Paola Escudero

Linguistic Perception and Second Language Acquisition

In *Linguistic Perception and Second Language Acquisition*, Paola Escudero provides a detailed description, explanation, and prediction of how optimal second language (L2) sound perception is acquired, and presents three empirical studies to test the model's theoretical principles.

The author introduces the *L2 Linguistic Perception* (L2LP) model, a new formal and comprehensive proposal which integrates, synthesizes, and improves on previous studies, and therefore constitutes the most explanatorily adequate account of the whole process of L2 sound acquisition. More specifically, it proposes that the description of optimal L1 and L2 perception allows us to predict and explain the initial state, the learning task, and the end state that are involved in the acquisition process. It advances the hypothesis of *Full Copying* which constitutes a formal linguistic explanation for the prediction that learners will initially manifest an L2 perception that matches their optimal L1 perception. It also predicts that the *degree of mismatch* between perception grammars will define the number and nature of the learning tasks. With respect to L2 development, it posits that learners will either need to create new perceptual mappings and categories, or else adjust any existing mappings through the same learning mechanisms that operate in L1 acquisition. Finally, the model's hypotheses of *separate perception grammars* and *language activation* predict that learners will achieve optimal L2 perception while preserving their optimal L1 perception.

This book addresses questions of speech perception, phonetics, phonology, psycholinguistics, and language acquisition, and should therefore be of interest to researchers working in any of these areas.

: LOT Netherlands Graduate School of Linguistics

ISBN 90-76864-80-2

: LOT 113

Paola Escudero

Linguistic Perception and Second Language Acquisition

UiL OTS

Paola Escudero

Linguistic Perception and Second Language Acquisition

Explaining the attainment of optimal phonological categorization

Utrecht Institute of Linguistics OTS

: LOT Netherlands Graduate School of Linguistics

Landelijke Onderzoekschool Taalwetenschap

Linguistic Perception and Second Language Acquisition

Explaining the attainment of optimal phonological categorization

Published by
LOT
Trans 10
3512 JK Utrecht
The Netherlands

phone: +31 30 253 6006
fax: +31 30 253 6000
e-mail: lot@let.uu.nl
http://wwwlot.let.uu.nl/

Cover illustration: painting by Mike Sharwood Smith

ISBN 90-76864-80-2

NUR 632

Linguistic Perception and Second Language Acquisition

Explaining the attainment of optimal phonological categorization

Linguïstische Perceptie en Tweedetaalverwerving,
of hoe men leert optimaal fonologisch te categoriseren

(with summaries in Spanish, English, and Dutch)

Proefschrift

ter verkrijging van de graad van doctor
aan de Universiteit Utrecht op gezag van
de Rector Magnificus, Prof. dr. W. H. Gispen,
ingevolge het besluit van het College voor Promoties
in het openbaar te verdedigen
op dinsdag 8 november 2005
des middags te 12:45 uur

door

Paola Rocío Escudero Neyra
geboren op 5 december 1976 te Lima, Perú

Promotoren: Prof. dr. W. Zonneveld
Prof. dr. P.P.G. Boersma (UvA)

Co-promotor: dr. R.W.J. Kager

A Marco y Rocío,

los cimientos y pilares de mi vida

Contents

0 Introduction 1
0.1 Why L2 *perception*? 1
0.2 Contribution and outline 4

PART I: LINGUISTIC MODELLING OF SOUND PERCEPTION AND ITS ACQUISITION

1 Modelling speech perception 7
1.1 Modelling speech perception as an auditory mapping 9
1.1.1 Speech perception as a single universal mapping 9
1.1.2 Speech perception has a universal and a linguistic component 11
1.2 Evidence for the linguistic nature of speech perception 13
1.2.1 Auditory perception versus linguistic perception 14
1.2.2 Language-specific one-dimensional sound categorization. 17
1.2.3 Language-specific auditory cue integration 21
1.3 Modelling speech perception as a language-specific phenomenon 26
1.3.1 Language-specific perception within phonetics 27
1.3.2 Language-specific perception within psycholinguistics 29
1.3.3 Language-specific perception within phonology 32
1.4 Summary and implications 35
1.4.1 Resolving the nature of sound representation 36
1.4.2 How to model linguistic perceptual mappings 37
1.4.3 Implications for a comprehensive model of sound categorization 38

2 Linguistic Perception (LP): a phonological model of sound perception 41
2.1 The elements of Linguistic Perception (LP) 42
2.1.1 Perceptual mapping component: the *perception grammar* 44
2.1.2 Representational component: the *perceptual input* 49
2.2 The optimal perception hypothesis 52
2.2.1 Optimal one-dimensional categorization 53
2.2.2 Optimal cue integration 58
2.3 Acquiring optimal L1 linguistic perception 65
2.3.1 Initial perception grammar 66

2.3.2 The Gradual Learning Algorithm (GLA) 68
2.3.3 Learning mechanism 1: one-dimensional auditory-driven learning 71
2.3.4 Learning mechanism 2: lexicon-driven learning and cue integration 75
2.4 The proposal for word recognition 77
2.4.1 Lexical representations and recognition grammar 77
2.4.2 The L1 acquisition of optimal L1 recognition 79
2.4.3 Summary: adult Linguistic Perception and its L1 acquisition 81

PART II: MODELLING THE L2 ACQUISITION OF SOUND PERCEPTION

3 The Second Language Linguistic Perception (L2LP) model 85
3.1 The L2LP model: five ingredients 85
3.1.1 Distinction between perceptual mappings and sound categories... 86
3.1.2 L2LP ingredient 1: optimal L1 perception and optimal target L2 perception 87
3.1.2.1 L2LP ingredient 1: prediction and explanation 88
3.1.2.2 L2LP phonological/phonetic description 89
3.1.3 The logical states of L2 sound perception and the L2LP model 94
3.2 L2LP ingredient 2: the L2 initial state 97
3.2.1 L2LP prediction: L2 initial equals cross-language perception 98
3.2.2 Background explanation: L1 Transfer 99
3.2.3 L2LP explanation/description 100
3.2.3.1 Full Copying of L1 perceptual mappings 101
3.2.3.2 Already-categorized versus non-previously categorized dimensions 102
3.2.3.3 Phonemic equation and category re-use 104
3.3 Ingredient 3: The L2 learning task 105
3.3.1 Prediction: learning task equals cross-language difference 105
3.3.2 Explanation/description: perceptual and representational tasks 107
3.3.2.1 L2LP perceptual task: Changing and creating mappings 107
3.3.2.2 L2 representational task: Changing the number of L2 categories 109

3.4 Ingredient 4: L2 development 109
3.4.1 L2LP prediction: L2 development equals L1 development 111
3.4.2 Background explanation: access to development and learning mechanisms 111
3.4.3 L2LP explanation/description: Full Access to the GLA.................112
3.4.3.1 GLA category formation in L2 development 112
3.4.3.2 GLA category boundary shifts in L2 development 113
3.5 Ingredient 5: the L2 end state 113
3.5.1 L2LP prediction: optimal L2 and optimal L1 114
3.5.2 Background explanation: limitations for the L2 end state...............115
3.5.2.1 The role of cognitive plasticity and the L2 input....................115
3.5.2.2 The interrelation between the L1 and the L2……… …. 116
3.5.3 L2LP explanation/description: Input versus plasticity 117
3.5.3.1 Rich L2 input overrules small cognitive plasticity 117
3.5.3.2 The hypothesis of separate perception grammars 118
3.6 Summary and L2LP sound perception scenarios 121
3.6.1 Learning scenarios: L2LP prediction/explanation............................123
3.6.2 Scenarios: L2LP description of the different learning tasks 124

4 A review of other L2 sound perception models 127
4.1 Aim and scope of five L2 perception models...127
4.2 Speech perception and its acquisition 128
4.2.1 Speech perception in phonological models of L2 sound perception 129
4.2.2 Speech perception within phonetic models of L2 perception 130
4.2.3 L1 acquisition within the five models 131
4.2.4 Comparison with the L2LP's framework model..............................132
4.3 L2 sound perception 135
4.3.1 L2 initial state 135
4.3.1.1 Major's OPM and Brown's PIM..135
4.3.1.2 PAM, NLM, and SLM.. 136
4.3.1.3 Comparison with the L2LP initial state 137
4.3.2 L2 development 139
4.3.2.1 OPM and PIM's developmental proposals……………… 139
4.3.2.2 PAM, NLM, and SLM's developmental proposals..…….. 140
4.3.2.3 Comparison with the L2LP developmental state .. . …… 141

4.3.3 L2 end state 143
4.3.3.1 Comparison with the L2LP end state.... 145
4.3.4 L2 sound perception scenarios....146
4.3.4.1 Comparison with the L2LP scenarios....149
4.4 Summary and general comparison with the L2LP model....150

PART III: EMPIRICAL TESTS OF THE L2LP MODEL

5 Learning NEW L2 sounds 155
5.1 What does learning to perceive NEW sound categories involve? 158
5.2 L2 Linguistic Perception in a NEW scenario....161
5.2.1 Ingredient 1: predicting L1 and target L2 optimal perception....161
5.2.2 Ingredient 2: predicting cross-language and initial L2 perception 170
5.2.3 Ingredient 3: predicting the L2 learning task 173
5.2.4 Ingredient 4: predicting L2 development....175
5.2.5 Ingredient 5: predicting the L2 end state....180
5.3 Evidence: Spanish learners of Southern British English (SBE)....181
5.3.1 Model ingredient 1: Spanish and SBE perception data....183
5.3.2 Model ingredient 2: cross-language and initial L2 perception data....187
5.3.3 Spanish learners' development and end state 195
5.3.4 Discussion....198
5.4 Learning new sounds: L2LP predictions versus the evidence....200

6 Learning SUBSET L2 sounds....203
6.1 Is there a learning task in a SUBSET L2 perception scenario.... 204
6.2 Ingredients of L2 linguistic perception in a SUBSET scenario....209
6.2.1 Ingredient 1: predicting optimal perception from environmental production....209
6.2.2 Ingredient 2: predicting cross-language and initial L2 perception 214
6.2.3 Ingredient 3: predicting the L2 learning task 218
6.2.4 Ingredient 4: predicting L2 development....220
6.2.5 Ingredient 5: predicting the L2 end state....236

6.3 Evidence: Dutch learners of Spanish 238
6.3.1 Model ingredient 1: Dutch and Spanish perception data.......... 241
6.3.2 Model ingredient 2: cross-language and initial L2 perception data....................245
6.3.3 Dutch learners' L2 perception data....................246
6.3.4 Discussion....................251
6.4 Learning SUBSET sounds: the predictions versus the evidence..........251

7 Learning SIMILAR L2 sounds 255
7.1 Is there an L2 learning task in a SIMILAR scenario?....................257
7.2 Ingredients of L2 linguistic perception in a SIMILAR scenario....................260
7.2.1 Ingredient 1: predicting optimal perception from environmental production....................261
7.2.2 Ingredient 2: predicting cross-language perception and initial L2 perception 271
7.2.3 Ingredient 3: predicting the L2 learning task 273
7.2.4 Ingredient 4: predicting L2 development....................276
7.2.5 Ingredient 5: predicting the L2 end state....................280
7.3 Empirical evidence A: Spanish learners of Scottish English (SE)....................283
7.3.1 Scottish English (SE) and Spanish perception 283
7.3.2 Cross-language and initial L2 perception 286
7.3.3 L2 development in Spanish learners of Scottish English (SE) 288
7.3.4 Discussion....................291
7.4 Empirical evidence B: Canadian English (CE) learners of Canadian French (CF)....................292
7.4.1 Canadian English (CE) and Canadian French (CF) perception 292
7.4.2 Cross-language perceptual mismatch and L2 initial state 295
7.4.3 L2 development in Canadian English (CE) learners of Canadian French (CF)....................297
7.4.4 Discussion....................301
7.5 Learning similar sounds: the L2LP predictions versus the evidence....................302

8 Evaluation and conclusion....................305
8.1 Why a linguistic model of sound perception?....................305

8.2 What does the L2LP model provide? 308
8.2.1 A thorough description of the learner's L1 and target L2 308
8.2.2 A linguistic model for the L2 initial state 309
8.2.3 A thorough description of the L2 learning task 310
8.2.4 An explicit and comprehensive proposal for L2 development 311
8.2.5 An explanation for the attainment of optimal L2 sound perception 313
8.2.6 Three different scenarios and their comparative learning paths 313
8.3 Overall contribution 315
8.4 Future research 316

Resumen 319
Summary 325
Samenvatting 331
References 337
Acknowledgements 349
Curriculum Vitae 351

0 Introduction

It is well known that second language (L2) learners have great difficulty when attempting to learn L2 sounds. This difficulty is clearly observed in the phenomenon commonly known as 'foreign-accented speech' which seems to be characteristic of most adult L2 learners. Typically, the latter are outperformed by infants and young children when the task is to learn the sounds of a language. That is, every child learns to produce and perceive ambient language sounds resembling adult performance in that language. In contrast, adult learners struggle to acquire native-like performance and commonly maintain a foreign accent even after having spent several years in an L2 environment. This paradoxical situation has sociological consequences since the general abilities of adult L2 learners are commonly judged on the basis of their language skills. Therefore, if their speech is not intelligible or 'accented', it may impede communication and even prevent integration into the community of native speakers.

The primary objective of the present study is to provide a comprehensive description, explanation, and prediction of how L2 sound perception is acquired. Below, I will first discuss the arguments in favour of focusing on L2 perception and then explain the difficulties involved in L2 production. Finally, I will outline the contents of this study.

0.1 Why L2 *perception*?

In early phonological theory, the role of perception in explaining the performance of L2 speakers was taken very seriously. This approach was manifested in the writings of esteemed researchers such as Polivanov & Trubetzkoy in the first half of the 20th century. Polivanov (1931) provided several anecdotal examples of how the phonemes of an L2 are perceived through the L1 system. These examples could be taken to mean that the difficulties in the production of L2 sounds arise from the influence of L1 perception. In addition, Trubetzkoy (1939/1969) also suggested that the inadequate production of L2 sounds had a perceptual basis since he considered that the L1 system acted as a 'phonological filter' through which L2 sounds are perceived and classified. However, due to the comparative ease of collecting empirical data for L2 production, the phenomenon of 'foreign accented speech' was almost exclusively addressed and explained from the point of view of production difficulties. The most prominent early exemplars of this tradition are, among others, Lado (1957), Eckman (1977, 1981), and Major (1987).

Although most observations and explanations of L2 segmental phonology have been based on production data, approaches based on perceptual difficulty have also been considered, though mainly in the field of phonetics. Cross-linguistic speech perception research performed in the 1960s showed that L2 learners also have 'perceptual foreign accents', i.e., their perception is shaped by the perceptual system of their first language (cf. Strange 1995: 22, 39). This seems to suggest that the origin of a foreign accent is the use of language-specific perceptual strategies that are entrenched in the L2 learner and that cannot be avoided when encountering L2 sound categories. In other words, problems producing L2 sounds could originate in large measure from difficulties in perceiving such sounds accurately, that is, in a native-like fashion. I argue that a full account of L2 segmental phonology should explain the way in which L2 speakers manage to learn how L2 segments should sound before explaining how they achieve accurate L2 production. This is because the accurate knowledge of L2 sounds can only emerge from the learner's ability to perceive such sounds correctly and to form appropriate representations of them.

Several researchers have addressed the controversy surrounding the interplay between the perception and production of L2 sounds, and compilations of the studies that consider such an interrelation are abundant. For instance, Llisterri (1995) and Leather (1999), among others, reviewed a number of studies supporting the argument that the L2 development of perception precedes that of production, and that accurate perception is a prerequisite for accurate production. Borden, Gerber & Milsark (1983) found that Korean learners of the English /r/-/l/ contrast had more native-like phonemic identification and self-perception than production, and suggested that perceptual abilities might be a prerequisite for accurate production. Neufeld (1988) described his findings as representing a 'phonological asymmetry' since his learners often showed to be much better at perceptually detecting sound errors than at avoiding producing them. Barry (1989) and Grasseger (1991) found that learners who showed "well-established perceptual categories" also manifested accurate production, arguing that perceptual tests can be a good means for detecting difficulties in producing L2 vowels and consonants. Further support for the hypothesis that L2 perception develops before and is a prerequisite to L2 production is also provided in Flege (1993) and Rochet (1995).

However, some studies have challenged this intuitive and widely evidenced property of L2 sound acquisition. For instance, Goto (1971) and Sheldon &

Strange (1982) found that, for Japanese learners of English, perceptual mastery of the English /r/-/l/ contrast does not necessarily precede and may even lag behind acceptable production. Sheldon (1985) reanalysed Borden et al's (1983) results and argued that their conclusion did not apply to all learners, given her findings that the longer an exposure to the L2 learners had had, the less possible it became to find that their perception was superior to their production. Flege & Eefting (1987) found that their Dutch learners produced substantial differences between stop consonants in their two languages but that they had only a small shift in the location of the category boundary when identifying the stops in the two languages. This suggested that the distinction between the two languages was not as clear in perception as in production. Furthermore, bilingual studies (Caramazza et al. 1973, Elman et al. 1977, Mack 1989) have shown that production can be more accurate than perception. For instance, Caramazza et al. (1973) tested the perception and production of voiced and unvoiced consonants among Canadian English-French bilinguals, and found that the production of their less proficient or non-dominant language was better than its perception.

Although these types of arguments may to some extent contradict the fact that L2 perception develops before production and that the former ability should be in place before the latter is mastered, these experimental studies evince shortcomings that may have influenced the conclusions that were drawn from them. For instance, Flege & Eefting's findings along with those of the bilingual literature may be due to a problematic manipulation of the 'language set' variable resulting in the activation of two languages (cf. Chapter 3). From the results of this study, it can be inferred that the lack of rigorous control in language set affected the learners' perception abilities more than their production abilities. Therefore, given the weight of the evidence, it can be concluded that perception develops first and needs to be in place before production development can occur, and also that the difficulties with L2 sounds have a perceptual basis such that incorrect perception leads to incorrect production. This means that prioritizing the role of perception in explaining the acquisition of L2 sounds seems to be valid and is perhaps the most propitious way of approaching the phenomenon. In fact, many L2 proposals mainly from the field of phonetics assume that a learner's ability to perceive non-native sounds plays a crucial role in the acquisition of L2 segmental phonology.

0.2 Contribution and outline

This study is intended to constitute a theoretical and empirical contribution to the fields of second language acquisition and phonetics/phonology.[1] With respect to the theoretical contribution, it advances a *linguistic* model of L2 sound perception, which is a phenomenon that has often been considered outside the domain of linguistic theory proper and the subject matter of disciplines such as phonetics and psycholinguistics.

There are three main parts to this study. Part I discusses the general phenomenon of speech perception and the first language (L1) acquisition of speech perception, Part II introduces a new model of L2 sound perception and examines the models that have preceded it, and Part III presents empirical data to test and evaluate the L2 proposal. Part I comprises two chapters which motivate the theoretical assumptions of the L2 model advanced within Part II of this study. In Chapter 1, I discuss the ways in which speech perception has been modelled in the literature, the evidence in favour of bringing speech perception into the domain of phonological theory, and the criteria that are required for a comprehensive model of sound perception. In Chapter 2, I discuss in detail the *Linguistic Perception* (LP) model, which I consider to be the most explanatorily adequate proposal for speech perception and its acquisition. This model's general speech perception proposal is based on Boersma (1998) and on Escudero & Boersma (2003), and the first language (L1) acquisition proposal is based on Boersma, Escudero & Hayes (2003). Chapter 2 contains my personal interpretation and explanation of the speech perception proposal as well as the language acquisition issues raised in these three articles. Throughout the chapter, it is clearly stated how this version differs from the original proposals.

Part II of this study deals with theoretical proposals for L2 sound perception. In chapter 3, I advance a linguistic model for L2 sound perception which aims at describing, explaining, and predicting L2 performance in the three logical states of language acquisition, namely the initial state, the developmental state, and the end state. This is the essence of the Second-Language Linguistic Perception (L2LP) model. This model has five theoretical ingredients, which are also methodological phases, and these ingredients allow for a thorough handling of L2 sound percep-

[1] My research has been funded by the Utrecht Institute of Linguistics since October 2001, but some of my work on this subject dates from 2000, and many of my articles written (or co-written) between 2000 and 2004 are the result of previous research.

tion. Most importantly, it provides a connection between the acquisition states in L2 sound perception through the proposed rigorous description of the learner's L1 and target L2, and through an explicit account of the L2 learning task. In chapter 4, I review five models of L2 sound perception and compare them to the L2LP model with respect to their general speech perception and L2 acquisition proposals. It is concluded that the L2LP synthesizes previous proposals and improves on their explanatory adequacy. In this chapter, the comparison is made only on theoretical grounds but the models' predictions for L2 sound perception in diverse learning scenarios are clearly stated so that the reader can evaluate their validity in view of the L2 perception data presented in last part of the study.

Part III constitutes the empirical portion of this study. It presents L2 sound perception data that document three different learning scenarios in three different chapters. Two well-attested L2 sound categorization scenarios are considered: a NEW scenario in which learners are confronted with L2 phonological categories (i.e., phonemes) that do not exist in their L1, and a SIMILAR scenario in which learners are confronted with L2 phonemes that have counterparts in their L1. Moreover, it is proposed that there exists another scenario called SUBSET which has not previously been considered in other models of L2 sound perception. In this scenario, learners are confronted with L2 phonological categories that have more than one counterpart in their L1, and which therefore constitute a subset of their L1 categories. Although previous research has not found this third scenario to constitute a learning problem, the L2LP model predicts that L2 learners will encounter difficulties if the L2 sounds form a subset of their L1 sound categories. This model gives specific predictions, explanations, and descriptions, and it proposes a comparative level of L2 difficulty for each of the three scenarios. In each empirical chapter (cf. Chapters 5 to 7), cases illustrating these specific learning scenarios are theoretically problematized and empirically tested.

Finally, Chapter 8 provides a general discussion of the findings as they relate to the proposed L2LP model as well as to the other L2 sound perception models reviewed in this study. In addition, it contains the conclusions that can be drawn from the theoretical and empirical issues raised in this study as well as its foreseeable potential impact on the fields of language acquisition, phonology, phonetics, and psycholinguistics. This final chapter also addresses some potential shortcomings of the model and touches on the research that is currently envisaged to improve and further test the L2LP's theoretical and methodological proposals

PART I:

LINGUISTIC MODELLING OF SOUND PERCEPTION AND ITS ACQUISITION

1 Modelling speech perception

In this chapter, I review the types of proposals found in the literature for the modelling of speech perception. Speech perception has commonly been modelled within phonetics or psycholinguistics. However, linguistic proposals for this phenomenon also exist. The reason for considering the current status of speech perception within linguistic modelling is that the present study promotes a phonological model for describing, explaining, and predicting L2 sound perception. Before discussing modelling issues, let us start with a general definition of speech perception.

Listeners have the task of connecting the speech signal to the stored forms and their meanings in order to understand words in their language. It is through speech perception that the decoding of the speech signal into meaningful linguistic units occurs. Thus, speech perception is the act by which listeners map continuous and variable speech onto linguistic targets. Such 'mapping' of the speech signal is depicted by the connecting lines in Figure 1.1 where the nature of the speech signal is represented by the auditory continuum on the left, and the 'linguistic units' represent the targets of the perceptual mapping.

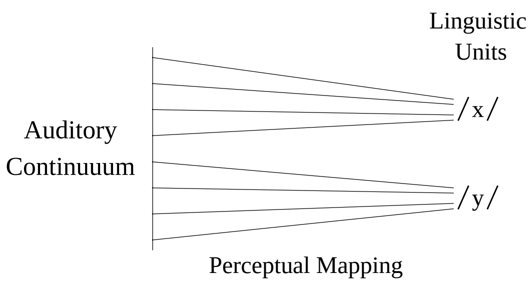

Fig. 1.1. The mapping of the auditory values of the speech signal onto linguistic units.

In this study, I concentrate on the mapping of the signal onto the phonological elements that constitute the words in a language, that is, on how the continuous and variable speech signal is mapped onto discrete and abstract phonological units, such as phonemes, phonological segments, phonological features, autosegments, or prosodic structures. Within linguistics, the decoding of the signal can be viewed as generating the mappings and representations shown in (1.1).

(1.1) Linguistics: Two mappings and three representations for comprehension.

	Mapping 1		Mapping 2	
	OF to SF		SF to UF	
[**O**vert **F**orm]	→	/**S**urface **F**orm/	→	/**U**nderlying **F**orm/

This linguistic model for speech comprehension has two mapping components, as depicted by the arrows, and three levels of representation. The first representation, the Overt Form (OF) or Phonetic Form (PF), refers to the phonetic description of a word, i.e., a detailed specification of how speech is actually pronounced, which is commonly written between brackets. For example, the word *sheep* is represented as [ʃip]. The second representation, the Surface Form (SF), refers to the phonological structure of a word, i.e., the discrete, abstract, and invariant aspects that listeners extract from the signal, which is commonly written between slashes, as in /ʃip/. The last form, the Underlying Form (UF), represents a word as it is stored in the listener's mental lexicon, i.e., the abstract and word-sized phonological form of a word paired with its meaning. This is commonly written between slashes together with its semantic meaning, which is itself commonly written between quotes, as in /ʃip/ 'fluffy animal'. Given that speech perception refers to the mapping of the signal onto phonological structure, it is considered to occur in the first mapping, i.e., OF to SF in (1.1).

In the sections below, two main issues that relate to the linguistic modelling of speech perception are discussed, namely the nature of the perceptual mapping and the nature of the targets of such a mapping. With respect to the perceptual mapping, I discuss the two basic possibilities for modelling speech perception, namely as a general auditory or language-specific process. That is, speech perception could be regarded as a mapping performed by the human auditory system, something that would imply that no linguistic knowledge is involved. Alternatively, it could be considered part of linguistic knowledge, which would imply that experience with a language results in abstract, systematic, and language-specific speech decoding.

In § 1.1, I begin by discussing proposals embedded within the most common approach to phonology which assume the general auditory or extra-linguistic nature of speech perception. In § 1.2, I discuss empirical evidence for the language-specificity of the perceptual mapping of the speech signal. Given the weight of this evidence, I argue that experience with a language results in language-specific perceptual knowledge, which means that speech mappings can be, and perhaps should be, modelled as linguistic knowledge. In § 1.3, I discuss phonetic, psycholinguistic, and phonological proposals that assume the language specificity of speech perception. Finally, in § 1.4, I examine how mapping and representations relate to each other in order to establish what sorts of forms we talk about when we refer to the 'units', 'objectives', or 'targets' of speech perception. From this discussion, I draw the components that need to be incorporated into a comprehensive linguistic model of sound perception.

1.1 Modelling speech perception as an auditory mapping

The most common approach to the modelling of speech perception assumes that this phenomenon represents a general auditory, extralinguistic, and universal capability. This assumption is illustrated, for instance, in most of the phonological proposals included in Hume & Johnson's (2001a) volume on the role of perception in phonology which contains contributions that may be considered representative of the most prevalent views in this field. Central to the auditory approach to speech perception is the idea that external phenomena, such as speech perception, interplay with but do not constitute linguistic knowledge. This view is based on a distinction between cognitive, abstract, and symbolic phenomena, on the one hand, and general physiological phenomena, on the other.

In § 1.1.1, I analyze two articles that interpret the nature of speech perception as the single universal (i.e., extra-linguistic) mapping of the speech signal. However, since not all phonological proposals that assume the universality of speech perception regard the entire mapping of the signal onto phonological representations as extra-linguistic or universal, this is followed in § 1.1.2 by a discussion of a model that explicitly suggests that speech perception has both universal and language-specific components.

1.1.1 Speech perception as a single universal mapping

Hyman (2001: 145) defines phonetics as a discipline that deals with the production, transmission, and perception of speech sounds, while he views synchronic phonology as dealing with the universal properties of sound patterns in languages and with what goes on in the minds of speakers with respect to sound patterns (p. 149). Thus, he considers speech perception to be a part of the universal component of phonetics and argues that speakers do not need to 'know' phonetics when dealing with sound patterns because no evidence is available to show that phonology is stored in phonetic terms.

However, Hyman's conclusion that "universal phonetics determines in large part what will become a language-specific phonetic property, which ultimately can be phonologized to become a structured, rule-governed part of the grammar" (Hyman 2001: 149) seems puzzling. This is because it is not obvious whether universal and language-specific phonetics each interact with phonology in the same way, nor is it evident where universal phonetics stops and where language-specific phonetics begins. What *is* clear, however, is his belief that phonetic grounding is not needed for phonological rules. However, if language-specific phonetic properties are rule governed, it seems quite likely that some kind of phonetic grounding would underlie many phonological rules. Hyman's claims about the universality of speech perception are based on the absence of evidence to the effect that listeners possess phonetic knowledge. Evidence contesting this position will be presented in § 1.2.

Not unlike Hyman (2001), Hume & Johnson (2001b) argue that speech perception is an 'external force' whose elements are tied up with physical acoustic descriptions of speech sounds and with the auditory transduction of speech sounds in the auditory periphery. They view phonology as an internal phenomenon because it deals with the cognitive symbolic representation of sound structure whose elements are dissociated from any particular physical event in the world (cf. pp. 11-12). They refer to this dichotomy as an instance of the mind/body problem, a distinction which is also found in Hale & Reiss (1998). Although Hume & Johnson propose that speech perception has a direct influence on sound patterns, they claim that this so-called external factor should not be included in phonological theory because it is not exclusive to language or, stating that "speech perception uses perceptual abilities that are also relevant to general auditory and visual perception" (p. 15). Thus, they assume that general auditory and even general perceptual mechanisms handle speech perception so that it would be erroneous to directly incorporate the mechanisms underlying speech perception into phonological analysis because this would imply that such mechanisms belong exclusively to language (cf. p. 14). However, it will be shown in § 1.2 that the perception of speech stimuli triggers different mechanisms than those of other auditory or visual stimuli, which suggests that speech perception is part of linguistic knowledge.

These phonological/linguistic proposals assume that perception may have a role to play in shaping phonological systems but that it should not be included in the linguistic component of language-specific sound structure. Within this approach, the mapping from an Overt Form (OF) to discrete categories, i.e., the first mapping in (1.1), is an automatic result of the physiological properties of the human auditory system. This automatic and extra-linguistic perceptual mapping is depicted as a double arrow in (1.2), which contains the same first mapping as in (1.1) except for the addition of the nature of this mapping.

(1.2) Speech perception as a single auditory mapping

Mapping 1:

Auditory/universal

OF ⇒ Surface Form (SF)

1.1.2 Speech perception has a universal and a linguistic component

Brown (1998) offers a proposal for speech perception that is similar to that of Hyman (2001) and Hume & Johnson (2001b) because she likewise proposes that the speech signal is first handled by universal phonetics and only afterwards by a phonological component. Crucially, all three sources refer to the initial categorization of the signal as an extra-linguistic factor, i.e., a mapping that is driven by perceptual capabilities common to all human beings and therefore part of the set of universal or general auditory capabilities.[2]

Among these, Brown (1998) contains a more developed proposal that views speech perception as a two-step mapping. She adduces the speech perception results reported in Werker & Logan (1985) as support for the traditional distinction between the phonetics and the phonology of sound patterns. These results showed that English listeners could perceive the difference between dental and retroflex Hindi stops when the inter-stimulus interval between tokens was short enough to enable auditory perception. Hence, Brown argues that universal phonetics and

[2] A similar view can be found in Steriade's (2001: 236) proposal of an external or extralinguistic *perceptability map* (P-map) to formalize the universal perceptual similarity constraints that have an effect on phonological sound patterns observed in production, such as place assimilation phenomena. Steriade's proposal is not fully discussed here because it clearly refers to production and does not give an explicit account of the nature and elements of speech perception.

phonology occur at two different levels of representation, as shown in Figure 1.2 and in (1.3). Crucially, she claims that these two levels occur sequentially during the same act of speech perception. That is, the acoustic signal is first divided into phonetic categories through a universal phonetic mapping only to be subsequently classified into native phonemic categories through the speakers' phonological structure, i.e., their feature geometry.

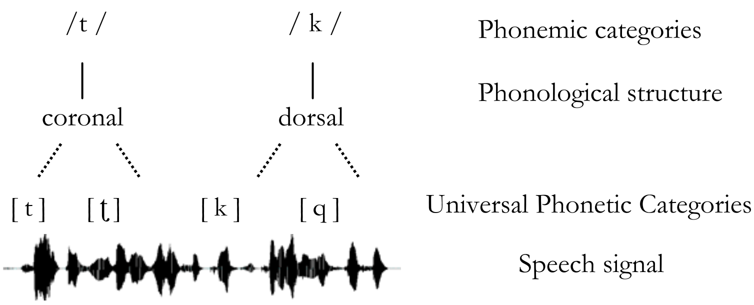

Fig. 1.2. A model of English speech perception, adapted from Brown (1998: 149).

What is noticeable in Figure 1.2 is that the mapping between the signal and the universal phonetic categories has no connecting line. This is because this mapping is considered to be an automatic result of man's general auditory system. Also, the connecting lines between the phonetic categories and the phonological structure are non-directional because Brown proposes that the phonological structure maps the phonetics, a claim that seems to imply a top-to- bottom mapping.

(1.3) Speech perception as two consecutive mappings: auditory then phonological

	Mapping 1a Auditory/universal		Mapping 1b Phonological	
OF	⇒	Universal Phonetic Form (UPF)	→	SF

Brown's model can be seen as the perceptual counterpart of Keating's (1984) production model, which also proposes the existence of an intermediate universal level of representation, as shown in (1.4).

(1.4) Keating's model for speech production

Phonological categories → UPC ⇒ OF

Although in speech production the mappings go in the opposite direction to that of speech perception, i.e., from abstract categories to the speech signal, Keating's model also proposes a two-way mapping with a universal and a language-specific component. This model, just like Brown's speech perception model, crucially suggests that speakers choose the forms they produce in their language from a finite number of universal categories, i.e., from discrete Universal Phonetic Categories (UPC). As an example of finite universal phonetic categories, Keating gives the three values for plosive consonants, viz., voiced (e.g., [b]), voiceless unaspirated (e.g., [p]), and voiceless aspirated (e.g., [p^h]). However, Cho & Ladefoged (1999) found no evidence for discrete universals in the VOT productions of 18 different languages. In fact, their data could be interpreted as a continuous distribution of VOT values across languages (cf. Boersma 1998: 276).

Thus, it would seem that although some phonological feature values appear to be organized in finite clusters across the languages of the world, there is no concrete empirical evidence to suggest that specific values are actually instantiated in these languages. Therefore, on the basis of concrete examples such as these, it can be concluded that, at least for speech production, the existence of UPCs is not borne out. This is because the production of sound categories does not yield discrete universal properties but rather yields a continuum of language-specific realizations. In the next section, I discuss the empirical evidence underlying Brown's proposal for a universal level of representation in speech perception, and I argue that this evidence is best interpreted as reflecting two modalities of perception rather than a sequence of a universal and language-specific perception.

In sum, proposals like those discussed in this section view the initial perceptual mapping of the acoustic signal onto discrete categories as the automatic result of the human auditory system. Consequently, only some so-called general auditory speech perception effects are included in their phonological proposals in order to explain various universal tendencies in the phonological system of human language. However, the actual perceptual mapping escapes phonological or linguistic modelling because it is considered to lie outside the scope of phonological theory, given its non-linguistic, non-language-specific, and automatic nature.

1.2 Evidence for the linguistic nature of speech perception

In this section, I present evidence in support of the linguistic nature of the decoding of continuous speech into language-specific sound categories. First, in § 1.2.1, I report on studies that differentiate between general auditory perception and speech perception where it is argued that the perception of sound segments is shaped by language experience and guided by perceptual mappings that are specific to the language at hand. Then, in sections § 1.2.2 and § 1.2.3, I illustrate how the decoding of the speech signal into vowels and consonants, i.e., sound categorization, is indeed language-specific. The language-specificity of sound categorization is demonstrated with the cross-linguistic differences in the classification of the same acoustic continua found in the speech signal. I also discuss the cross-linguistic differences in the integration of the same auditory dimensions in vowel categorization.

1.2.1 Auditory perception versus linguistic perception

Speech perception does not work in the same way for all listeners. Rather it gets warped or attuned to best cope with the acoustic-phonetic properties of a particular language environment. This language specificity of speech perception can best be illustrated with the differences found between the perception of sounds as acoustic reality and their interpretation as the speech of one's native language. For instance, Miyawaki et al. (1975) showed that American English and Japanese listeners differed significantly in their perception of /ra/ and /la/ if tokens of these syllables were presented within a speech context but not if they were presented in a non-speech context. That is, Japanese and American English listeners performed equally well when perceiving the main acoustic dimension that differentiates the two English consonants /r/ and /l/, (i.e., F3) when played in a non-speech context. These seemingly contradictory findings can only be explained as the workings of two different kinds of stimuli decoding, namely linguistic (because it is language-specific) versus general auditory (because it is universal). Thus, when the listeners heard the acoustic dimension that differentiates two tokens in a speech context, their language-specific knowledge guided their discrimination between such tokens, whereas when the auditory difference was placed within a non-speech context, general auditory processing guided their discrimination.

A similar result has recently been obtained for the perception of phonotactics in French and Japanese listeners. Jacquemot et al. (2003) showed that these listeners phonologically discerned the differences allowed in their linguistic systems while they auditorily discerned illegitimate differences. They tested the dissimilarity of linguistic perception and auditory perception by comparing the same two sets of stimuli across the two languages. Thus, listeners were presented with the same two contrasts, viz., the pair ebza-ebuza which receives two representations in French but only one in Japanese, and the pair ebuuza-ebuza which receives two phonological forms in Japanese but only one in French. With respect to the perception performance, the authors found that the perception of a phonological contrast, i.e., ebza-ebuza for the French and ebuuza-ebuza for the Japanese, yielded significantly better results than the perception of the auditory contrast condition. These findings demonstrate the difference between speech perception and auditory perception because the listeners perceived the phonological changes differently from the auditory changes.

Perhaps more interestingly, Jaquemot et al. (2003) also investigated brain activation when the French and Japanese listeners discriminated the tokens of their respective phonological and auditory conditions. It was found that perception in the phonological condition yielded significantly more activation in two specific areas of the brain than did perception in the auditory condition. Moreover, the two areas with more activation during phonological changes could be linked, one with the decoding of complex auditory input that is computed into abstract representations, and the other with the performance in experimental tasks involving phonological short-term memory. Therefore, both brain imaging and behavioural data were found to support the difference between auditory and phonological perception. With respect to sound perception, the authors suggested that the two brain regions involved in the perception of phonotactics might also be involved in the categorization of the speech signal into vowels and consonants. In sum, similar phonological processing may very well underlie the decoding of phonologically viable sequences of sounds as well as the decoding of segmental units.

In addition, it would seem that under certain time conditions, speech sound discrimination could go from phonological to general auditory. For instance, Werker & Logan (1985) showed that English listeners could perceive the difference between dental and retroflex Hindi stops when the time between the speech stimuli to be discriminated was reduced. That is, under a short Inter Stimulus Interval (ISI) condition, the English listeners could hear the differences between sounds that do not occur in their language. Strange (1995) interprets this result as the workings of auditory perception versus phonological perception. That is, when stimuli are closely adjacent, the auditory properties can be used to differentiate the sounds, whereas when a long silence is placed between them, listeners can only rely on abstract phonological representations. From this, it may be argued that the differential type of perception shown in (1.5) below is a more plausible interpretation of

Werker & Logan's findings than Brown's proposal of a two-step sequential perception of speech sounds, the first being extra-linguistic and the second phonological.

(1.5) Mappings and representations for speech and non-speech perception (single arrow = language-specific mapping, double arrow = universal mapping, bi-directional arrow = comparison, rep. = representation)[3]

a. Speech perception: acoustics ⇒ Auditory rep. → Phonological rep.
b. Non-speech perception: acoustics ⇒ Auditory rep.
c. Speech discrimination with short ISI: Auditory rep1 ↔ Auditory rep.
d. Speech discrimin. with long ISI: Phonological rep1 ↔ Phonological rep.

In addition, Werker & Logan's results suggest a difference between discriminating and identifying speech sounds, something which has been shown to exist in the perceptual learning of novel categories. Thus, Guenther et al. (1999) found that discrimination training led to an increase in the ability to differentiate between sounds in a particular acoustic region while identification training led to a decrease in the ability to do so in the same region. Based on these findings, I argue that different auditory stimuli and tasks yield different processing paths, as illustrated in (1.5), where a double arrow represents a mechanical/automatic auditory mapping and a single arrow represents language-specific or phonological mapping.

Developmentally, language-specific sound perception is found in pre-verbal infants during their first year of life (cf. Werker & Tees 1984; Jusczyk, Cutler, & Redantz 1993; Polka & Werker 1994). Kuhl (2000) argues that with language experience, infants develop from universal auditory discrimination to filtered or warped language-specific perception. This language-specific filtering or mapping of speech input alters their attention to the acoustic dimensions of speech in order to

---

[3] In this chart, it is assumed that the auditory representations for speech and non-speech perception are the same, viz., the output of psychoacoustic processing. This would mean that the auditory representation for speech is continuous given the general continuity of psychoacoustic scales, e.g., as shown, for instance, by the fact that the human ear can distinguish more than a thousand different pitch values (cf. Kewley-Port 1995). Although it remains an empirical question whether this is the case, the answer to this question is not relevant here. The only important point is that the auditory representations for speech discrimination are continuous, as shown in Schouten, Gerrits & van Hessen (2003). They are not discrete universal phonetic categories, as Brown interprets them to be from Werker & Logan's findings. However, their findings for short ISIs can just as well be interpreted as psychoacoustic perception, i.e., the discrimination of auditory differences.

highlight differences between the categories of their native language. Hence, Kuhl claims that "no speaker of any language perceives acoustic reality; in each case, perception is altered in the service of language" (p. 11852). However, this altering of the perceptual space seems to apply to speech only because listeners do not lose their ability to perceive auditory differences in completely non-speech contexts, such as those used by Miyawaki et al. (1975) or in contexts that trigger auditory perception, such as those involving non-phonological contrasts.

Given the weight of the evidence, it can be concluded that the decoding of the speech signal into vowels and consonants is performed through a language-specific, and therefore phonological, mapping. Of course, this view has long been the implicit norm in the field of speech perception (cf. Strange 1995 and Kuhl 2000), but not in phonology. If the language specificity of speech perception is a fact, monolingual adult listeners should exhibit a sound categorization performance that is appropriate for their own native language only, just as they exhibit the language-specific perception of phonological sound sequences. Alternatively, the decoding of sound segments may be universal so that listeners with the same vowels and consonants could very well categorize any speech stimuli in the same manner because the categories themselves might be responsible for such perceptual mapping. The next section presents cross-linguistic perceptual data that supports the language specificity of sound categorization.

1.2.2 Language-specific one-dimensional sound categorization

Cross-linguistic studies constitute a promising research area to answer questions concerning the language-specific (and therefore linguistic) or universal (and therefore psychoacoustic) nature of sound perception. For years, it has been well known that the sound systems of different languages can differ significantly from one another, and that such mismatches usually lead to the difficulties learners encounter when dealing with non-native sounds. Using phonemes, i.e., abstract phonological representations, to describe and explain segmental phonology, it was initially proposed that non-native sounds that had native counterparts would be easy to learn, whereas non-native phonemes with no such counterparts would be difficult (cf. Lado 1957). This surmise accounts for the well-attested difficulty that Japanese listeners have when trying to differentiate the English sounds /r/ and /l/ (cf. Best & Strange 1992) as well as for the comparative ease with which they can discriminate between English /r/-/w/ (cf. Halle, Best & Levitt 1999), the reason being that

Japanese does not have /l/ but does have phonemes similar to English /r/ and /w/. However, even when two languages possess phonemically equivalent sounds, difficulties may still arise because instances of such sounds may differ in narrow phonetic detail.

Abrahamson & Lisker (1970) found that although Spanish and English speakers used the same two phonemes /b/ and /p/ to categorize synthetic tokens, several of the tokens that were identified as /b/ by English listeners were identified as /p/ by Spanish listeners. Phonetically, sounds such as /b/ or /p/ are characterized by voicing properties that can be captured by the acoustic dimension of Voice Onset Time (VOT) as measured before and after the release of the stop consonant. These authors investigated the possible cross-linguistic variation in the perception of VOT in English versus Spanish listeners. To that end, they used synthetic tokens that varied from an extremely pre-voiced consonant with a voicing murmur preceding the release of the stop consonant by 150 milliseconds (–150 ms VOT) to a an extremely post-voiced stimulus that included an aspiration noise that lasted for 150 ms after the release of the stop consonant (+150 ms VOT). The synthetic series thus included pre-voiced stop consonants (e.g., [b]), voiceless non-aspirated (or short voicing lag) stops (e.g., [p]) and voiceless aspirated (or long voicing lag) stops (e.g., [p^h]).

