seen in the behaviour of the excrescent schwa in these two languages in Sections 3.1.2 and 3.1.3. Examples (5) and (6) from Cook (1993: 155) show the lowering that occurs before /q/.

(5) /niqin/ [neqᵊin] ‘we paddled’
(6) /tŝ˷iqi/ [tsˤ˷eqᵊi]˟ [tsˤ˷ᵊiqᵊi] ‘woman’

Thus, we must categorize Chilcotin as a type D language in our Table 1.

Note that both vowel lowering preceding /q/ and the excrescent schwa following /q/ occur in the word in (5). Also note that in (6), where a retracted consonant both precedes and follows the (first) /i/, there is no clear precedence: the realization of /i/ varies between [e] and [ᵊi].

3.2.4. Skye Scots Gaelic

According to the description of Borgstrøm & Oftedal (1941: 18), Skye Scots Gaelic lowers and retracts /i/ both before and after /l/. While no primary data on this language is available to the authors at present, if this description is indeed correct, then this language constitutes the only known case of symmetric responses to this conflict (i.e., where the same strategy is employed in both pre- and post-vocalic positions), and must be categorized as a type E language in Table 1.

3.3. Conflict avoidance[3]

In addition to the two main effects discussed in the above sections, it is worth mentioning an additional response to conflict: systematic avoidance (i.e., cases where sequences of ATR vowels + liquids or other retracted consonants are systematically avoided). Examples of this type of language include Yoruba, where /r/ is deleted altogether before /i/ (anonymous reviewer, p.c.); Cook Island Maori, where that language's usual epenthetic vowel, /i/, surfaces as [a] following /r/ (Kitto and de Lacy 1999); and Bernera Scots Gaelic, where tense /l/ does not appear before or after a stressed front vowel (Borgstrøm and Oftedal 1940: 68). These languages thus all fall into category I in Table 1.

4. Summary and Conclusion

The results of the survey in Section 3 are summarized in in the right-most column of Table 1. It can be seen from Table 1 that only patterns of type C (serial, symmetric) remain unattested in our sample. Given the small sample size in the present study, we are optimistic that a more thorough search will reveal languages of this type. If, however, such a language were found not to exist, it would be interesting to consider the possible relationship between this and previous observations relating to gestural positional assymmetries in syllable structure (e.g., Krakow 1999; Redford 1999). A specific question that emerges from this study relates similarly to syllable assymetries: Why are the back/vowel gestures of liquid consonants in syllable-initial allophones of liquids in most languages relatively smaller in magnitude than those in final allophones (Gick, Campbell, Oh and Tamburri-Watt in press)? Further research in these areas will be needed to clarify this relationship between the phonetic and phonological asymmetries in syllable production.

It has been argued in this paper that the apparent schwa in English tense vowel + liquid combinations is the incidental result of the tongue moving through a schwa-like configuration during the transition forced by conflicting tongue root targets. This phenomenon has helped to identify other similar cases of conflicting targets, allowing for a cross-linguistic comparison of strategies employed in resolving such articulatory conflicts. To accommodate these data, a model for gestural coordination in syllable codas has been presented whereby coda timing is determined by the relationship between syllable peak and consonant closure, but where timing is unaffected by the number of intervening vocalic events. Finally, it has been shown that the different strategies typically employed in cases of gestural conflict may be straightforwardly described in terms of the ranking of phonological constraints on articulatory gestures, and that the predictions of this constraint-based approach are supported in a cross-language survey.

An additional implication not discussed above involves the perception-production mismatch in the number of syllables in the excrescent schwa cases. If the argument in the present paper is correct, then the apparent "extra" syllable in these cases is most clearly explained as being a purely physical by-product of gestural conflict, and itself has no phonological status. Thus, for example, *feel* and *feed* are most efficiently represented as being monosyllabic, contrary to perception. This type of case will need to be addressed explicitly by theories of phonology where perception plays a more prominent role (e.g., Ladd, this volume).