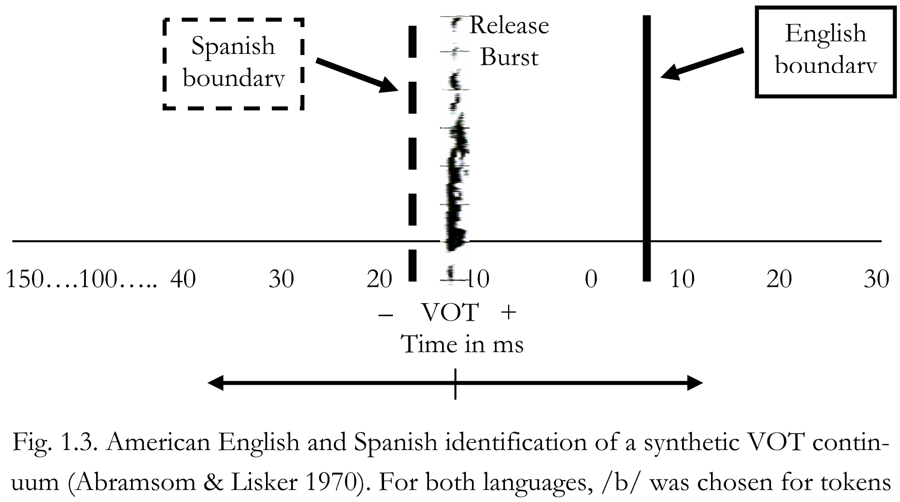

Fig. 1.3. American English and Spanish identification of a synthetic VOT continuum (Abramsom & Lisker 1970). For both languages, /b/ was chosen for tokens to the left of the boundary and /p/ for tokens to the right of the boundary.

As shown in Figure 1.3, it was found that although Spanish and English listeners divided the VOT continuum into voiced and voiceless stops, the category boundary line between these two phonemes fell in different locations for each language. That is, English listeners categorized both pre-voiced (+150 ms to 0 ms) and short lag stimuli (0 ms to –30 ms) as /b/, whereas Spanish listeners categorized short lag stimuli as /p/. Although the cross-linguistic perceptual difference seemed to be caused by a language-specific categorization of the VOT dimension, it cannot be ruled out that the consonant representations may be different. That is, we can either assume that the sound representations in the two languages are different or, alternatively, that the sounds are equivalent at an abstract level but that their realizations are processed differently in each language. To properly evaluate each of these two alternatives, we must go beyond a phonological abstract description of the sounds and examine whether the differences between them lie in their language-specific acoustic-phonetic production characteristics.

Best (2003: 2889) suggests that vowels may be particularly useful to shed light on the question of whether sound categorization is truly language-specific because these segments are produced with higher intensity, longer duration, and more acoustic dimensions than consonants. They are also fewer in number, which makes them much more variable than consonants among languages and even among dialects. Therefore, by looking at the perception of vowel segments, it may be possible to establish the language specificity of the perceptual mapping of acoustic-phonetic properties. Escudero (in progress a) tested the perception of 64 monolingual Peruvian Spanish listeners who were presented with natural tokens of Scottish English (SE) and Southern British English (SBE) of /i/ and /ɪ/,[4] which were drawn from a corpus obtained by Escudero & Boersma (2003). The listeners categorized a total of 96 target tokens, i.e., 24 tokens per vowel in each dialect. To simulate a more natural perceptual environment, 120 CVC fillers with syllables containing different vowels and consonants were also included in the stimulus set. The listeners performed a forced-choice vowel categorization task in which they were asked to choose one of the five Spanish vowel monophthongs /a, e, i, o, u/. Figure 1.4

[4] This perceptual study was conducted in Lima during my affiliation with Utrecht University with the support of a personal travel grant awarded by the Netherlands Organisation for Scientific Research (NWO). Special thanks go to Professor Jorge Perez and to Jorge Acurio for their help in the data collection process.

shows the F1 values of the SE and SBE tokens as well as the average production of the two acoustically closest Spanish vowels, viz., /i/ and /e/.

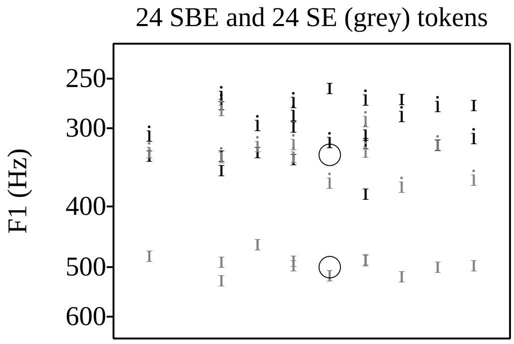

Fig. 1.4. F1 values for the SE and SBE /i/ and /ɪ/ tokens. Circles: Spanish mean productions for /i/ and /e/.

If sound perception is based on the language-specific mapping of fine-grained acoustic-phonetic information, SE and SBE /i/ tokens should be perceived as Spanish /i/, and SE and SBE /ɪ/ should be differentially perceived as Spanish /e/ and /i/ respectively. Table 1.1 shows that the majority of /i/ tokens were indeed perceived as Spanish /i/, and that SE /ɪ/ was mostly perceived as Spanish /e/, while SBE /ɪ/ was mostly perceived as Spanish /i/.

Table 1.1. Spanish (Sp.) categorization of /i/ and /ɪ/ produced in two English dialects.

	Sp. **/i/** Mean	Sp. **/e/** Mean	Sp. **/u/** Mean	Sp. **/o/** Mean	Sp. **/a/** Mean	No. of tokens	Spanish listeners
SE **/i/**	22.3	1.2	0.2	0.3	0	24	64
SE **/ɪ/**	5.2	17.5	0.4	0.7	0.2	24	64
SBE **/i/**	20.5	2.6	0.5	0.2	0.2	24	64
SBE **/ɪ/**	16.4	5.4	1.8	0.4	0	24	64

From this, it can safely be concluded that native Spanish categorization takes into account the acoustic values with which foreign tokens are produced in that it exhibits a language-specific mapping of acoustic information. In addition, other cross-linguistic studies have produced similar findings. For instance, Rochet (1995) showed that although both Portuguese and English have only two high vowels, viz., /i/ and /u/, Portuguese listeners categorize French /y/ as their own /i/ whereas English listeners categorize it as their own /u/. This was interpreted to mean that the vowel's second formant (F2) was perceived differently in each language, thereby providing further evidence that vowel categorization exhibits a language-specific mapping of the same auditory continuum.

Even more compelling support for the language-specific nature of the decoding of the acoustic signal into vowels and consonants is given by the integration of multiple auditory dimensions in sound categorization. That is, although the same acoustic dimensions may be involved in the production of sounds in various languages or language varieties, these dimensions contribute differently to language-specific categorization. When several dimensions are involved, the number of logically possible combinations increases, making it more difficult to universally and randomly select one of these combinations. The next section presents examples of language-specific perceptual cue integration and perceptual cue weighting in sound categorization.

1.2.3 Language-specific auditory cue integration

Typically, more than a single piece of acoustic-phonetic information is involved in distinguishing phonological segments in a given language environment, and listeners use those multiple sources of information when identifying or categorizing the sounds of their language. For instance, the English high front vowels /i/ and /ɪ/ combine vowel height, whose acoustic correlate is the first formant (F1), with length, whose acoustic correlate is vowel duration, because these vowels differ in F1 (cf. Peterson & Barney 1952) and in duration (cf. Peterson & Lehiste 1960). English listeners rely on both of these auditory cues when identifying these vowels, as was shown by Bohn & Flege (1990). Thus, the cross-linguistic and developmental differences in cue integration should show the language-specificity of sound categorization. Here I present examples from my own research that support the systematic and differential nature of the integration of multiple auditory continua across languages and language varieties.

Picard (1987) gives a comparative phonological and phonetic description of Canadian English (CE) and Canadian French (CF) sound inventories. According to this author, the same two IPA symbols, namely /æ/ and /ɛ/, can be used to describe the low front and mid front vowels in CE and CF. In addition, Picard predicts no cross-language difficulty for these two vowel sounds in consonant-vowel-consonant (CVC) contexts, at least in closed syllable contexts (cf. pp. 64-67). Escudero & Polka (2003) presented the same tokens of CF /æ/ and /ɛ/ to CE and CF listeners.[5] This study aimed at investigating whether listeners with the same vowel sounds used the acoustic dimensions involved in production differently. Thus, eight monolingual CE and eight monolingual CF listeners were asked to categorize 30 CVC tokens containing /æ/ and /ɛ/ produced by six adult (3 male and 3 female) CF speakers. Figure 6 shows the mean F1 and duration of the target tokens.

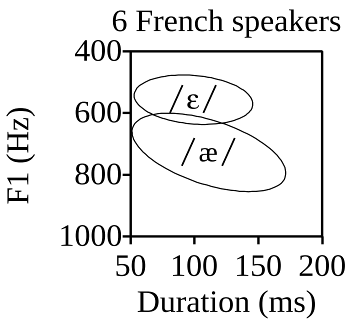

Fig. 1.5. Mean F1 and duration of the /æ/ and /ɛ/ target stimuli. Ellipses: Production distributions (one standard deviation from the mean).

As shown in Figure 1.5, the average productions of the target tokens differ in F1 and duration because /ɛ/ has a lower average F1 production and a shorter duration. However, their distributions, as depicted by the ellipses, show that /æ/ can also be produced with a short duration. It was predicted that if CE and CF listeners relied only on the abstract representations of the vowels, they would classify the tokens similarly. However, if they also relied on language-specific vowel categoriza-

5 This study was conducted during my affiliation in the School of Communication Sciences and Disorders of McGill University in collaboration with Dr. Linda Polka. It was funded by Dr. Polka's personal research grant and by a Graduate Studies Fellowship (McGill University) awarded to myself. Special thanks go to Stephanie Blue for her extensive help during the data collection process.

tion, they would exhibit differences in the classification of the /æ/ and /ɛ/ tokens. During the perception experiment, the stimuli were presented as being either English or French syllables, depending on the listeners' language background. The listeners were asked to classify the vowels in the CVCs by clicking on one of the five response options that appeared on a computer screen. The options were different for each language group in that the French listeners had the French vowel spellings for /æ/, /ɛ/, /e/, /i/, [ɪ] (an allophone of /i/ that occurs in closed syllables), while the English listeners had English keywords containing the five vowels /æ/, /ɛ/, /eɪ/, /i/, /ɪ/.

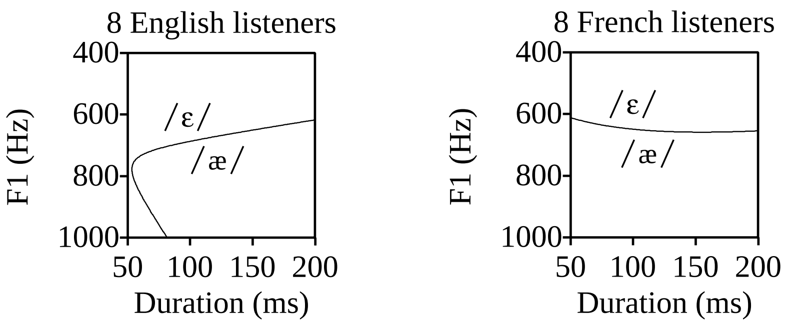

Fig. 1.6. Categorization of CF /æ/ and /ɛ/ by CF (left) and CE (right) listeners (adapted from Escudero & Boersma 2004a).

Figure 1.6 shows Escudero & Boersma's (2004a) analysis of Escudero & Polka's data,[6] adapted to show only the average perception for each language group. The solid curve in the square is the mean category boundary line which estimates where the subjects were, on average, equally likely to respond /æ/ or /ɛ/.[7] If the boundary is completely vertical, this indicates that the listeners used only vowel duration differences to categorize the tokens; if the boundary is completely diagonal, listeners integrated both F1 and duration differences to the same extent; and if the boundary is completely horizontal, listeners used F1 differences only. Thus, given the shape of their boundaries, one may assume that the CE lis-

[6] This analysis was conducted during my affiliation with the Utrecht Institute of Linguistics with funding from Utrecht University and from Boersma's NWO grant.

[7] See § 2.1.2 for an explanation of how perceptual category boundaries and auditory cue reliance have been computed. The same methodology was used as in Escudero and Boersma's (2004a) analysis of these CE and CF perceptual data.

teners used both duration and F1 to categorize the vowel tokens (diagonal boundary), while the CF listeners used mainly F1 (quasi horizontal boundary). This is shown in the differential categorization of particular tokens. For instance, tokens with a vowel duration of less than 110 ms were mostly categorized as /ɛ/ by CE listeners but as either /æ/ or /ɛ/ by CF listeners. Also, tokens with F1 values between 600 and 780 Hz and durations shorter than 110 ms were categorized as /ɛ/ by English listeners but as /æ/ by French listeners. Furthermore, CF listeners categorized /æ/ tokens as /æ/ 92% of the time but CE listeners categorized the same /æ/ tokens as /æ/ only 64% of the time, thus producing a significant categorization difference ($p < 0.01$). All in all, then, this means that the integration of the auditory information, the resulting category boundary, and the perceived distributions of the same vowel stimuli were reliably different for the two groups of listeners.

In addition, not only can we find sound categorization differences between languages but also between varieties of the same language. As an example, I present the cross-dialectal categorization of the same synthetic stimuli with the acoustic properties of the English vowels /i/ and /ɪ/. Escudero & Boersma (2003)[8] report on the vowel categorization of 20 SE speakers and 21 SBE speakers who were presented with 10 repetitions of the 37 synthetic tokens represented in Figure 1.7 (cf. § 2.1.2, Chapter 5, and Chapter 7). The bottom left and the top edge of the stimuli square were based on the spectral and durational properties of natural exemplars of /ɪ/and /i/ produced by SE speakers. The six vertical steps, which led to seven spectrally different stimuli, were equal on Mel scale (cf. Stevens, Volkmann & Newman 1937). Six horizontal fractional steps of 1.1335 were also considered and led to the seven duration values in the figure.

[8] This research project started in January 2000 and the first version of the article mentioned in the text was written before my affiliation with the Utrecht Institute of Linguistics.

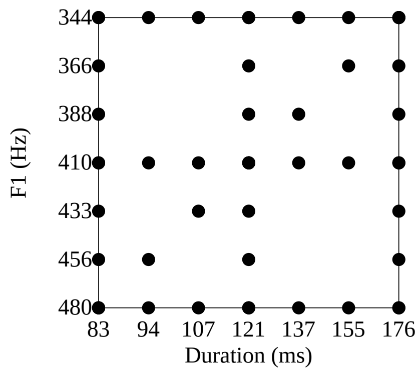

Fig. 1.7. The 37 isolated, synthetic vowels presented to SE and SBE listeners.

The 41 subjects were asked to press either of two buttons, one representing /i/ and the other /ɪ/ depending on the vowel that they thought they heard. Figure 1.8 shows the mean category boundary line for the two types of listeners who can be seen to exhibit dialect dependent vowel categorization.

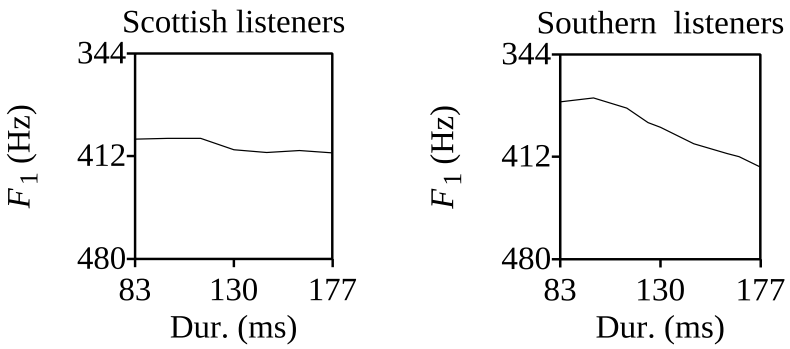

Fig. 1.8. Perceptual boundaries in the average vowel categorization of SE (left) and SBE (right) listeners.

The average SE category boundary line is almost horizontal, which means that these listeners mainly used F1 differences to classify the stimuli. In contrast, the SBE average category boundary line is diagonal, which means that these listeners used both F1 and duration differences to categorize the stimuli. Also, the individual results show that the majority of SE listeners (16 of 20) had a completely horizontal boundary while the majority of the SBE listeners (15 of 20) had a diagonal boundary. A one-tailed two-sample Kolmogorov-Smirnov test conducted on the individual use of F1 and duration differences confirms that the SE and SBE perception of /i/ and /ɪ/ are reliably different ($p < .003$). That is, the categorization of the same synthetic tokens is different for listeners that have been exposed to two different varieties of English. Therefore, it can be concluded that the integration of multiple auditory dimensions in sound categorization is not only language-specific but also specific to the variety of the language to which the listener has been exposed.

In sum, the evidence put forth in this section shows that human listeners have two different ways of hearing the acoustic properties of environmental input, namely as auditory stimuli or as speech stimuli. When receiving auditory input, the listener's general auditory capabilities handle the perception task in a way that is common to all human beings. In contrast, when hearing speech, her speech perception system, learned and shaped with language exposure, handles the perception task in ways that are only appropriate for her specific language environment. In other words, adult listeners have acquired systematic ways of listening to their native language, and these should be represented somewhere in their minds given the repeated task of having to map speech input onto abstract phonological structures. Thus, the attested language-specific perceptual mapping can be considered part of the linguistic knowledge that underlies the decoding of the continuous and variable speech signal into sound categories. In the next section, I discuss some of the proposals that have taken into account the language-specificity of perceptual mappings to model speech perception as language-specific knowledge.

1.3 Modelling speech perception as a language-specific phenomenon

In this section, I show how three different disciplines, viz., phonology, phonetics, and psycholinguistics, have modelled the perceptual mapping of the speech signal as a language-specific phenomenon. Crucially, phonetic and psycholinguistic proposals have long taken into account the evidence shown in § 1.2 in assuming that listeners' perception varies according to the specific language environment. On the other hand, phonological proposals that take into account that speech perception is a linguistic mapping, i.e., a learned language-specific phenomenon, have emerged only recently. Thus, apart from the very early work of Polivanov (1931), it was not until the late 1990s that phonologists started to acknowledge the linguistic nature of speech perception. Below, I review how phonetic, psycholinguistic, and phonological proposals model language-specific sound perception.

This review provides a description of the means by which the different disciplines have modelled perceptual mappings as well as the assumptions they have made with respect to the targets of sound perception, i.e., the discrete (and likely abstract) categories that constitute the targets or units of the perceptual mapping of the signal. Importantly, phonetics, psycholinguistics, and phonology agree that sound categories need to have some level of abstraction though the specifics are still a matter of debate.

1.3.1 Language-specific perception within phonetics

Research embedded in phonetics aims at describing the precise nature of the acoustic dimensions found in the speech signal as well as their physiological and auditory correlates. For instance, we know through phonetics that the acoustic correlate of the phonological feature of vowel height is the vowel's first formant frequency (F1) measured in the physical scale of Hertz. This physical property can also be expressed in perceptual terms with an auditory scale such as Mels or Barks. Phonetics has also demonstrated that the speech signal is continuous and that it contains great variability due to within- and between-speaker production differences.

Phonetic research has also shown that there can be a one-to-one, a many-to-one, or a one-to-many relationship between the acoustic dimensions that constitute a sound and the way those dimensions are used to classify speech sounds. For instance, vowel duration has a one-to-many relationship within English sound segments because it is used to identify both vowels and consonants. Crucially, the tests on auditory versus language-specific perception shown in § 1.2 have been conducted within phonetics. Here, speech perception is modelled as the phonetic mapping of the signal onto phonetic categories. This is illustrated in (1.6) which differs from the formulation in (1.2) in that the perceptual mapping is considered to be language-specific (as depicted by the single arrow).

(1.6) Phonetic model for the nature and elements of speech perception
Acoustic signal → phonetic categories

Most phonetic proposals implicitly model speech perception as a language-specific phenomenon that consists of perceptual mappings and phonetic categories that evince a certain level of abstraction from the signal. Johnson & Mullenix (1997) note that abstract representations such as prototypes, which are either described as articulatory or auditory abstract entities, require a complex mapping from the signal. Those mappings have normally been modelled with simulated neural networks in which auditory neural maps are tuned through a proposed sensitivity to the acoustic dimension of speech (cf. Guenther & Gjaja 1996). However, Diehl et al. (2001) argue that these kinds of neuro-phonetic mappings are not needed if a category can be defined by the boundaries that separate it from other categories. This can be done if category boundaries are simple in form, so that both mental representations and stimulus mapping can be described in theoretically simple terms. These authors refer to the model proposed by Ashby & Gott (1988) and Ashby & Maddox (1998) in which categorization depends on the distance from a decision boundary separating the competing categories in the perceptual space. Nevertheless, it would still be of interest to see if a simple phonetic mapping could be proposed, one that could output abstract categories and perceptual boundaries.

On the other hand, the findings of Pisoni et al. (1994) seem to suggest that the mental representation of a sound includes a large sample of instances of such a sound rather than an abstract representation, thus suggesting that a mapping procedure from the linguistic input may be trivial depending on how numerous and representative the stored exemplars are. However, Kingston's (2003a) cross-language findings, for instance, show that some level of abstraction is evidenced in speech perception so that mappings from raw input onto abstract categories are, in fact, needed. In § 1.4.2, I offer a proposal for the perceptual mappings involved in sound perception which assumes that the targets of the process (i.e., phonetic categories) have some level of abstraction to allow for economical storage and the perceptual integration of multiple acoustic dimensions.

In contrast to phonological proposals where stored representations are considered to be abstract (cf. § 1.3.3), symbolic, and distinctive, phonetic categories are discrete though phonetically detailed. Importantly, speech perception theories that are embedded in phonetics such as the Motor Theory (cf. Liberman & Mattingley 1985) and the Direct Realist Theory (cf. Fowler 1986, 1989, and Best 1995) claim that listeners perceive either articulatory gestures or the neural commands underlying such gestures. Alternatively, auditory features can be viewed as the result of listeners' perception given that Diehl et al. (2001) have demonstrated that the primary objects of speech perception are auditory events. That is, there is controversy as to the exact nature of the phonetic categories that result from the processing of the speech signal. However, describing phonetic categories as either articulatory gestures or auditory events comprises a certain level of abstraction (cf. Strange 1995: 8) and so does the assumption that *prototypes* of sounds are stored (cf. Kuhl 1991). Therefore, these proposals should also account for the way in which the continuous speech signal is mapped onto such abstract perceptual entities. In § 1.4.1, I offer a solution to the problems surrounding both the level of abstraction for the units of perception and the nature of perceptual mappings.

1.3.2 Language-specific perception within psycholinguistics

Psycholinguistics aims at describing and explaining the online processing (or perceptual mapping) of speech sounds. Most psycholinguistic models distinguish between the mapping from the acoustic signal onto phonemic categories (called speech perception) and the mapping performed for lexical access (called speech recognition). Several psycholinguistic studies have shown that listeners process the signal through an intermediate pre-lexical level that mediates between the raw acoustic information in the signal and the words in the lexicon (cf. Miller & Dexter 1988, Schacter & Church 1992, and Pitt & McQueen 1998). Thus, many psycholinguistic models, an overview of which can be found in McQueen (2004), assume that fine-grained acoustic-phonetic information is first analysed into abstract pre-lexical categories rather than being directly mapped onto the lexicon. Figure 1.9 illustrates a psycholinguistic model for word comprehension with two levels of representation and two processes or mappings.

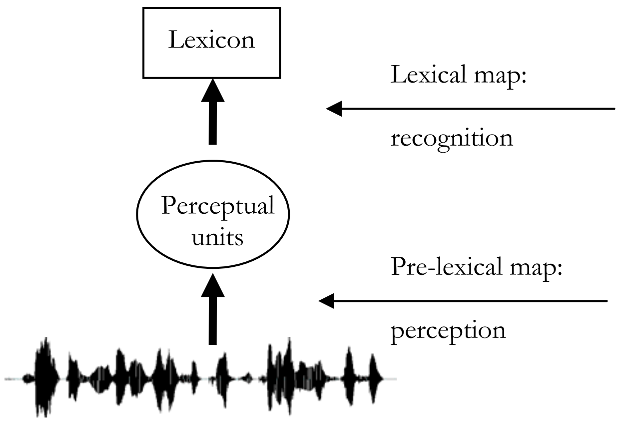

Fig. 1.9. Illustration of a psycholinguistic model with pre-lexical processing.

The main body of evidence in favour of the pre-lexical decoding of the speech signal is derived from the listener's compensatory effects that result from the processing of co-articulation and speech rate, which have been shown to have a pre-lexical locus (cf. Pitt & McQueen 1998, Miller & Dexter 1988). Likewise, the normalization of between-speaker variation has been shown to occur through abstract pre-lexical processing and representations (cf. Schacter & Church 1992). Pre-lexical processing is assumed to be language-specific, just as it is in phonetic modelling.

Importantly, psycholinguistic research has also shown that the online pre-lexical mapping of the speech signal occurs without access to meaning, i.e., without the intervention of the lexicon (cf. Miller & Dexter 1988, Pitt & McQueen 1998, Dupoux, Pallier, Kakehi & Mehler 2001). This is because speech perception does not seem to benefit from lexical access, and several pieces of evidence support the dissociation of lexical biases and low-level speech processing (cf. Miller & Dexter 1988 and Pitt & McQueen 1998). Also, Burki-Cohen, Miller & Eimas (2001) showed that only when acoustic information is degraded do listeners rely on lexical information during speech perception. Crucially, Dupoux et al. (2001) demonstrated that Japanese listeners perceive sound sequences irrespective of their lexical knowledge. They showed that vowel epenthesis took place pre-lexically when Japanese listeners perceived non-words containing impermissible sound sequences. For instance, the Japanese lexicon would lead to the insertion of /u/ for non-words such as /sokdo/ but the insertion of /a/ in words like /mikdo/ because *sokudo* and *mikado* are Japanese words. However, the Japanese listeners tested in Dupoux et al. (2001) reported hearing /u/ in both cases, which shows that the listeners decoded the speech signal using phonological knowledge that is not affected by lexical candidates.

Given the weight of the evidence, it is not surprising that eight out of the ten psycholinguistic models recently reviewed in McQueen (2004) suggest the pre-lexical and bottom-up (as depicted by the direction of the arrows in Figure 10) processing of the speech signal. That is, they propose that rather than being directly mapped onto the lexicon, the fine-grained acoustic-phonetic information contained in the speech signal is first analysed into abstract pre-lexical categories prior to lexical access. Perhaps most importantly for the goal of the present study, Norris, McQueen & Cutler's (2000) Merge model, which specifically addresses sound perception, proposes that this perception is not affected by lexical feedback during online speech processing.

Several psycholinguistic models (e.g., Norris et al. 2000) also propose that, at the time a phonemic categorization response is given, lexical and perceptual information can merge. In addition, the lexicon can influence perception during offline perceptual learning (cf. McQueen & Cutler 1999 and § 2.1.3). It is important to mention that the Merge model assumes that perceptual units or representations are abstract symbolic segments i.e., phonemes. Figure 1.10 shows a representation of the Merge model that is similar to that in Figure 1.9 but includes lexical intervention during offline perception, i.e., when giving a categorization response or during perceptual learning.

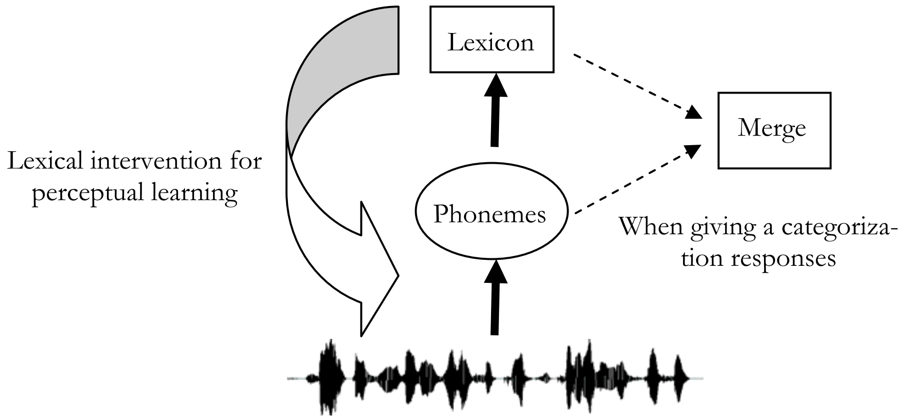

Fig. 1.10. Representation of the Merge model: Pre-lexical bottom-up processing (solid arrows), off-line lexical intervention (white and grey arrow), merge (dotted arrows).

With respect to the formalization of psycholinguistic modelling, most proposals are computationally implemented so that they incorporate processing mechanisms that can simulate pre-lexical speech perception. Although it is assumed that speech perception is language-specific, the processing of the speech signal is modelled by means of general neural networks that map the fine-grained acoustic detail onto perceptual representations. However, such neural networks are trained on language-specific stimuli and therefore result in language-specific processing. In the next section, I argue that linguistic modelling is also able to adequately describe and explain language-specific perception. To that end, I present a proposal that posits a combination of phonetic and phonological components for language-specific perceptual mappings.

Although the existence of a pre-lexical level of perception is a notion that is agreed upon, there is no consensus within psycholinguistic modelling as to the specific nature and level of abstractness of such units because empirical evidence appears to allow for several possibilities. Among those possibilities are syllables (cf. Massaro 1987 and Mehler 1981), phonemes (cf. Neary 2001 and Norris 1994), context-sensitive allophones (cf. Wilckelgren 1969 and Luce et al. 2000), articulatory gestures (cf. Liberman & Mattingley 1985), and acoustic-phonetic features (cf. Lahiri & Marslen-Wilson 1991 and Stevens 2002). Nevertheless, McQueen & Cutler (1997: 570) argue that the nature of the pre-lexical representations is constrained by the fact that they need to share the same vocabulary with lexical representations, i.e., they need to be 'commensurable' enough so that the mapping between lexical and perceptual representations can be easily achieved. Consequently, perceptual representations should at least be abstract and discrete. Crucially, Pisoni & Luce (1987) have shown that listeners can use different 'units of perception' depending on the demands of the listening situation, which is something that should be taken into account in the modelling of pre-lexical speech perception. In § 1.4.1, I propose a solution to the debate concerning the units of perception, i.e., perceptual units within psycholinguistic proposals.

1.3.3 Language-specific perception within phonology

As discussed in previous sections, most phonological proposals model perceptual mappings as universal or extralinguistic. With respect to sound representations, phonological models consider the representation of sounds as a phonological structure which is 'discrete' and 'highly abstract' because it has no relation to the acoustic-phonetic properties of the signal. Also, phonological categories are considered 'distinctive' because they exist only if they convey a difference in meaning, i.e., if they form minimal pairs. Thus, phonological theory views the representation of a sound as a contrastive unit because it contains only features that distinguish it from the representations of other sounds. This means that phonological categories do not contain other non-contrastive acoustic-phonetic properties with which sounds are produced.

Going back to perceptual mappings, another option within phonological modelling would be to assume that speech perception is a single linguistic mapping between the acoustic signal and abstract phonological representations. This linguistic mapping proposal contrasts with the single auditory mapping of most phonological proposals that constitute the mainstream phonological view on speech perception, as discussed in § 1.1. It also contrasts with the two-mapping proposal that refers to an auditory mapping followed by a linguistic one, as proposed by Brown (1998). Table 2 shows the three proposals for the nature of perceptual mappings within phonological theory.

Table 1.2. Three different phonological proposals for modelling speech perception.

Modelling the nature of speech perception		
Single auditory mapping	Two-way mapping	Single linguistic mapping
Auditory ⇒ SF	Auditory ⇒ Universal F → SF	Auditory → SF

Recall that the single auditory mapping and the two-way mapping proposals were formulated in (1.2) and (1.3) respectively (cf. § 1.1), and that these different ways of modelling the nature of speech perception were ascribed to specific authors. In the table, the single linguistic mapping option not only refers to speech perception as a language-specific phenomenon, just like in phonetic and psycholinguistic modelling, but it also incorporates a processing or mapping phenomenon into the domain of phonology, i.e., it interprets it as linguistic knowledge. The question now is whether phonologists have explicitly modelled speech perception as linguistic knowledge. Though such proposals are uncommon, they do exist. The earliest such proposal can be found in Polivanov (1931/1964) where the Japanese perception of drama as /dorama/ or /dzurama/ is explained by word formation rules which eliminate forms like /drama/ and /durama/ because they go against the way in which the Japanese language forms words. In addition, Polivanov proposed that rules that preserve the perceptual identity of the input interact with word formation rules because the former tell the listener that the two well-formed candidates /dzurama/ and /dorama/ differ from the input.

Despite being so close to an explicit and adequate model for describing speech perception through phonological means, Polivanov's proposal was not supported by rigorously collected empirical evidence. That is to say, the proposal was mainly based on anecdotal evidence. For one reason or another, phonological modelling did not take into account the language specificity of speech perception for many years, either because the evidence was not available or because it was interpreted as being insufficient to lead to the phonological modelling of a phenomenon that had been considered peripheral.

Only in the late 1990's did phonological proposals address the linguistic modelling of speech perception. For instance, Tesar & Smolensky (1998, 2000) introduced a linguistic grammar mapping for the perception of syllable structure. They called the process by which syllables are mapped onto feet robust interpretative parsing, a phenomenon that can be equated to the perception of syllable structure because both procedures refer to the perceptual mapping of the speech signal, as was suggested in Boersma & Levelt (2003). One model that seeks to explain language-dependent sound perception is that of Boersma (1998) who proposes that a linguistic (or phonological) perception grammar handles the mapping of the acoustic signal onto discrete segmental units. There are also other phonological proposals that have modelled speech perception phenomena through linguistic means in that they have assumed that speech perception constitutes linguistic knowledge. Among these, we find Kenstowicz (2001) and Broselow (2003) who model loanword adaptation with linguistic perception grammars, and Pater (2004) who argues that a linguistic grammar can explain the knowledge underlying the perception and first language acquisition of phonotactics.

Crucially, Boersma's (1998) phonological proposal is the only one that can handle language-specific (or language-dependent) sound perception, i.e., the mapping of the acoustic signal onto phonological segmental representations. That is, Boersma's work constitutes the most promising framework for describing and explaining sound perception because it seeks to incorporate phonetic detail. His proposal of auditory-to-auditory mapping constraints in his perception grammar can provide the linguistic mechanism that underlies the systematic and language-specific processing of the acoustic-phonetic properties in the input, as will be argued in § 1.4.3. In the next section, I will anlyse the properties of the elements of language-specific perception that should be taken into account for the comprehensive linguistic modelling of sound perception.

1.4 Summary and implications

As previously noted, the present study aims at providing a comprehensive linguistic model for L2 sound perception. The choice for this type of model is based on the language specificity of the perceptual mapping of the speech signal which renders this phenomenon a subject matter of linguistic modelling. In §1.3.3, it was shown that attempts at modelling speech perception through linguistic means have been made and that it is possible to provide a phonological account of so-called phonetic phenomena, such as the production and perception of the sounds of a language.

In § 1.3, we saw that the nature of perceptual mappings and sound representations differs between phonological and phonetic modelling because while several phonological proposals regard it as universal, most phonetic proposals assume their language-specific nature. According to most phonologists and phoneticians, the study of sound segments within each of these disciplines refers to different phenomena so that these disciplines constitute different but complementary subjects of study. However, the nature of perceptual mappings suggests that phonetics and phonology may describe a single phenomenon because universal speech perception is a highly unlikely concept, as was shown in Cho & Ladefoged (1999) for speech production, and because of the issues discussed in § 1.2. Therefore, sound perception can be viewed as a single perceptual mapping from the acoustic signal onto abstract representations that constitute the phonological structure of a given language.

With respect to the nature of abstract representations and perceptual mappings, it seems that phonetic, phonological, and psycholinguistic models do not fully concur on the precise level of abstraction that phonological categories have. Although most models typically make use of phonemic-like representations when modelling sound perception, other less abstract categories have also been proposed. In the next section, I summarize the proposed possibilities for sound representations and provide an attempt to resolve the nature of the targets of speech perception. In § 1.4.2, I discuss the perceptual mapping properties that an adequate model of sound perception needs to consider, given the empirical evidence provided by phonetic and psycholinguistic research. Finally, in § 1.4.3, I list the necessary criteria for a unified account of language-specific sound perception that is embedded in linguistic theory.

1.4.1 Resolving the nature of sound representation

The abstract representations onto which the acoustic signal should be mapped, or their level of abstraction from the signal, seem to be crucial in defining the perceptual mappings needed for language-specific sound perception. As we saw in the previous section, this issue remains unresolved because many types of representations have been proposed and empirically evidenced. These different proposals for describing the representational knowledge underlying the sounds of languages can be organized in terms of their degree of abstraction from the acoustic-phonetic information. Figure 1.11 shows a possible way to organize them from least abstract (right) to most abstract (left) while including the traditional distinction between phonetic categories and phonological segments.

1	2	3	4	5	6
acoustic features	articulatory features	phonetic prototypes	context allophones	phonological features	pho-nemes
		Phonetic categories		Phonological categories	

Fig. 1.11. Proposed sound representations, their degree of abstraction from the signal, and their traditional classification.

I wish to propose that an adequate description of how the sounds of a language are represented should be connected to an explanation of the way in which those representations are extracted from the acoustic-phonetic properties of the input. This is because sound representation is a consequence of language-specific perceptual mappings which have developed through language experience. This proposal also seems to underlie Boersma's explicit definition of speech perception as "the construction of a discrete phonological structure from raw acoustic material" (2000: 10) as well as Pater's assumption that representations are "constructed, and/or accessed, on the basis of the acoustic signal" (2004: 219). Hence, there currently seems to be a convergence regarding the idea that sound categories are the result of the decoding of the speech signal. Thus, the uncertainties about the target or unit of sound perception could be solved if the level of abstraction for perceptual categories turned out to depend on the combinations of auditory dimensions and their frequency distributions in the production environment.

Under the assumption that speech perception is learned through language experience, the number and nature of distinctive sound categories, i.e., the sound inventory of a language, can be seen to result from the mapping mechanism that has been developed to accurately classify the speech signal produced in a given language environment. Within such a developmental approach, perceptual mappings are considered to underlie listeners' perceptual behaviour, which leads to observed phenomena within sound categorization such as perceptual boundaries, prototypical members of a category, and sound inventories. Thus, adult listeners have perceptual mappings that allow them to categorize acoustic-phonetic information as language-specific sounds and so to perceive the appropriate number of sound contrasts. Likewise, adult listeners have appropriate perceptual boundary locations and they exhibit prototypical effects which result from the learned linguistic perception of their environmental speech signal.

1.4.2 How to model linguistic perceptual mappings

If categories have some level of abstraction and if that level of abstraction depends on the extraction of language-specific linguistic properties from the signal, perceptual mappings must mediate between the continuous and variable acoustic signal and sound representations. What sort of perceptual mappings could provide such mediation? Given the auditory properties of the speech signal, perceptual mappings should be able to process a variety of auditory values that are shared between production environments but that have different distributions and are used in distinctive ways in these different environments. That is, perceptual mappings across languages may behave similarly in that they need to process the same auditory dimensions given the common properties of the speech signal across languages. However, they also convey the specific ways in which the sounds of a particular language should be optimally perceived. The question, then, is how we can model such a universal and language-specific interaction of the mappings involved in sound perception.

Given that the perceptual mapping from acoustics to the abstract representation of sounds is also language-specific, and therefore represents linguistic knowledge, it should undergo phonological modelling, as has been proposed in Boersma (1998), Tesar & Smolensky (2000), Broselow (2003), Kenstowicz (2001), and Pater (2004). Some phonological proposals, such as those offered in Hyman (2001), Hume & Johnson (2001b), and Steriade (2001), and some phonetic proposals, such as the one in Keating (1984), seem to argue in favour of a division of labour between phonetics and phonology when dealing with speech production and perception.

However, both phonetic and phonological facts need to be conveyed to lead to a more adequate model for explaining and describing the knowledge underlying speech perception. This is because the nature of the speech signal requires some kind of phonetic mapping that could also be encoded as phonological knowledge, given the language specificity of perceptual mappings. Kingston argues that because the forces that underlie speech perception are regulatory and evaluative, they are "in the mind and not in the vocal tract or ear", adding that "nothing stands in the way of their incorporation into other mental constructions or operations, such as the grammar of a particular language" (2003b: 288). This means that the linguistic grammar not only helps speech perception, as seems to have been suggested in phonological proposals such as the one found in Chomsky & Halle (1968), but also that perceptual mappings need to be encoded as part of the grammar. In the next section, I summarize the properties of speech perception described in this chapter and outline a proposal for an explanatorily adequate model for the decoding of the speech signal.

1.4.3 Implications for a comprehensive model of sound categorization

Four main properties of the perception of speech sounds have been discussed in this chapter. First, speech perception refers to the decoding of the variable and continuous acoustic properties of the speech signal. Second, it involves abstract representations and perceptual mappings. Third, it is language specific and language dependent, i.e., the decoding of the speech signal is developmentally shaped by a language environment, and therefore it is only appropriate for such an environment. Fourth, it involves phonological representations whose degree of abstraction should depend on the acoustic properties of the signal and the way in which such properties are encoded in the perceptual mappings. Together, these four basic properties have a crucial impact on the way speech perception could most adequately be modelled. Table 1.3 shows a possible model that incorporates these properties.

Table 1.3. Properties of speech perception and proposal to model them.

Speech perception properties	Model proposal
Definition: Decoding of acoustics	Phonetic parsing: Phonetically-grounded mappings
Elements: Abstract categories + mappings	Perceptual mappings connect the signal with the listener's abstract representations
Nature: Shaped by and targeted to a specific language environment	Linguistic knowledge underlies speech perception: Grammatical rules or grammatical constraints
The nature of categories depends on the signal and the mappings	The input generates the mappings and they, in turn, generate sound representations.

As can be seen, this model would integrate phonetic and phonological approaches to speech perception. However, psycholinguistic modelling would be missing. Two of the psycholinguistic constructs discussed in this chapter would seem to be crucial for the modelling of speech perception. First, as psycholinguistic research has shown, listeners map the signal onto pre-lexical abstract representations so that the decoding of the acoustic properties into pre-lexical representations precedes the access to meaning. Second, not only does this pre-lexical level precede lexical access but also it also takes place without the intervention of lexical knowledge. This means that the speech signal is parsed in a bottom-up fashion without feedback from the lexicon which is at a higher level. In other words, there is no top-down processing. Thus, within psycholinguistic modelling, speech perception is viewed as the decoding of the speech signal prior to the access of lexical items. These psycholinguistic postulates need to be integrated into a model that would comprehensively and most adequately explain speech perception.

In the next chapter, I show that Boersma's (1998) phonological proposal constitutes the most promising framework for describing and explaining sound perception because it models all the properties of speech perception listed in Table 1.3 in addition to incorporating psycholinguistic constructs. In particular, his proposal of auditory-to-auditory phonetic constraints, which constitute a perception grammar that performs the categorization of auditory continua, can provide the linguistic mechanism underlying the systematic and language-specific processing of the acoustic-phonetic properties in the input. This grammar is considered to work without the intervention of the lexicon in that it constitutes pre-lexical linguistic knowledge that is used to initially decode the signal. In Chapter 2, it will be demonstrated that this phonological model together with its recent extensions meet the criteria that have been put forward for an explanatorily adequate model of speech perception and, crucially, of the acquisition of this perception.

2 Linguistic Perception (LP): a phonological model of sound perception

This chapter presents the Linguistic Perception (LP) model which is a phonological proposal for explaining speech sound perception. This model is based on the perception component of Functional Phonology (cf. Boersma 1998) and constitutes a unified framework for describing, explaining, and predicting adult sound perception and how first language learners acquire it. It is shown that the LP model most adequately accounts for cross-linguistic variation in sound categorization, i.e., the perceptual differences between listeners with different language backgrounds, as was discussed in § 1.1.2, as well as for the specific developmental patterns attested in the infant and child sound perception literature. Importantly, the model's success lies in its comprehensiveness in that it unifies phonological, phonetic, and psycholinguistic modelling, and incorporates all the criteria that have been put forward in § 1.5.3 to produce an adequate model of sound perception.

There are five sections in this chapter. In § 2.1, I discuss the model that is assumed for explaining speech comprehension, the status of speech perception within such a model, and the nature of the elements that handle the decoding of the speech signal, namely a linguistic grammar and perceptual representations. In § 2.2, a principle that organizes the linguistic grammar, viz., optimal perception, is postulated and illustrated with the linguistic categorization of an auditory dimension as well as with the integration of multiple auditory dimensions in linguistic perception. These two sections combine some of the ideas presented in Boersma (1998) and Escudero & Boersma (2003) to which are added new discussions and explanations.

Given that the L2 model to be proposed in Chapter 3 is based on the general LP model described here, § 2.3 discusses the proposal for the L1 acquisition of optimal perception advanced in Boersma, Escudero & Hayes (2003). This L1 proposal constitutes the perceptual learning view assumed within the LP model which applies to both L1 and L2 acquisition. Next the proposal for the mapping from perceptual representations to words in the lexicon is described in § 2.4. Although the scope of the present study primarily concerns perception, the L1 acquisition model which is assumed here proposes that lexical knowledge is involved in perceptual learning once the lexicon starts to be in place. Therefore, it is important to discuss the nature of lexical representations and the procedure that allows a listener to gain access to them, a procedure which is commonly known as speech recognition among psycholinguists. Finally, the overall theoretical framework for adult sound perception and its acquisition, which underlies the remainder of this study, is summarized in § 2.5.

2.1 The elements of Linguistic Perception (LP)

The LP model, just like the proposals in Boersma (1998), Tesar & Smolensky (2000) and Pater (2004), assumes that linguistic knowledge underlies speech perception because the mapping of the speech signal has a systematic and language-specific nature, as discussed in Chapter 1. The LP model provides an explicit and comprehensive phonological proposal for describing and explaining linguistic sound perception. It is proposed that a language user handles the speech signal by means of a linguistic grammar, which is the mapping component, and perceptual categories, which is the representational component. Figure 2.1 shows the elements of the LP model which are based on the speech comprehension model advanced in Boersma (1998). The perceptual part of the model comprises three elements. Starting from the bottom, we have the auditory signal, the device that decodes this signal, i.e., the perception grammar, and the discrete output of the perceptual mapping, i.e., the perceptual representations.

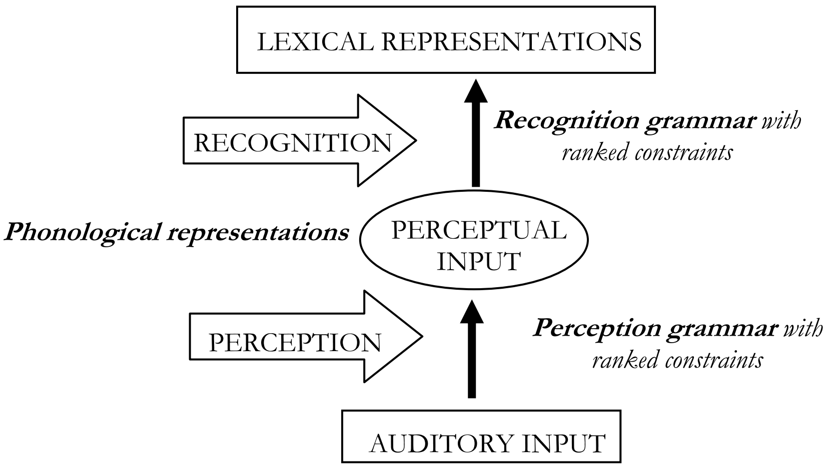

Fig. 2.1. The LP elements for speech comprehension: auditory input, two mappings, and two levels of abstract representations.

Thus, within the LP model, speech perception is considered to mediate between the auditory input and the internal stored representations of words, i.e., the underlying or lexical representations. Perceptual representations act as input to the recognition grammar which further decodes the information provided by perception so that words may be accessed in the lexicon. Importantly, the terms perception grammar and perceptual input in Figure 1 are equivalent to the two psycholinguistic elements involved in the pre-lexical processing of speech, namely a processor and a pre-lexical level of representation (cf. § 1.3.3). In addition, perceptual inputs can be equated with *phonological representations*, and therefore both notions will be used in the present study to refer to the discrete arbitrary symbols that listeners use to represent the sounds of their language.

Also, just as in psycholinguistic modelling, the LP model assumes that perception is a pre-lexical, low-level (albeit linguistic and language-specific), and automatic process that leads to a level of representation that cannot be consciously accessed. In contrast, speech recognition is viewed as a high-level mapping that not only involves the interpretation of semantics and pragmatics but also generates a representational level, viz., lexical items, which can be accessed consciously. Crucially, the LP model assumes that the signal is linguistically analysed in a two-step and bottom-up fashion, as depicted by the direction of the arrows in Figure 1. This entails that the lexicon does not intervene in perceptual mappings. This is also in line with the psycholinguistic models that assume pre-lexical processing without lexical intervention, such as the Race and Merge models discussed in § 1.3.3. With respect to the linguistic modelling of the difference between lexical and pre-lexical mappings in speech comprehension, Boersma (1998) proposed that a separate recognition grammar performs the mapping from perceptual representations to lexical representations. This linguistic proposal for recognition grammars will be discussed in § 2.4.

2.1.1 Perceptual mapping component: the *perception grammar*

The LP model assumes that the language-specific decoding of the speech signal into discrete categories is handled by a *perception grammar*. This linguistic grammar performs the mapping of the signal through constraints that map or connect the acoustic properties of the input with sound representations. Within the LP model, four types of mapping constraints are proposed, and these can be seen as describing four different stages in the development of linguistic perception (cf. § 2.3). These four types of mapping constraints differ in that they map onto different discrete units, i.e., onto perceptual representations with different degrees of abstraction from the speech signal (cf. § 2.1.2).

The first type of mapping constraints was proposed by Boersma (1998). These constraints map values along an auditory continuum onto perceived values along the same auditory continuum, that is to say, they evaluate the continuous auditory input and they output discrete auditory categories. Given the close relation between their input and their output, constraints of this type can be called auditory-to-auditory mapping constraints, the formulation of which is given in (2.1).

(2.1) Auditory-to-auditory mapping constraints

PERCEIVE (*f: x*) constraint family

'Map the value *x* along the continuum *f* to some value along that same continuum'

*CATEG (*f: y*) constraint family

'Do not perceive anything as the value y along the continuum *f*'

*WARP (*f: d*) constraint family

'Don't perceive a value along a continuum *f* as a value that is a distance *d* (or more) away along that same continuum'

Thus, an interaction of PERCEIVE, *CATEG, and *WARP constraints generates the dotted lines in Figure 2.2 which represent an auditory-to-auditory mapping. For instance, the mapping line from the F1 value 900 to the F1 category 750, i.e., the auditory-to-auditory mapping of F1 = 900 onto F1= 750, results from satisfying the PERCEIVE [900] constraint, from violating the *CATEG 750 constraint, and from violating *WARP [150]. In § 2.3, I show how these constraints are ranked in *infant* perception grammars.[9] These auditory-to-phonetic constraints can be called 'one-dimensional' because they map an auditory continuum onto perceived auditory categories along the same single continuum.

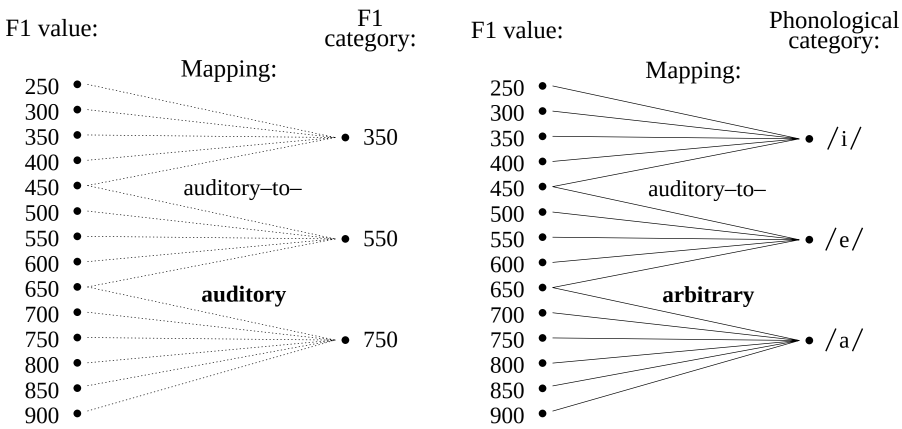

Fig. 2.2. Auditory-to-auditory mapping (cf. Boersma 1998) and auditory-to-arbitrary mappings (cf. Escudero & Boersma 2003).

However, adult listeners have been shown to combine a number of auditory dimensions when perceiving sound categories. In other words, they exhibit perceptual cue integration. This means that the targets of perception need to be abstract enough to allow for the integration of multiple cues and not just for the mapping of one auditory continuum onto a perceived auditory category along that same continuum. Thus, constraints that can map the signal onto abstract phonological categories are required to more adequately describe the contents of an adult perception grammar. These constraints are nowadays called cue constraints (see Boersma 2005). The simplest of them map values along an auditory continuum

[9] Note that *adult* perception grammars contain different constraints which will be described below.

onto an abstract phonological category that refers to a single abstract scale or single phonological feature and were proposed in Boersma, Escudero & Hayes (2003) as part of the phonological development towards adult-like sound perception (cf. § 2.3). Given the nature of their output, these constraints can be called one-dimensional auditory-to-feature constraints. A formulation and examples of this second type of constraints are given in (2.2).

(2.2) One-dimensional auditory-to-feature constraints

'A value *x* on the auditory continuum *y* should not be perceived as the phonological feature /*z*/'

[F1] → /F1 = height/

[Duration] → /Duration = length/

[F2] → /F2 = backness/

/*z*/: An arbitrary label for the phonological feature that refers to an auditory continuum.

As seen in (2.2), then, the abstract category is a phonological feature that is arbitrary in nature because the label 'height' could be replaced by any other label such as 'uh'. Examples of these cue constraints are 'an F1 value of 300 Hz is not /height 1/', 'a duration value of 120 ms is not /long/', 'an F2 value of 2500 is not /back/', and so on for every value in every auditory continuum.[10] In this case, the perceptual targets or representations refer to a single abstract scale or feature. However, it is important to bear in mind that, in reality, these constraints are only a step towards adult-like constraints, as proposed in Boersma, Escudero & Hayes (2003) (cf. § 2.3). A type of cue constraints that might be used to adequately describe adult perception is given in (2.3). This third type of constraint results in the perceptual auditory integration that is commonly performed by adult listeners because any auditory input could map onto any phonological feature. This is because these constraints map auditory continua onto more abstract arbitrary categories with no a priori connection with any auditory continua. Given that they relate any auditory continuum to any phonological feature, rather than relating a single auditory con-

[10] Depending on the number of categories produced along the F1 continuum, the feature /height/ could refer to a number of categories. In a language with three vowel heights, the categories could be /height 1/, /height 2/, and /height 3/, or they could be /high/, /mid/, and /low/. Given that the feature label is arbitrary, either notion is correct provided that it refers to a single discrete scale.

tinuum to a single abstract scale, they can be called multidimensional auditory-to-feature constraints.

(2.3) Multidimensional auditory-to-feature constraints

'A value *x* on the auditory continuum *y* should not be perceived as the phonological feature /*z*/'

[F1] → /height/

[F1] → /length/

[F1] → /backness/

[Duration] → /length/

[Duration] → /height/

[Duration] → /backness/

[F2] → /backness/

[F2] → /height/

[F2] → /length/

/*z*/: any phonological feature that refers to any auditory continua.

Although the two previous constraint families lead to phonetic-to-phonological perceptual mappings because they map raw auditory input onto abstract categories, they are still not abstract enough to handle adult sound categorization optimally. Escudero & Boersma (2003) proposed that adult auditory cue integration is performed through cue constraints that refer to highly arbitrary units, such as vowels and consonants, which have no connection with auditory continua. These constraints can also be called multidimensional auditory-to-segment mapping constraints, and their formulation is given in (2.4). The solid connecting lines on the left side of Figure 2 show the perceptual mapping that results from this type of cue constraints.

(2.4) Multidimensional auditory-to-segment constraints

'A value *x* on the auditory continuum *y* should not be perceived as the phonological segment /*z*/'

[F1] → /i/

[F1] → /e/

[F1] → /u/

[Duration] → /i/

[Duration] → /e/

[Duration] → /u/
[F2] → /i/
[F2] → /e/
[F2] → /u/
/*z*/: a completely arbitrary label for a feature combination, e.g., /x/ or /i/, which has no relation with the auditory world.

Consider, for instance, the English vowels /i/ and /ɪ/ which are produced with a combination of at least two acoustic continua, viz., duration and the first formant (F1). In English, speakers produce /i/ with a longer duration and a lower F1 than /ɪ/. This means that English listeners have a perception grammar with constraints that integrate duration and F1 values for the categorization of vowel segments such as /i/ and /ɪ/. Such an integration can be performed by relating both F1 and duration values to the same phonological segment. Thus, an adult English perception grammar contains phonological cue constraints such as 'an F1 of 300 Hz is not /i/' or 'a duration of 120 ms is not /i/', as well as similar constraints that map F1 and duration values onto /ɪ/.[11]

It is important to mention that by assuming a linguistic grammar for sound perception, the LP model synthesizes phonological, psycholinguistic, and phonetic modelling proposals into a comprehensive explanation of the perceptual mapping that is involved in sound perception. Phonological modelling is incorporated because language-specific perception is modelled as a formal linguistic grammar. Psycholinguistic modelling is incorporated because the perception grammar performs the online processing of the speech signal. And phonetic modelling is included because the perception grammar contains constraints that refer to auditory-

---

[11] The reader may wonder why using segmental units, such as /i/, should be preferred over a combination of features, such as /high, front/, for describing the constraints in an adult perception grammar. My work in collaboration with Paul Boersma has addressed that question. For instance, the modelling of the perception of the Dutch vowel system in Boersma and Escudero (2004) would not have been possible with constraints that map F1 to /high/ and F1 to /front/, separately. In Boersma, Escudero & Hayes (2003), it is shown that learners may need extra structural constraints, e.g., */high, short/ in the SBE case, against feature combinations that do not occur in their language, but this would complicate the model considerably and it is not clear whether it would work anyway. Adult perceptual cue integration *would* be possible with constraints that map F1 to feature combinations such as 'an F1 of 300 Hz is not /high, front/'. However, in this case, the feature combination would act as a single higher order structure which could just as well be abbreviated to /i/.

phonetic properties. These constraints can decode the continuous and variable speech signal into segmental units.

To sum up, at least four types of constraints for the perceptual mapping of the signal onto discrete representations have been proposed. From these constraints, the ones called cue constraints can map the signal onto vowels and consonants. These constraints have been regarded as the optimal constraints in an adult perception grammar, and they are the ones responsible for the integration of multiple auditory dimensions in phonological sound perception. In § 2.3, I discuss how the three other types of constraints, namely auditory-phonetic, one-dimensional-to-feature, and multidimensional-to-feature, emerge during L1 perceptual development. The next section shows how the LP's representational proposal resolves the issues regarding the 'unit of perception' discussed in § 1.4.1.

2.1.2 Representational component: the *perceptual input*

The decoding of the speech signal leads to the construction of phonological representations which in Functional Phonology (cf. Boersma 1998) are referred to as *perceptual input*, a term that suggests a direct connection with word recognition. This representational component of the LP model deals with the two main issues related to sound representation, viz., its status with respect to the mapping component and its degree of abstraction from the signal. The perceptual representations are posited to be the product or result of the ranking of the perceptual mapping constraints. This means that their status depends on the constraints that perform the mapping of the signal onto these representations. Given the four types of constraints described in the previous section, it is proposed that the signal can be mapped onto four different kinds of perceptual representations. Table 2.1 shows the specific representation that results from each constraint type.

Table 2.1. Mapping constraints, resulting representations, and their properties.

Mapping Constraint	Resulting representation	Discrete	Abstraction	Complexity	Relation with signal
Auditory-to-auditory	Phonetic	Yes	None	one-dimensional	Strong
1-dimensional auditory-to-phonological	Phonological feature	Yes	Small: auditory relation	one-dimensional	Mild
Multi-dimensional auditory-to-phonological	Phonological feature	Yes	Large: completely arbitrary	multi-dimensional	None
Auditory-to-arbitrary phonological	Segment	Yes	Very large: completely arbitrary, and higher-order	multi-dimensional	None

As shown above, perceptual representations can potentially have a diverse nature and degree of abstraction depending on the acoustic properties of the auditory events in the language environment at hand, with such properties being conveyed in the constraints. In § 2.1.1, we saw that four types of constraints can be found in the perception grammar and three different kinds of representations can result from them. This is because perceptual representations are the products of the mapping performed by the constraints in the perception grammar. The three different types of representations can share the nature of at least three 'units of perception' proposed in the literature, namely auditory discrete events as in exemplar-based proposals (cf. Pisoni et al. 1994), abstract phonetic categories or phonological features, and segments or phonemes. If the speech signal produced in a particular language frequently contains combinations of diverse acoustic-phonetic continua, the perception grammar will map the speech signal onto representations that refer to those combinations (cf. Boersma 1999), such as in the fourth type of representations in Table 1. However, if a sound category is produced with a single dimension, then the representation of such a sound will likely be a phonological feature such as the second type of representation in Table 1.

Importantly, adult perception grammars map the acoustic signal onto abstract representations for purposes of economical lexical storage (cf. § 2.3). Abstract categories also allow for the integration of acoustic dimensions because once categories are abstract, *any* acoustic dimension could be mapped onto them. This leads to phonological cue constraints which output representations that have *no* connection with the acoustic properties of the signal so that they can be considered arbitrary. For instance, in many languages vowel height (F1) frequently combines with vowel duration in the production of vowel segments such as /i/. Examples of these arbitrary cue constraints, also called auditory-to-phonological constraints, are '500 Hz is not /i/' or '120 ms is not /i/'. Thus, it is proposed that the perceptual representations that are the output of the adult perception grammars can be highly abstract and arbitrary because the phonetic information needed to map the speech signal is conveyed in the cue constraints. Also, the number of perceived segmental categories that can be output by the adult grammar defines the sound inventory of a language which, in turn, determines the possible abstract linguistic units that can be used to form words. Finally, with respect to the status of perceptual representations within word comprehension (cf. Figure 1), the LP model posits that the output of perception is pre-lexical because it is the result of the pre-lexical decoding of the speech signal. This pre-lexical, perceptual and abstract code is later used as an input to lexical access.

In sum, the main difference between the LP model's sound categories and the representations assumed in other proposals, such as the ones described in Chapter 1, is that the nature and number of categories depends on the perceptual mappings found in the perception grammar. A sound system can therefore be interpreted as the set of segmental representations that the adult perception grammar is able to construct or categorize from the speech signal of a language in which several acoustic dimensions are commonly integrated. As will be discussed in § 2.3, frequent acoustic combinations lead to the learning of mappings which output segments incorporating such combinations. Before going into how perceptual mappings and their respective representational outputs are learned, however, I will first present the LP hypothesis for the ranking of constraints in the perception grammar which will be referred to as the *optimal perception hypothesis*.

2.2 The optimal perception hypothesis

The LP model proposes that the perceptual mapping of the signal depends on the particular characteristics of the listener's production environment. Thus, the optimal way to perceive the sounds of a language is by making categorization decisions that lead to *maximum-likelihood behaviour* (cf. von Helmholtz 1910) which minimizes the possibilities of misunderstanding a speaker (cf. Boersma 1998: 337, 340, 371). It is hypothesized that experience with the particular ways in which sounds are produced results in *optimal perception* whereby listeners learn to categorize the speech signal into the vowels and consonants that match the ones intended by the speaker.

In other words, the constraint rankings in the perception grammar are based on the distributions of the acoustic values with which sound categories are produced, that is to say, on how likely it is that acoustic values were intended as a given sound category. For instance, if an F1 value of 280 Hz is used to produce the vowel /i/ and never used in the production of a different vowel, it means that such an F1 value has a 100% probability of having been intended as the front vowel /i/ in a given environment. Therefore, an optimal listener will always perceive [F1 = 280 Hz] as /i/ in a language having only front vowels. However, the variation in the productions of a sound category result in a differential likelihood or probability percentage for different F1 values. This range of F1 values with different frequencies and different probabilities of intention constitutes the production distribution of a sound in a given production environment. In Escudero & Boersma (2003), it was argued that if the token distributions of two sound categories are given, their optimal perception can be computed and the perception grammar that underlies such optimal behaviour can be described.

The predicted perception can thus be compared with that of simulated and real listeners in order to test the optimal perception hypothesis. In § 2.2.1, I discuss a case of optimal one-dimensional sound categorization as illustrated by the perception of F1 values as Spanish /i/ or /e/. In § 2.2.2, I discuss a case of optimal cue integration of F1 and vowel duration values in the categorization of Southern British English (SBE) and Scottish English (SE) /i/ and /ɪ/. In both cases, I describe how we can measure the production distributions and how we can predict the optimal perception and optimal constraint rankings. Crucially, it is shown that the hypothesized optimal perception compares well with the cross-linguistic findings reported in § 1.1.2 to the effect that the proposed optimal perception grammars may be the knowledge underlying the listeners' perception.

2.2.1 Optimal one-dimensional categorization

The Spanish vowels /i/ and /e/ differ in the F1 values with which they are produced. To be able to predict how an optimal Spanish listener perceives F1 values as either of these two vowels, we need to establish the F1 values that characterize the productions of each one. In Escudero & Boersma (2003), it was assumed that the production distributions of sound categories can be calculated if the variations of the average productions are known. For instance, Cervera et al. (2001) measured the productions of Spanish /i/ and /e/ of ten male speakers and computed the mean F1 values as well as the standard deviations from the mean. Figure 2.3 shows the measured mean values of Spanish /i/ and /e/, along a base-10 logarithmic F1 continuum with a Gaussian token distribution with a standard deviation of 0.166 octaves.[12]

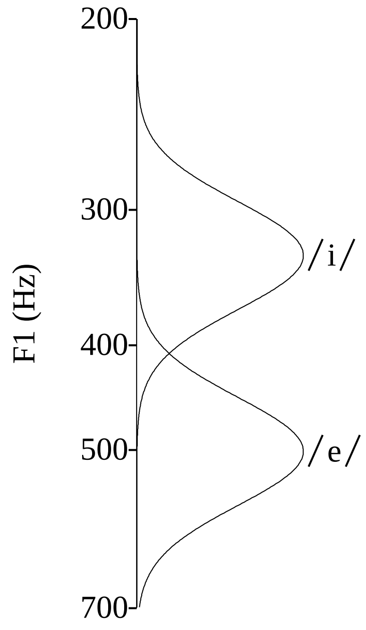

Fig. 2.3. Production averages and distributions of Spanish /i/ and /e/.

[12] It is also assumed that both vowels are equally frequent so that the peaks in the distributions are at the same height. The extent of the peak in the distribution along the horizontal axis shows the probability that a value has been intended as a particular vowel.

We can compute three different values for this production environment which will allow us to predict the optimal Spanish perception. First, we can compute the F1 distance between the two average values for the two vowels and express them in standard deviations. Given that the auditory correlates of acoustic dimensions are best described with logarithmic-like scales, as shown in the psychoacoustic literature where the mels or barks scales are used to describe auditory phenomena, the distances between sound categories can be measured using a logarithmic scale such as *octaves*. Thus, the F1 distance between the average production of Spanish /i/ and /e/ is log2(502) – log2(331) = 0.6 octaves. In Figure 2.4, the F1 distance is represented as the connecting dashed line drawn from /i/ to /e/.

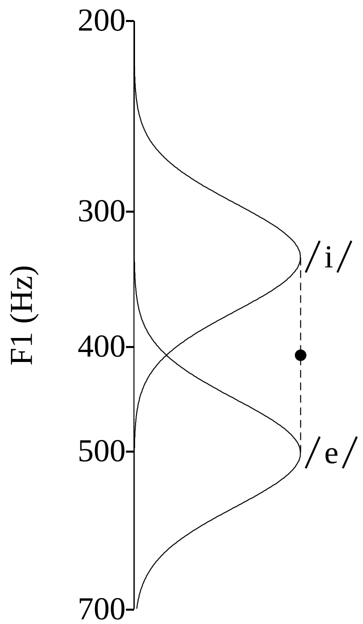

Fig. 2.4. Distance and midpoint between the F1 values of Spanish /i/ and /e/.

Secondly, from this distance we can compute the F1 value that has the same probabilities of having been intended as either of the two vowels, i.e., 50% of the time as /i/ and 50% of the time as /e/. This value can be referred to as the *equal-likelihood production*. This equal-likelihood value is represented as the crossing point of the two distribution curves. Given that these vowel distributions have equal standard deviations of 0.166 octaves and that the vowels are assumed to be equally frequent, the crossing point of their distributional curves is the middle point between their F1 values. Thus, the equal-likelihood point or crossing point is located at 0.6 / 2 = 0.3 octaves from the mean which is 407 Hz or log2(502) - 0.3 octaves converted back to Hz = $2^{8.67}$= 407 Hz. Finally, we can compute the percentage of tokens that are produced in the overlapping area delimited by the crossing point in the distributions. If we draw 1000 tokens from each of the distributions of the two vowels, for example, we can compute the number of tokens that fall in the overlapping production area first as 0.3/0.166 = 1.81 standard deviations, and then as gaussQ(1.81) = 0.035 = 3.5 %.

With these three computations for the distributions of F1 values in the production of Spanish /i/ and /e/, the optimal perception can be predicted. That is, an optimal listener will have a perception that matches the production distributions described above, i.e., she will follow a maximum-likelihood strategy. Such a Spanish maximum-likelihood listener has a perceptual boundary that coincides with the equal-likelihood point in production which is located at 407 Hz, and this means that she will perceive every token above the boundary as /i/ and every token below the boundary as /e/. This listener will be correct in 96.5% of the time when categorizing F1 values into Spanish /i/ and /e/, which is the percentage of tokens that are not produced with overlapping values (i.e., 100%-3.5%). Therefore, this optimal Spanish listener will misperceive only 70 tokens when categorizing 1000 tokens of Spanish /i/ and 1000 tokens of Spanish /e/ that are drawn from the distributions depicted in Figures 3 and 4. We can represent the optimal perceptual boundary in the same acoustic space shown for production by drawing a line located at the F1 value of the crossing or equal-likelihood point, as illustrated in Figure 2.5.

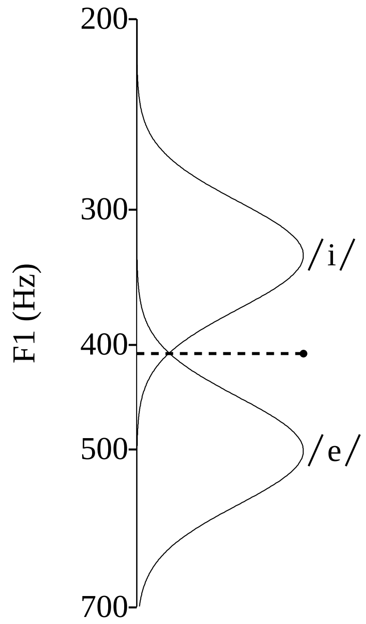

Fig. 2.5. Optimal perceptual category boundary between Spanish /i/ and /e/.

Thus, an optimal Spanish listener perceives every F1 value below the perceptual boundary as /i/ and everything above it as /e/ because values below such a boundary are most likely to be /i/ and values above it are most likely to be /e/ in the given production environment. The LP model thus proposes that an optimal perception grammar underlies the optimal perception of F1 values in Spanish. Such an optimal grammar can be represented as the continuous rankings of constraints against perceiving F1 values as /i/ or /e/, as shown in Figure 2.6. For instance, the constraints against perceiving F1 values that fall above the crossing point as /i/ are ranked lower than the constraints against perceiving those same values as /e/. Therefore, an optimal Spanish listener will perceive those values as /i/.

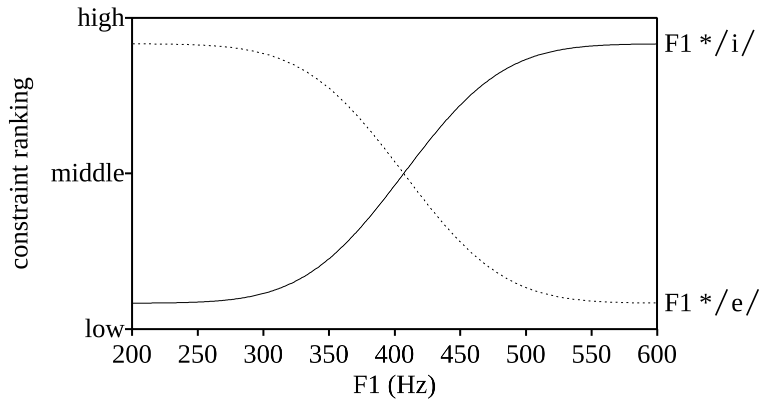

Fig. 2.6. Continuous constraint rankings of a Spanish perception grammar that leads to the optimal categorization of F1 values into /i/ and /e/.

With this optimal ranking in the Spanish perception grammar, we can also explain the cross-language categorization of English tokens by monolingual Spanish listeners, which was reported in § 1.2.1, and which is based on their L1 perception. In the formalization of the optimal perception grammar, the decision scheme works according to the constraint-based framework of Optimality Theory (OT) (cf. Prince & Smolensky 1993), and more specifically, its probabilistic version which is called Stochastic OT (cf. Boersma 1998).

The two OT tableaux below give the rankings that can be read off Figure 2.6 for the categorization of the F1 values 337 and 485 Hz respectively. The top-left cell in the tableaux shows the F1 value that enters the perception grammar, i.e., the input to the grammar. The constraints and their rankings are shown in the rest of the first row. In these tableaux, only the constraints that refer to the average English productions are shown, but the continuous constraint family includes constraints against perceiving any F1 value as either /i/ or /e/. In the Spanish perception grammar, the highest ranked constraint, i.e., the leftmost constraint, is 'do not perceive 337 Hz as /e/' because of the large difference between 337 Hz and the mean production of /e/ which is 502 Hz. In Tableau 2.1, the high ranking of this constraint bans the perception of 337 Hz as /e/ and so the winner is /i/. The second highest constraint is 'do not perceive 485 Hz as /i/' because of the large difference between 485 Hz and the mean production for /i/ which is 332 Hz. Tableau 1.2 shows that this constraint ranking bans the perception of 485 Hz as /i/ so that this F1 value is categorized as /e/.

[F1= 337 Hz]	337 Hz not /e/	485 Hz not /i/	485 Hz not /e/	337 Hz not /i/
☞ /i/				*
/e/	*!			

Tableau 2.1. A typical SBE /ɪ/ token categorized by the optimal Spanish grammar.

[F1= 485 Hz]	337 Hz not /e/	485 Hz not /i/	485 Hz not /e/	337 Hz not /i/
/i/		*!		
☞ /e/			*	

Tableau 2.2. A typical SE /ɪ/ token categorized by the optimal Spanish grammar.

From this, it is safe to conclude that an optimal Spanish perception grammar can account for the knowledge that underlies the differential categorization of SBE and SE /ɪ/ by Spanish listeners, as was reported in § 1.1.2. In the next section, I show a more common but more intricate case of sound categorization which involves the combination of more than one acoustic dimension in the production of sound categories.

2.2.2 Optimal cue integration

Typically, more than one acoustic dimension is involved in the production of sound categories, and listeners use those multiple sources of information in sound categorization. The production of English /i/ and /ɪ/, for instance, combines F1 differences (/ɪ/ tends to have a lower F1 than /i/) and vowel duration differences (/i/ tends to be longer than /ɪ/). However, the two dimensions are not used in the same way in different English dialects. For instance, it was shown in Escudero & Boersma (2003) that SE speakers pronounce these vowels with almost the same duration but with a very different F1, whereas SBE speakers make a large duration distinction and a smaller F1 distinction between the two vowels. Thus, it is proposed that optimal listeners will perceptually combine auditory cues according to the way they are combined in production.

In order to predict the optimal perception in two different English dialects, we need to accurately compute the use of both cues in each environment, starting with the three measures described in the previous section. Table 2.2 shows the production averages and standard deviations of SE and SBE /i/ and /ɪ/, as reported in Escudero & Boersma (2003).

Table 2.2. F1 and duration average values and standard deviations (s.d.) for SBE and SE vowels.

SBE		Average	s.d.	SE		Average	s.d.
/ɪ/	duration	59.7 ms	0.4 dou.	/ɪ/	Duration	84.8 ms	0.4 dou
	F1	337 Hz	0.2 oct.		F1	485 Hz	0.2 oct.
/i/	duration	104.6 ms	0.4 dou.	/i/	Duration	94.0 ms	0.4 dou.
	F1	292 Hz	0.2 oct.		F1	343 Hz	0.2 oct.

The standard deviations in Table 2.2 were chosen in order to ensure that the environments contain a wide range of F1-duration pairs. With these averages and standard deviations, we can compute the three production values that will allow for a prediction of the optimal perception in each environment. Figure 2.7 shows the average productions in each dialect plot on an F1-duration acoustic plane.

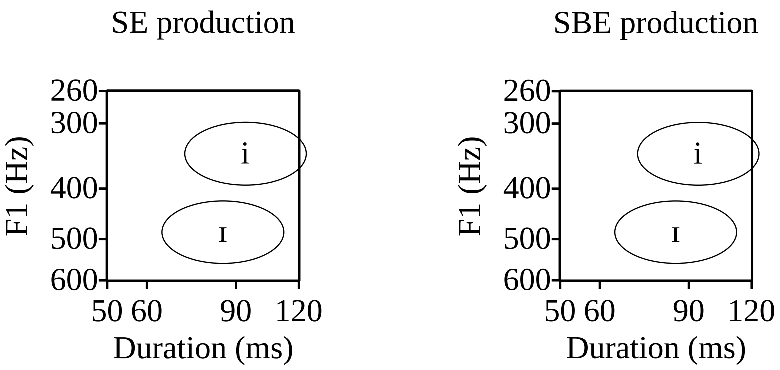

Fig. 2.7. Average and standard deviations (the ellipses) for the productions of SE and SBE /i/ and /ɪ/.

In this two-dimensional case, we first compute the midpoint of the line that connects the average productions of the two vowels, as shown in Figure 2.8 below. The midpoint of the line that connects SE /i/ and /ɪ/ is located at [408 Hz, 89 ms][13] and the one that connects the SBE vowels is located at [314 Hz, 79 ms].[14] This midpoint is on the equal-likelihood line because the standard deviations are equal, and this represents the most frequent F1-duration pair that is intended as /i/ 50% of the time and as /ɪ/ 50% of the time. When two dimensions are involved in the production of sound categories, the equal-likelihood production is not a point, as was explained in § 2.2.1, but a line, as shown in Figure 2.8. This line connects all the F1-duration pairs that have a 50-50% intention.

[13] The F1 value of the SE midpoint was computed as $2^{(\log2(485)\ -\ (\log2(485/343)/2))} = 407.86$ Hz and the duration value was computed $2^{(\log2(94)\ -\ (\log2(94/84.8)/2))} = 89.28$ ms.

[14] That is, $2^{(\log2(337)-(\log2(337/292)/2))} = 313.69$ Hz and $2^{(\log2(104.6)\ -\ (\log2(104.6/59.7)/2))} = 79.02$ ms.

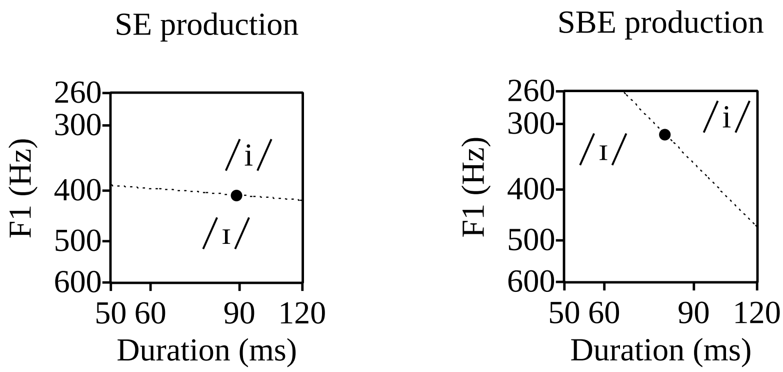

Fig. 2.8. Midpoint and equal-likelihood line (dotted line) for the production of SE and SBE /i/ and /ɪ/.

The slope of the *equal-likelihood line* can be computed as the ratio of the F1 and duration acoustic distances between the two vowels in each language multiplied by the squared ratio of the F1 and duration standard deviations. The F1 distance between SE /i/ and /ɪ/ is log2(485/343) = 0.500 octaves and the duration distance is log2(94/84.8) = 0.149 duration doublings. The standard deviations for the two dimensions are 0.2 and 0.4 respectively, as shown in Table 2. Thus, the SE equal-likelihood line has a slope of $(0.149/0.500) \cdot (0.2/0.4)^2 = 0.075$ octaves per duration doubling. Following the same computation, the SBE equal-likelihood line has a slope of $(0.809/0.207) \cdot (0.2/0.4)^2 = 0.98$ octaves/duration doubling. With the midpoint and the slope, we can compute the location of the points that cross the edges of the given F1-duration acoustic space, which gives us the location of the entire equal-likelihood line.[15] Figure 2.8 shows the equal-likelihood line in the production of SE and SBE /i/ and /ɪ/ as the dotted line that crosses the equal-

[15] The SE line extends from left to right (ranging from 50 to 120 ms) so that we need to compute its F1 location at the left and right edges. To compute the location of the F1 value at the left edge, we measure the duration distance between the midpoint and the left edge, and then multiply it by the slope of the line, i.e., log2(89/50) * 0.075 = 0.062 octaves, finally subtracting that distance from the F1 location of the midpoint, i.e., log2(408)= 8.672 – 0.062 = 8.61 octaves or $2^{8.61}$ = 390.83 Hz. The SE point on the right edge is located at [120 ms, 417 Hz], i.e., log2(408) = 8.672 + log2(120/89)*0.075 = 8.71 or 417.25 Hz. The SBE line extends from top left to centre right, and the location of the points at the top and right side are [260 Hz, 66 ms], i.e., log2(79)= 6.30 -log2(314/260)*0.98 = 65.66 ms, and [473Hz, 120 ms], i.e., log2(314)+log2(120/89)*0.98 = 472.99 Hz, respectively.

likelihood point in each environment. It can be concluded that the SBE slope is 13 times steeper that the SE slope.

Finally, to calculate the *relative use* of each dimension in the production of /i/ and /ɪ/, we first express the distances in standard deviations and then compare such distances by taking their ratio. Thus, the SE F1 distance between the two vowels expressed in standard deviations is 0.500/0.2 = 2.5 and the SE duration distance expressed in standard deviations is 0.149/0.4 = 0.37, which means that the F1-duration relative use is 2.5/0.37 = 6.76. For SBE, the F1 and duration differences expressed in standard deviations are 1.04 and 2.01 with a ratio of 1.95. This means that the F1 use is almost seven times the duration use in SE, while in SBE the duration use is twice that of the F1 use.

Following the optimal perception hypothesis, it is proposed that optimal listeners perceive the vowels according to the way they are produced in their specific environment. This means that the optimal perception will match the values computed from the production environment. Specifically, three main perceptual values will match their production counterparts if the optimal perception hypothesis holds. First, the location of the optimal perceptual boundary between the vowels, which estimates the F1-duration pairs that are equally likely to be categorized as /i/ or /ɪ/, will coincide with the location of the equal-likelihood line in production. Second, the shape of the perceptual boundary will match that of the production equal-likelihood line. In other words, the slope of the perceptual boundary line will have the same size as that of the equal-likelihood line. Figure 2.9 shows the predicted optimal categorization of F1-duration pairs as /i/ or /ɪ/ in SE and SBE.

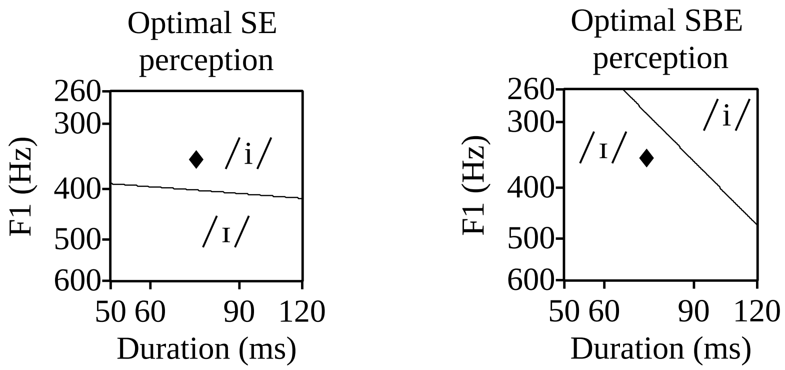

Fig. 2.9. Predicted optimal category boundary and categorization for tokens of /i/ and /ɪ/ in SE and SBE listeners. Diamond: [349, 74 ms].

The solid line in the figure represents the predicted perceptual category boundary line which defines the vowel productions that are perceived as /i/ 50% of the time and as /ɪ/ the other 50% of the time. This optimal perceptual boundary coincides with the equal-likelihood line, which is represented by the dotted line in Figure 2.8. Therefore, it is proposed that optimal listeners will categorize anything above their boundary line as /i/ and anything below it as /ɪ/.[16] For instance, the F1-pair [349 Hz, 74 ms], which is represented as a diamond in Figure 2.9, falls above the SE optimal category boundary line, and this means that an optimal SE listener will perceive this token as /i/. The same F1-duration token falls below the SBE optimal boundary, and therefore an optimal SBE listener will perceive this token as /ɪ/.

This optimal categorization of F1-duration pairs results from the optimal ranking of phonological cue constraints in the SE and SBE perception grammars. For this case of cue integration, the optimal perception grammar contains a family of continuous cue constraints that map both F1 and duration values onto phonological segments. Tableaux 2.3 and 2.4 show the optimal constraints and optimal constraint rankings that underlie the optimal SE and SBE perception of the auditory event [349 Hz, 74 ms]. In the SE perception grammar, the highest ranked constraint is '349 is not /ɪ/' because of the large distance between 349 Hz and the

[16] In this study, optimal listeners are equated to maximum-likelihood listeners. But see Escudero & Boersma (2003, 2004b) for a discussion of the difference between maximum-likelihood and probability matching listeners.

average production of /ɪ/ (cf. Table 2.2). Therefore, the optimal SE grammar outputs the category /i/ when confronted with the token [349 Hz, 74 ms], as shown by the pointing finger in Tableau 3. In contrast, a SBE perception grammar has the same constraint ranked the lowest and has '349 is not /i/' ranked the highest because 349 Hz is very different from the average SBE production but very close to the average /ɪ/ production. Therefore, the optimal SBE grammar will output /i/ for the same [349 Hz, 74 ms] token, as shown by the pointing finger in Tableau 2.4.

[349 Hz, 74 ms]	349 Hz not /ɪ/	74 ms not /ɪ/	74 ms not /i/	349 Hz not /i/
☞ /i/			*	*
/ɪ/	*!	*		

Tableau 2.3. Categorization of the F1-duration pair [349 Hz, 74 ms] by the optimal SE perception grammar.

[349 Hz, 74 ms]	349 Hz not /i/	74 ms not /i/	74 ms not /ɪ/	349 Hz not /ɪ/
/i/	*!	*		
☞ /ɪ/			*	*

Tableau 2.4. Categorization of the F1-duration pair [349 Hz, 74 ms] by the optimal SBE perception grammar.

As discussed in § 1.2.3, real-life SE and SBE listeners have an optimal perception of /ɪ/ and /i/ because they manifest a perceptual category boundary that resembles the equal-likelihood line in their specific production environment, as can be seen in the comparison of Figures 2.8 and 2.9. Therefore, it can be concluded that native speakers do manifest optimal perception when categorizing multiple auditory dimensions as vowels. This means that the principles that underlie sound perception can be adequately described by the perception grammar proposed within the LP model, i.e., with phonological cue constraints and their optimal rankings.

We have seen that two of the perceptual values for an optimal listener, viz., the location and the slope of the category boundary, match their production counterparts. The third, and last, perceptual value that matches the production environment is the relative perceptual use of the auditory dimensions. That is, for the categorization of vowels and consonants, the optimal listener will perceptually integrate auditory dimensions according to the way they are combined in production. Thus, it is predicted that optimal SE listeners will categorize /i/ and /ɪ/ mainly on the basis of their F1 differences and only to a small extent on the basis of their duration differences. This is because the use of acoustic dimensions in production shapes their optimal use in perception such that the optimal SE listener will have a perceptual relative use or reliance on F1 and duration of 7:1 because that is the language-specific relative use in production. On the other hand, the optimal SBE listeners will have a relative perceptual reliance on F1 and duration for the categorization of /i/ and /ɪ/ of 1:2 because that is the language-specific relative use of the dimensions. The computation of relative cue reliance will be further discussed in Chapter 3, where English vowels will be compared to Spanish vowels.

In sum, the LP model's adult perception grammar successfully models cross-linguistic differences in sound categorization. However, the categorization of continuous acoustic dimensions and combinations of dimensions implies a large number of constraints and constraint rankings. Therefore, the model should also explain how adult listeners arrive at optimal constraints and optimal constraint rankings. It is argued that the evidence gathered by speech perception research and the learning mechanisms attested in L1 sound perception can be incorporated into the LP model in order to provide a proposal that can most accurately deal with the perceptual learning of sound categories. The following section discusses the extension to the LP model and explains how linguistic sound perception comes about, i.e., how its elements emerge and develop in L1 acquisition.

2.3 Acquiring optimal L1 linguistic perception

This section is based on Boersma, Escudero & Hayes' (2003) L1 acquisition extension to the proposals put forth in Boersma (1998) and Escudero & Boersma (2003). Central to this acquisition proposal is the Functional Phonology assumption that sound categories and linguistic perception emerge from the learner's interaction with her production environment and are not present at birth. It is proposed that the L1 acquisition of sound perception involves the formation of language-specific abstract representations through the creation of language-specific perceptual mappings. What these authors postulate is that the child builds a linguistic perception grammar to map the auditory events in her production environment onto sound categories which are initially auditory-phonetic and which later become abstract arbitrary symbols. This L1 model thus provides a formal account of the transition from infant auditory-phonetic perception to adult phonological categorization.