Notes

* The authors wish to thank Shaffiq Rahemtulla, Sunyoung Oh, and Christine Boggs for assistance with data collection and analysis; Lisa Davidson, Abigail Cohn, Lisa Lavoie, and Doug Whalen for helpful discussion and comments; the editors of this volume and two anonymous reviewers for their insightful comments on an earlier version; previous audiences who heard earlier versions of this material and contributed their questions and insights; and our subjects for their participation. This work was funded by a grant from the Natural Science and Engineering Research Council (NSERC) of Canada, an equipment grant from the Canadian Foundation for Innovation (CFI), and NIH Grant DC-02717 to Haskins Laboratories.

1. Pittsburgh is unique in that the non-low tense vowels are laxed not only before /l/ (e.g., *feel* [ɪ], *fool* [ʊ], *fail* [ɛ], etc.) and /r/ (e.g., *here* [ɪ]), but also before /g/ (e.g., *league* [ɪ]), etc. While this case will not be pursued in detail in this paper, it is particularly interesting as it implies the testable prediction that (as we shall see below) the Pittsburgh /g/ is somehow at articulatory odds with the high tense vowels (e.g., perhaps the /g/ is produced at a more posterior place of articulation than in other dialects of English).
2. sic [r] (Pulleyblank 2003) – retroflex approximant [ɻ] in IPA.
3. Many thanks to an anonymous reviewer for suggesting this additional possibility, and for providing the Borgstrøm & Oftedal (1940, 1941) references.

References

Borgstrøm, C. H. and M. Oftedal
1940 *A Linguistic Survey of the Gaelic Dialects of Scotland: Vol 1 - The Dialects of the Outer Hebrides*. Oslo: Oslo Aschehoug.

1941 *A Linguistic Survey of the Gaelic Dialects of Scotland: Vol. 2 - The Dialects of Skye and Ross-shire*. Oslo: Oslo Aschehoug.

Browman, C. P. and L. M. Goldstein
1992 Articulatory phonology: An overview. *Phonetica* 49: 155–180.
1995 Gestural syllable position effects in American English. In: F. Bell-Berti & L. J. Raphael (eds.), *Producing Speech: Contemporary Issues. For Katherine Safford Harris*, 19–34. Woodbury, NY: American Institute of Physics.

Cohn, A. C. and L. M. Lavoie
2000 English vowel-liquid monosyllables: a case of trimoraic syllables. Paper presented at University of Massachusetts, Amherst, Linguistics Colloquium Series. March, 2000.

Cook, E.-D.
1983 Chilcotin flattening. *Canadian Journal of Linguistics* 28: 123–132.
1993 Chilcotin flattening and autosegmental phonology. *Lingua* 91: 149–174.

Delattre, P. C. and D. C. Freeman
1968 A dialect study of American r's by x-ray motion picture. *Linguistics* 44: 29–68.

Di Paolo, M. and A. Faber
1991 Phonation differences and the phonetic content of the tense-lax contrast in Utah English. *Language Variation and Change* 2: 155–204.

Gafos, A. I.
2002 A grammar of gestural coordination. *Natural Language & Linguistic Theory* 20: 269–337.

Gick, B.
1999 The organization of segment-internal gestures. In: *Proceedings of the XIVth International Congress of Phonetic Sciences*, 1789–1792. San Francisco, CA. August, 1999.
2002 The use of ultrasound for linguistic phonetic fieldwork. *Journal of the International Phonetic Association* 32(2): 113–121.
2003 Articulatory correlates of ambisyllabicity in English glides and liquids. In: J. Local, R. Ogden & R. Temple (eds.), *Papers in Laboratory Phonology VI: Constraints on Phonetic Interpretation,* 222–236. Cambridge: Cambridge University Press.

Gick, B., F. Campbell, S. Oh and L. Tamburri-Watt
in press Toward universals in the gestural organization of syllables: A cross-linguistic study of liquids. *Journal of Phonetics*.

Gick, B., A. M. Kang and D. H. Whalen
2002 MRI evidence for commonality in the post-oral articulations of English vowels and liquids. *Journal of Phonetics* 30: 357–371.