In § 2.3.1, I present the proposal for the initial state of infant perception as well as the learning mechanisms that allow perceptual development. In § 2.3.2, I discuss the learning device that is taken to underlie the auditory and phonological categorization of the speech signal as well as the mechanisms that such a device is able to implement. In § 2.3.3 and § 2.3.4, the two sequential learning mechanisms that are at play in L1 perceptual learning will be illustrated with the acquisition of vowel categorization.

2.3.1 Initial perception grammar

In Boersma, Escudero & Hayes (2003), it is proposed that the three constraint families described in (2.1) are found in the perception grammar of an infant that is beginning to learn the perception of the sounds of her L1. These constraints are repeated in (2.5) for convenience:

(2.5) Auditory-to-auditory constraints in the infant perception grammar

PERCEIVE (*f*: *x*)

'Map the value *x* along the continuum *f* to some value along that same continuum'

*CATEGORIZE (*f*: *y*)

'Do not perceive anything as the value *y* along the continuum *f* ')

*WARP (*f*: *d*)

'Do not perceive a value along a continuum *f* as a value that is a distance *d* (or more) away along that same continuum').

It can be assumed that the constraints in (2.5) emerge in the infant's developing perception grammar. That is, every time a baby hears an auditory value [x], she creates a PERCEIVE and a *CATEG constraint for such a value along its auditory continuum. Through the introduction of these constraints, perceived auditory candidates are generated. For auditory values to be mapped onto one of these candidates, they should violate *WARP constraints which are introduced as required. This follows from the assumption that the creation of constraints that results in the generation of an auditory candidate set is driven by the general classification device with which humans are born. This device allows them to group tokens of objects in the world in order to optimally cope with all sorts of sensory input. This classification device is the Gradual Learning Algorithm (GLA) which has been described in Boersma (1998) and Boersma & Hayes (2001). This is the algorithm that is responsible for the perceptual learning described in § 2.3.2.

As for the initial ranking of the three continuous constraint families with respect to one another, Boersma, Escudero & Hayes (2003) propose that all *CATEG constraints are ranked higher than the PERCEIVE constraints, and also above *WARP constraints that do not change the identity of the input in an auditorily noticeable way, i.e., below what is known as the just noticeable difference for vowel formants, as reported in Kewley-Port (1995). Tableau 2.5 shows an example of the constraints and constraint rankings which, for instance, categorize F1 values onto F1 auditory-phonetic categories.

[320 Hz]	*CATEG (/320/)	*CATEG (/340/)	PERCEIVE ([320])	*WARP (20)
/320 Hz/	*!			
/340 Hz/		*!		*
☞ /–/			*	

Tableau 2.5. The null perception at the initial state in learning to perceive sound categories.

In general, it can be said that the interaction of the constraints at any given time in the infant's development leads to the auditory mapping of the signal onto a specific perceived auditory category. In the example in Tableau 5, the baby will have a 'null perception', that is to say, she will not categorize the input into any category because the constraints that ban the categorization of the two perceived auditory categories /320 Hz/ and /340 Hz/ are the highest ranked. Thus, Boersma, Escudero & Hayes propose that the infant will not be content with not classifying what she perceives, and that therefore perceptual development will occur through the lowering of the restrictions for some of her *CATEG constraints. In the next section, I discuss the proposed learning device that leads to optimal perception in addition to the types of learning that it is able to perform, viz., lexicon-driven and auditory-driven perceptual learning.

2.3.2 The Gradual Learning Algorithm (GLA)

Within the LP framework, it is assumed that the GLA is the learning device that underlies the L1 acquisition of linguistic perception and sound representation. This learning algorithm is a general learning device, innate and blind, that acts upon different kinds of input. It works upon input events to change constraint rankings via the decision-making framework of Stochastic or Probabilistic OT (cf. Boersma 1998) which is incorporated in the optimal perception hypothesis described in § 2.2. Recall that the LP model poses the optimal perception hypothesis to explain how constraints are ranked in the adult perception grammar which, as we have seen, operates by taking into account production distributions.

With respect to L1 perceptual learning, it is proposed that the LP model together with the GLA can explain how adult listeners arrive at the constraints and optimal constraint rankings that are found in their perception grammar. Escudero & Boersma (2003, 2004b) proposed that the GLA performs the perceptual learning that occurs as soon as the infant has a lexicon. Here I discuss a simplified version of the authors' example of an infant SE listener. At some point in her development, this infant may have a constraint ranking in her perception grammar that is not completely appropriate for her production environment. This non-optimal ranking may result in the perception of a vowel token with an F1 of [349 Hz] as /ɪ/, as shown by the pointing finger in Tableau 2.6.

[349 Hz]	349 Hz not /i/	349 Hz not /ɪ/
/i/	*!	
☞ /ɪ/		*

Tableau 2.6. Categorization of 349 Hz as /ɪ/ by a non-optimal SE perception grammar.

However, in the SE production environment, a vowel token with such an acoustic value is most likely to have been intended as the vowel in *sheep* rather than the one in *ship*. This is because an F1 value of 349 Hz is closer to the average F1 production of SE /i/ than to that of /ɪ/ and, in this language, F1 differences are more important than duration differences, as was shown in § 2.2.2. It is proposed that the infant's *recognition grammar,* which will be discussed at length in § 2.4, will detect a mismatch between the perceived representation and the phonological form of the lexical item that she accesses contains |i|.[17] This is because the semantic context intended by the speaker was 'animal' and not 'means of transportation'. Thus, this mismatch between the output of perception and the output of recognition tells the infant that the correct perception should have been /i/ (which is preceded by a check mark in Tableau 2.7), and that /ɪ/ is a perception error (which is represented as '* *').

[349 Hz] Intended as |i|.	349 Hz not /i/	349 Hz not /ɪ/
√ /i/	*!→	
* ☞ * /ɪ/		←*

Tableau 2.7. One-dimensional lexicon-driven constraint re-ranking.

17 Within the LP model, lexical forms are written between pipes to differentiate them from phonological representations which are written between slashes.

To repair the error that causes the perception-recognition mismatch, the child's GLA changes the perception grammar by raising the rankings of all the constraints that are violated that in the incorrect winner, by lowering the rankings of all the constraints that are violated in the form that she now considers correct, and also by raising the constraints of the form that she knows she should have perceived, as depicted by the arrows in Tableau 2.7. This increases the probability that she will perceive /i/ the next time she hears an F1 of 349 Hz. The rankings are changed by only a small *step* along the continuous ranking scale of stochastic OT, but after the occurrence of a large number of perception errors involving auditory events containing 349 Hz, the rankings of the constraints will have become similar to those of the adult SE listener.

There exists empirical evidence in support of this type of perceptual development which is guided by the lexicon. For instance, Nittrouer (1996) showed that once the lexicon is in place, young children developmentally adjust their category boundaries and their perceptual use of acoustic dimensions to match those of adult-like category boundaries for one-dimensional mappings and multi-dimensional mappings. That is, infants change the location and shape, i.e., the slope, of their category boundary to make it more appropriate to their language environment. This suggests that infants and young children further adjust their perception to more closely resemble adult perception once they have abstract representations of words in their lexicon. In other words, they go from relying solely on auditory input to taking into account speakers' intended meanings. Thus, the proposed GLA lexicon-driven learning can be viewed as the formalization of the learning mechanism that results in developmental category boundary shifts.

Escudero & Boersma's (2003) modelling assumes the worst possible starting point for lexicon-driven learning, which is that babies have a 50-50% categorization at the onset of the type of perceptual learning that occurs with the help of the lexicon, that is, the error-driven learning that results from a perception-recognition mismatch and which is shown in Tableau 2.7. This situation can only exist if no perceptual learning has occurred prior to the establishment of a lexicon that contains abstract phonological forms. However, this is unlikely because infant perception research has shown that babies as young as six months start to attune their perception to the sounds of their ambient language. This means that language experience leads to early perceptual learning that does not occur with the aid of the lexicon. For instance, Maye, Werker & Gerken (2002) have shown that infants have a remarkable capacity to calculate the frequency distributions of the auditory-phonetic information in their linguistic input, and have proposed that this leads to the formation of phonetic categories. In addition, Kuhl (1991) has shown that very young infants exhibit a warping of their perceptual space in favour of the most common or most prototypical realization of sound categories in their language. In her estimation, this leads to the *perceptual magnet effect* in speech perception which manifests itself as the formation of *prototypes* in the developing infant's perception. Such prototypes attract the perception of less common tokens, which means that they allow the mapping of multiple productions onto a finite set of discrete prototypical categories.

To bridge the gap between lexicon-driven perceptual learning and the auditory-phonetic learning that occurs prior to lexical availability, Boersma, Escudero & Hayes (2003) put forth an integrated model of the two perceptual learning processes. They claimed that the GLA and a perception grammar with auditory-phonetic constraints, i.e., the three constraint families discussed in § 2.3.1, can explain the perceptual learning that leads to the formation of phonetic categories in infants. At this initial stage of perceptual learning, the GLA is seen to as an identity matching device and as a frequency-driven learning device. Thus, the infant's gradual learning device will first allow her to match an incoming F1 value to a perceived auditory category that preserves its identity. Later on, based on the higher frequency of occurrence of certain auditory values in the environment, the infant's GLA will be able to map not just the most frequent values onto their perceived counterparts but also other less frequent values onto the perceived categories of frequent values. This is what is called *distributional learning*, the consequence of which is the frequency-driven ranking of *CATEG constraints in the infant's perception grammar. As a result, the most frequent values will be the preferred categories because their constraints will be the lowest ranked.

2.3.3 Learning mechanism 1: one-dimensional auditory-driven learning

Given the contents of the previous section, we can see that perceptual learning takes place through two types of learning mechanisms which occur sequentially. First, the GLA performs auditory-driven learning during which it acts as an identity-matching and distributional learning device fed by the frequency distributions of the auditory events in the linguistic input. Consider, as an example, a SE infant learning to categorize the F1 continuum into English /i/ and /ɪ/. Figure 2.10

shows the distributions of the F1 values with which the vowels are produced, and which are taken from the averages and standard deviations in Table 2.2.

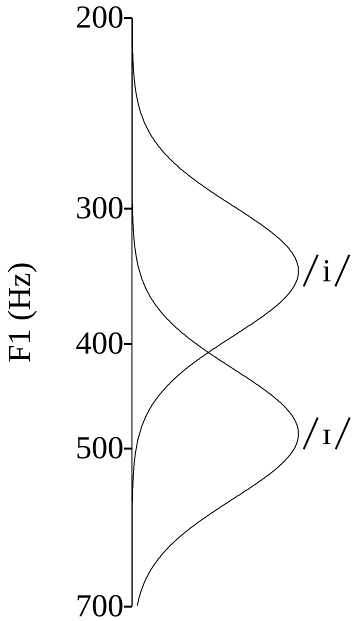

Fig. 2.10. F1 distributions for the SE /i/ and /ɪ/ vowels.

Tableau 2.8 shows how the infant's innate distributional learning device or GLA changes a proposed initial state in which all *CATEG constraints are ranked high and all PERCEIVE constraints are ranked low into an 'identity-matching' perception in which the learner will want to map an incoming value onto its 'identical' perceived auditory value so that, for instance, an incoming [300 Hz] will be perceived as /300/. That is, the infant's initial preference for not perceiving incoming F1 values at all will turn into a perception of such values as themselves, i.e., the child's GLA will perform an identity matching of auditory events.

[300 Hz]	*CATEG (/300/)	*CATEG (/320/)	PERCEIVE ([300])	*WARP (20)
/320 Hz/		*!		*
√ /300 Hz/	*!→			
☞ /–/			←*	

Tableau 2.8. Auditory-driven perceptual learning.

Given the SE distributions of /i/ and /ɪ/ shown in Figure 9, this infant is likely to be confronted with a great number of F1 values though some of them will be more frequent than others due to the probability of values represented by the Gaussian curves shown in § 2.2. This will be the case if we assume that the SE productions of /i/ and /ɪ/ have Gaussian shapes with peaks or most frequent productions at the vowel averages, i.e., at 343 Hz and 485 Hz respectively, as shown in Figure 9. Given the frequency with which F1 values are produced in her environment, the infant's perceptual space will be warped in such a way that she will learn to map incoming F1 values onto the auditory-phonetic categories with the values of the centres, the ones that are more frequent. This warping of the perceptual space is also leads to discrimination curves and, ultimately, to the creation of auditory-phonetic categories.

Once phonetic categories have become discrete, they are automatically turned into abstract perceptual categories, which in the case of the two vowel heights could be /high/ and /semi-high/. Also, during this phonological development, additional constraints that map acoustic events to the newly created abstract phonological categories are created. These are part of the one-dimensional auditory-to-feature constraint template that was formulated in (2.2), i.e., 'a value x on the auditory continuum y should not be perceived as the phonological feature /z/'. Thus, for the learning of F1 phonological categories, phonetic-to-phonological constraints such as '[300 Hz] is not /high/' or '[500 Hz] is not /high/' are initially incorporated into the infant perception grammar.[18] These constraints map the input onto phonological features, thus enabling the infant to use abstract phonological categories that, at this stage, are arbitrary one-dimensional features such as

[18] See Boersma, Escudero & Hayes (2003: § 3.1) for a discussion of how the original auditory-to-auditory constraints, e.g., *WARP, get translated to accommodate the newly introduced phonetic-to-phonological constraints.

/height/. As a result, phonological features that refer to only one auditory dimension are used to economically store words in her lexicon. At this point in the learning process, the task of mapping acoustic events to discrete abstract categories has been achieved, resulting in a reasonably good categorization performance, as shown in Tableau 2.9.

[300 Hz]	PERCEIVE ([300])	300 is not /semi-high/	300 is not /high/	*CATEG (/320/)
☞ /high/			*	
/semi-high/		*!		

Tableau 2.9. The result of distributional learning with the new one-dimensional cue constraints.

Within this L1 model, then, the phonological abstraction in the representations and in the mapping constraints constitutes the transition between auditory-driven and lexicon-driven learning. It is proposed that the types of constraints described in § 2.1.1 emerge in the infant's grammar at different stages of perceptual development. We have seen that auditory-to-auditory constraints, as formulated in (2.1), are the first type of constraints that are introduced in the infant grammar. Figure 2.11 shows the proposed developmental sequence for all four types of constraints.

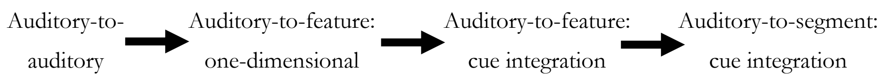

Fig. 2.11. Transition in the type of constraints present in the learner's developing perception grammar. These constraints were formulated in (2.1), (2.2), (2.3), and (2.4), respectively.

The next section shows how the abstract representations of words in the lexicon help the SE infant to acquire optimal one-dimensional constraint rankings. In addition, it is discussed how the integration of F1 and duration in the perception of /i/ and /ɪ/ comes about.

2.3.4 Learning mechanism 2: lexicon-driven learning and cue integration

As explained in § 2.3.2, once abstract form-meaning pairs are stored in the lexicon, they can trigger re-rankings in the perception grammar that will lead to optimal one-dimensional perception. In Tableau 5, for instance, the dotted line between the two phonetic-to-phonological constraints suggests that they do not yet have a strict ranking with respect to one another, perhaps because 300 Hz could be used to produce /ɪ/. For those occasions in which the speaker intends to produce a /semi-high/ vowel but does so with an F1 value that is commonly associated with a /high/ vowel, the infant learner will have to rely on an extra type of information to be able to perceive the vowel intended by the speaker. Following the optimal perception hypothesis, an optimal perception grammar is able to match the speaker's intended production. This optimal perception is manifested as the optimal location of the perceptual boundary which needs to coincide with the production environment (cf. § 2.2).

It is proposed that the GLA, at this point in development, acts as a lexicon-driven learning device that reranks the perception grammar constraints when mismatches occur between the output of perception and the lexicon. For instance, imagine that a [349 Hz] production is intended as a /high/ vowel rather than as a /semi-high/ vowel, something that would be more common in a SBE environment. In this case, the semantic context reveals that the speaker intended a word containing a /high/ vowel, such as *sheep*. At this point, that is, after distributional learning has taken place, the infant has a constraint ranking that maps [349 Hz] onto /semi-high/ because this F1 value is frequently used to produce a /semi-high/ vowel in SBE. Therefore, the learner will perceive /semi-high/ when categorizing [349 Hz]. However, her knowledge of the semantic context will tell her that the speaker intended a word containing a /high/ vowel and not her perceived /semi-high/ category, thus resulting in a perception-lexicon mismatch. It is proposed that the listener's GLA acts upon such a mismatch by ensuring that she will be more likely to perceive the next [349 Hz] token as /high/. This is achieved by lowering the constraint against perceiving the acoustic value as /high/ and by simultaneously raising the one against perceiving the same value as /semi-high/, as in § 2.3.2. This procedure continues until the grammar performs an optimal mapping of the production environment.

Recall that adult perception is characterized by the integration of multiple dimensions into abstract phonological categories. In § 2.1.1, it was argued that cue integration can only be possible if the phonetic-to-phonological constraints can map onto any abstract phonological feature. The L1 acquisition model proposes that once sound categories have abstract labels in the lexicon, the child can consider the relations of each category with all auditory continua. These phonetic-to-phonological constraints are part of the multidimensional auditory-to-feature constraints that were formulated in (2.3), viz., 'a value *x* on the auditory continuum y should not be perceived as the phonological feature /*z*/'. They provide the phonetic-to-phonological mapping of any auditory value to any of the candidate phonological features. Examples of these constraints are 'an F1 of 349 Hz is not /short/' and 'a duration of 91 ms is not /semi-high/'. These constraints also allow for the perceptual integration of auditory dimensions, thus resulting in a perception that is more adult-like (cf. § 2.1.1), and this means that only at this point can the learner manifest the kind of cue integration that commonly takes place in adult perception.

Later on, the infant will introduce constraints that map onto segmental units in order to optimally cope with her production environment. As was argued in § 2.1.1, phonetic-to-phonological constraints that map auditory values onto segments are likely to be the optimal constraints in an optimal adult perception grammar. This is because, for instance, each of the 10 SBE monophthongs (namely /i, ɪ, ɛ, æ, ʌ, ɑ, ɒ, ɔ, ʊ, u/) occupies its own unique position in an F1-F2 plane, as opposed to, for instance, Japanese where every short vowel has a long counterpart with identical F1 and F2 values. Therefore, an optimal SBE listener will be better off with a direct mapping of F1 and duration onto vowel segments rather than onto feature combinations such as /high, long/.

As for the empirical evidence in favour of developmental cue integration, many studies have shown that perceptual cue use is different for adults, infants, and children (cf. Gerrits, 2001; Nittrouer & Miller, 1997, Jones 2003). For instance, Gerrits (2001) showed that Dutch four-year-olds pay attention to duration cues much more than adult listeners do for both vowels and consonants. Likewise, Jones (2003) showed that young children use either vowel duration or F1 onset when perceiving final obstruent voicing in American English, while adult listeners integrate both types of information. This suggests that infants first perform one-dimensional categorization and, only later, integrate acoustic dimensions in the perception of sound categories. Thus, the developmental sequence proposed by Boersma, Escudero & Hayes (2003) seems to constitute an adequate model for L1 perceptual learning. In the next section, I present a brief summary of the proposal for word recognition, given that perceptual learning depends on optimal word recognition.

2.4 The proposal for word recognition

In the previous section, it was shown that once the lexicon is in place, the infant can adjust the ranking of constraints in her grammar to achieve optimal perception through GLA lexicon-driven perceptual learning. This type of perceptual learning assumes that the infant is able to retrieve the lexical representation that corresponds to the speaker's intended word. However, this word retrieval is not an automatic result of perception but involves a mapping from perceived representations onto underlying lexical representations that are connected to meaning. In Boersma (1998, 2001), it was proposed that the mapping of perceptual representations onto lexical items is performed through a *recognition grammar*. In this section, I will summarize his proposal for this mapping and, crucially, for its acquisition.

2.4.1 Lexical representations and recognition grammar

Recall that the word comprehension model in Figure 1 includes a recognition system that connects the output of perception with lexical representations. Thus, Boersma proposes that the word recognition system could be formalized as an OT recognition grammar that maps the perception of an utterance to the underlying phonological form stored in the lexicon so that meaning can be accessed. Furthermore, this recognition grammar contains constraints against changing the form of the perceived form when accessing lexical items, i.e., the FAITH constraint family, and constraints against gaining lexical access, i.e., the *LEX constraint family. Word recognition is thus achieved through the interaction of these two constraint families.

Boersma illustrates his proposal with a case of phonological alternation, viz., final devoicing in Dutch. He shows that the use of constraints that directly evaluate candidate lexical forms in comprehension allows us to account for the recognition of two lexical items like Intended as |rɑt| and |rɑd| from a single perceived or phonological representation /rɑt/. In addition, based on psycholinguistic evidence, he proposes that the ranking of *LEX constraints is provided by word frequency in that words with a higher frequency of occurrence are recognized better than words with a lower frequency. Thus, because the Dutch word |rɑt| 'rat' is more common than the word |rɑd| 'wheel', a token produced as [rɑt] will lead to the perception of / rɑt / and thus to the recognition of |rɑt| if the context is not taken into account. This frequency-driven ranking can be expressed as *LEX (|rɑd| 'wheel') >> *LEX (|rɑt| 'rat' (where '>>' means 'is higher ranked than'). Given that the recognition grammar outputs the lexical form that violates the lowest ranked constraints, the form |rɑt| 'rat' will be the one recognized.

*LEX constraints are also ranked on the basis of the semantic context. That is, if the semantic context within which [rɑt] is produced suggests 'turn', the recognition of |rɑd| 'wheel' will be preferred. This semantic conditioning of word recognition is proposed to be conveyed by the constraint ranking shown in (2.6).

(2.6) *Semantic conditioning of lexical access*

*LEX (|rɑt| 'rat' in context = 'turn') >> *LEX (|rɑd| 'wheel' in context = 'turn')

Thus, *LEX constraints are ranked on the basis of word frequency and semantic context, and this entails that the ranking in (2.7) is the optimal one for the *LEX constraints in an adult recognition grammar.

(2.7) Word frequency and semantic ranking of *LEX in the recognition grammar

*LEX (|rɑt| 'rat' in context = 'turn')
>> *LEX (|rɑd| 'wheel' in context = 'turn') (ranked by semantic context)
>> *LEX (|ʋil| 'wheel' in context = 'turn') (ranked by word frequency)

Tableau 2.10 shows how an optimal Dutch recognition grammar selects the lexical form when a speaker says [rɑt] in the context 'turn' and means |rɑt| 'wheel'. An optimal recognition grammar in this case exhibits an interaction of phonological and semantic constraints so that in order to make |rɑd| 'wheel' the winner when the input to the grammar is [rɑt] in the context 'turn', phonological faithfulness is ranked between two semantic constraints. The FAITH constraint in the tableau must refer to vowel height ('do not change /ɑ/ to |i|') and consonant sonority ('do not change /r/ to |ʋ| or /d/ to |l|'). The FAITH constraint against changing the voicing feature of the final consonant should be ranked low so that [rɑt] can be recognized as |rɑd|.

[rɑt] context = 'turn'	*LEX |rɑt| 'rat' / 'turn'	FAITH	*LEX |rɑd| 'wheel'/'turn'	*LEX |ʋil| 'wheel'/'turn'
|rɑt| 'rat'	*!			
☞ |rɑd| 'wheel'		*	*	
|ʋil| 'wheel'		***!		*

Tableau 2.10. Phonology-semantic interaction in an optimal Dutch recognition grammar.[19]

Here we see that the first lexical candidate (going from top to bottom) is not chosen because it is semantically too distant from 'turn', and the last candidate is not chosen because its phonological form is too different from [rɑt]. Thus, the optimal Dutch recognition grammar has the following constraint ranking *LEX (|rɑt| 'rat' / context = 'turn') >> FAITH >> *LEX (|rɑd| 'wheel' / context = 'turn'). With this optimal ranking, |rɑd| 'wheel' will be the recognized or accessed form, the one matching the speaker's intended form. The question now is how an adult Dutch listener acquires this optimal recognition grammar. The next section discusses Boersma's (2001) proposal for the L1 acquisition of the rankings of phonological and semantic constraints in the recognition grammar.

2.4.2 The L1 acquisition of optimal L1 recognition

Boersma's proposal for the ranking of *LEX constraints says that it is driven by word frequency because it leads to a lower semantic misunderstanding than if word frequency were not taken into account. For instance, if [rɑt] productions mean 'rat' 70% of the time and 'wheel' the other 30%, an L1 learner who always recognizes |rɑt| will misunderstand a speaker only 30% of the time only because she has the grammar depicted in Tableau 2.10. It is proposed that this word frequency-driven reranking of *LEX constraints is also performed by the GLA, which in this case reranks lexical constraints on the basis of instances of recognition errors, i.e., mismatches between the context in which a word is produced and the recognized

[19] For simplicity purposes, this tableau does not include the structural constraint against final voiced codas (*VOICED CODA) ranked above FAITH which is found in Boersma (2001: 31).

form-meaning pair. This learning mechanism is similar to the error-driven perceptual learning that is guided by the lexicon, as was shown in § 2.3.2. That is, if a child has a non-frequency driven grammar and she recognizes |rɑd| 'wheel' when the speaker intended |rɑt| 'rat', her GLA will rerank the *LEX constraints so that the most common word will be most likely to win out. This will therefore result in the 70-30% recognition of the two words.

However, since this frequency-driven recognition still leads to a high percentage of errors, i.e., it leads to occasions in which the child recognizes a word that was not intended by the speaker, an extra source for the ranking of *LEX constraints should be available, and that source is semantic context. This means that the child will need to learn that a [rɑt] production within the semantic context 'turn' should be recognized as |rɑd| and not as the lexical item that her word frequency ranking tells her to recognize, i.e., |rɑt|. It is thus proposed that the GLA will also allow the learner to rerank constraints on the basis of recognition errors due to the semantic context. For instance, she may entertain a ranking of *LEX |rɑt| 'rat' that is lower than *LEX |rɑd| 'wheel' when confronted with the speaker's production of [rɑt] because the first word is more common than the latter. However, if the word is produced in the semantic context 'turn', the learner will commit a semantic error because she will recognize 'rat' and not 'wheel'. As a result, she will have the optimal semantic ranking that was shown in (2.6).

There is more to this, however, for the learner still needs to incorporate phonological faithfulness (FAITH) into her recognition grammar so that she can compare the phonological form of her recognition with the phonological form of the incoming perceived utterance. This optimal ranking of FAITH and *LEX constraints, which leads to their optimal interaction in the recognition grammar, is also achieved through GLA error-driven learning. That is, if the child has an underdeveloped sensitivity to context or an excessive sensitivity to phonological faithfulness, i.e., if she has the ranking *FAITH >>*LEX (|rɑt| 'rat' / context = 'turn'), she will access |rɑt| 'rat', and this will result in a semantic discrepancy between the intended production and the child's comprehension. This discrepancy will lead to the constraint reranking shown in Tableau 2.11.

[rɑt] context = 'turn'	FAITH	*LEX |rɑt| 'rat' / 'turn'	*LEX |rɑd| 'wheel'/'turn'
☞ |rɑt| 'rat'		←*	
√ |rɑd| 'wheel'	*!→		*→

Tableau 2.11. Learning the optimal phonology-semantic interaction through GLA error-driven learning.[20]

2.4.3 Summary: adult Linguistic Perception and its L1 acquisition

The LP model describes, explains, and predicts adult sound perception by proposing that a perception grammar with auditory-to-segment cue constraints can optimally cope with the production environment. The ranking of the constraints in the adult perception grammar follows the optimal perception hypothesis which states that an optimal listener perceives an auditory value as the phonological segment that was most likely to have been intended by the speaker. Thus, an optimal listener's perception matches the production distributions of the environment at hand because the optimal perceptual boundary between two sound categories coincides in both location and shape with the equal-likelihood in production. The LP proposal for adult sound perception is summarized in Table 2.3.

Table 2.3. The LP proposal for adult sound perception and its L1 acquisition.

Adult sound perception		L1 acquisition	
Type of mapping	Constraint ranking	Initial state	Development
Auditory-to-segment cue constraints	Optimal perception hypothesis	Auditory-to-auditory constraints	Distributional, and then lexicon-driven learning

[20] In this case, FAITH refers to the voicing value of the final consonant.

With respect to L1 acquisition, the LP model has been extended to explain how optimal sound perception is acquired. The claim is that infants first create auditory-phonetic constraints and categories and then gradually develop into optimal adult listeners. It was shown that the GLA is responsible for the perceptual development that results in the auditory-to-auditory mapping of the signal and in the adjustment of category boundaries once the infant has created phonetic-to-phonological mappings as well as abstract phonological categories that are used in the lexicon. The L1 proposal also predicts a specific gradual developmental path from infant perception to optimal adult perception. Table 2.4 summarizes the proposed learning steps for the development of optimal sound perception.

Table 2.4. Proposed developmental path for the learning of L1 linguistic perception and sound representation.

	Auditory-driven learning	Lexicon-driven learning
Step 1	Create auditory-to-auditory constraints & auditory categories	
Step 2	Distributional learning: constraint rankings match the production distributions	
Step 3	Abstraction 1: turn auditory constraints and categories into one-dimensional cue constraints and phonological features	
Step 4	Phonological features are copied to the lexicon	
Step 5		One-dimensional constraint rerankings and category boundary shifts

Step 6	Abstraction 2: turn one-dimensional constraints into multidimensional ones (every auditory dimension maps onto every feature)
Step 7	Initial cue integration
	Abstraction 3: turn features into segments and store them in the lexicon. Turn featural cue constraints into segmental ones
Step 8	Optimal cue integration (adult-like)

These proposed L1 learning mechanisms and developmental path form the basis of the L2 proposal that is advanced in the next chapter. This L2 model combines the LP model with the issues involved in the acquisition of second languages.

PART II:

MODELLING
THE L2 ACQUISITION
OF SOUND PERCEPTION

3 The Second Language Linguistic Perception (L2LP) model

This chapter addresses the cross-linguistic and developmental variation in L2 sound perception which is attested in the research in this field. In this literature, it has been shown that L2 learners use the auditory information that differentiates sound categories differently from native listeners. For instance, when categorizing /i/ and /ɪ/, American English listeners prefer spectral (i.e., vowel quality) information, Mandarin learners of English prefer temporal (i.e., vowel duration) information, and Spanish learners of English use the two dimensions equally (cf. Bohn 1995; Flege, Bohn & Jang 1997). Moreover, it is known that L2 learners change their auditory cue weighting as their experience with the L2 increases. For instance, Morrison (2002) showed that that the cue weightings for the /i/-/ɪ/ contrast of Spanish learners of Canadian English change from a poor reliance on duration after their first month in Canada to a reasonably good reliance on spectrum or duration (but not on both) after their sixth month. The question of interest to language acquisition researchers, then, is how best to explain the acquisition process in L2 sound perception.

This chapter advances a linguistic model for L2 sound perception and its acquisition which will be referred to as *Second Language Linguistic Perception (L2LP)* because its general framework is the LP model discussed in Chapter 2. This model provides a theoretical and methodological framework to describe, explain, and predict the acquisition of L2 sound perception. It is composed of five theoretical ingredients which offer a proposal for each of the components of the L2 acquisition process, namely the description of the two languages involved, the initial state, the learning task, the development, and the end state.

3.1 The L2LP model: five ingredients

The L2LP model offers both a theoretical and methodological proposal because it composed of five theoretical ingredients that are at the same time a sequential methodology for testing and evaluating the model's predictions and explanations. In § 3.1.1, I show how the model's distinction between perceptual mappings and phonological representations reveals differences between languages that are crucial for describing, explaining, and predicting L2 sound perception. In § 3.1.2, I introduce the model's first ingredient, which is also its first methodological phase, viz., the prediction, explanation, and description of the optimal perception in the two languages at hand. In § 3.1.3, I present an overview of the model's four remaining ingredients and a discussion of how they relate to the logical states of language acquisition, namely the initial, developmental, and end states, which must all be part and parcel of any adequate model of language acquisition.

Each of the model's five ingredients will be presented as follows. First, I state the model's prediction for monolingual, cross-language, or L2 sound perception. Second, I present the model's theoretical explanation that underlies the predicted perception, including a discussion of previously proposed explanations and well-known concepts in L2 acquisition research. Finally, I provide an analysis of the predicted perceptual category boundaries and cue use, as well as of the knowledge that underlies such a predicted perception.

3.1.1 Distinction between perceptual mappings and sound categories

The L2LP model is based on the LP model, which means that L2 perception is described using linguistic perception grammars and phonological categories. Importantly, one of the L2LP's basic claims is that a separation of perceptual mappings from sound representations leads to an adequate comparison of the perception systems involved. In this section, I seek to exemplify how the principled separation of mappings and categories is crucial to the description, explanation, and prediction of L2 sound perception.

Canadian English (CE) and Canadian French (CF) have /æ/ and /ɛ/ in their vowel inventories. Following the optimal perception hypothesis, we can compute the language-specific midpoints between the vowels of the two languages (cf. § 2.3.1). To demonstrate how differentiating between mappings and categories leads to a better description of L2 sound perception, Figure 3.1 shows the F1 midpoints for /æ/ and /ɛ/ which are represented as horizontal lines in the two languages.[21] The average F1 productions for the CE vowels are 840 Hz and 681 Hz respectively, and the acoustic distance between them is log2 (840/681) = 0.3 octaves, so that the boundary point is located at log2 (840) - 0.15 = 9.563 octaves above 1 Hz, which is to say at 756 Hz. For the CF vowels, the averages are 728 Hz and 557 Hz, the F1 distance is = 0.39 octaves, and the boundary point lies at 637 Hz. The figure

[21] For a discussion of two-dimensional category boundary lines see Chapters 5 and 7.

thus shows how optimal CE and CF listeners categorize vowel tokens with F1 values ranging from 550 to 950 Hz. Note that these boundary points are optimal only if we assume that the input tokens differ along the F1 dimension only, i.e., if they have ambiguous duration values.

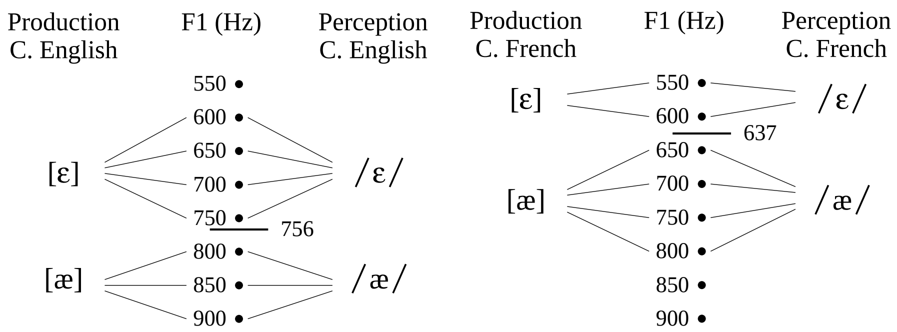

Fig. 3.1 CE and CF perceptual mappings and sound representations for the productions of /æ/ and /ɛ/.

What we observe here is that although the phonological categories in the two languages are the same, or are described with the same abstract symbols, the optimal perception hypothesis predicts that their perceptual mappings will be different. Thus, tokens between 637 and 757 Hz will be categorized as /ɛ/ by CE listeners but as /æ/ by CF listeners. In other words, the CE and the CF optimal perception of /æ/ and /ɛ/ may result in the same or similar abstract sound categories but in different category boundaries. This separate comparison between sound categories and perceptual mappings allows for the prediction of a double L2 learning task, one perceptual and one representational, as will be discussed in § 3.3. Crucially, when perceptual mappings are the only source of difference between the L1 perception and the target L2 perception, the L2 learner will have a perceptual learning task but not a representational one.

3.1.2 L2LP ingredient 1: optimal L1 perception and optimal target L2 perception

The L2LP model proposes that the first step in accounting for L2 sound perception is to conduct a thorough analysis of the optimal perception in each of the languages involved. Table 3.1 summarizes the L2LP predicted perception, the underlying explanation for such a perception, and the proposed tools for describing perception and its underlying linguistic knowledge.

Table 3.1. Prediction, explanation, and description in the L2LP's first ingredient.

L2LP model	Prediction	Explanation	Description
Ingredient 1: L1 and target L2 perceptions	Native listeners are optimal perceivers	Optimal perception hypothesis (cf. § 2.2)	Native listeners have optimal perception grammars

In Chapter 2, we saw that the LP model predicts that listeners are optimal perceivers. This prediction finds an explanation in the *optimal perception hypothesis* which states that human listeners maximize their probabilities of understanding speakers by making perceptual decisions that match their intended message (cf. § 2.2). This leads to a strong dependency of perception on the production environment because an optimal listener manifests a sound perception that matches the production of sounds in her environment. In § 3.1.2.1, I demonstrate how this general sound perception mechanism is used for the prediction and explanation of L2 sound perception. In § 3.1.2.2, I give a description of the optimal perception grammar in the languages involved, i.e. the learner's L1 and target L2 perceptions. It is proposes that this analysis of monolingual optimal perception is the first step towards a comprehensive explanation of L2 sound perception.

3.1.2.1 L2LP ingredient 1: prediction and explanation

Within the L2LP model, the term *L1* refers to the learner's native language. The term *target L2* refers to the language to be learned and defines the goal of the learning process or the L2 learning task. The term *L2* or *learner's L2*, which is often referred to as *interlanguage*, is used for the system formed upon exposure to a new language. For purposes of simplicity, the model refers to a language acquisition process in which the L2 learner has a single L1, learns a single target L2, and therefore develops a single L2 system.[22] Following the optimal perception prediction and hypothesis shown in Table 1, the L2LP model defines *optimal L1 perception* as the best possible way to perceive sound categories in the learner's first language. The *target L2 optimal perception* is the best possible way of perceiving the learner's target L2, which is predicted to be found in native speakers of the target language.

The L2LP model proposes that the description of the optimal L1 perception and optimal target L2 perception allows us to predict and explain three different aspects of L2 sound perception. These aspects correspond to three different ingredients of the L2LP model, in other words, it is proposed that the description of the optimal L1 perception leads to predicting the initial state of the L2 learning process, i.e., the perceptual system that learners will initially use in their L2. Likewise, computing the optimal target L2 perception allows us to determine the precise nature of the L2 learning tasks that the learner needs to perform in order to attain optimal target L2 perception. Finally, the L1 optimal perception will give a reliable estimate of the system that the learner needs to preserve in order to be able to best cope with her L1 environment while she develops into an optimal L2 listener.

3.1.2.2 L2LP phonological/phonetic description

We can start by examining how the L2LP model describes the linguistic knowledge underlying the optimal perception of the learner's L1 and target L2.[23] Recall that the LP described in Chapter 2 proposes that an OT perception grammar represents the linguistic knowledge that underlies sound perception. It is important to remember that adult perception grammars contain continuous cue constraints that evaluate incoming speech input and classify them as the most likely phonological representation. Tableaux 3.1 and 3.2 show a phonological description of the knowledge underlying the categorization of the F1-duration pair [349 Hz, 74 ms] by optimal Spanish and Southern British English (SBE) listeners respectively.

---

[22] Although this may sound like an idealization of L2 acquisition, it is argued that this basic situation needs to be thoroughly explained before we can move on to cases in which more than two languages are involved.

[23] This proposal is an extension of the analysis and computations provided in Escudero & Boersma (2003, 2004b).

[349 Hz, 74 ms]	349 Hz not /e/	74 ms not /i/	74 ms not /e/	349 Hz not /i/
☞ /i/		*		*
/e/	*!		*	

Tableau 3.1. The optimal perception grammar that underlies the perception of [349 Hz, 74 ms] by an optimal Spanish listener.

[349 Hz, 74 ms]	74 ms not /i/	349 Hz not /i/	349 Hz not /ɪ/	74 ms not /ɪ/
/i/	*!	*		
☞ /ɪ/			*	*

Tableau 3.2. The optimal perception grammar that underlies the perception of [349 Hz, 74 ms] by an optimal SBE listener.

Recall that the LP posits that cue constraints are ranked according to the optimal perception hypothesis. In the tableaux above, we see the ranking of only the relevant constraints and candidate representations. Constraints that do not refer to 74 ms or to 349 Hz are ignored as are vowel categories other than /i/ and /ɪ/. This perception grammar thus represents the linguistic knowledge underlying optimal perception which we can observe through perceptual experiments. It is proposed here that at least three results of optimal perception grammars can be measured and compared across the languages that are involved in any given case of L2 sound perception. These are the location of category boundaries, the shape of category boundaries, and the relative use of auditory dimensions. Following the optimal perception hypothesis, the optimal perception of a language matches its production distributions. To make the description of two languages comparable, we need to consider the same number of acoustic dimensions. Figure 3.2 shows an F1-duration plane with the boundary line and midpoint, which is the most ambiguous F1-duration pair in the optimal categorization of the Spanish and SBE vowels.[24]

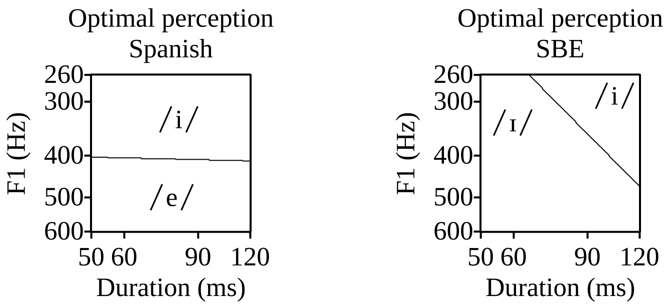

Fig. 3.2: Optimal perception of Spanish /i/ and /e/ and of SBE /i/ and /ɪ/.

The midpoint between the Spanish vowels is located at [407 Hz, 80 ms], while the SBE midpoint is located at [314 Hz, 79ms]. With the midpoint and the slope of the boundary line, we can compute the F1-duration points at which the boundary line crosses the edges of a given F1-duration plane, as seen in § 2.2.2. For instance, to compute the crossing point at the right edge of the SBE square, we only need to know its F1 location because it has already been determined that its location along the duration dimension (*x* axis) is 120 ms. As shown in § 2.2.2, the F1 distance between the boundary midpoint and the crossing point at the right edge of the SBE square is computed as the distance between the midpoint and the edge point along the duration dimension multiplied by the slope of the SBE boundary line[25]. This means that the optimal SBE boundary crosses the right edge of the square at [473 Hz, 120 ms]. For the other crossing point, we need to compute its duration location because this line starts at the top edge of the square, i.e., at 260 Hz (cf. Figure 3.2). Thus, the top-left crossing point for the optimal SBE boundary is located at

---

[24] The Spanish duration values are 81 ms. for /i/ and 78 ms. for /e/, as reported in Cervera et al. (2001).

[25] The F1 distance between the midpoint and the right edge is computed as log2 (120/79) · 0.98 = 0.591 octaves, and the F1 value at the right edge as log2 (314) + 0.591 = 8.886 octaves above 1 Hz = $2^{8.886}$ Hz = 473 Hz.

[260 Hz, 66 ms].[26] With respect to the optimal Spanish boundary, the same computations allow us to determine the locations of the right and left crossing points which are located at [403 Hz, 50 ms] and [411 Hz, 120 ms] respectively.[27]

Once we know these crossing points for the optimal boundaries, we can describe how the acoustic dimensions are used in the optimal vowel categorization of the two languages. To compute the use of an acoustic dimension in perception, i.e., its *cue reliance*, we need to know the categorization responses for the tokens delimited by a given acoustic space. For instance, the F1-acoustic space for the SBE and Spanish vowels is delimited by the squares of Figure 2 which have F1 values ranging from 260 to 500 Hz and duration values from 50 to 120 ms. To be able to count categorization responses, we need to divide the continuous acoustic space into discrete F1-duration points. Figure 3.3 shows a 7-point logarithmic division of the F1 and duration dimensions, which leads to 28 different F1-duration points at the square edges.

---

[26] The duration distance between the midpoint and the top-left edge of the optimal Southern boundary is computed as log2 (314/260) $\cdot$ 0.98 = 0.267 doublings, which means that the duration value of the top-left crossing point is log2 (79) - 0.267 = 6.037 doublings above 1 ms. or 66 ms.

[27] Assuming standard deviations of 0.2 octaves and 0.4 doublings (as in § 2.2), the Spanish boundary slope is (0.055/0.601) $\cdot$ $(0.2/0.4)^2$ = 0.023. The category boundary slope is computed as the ratio between the duration and F1 acoustic distances multiplied by the squared ratio between the F1 and duration standard deviations (cf. § 2.2.2). Thus, the F1 value for the crossing point at the left edge is log2 (407) – (log2 (80/50) $\cdot$ 0.023) = 8.653 octaves or 403 Hz and the crossing point at the right edge is log2 (407) + (log2 (120/80) $\cdot$ 0.023) = 8.682 octaves or 411 Hz.

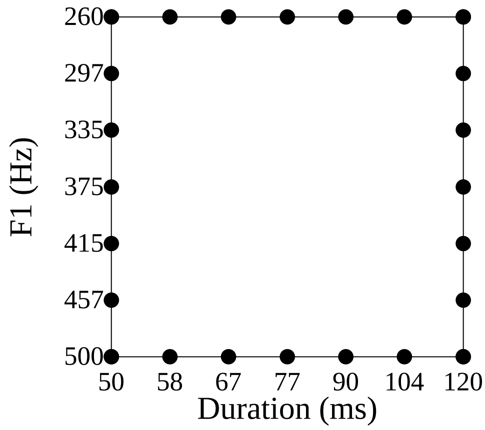

Fig. 3.3. The 28 tokens that result from a 7-point division of the F1-duration square together with the SBE and Spanish optimal category boundaries.