Gick, B. and I. Wilson
2001 Pre-liquid excrescent schwa: What happens when vocalic targets conflict. In: P. Dalsgaard, B. Lindberg and H. Benner (eds.), *Proceedings of the 7th European Conference on Speech Communication and Technology (Eurospeech 2001)*, 273–276. Aalborg, Denmark: Center for Personkommunikation. September 2001.

Halle, M. and W. Idsardi
1997 r, hypercorrection, and the Elsewhere Condition. In: I. Roca (ed.), *Derivations and Constraints in Phonology*, 331–348. Oxford: Clarendon Press.

Kager, R.
1999 *Optimality Theory*. Cambridge, U.K.: Cambridge University Press.

Kitto, C. and P. de Lacy
1999 A correspondence theory of epenthetic quality. In: C. Kitto and C. Smallwood (eds.), *Proceedings of the Austronesian Formal Linguistics Association (AFLA) VI*, 181–200. Toronto: Toronto Working Papers in Linguistics.

Krakow, R. A.
1999 Physiological organization of syllables: A review. *Journal of Phonetics* 27: 23–54.

Ladd, D. R.
this vol. "Distinctive phones" in surface representation.

Lavoie, L. M. and A. C. Cohn
1999 Sesquisyllables of English: The structure of vowel-liquid syllables. In: *Proceedings of the XIVth International Congress of Phonetic Sciences*, 109–112. San Francisco, CA. August, 1999.

Levin, Juliette
1987 Between epenthetic and excrescent vowels. In: M. Crowhurst (ed.), *Proceedings of the West Coast Conference on Formal Linguistics (WCCFL) 6*.

MacKay, I. R. A.
1977 Tenseness in vowels: An ultrasonic study. *Phonetica* 34: 325–351.

McCarthy, J. J.
1991 Synchronic rule inversion. *Proceedings of the Berkeley Linguistics Society* 17: 192–207.

McCarthy, J. J. and A. S. Prince
1993 Generalized alignment. In: G. E. Booij and J. van Marle (eds.), *Yearbook of Morphology 1993*, 79–153. Dordrecht: Kluwer.

Narayanan, S. S., A. A. Alwan and K. Haker
1997 Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data, Part I: The laterals. *Journal of the Acoustical Society of America* 101(2): 1064–1077.

Orgun, C. O.
2001 English r-insertion in optimality theory. *Natural Language & Linguistic Theory* 19: 737–749.

Prince, A. S. and P. Smolensky
1993 Optimality theory: Constraint interaction in generative grammar. Manuscript, Rutgers University, New Brunswick, NJ, and University of Colorado, Boulder, CO.

Pulleyblank, E.
2003 Pharyngeal glide. Ms., University of British Columbia.

Redford, M. A.
1999 An articulatory basis for the syllable. Ph.D. dissertation, University of Texas, Austin.

Sproat, R. and O. Fujimura
1993 Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics* 21: 291–311.

Steriade, D.
1982 Greek prosodies and the nature of syllabification. Ph.D. dissertation, Massachusetts Institute of Technology.

Walsh Dickey, L.
1997 The phonology of liquids. Ph.D. dissertation, University of Massachusetts, Amherst.

Wood, S. A. J.
1996 Assimilation or coarticulation? Evidence from the temporal co-ordination of tongue gestures for the palatalization of Bulgarian alveolar stops. *Journal of Phonetics* 24: 139–164.

Author index

Subject index

Language index

Note: Ethnologue codes for each language are given in brackets. See Gordon, Raymond G., Jr. (ed.), 2005, *Ethnologue: Languages of the World.* Fifteenth edition. Dallas, Texas: SIL International. Online version: http://www.ethnologue.com/.