Cue reliance values can be obtained using the same type of analysis that was utilized in previous studies of L2 perceptual cue weighting (cf. Bohn 1995; Flege et al 1997). In these studies, the *duration reliance* was computed as the categorization percentage along the right edge of the acoustic square, i.e., the number of tokens categorized as one of the vowels divided by the number of stimuli points and minus the categorization percentage along the left edge. On the other hand, the *F1 reliance* was computed as the categorization percentage along the top edge minus the ones along the bottom edge. In the case of SBE and Spanish, we can compute the number of tokens that are categorized as /i/ along the edges by taking into account the location of the optimal boundary line at those edges, and then dividing it by 7, i.e., the number of dividing points. Figure 3.4 shows the number of /i/ responses if, for instance, 10 repetitions of each of the 28 edge points are presented as vowel tokens to the optimal listeners of the two languages. Note that the points without a response number have no /i/ responses.

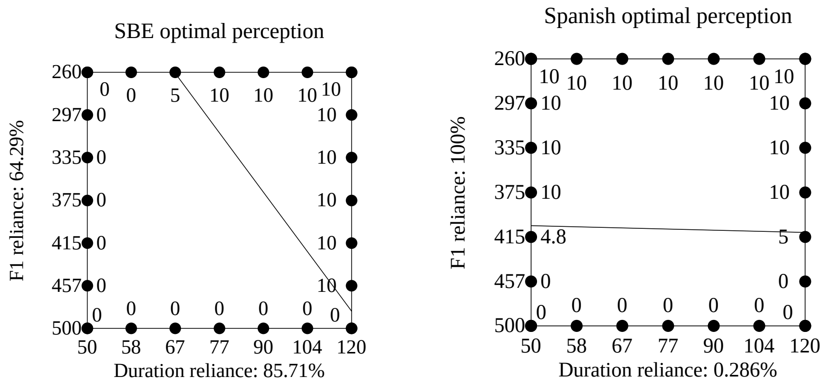

Fig. 3.4. Optimal /i/ categorization responses (inner numbers) and cue reliance.

As can be seen, the duration and F1 reliance for an optimal SBE listener are ((10+10+10+10+10+10= 60 /i/ responses) - 0 /i/ responses) / 7 = 85.71% and (45 /i/ responses - 0 /i/ responses)/ 7 = 64.29% respectively. For the optimal Spanish listener, the duration reliance is 0.286% and the F1 reliance is 100%. Thus, an optimal SBE listener uses the duration dimension 1.33 times more often than the F1 dimension when categorizing SBE /i/ and /ɪ/. With respect to the relative cue reliance in the categorization of Spanish /i/ and /e/, an optimal Spanish listener relies on the F1 dimension 286 times more than on the duration dimension.

Another possible computation for the optimal use of auditory dimensions is the specific way in which they are correlated to produce sound differences in a particular language environment. It turns out that there are cross-linguistic differences in the way in which individual acoustic dimensions combine and *trade*, a term used to define what happens to one of the cues if a change in the other occurs. This type of interaction between dimensions can be referred to as *cue trading* or *cue correlation*, and the optimal cue correlation in perception can be computed from the correlation of the acoustic dimensions in production.

3.1.3 The logical states of L2 sound perception and the L2LP model

Following the LP model, learning to perceive speech sounds in an L1 or in an L2 involves arriving at optimal perceptual mappings and phonological representations. Also, the acquisition of both L1 and L2 linguistic sound perception is characterized by the same logical stages. That is, both L1 and L2 acquisition have an initial state that commonly differs from the optimal perception, a developmental state in which the perceptual system approximates but is not identical to optimal perception, and an end state where the perception system somehow stabilizes. However, the properties of each state may differ in L1 and L2 acquisition. For instance, L2 learners have linguistic experience whereas L1 learners do not yet have a fully developed language system, and this certainly results in different initial states. I argue that an adequate and comprehensive model of L2 sound perception must be able to describe, explain, and predict the properties of the learner's system at each of the three logical states of L2 acquisition. Such a model should also be able to compare the logical states of L2 acquisition to those attested in L1 acquisition.

The L2LP model is composed of five theoretical ingredients that give an explicit prediction, linguistic explanation, and phonetic/phonological description of L2 sound perception at the three logical states of the acquisition process. Specifically, Ingredients 2, 3, and 5 contain the model's proposal for the initial, developmental, and end states in L2 sound perception respectively. As discussed in § 3.1.2, the model's Ingredient 1, namely the description of the optimal perception in the learner's L1 and target L2, is posited to directly enable the explanation of the initial and end states, which are the model's Ingredients 2 and 5. The remaining component, viz., Ingredient 4, refers to the explicit proposal for the L2 learning task. This ingredient does not directly address the logical states in L2 acquisition but instead portrays the L2 development that needs to occur for the attainment of optimal L2 sound perception. Figure 3.5 shows the sequential organisation of the L2LP theoretical ingredients for L2 sound perception.

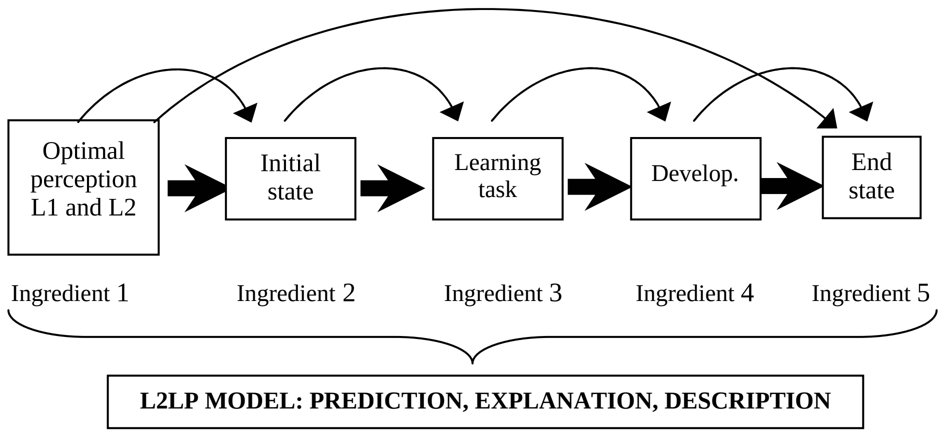

Fig. 3.5. The L2LP theoretical ingredients.

Here the solid arrows between the squared boxes represent the sequential nature of the model's ingredients and the curved arrows represent the relation between them, which means that they are not only sequentially organized but also in a mutually dependent way. The model's sequential organization implies that, for instance, the proposal for the initial state must follow ingredient 1 because only after we have described the optimal perception of the languages involved can we move on to describe, explain and predict the L2 process.

With respect to the particular relations between the L2LP's ingredients, Ingredient 1 directly feeds the proposal for two of the logical states of acquisition, viz., the initial and end states, as shown by the curved arrows in Figure 3.5. In § 3.1.2, it was also mentioned that the model's Ingredient 1 enabled the description of the L2 learning task. However, the relation between the first and third ingredients is mediated by the proposal for the initial state because the learning task (Ingredient 3) refers to how much the initial L2 (Ingredient 2) needs to develop to match the target L2 perception. Also, both the initial state and the learning task (Ingredients 2 and 3) are directly related to the developmental state (Ingredient 4) because they establish the starting point of the developmental path and the learning procedure that needs to take place for the attainment of optimal target L2 perception.

At the bottom of Figure 5, we see the three-way L2LP proposal for modelling L2 sound perception. First, the model considers the *prediction* of the perceptual behaviour within each of its ingredients. In other words, an explicit prediction of what perception looks like is formulated, an example of which would be 'L2 sound perception at the initial state will have *x* characteristics'. Second, the model provides a *linguistic explanation*, which means that its predictions always have an underlying explanation that is formalized through linguistic means. Third, the model provides a *phonological/phonetic description* of the predicted perceptual behaviour in that it describes the underlying optimal perception grammar and the perceptual phenomena that result from such a grammar, viz., optimal category boundaries and cue reliance (cf. § 3.1.2).

Next, Figure 3.6 shows a set of boxes with methodological phases, the purpose of which is to show that the model's components also represent a sequential order for conducting L2 sound perception experiments. The solid arrows represent the order in which experiments need to be conducted in order to both adequately investigate L2 sound perception and evaluate the model's theoretical proposal. I argue that this methodological proposal can be used for conducting empirical research independently of the theoretical proposal.

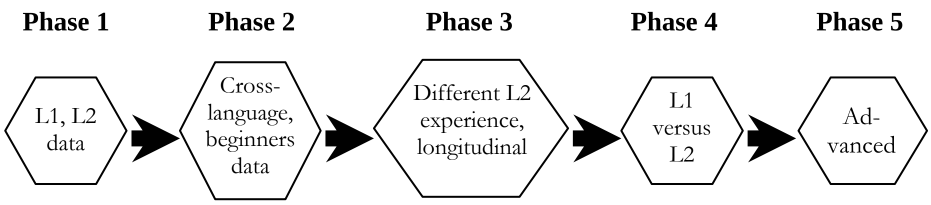

Fig. 3.6. The L2LP methodological proposal for conducting L2 sound perception studies.

To sum up, the L2LP model aims at providing a comprehensive description, explanation, and prediction of the acquisition process involved in L2 sound perception. The model's five sequential ingredients offer a proposal for each of the components of the L2 process. As a first element, such ingredients include the comparison of the adult optimal perception in the two languages involved, which is obtained through the computations of perceptual cue use demonstrated in the previous section. This first ingredient provides a window to the starting point and the learning task for L2 learners, and it connects the general LP model to its L2 version. In addition, the model's comprehensive L2 proposal comprises three main tenets that refer to L2 acquisition, as shown in Table 1. Finally, the model's Ingredient 3 gives an explicit proposal for the target of L2 development, i.e., the predicted learning task and its predicted solution.

3.2 L2LP ingredient 2: the L2 initial state

The second L2LP theoretical ingredient (cf. Figure 3.5) and methodological phase (cf. Figure 3.6) refer to the onset of learning and is based on Escudero & Boersma's (2004b) interpretation of L1 transfer (cf. 2004b: 579-580). Table 3.2 summarizes the model's three-way proposal for the initial state in L2 sound perception.

Table 3.2. The L2LP's three-way proposal for initial L2 sound perception.

L2LP model	Predicted perception	Linguistic explanation	Phonol. description
Ingredient 2: The L2 initial state	L2 initial equals cross-language perception	Full Copying hypothesis • Full Copying of L1 perceptual mappings • Phonemic equation of L2 and L1 sounds	L1 optimal boundaries and categories constitute the initial L2 perception

In § 3.2.1, I discuss the L2LP prediction which states that cross-language perception, i.e., the perception of a language by a listener with no prior knowledge of that language, constitutes the L2 initial state. In § 3.2.2, I discuss previous explanations that suggest a similar initial state for L2 acquisition. In § 3.2.3, I present the underlying explanation for predicted initial state. This is accompanied by a phonological description of the initial L2 perception, i.e., the perceptual mappings and phonological representations that result from the initial L2 perception grammar.

3.2.1 L2LP prediction: L2 initial equals cross-language perception

Speech perception research has shown that it is possible to measure how listeners perceive the sounds of a language that they do not know. This is called foreign, non-native, or cross-language perception. Within the L2LP model, it is hypothesized that cross-language perception represents the starting point for L2 learners. In other words, it is proposed that describing how optimal listeners of one language categorize the sounds of another language automatically provides a measure of the L2 initial state. This leads to the model's *cross-language equals initial L2* prediction which states that L2 learners will initially categorize the sounds of the target L2 in the same manner they categorize the sounds of any foreign language.

Cross-language perception can be described in the same way optimal perception was described in Chapter 2 and in § 3.1.1. Here I illustrate this mechanism with the perception of SBE /i/ and /ɪ/ by a monolingual optimal Spanish listener. Figure

3.7 shows the two-dimensional categorization of SBE /i/ and /ɪ/ by an optimal Spanish listener. Recall that the optimal Spanish listener perceives most tokens that fall above the category boundary line as Spanish /i/ and most tokens that fall below it as Spanish /e/.

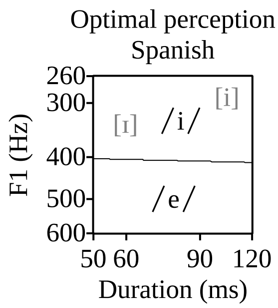

Fig. 3.7. Categorization of SBE /i/ and /ɪ/ by an optimal Spanish listener.

As can be seen, the location of the Spanish optimal boundary leads to the categorization of the two SBE vowels as a single Spanish vowel. This scenario is comparable to the *single-category assimilation* scenario considered in Best's (1995) Perceptual Assimilation Model which will be discussed in Chapter 4. However, not all cases of L2 learning have the same cross-language history. Given that optimal perception grammars can differ in various ways, the L2LP model acknowledges that cross-language perception and, consequently, the L2 initial state can exhibit different patterns depending on how the native language and the L2 compare to one another.

3.2.2 Background explanation: L1 Transfer

It has been widely observed that L2 sound perception is highly constrained by linguistic experience, i.e., by the sounds and perceptual processes of the native language, and a standard approach has been to describe, predict, and explain L2 perception by referring to the L1. For instance, early on Trubetzkoy (1939/1969) and Polivanov (1931) observed that L2 learners tend to associate the sounds of the new language to the sounds of their own system, and they regarded this association as the cause of their divergence from native speakers of the target L2. Early research on L2 acquisition also acknowledged that the L1 system plays a substantial role in the initial formation of the learner's new system or *interlanguage*, a term coined by Selinker (1972). Performance-oriented approaches such as the *Contrastive Analysis Hypothesis* (cf. Lado 1957) suggest that L1 habits are used in the process of learning a second language, and that these have negative or positive results depending on whether they are similar or different to the habits used in the target L2. The influence of the L1 system on L2 learning has been used as an important construct to explain the differences between L1 and L2 acquisition. Two widely used terms to describe this construct are *L1 transfer* and *cross-linguistic influence*.

However, despite the attraction of using a transfer explanation for phonology, especially given that nowhere else in the learner's L2 is L1 influence more obvious, the specific degree and nature of L1 transfer remains controversial (cf. Archibald & Young-Scholten 2003). An assumption of whether no transfer, partial transfer, or full transfer best represents the initial state of L2 learners will have a bearing on how we view the L2 learning task and development. It is important to consider that the concept of transfer can have different meanings, as suggested by Hammanberg (1997), so that it could refer to a learner's conscious or unconscious *strategy,* to the *proces*s of transferring L1 knowledge onto L2 learning, or to the *result* of such a process. However, most L2 proposals seem to combine all three possible interpretations in their use of the concept of L1 transfer.

3.2.3 L2LP explanation/description

The linguistic hypothesis of Full Transfer, as defined by Schwartz & Sprouse (1996), proposes that at the onset of L2 acquisition learners transfer L1 representations and their mappings to and from the speech signal. Escudero & Boersma (2004b) provide an interpretation of this hypothesis for sound perception, which states that the L1 abstract categories and perception grammar are transferred. Within the L2LP model, it is further proposed that L1 transfer results in the cognitive representation of a *copy* or *duplicate* of L1 perception that will henceforth constitute L2 perception. That is, the L2LP hypothesis of *Full Copying* constitutes a formal linguistic explanation for the prediction that L2 learners will initially manifest an L2 perception that matches their optimal L1 perception.[28]

---

[28] It is important to mention that L1 *copying* is considered to be an automatic and optimal initial strategy in L2 learning because starting out with L1 categories and L1 mappings gives the L2 learner a head start. This is because it is better to be able to perceive the number of categories allowed by L1 perception than no categories at all.

Moreover, the L2LP model claims that the Full Copying hypothesis can only be tested if the cross-language perception of the target L2 is compared to the L2 perception of L2 learners in an *absolute* beginning stage. In the next subsections, I discuss the L2LP description and explanation of how the L1 optimal perception grammar and the phonological representations are copied to the L2 initial state.

3.2.3.1 Full Copying of L1 perceptual mappings

As a result of the Full Copying, L2 learners will reuse their L1 perceptual mappings at the onset of their L2 acquisition process. For instance, Spanish learners of SBE /i/ and /ɪ/ will reuse the constraints that map the acoustic values of the English vowels onto sound representations. Figure 3.8 shows how the initial L2 grammar of Spanish learners maps the F1-duration pairs that are possible SBE /i/ and /ɪ/ tokens.

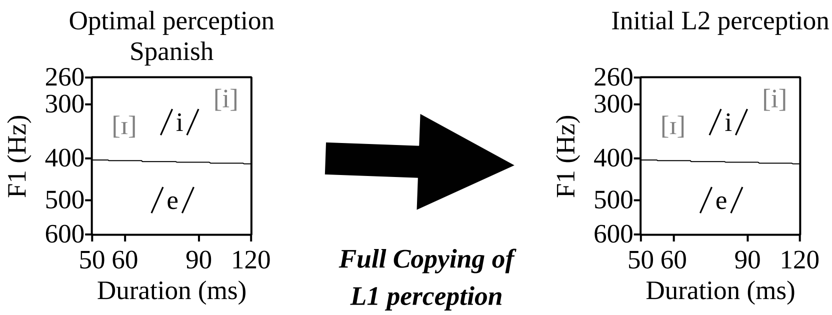

Fig. 3.8. L2LP model's initial state.

Here we observe that the copying of the optimal Spanish grammar results in the perception of two target L2 vowels as a single initial L2 vowel category. This is because the copied L1 grammar has mapping constraints that result in the categorization of the average production of SBE /i/ and /ɪ/ as the vowel category /i/. Also, the duration values with which the target L2 vowels are produced do not seem to result in any perceptual difference for a beginning Spanish learner of SBE. The next section presents a discussion of the status of vowel duration in the initial state for this learner.

3.2.3.2 Already-categorized versus non-previously categorized dimensions

Within the L2LP model, there are two types of perceptual mappings that can be copied to the L2 initial state, namely mappings that refer to *already-categorized* or *non-previously-categorized* dimensions in the learner's L1. In Spanish learners, the F1 continuum is an example of the former while vowel duration is an example of the latter. In both cases, the learner starts off with fewer categories than required to optimally perceive SBE high front vowels. However, in the already-categorized case, she does so with a single category whereas in the non-previously-categorized case there are *no* initial categories. For instance, Spanish learners will categorize English vowels /i/ and /ɪ/ as Spanish /i/ because of the F1 values of the English vowels. Other possible cases of already-categorized dimensions which lead to the single-category assimilation of L2 sounds are the perception of English /æ/ and /ɛ/ by Dutch learners and that of French /i/, /y/, /u/ by English learners, in which the already-categorized dimension is F2.[29]

With respect to Spanish learners of English vowels, it is of importance to note that English vowels also differ in vowel duration values. This means that during L1 SBE perceptual development, vowel height becomes integrated with vowel length to form vowel segments such as /i/ and /ɪ/. However, the Spanish perception grammar does not have constraints that map vowel duration to *length* categories. Escudero & Boersma (2004b) proposed that optimal Spanish listeners have no vowel duration categories because this is a 'blank slate' or 'uncategorized' dimension in Spanish. Consequently, vowel duration is encoded as a non-previously-categorized dimension with no constraints in the perception grammar. Another example of the non-previously-categorized dimension which results in a null L1 categorization are phonological tone differences for Spanish learners of Chinese because these learners do not posses any categorized dimension that can be matched to tone differences.

Therefore, the L2LP model predicts two different sub-scenarios in the NEW L2 sound perception scenario because their initial states will be different. That is, learners can either start with a single category along an already-categorized L1 dimension, or else start with no categories along a non-previously-categorized dimension. In § 3.3, it will be shown that these two different subscenarios result in differ-

[29] Vowel duration may be a case of already-categorized dimensions in Mandarin Chinese learners of English because, according to Flege et al. (1997), this dimension is mapped onto Chinese tone categories.

ent L2 learning tasks. However, Bohn (1995) suggested that vowel duration has the same status as F1 in Spanish learners because both F1 and vowel duration are already-categorized dimensions. Table 3.3 compares Bohn's view to the L2LP proposal.

Table 3.3. Two different views on vowel duration in Spanish listeners.

Vowel duration in Spanish listeners	Bohn (1995)	L2LP model
Status	Already-categorized	Non-previously-categorized
Number of categories	One	None

The L2LP model thus proposes that, for example, Spanish, French, and Brazilian Portuguese listeners, among others, do not have *one* vowel duration category, they have *no* vowel duration category at all. Empirical evidence in support of this L2LP assumption comes from the auditory perception of vowel duration by French listeners. Jaquemot et al. (2003) found that these listeners perceived the differences between /ebuuza/ and /ebuza/ as psychoacoustic, whereas they perceived the differences between /ebza/ and /ebuza/ as phonological. These authors' findings resulted from perceptual experiments as well as brain imaging studies. One must also consider that vowel duration categories are not universal because there are languages such as Estonian that have more than two such categories. Crucially, the use of this auditory dimension in perception varies among the languages that employ it to categorize vowels, e.g., Scottish English (SE) versus SBE or CE versus CF (cf. § 2.2).

In sum, this evidence would seem to support the *non-previously-categorized* status of vowel duration in languages that do not use this auditory dimension to signal vowel identity. It is proposed that the non-previously categorized dimension in Spanish listeners is part of the L1 optimal perception grammar that is copied to the L2 initial state. That is, Full Copying of *non-previously-categorized* dimensions predicts that, when listening to SBE vowels, Spanish learners at an absolute beginning stage will perceive the vowel duration values in a psychoacoustic/auditory way, and not as a mapping to a single discrete duration category, as Bohn (1995) seemingly assumes.

3.2.3.3 Phonemic equation and category re-use

As a consequence of an L2 initial state that is a duplicate of L1 perception, beginning L2 learners will be able to perceive only L1 perceptual categories when listening to the acoustic events in the L2 production environment. This means that L2 learners will equate L2 phonological representations or phonemes with the L1 categories onto which those L2 representations are initially mapped. For instance, Spanish learners of SBE should equate the target L2 vowels /i/ and /ɪ/ with a single L1 vowel, as shown in Figure 3.9.[30]

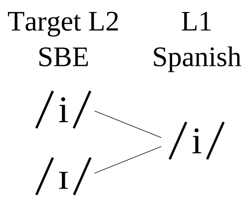

Fig. 3.9. Phonemic equation of SBE /i/ and /ɪ/ to Spanish /i/ by Spanish listeners and beginning learners of Southern English.

This phonemic equation has direct implications for lexical storage, i.e., for the representation of words in the mental lexicon in that learners are expected to reuse the equated L1-L2 phonological representations when storing L2 words. For instance, Spanish learners of SBE will initially store L2 words such as *sheep* and *ship* with the same phonological form, viz., /ʃip /. Importantly, the idea of category reuse is compatible with Flege's (1995) Speech Learning Model (SLM) which posits that the L2 initial state consists of L1 categories only.[31]

[30] When Full Copying and phonemic equation are combined, it can be predicted that beginning Spanish learners of English will initially phonemically equate the 5 Spanish vowels /i/, /e/, /a/, /o/, and /u/ to the 12 English vowels. Therefore, they will use only those five vowels to store English words.

[31] This and other models of L2 sound perception will be discussed in Chapter 4.

3.3 Ingredient 3: The L2 learning task

The third ingredient of the L2LP model refers to the L2 learning task that results from the differences between the initial L2 and the target L2 perception. That is, the L2LP not only provides an explicit proposal for the logical states of L2 sound perception but also postulates a connection between the initial state, the development, and the end state in the acquisition process. This is because describing, explaining, and predicting the L2 learning task allows us to establish the target of development as well as the possible learning mechanisms involved in the process of attaining optimal target L2 perception. It is proposed that an exhaustive investigation of the cross-linguistic differences between the optimal L1 and target L2 perceptions provides an adequate description of the L2 learning task, given that cross-language perception is identical to the L2 initial state, as predicted by the Full Copying hypothesis. In this section, I demonstrate how we can describe the perceptual and representational tasks that L2 learners face when learning to perceive the sounds of a new language. Table 3.4 presents a summary of the L2LP's three-way proposal for the learning task in L2 sound perception.

Table 3.4. The L2LP learning task.

Ingredient 3:	Predicted perception	Explanation	Description
L2LP learning task	L2 learning task equals cross-language difference	L2 learning involves approximating the optimal target L2 perception	The cross-language difference results from comparing the optimal L1 and L2 perceptions

3.3.1 Prediction: learning task equals cross-language difference

Given the predicted L2 initial state, which is a copy of the optimal L1 perception, L2 learners are likely to have problems perceiving L2 sounds. As shown in § 3.1, even in those instances where the L1 and L2 have the same number of phonological representations, it is quite common to observe *cross-language perceptual differences*, as in the case of Canadian English and Canadian French /æ/ and /ɛ/. It is predicted that the *degree of mismatch* between perception grammars will constitute the L2 learning task. Thus, learners will exhibit different levels of non-optimal perception depending on the degree of perceptual mismatch between the languages. A non-optimal L2 perception grammar will result in category boundary mismatches as well as mismatches in the number of phonological representations.

The L2LP model predicts that the L2 learning task will be equal to the differences between the L1 optimal perception and the target L2 perception. It is proposed that a rigorous computation of the differences between the learner's initial state, i.e., the optimal L1 perception, and the optimal target L2 perception will allow us to describe, explain, and predict the learning tasks that L2 sound perception requires. It is argued that two types of cross-language differences are at play, namely perceptual and representational differences, and that these result in two different learning tasks. In order to measure the differences between the languages involved, we first need to establish the region in the acoustic space that leads to perceptual mismatches. Figure 3.10 shows the cross-language mismatch (which is represented as the grey region in the square on the right) between the L1 optimal perception and the optimal target L2 perception for CF learners of CE.

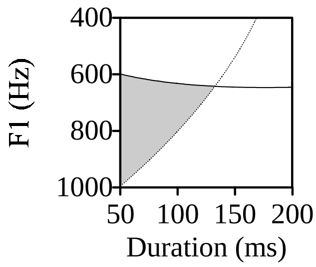

Fig. 3.10. Cross-language perceptual difference and region of perceptual mismatch (in grey) for CF learners of CE. Quasi horizontal curve: Optimal CF perceptual boundary. Diagonal curve: Optimal target L2 boundary.

Thus, the tokens in that region will be categorized as /æ/ by an optimal CF listener, and therefore by the beginning CF learner of CE, but as /ɛ/ by the target L2 optimal listener. This means that there is a difference in the mapping of F1 values as well as in the mapping of duration values. That is, assuming optimal perception and Full Copying, CF beginning learners of CE are predicted to miscategorize the CE tokens that fall in the region of mismatch. As we will see, the mismatch between any two languages may be due to differences in perceptual mappings, differences in the number of sound representations, or both.

3.3.2 Explanation/description: perceptual and representational tasks

The reason why L2 learners are faced with a learning task is evident: they need to be able to cope with the new production environment optimally well. In other words, they should change their L2 initial perception to match the optimal target L2 perception. To that end, they need to be able to bridge the cross-language differences between their initial L2 and the target L2. The L2LP model assumes a principled distinction between perceptual mappings and phonological representations, and therefore the model's description for the L2 learning task also incorporates such a distinction. Thus, it is argued that there are two types of learning tasks in L2 sound perception, namely a *perceptual task* and a *representational task*.

3.3.2.1 L2LP perceptual task: Changing and creating mappings

The learning task for beginning CF learners of CE involves a boundary shift and the categorization of the vowel duration dimension. This is because they need to change the shape of their L1 boundary from a quasi-horizontal one to a diagonal target L2 boundary, as shown in Figure 3.10 above. This multidimensional adjustment in perceptual boundaries will be first performed through the categorization of F1 values between 600 and 780 Hz into the learners' L2 /ɛ/ category, and through the creation of vowel length categories. In addition, the learners will need to integrate the two dimensions to finally attain a diagonal perceptual boundary.

In other L2 sound perception cases, the perceptual mismatch may be different from this case so that the L2 initial grammar may need to be adjusted in different ways. For instance, a small perceptual mismatch can be found in the case of Spanish learners of SE /i/ and /ɪ/, while as will be described in Chapter 7. A large perceptual mismatch can be found in Spanish learners of SBE /i/ and /ɪ/ as well as SBE learners of Spanish /i/ and /e/, as will be discussed in Chapters 5 and 6 respectively. It is proposed that in cases of a large cross-language perceptual difference the L2 development will involve a perceptual and a representational task, while cases of a small cross-language difference will involve a perceptual task only. The latter is because a small perceptual difference is commonly accompanied by the same number of categories in the L1 and the target L2.

Following the hypothesis of Full Copying of blank slates, the L2LP model proposes that when the target L2 produces sounds with auditory dimensions that were not previously categorized in the learner's L1 perception grammar, her perceptual task is to create new mappings to cope with the new production distributions. Figure 3.11 shows the L2LP proposed learning task for a case in which the cross-language difference involves a non-previously-categorized dimension in the learner's L1, i.e., vowel duration for Spanish learners of SBE high front vowels.

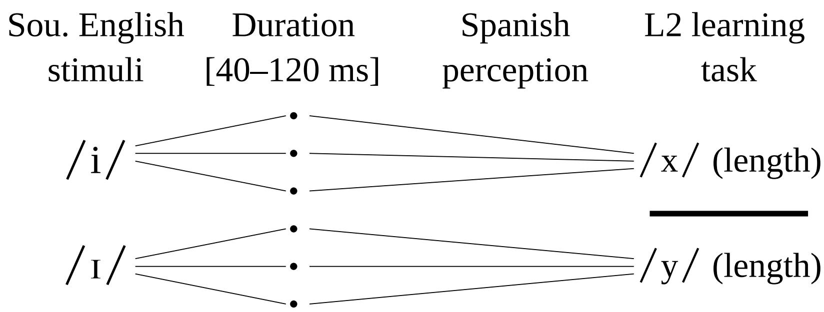

Fig. 3.11. Perceptual task (non-previously-categorized): *Creating* new perceptual mappings.

Following the hypothesis of Full Copying of L1 perceptual mappings, the L2LP model proposes that when the target L2 produces sounds with already-categorized dimensions in the learner's L1 perception grammar, her perceptual task is to generate extra categories from those that already exist through the *redistribution* or *splitting* of L1 perceptual mappings. Figure 3.12 shows the L2LP proposed learning task for a case in which the cross-language difference involves an already-categorized dimension in the learner's L1, i.e., F1 or /height/ for Spanish learners of SBE. In the figure, the thick solid line that runs horizontally from the middle of the F1 values to the right represents the split of the Spanish category /i/ along the height continuum. In addition, the dotted lines represent the redistribution of F1 values.

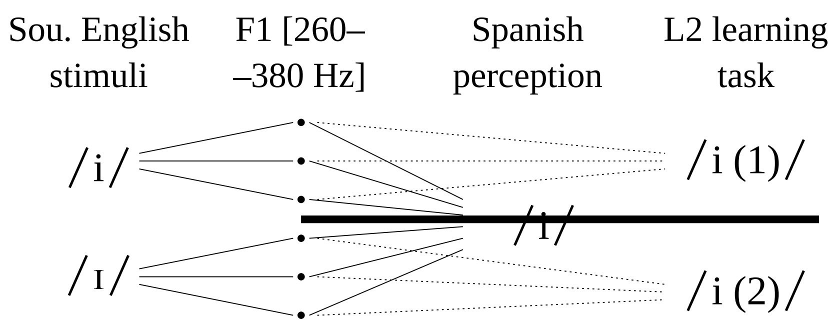

Fig. 3.12. Perceptual task (already-categorized): *Splitting* L1 perceptual mappings.

3.3.2.2 L2 representational task: Changing the number of L2 categories

As for the representational task, a Spanish learner of SBE will initially perceive the same vowel in words like *sheep* and *ship,* and therefore her L2 lexicon will contain the same lexical representation for the two words. Thus, she will start with a non-optimal phonological lexicon that leads to non-optimal word recognition. Consequently, she will be led to assume that tokens of the two words differ semantically but not phonologically, i.e., that they are *homophones*, because her recognition grammar will tell her that two semantic categories are linked to the same phonological form /ʃip/. The logical way of solving this representational problem would be to have a new category available for one of the SBE vowels. This non-optimal lexical storage will leave the learner with the sole option of relying on the semantic or pragmatic context to access the correct meaning.

The question at this point is how a new L2 category becomes available. It is proposed that in order to use two different words in her lexicon, the L2 learner first needs to perceive the difference between the L2 vowels. In the previous section, we saw that Spanish learners of SBE can accomplish their representational L2 task by creating new vowel duration mappings because this dimension is a non-previously-categorized one.

3.4 Ingredient 4: L2 development

The L2LP model proposes that in order to accomplish the L2 learning task, the learner will either need to create new perceptual mappings that will lead to new phonological representations or adjust the existing perceptual mappings. Most importantly, it is argued that the mechanisms underlying the learning of optimal L1 linguistic perception will also underlie L2 linguistic perception, as explained in § 3.4.1. Table 3.5 summarizes the L2LP proposal for how learners perform their learning task in L2 sound perception.

Table 3.5. L2LP three-way proposal for the development of L2 sound perception

Ingredient 4:	Predicted perception	Explanation	Description
L2LP development	L2 perceptual development equals L1 perceptual development	Full GLA Access: L1-like development • Auditory-driven category formation • Lexicon-driven constraint re-rankings	• Category learning: New acoustic dimensions lead to new mappings & categories • Perceptual learning: Re-rankings & boundary shifts

Due to the Full Copying of her L1 perception, the beginning L2 learner will not be able to optimally cope with the L2 if the copied perception differs from the optimal perception of the new production environment. To become an optimal L2 perceiver, the beginning learner will need to acquire the number of L2 categories that optimal listeners posses. Following the hypothesis of Full Access, as defined by Schwartz & Sprouse (1996:41) it is proposed that L2 learners have access to L2 development and are able to become optimal perceivers. Importantly, the claim made by these authors to the effect that restructuring draws from options of Universal Grammar (UG) implies that both the creation of new L2 sound representations and the adjustment of the L2 initial perception grammar must go through developmental stages that are found in the development of L1 perception. This was also claimed in Escudero & Boersma (2004b). Thus, it is hypothesized that L2 learners have automatic access to the L1-like learning device, the GLA, that allowed them to acquire their L1 perception. This device is also available for L2 acquisition so that learners may gradually adjust their L2 perception to match that of optimal listeners of the new production environment. This means that L2 learners will thus create new categories and adjust their category boundaries in the same way infants and children do so in their L1 perception.

3.4.1 L2LP prediction: L2 development equals L1 development

The L2LP model predicts that L2 learners will develop following the same learning mechanisms used in the L1 acquisition of optimal sound perception. That is, L2 learners will learn to categorize new sounds through distributional learning, and they will also come to adjust their perceptual mappings to match those of the target L2 with the help of their lexical representations. In the next section, I discuss some of the general L2 issues that are related to this prediction.

3.4.2 Background explanation: access to development and learning mechanisms

The question of how L2 learners develop from an initial state of L1 transfer to more closely approximate native-like knowledge and performance leads to the notion of *learnability constraints* in L2 acquisition. If second languages can be learned irrespective of how much the learner approximates native-like performance, some kind of a learning mechanism must be in place. Following a generativist perspective of language learning, researchers have proposed that L2 learning is also guided by the principles that compose UG, which are deemed to be innate and specific to the faculty of language. In a UG approach, then, it seems quite natural to postulate some degree of access to such principles if L2 learners develop, and researchers such as Schwartz & Sprouse (1996) and White (2000) have in fact proposed the possibility of *full access* to UG in L2 acquisition.

However, as White (2003) rightly points out, not only does the learner need a set of universal restrictions with respect to the grammars that she may develop, she also needs a learning device. From a *nativist* or UG perspective, such a device will function as a triggering or accessing device. In the realm of current phonological theory, learning devices in the form of algorithms have been proposed to change the rules or rerank the constraints that constitute the developing grammar. An example of such a learning algorithm within a nativist or UG perspective is Tesar & Smolensky's (2000) *Constraint Demotion Algorithm* which works within an OT description of phonological knowledge. Alternatively, an *emergentist* and *domain general* perspective presupposes that learning occurs in a bottom-up fashion in that the learner's knowledge of an L2 changes with environmental input by means of a general-cognition learning device in the form of a learning algorithm. In Chapter 4, I will discuss how learnability assumptions shape the proposals of L2 perception, and I will assess their predictions about the levels of proficiency that L2 learners can achieve.

3.4.3 L2LP explanation/description: Full Access to the GLA

The GLA is a blind device which does not involve any complex learning mechanisms involving general cognition or meta-linguistic knowledge. Neither does it apply exclusively to linguistic input. With respect to speech comprehension, the GLA is only responsible for the types of learning mechanisms that occur in L1 learning, viz., category formation, category boundary shifts, and recognition learning (cf. § 2.3 and § 2.4). It is proposed here that an L2 learner does not need other types of high-level complex strategies when learning to perceive and recognize L2 sounds. Rather, she will develop into an optimal target L2 listener through the L1-like GLA learning mechanisms that lead to the creation of new L2 categories, to lexicon-driven L2 perceptual learning, and to message-driven recognition learning.

3.4.3.1 GLA category formation in L2 development

As a consequence of L1-like development, it is predicted that L2 learners will form categories along dimensions that have never been used to classify sounds in their L1. This is because distributional learning or, as it is called in Boersma, Escudero & Hayes (2003), auditory-driven learning, only applies to non-previously-categorized dimensions such as vowel duration for learners with no L1 vowel length distinctions. Thus, L2 learners will create auditory mappings for new dimensions that have two distributions in the target L2.

For instance, following Full Access to GLA auditory learning, the durational distributions of SBE /i/ and /ɪ/ will lead to a categorization of the vowel duration continuum into two vowel length categories. This will happen in the same way as the auditory-driven learning described in § 3.3 for L1 learners. Thus, a Spanish learner of SBE will create perceptual mappings that link vowel duration values to two newly created vowel length categories, namely /short/ and /long/. Tableau 3.3

shows how a Spanish learner of SBE will now perceive as /i, short/ the vowel token [349 Hz, 74 ms] she initially categorized as /i/.

[349 Hz, 74 ms]	349 Hz not /e/	74 ms not /long/	349 Hz not /i/	74 ms not /short/
/i, long/		*!	*	
☞ /i, short/			*	*
/e, long/	*!	*		
/e, short/	*!			*

Tableau 3.3. Categorization of a typical SBE /ɪ/ token by a Spanish learner who has created a length contrast.

3.4.3.2 GLA category boundary shifts in L2 development

L2 learners also have access to the GLA constraint reranking that allows for the adjustment of L2 category boundaries. Thus, for Spanish learners of SBE, this category boundary shifting will be guided by the newly created lexical categories /short/ and /long/, just as in L1 acquisition (cf. § 3.3). Category boundary shifts have actually been attested in L2 sound perception. For instance, Caramazza et al. (1973) reported that native speakers of French who had begun to acquire English before their seventh birthday exhibited a large shift of the /b/-/p/ boundary, while Flege & Eefting (1987) showed that the perceptual /b/-/p/ boundary of Dutch learners of English depends on the language the learners think they hear. However, these two studies also reported that L2 learners do not exhibit category boundary shifts that lead to an optimal L2 sound perception. In the next section, I present the L2LP proposal for the end state in L2 sound perception which predicts that L2 learners will shift their L2 category boundaries to match those of the optimal target L2 listener.

3.5 Ingredient 5: the L2 end state

Now that we know how learners approach the task of becoming optimal L2 listeners, the next question is whether they actually reach optimal target L2 perception.

This is especially interesting because the L2 acquisition literature has extensively shown that L2 learners are not likely to reach their target of having native-like proficiency, and that they stop their learning process without fully accomplishing the L2 learning task. This phenomenon is called *fossilization* and it commonly refers to only one state in L2 proficiency. On the other hand, the term *end state* is used to describe a number of L2 proficiency levels (cf. Klein 1986: 51) that may or may not match the target language.

Given that most examples of L2 fossilization come from the L2 production literature, such a non-target like end state in L2 sound perception patterns has yet to be established. With respect to the L2LP proposal for the end state in L2 sound perception, Table 3.6 presents a summary of the three-way proposal that allows the model to predict that L2 learners can achieve optimal target L2 perception.

Table 3.6. The L2LP end state

Ingredient 5:	Predicted perception	Explanation	Description
L2LP End state	1. L2 end state = Optimal target L2, if rich L2 input	1. Rich L2 input outweighs small cognitive plasticity	1.1 Plasticity = GLA learning step 1.2 Rich input = L1-like
	2. L1 perception remains optimal	2. Separate perception grammars hypothesis	2. Language mode activation hypothesis

3.5.1 L2LP prediction: optimal L2 and optimal L1

The proposal of two separate grammars has crucial implications in L2 development as well as in L1 optimal perception. For both L1 and L2 categorization to be optimal and for the two languages not to influence each other in their representations, they must be separate systems. If two separate systems underlie the perception of two languages, it is proposed that L2 development need not affect the already optimal L1 perception provided that sufficient input for both languages is received. In other words, the current L1 model proposes that learners who use their two languages to similar extents will exhibit L2 development as well as L1 stability. It is argued that any intermediate L1-L2 perception performance is the consequence of activating the two perception grammars at the same time. Thus, one would expect that an L2 learner's linguistic perceptions will resemble those of the monolingual speakers of each of her two languages.

3.5.2 Background explanation: limitations for the L2 end state

Within L2 research, it is commonly assumed that the main cause for there being a *plateau* in L2 development, i.e., fossilization, is the lack of cognitive plasticity and/or sufficient linguistic input. Importantly, the level of interrelation between the L1 and the L2 results in specific predictions for the L2 end state and for the effects of L2 development on the L1. In this section, I discuss the current state of affairs of these issues in the field of L2 acquisition.

3.5.2.1 The role of cognitive plasticity and the L2 input

Researchers have always observed that adults are not nearly as efficient L2 learners as children. As a result, most explanations for this lack of language learning efficiency in adults have been based on biological and neurological limitations. For instance, Penfield & Roberts (1959) claimed that an innate biological clock for language learning allows direct learning from the input until approximately the age of nine whereupon acquisition begins to result in poorer attainment levels. Likewise, Lenneberg (1967) suggested that this loss of predisposition for language learning has a biological basis since it is due to the completion of hemispheric lateralization around puberty. He labelled the period between two years of age and puberty the *critical period* for language acquisition. This idea led him to formulate the well-known *Critical Period Hypothesis* (CPH) which claims that only before puberty can learners acquire an L2 from mere exposure to the input without conscious and laboured effort. Other less radical notions have also been proposed such as the *sensitive period* (cf. Long 1990).

However, despite the obvious need for postulating age-related constraints on language acquisition, it seems controversial to claim that these types of constraints act alone. Specifically, it may be the case that social, psychological, input, and language-use factors could be correlated with the observed decline in language acquisition capabilities. Moreover, there may be no categorical loss of language acquisition abilities at a specific age limit, but rather a continuous decrease of the probability of mastering an L2 with a native-like level of proficiency. Indeed, this is what Hyltenstam & Abrahamsson's (2003) recent review and proposal on maturational constraints in L2 acquisition suggest. These authors claim that there is a continuous *maturational period* that predicts that acquisition will be increasingly difficult with age. However, they remain neutral with respect to the exact levels of L2 attainment because other non-maturational constraints can influence the end result (cf. pp. 575-6).

3.5.2.2 The interrelation between the L1 and the L2

If the L1 and L2 sound categories and perceptual processes are represented as knowledge in our minds, the next natural question is how these two systems relate to each other. Presumably they both belong to our linguistic faculty, but do L2 learners have one or two perception grammars? The degree of integration or separation assumed between the L2 learner's phonological systems influences the level of perceptual proficiency that she can have in both languages. Following Francis (1999), Cook (2002) proposes that there are three logical possibilities for the representational status of two or more language systems. Figure 3.13 is an adaptation of Cook's schema which includes an extra division for *mixed representations* (which Francis called integrated representations).

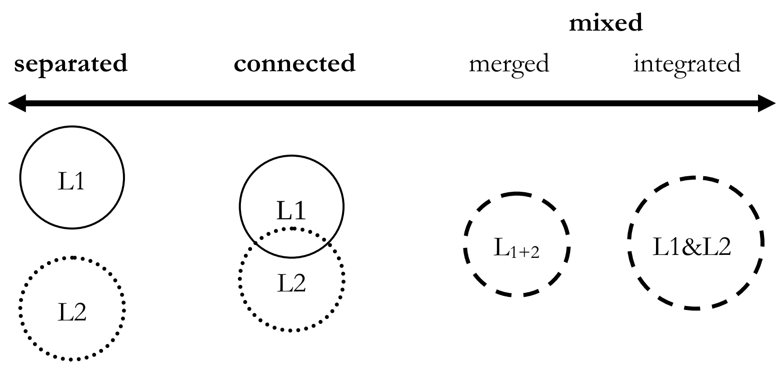

Fig. 3.13. Possible cognitive status of sound categories and perception processes in L2 learners and bilingual speakers, adapted form Cook (2002).

It is a matter of debate as to which of these possibilities best describes the knowledge or performance of L2 and bilingual speakers, and we will see that assuming any one of them crucially shapes the differing views of L2 perception to be discussed in the next section. On the one hand, in a *separate* systems outlook, L1 and L2 sound categories are considered to belong to autonomous systems. On the other hand, the *mixed* view advocates that L1 and L2 sound systems are, in reality, a single representational system. In turn, this last perspective comprises two possibilities, viz., *merged* systems which imply no language differentiation, and *integrated* systems which entail language specification within a single space. In a less extreme perspective, such as the *connected* view, L1 and L2 representations are mostly distinct but may share some elements or properties, as shown by the intersection in Figure 3.14. In Chapter 4, I will explain how other L2 sound perception models interpret the L2 constraints presented here and then compare their views to the L2LP proposal outlined below.

3.5.3 L2LP explanation/description: Input versus plasticity

In this section, I present the L2LP's interpretation of cognitive plasticity, L2 input, and the representation of L1 and L2 knowledge. This interpretation is used to explain the L2LP prediction that claims that L2 learners can achieve optimal L2 perception and at the same time maintain their optimal L1 perception.

3.5.3.1 Rich L2 input overrules small cognitive plasticity

The L2LP model interprets the continuously decreasing ability to learn languages as the size of the learning step that the GLA takes when performing auditory or error-driven perceptual learning. In the model's computer simulations (cf. Escudero & Boersma 2003, 2004b), the plasticity of a virtual infant learner is always set to a value that allows for large learning steps, which means that an L1 learner will learn fast but will also make more mistakes. In contrast, the plasticity of a virtual adult learner is set to a much smaller value so that she will be much slower. However, small plasticity results in more accuracy, and therefore it is predicted that adult learners will be more accurate than infant learners who take large learning steps. Thus, depending on the L2 initial state, an adult L2 learner could potentially outperform an infant L1 learner at some developmental stages. This adult advantage is found in simulated Spanish learners of SE (cf. Escudero & Boersma 2003) and simulated Dutch learners of Spanish (cf. Boersma & Escudero 2004) but it would need to be tested by comparing real child learners to adults.

On the other hand, the *input* factor in language acquisition refers to the linguistic evidence needed for learning to occur. It seems that L1 learners only need *positive evidence*, i.e., environmental speech, in order to master their native language, while L2 learners seem to need *negative evidence*, i.e., corrections or specific instruction, to acquire an L2. However, it is evident that the type, quality, and quantity of input differ between L1 and L2 acquisition, especially when it relates to perceptual learning. Within the Native Language Magnet model (cf. Kuhl 2000: 11855), it is argued that L2 learners need the 'right' kind of perceptual input, namely enhanced acoustic cues, multiple instances of the same sound, and massive listening experience to learn to perceive L2 sounds. Crucially, this appropriate input resembles the features of child-directed speech (also called *motherese*) which is abundant in the input to L1 perceptual learning. I argue that the end states of L1 and L2 sound perception can only be compared if the two acquisition processes involve the same type, quality, and quantity of phonetic input. The L2LP model incorporates this crucial input factor because the learning algorithm associated with the model, i.e., the GLA, depends on auditory input in order to change the developing perception grammar. Importantly, the simulations of L2 learning shown in Boersma & Escudero (2004) suggest that the more input the L2 learner gets, the more likely the optimal target L2 perception will be attained because the result of the simulations depends on the type and quality of input fed to the GLA.

In sum, it is argued that the role of the input is more important than that of cognitive plasticity. This is because rich L2 input can outweigh the reduced level of plasticity in adult learners. Given that L2 learners are unlikely to get the same type of input as L1 learners, an end state that does not match the optimal target L2 perception cannot be ascribed to maturational constraints that lead to lack of cognitive plasticity until we can be sure that these learners have been exposed to a rich production environment.

3.5.3.2 The hypothesis of separate perception grammars

It follows from the Full Copying hypothesis that L1 and L2 perception constitute two separate systems. In other words, Full Copying makes explicit the notion of transfer by proposing that a copy or duplicate of the L1 is created because once the L1 knowledge is transferred, it becomes a separate system. As mentioned above, the L2 perception system is initially equal to that of the L1 but it later follows its own developmental path based on the L2 production environment. That is, L1 and L2 perception are represented as two separate sets of sound categories and perception grammars which handle each of their two production environments independently. The question now is whether there is empirical evidence to support these hypothesized separate linguistic perceptions for L2 learners.

To test the separate perception grammars hypothesis, we need to construct a falsifiable null hypothesis which, in this case, is that L2 learners possess a single perception grammar. Perception experiments such as those of Caramazza et al. (1973), Elman et al 1977, and Flege & Eefting (1988) would seem to indicate that bilinguals and L2 learners do have a single perception grammar because they exhibit category boundaries that are situated at intermediate locations when compared to monolingual L1 and L2 boundary locations. This type of evidence would seem to disconfirm the separate perception grammars hypothesis. However, the L2LP model explains these intermediate results through the notion of *language modes*. This notion is found in Grosjean (2001) who refers to a continuum ranging from a purely monolingual mode in which only one system is in use to a completely bilingual mode in which both systems are fully operative. Figure 3.14 shows a possible representation of his language mode continuum, one which evinces a bilingual mode in the middle of two monolingual modes, i.e., the bilingual's L1 and L2 modes.

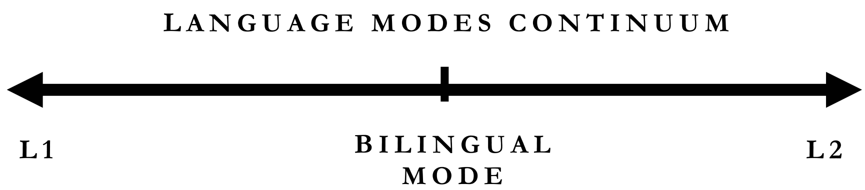

Fig. 3.14. The bilingual's language continuum.

Within the L2LP model, Grosjean's proposal is interpreted as suggesting that L2 learners and bilinguals have different language modes that can be active during on-line linguistic perception. Thus, the degree of activation of each perception grammar depends on the amount of evidence provided by *language setting variable*s, such as the language of the instruction, the language of the stimuli, the experimenter's language, the language of the response categories, the language the task requires, etc. This L2LP interpretation of Grosjean's language mode continuum can be referred to as the *language mode activation hypothesis*. According to this hypothesis, the parallel activation of separate systems can lead to intermediate responses because the output of the two perception grammars will merge when a categorization response is given.[32] That is, intermediate L1-L2 sound perception is a consequence of the gradient and parallel activation of the learner's two perception grammars during online perception, rather than being a consequence of having a single set of categories and perceptual mappings for the two languages. Thus, the L2LP model predicts that L2 learners have two different perception grammars and that any intermediate perception is the result of parallel activation during online speech perception. Figure 3.15 shows three different hypothetical language modes which are predicted to occur depending on whether the learners are in a monolingual L1 mode, a completely bilingual more, or a monolingual L2 mode.

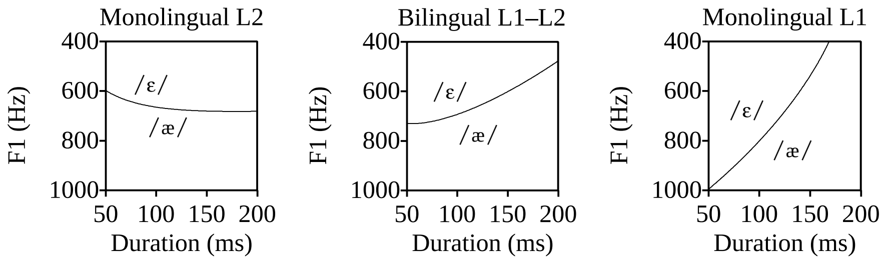

Fig. 3.15. Predicted perception in advanced Canadian French learners of Canadian English when confronted with three different settings for language mode activation.

Following more general psycholinguistic proposals for language activation in L2 learners and bilinguals such as those in Dijkstra, Graingeer & Van Heuven (1999), Jared & Kroll (2001), and Kroll & Sunderman (2003), it is suggested here that there may be a 'race' between the two perception grammars during online linguistic perception to output a perceived sound category. Under task demands, the learner may merge the information given by the two grammars when performing a sound categorization test. However, a bilingual mode or the simultaneous activation of

[32]This is the same proposal found in the psycholinguistic model called Merge (cf. Norris et al. 2000) which was mentioned in § 1.3.2.

two languages is limited by the amount of evidence in favour of activating a single grammar, which is referred to as *language control* (cf. Green 1988). Language control or deactivation during speech perception means that the L1 or the L2 can be deactivated or inhibited appropriately, depending on the given language setting variables.

As shown in Figure 3.16, the L2LP model predicts that advanced L2 learners in an L2 monolingual mode will exhibit an L2 perception similar to that of monolingual native listeners. This means that no fossilization in L2 sound perception and no attrition in L1 sound perception will be attested. In other words, under the proper circumstances, L2 sound perception can develop to reach the optimal target L2 perception level. This L2 development will occur without affecting the optimal L1 perception which will remain stable if the learner is exposed to sufficient L1 input. Importantly, the L2LP model predicts that the speed and path of development will be different depending on the specific L2 perception task the learner needs to face, as described in the summary below.

3.6 Summary and L2LP sound perception scenarios

A summary of the L2LP's five ingredients for predicting, explaining, and describing L2 sound perception appears in Table 3.7. Recall that three of these ingredients refer to the three main elements of L2 acquisition.

Table 3.7. A summary of the L2LP's theoretical ingredients.

L2LP	Prediction	Explanation	Description
Optimal L1 & target L2	Human beings are optimal listeners	Optimal listeners handle the environment maximally well	L1 and L2 optimal category boundaries: Location & shape
Initial state	= Cross-language perception	Full Copying	L1 boundary location and shape
Learning task	= Reach the optimal target L2 perception	L2	Bridging mismatches between L1 and target optimal perception
Development	= L1-like	Full GLA Access	Category formation and boundary shifts
End state	Optimal L1 perception and optimal L2 perception	Input overrules plasticity Separate grammars	Language activation modes, through language setting variables

The L2LP model predicts that L2 learners will start with a copy of their optimal L1 perception and will have access to L1-like learning mechanisms in developing into optimal L2 listeners. With respect to development, the L2LP's hypothesis of Full GLA Access states that L2 learners will perform their learning task through auditory-driven category formation and lexicon-driven category boundary shifting. These are the same mechanisms L1 learners use to arrive at optimal L1 perception (cf. § 2.3). As for the end state of L2 sound perception, the model's hypotheses on separate perception grammars and language activation modes predict that L2 learners will reach optimal L2 perception and that, at the same time, their L1 perception will remain optimal.

3.6.1 Learning scenarios: L2LP prediction/explanation

Most explanations of L2 sound perception acknowledge that L2 acquisition will be different depending on how the learner's L1 compares to the target L2. They usually presuppose two possible scenarios, namely the acquisition of NEW and SIMILAR sounds(cf. Flege 1987). In the first scenario, the learner usually perceives fewer sounds than the ones produced in the target language because her L1 has fewer sound categories than the ones found in the L2. The most common case of this type involves the perception of two L2 sounds as a single L1 sound, which results in an L2 contrast neutralization as in the case of Japanese learners of English /r/ and /l/.

In the second scenario, the learner usually perceives the same number of sounds as those produced in the target language because her L1 has the same number of sound categories. The most common case here is the perception of an L2 contrast as a corresponding contrast in the L1. This is where phonetic differences between the L1 and L2 sound categories result in slight differences in sound categorization, as in the case of French learners of English /d/ and /t/.

The L2LP model predicts that the L2 initial state will exhibit different kinds of mismatches when compared to the optimal target L2 perception, and that therefore different L2 learning tasks will arise. This is because L2 sound perception cases exhibit diverse cross-language perceptual differences/mismatches with respect to the languages involved. For instance, Spanish learners who start with a copy of their L1 perception grammar will face different problems depending on how it compares to the optimal perception of different target L2s. There are at least three such scenarios that can be attested in L2 sound perception.

The first of these occurs if the L1 perception grammar outputs fewer perceptual categories than those required to optimally perceive the target L2. This is called the NEW scenario because the L2 environment produces phonological differences that do not exist in the L1. If the L1 perception grammar outputs more categories than required in the L2, the learner faces a SUBSET scenario because the L2 categories constitute a subset of her L1 categories. It is worth emphasizing that this second scenario has not been considered in previous L2 perception models which commonly consider only the first two scenarios. Finally, if the L1 perception grammar outputs the same number of categories than output by the optimal target L2 grammar, the learner faces a SIMILAR scenario because the L1 and L2 categories are phonologically equivalent. Figure 3.17 exemplifies the three predicted L2 scenarios

narios that are posited to emerge as a consequence of Full Copying and phonemic equation.[33]

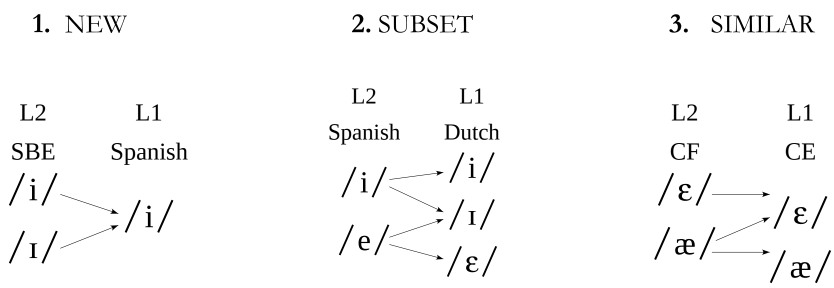

Fig. 3.17. Three scenarios for L2 sound perception. Southern British English (SBE), Canadian English and Canadian French (CE and CF).

Within the NEW scenario, there are two different types of sub-scenarios, namely the association of new L2 sounds to already-categorized dimensions and their association to non-previously categorized dimensions. As described in § 3.2.3.2, the F1 continuum is an example of an already-categorized dimension for Spanish learners whereas, according to the L2LP model, vowel duration is a non-previously-categorized dimension. These two subscenarios share the main characteristic of the NEW scenario in that the L1 categories are fewer in number than the target L2 categories. However, one of the two L2 sounds will already be present in the L1 in the already-categorized case, whereas none of the L2 sounds will be present in the L1 in the non-previously-categorized case. Thus, these two sub-scenarios of the NEW L2 sound perception scenario will result in different perceptual tasks that aim at changing the number of L2 categories.

3.6.2 Scenarios: L2LP description of the different learning tasks

The L2LP model offers specific predictions for the learning task in each scenario along with specific predictions of the learner's performance and underlying knowl-

[33] Although I am aware that other scenarios or even variants of these three scenarios may occur, I will limit the scope of this study to an in-depth discussion of the three predicted scenarios which I consider to be the most common ones. Importantly, the discussion of the NEW scenario in Chapter 5 will be limited to the non-previously-categorized sub-scenario for the prediction of L2 development and end state.

edge ranging from her initial state to her end state. Also, the model explains the different L2 performances in each of the scenarios by proposing that specific learning mechanisms are at play. As a consequence, different learning tasks will be attested in each scenario, as shown in Table 3.8.

Table 3.8. Predicted initial states and learning tasks for the three L2LP scenarios.

	NEW	**SUBSET**	**SIMILAR**
Initial state	Too few categories	Too many categories	Non-optimal mappings
Perceptual task	Two tasks: Creation and integration	One task: Category boundary shift	One task: Category boundary shift
Representational task	Two tasks: Create features and turn them into segments	Two tasks: Reduce lexical and perceived categories	none
Relative difficulty	Most difficult	Medium difficulty	Less difficult

As can be seen, the NEW and SUBSET scenarios, in which learners start with a different number of sound categories than the ones perceived by the optimal target L2 listener, are predicted to have two L2 learning tasks, namely a perceptual and representational task. However, the representational task in these two scenarios is postulated to be different. In contrast, the SIMILAR scenario, in which learners start with the same number of perceived categories, only involves a perceptual L2 learning task. The model predicts that the number and nature of the learning tasks to be performed defines the relative level of difficulty between scenarios, as shown in Table 8. The NEW scenario is predicted to be more difficult than the SUBSET scenario because of the nature of the tasks involved. This is mainly because the NEW scenario not only involves the *creation* of new categories and perceptual mappings but also the integration of the new categorized dimensions with already categorized dimensions.

In the following chapters, I will show how the L2LP three-way proposal and its five theoretical ingredients can be applied to the specific characteristics of each of the three scenarios that the model presupposes. These chapters will also present empirical data which follow from the model's methodological proposal of sequential experimental phases. However, before presenting the detailed predictions for each L2 sound perception scenario, the next chapter examines previous L2 perception models and compares them to the L2LP model proposed here.

4 A review of other L2 sound perception models

In this chapter, I will review five models of L2 sound perception that precede the L2LP model presented in the previous chapter. Historically, they constitute the most influential and promising models of L2 sound perception. The discussion of these models will be divided in two sections. First, in § 4.1 and § 4.2, I will discuss the models' frameworks for modelling sound perception and its L1 acquisition, and compare them to the L2LP model's theoretical framework. Then, in § 4.3, I will review the models' proposals for L2 sound perception with respect to the three logical states in language acquisition, viz., the initial state, the developmental state, and the end state. The predictions and explanations found in the five models for each state in L2 sound perception will be compared to the L2LP proposal. Finally, in § 4.4, a summary of the proposals as well as their comparative advantages and limitations will be presented, and it will argued that, in most instances, the L2LP model synthesizes and improves on the proposals put forth by all five previous models.

4.1 Aim and scope of five L2 perception models

Despite some convergence, the five models discussed here rely either on phonological or on phonetic frameworks. Therefore, they differ in their approach to L2 sound perception and in the claims that are made. The phonological approach to L2 segments is represented by Brown's Phonological Interference Model (PIM) and Major's Ontogeny Phylogeny Model (OPM). The PIM presented in Brown (1998, 2000) aims at explaining the origin of the influence of L1 phonology on the acquisition of L2 segments as well as identifying the level of phonological knowledge involved in such L1 influence. The OPM as outlined in Major, (2000, 2001, 2002) endeavours to describe the principles involved in the formation of L2 phonological systems, the change in an L1 that results from exposure to an L2, and language contact phenomena such as bilingualism and multilingualism. As for the scope of these two models, the OPM explicitly refers to the three logical states in L2 acquisition, while the PIM indirectly refers to those three states because the influence of the L1 can, in principle, be described and explained within each of these three states.

In the realm of phonetics, the three most influential models that aim at explaining L2 sound perception are Best's Perceptual Assimilation Model (PAM), Kuhl's

Native Language Magnet (NLM) model, and Flege's Speech Learning Model (SLM). The PAM seeks to account for the observed performance in the perception of diverse non-native contrasts. The NLM model attempts to explain the development of speech perception from infancy to adulthood. As for the SLM, it has been primarily concerned with the ultimate attainment of L2 production (cf. Flege 1995: 238) though it has recently begun to show an interest in the ultimate attainment of L2 perception (cf. Flege 2003). As for the scope of these phonetic models, the PAM and the NLM model are mainly interested in explaining the L2 initial state through non-native perception but they still offer suggestions as to how the L2 development and end state can be achieved. The SLM mainly deals with the end state but its claims regarding why L2 learners may not have a native-like end state are directly connected to an explanation of the initial and developmental states in L2 sound perception.

4.2 Speech perception and its acquisition

In general, phonological approaches try to account for L2 acquisition by assuming that learners have a formal knowledge that underlies their observable linguistic behaviour or performance. They base this assumption on the general proposal of generative linguistics (cf. Chomsky 1957) that performance is not always equal to competence because it can be constrained by non-linguistic factors that may be sociological or psychological in nature. The knowledge that underlies performance in the area of segmental phonology can be viewed as a system of structures (rules, features, hierarchies, or constraints) that is represented in learners' minds. Also, phonological proposals consider 'distinctive' segments or phonological features to be the units of analysis for describing phonological systems.

Phonetic approaches do not rely on abstract systems that shape the learner's performance. Instead, they consider the actual phonetic components of the acoustic signal to be stored in a somewhat abstract way in the learner's long-term memory. Thus, a certain level of abstraction is presupposed but it relates to the speech signal more straightforwardly and it is constituted by the actual sound categories. Moreover, there is no proposal for another device that performs the connection between the speech signal and those categories. However, within recent phonetic approaches, there has also been an interest in discovering the brain functionality behind the processing of sound categories. These proposals deal mainly with 'position-dependent allophones' (cf. Flege 1995, Rochet 1995) that are called 'phonetic' categories.

4.2.1 Speech perception in phonological models of L2 sound perception

The OPM is purposely not presented in terms of any formal linguistic framework, thus supposedly enabling it to survive any possible outdating or refutation of current theories (cf. Major 2002: 88). As a result, explicit proposals for sound representation, the mapping of the acoustic signal onto representations, or the exact workings of the phonological system are not presented. Major (2002) states that his model makes very general claims which do not necessitate any details concerning specific phenomena such as fine-grained phonetics. He argues that this is a virtue rather than a weakness because the OPM provides a macroscopic framework for testing individual phenomena. Consequently, there is no clear proposal for the nature of either language-specific or universal sound perception, or for any of its elements whether they are solely phonological representations or perceptual mappings too. As stated in Chapters 1 and 2, only an explicit proposal of how the speech signal is mapped onto phonological representations can be considered a proper framework for describing and explaining sound perception and its acquisition. Although this is clearly not the aim of the OPM, I believe that a comprehensive model should always aim at being as explicit as possible.

As for Brown's PIM, it assumes the traditional distinction between universal phonetics and language-specific phonology. This distinction is used to represent sound perception as two different levels in the perceptual mapping, as shown in (4.1).

(4.1) PIM's proposal for sound perception

Acoustic signal ⇒ Universal Phonetic Categories → Feature Geometry → Phonemes

Here the first arrow corresponds to the perceptual mapping of the speech signal onto discrete categories, a process which is commonly known as speech perception. According to the PIM, this procedure is non-linguistic and common to all human listeners because it is the automatic result of the auditory processing performed by the human ear, as depicted by a double arrow. Therefore, this model provides no formal phonological description or explanation for this initial mapping of the signal. However, in Chapter 2 it was argued that the mapping of the speech signal is likely to be performed as a single procedure that connects the signal with language-specific phonological representations.

The second arrow in (4.1) goes from the universal phonetic categories to the feature geometry because Brown proposes that the listener's phonological structure turns universal categories into phonological structure. In the non-linear phonological framework of the PIM, which is Feature Geometry (cf. Clements 1985), it is assumed that phonemes have an internal structure composed of a hierarchy of phonological features that are contained in the phonological component of Universal Grammar (UG). The representation of a given sound segment in a given language is a subset of universal feature geometry. In other words, the representation of a sound always makes it distinctive within a particular language following Rice & Avery's (1991) proposal of Minimally Contrastive Underspecification.

4.2.2 Speech perception within phonetic models of L2 perception

The PAM proposes that adult listeners have no mental representations or mental perceptual mappings for sound perception, and that they directly seek and extract the invariants of articulatory gestures and gestural constellations. This proposal is based on the frameworks of Articulatory Phonology (cf. Browman & Goldstein 1989) and the ecological approach to speech perception, also called direct realism (cf. Best 1984, Fowler 1986). The gestures that we perceive are complex articulatory events specified by higher order invariants in the signal, e.g., bilabial, stop, high, front, etc. This direct-realist and non-mentalist proposal contrasts with the mentalist and abstract speech-perception proposals found in Major's OPM and Brown's PIM. With respect to the status of perception, the PAM explicitly proposes that sound perception is language specific because it is the product of perceptual learning. This means that listeners can only efficiently pick up gestural invariants of their native language environment.

Turning to the NLM model, it argues that complex neural perceptual maps underlie sound perception and that such neural mappings result in a set of abstract phonetic categories. Adult perception is seen as language specific because it is shaped by earlier linguistic experience (cf. Kuhl 2000: 11854). Unlike the PAM proposal, the NLM claims that perceptual representations are stored in memory. Perceptual mappings differ substantially for speakers of different languages so that the appropriate perception of one's primary language is completely different from that required for other languages (cf. Iverson & Kuhl 1995, Iverson & Kuhl 1996). Kuhl emphasizes that perception is language specific, claiming that "no speaker of any language perceives acoustic reality; in each case, perception is altered in service of language" (2000: 11852).

Within the SLM, sound perception is defined as the discrimination of the phonetic features or properties in the signal in order to identify the appropriate *position-defined allophones* or *phonetic categories* which are stored in long-term memory. Perceptual mappings and sound categories seem to be conflated because the stored categories are seen to perform the mapping of the speech signal. Therefore, the model does not provide an explicit proposal for how phonetic discrimination or processing works. Neither does it explain how the degree of *perceived phonetic distance* can be measured, although it suggests possibilities such as auditory, gestural, or phonological metrics. As in the PAM and NLM models, perception in the SLM is assumed to be language specific because there are cross-linguistic differences in the processing of phonetic features or properties, and also because L2 perception problems do not have a general auditory basis, as noted in Flege (1995: 266).

4.2.3 L1 acquisition within the five models

The two phonological models, namely the OPM and the PIM, propose that language acquisition is aided by *innate linguistic universals*. The OPM argues that the set of universals with which a child is born includes not only (some version of) UG but also learnability principles, markedness, underlying representations, rules, processes, constraints, and stylistic universals. The PIM holds that a child starts out with a universal feature geometry which is provided by UG and expanded over the course of acquisition until the adult feature geometry for a particular language is acquired (cf. Brown & Matthews 1997). In this model, L1 development is guided by the particular dependency and constituency relations encoded in UG, i.e., its phonological component that is represented by a feature geometry (cf. Brown 1998: 144) as well as by a child's detection of the contrastive use of features in the input (cf. Rice & Avery 1995). However, it is not clear exactly how a child would 'detect' a contrast in the input so that any additional structure could be incorporated into her developing phonological grammar.

On the other hand, the phonetic models assume that the emergence of language-specific sound perception is the result of exposure to a particular language environment. In other words, infants are not born with innate language categories or perceptual mappings. The PAM proposes that in the beginning, infants hear and detect every articulatory gesture, and that later on they learn to detect only high-level features, i.e., those that signal sound contrasts in their native language. Once a child is able to process the high-level features that form the phonological system of her language, the task of perceiving L1 sounds becomes easier and more adult-like. However, the learning mechanism that allows the infant to 'pick up' high level articulatory gestures is left undefined.

Kuhl's NLM model does consider the mechanisms underlying the learning of language-specific perception, and so proposes that the child engages in learning processes that lead to the emergence of a speech perception system and perceptual representations. In Kuhl (2000), she puts forward a body of evidence (cf. Godsitt, Morgan & Kuhl 1993, Saffran, Aslin & Newport 1996) showing that infants acquire sophisticated information from the signal through the detection of the distributional and probabilistic properties of the ambient language. In sum, it is argued that infants' perception becomes language specific through the categorization, statistical processing, and perceptual warping of acoustic dimensions, all of which take place within their first year.

The SLM assumes the same learning processes and mechanisms as those proposed by the NLM model, viz., the ability to accurately perceive featural patterns in the input and to categorize a wide range of segments (cf. Flege 2003). However, no formal proposal for the mechanisms behind the learning of L1 perception can be associated with this model, apart perhaps from the claim that perception is dominated by 'equivalence classification', a mechanism that leads to the classification of acoustically different tokens into the same abstract category (cf. Flege 1987, 1995).

4.2.4 Comparison with the L2LP's framework model

Table 4.1 summarizes the framework that previous models have used to describe and explain speech perception and its acquisition. It also describes how their assumptions compare to the LP model which underlies the L2LP framework.

Table 4.1. Speech perception and its acquisition in models of L2 sound perception.

L2 Models	Speech perception framework: Nature	Speech perception framework: Elements	Proposal for L1 acquisition: Initial state	Proposal for L1 acquisition: development
OPM	Not explicit	Not explicit	U = universal principles only	U = Decreases L1 = Increases
PhIM	First mapping of the signal is extra-linguistic	Auditory mappings, universal categories	UG: universal feature geometry	Detection of contrasts
L2LP	Linguistic, single mapping acoustic-to-phonological	Linguistic mappings, phonological categories	Auditory perception	Distributional learning, and then lexicon-driven learning
PAM	Language-specific	Articulatory gestures	Articulatory perception	Pick up high-level articulatory gestures
NLM	Language-specific	Neural maps, phonetic categories	Auditory perception	Categorization, distributions, warping
SLM	Language-specific	Phonetic categories, phonetic discerning	Auditory perception	Categorization, equivalence classification

Following the criteria for the comprehensive modelling of speech perception that were set down in § 1.5, and which state that perception involves abstract phonological representations and perceptual mappings, I claim that the only proposal capable of modelling speech perception and its acquisition in an explanatorily adequate way is the LP model. This claim ensues from the following considerations.

As shown in Table 4.1, Brown's PIM and Flege's SLM do not model the difference between perceptual mappings and sound representations because in both cases the phonological structure performs the categorization of the signal without there being a separate mapping procedure.[34] Although the SLM suggests that sound perception refers to the extraction of phonetic information, this is performed by sound categories and not by perceptual mappings. In contrast, Best's PAM provides a proposal for how the speech signal is mapped onto higher-level units based on the theoretical framework of Articulatory Phonology (cf. Browman & Goldstein 1989). Although this represents an alternative to the LP's mental representations and auditory mappings, the organization of articulatory features and their mapping mechanism are not articulated. Neither is the way in which listeners remember features and high-level units if they are not represented in their minds.

The closest to an adequate proposal for speech perception is found in the neural mappings that are posited in Kuhl's NLM model whose organization and mapping mechanism have been modelled by means of connectionist neural networks (cf. Guenther & Gjaja 1997). The NLM model also assumes a set of abstract phonetic categories that result from the neural mappings. However, the LP model constitutes the linguistic version of the NLM model, and that it therefore brings speech perception, a process commonly described within neural, phonetic, or psycholinguistic modelling, into the domain of phonological theory.

As for the L1 acquisition of speech perception, I argue that the proposal based on the LP model (cf. Boersma, Escudero & Hayes 2003) provides a phonological formalization of the mechanisms underlying this process. The learning mechanisms performed by the model's GLA (cf. § 2.3.2), i.e., auditory-driven (or distributional) learning and learning guided by lexical knowledge, provide a formal phonological account of the mechanisms of underlying category formation and boundary shifts because they apply in the context of a perception grammar. One of the only other models that explicitly refer to the mechanisms underlying the learning of language specific perception is Kuhl's NLM model. This model assumes that categorization, distributional learning, and perceptual warping of acoustic dimensions result in perceptual development between the ages of 6 and 12 months. On the other hand, the SLM assumes similar learning mechanisms as the NLM, namely category formation and equivalence classification, which lead to classifying acoustically differ-

[34] The OPM as found in Major (2000, 2001) does not provide an explicit proposal for describing or explaining sound perception.

ent tokens as the same abstract category (cf. Flege 1987, 1995). However, as mentioned before, the LP model is alone in providing a linguistic formalization of the NLM and SLM learning mechanisms.

4.3 L2 sound perception

In this section, I review the various models' proposals for L2 sound perception with respect to the three logical states in language acquisition, viz., the initial, developmental, and end states, which all five models address in one way or another. I then compare them to the L2LP model's predictions, explanations, and descriptions.

4.3.1 L2 initial state

Both phonetic and phonological models assume that the initial state in L2 sound perception is shaped by the learners' previous linguistic experience. However, they differ in the exact details and level of elaboration of their proposals for this first state in language acquisition, as described below.

4.3.1.1 Major's OPM and Brown's PhIM

Within the OPM, the learner's system is seen as an *interlanguage* (IL) which which has three components, elements, viz., the L1, the L2, and universal factors, as shown in (4.2). The L1 component is by definition one's native language (NL) while the L2 component is comprised of the external linguistic influences on one's NL. The L1 and L2 interactions in the L2 learner are affected by U which, as shown in Table 1, is the set of universal principles governing language acquisition and change.

(4.2) Major's proposal for the components of the learner's interlanguage (IL):
IL = parts of L1 + parts of L2 + parts of U = 100%

This model proposes that, initially, the ideal learner has only the L1 because there is no L2 yet and, in addition, U is dormant, i.e., there is no observable component of U that is not already part of the L1 grammar (cf. Major 2002: 69). Only the L1 can be observed at the beginning because its existence prevents U from surfacing. The *L1 transfer* component makes the learner start with a 100% L1 and 0% for both L2 and U. Thus, the OPM presupposes that the strong role of L1 transfer follows from research that demonstrates that one relies on existing cognitive structures when learning new ones. In L2 phonology, this means that L1 transfer will dominate the initial state. It is claimed that the very notion of a heavy foreign accent constitutes strong support for the hypothesis of L1 transfer even without having to rely on empirical grounds because one can easily identify the speaker's L1 when she speaks the L2. Therefore, transfer is considered to be active (cf. Major 2002: 71).

Brown's PIM proposes that the L1 feature geometry, or L1 phonological structure, filters the L2 input and eliminates the ability to perceive cues in the acoustic signal that could trigger L2 acquisition. The model claims that the L2 initial state is the L1 phonological structure, which means that L1 transfer is assumed to occur. That is, the L2 learner transfers her L1 feature geometry to her IL. It seems that the PIM, unlike the OPM, assumes that L1 transfer occurs only once at the initial state, which makes it similar to the L2LP model's Full Copying hypothesis.

4.3.1.2 PAM, NLM, and SLM

The PAM claims that the high-level features and categories of the L1 are used to handle new language environments. The attunement to the language specific organization of articulatory events leads to a lack of efficiency in finding familiar gestural invariants in non-native speech. This in turn leads to the model's central premise which is that listeners *assimilate* non-native sounds to the native sounds they perceive as most similar. The model defines 'perceptual similarity' in terms of dynamic articulatory information, i.e., the ways in which articulatory gestures shape the speech signal. It proposes that accuracy in the discrimination of non-native sounds depends on the way they are assimilated to the L1 sounds. L2 speakers have already tuned their linguistic perceptual device to particular high-level features and will therefore have difficulty in detecting the features in the new language. In other words, the target language (TL) may have high-level features that signal contrasts that are actually low-level ones for the L2 speakers. Consequently, the TL sounds will get assimilated to the ones in the L1.

As for the NLM, it proposes that the L1 language-specific filter will make the acquisition an L2 much more difficult because future learning is constrained by the initial mental mappings that have committed neural structure. Therefore, learning to perceive L2 sounds is constrained by the initial mapping, i.e., the native-language sound mapping, that has taken place. Moreover, this constraint operates independently of any critical period. However, it is still claimed that the older the learner is, the more neural commitment she has to the native language mappings. The native-language mental maps thus interfere with the creation of new mappings for the new language input.

The SLM claims that an L2 learner starts with her L1 categories, L1 sub-segmental properties, and L1 properties. This initial state may lead to a failure to discriminate the phonetic differences between pairs of L2 sounds or between L2 and L1 sounds. It is argued that learners relate L2 sounds to L1 positional allophones such that L2 perceptual failure happens because the L1 phonology filters out L2 sound features or properties. However, the exact operation of this perceptual filtering is not accounted for.

4.3.1.3 Comparison with the L2LP initial state

Table 4.2 summarizes the various models' proposals for the L2 initial state as compared to the L2LP (in grey).

Table 4.2. Proposal for the L2 initial state in six models of L2 sound perception.

L2 models	Explanation	Connection with cross-language perception	Prediction
OPM	Transfer	Not explicit	L2 initial = L1
PhIM	L1 Transfer	Not explicit	L2 initial = L1 feature geometry
L2LP	Full Copying	Yes	L1 grammar & L1 categories
PAM	Linguistic experience	Yes	L1 assimilation
NLM	Linguistic experience	Yes	L1 maps & L1 categories
SLM	Linguistic experience	Not explicit	L1 categories

What this shows is that the L2LP model proposes the Full Copying hypothesis (cf. § 3.2.3) which is a re-interpretation of the Full Transfer/Full Access hypothesis proposed for syntax by Schwartz & Sprouse (1996). According to Full Copying, the learner automatically uses her entire L1 perception grammar and the L1 categories that result from it when starting to acquire an L2 (cf. § 3.2). Crucially, the L2LP model suggests that the Full Copying hypothesis can only be tested if the cross-language or non-native perception of the target L2 is compared to the L2 perception of absolute L2 beginners. For instance, the cross-language perception of Southern British English vowels by Spanish listeners must be compared to the L2 perception of beginning Spanish learners of Southern British English. This cross-language connection is not found in the SLM, OPM, or PIMs which all fail to account for the perception of beginning learners or the non-native perception of monolingual listeners.

In addition, the L2LP Full Copying hypothesis underlies the NLM's L1 neural filter, the SLM's equivalence classification, and the PAM's perceptual assimilation, all of which result from the listeners' linguistic experience and their categorization of L2 sounds. This linguistic hypothesis is also compatible with the OPM and PIM proposal that the L1 grammar or the L1 feature geometry is the starting point of L2 acquisition. However, the L2LP model not only describes the L2 initial state but can also predicts it through its explicit connection of cross-language perception with initial L2 perception (cf. § 3.1 and § 3.2).

4.3.2 L2 development

All the models claim that L2 learners have access to development, which means that none of them assumes the complete loss of plasticity for adult L2 learners or a critical period for language acquisition. Also, they presuppose that learners develop so as to approximate the target L2 performance. Here I review the models' specific proposals for how L2 learners turn their initial state, which is dominated by their L1 perception, into a more target-like state.

4.3.2.1 OPM and PIM's developmental proposals

As mentioned in the previous section, the OPM proposes that learners start with their L1 phonological knowledge, their L2 being nonexistent and their U dormant. It predicts that L2 development occurs through the decrease of the L1 component, the increase of the L2 component, and the increase of the U component. The growth of U in the learner's interlanguage is evidenced by the fact that their speech is neither like the L1 or the L2. However, the model does not demonstrate that *universal* patterns cannot be related to either L1 or L2, or to both. Also, the model predicts that with more L2 proficiency, the U component will decrease and the learner's forms will be replaced by L2 forms (cf. § 3.3.4 for different developments depending on the L2 scenario). Major argues that Hancin-Bhatt & Bhatt (1997: 386) provide empirical evidence and theoretical support for the OPM because their Optimality-Theoretic account of the data shows that L2 learners start with mostly transfer-related errors. Later on, they show developmental errors and when they become more advanced, they have neither of the previous two types of errors and have a more native-like performance.

Brown's PIM proposes that the L2 learner develops by means of a mechanism she terms 'redeployment' which involves reusing L1 features in order to distinguish

L2 contrasts. This way, the learner can only distinguish a non-native contrast if the pertinent distinctive feature already exists in the L1 feature geometry. Once the contrastive use of the non-native segments is detected, the L1 feature can be reused or redeployed to arrive at the appropriate L2 phonological structure and phonemes. In support of her redeployment hypothesis for L2 development, Brown tested the perception of English /r/ and /l/, a contrast which is specified by the feature [coronal],[35] in Japanese and Chinese listeners. Unlike Japanese, Chinese also has a [coronal] contrast in the form of alveolar versus retroflex sibilants. Brown's (1998: 155-170) results showed that Chinese learners were able to accurately perceive the English contrast whereas Japanese learners were not. However, given that the cross-language perception of monolingual Chinese and Japanese listeners was not tested her results might be interpreted as showing that Chinese subjects map the English consonants onto two different L1 categories whereas the Japanese do so onto only one. Therefore, we cannot know whether the Chinese learners already associated the two English sounds to two L1 categories, and were thus already optimal English listeners before learning the language.

4.3.2.2 PAM, NLM, and SLM's developmental proposals

Best's PAM (1995) suggests that at some point in the L2 developmental process, learners will be able to perceive a non-native contrast by *splitting* their L1 categories. Also, Best & Strange (1992) suggest that exposure to L2 input may lead to the reorganization of assimilation patterns in cross-language perception. However, the learning mechanisms that allow the splitting of categories or the reorganization of assimilation patterns are not provided.

Although the NLM model proposes that L2 learners can create new mappings for the perception of L2 sounds, it is not clear whether the creation of such new mappings is achieved through the same means as in L1 acquisition or through some other mechanisms. Kuhl (2000: 11856) suggests that the creation of L2 mappings differs from that which occurs during L1 acquisition, and that therefore other ways of achieving development may be needed. However, no other types of learning mechanisms are proposed.

---

[35] In Brown's view, the two English phonemes have different internal structures, i.e., different feature geometries. This difference is based on the feature [coronal] which is present in the featural representation of /r/ but absent in that of /l/.

As for the SLM, it concentrates on the end state and therefore does not elaborate on the other two states in L2 sound perception. With respect to development, this model presupposes that adults retain the capacity used by infants and children to acquire their L1, including the learning of an accurate perception of the properties of L2 speech sounds and the formation of new phonetic categories (cf. Flege & MacKay 2004). The basic claim is that L2 learners develop by creating categories so that development depends on the learners' discrimination of the phonetic differences between L2 sounds, or between L2 and L1 sounds, depending on the degree of perceived cross-language phonetic similarity. Thus, the greater the perceived phonetic dissimilarity is between an L2 sound and the closest L1 sound, the more likely it is that a new L2 category will be formed (cf. Flege 2003). For instance, Flege (1987) attempted to show that native English learners of French could produce /y/ more accurately than /u/ because /y/ is auditorily more distant from French /u/ than from English /u/. However, no cross-language data were gathered to support the model's claims (cf. § 4.3.4 for a further discussion of the model's learning scenarios).

4.3.2.3 Comparison with the L2LP developmental state

Table 4.3 compares the L2LP proposal for L2 development with those of the other models that we have examined.

Table 4.3. Proposal for the L2 development in six models of L2 sound perception.

L2 models	Explanation	Learning mechanisms	Prediction
OPM	Access to U	Not explicit	L2 is acquired
PhIM	Access to development	Redeployment of L1 features	L1 structure reused for L2 phonemes
L2LP	Full GLA Access: L1-like development	Auditory-driven Lexicon-driven	Category formation, mapping adjustment
PAM	Access to develop-ment	Not explicit	Category split Reor-ganization
NLM	Access to develop-ment	Different from L1	Creation of L2 maps & categories
SLM	Access to develop-ment	Not explicit	Category formation, category merging

As can be seen, the OPM assumes access to U since the IL frequently evinces phenomena that are neither L1 nor L2. However, many of these can be attributed to the influence of the L2 on the L1, i.e., to starting with the L1 and then changing towards the L2, as was shown by the copying of non-previously-categorized dimensions in § 3.2. More importantly, the OPM does not propose any learning mechanisms that might trigger the increase and decrease of the L2 components, nor does it offer an account of the role of the input in this development.

On the other hand, the L2LP model's explicit learning mechanisms as applied to phonological perception grammars constitute the formalization and implementation of the predicted development of L2 sound perception (cf. § 3.4). In other words, auditory-driven learning underlies the category and mapping creations mentioned within the NLM model and the SLM.

4.3.3 L2 end state

In this section, I review the models' proposals for the end state in L2 sound perception. That is, I examine their explanations of the extent to which L2 learners develop as well as their predictions of what the final state in L2 acquisition looks like. These claims refer to the role of the L1, the L2 input, the age factor (for adult learners), and the interrelation between the L1 and L2 systems in the attainment of native-like perception.

The OPM provides an explanation for the phoneme boundaries in bilingual perception that Caramazza et al. (1973) and Williams (1977) found to be intermediate between monolingual L1 and L2 phoneme boundaries. Accordingly, this intermediate performance in perception can be explained by the OPM's proposal of partially merged L1 and L2 systems because each of the bilinguals' phonological systems has a component of the other system (cf. Major 2002: 82). However, it is not clear whether the intermediate perception performance documented in the bilingual literature is due to the influence of one language over the other at the cognitive representational level or at the on-line performance level.

The PIM claims that the input is parsed or mapped through the L1 structure, and therefore presupposes a difference between *input* and *intake*. The latter is the actual input to the acquisition device which results from the mapping or parsing of the ambient input through the existing L1 structure. However, no explicit proposal for the interplay between the L1 feature geometry and the developing L2 geometry is provided. Therefore, it is not clear whether L1 structures will be modified as they get redeployed to constitute L2 representations.

The PAM does not address the L2 end state, and therefore the effects of L2 development on L1 perception are not considered. Crucially, a proposal for the discovery and extraction of the articulatory features for two languages is not mentioned, and neither is it explained how a learner will cope with two types of higher order invariants at the same time.

The NLM model proposes that perceptual experience constrains future perceptual learning independently of a critical period for language acquisition. This argument gives an alternative explanation to the existence of a 'critical period' for language learning, which have been commonly proposed to account for the fact that adults do not learn languages as naturally and efficiently as children do. However, Kuhl (2000) also suggests that early in life, interference effects are minimal so that two different mappings can be acquired, whereas when a second language is learned after puberty another form of separation between the two perceptual systems may be required to avoid interference. This difference has been shown in brain imaging studies which find that adult bilinguals who acquire both languages early in life activate overlapping regions of the brain when processing the two languages, whereas late learners activate two distinct regions of the brain (cf. Kim et al. 1997). However, this model does not propose an explanation for the ways in which the separation of perceptual mappings for two languages and the activation of overlapping regions of the brain may be influenced by L2 proficiency.