List of contributors

Harald Baayen
University of Nijmegen and Max Planck Institute for Psycholinguistics
Netherlands
baayen@mpi.nl

Mary Baltazani
University of Ioannina
Greece
mbaltaz@cc.uoi.gr

Catherine Best
MARCS Auditory Laboratories, University of Western Sydney/Bankstown and Haskins Laboratories
Australia / U.S.A.
c.best@uws.edu.au

Diane Brentari
Purdue University
U.S.A.
brentari@purdue.edu

Dani Byrd
University of Southern California
U.S.A.
dbyrd.usc.edu

Taehong Cho
Hanyang University
Korea
tcho@hanyang.ac.kr

David P. Corina
Center for Mind and Brain
University of California, Davis
U.S.A.
corina@ucdavis.edu

Audra Dainora
Massachusetts Institute of Technology
U.S.A.
dainora@mit.edu

Rory DePaolis
James Madison University
U.S.A.
depaolra@cisat.jmu.edu

Gerard Docherty
University of Newcastle
United Kingdom
G.J.Docherty@newcastle.ac.uk

Matthew W. G. Dye
University of Bristol
United Kingdom
matt.dye@rochester.edu

Mirjam Ernestus
Max Planck Institute for Psycholinguistics and Radboud University Nijmegen
Netherlands
Mirjam.Ernestus@mpi.nl

Paul Foulkes
University of York
United Kingdom
pf11@york.ac.uk

Adamantios I. Gafos
New York University and Haskins Laboratories
U.S.A.
adamantios.gafos@nyu.edu

Bryan Gick
University of British Columbia and Haskins Laboratories
Canada / U.S.A.
gick@interchange.ubc.ca

Louis Goldstein
Yale University and Haskins Laboratories
U.S.A.
louis.goldstein@yale.edu

Harry van der Hulst
University of Connecticut
U.S.A.
harryvanderhulst@uconn.edu

Darya Kavitskaya
Yale University
U.S.A.
darya.kavitskaya@yale.edu

Heather P. Knapp
University of Washington
U.S.A.
hknapp@u.washington.edu

Alexei Kochetov
Simon Fraser University and Haskins Laboratories
Canada / U.S.A.
akocheto@sfu.ca

Els van der Kooij
Radboud University
Netherlands
e.van.der.kooij@let.ru.nl

D. Robert Ladd
University of Edinburgh
United Kingdom
bob@ling.ed.ac.uk

Gaurav Mathur
Haskins Laboratories
U.S.A.
mathur@haskins.yale.edu

Satsuki Nakai
University of Edinburgh
United Kingdom
satsuki@ling.ed.ac.uk

David M. Perlmutter
University of California, San Diego
U.S.A.
prlmttr@ling.ucsd.edu

Janet B. Pierrehumbert
Northwestern University
U.S.A.
jbp@babel.ling.northwestern.edu

Christian Rathmann
The Ohio State University
U.S.A.
rathmann.1@osu.edu

Daniel Recasens
Universitat Autònoma de Barcelona and
Institut d'Estudis Catalans
Spain
Daniel.Recasens@uab.es

Wendy Sandler
University of Haifa
Israel
wsandler@research.haifa.ac.il

James M. Scobbie
Queen Margaret University College
United Kingdom
JScobbie@QMUC.ac.uk

Stefanie Shattuck-Hufnagel
Massachusetts Institute of Technology
U.S.A.
stef@speech.mit.edu

Shui-I Shih
University of Southampton
United Kingdom
sis@soton.ac.uk

Daniel Silverman
McGill University
Canada
daniel.doron.silverman@gmail.com

Rajka Smiljanić
Northwestern University
U.S.A.
rajka@babel.ling.northwestern.edu

Jenny Tillotson
University of York
United Kingdom
jmtillotson@jmtillotson.freeserve.co.uk

Shelley Velleman
University of Massachusetts at Amherst
U.S.A.
velleman@comdis.umass.edu

Marilyn May Vihman
University of Wales, Bangor
United Kingdom
m.vihman@bangor.ac.uk

Dominic Watt
University of Aberdeen
United Kingdom
d.watt@abdn.ac.uk

Douglas H. Whalen
Haskins Laboratories
U.S.A.
whalen@haskins.yale.edu

Ian Wilson
University of Aizu
Japan
Wilson@u-aizu.ac.jp