The SLM argues that the state of development of L1 categories at the time of L2 acquisition will affect the native-like attainment of L2 perception. This is because the more L1 categories are developed, the more likely they are to block the formation of new categories for L2 sounds. Thus, native-like L2 perception will be more likely to be found in learners that have an early (normally pre-pubescent) age of arrival (AOA) in the L2 community than in learners with a late (or post-pubescent) AOA (cf. Flege & MacKay 2004). It is also predicted that learners who frequently use their L1 will be less likely to attain native-like L2 perception than those who rarely do so (cf. Flege, MacKay & Meador 1999; Piske, MacKay & Flege 2001, Flege & MacKay 2004).

Crucially, the SLM proposes that L1 and L2 phonetic categories are represented in a common phonological space so that both systems mutually influence one another. As a consequence, it is predicted that when a new phonetic category is established for an L2 sound that is close to an L1 sound, it will *dissimilate* (cf. Flege 2002). As a consequence, the L1 and L2 categories of bilinguals will be different from those of native speakers of the two languages, and so will their L1 and L2 perception (cf. Flege, Schirru & MacKay 2003). Thus, if a new category is not established for an L2 sound that differs audibly from the closest L1 sound, an experienced L2 learner will be expected to develop a 'composite' (or merged) category that contains both the L1 and L2 categories, such a situation resulting from *assimila-*

tion (cf. Flege 1987, MacKay, Piske & Schirru, 2001).[36] In addition, Flege (1995, 2002) argues that the principles of assimilation and dissimilation as well as the existence of a common system may underlie Grosjean's (1989) claims that the bilingual's two systems are always engaged at the same time so that the 'mixing' of L1 and L2 is inevitable.

4.3.3.1 Comparison with L2LP end state

Table 4.4 compares the L2LP proposal for the L2 end state with those of the other models that we have examined.

Table 4.4. Proposal for the L2 end state in six models of L2 sound perception. > = 'is more important than'

L2 models	Explanation	Prediction
OPM	Connected L1 and L2 phonologies	Possibly intermediate
PhIM	Not explicit	Not explicit
L2LP	Input > Plasticity (age factor) Separate perception grammars Language activation	L2 can be optimal L1 is not affect = Optimal Language modes
PAM	Not explicit	Not explicit
NLM	Linguistic experience	L1 maps & L1 categories
SLM	Age factor Single/common phonological space	Early = More native-like No perfect/optimal bilingual No monolingual modes

[36] For a more thorough discussion of the principles of category dissimilation and assimilation advocated in the SLM, see Flege (2002, 2003).

As we have seen, the L2LP model advances the separate perception grammars hypothesis which states that L2 learners and bilingual speakers have separate systems for perceiving each of their two languages. In contrast, Flege's SLM proposes that the perception of L1 and L2 sounds is performed by a single phonological system that is composed of a single set of sound categories. These two positions yield completely different predictions for the L2 end state. That is, the L2LP model predicts that a learner can attain optimal target L2 perception and, at the same time, maintain her optimal L1 perception. This is because the two languages have different perception grammars which allow for L2 development without there being any negative effects on the already optimal L1 perception. In contrast, the SLM predicts that L2 development in the form of category formation or boundary adjustments will inevitably affect L1 perception.

Is there any way of evaluating these two contradictory predictions? The solution may lie in Grosjean's (2000) concept of language modes (cf. § 3.5) which represent a continuum between an L1 and an L2 monolingual modes where only one language is activated. This continuum thus has a bilingual mode in the middle where two systems are fully activated at the same time. The L2LP model's interpretation of language modes, on the other hand, is that two separate systems can be activated at the same time which means that intermediate L1-L2 sound perception is the result of the parallel activation of the learner's two grammars during on-line perception. Therefore, according to the L2LP model, intermediate perception is not the result of having a single set of categories and perceptual mappings for two languages. Unlike the SLM, it predicts that advanced L2 learners will have optimal L1 and L2 perceptions in the monolingual setting of each language. To test this prediction, the language modes hypothesis needs to be made operative by controlling the sub-variables that constitute a language setting, e.g., stimuli, language of instructions, language of responses, etc. Chapters 6 and 7 will discuss L2 perception studies that manipulate the language activation variable in order to directly test the L2LP model's hypotheses of separate perception grammars and language activation.

4.3.4 L2 sound perception scenarios

The OPM states that L2 learners can be faced with two main scenarios depending on whether new or similar linguistic phenomena are involved, and it is predicted that L2 development will be different in each of them. It is also argued that new phenomena can be either normal or marked, i.e., relatively rare (cf. Major 2002: 76). Both similar and new phenomena that are marked are predicted to be acquired with difficulty. Thus, during the stage at which L2 normal new phenomena will be acquired, the L1 will still be present in similar phenomena, and U will still be present in new marked phenomena, as shown in Table 4.5.

Table 4.5. OPM's L2 sound perception scenarios.

Learning stage	Learning factor	Normal Phenomena	Similar Phenomena	Marked Phenomena
Early stages	L1	Dominance & decrease	Dominance & slow decrease	Dominance & decrease
	U	Minimal influence & increase	Minimal influence & slow increase	Minimal influence & rapid increase
Later stages	L1	Decrease	Slow decrease	Slow decrease
	U	Increase	Slow increase	Rapid increase
End state	L2	L2 acquired	L2 acquired slowly	L2 acquired slowly

On the other hand, the PIM says that the study of similar contrasts will not contribute to our understanding of how a novel L2 representation is acquired, and the study of completely new contrasts, such as Zulu clicks, will only tell us whether UG plays a role in L2 acquisition (cf. Brown 1998: 150-151). Therefore, it is argued that we should concentrate on non-native contrasts in which only one of the members is a phoneme in the learner's L1 because it is the only scenario in which we can see whether learners can develop new segments on the basis of already existing features.

As for the PAM, it argues that non-native sounds can be mapped onto L1 sounds following one of the patterns shown in Table 4.6. As we can see, if two sounds in the L2 are assimilated to two different sounds, discrimination is predicted to be excellent, whereas if two L2 sounds are assimilated to a single L1, sound discrimination will be very poor. The situation called 'category goodness' refers to the assimilation of two L2 sounds to a single L1 category with one L2 sound being a poor exemplar of the L1 category. The last two scenarios suggest that L2 sounds may not have a perceived similarity with any L1 sound so that they do not assimilate. However, Rochet (1995) suggests that most L2 speech sounds will be perceived as some L1 category. It may the case that sounds like Zulu clicks are perceived as non-speech but the great majority of L2 sounds, especially the ones that most learners have to deal with, may (inevitably) be perceived as the most similar L1 sound.

Table 4.6. PAM's non-native perception scenarios and predicted degrees of discrimination.

Perceptual Assimilation	Single-category	Two-category	Category-goodness	Non-categorized	categorized-non-categorized
Pattern	2 L2 sounds onto 1 L1	2 L2 sounds onto 2 L1	2 onto 1, but 1 L2 is deviant	2 L2 sounds onto 0 L1	2 L2: 1 onto 0 L1, 1 onto 1 L1
Predicted Discrimination	Poor	Excellent	Moderate to very good	Poor to very good	Very good

Finally, the SLM posits that new sounds can be acquired while old or similar sounds are very difficult or almost impossible to learn at a native-like level. This is because the assimilation of a similar L2 sound into an L1 category will inevitably lead to a change in the latter in order to cope with the L2 productions (cf. Flege 1995). One of the main ideas in the SLM is the fact that L2 sounds that are similar to L1 sounds are less easily perceived in a native-like way than are new sounds.

This relates to the idea of *perceptual equivalence*: an L1 sound and an L2 sound are merged into one category because they are perceived as equal. Thus, a single phonetic category will be used to perceive sounds that are perceptually similar which means that the closer an L2 sound is to an L1 sound, the more difficult it will be to form a new category for it.

4.3.4.1 Comparison with the L2LP scenarios

The five models that have been reviewed here all consider two different scenarios for L2 sound perception, namely learning to perceive new sounds and learning to perceive similar sounds. The PAM calls these scenarios single-category and two-category assimilation, as shown in Table 6. The L2LP model also includes the description, explanation, and prediction of the three logical states in L2 sound perception that apply to these two scenarios. In proposing that because the initial state is different in these two scenarios, it predicts that development will also be different. Importantly, unlike the other models, it also considers a third situation that constitutes the reverse of the NEW scenario, as in the case of English or Dutch learners of Spanish vowels. Within the L2LP model, this L2 sound perception scenario is called SUBSET because the L2 sounds to be learned represent a portion of the L1 sounds.

In the other five models of L2 sound perception, the comparative difficulty of the two scenarios they consider is unclear in that they do not agree on the exact challenge for L2 acquisition that is involved in the learning of NEW sounds as opposed to that of SIMILAR sounds. That is, the PIM, PAM, and NLM suggest that a SIMILAR scenario poses *no* L2 learning challenge, whereas the OPM and SLM claim that it poses the greatest L2 challenge. The L2LP model resolves this contradiction by showing that both scenarios pose a learning challenge, though the tasks are different in each case. In this view, an L2 learner faced with a NEW scenario needs to perform both a perceptual and a representational task, i.e., to create new perceptual mappings and to form new categories (cf. Chapter 5). In contrast, a SIMILAR scenario only presupposes a perceptual task because L1 categories that are similar to the target L2 categories can be reused (cf. Chapter 7). In sum, the number and nature of the tasks lead to the L2LP prediction that a NEW scenario will be more difficult than a SIMILAR scenario, as was described in § 3.6.2.

4.4 Summary and general comparison with the L2LP model

I have reviewed five current models that aim at explaining non-native or L2 sound-perception, viz., the two most promising models embedded in traditional phonological theory and the three current models in speech perception research. In § 4.1, I discussed their proposals for speech perception and its L1 acquisition, and I compared them to the LP model which forms the framework for the L2LP model. Here I summarize the aims and proposals of the five models, and then compare them to the L2LP model with respect to the three logical states in language acquisition as well as to the L2LP's two other ingredients, i.e., the description of the L1 and L2 optimal perception and the explicit proposal for the L2 learning task. I then compare the predictions that all the models make for L2 sound perception at each state of the acquisition process.

Table 4.7 presents a summary of the objectives as well as the L2 theoretical proposals of all six models described above.

Table 4.7. Aims and theoretical proposals for six models of L2 sound perception.

L2 model	AIM	Initial state	Development	End state
OPM	Describe & explain L2 phonology	Process of L1 transfer: L1 grammar	Increase & decrease	L1 and L2 are partly connected
PhIM	Describe & explain the role of the L1 in L2 phonology	L1 transfer: L2 = L1 feature geometry	L1 feature redeployment	No explicit proposal
L2LP	Describe, explain & predict the entire L2 process	Full Copying: L1 grammar/ reuse of L1 sounds	Full GLA Access: Create & adjust	Input > plasticity L1 & L2 are separate
PAM	Describe, explain cross-language perception	L1 categories	Split & reorganize	No explicit proposal
NLM	Describe, explain cross-language perception	L1 categories and L1 neural maps	Create	No explicit proposal
SLM	Describe, explain, predict the L2 end state	L1 categories	Create or merge	Age factor L1 and L2 are merged

The L2LP model has the most ambitious goal and therefore the largest scope of all six models because it gives an explicit proposal for all three states of L2 sound perception. Furthermore, the L2LP's first ingredient, viz., the thorough description of L1 and L2 optimal perception, is not considered in any of the other models. Crucially, this ingredient leads to a highly explicit proposal for the initial and end states because the former equals the optimal L1 perception and the latter should equal the optimal target L2 perception. Also, the L2LP advances an explicit proposal for the learning task, something which is mentioned but not explicated in the other models. All in all, then, the L2LP model seems to constitute a comprehensive proposal that integrates, synthesizes, and improves on the other L2 sound perception models. That is, the theoretical ingredients of the model provide an explicit and valid theory for the entire acquisition process. However, it remains to be shown whether the model is explanatorily adequate, i.e., whether its hypothesized explanations lead to predictions that are confirmed or borne out by real L2 perception data. This question will be addressed in the next chapters.

Before looking at L2 sound perception data, I provide a summary of the predictions that follow from the L2LP proposal and how they compare to those made by the other five models. Table 4.8 shows the predicted L2 sound perception in each of the three logical states of the acquisition process.

Table 4.8. Six models' predictions for L2 sound perception at the initial, developmental, and end states in the acquisition process.

L2 models	Initial state	Development	End state
OPM	L2 initial = L1	L2 is acquired	Possibly intermediate
PhIM	L2 initial = L1 feature geometry	L1 structure reused for L2 phonemes	Not explicit
L2LP	L1 grammar & L1 categories	Category formation, mapping adjustment	L2 can be optimal L1 is not affected = Optimal Language modes
PAM	L1 assimilation	Category split Reorganization	Not explicit
NLM	L1 maps & L1 categories	Creation of L2 maps & categories	L1 maps & L1 categories
SLM	L1 categories	Category formation, category merging	Early = More native-like No perfect/optimal bilingual No monolingual modes

PART III:

EMPIRICAL TESTS OF THE L2LP MODEL

5 Learning NEW L2 sounds

The most common and widely attested type of L2 sound perception involves the learning of NEW categories. This scenario occurs when the learner's L1 perception leads her to identify fewer sounds than the ones produced in the L2 environment. As a consequence, she fails to hear the differences between certain L2 sounds because she perceives a single sound when confronted with two different L2 sounds. As discussed in Chapter 4, this is regarded as problematic for L2 learning within all the phonological and phonetic models that aim at explaining L2 sound perception. The L2LP model also considers this situation a challenging learning problem because it predicts that a learner will copy her L1 categories and perception grammar to cope with the L2 environment. With respect to L1 categories, the NEW scenario is characterized by the *phonemic equation* of two L2 sound categories with a single L1 category, as shown in the examples of Figure 5.1. Thus, a learner will use the same L1 phoneme for the lexical representation of L2 words that contain two different phonemes.

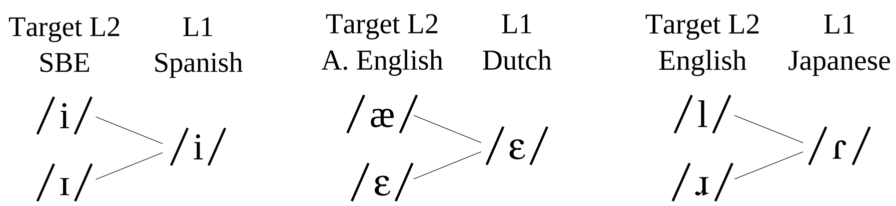

Fig. 5.1. Phonemic equation in the NEW L2 perception scenario. SBE = Southern British English. A = American.

The LP model makes a principled distinction between abstract phonological representations and perceptual mappings, while the L2LP model proposes that such a difference allows us to most accurately describe, explain, and predict L2 sound perception. This is because the development of L2 sound perception is defined by how L2 sound categories relate to the L1 perceptual mappings and categories. In the NEW L2 sound perception scenario, the equation of two L2 categories with a single phonologically equivalent L1 category, which occurs at the abstract phonemic level, is accompanied by the perceptual mapping of the majority of the tokens of both L2 categories onto the same L1 category. In other words, most phonetic realizations or auditory events of the two L2 phonological representations will be perceived as a single L1 category. An example of the typical perceptual mapping in a NEW scenario is given in Fig. 5.2, where the thick lines represent the most common mapping of an L2 sound onto an L1 category.

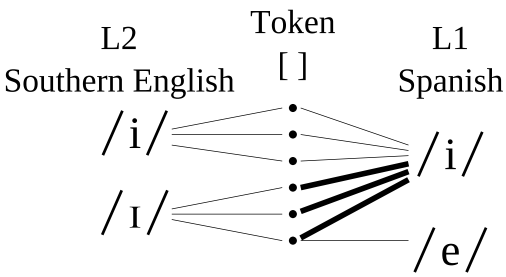

Fig. 5.2. Perceptual mapping in a NEW L2 scenario.[37]

Table 5.1 provides a summary of the L2LP model's five ingredients which were discussed in Chapter 3. With respect to the logical states in L2 acquisition, the Full Copying hypothesis predicts that a copy of the L1 perception grammar and categories will initially underlie L2 perception.

[37] Although most tokens of two L2 sounds are perceived as a single L1 category in a NEW scenario, some tokens could also be perceived as another acoustically close L1 category. Also, not all NEW cases are the same because some will have complete, or almost complete, single-category assimilation whereas others will have more tokens being perceived as a second L1 category.

Table 5.1. Summary of the proposal for L2 linguistic perception and respective predictions for L2 performance and knowledge.

L2LP	Prediction	Explanation	Description
Optimal L1 & L2	Human beings are optimal listeners	Optimal listeners handle the environment maximally well	L1 and L2 optimal category boundaries: location & shape
Initial state	= Cross-language perception	Full Copying	L1 boundary location and shape
Learning task	= Reach the optimal target L2 perception	L2 learners want to reach target	Bridging mismatches between L1 and target optimal perception
Development	= L1-like	Full GLA Access	Category formation and boundary shifts
End state	Optimal L2 perception and optimal L2 perception	Input overrules plasticity Separate grammars	Language activation modes, through language setting variables

The L2LP model interprets the hypothesis of Full Access as the total accessibility to the Gradual Learning Algorithm (GLA). This predicts that L2 development will occur and that it will be guided by the same learning mechanisms that the algorithm performs in L1 perceptual learning (cf. § 2.1.3). Also, the proposal of two separate perception systems for L1 and L2 predicts the maximum level of L2 perceptual success. That is, the L1 and L2 perceptions of advanced learners will match those of optimal perceivers in the two languages because L2 development is achieved without affecting the already optimal L1 perception which remains stable provided the learner continues to be exposed to her L1. The model explains possible intermediate L1-L2 perception through its language activation modes hypothesis which posits that L2 learners can activate their two languages (to different degrees) at the same time, depending on the language setting they find themselves in.

In this chapter, I present the specific L2LP model predictions when NEW sounds are involved. These are L2 sounds that are produced with at least one auditory dimension that has not been previously incorporated into the learner's L1 linguistic perception. Recall that a distinction was made between *already-categorized* and *non-previously-categorized* auditory dimensions (cf. § 3.2.3.2), and that it was proposed that in the latter situation the learning task and the L2 development follow L1-like development. What will be shown here is how the general L2LP proposal can be applied to the acquisition of Southern British English (SBE) /i/ and /ɪ/ by Spanish learners wherein the new auditory dimension is vowel duration. In § 5.1, I discuss the elements involved in learning to perceive NEW sounds. In this chapter, I offer specific predictions and explanations for a NEW scenario as well as experimental evidence to support such predictions. In § 5.2, I apply the five ingredients of the L2LP model to an example of this NEW scenario in order to draw specific, concrete, and falsifiable predictions. In § 5.3, which is the empirical part of the chapter, I report perceptual data to test each of the five specific predictions formulated in § 5.2. Finally, in § 5.4, I provide a concise evaluation of the model's predictions in view of the evidence presented.

5.1 What does learning to perceive NEW sound categories involve?

The NEW scenario, in which the learning task is to perceive *more* categories through the formation of new phonological representations, involves both the category-formation and the boundary-shifting L1 learning mechanisms described in § 2.1.3. Here I present a summary of the steps that are needed to acquire perceptual proficiency in the NEW L2 sound scenario. The predictions for the initial state and the learning tasks apply to cases of already-categorized dimensions (such as F1 or F2 for Spanish learners of English /i/ and /ɪ/ and Dutch learners of English /æ/ and /ɛ/) as well as to cases of non-previously-categorized dimensions (such as vowel duration for L1 Spanish, French, Polish, or Portuguese listeners).

As a first step on her way to becoming an optimal L2 listener, the learner will copy her L1 perception onto the system of her new language. As a result, her initial L2 perception will be equal to her L1 perception, i.e., it will have L1 perceptual mappings and L1 phonological categories. In a NEW scenario, the L2 learning task is proposed to involve two components, namely perceptual learning and representational learning (cf. § 3.3.2). Table 5.2 shows the exact nature of these two types of tasks and the predicted degree of difficulty for the performance of this double L2 learning task.

Table 5.2. The initial state and the learning tasks in a NEW scenario.

	L2LP proposal	**Prediction for NEW**
Initial state	*Full Copying*	Too few categories
Perceptual task	yes	*Create* or *split* perceptual mappings *Integrate* auditory cues
Representational task	yes	*Create* or *split* categories *Create* or *split vowel* segments
Degree of difficulty	Large: two complex tasks	Requires extensive exposure to rich L2 input

The L2LP model proposes that learners who face a NEW scenario involving a non-previously-categorized dimension (such as vowel length for Spanish learners of English vowels, or perhaps F3 for Chinese, Japanese, and Korean learners of English /r/ and /l/) will have access to an acquisition device that makes L2 development available through two sequential L1-like learning mechanisms. First, the learner's L1 acquisition device, i.e., the GLA, which is also available for L2 learning, will be responsible for the categorization of the new auditory dimension (s) in order to create new perceptual mappings and resulting new phonetic categories. These new L2 phonetic categories will be turned into abstract one-dimensional phonological categories and will be copied to the L2 lexicon, a process which is identical to that followed by L1 learners (cf. § 2.3). Table 5.3 summarizes the results of Full Access to auditory-driven L1 learning in the NEW L2 scenario.

Table 5.3. Category formation and storage in a NEW scenario.

Learning path	L2LP proposal	L2 result
Step 2	*Full GLA Access 1)*	New phonetic mappings and phonetic category formation
Step 3	*auditory-guided learning*	New phonological mappings and new phonological categories
Step 4		New phonological representations to store L2 words

Once the new categories are used to store L2 words in the lexicon, additional learning to match the optimal perception of the L2 environment can occur, just as in the L1 learning of sound perception (cf. § 2.3.4). The L2LP model proposes that the new forms in the lexicon can supervise the adjustment of perceptual category boundaries, as shown in Table 5.4 where the next four steps of the proposed L2 development are described.

Table 5.4. Perceptual category boundary shift and cue integration in the NEW scenario.

Learning path	L2LP proposal	L2 result
Step 5	*Full GLA Access:*	Constraint reranking: boundary adjustment
Step 6	*2) lexicon-guided learning*	Low level integration of acoustic dimensions
Step 7		Using more abstract categories in the lexicon
Step 8		Further boundary adjustments

In this table, we observe that Full Access to the GLA also implies that the L2 learner will have access to lexicon-guided learning in which the algorithm acts as a boundary shifting device when there are mismatches between pre-lexical and lexical recognition (cf. § 2.3.2). When lexicon-driven learning is available to the learner, more complex tasks such as the integration of multiple auditory dimensions can occur. These tasks are obviously needed in the NEW scenario because a typical case of NEW L2 sounds will not only require the creation of new perceptual mappings but also the integration of new elements into the already-categorized continua.

5.2 L2 Linguistic Perception in a NEW scenario

The aim of this section is to illustrate the model's specific predictions for the NEW scenario. The L2LP model postulates a device, the GLA, that can only perform L1-like learning, such as auditory-driven category formation and lexicon-driven category boundary shifts. Therefore, the predictions for the L2 development and the L2 state in a NEW scenario are restricted to cases that involve at least one non-previously-categorized auditory dimension, such as vowel duration in Spanish learners of SBE vowels.

In the following sections, I will apply the five ingredients that constitute the L2LP model, as shown in Table 1, to the case of Spanish learners of SBE /i/ and /ɪ/. As discussed in § 3.1.3, while four of the model's ingredients are clearly related to L2 acquisition, the first one, viz. optimal perception, involves the L1 perception of each language. According to the model's proposal, the measure of the optimal perception in the target L2 will allow us to determine the learning task and the precise nature of the target of perceptual learning. Also, according to the model's Full Copying hypothesis, measuring the optimal L1 perception will allow us to determine the perception system that the learner brings to the L2 task. And, finally, the L1 optimal perception will give a reliable estimate of the system that the learner needs to maintain in order to optimally cope with her L1 environment while she is developing into an optimal L2 listener.

5.2.1 Ingredient 1: predicting L1 and target L2 optimal perception

The LP model presupposes that the first step in explaining the perceptual differences between speakers of different languages is to have a firm understanding of the production environments at hand. For the NEW L2 case of Spanish learners of SBE /i/ and /ɪ/, this involves measuring the production of SBE /i/ and /ɪ/ and that of the closest Spanish vowels, viz., /i/ and /e/.[38] This is because the optimal perception hypothesis proposes that perception strongly depends on the specific production environment such that the optimal way of perceiving the sounds of a language depends on how such sounds are produced in that language. The computation of the relevant measurements for the production environments that can later be compared to perception involves providing the mean productions and the standard deviations for the sounds involved. Table 5.5 shows the production averages and standard deviations for the SBE and Spanish vowels along two dimensions.

Table 5.5. F1 and duration average values for SBE and Spanish vowels.

SBE		Average	s.d.	Spanish		Average	s.d.
/ɪ/	duration	59.7 ms	0.4 dou.	/e/	duration	78 ms	0.4 dou
	F1	337 Hz	0.2 oct.		F1	502 Hz	0.2 oct.
/i/	duration	104.6 ms	0.4 dou.	/i/	duration	81 ms	0.4 dou.
	F1	292 Hz	0.2 oct.		F1	331 Hz	0.2 oct.

The F1 and duration geometric averages are taken from Escudero & Boersma (2003) for the SBE vowels and from Cervera, Miralles & Gonzalez-Alvarez (2001) for the Spanish vowels. For the sake of simplicity, the standard deviations are set at the same values in the two languages. Although such standard deviations were not measured, they are used here to ensure that these environments contain a wide range of F1-duration pairs. Figure 5.3 shows these values in an F1-duration plane.

[38] The Spanish five-vowel system which is made up of the sound categories /i, e, a, o, u/ is much smaller than that of SBE which has 12 vowel monophthongs, namely /i, ɪ, ɛ, ∈, æ, a, ɑ, ɒ, ʌ, ɔ, ʊ, u/.

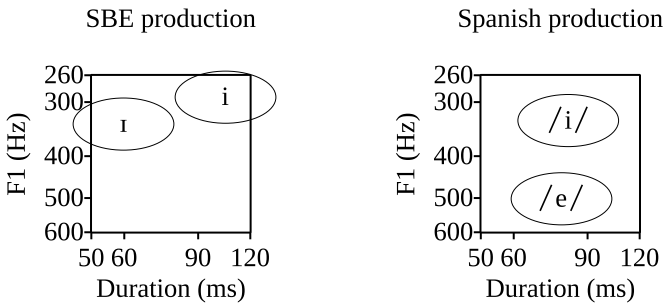

Fig. 5.3. SBE and Spanish production distributions.

As described in Chapter 2, there are three different computations that will allow us to establish a production-perception relationship. First, we need to compute the location of the midpoint, which in this case has an F1 and a duration location situated between the two vowels. To establish the location of the midpoint, we first compute the distance between the vowels along the F1 and duration axis. Such distances can be expressed in the base-2 logarithmic units of octaves and duration doublings. For the SBE vowels, going from /ɪ/ to /i/ amounts to a falling F1 of log2(337/292) = 0.207 octaves and to a rising duration of log2(104.6/59.7) = 0.809 duration doublings. The SBE midpoint has an F1 location that lies in the middle of the F1 distance and a duration location that lies in the middle of the duration distance, which is log2(337) - 0.207/2 octaves and log2(104.7) - 0.809/2 duration doublings, or equivalently [314 Hz, 79 ms].[39] For the Spanish vowels, going from /e/ to /i/ amounts to a falling F1 of log2(502/331) = 0.601 octaves and to a rising duration of log2(81/78) = 0.055 duration doublings, thus leading to a midpoint located at [408 Hz, 80ms].[40] Figure 5.4 shows the midpoints in the two languages.

[39] The conversion from octaves to Hertz is $2^{(\log2(337) - 0207/2)}$ = 313.67 Hz and the conversion from duration doublings to milliseconds is $2^{(\log2(104.6) - 0.809/2)}$ = 79.03 ms.

[40] That is, $2^{(\log2(502)-0.601/2)}$ = 407.61 Hz and $2^{(\log2(81)-0.055/2)}$ = 79.47 ms.

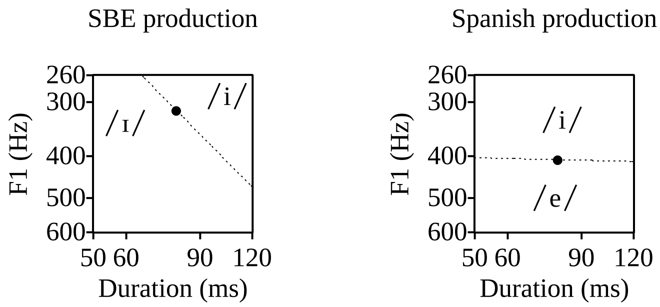

Fig. 5.4. Midpoint and equal-likelihood line for the SBE and Spanish vowels.

Second, we need to compute the equal-likelihood line in production, i.e., the line (which is represented as a dashed line in Figure 4) that connects all the F1-duration pairs that are 50% of the time intended as one vowel and 50% as the other. The slope of the equal-likelihood line can be computed as the ratio of the F1 and duration acoustic distances between the two vowels multiplied by the squared ratio of the F1 and duration standard deviations. The F1 and duration distances are known from the previous computation, and the standard deviations for the two dimensions in each language are 0.2 and 0.4 respectively (cf. Table 2). Thus, the SBE equal-likelihood line has a slope of (0.809/0.207) · (0.2/0.4)2 = 0.98 octaves/duration doubling and the Spanish equal-likelihood line has a slope of (0.055/0.601) · (0.2/0.4)2 = 0.021 octaves/dur. doublings.

In this case, the midpoint is on the equal-likelihood line because the standard deviations are the same for the two vowels. Therefore, with the midpoint and the slope, we can compute the location of the entire equal-likelihood line. To do this at one of the edges of the F1-duration plane, we need to measure the duration distance between the midpoint and the edge, and then multiply it by the slope of the line. The SBE line extends from the top to the right of the square, which means that we need to compute its duration location at the top edge and its F1 location at the right edge. Thus, the location of the crossing point at the top edge is [260 Hz, 66 ms], i.e., (log2(79) = 6.30 octaves) - log2(314/260)*0.98 = 65.66 ms, and the crossing point at the right edge is [473Hz, 120 ms], i.e., log2(314) + log2(120/89)*0.98 = 472.99 Hz. The Spanish line extends from left to right so that we need to compute its F1 location at the left and right edges. The Spanish left edge point is located at [404 Hz, 50 ms], i.e., log2(408) - log2(80/50)*0.023 =

403.61 Hz, and the right edge point is located at [412 Hz, 120 ms], i.e., log2(408) + log2(120/80)*0.023 = 411.82 Hz.

The last computation is that of the relative use of the F1 and duration for signalling the differences between the two vowels in each language, which we can calculate from the slope of the equal-likelihood line. For SBE, the duration difference as expressed in standard deviations of duration is approximately twice as large as the F1 difference, i.e., (0.809/0.4)/(0.207/0.2) = 1.95, whereas in the Spanish environment the F1 difference as expressed in standard deviations of F1 is approximately 22 times larger than the duration difference, i.e., (0.601/0.2)/(0.055/0.4) = 21.9. Table 5.6 shows the production measurements for the two languages.[41]

Table 5.6. Production values for the Southern British English (SBE) and Spanish production distributions (oct. = octaves, durdou = duration doublings).

	Vowels	Midpoint	Equal-likelihood slope	Equal-likelihood edge points	F1/dur use
SBE	/i/-/ɪ/	[314Hz, 79ms]	0.98 oct./durdou.	[260 Hz, 66 ms]; [473 Hz, 120 ms]	1/2
Spanish	/i/-/e/	[408Hz, 80ms]	0.0021 oct./durdou.	[404 Hz, 50 ms]; [412 Hz, 120 ms]	22/1

The optimal perception hypothesis implies that an optimal listener perceives vowels according to the way they are produced in her specific environment. Hence, the optimal perception will match the values computed from the production environments in the two languages. Specifically, three main perceptual values will match their production counterparts if the optimal perception hypothesis holds. First, the location of the optimal perceptual boundary between the vowels, which estimates the F1-duration pairs that are equally likely to be categorized as /i/ or /ɪ/, will coincide with the location of the equal-likelihood line in production. Second, the shape of the perceptual boundary will match that of the production equal-likelihood line. In other words, the slope of the perceptual boundary line will be the

[41] Note that the slope and edge points refer to the same computation, i.e., the equal-likelihood line in production.

same as that of the equal-likelihood line in production. Figure 5.5 shows this equal-likelihood line as the predicted optimal category boundary in perception, which entails that the same values in Table 5.6 above apply to this boundary line.

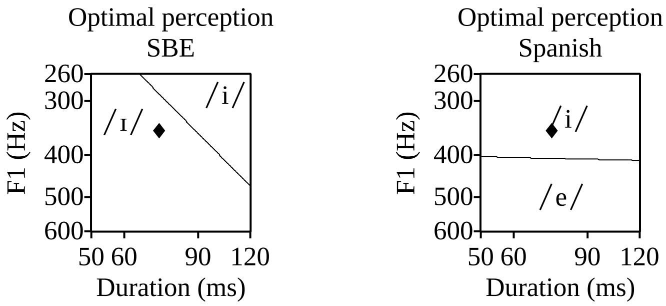

Fig. 5.5. The predicted optimal perception for SBE /i/ and /ɪ/ and for Spanish /i/ and /e/. Solid line: optimal perceptual category boundary = Equal-likelihood line. Diamond: F1-duration pair with values [349 Hz, 74 ms].

With respect to sound categorization, it is predicted that an optimal SBE listener will perceive F1-duration pairs with values that fall above her optimal category boundary as /i/ and F1-duration pairs with values that fall below such an optimal boundary as /ɪ/. The optimal Spanish listener will perceive F1-duration pairs that fall below her category boundary as /i/ and F1-duration pairs that fall below it as /e/. For instance, a token with values [349 Hz, 74 ms], which is represented by the diamonds in Figure 5.5, will be interpreted as /ɪ/ by an optimal SBE listener and as /i/ by an optimal Spanish listener.

The third perception value that is predicted to match its production correlate is the use of the F1 and duration dimensions in perception, i.e., their cue reliance. As described in § 3.1.2, the perceptual use of the duration dimension for categorizing vowels, i.e., the *duration reliance*, is computed as the percentage of /i/ responses along the right edge of the square minus the ones along the left edge. On the other hand, the *F1 reliance* is computed as the percentage of /i/ responses along the top edge minus the ones along the bottom edge. Figure 5.6 shows the F1 and duration reliance for SBE and Spanish optimal listeners.

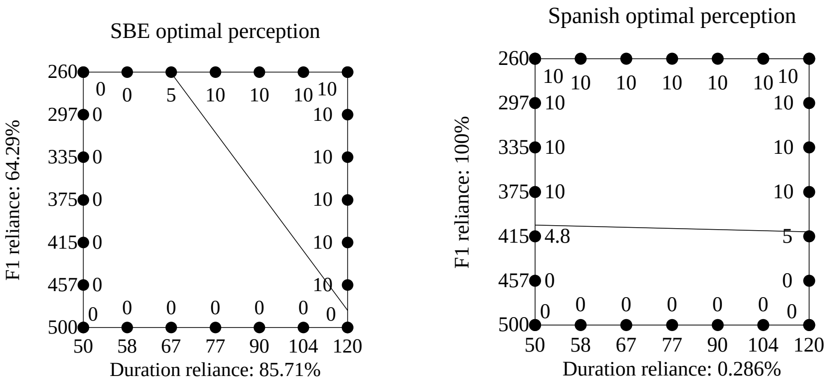

Fig. 5.6. Optimal cue reliance for SBE and Spanish listeners.[42]

With the cue reliance percentages we can compute the category boundary slope, which in this case should be the same as, or very similar to, the equal-likelihood slope, though it is also dependent on the stimulus square presented to the listeners. Escudero & Boersma (2003: 81) argued that the ratio of the duration and spectral reliances expressed in terms of the F1 and duration ranges provides a good estimate of the slope of the category boundary line. This means that the SBE optimal boundary has a slope of (85.71% · log2(500/260))/(64.29% · log2(120/50)) = 0.996 oct./dur. doubling and the Spanish optimal boundary has a slope of 0.0021 oct./dur. doubling. Both of these are the same as the equal-likelihood slopes of 0.98 and 0.0021 shown in Table 6 above.

Following the LP model described in Chapter 2, it is proposed that a language-specific perception grammar underlies optimal categorization, category boundaries, and cue reliance. Such a perception grammar contains *cue constraints* that are ranked in language-specific ways and which were learned during L1 acquisition (cf. § 2.1.3). Figure 5.7 shows a representation of the continuous constraint ranking in the adult SBE perception grammar which optimally categorizes F1 values into /i/ and /ɪ/.

---

42 The 2.86% duration reliance in the Spanish optimal perception grammar is a product of a very small difference between the duration values of Spanish /i/ and /e/, viz., three ms. It is assumed here that such a duration cue reliance cannot be considered reliably different from zero. Therefore, it is predicted that the optimal Spanish listener will have a zero duration reliance or close to it, though such a prediction needs to be empirically tested. The perceptual data presented in § 5.3.1 may give an answer to this question.

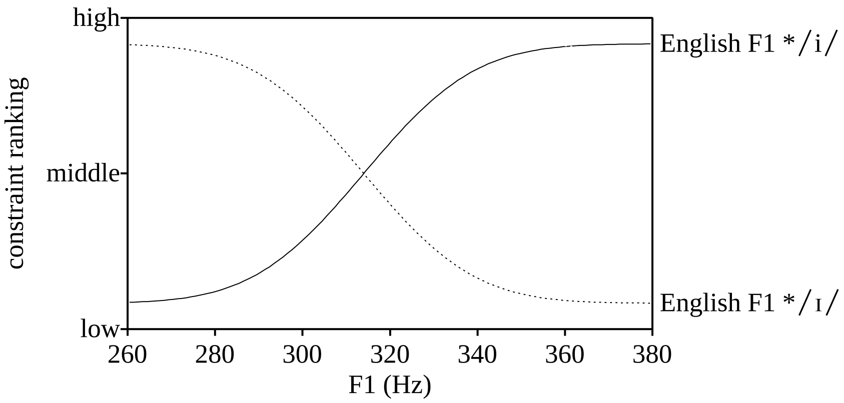

Fig. 5.7. The cue constraints in the optimal SBE perception grammar for the categorization of F1 values (260-380 Hz) as /i/ (solid) or /ɪ/ (dotted).

Along the *y* axis in the figure, we see the constraint rankings that provide the likelihood of a value to be categorized as either of the two sound segments if duration is held constant at the most ambiguous value. This is in accordance with the decision-making mechanism of stochastic OT (cf. § 2.1.2). The two curves correspond to the continuous ranking of cue constraints against the categorization of F1 values as either of two English vowels. The solid line represents the rankings for the cue constraint family 'an F1 value should not be perceived as /i/' or, in short, 'F1 */i/', while the dotted line represents the continuous cue constraints for F1 */ɪ/. Thus, because the constraint 'do not perceive 200 Hz as /i/' is ranked low in this optimal SBE grammar, such an F1 will be categorized as /i/ almost 100% of the time. Also, because the constraint against perceiving 337 Hz as /ɪ/ has a low ranking, a token with such an F1 value will be perceived as /ɪ/. The crossing of the cue constraints at the middle point of the ranking will result in a 50-50% perception. This crossing point coincides with the F1 value of the SBE midpoint in Figure 4, i.e., 314 Hz.

With respect to vowel duration, Figure 5.8 shows the constraints and constraint rankings for the categorization of vowel duration in an optimal SBE grammar, that is to say, how these English listeners categorize duration values as the two English vowels if F1 is held constant at the most ambiguous value. This crossing of the constraints is at the duration value of the SBE midpoint in Figure 5.4, i.e., 79 ms.

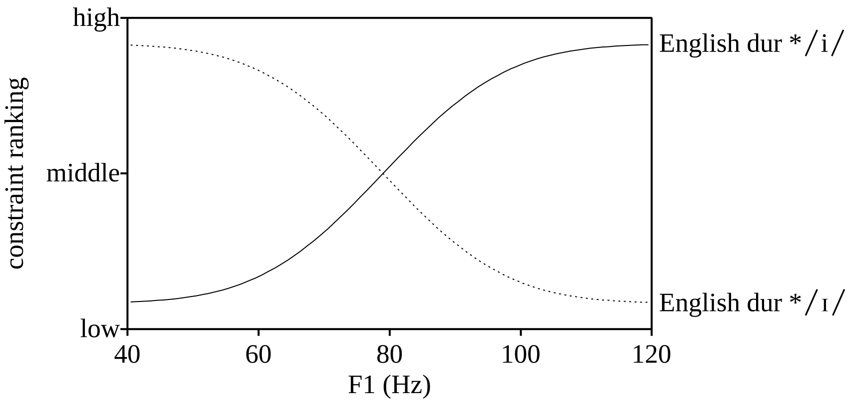

Fig. 5.8. The cue constraints in the optimal SBE perception grammar for the categorization of vowel duration values (40-120 ms) as /i/ (solid) or /ɪ/ (dotted).

For the optimal Spanish perception grammar, Figure 5.9 below shows the proposed constraints and constraint rankings for the categorization of F1 values into Spanish /i/ and /e/. We can observe that most of the values are categorized as /i/ because the cue constraints against perceiving F1 values lower than 380 Hz as /e/ are highly ranked.

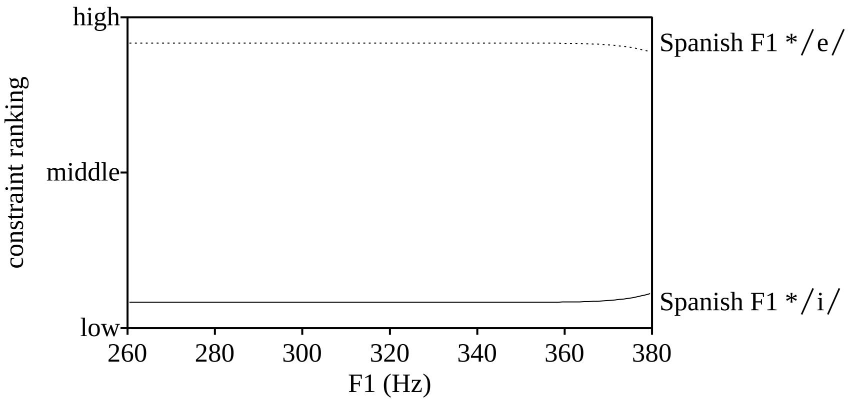

Fig. 5.9. The cue constraints in the optimal Spanish perception grammar for the categorization of F1 values (260-380 Hz) as /i/ (solid) or /e/ (dotted).

With respect to vowel duration, an optimal Spanish perception grammar has a middle ranking for the cue constraints against perceiving vowel duration values as either of the two vowels, as shown in Figure 10 below.

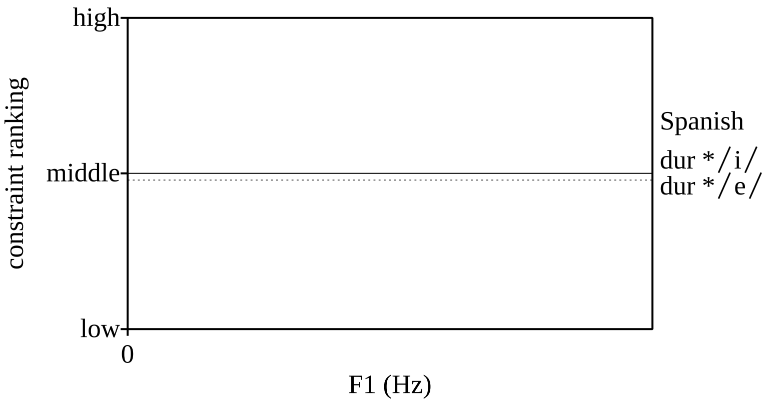

Fig. 5.10. The cue constraints in the optimal Spanish perception grammar for the categorization of vowel duration values (40-120 ms) as /i/ (solid) or /e/ (dotted).

The middle ranking of vowel duration in the optimal Spanish perception grammar results from the small differences between vowel duration values in the Spanish production environment. Crucially, these duration-to-vowel cue constraints do not refer to categories that contain vowel length because vowel length categories are never created by Spanish perception grammars. Rather, the cue constraints in Figure 5.10 originate from constraints that map any auditory dimension, e.g., VOT, F2, F3, duration, etc., onto height categories such as /high/ which are introduced during L1 perceptual development (cf. § 2.3.4). These constraints develop in order to allow for the integration of F1 and F2 in the perception of Spanish vowels so that height and backness are combined into more complex categories such as /height, backness/ or into vowel segments. This means that the developing Spanish grammar never contains duration-to-duration auditory constraints which means that there are never any duration-to-length cue constraints. In sum, it can be said that in Spanish auditory-driven learning does not apply to vowel duration, and therefore it represents a non-previously-categorize dimension for an optimal Spanish listener (cf. § 3.2.3.2).

5.2.2 Ingredient 2: predicting cross-language and initial L2 perception

In most cases, a NEW scenario manifests itself as a cross-language perceptual mapping of two foreign language sounds onto a single L1 category (cf. Figures 1 and 2). Likewise, the typical pattern in this scenario is the equation of two foreign phonemes with one L2 phoneme. Recall that the L2LP's Full Copying hypothesis states that cross-language categorization is the initial state of L2 perception. That is, both foreign language listeners and beginning L2 learners will automatically and unconsciously reuse the L1 categories that are equated with the L2 sounds, and they will reuse the L1 perception mappings that correspond to such L1 categories. Thus, it is proposed that cross-language perception will establish how absolute beginners will initially perform in their L2. In a NEW scenario, the copying of the optimal L1 perception will turn out to be quite problematic because it will lead to lexical and perceptual mismatches. In other words, the reused categories will be too few, and thus the corresponding perceptual mappings will categorize too few sounds when compared to the optimal perception. Figure 5.11 shows how the optimal Spanish perception grammar copes with the target L2 in a NEW scenario.

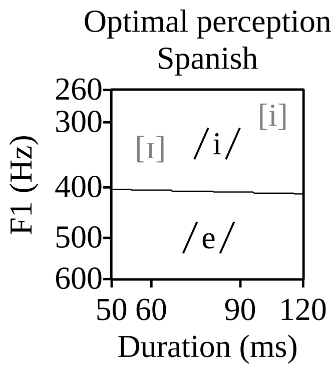

Fig. 5.11. Categorization of the SBE average productions (in grey and between slashes) of [i] and [ɪ] by the optimal Spanish listener.

As can be seen, the average SBE productions will be categorized as a single vowel by the optimal Spanish listener because they fall above the upper part of her boundary line.[43] In other words, the Spanish cue constraints against perceiving F1 values lower than 407 Hz as /i/ are ranked the highest, so that tokens of both SBE /i/ and /ɪ/ will be categorized as Spanish /i/. Tableau 5.1 shows how this optimal Spanish grammar categorizes the average SBE token intended as /ɪ/, i.e., as [337 Hz, 59.7 ms] as shown in Figure 5.9.

43 That does not mean there will be a failure to discriminate the auditory differences between tokens. However, perceiving them as speech sounds will deactivate the listeners' auditory perception, as discussed in § 1.1.2.

[337 Hz] SBE /ɪ/	337 Hz not /e/	59 ms not /i/	59 ms not /e/	337 Hz not /i/
☞ /i/				*
/e/	*!			

Tableau 5.1. Categorization of the average token of SBE /ɪ/ by the optimal Spanish listener.

Given the Full Copying of optimal perception, this cross-language perceptual behaviour will be found in the categorization of beginning Spanish learners through the reuse of L1 perception mappings. In addition, these learners will also manifest the phonemic identification shown in Figure 5.12 where the vowels of English words such as sheep and ship are equated with the first vowel in Spanish words such as *chica* or *Lima*. The figure also shows how the L1 corresponding phoneme is reused in the L2 lexicon to initially represent the two SBE words.

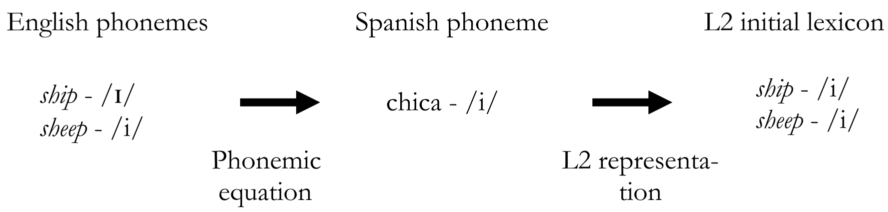

Fig. 5.12. Spanish cross-language phonemic equation and L2 initial lexicalization.

In sum, two predictions can be made regarding the initial state of new L2 perception scenario. One is that L2 learners will reuse a single L1 phonological representation to store L2 words that have two phonological representations in the lexicon of optimal L2 listeners. The second is that L2 learners will reuse their optimal L1 perception grammar, including non-previously-categorized dimensions. After the reuse of a single phonological representation and the mappings to a single perceived category, beginning learners will not be able to rely on F1 or duration to identify the differences between SBE /i/ and /ɪ/. Therefore, they will categorize and lexicalize only one L2 vowel, and in so doing render their perception and recognition extremely non-optimal. This situation will lead to the two L2 learning tasks described in the following section.

5.2.3 Ingredient 3: predicting the L2 learning task

There are two types of learning tasks in a new scenario, namely a representational task and a perceptual task. In the former, the learner will have the same phonological form for words like sheep and ship. In other words, the L2 lexicon of a beginning Spanish learner of SBE will contain the same phonological form in the lexical representation for the two English words. This non-optimal lexical storage will leave her with the sole option of relying on the semantic or pragmatic context in order to access the correct meaning.[44] The logical way of solving this L2 phonological recognition problem would be to have a new category available for one of the SBE vowels. The question is how such a category could become available. One possibility would be to create extra phonological categories by splitting already-existing L1 categories, e.g., by splitting /i/ into /high/ and /high-mid/, or by creating completely new categories along the non-previously categorized vowel duration dimensions, i.e., /long/ and /short/. It is proposed here that to be able to use two different words in her lexicon, the L2 learner will first need to perceive the difference between the L2 vowels.

The L2LP model proposes that vowel duration has a different status than that of F1 in the Spanish optimal perception grammar, thus triggering its use in the L2 perceptual task of Spanish learners of SBE vowels. That is, as mentioned in the previous section, an optimal Spanish listener has no vowel duration categories because this is a non-previously-categorized dimension whereas F1 is an already-categorized dimension. The model also proposes that non-previously categorized dimensions will be preferred to already-categorized dimensions when learning to perceive new L2 sounds. This means that a Spanish learner will use the non-previously-categorized status of vowel duration to create completely new perceptual categories instead of splitting the already-categorized F1 or F2 continua. Figure 5.13 shows the L2LP model's learning task for a case of the new scenario that involves a non-previously-categorized dimension.

---

[44] Given optimal semantic categorization, the learner's recognition grammar will tell her that two semantic categories are linked to the phonological form /ʃip/. Thus, she will assume that tokens of the two words differ semantically but not phonologically, i.e., that they are *homophones*.

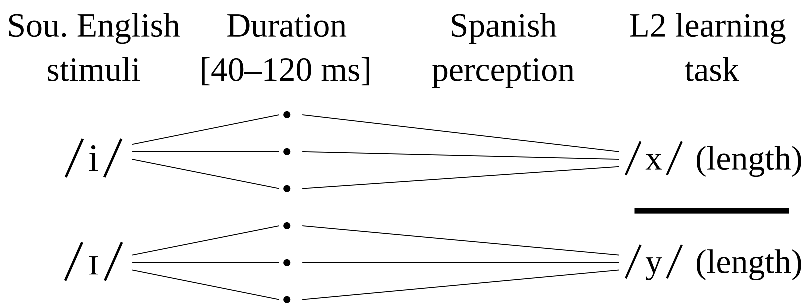

Fig. 5.13. The L2LP perceptual task for a new scenario: phonological categorization of vowel duration (as depicted by the solid lines).[45] Horizontal line: Vowel length perceptual boundary.

This proposal makes the L2 perceptual task equivalent to that of an infant learning to categorize L1 sounds. Crucially, such an L1 category formation has been suggested to occur through *distributional learning* (cf. Maye, Gerken & Werker 2002 and § 2.1.3) which is a mechanism based on the auditory distributions of non-previously categorized dimensions. Apart from proposing the L1-like nature of the L2 learning task, the L2LP model provides a formal account for the learning mechanism through which the learner executes the learning task in this scenario. In other words, the model provides a phonological formalization of auditory-driven category formation, as will be demonstrated in the next section.

---

[45] As described in Chapter 3, the L2LP proposal contrasts with Bohn's (1995) views on this in two respects, namely in the nature of vowel duration and in that of the L2 learning task. First, he suggests that Spanish listeners have a single vowel category, which makes this dimension equal to F1. Second, and as a result of the first argument, he claims that a Spanish learner will need to split her L1 vowel category. Importantly, the two proposals agree in positing that the learning task involves vowel duration. However, the reason for using this dimension is completely different in each argument. That is, while Bohn argues that vowel duration is used as a universal resource when F1 differences are not available, the L2LP proposes that its non-previously-categorized status leads to auditory-driven (or distributional) category formation, which is an L1-like learning mechanism.

5.2.4 Ingredient 4: predicting L2 development

Here I present the predicted developmental path and formal modelling of the learning mechanisms in involved in the current learning scenario. The graph in Figure 5.14 shows a summary of the L2LP proposal for the developmental path in a NEW scenario such as the acquisition of SBE /i/ and /ɪ/ by a Spanish learner. The squares in the figure show the proposed initial state and learning mechanisms that result from Full Copying and Full Access to the GLA (cf. § 3.2 and § 3.4 respectively), the circles represent the predicted learning tasks and L2 end state.

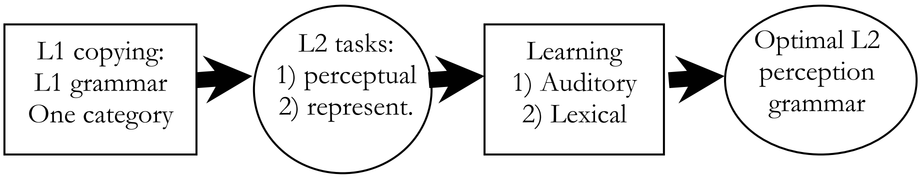

Fig. 5.14. Predicted developmental path for a NEW L2 perception scenario (represent. = representational)

We can see that the learner initially reuses a single L1 category and the perceptual mappings that output that single category alone. As mentioned in the previous section, this leads to a perceptual and a lexical task. As for the perceptual learning, it is proposed that the learner introduces constraints that allow for the perception of more than one category for SBE high front vowels. This perceptual task is accomplished by creating auditory constraints for vowel duration, which is a non-previously-categorized dimension in her initial L2 perception grammar.

Recall that L1 perceptual category formation has been posited to occur by means of distributional learning which is a mechanism that is performed by the GLA (cf. § 2.3.2). This formal proposal states that when the Spanish learner hears the duration distributions of SBE vowels, the GLA acts as a distributional learning device which creates new categorization constraints for vowel duration. Thus, the distributions are gradually mapped to the two most frequent duration values which emerge as discrete auditory phonetic categories (cf. § 2.3.3). Tableau 5.2 shows the formalization of distributional learning for the categorization of SBE duration values by Spanish L2 learners. It is proposed that *CATEG [*x* ms] constraints will be created, and they will be ranked depending on the frequency distributions of the duration values (cf. § 2.1.3).

[60 ms]	*CATEG (/60/)	*CATEG (/80/)	PERCEIVE ([60])	*WARP (20)
/80 ms/	*!			*
√ /60 ms/		*!→		
☞ /–/			←*	

Tableau 5.2. Auditory-driven constraint ranking for the creation of new phonetic categories.

Thus, once the values have become discrete categories in the learner's perception, phonological abstraction will lead to their conversion into abstract categories. In this case, the auditory-phonetic categories /60/ and /105/ will be transformed into /short/ and /long/ respectively. This category abstraction, in turn, will lead to the translation of the auditory constraints so that they can map auditory values onto the newly created abstract categories for vowel length. These new one-dimensional cue constraints are the ones that were proposed to emerge in the L1 perception grammar after auditory-learning has occurred (cf. § 2.3.3). The newly introduced cue constraints and constraint rankings are shown in the representation of the L2 developmental grammar in Figure 5.15.

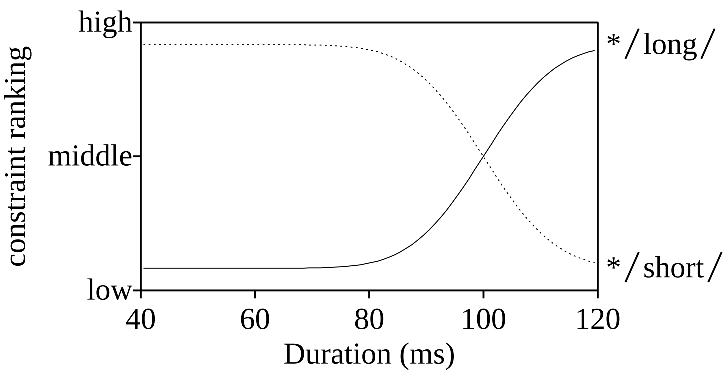

Fig. 5.15. The L2 perception grammar after categorizing duration: continuous cue constraints and rankings for perceiving duration values as /short/ (dotted) or /long/ (solid).

What the L2LP model proposes, then, is that the perception of high front SBE vowels in Spanish learners will develop from an initial single-category stage to a two-category stage through L1-like auditory-driven category formation. At this point, the learner will use phonological length to represent words in her L2 lexicon. That is, the phonological component of vowels in lexical items will contain vowel length. For instance, words like *ship* and *sheep*, which are both represented as /ʃip/ in the initial L2 lexicon, will now be represented as /ʃ-long-p/ and /ʃ-short-p/ respectively. Through these perceptual and lexical developments, the learner will perceive and lexicalize the optimal number of L2 vowels when confronted with SBE words containing /i/ and /ɪ/.

Although it is predicted that distributional learning can lead to reasonable category boundaries, i.e., the learner can cope optimally with most produced instances of vowel duration, some further boundary adjustments aided by the new lexical representations will need to be made. This is because the formation of phonological length categories may lead to a non-optimal category boundary between /short/ and /long/. This non-optimal perception results from a non-optimal ranking of the one-dimensional cue constraints that map duration values onto /short/ or /long/. The learning mechanism that underlies this further optimisation of L2 perception is demonstrated in Tableau 5.3.

[100 ms]	100 ms not /long/	100 ms not /short/
√ /long/	*!→	
* ☞ * /short/		←*

Tableau 5.3. Lexicon-driven adjustment of the L2 category boundary for vowel length.

Here we can see that the learner's boundary shifting device, viz. the GLA, will now gradually adjust the ranking of these cue constraints (as depicted by the arrows) every time the perceived length category does not match the intended item that the learner will recognize through the semantic component of her lexical representations. For instance, if an intended SBE /ʃip/ *sheep* is pronounced with an appropriate duration value *x* but the L2 perception grammar strongly prohibits a /long/ perception, i.e., if '*x* is not /long/' is highly ranked and facilitates a /short/

perception, the learner will have a non-optimal perception of such a token because she will likely perceive it as /ʃ-short-p/. However, the semantic component of the relevant lexical item will tell her that the intended word was /ʃ-long-p/ so that the output of her perception and the word that she recognizes will be in conflict. This perception-recognition mismatch will trigger the GLA constraint reranking which will reduce the instances of perception-recognition mismatches. Importantly, this lexicon-driven category boundary adjustment is also an L1-like learning mechanism as described in § 2.1.3.

Up to this point, the learner will have achieved quite an accurate perception of the two English high front vowels. However, there is a further development towards optimal perception, namely cue integration. Recall that the SBE vowels are produced with F1 and duration differences in such a way that an optimal SBE listener relies on both dimensions. The question now is how an L2 learner can learn to integrate F1 differences in the perception of these vowels. It is proposed that learners will later introduce more abstract cue constraints that relate any F1 value with any vowel height category and, crucially, with any vowel length category. As was mentioned in Chapter 2, these constraints are called multi-dimensional cue constraints. Also, constraints that relate duration values with length and height categories will be introduced. These more abstract cue constraints in the perception grammar are, for example, '300 Hz is not /short/' or '150 ms is not /long/'. The availability of such cue constraints shows that the categories /short/ and /long/ have become as abstract as the categories /ɪ/ and /i/ because they can be the targets of more than a single auditory dimension. Thus, learners can start to rely on both F1 and durational differences when categorizing vowels as /short/ or /long/.

On the other hand, in the case of Spanish learners, the optimal ranking of the new cue constraints to achieve a diagonal boundary in categorization, i.e., cue integration, will require the perceptual learning of all the vowel categories in the L2. This is because the relationship between F1 and duration values with all possible categories will need to be considered for the constraint reranking. Therefore, this development is predicted to require extensive L2 exposure. Escudero & Boersma (2004b: 579) suggested that because all SBE high vowels (/i u/) are long and all semi-high vowels (/ɪ ʊ/) are short, cue integration can be learned. That is, cue constraints such as '260 Hz is not /short/' and '380 Hz is not /long/' will become high-ranked because 260 Hz is unlikely to refer to a /short/ vowel and 380 Hz is unlikely to refer to a /long/ vowel. Thus, distributional learning will allow cue constraints that refer to F1 values below 314 Hz, which is the most frequent ambiguous production of SBE /i/ and /ɪ/, to be categorized as /long/ and values above it as /short/, as shown in Figure 5.16.

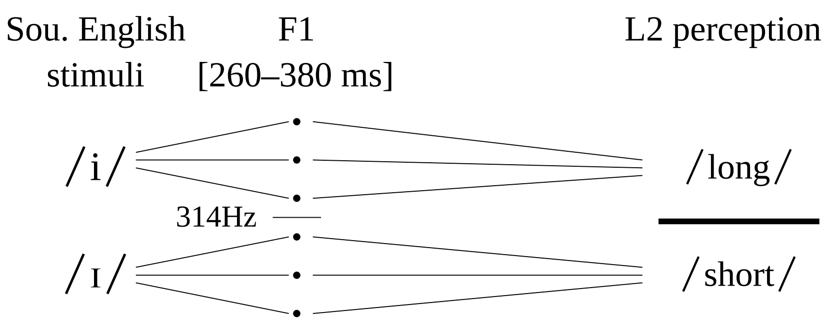

Fig.5.16. Distributional mapping of F1 values onto the abstract phonological categories /short/ and /long/.

The learning mechanisms and the L2 development described above can be seen as constituting a step-like developmental path, as was described in Escudero & Boersma (2004b). Such a path in the development of new L2 sounds can be summarized in the steps mentioned below which elaborate on those given in Tables 2, 3, and 4 (cf. § 5.1).

Step 1: L1 category and grammar copying (Full Copying)

This leads to the phonemic equation of two L2 sounds with one L1 sound as well as to perceptual mappings that output a single category.

Steps 2, 3, 4: auditory-driven category formation

The L2 perceptual task is resolved by creating two new categories along a non-previously-categorized dimension via GLA auditory learning. First, auditory ranked constraints and auditory-phonetic discrete categories are formed. Then, the categories are turned into abstract representations through phonological abstraction. To be able to map onto those abstract categories, the constraints are translated into one-dimensional cue constraints, such as ‘60 ms is not /short/’. At this point, the learners are able to use two phonological categories to represent words in their L2 lexicon.

Step 5: lexicon-driven category boundary adjustment

The learners then adjust or shift their duration category boundary through GLA lexicon-driven learning. That is, once the learners have vowel length categories in their L2 lexicon, the GLA will gradually change possible non-optimal constraint rankings due to perception-recognition mismatches.

Step 6: perceptual cue integration of F1 and duration values

The learners can achieve the optimal diagonal boundary through the introduction and optimal ranking of multidimensional cue constraints that map any auditory dimension onto any phonological category, e.g., '300 is not /short/'.

5.2.5 Ingredient 5: predicting the L2 end state

The L2LP model proposes that the role of the input is more important than the lack of cognitive plasticity. This is because rich L2 input can overrule the reduced level of plasticity in adult learners. Given that L2 learners are unlikely to get the same type of input as L1 learners, an L2 end state that does not match the optimal target perception cannot be ascribed to maturational constraints until we can be sure that the learner has been exposed to a rich production environment, i.e., one that is at least equal to the one found in L1 acquisition.

Most importantly, the L2LP model's interpretation of Grosjean's (2001) language mode continuum can be referred to as the *language mode activation hypothesis* according to which the parallel activation of separate systems can lead to intermediate responses. This is because the output of the two perception grammars can be merged when a categorization response is given. In other words, intermediate L1-L2 sound perception is a consequence of the activation of the learner's two perception grammars during online perception rather than being a consequence of having a single set of categories and perceptual mappings for the two languages. In sum, the L2LP model predicts that L2 learners have different perception grammars for their two languages (cf. § 3.5.3) and that intermediate perception is the result of parallel activation during online speech perception. Thus, three different language modes are predicted to be found in L2 learners depending on whether they are in a monolingual L1, completely bilingual, or monolingual L2 mode.

In the case of the Spanish learners of English /i/ and /ɪ/, the new length distinction will not be used when perceiving L1 vowels because this distinction was created in the L2 perception system to optimally cope with the L2 production environment. Therefore, it is predicted that no spurious contrasts will emerge in the L1 perception of Spanish learners who have SBE as their target L2. Figure 5.17 shows the three predicted modes for intermediate Spanish learners of SBE who have a length distinction in their L2 perception grammars.

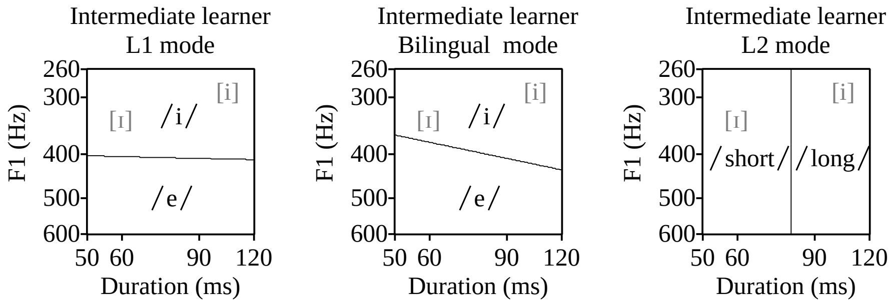

Fig.5.17. Predicted perception modes in intermediate Spanish learners of SBE confronted with three different language settings.

5.3 Evidence: Spanish learners of Southern British English (SBE)

Table 5.7 summarizes the L2LP model's predictions for each of the five ingredients involved in L2 sound perception, viz., the optimal perception in the two languages, the three logical states in language acquisition, and the L2 learning task. A summary of how the model describes the sound perception in these components is also given.

Table 5.7. Summary of the L2LP model's predictions for the new scenario.

L2LP ingredients	Predictions for new	Description
Optimal L1 & L2	SBE and Spanish listeners will exhibit optimal L1 perception	Optimal category boundaries and optimal cue reliance
Initial state	Beginning Spanish learners will be equal to monolingual Spanish when listening to SBE	Both types of listeners will exhibit L1 category boundaries, L1 cue reliance, and L1 categories
Learning task	Create new categories along the duration continuum + learn cue integration	
Development	L1-like auditory-guided learning followed by lexicon-driven learning	Category formation and category boundary shifts
End state	Spanish learners will attain optimal L2 perception and, at the same time, will maintain their optimal L1 perception	Two length categories in their L2 mode, a single height and no length category in their L1 mode

To test the model's five main predictions, I will provide perceptual data from the categorization of English /i/ and /ɪ/ by SBE listeners, monolingual Spanish listeners, and Spanish learners of SBE. Although some of these data were previously reported in Escudero (2001, 2002) and Escudero & Boersma (2004b), this is the first time that the full set is documented in the context of the L2LP model's theoretical and methodological proposal. In the following sections, each of the model's predictions will be tested against the available data. In § 5.3.1, the optimal perception hypothesis will be put to the test by showing that real SBE and Spanish listeners are optimal perceivers of their own production environments. In § 5.3.2, the cross-language and initial L2 perception of Spanish learners will be explained. Then, in § 5.3.3, data from non-beginning learners will be presented to test the predicted learning task and L2 development. Finally, a discussion of the findings will be provided in § 5.3.5.

5.3.1 Model ingredient 1: Spanish and Southern British English (SBE) perception data

To test the optimal perception hypothesis, I present here the results of studies that have measured the perception of the same synthetic vowel stimuli by SBE and Spanish listeners. Escudero (2001) tested 21 SBE listeners (10 females and 11 males) who were between 19 and 55 years of age and had lived in different areas of the South of England for most of their lives. For Spanish, Escudero (in progress a) tested the perception of 32 monolingual Peruvian Spanish listeners (16 females and 16 males) who had lived in Lima for most of their lives, had little or no knowledge of English, and were between 18 and 28 years of age. Ten of the SBE listeners were tested at the University of Edinburgh and the other 11 were tested at the University of Reading in Southern England, while the Spanish listeners were tested at the Pontificia Universidad Católica del Perú in Lima. Thus, most subjects were tested in their country of origin which, according to Beddor & Gotfried (1995), normally provides a good measure of monolingual perception.

These listeners were presented with the isolated synthetic vowels shown in Figure 5.18. Synthetic stimuli were used because they allow for the varying of acoustic dimensions separately and in equal steps, which is needed for relative cue reliance studies. Note that many other cue reliance studies have also used synthetic stimuli, e.g., Bohn (1995), Flege, Bohn & Jang (1997), Gerrits (2001), Nittrouer, Manning & Meyer (1993). Although Gerrits (2001) suggested that subjects tend to categorize natural stimuli more easily than synthetic stimuli, another study by Nittrouer (2001) has shown that neither children nor adults are adversely affected by the use of synthetic stimuli.[46]

[46] Also, the use of isolated vowels avoids the contextual effects reported in the cross-language perception literature, e.g., Gottfried (1984), Gottfried & Beddor (1988), and Strange et al.(2001), because it makes listeners rely on their abstract representation of the vowels involved.

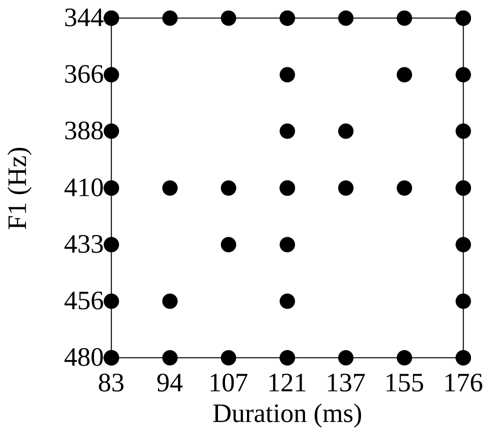

Fig. 5.18. The 37 synthetic stimuli presented to SBE and Spanish listeners.

Using the KLATT synthesizer, the stimuli were synthesized on the basis of the auditory properties of natural exemplars of the vowels /i/ and /ɪ/ which were produced 10 times each by two Scottish English (SE) speakers. The average F1 of the naturally produced vowels was 484 Hz for /ɪ/ and 343 Hz for /i/. The vowel synthesis also included the F2 dimension in order to produce more natural vowel tokens, the average F2 being 1890 Hz for /ɪ/ and 2328 Hz for /i/. These values were taken as the basis for the top and bottom edges of the stimulus rectangle. The six vertical steps, which led to seven different F1 values, were equal on the mel scale, ranging from 480 to 344 Hz, while the six horizontal steps, which led to seven duration values, ranged from 83 ms to 176 ms (in six equal fractional steps of 1.1335). However, this stimulus square does not match the production distributions of either of the languages. In particular, the F1 values are very different from the SBE productions, which range from 260 to 380 Hz (cf. Figure 3), while the duration differences do not match the Spanish environment which has only small differences in vowel duration. The SBE listeners were presented with the 37 synthetic stimuli shown in Figure 5.18. The Spanish listeners were presented with the same synthetic stimuli but they heard the 49 possible points in the square, i.e., the 37 tokens shown in the figure plus the missing 12 tokens that together form a 7 x 7 matrix.

The SBE listeners were presented with the 37 stimuli 10 times in different randomized orders. The Spanish listeners heard all 49 stimuli only three times because of the larger number of stimuli and subjects. All listeners were asked to categorize the stimuli they heard as either of two vowel categories, i.e., /i/ or /ɪ/ for the English listeners, and /i/ or /e/ for the Spanish listeners. The subjects were told to guess in the case of uncertainty, and to take as long as they wished to make a decision. If the optimal perception hypothesis is correct, we would expect the 21 Southern listeners and the 32 Spanish listeners to exhibit category boundaries and cue reliance equal to those associated with the optimal perception shown in § 5.2.1 (cf. Figure 5.3). However, it is worth noting that the SBE subjects had to categorize the tokens as either /i/ or /ɪ/ despite the fact that some of the F1 values in the stimuli (most likely the ones above 400 Hz) fell outside their production distributions, something which may have had an effect on their vowel categorization performance.[47]

Figure 5.19 shows the average boundary location and cue reliance for the two groups of SBE listeners and for the Spanish listeners. In the squares, tokens with values falling above the boundary line were mostly categorized as /i/ by all listeners, while tokens below this line were mostly perceived as /ɪ/ by the SBE listeners and as /e/ by the Spanish listeners. For the SBE subjects, the values in the 12 cells that were not measured were interpolated from the values in the neighbouring cells to get a continuous representation of their category boundary. The duration reliance is computed as the percentage of /i/ responses along the right edge of the stimulus rectangle minus the percentage of /i/ responses along the left edge, while the F1 reliance is computed as the percentage of /i/ responses along the top edge minus the percentage of /i/ responses along the bottom edge (cf. § 3.1.2 and § 5.2.1).

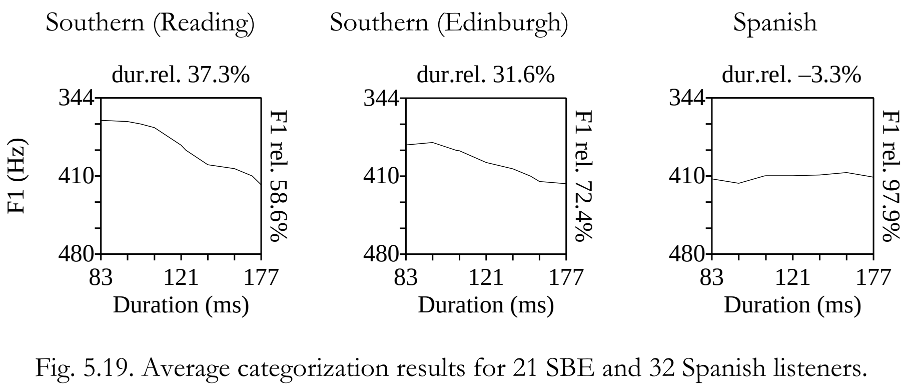

Fig. 5.19. Average categorization results for 21 SBE and 32 Spanish listeners.

[47] See § 5.3.4 for a discussion of this stimulus set effect in categorization.

Thus, as described in § 5.2.1, we can compute the slope of the boundary line from the cue reliance, which will allow us to compare real listeners to the predicted optimal perception in each language. The SBE listeners tested in Reading have a duration reliance of 37.3% and an F1 reliance of 58.6%, and the slope can be computed as the ratio of these percentages expressed in terms of the duration and F1 ranges of the stimulus square (cf. § 5.2.1), which is 37.3% · log2(480/344))/ (58.6% · log2(177/83)) = 0.28 octaves/duration doubling. The category slope for the SBE listeners tested in Edinburgh is 0.19. Both SBE slopes differ considerably from the predicted optimal slope of 0.98. Escudero & Boersma (2003) suggested that this difference was due to the particular ranges of F1 and duration which were used to define the stimulus square in this study, and which differ considerably from the ranges used to predict the optimal SBE perception. This and other reasons for the difference in the value of the slope of the category boundary between real and optimal listeners are further discussed in § 5.3.4. By contrast, the slope of 0.015 of the category boundary for the Spanish real listeners closely matches the predicted optimal boundary slope of 0.021.[48]

The values for the slope of the category boundary line can be taken to determine whether listeners rely on a single cue, i.e., either duration or F1, or on both these cues when categorizing the vowels. This way of analysing the data differs from the analysis in Escudero & Boersma (2004b) in that it only concentrates on a single-cue or cue integration pattern in the listeners' perception. That is, the analysis considered here simplifies Escudero & Boersma's division based on the amount of reliance on each cue by using only the number of cues. Thus, only three types of cue reliance are considered, namely 'single-cue: duration', 'single-cue: F1' and 'cue integration'. The use of duration as the single cue for categorizing vowels is characterized by a boundary slope with a value higher than 1.1 octaves/duration doubling and by a vertical boundary shape. If a boundary slope has a value very close to 0 or, more specifically, below 0.1, and if the shape of the boundary is a horizontal line, it can be said that a listener uses F1 as the single cue for categorizing vowels. A diagonal boundary shape at whatever location in the boundary line and a slope value between 0.1 and 1.1 show that the listener integrates both dimensions to categorize the vowels, although each dimension may be used to different degrees.

48 The actual value of the Spanish listeners' average slope is –0.015 octaves/duration doubling. However, here I use the absolute values of both the cue reliance and the slope, which differs from the analysis presented in Escudero and Boersma (2004b: 559).

We can apply these three types of listeners' perceptual cue use to classify the cue reliance in SBE and Spanish. According to their average results, both SBE groups are 'cue integration' listeners, and so is the optimal SBE listener. In contrast, both the Spanish real listeners and the optimal Spanish listeners have a 'single-cue: F1' perceptual cue use. Table 5.8 shows the individual results of the 21 SBE and the 32 Spanish listeners divided according to their type of perceptual cue use.

Table 5.8. Individual reliance type for the SBE and Spanish listeners.

Perceptual cue use	Slope of the category boundary	Southern (Edinburgh)	Southern (Reading)	Spanish
Single-cue: duration	> 1.1. oct./durdou	0	1	0
Cue integration	0.1 – 1.1 oct./durdou	7	7	0
Single-cue: F1	< 0.1 oct./durdou	3	3	32
	Total listeners:	10	11	32

Here we can see that the majority of SBE listeners integrate both dimensions whereas all Spanish listeners use only F1 to categorize vowels. Crucially, although the category slope of all 21 SBE listeners differs considerably from the predicted optimal SBE slope, the majority of the subjects are classified as 'cue integration' listeners.[49] Thus, it can be concluded that real listeners exhibit perceptual behaviour that is comparable to the predicted language-specific optimal perception.

5.3.2 Model ingredient 2: cross-language and initial L2 perception data

In this section, Spanish monolingual perception is compared to that of beginning learners of English in order to test the Full Copying hypothesis which states that

[49] See § 5.3.4 for a discussion on why the average SBE listener has a smaller than predicted duration reliance, i.e., 37.3% or 31.6% rather than 85.71 which was the predicted optimal duration reliance. This rather small duration reliance leads to a lower than predicted boundary slope.

beginning L2 learners will have the same perception as monolingual L1 listeners. The learners mentioned in this section were tested at the University of Edinburgh and their perception results were first reported in Escudero (2001). Escudero & Boersma (2004b: 562) labelled these listeners as beginners on the basis of their length of residence in an English-speaking country and their amount of English instruction. That is, a language background questionnaire revealed that 6 out of the 30 learners tested had lived in an English-speaking country for less than two months and had not received higher education in English, neither in the form of a university degree or a proficiency certificate. Table 5.9 shows the 6 beginning learners' origin, education in English, and time spent in four different English-speaking countries.

Table 5.9. Six beginning learners of English. - = no higher English education reported).[50]

Subject	Origin	High education in English	Months (Scotland)	Months (England)	Months (Ireland)	Months (Zimbabwe
Al	Spain	-	0	0	0	0
dmc	Spain	-	0	0	0	0
Jad	Spain	-	0	0	0	0
Jg	Spain	-	0	1	0	0
Mt	Spain	-	0	0	0	0
Mw	Argentina	-	1	0	0	0

These beginning learners were presented with 10 repetitions of the 37 synthetic stimuli shown in Figure 5.18. Most of them did not seem to know the difference between the pronunciation of the two vowels in question (though they must have learned the orthographies during formal instruction) so that they must have responded with the only distinction their L1 perception would allow, viz., that between Spanish /e/ and /i/. In addition, they must have used their L1 perceptual reliance, which is based on F1 differences alone. This is confirmed by the most of the results shown in Figure 5.20.

[50] Irish English and Zimbabwean English high front vowels have the same properties as SBE and SE high front vowels respectively (cf. Wells 1982a, 1982b).

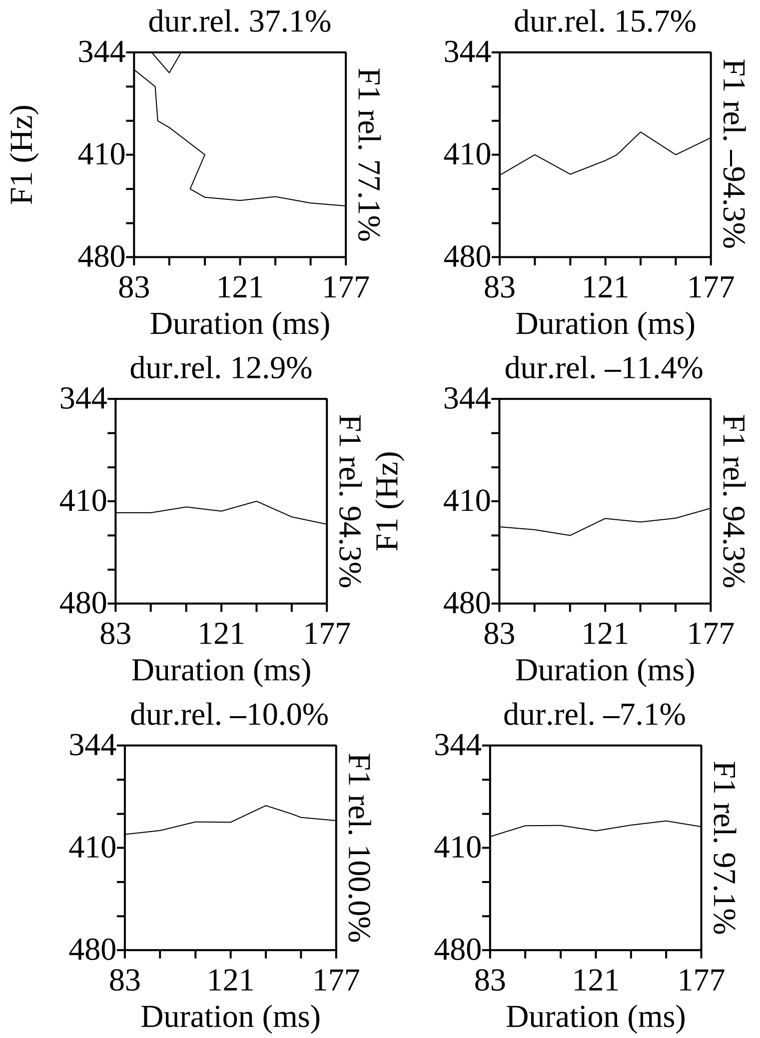

Fig. 5.20. Category boundaries and cue reliance for the six beginning learners.

Table 5.10 below shows the beginners' individual results together with those for the average monolingual Spanish listener. It can be said that most beginning learners show an L2 perception that matches the monolingual Spanish perception, i.e., a horizontal boundary shape and a 'single-cue: F1' perceptual cue use. This finding supports the L2LP model's prediction that initial L2 perception will be equal to cross-language or native perception. Given that their L2 perception matches the optimal perception of their L1, these learners differ considerably from the optimal target L2 perception if SBE is their target language. This is because they are 'single-cue: F1' listeners rather than 'cue integration' listeners like the SBE real and optimal listeners.

Table 5.10. L2 perception for six beginning learners.* = reversal of labels (*ship* and *sheep*). Adapted from Escudero & Boersma (2004b: 561).

Subject	Duration reliance (%)	Spectral reliance (%)	Slope of the category boundary	Perceptual cue use
Spanish	-3	98	0.01	Single cue: F1
Beginners:				
jg	37	77	0.21	Cue integration
*mw	16	-94*	0.08	Single cue: F1
al	13	94	0.06	Single cue: F1
*dmc	-11	94	0.05	Single cue: F1
mt	-10	100	0.04	Single cue: F1
jad	-7	97	0.03	Single cue: F1

The perceptual data presented above partly confirm the L2LP model's prediction for the cross-language perception and L2 initial state. However, the stimulus square presented to the Spanish monolinguals and the beginning learners did not match the specific production distributions of SBE because they were based on SE productions, as was described in § 5.3.1.[51] This entails that the cross-language and initial L2 predictions specific to SBE will need further investigation since it was predicted that SBE /i/ and /ɪ/ would be categorized as the single Spanish category /i/ by both monolingual Spanish listeners and beginning L2 learners. Crucially, this prediction has implications for the predicted learning task and development for Spanish learners of the SBE vowels. Therefore, a test was conducted to see whether, in fact, Spanish monolingual listeners would generally perceive a single

[51] Importantly, the Spanish optimal perception compares well to the Scottish English production distributions, as will be discussed in Chapter 7, something which suggests that the Spanish learners equated the response category /ɪ/ to their Spanish /e/.

vowel when listening to either natural or synthetic stimuli with the SBE production distributions.

To measure the monolingual Spanish perception of SBE /i/ and /ɪ/, natural and synthetic stimuli were presented as part of the study that appears in Escudero (in progress b). With respect to the natural stimuli, 64 Peruvian listeners (including the 32 mentioned in the previous section) were presented with 24 natural tokens of the SBE vowels /i/ and /ɪ/ which were drawn from the corpus reported in Escudero & Boersma (2003). Figure 5.21 shows how these natural SBE tokens compare to the Spanish vowel averages reported in Cervera et al. (2001).

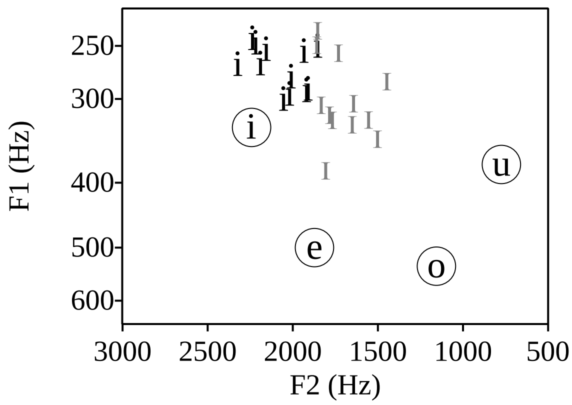

Fig. 5.21. Spanish average productions (circled) and the SBE /i/ and /ɪ/ tokens (in grey).

As for the synthetic stimuli, the 32 Peruvian listeners mentioned in the previous section were presented with the stimulus square in Figure 5.22. Note that this square has three steps of lower F1 values in comparison with the square in Figure 18. That is, the stimulus set is made up of the four continua of the previous square plus three extra continua with lower F1 values. These stimulus values for F1 fall within the production distributions of SBE, and therefore it was thought that the perception of these stimuli would give a further measure of how Spanish listeners perceive typical SBE /i/ and /ɪ/ tokens. The 7 new steps were also equal on the mel scale, ranging from 281 to 410 Hz. In addition, the duration values were kept equal in order to test if the Spanish listeners use the duration dimension at all.

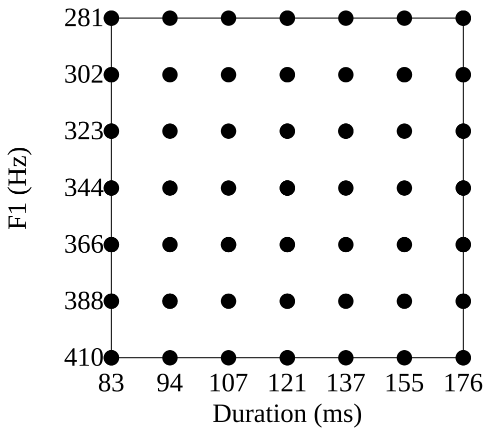

Fig. 5.22. The 49 synthetic stimuli that fall within SBE distributions for /i/ and /ɪ/.

Recall that the aim was to test whether natural and synthetic tokens of SBE /i/ and /ɪ/ would indeed be mostly mapped onto Spanish /i/, given that their F1 values fall within the distributions of that Spanish vowel (cf. Figure 5.3). Given that these tokens are typical instances of the SBE vowels, it was assumed that an optimal SBE listener would categorize most intended /i/ tokens as /i/, and most intended /ɪ/ tokens as /ɪ/. By contrast, it was hypothesized that the Spanish listeners would perceive SBE /ɪ/ tokens as Spanish /i/.

For the natural stimuli, the 64 Spanish listeners were presented with two repetitions of the SBE tokens, half of them intended as /i/ and the other half as /ɪ/ by a native speaker of SBE. They were provided with a computer screen that showed the five Spanish vowel monophthongs, namely <i, e, a, o, u>, and they were asked to click on the one they thought they had heard. Before the actual test, they had had a practice session of ten tokens. As for the task with synthetic stimuli, the 32 Spanish listeners heard the 49 stimuli three times and were asked to categorize them as either of the two Spanish vowels /i/ and /e/. Figure 5.23 shows the results of the Spanish listeners' categorization of the natural SBE tokens.

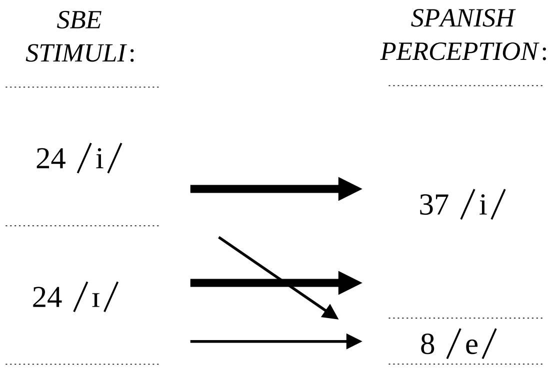

Fig. 5.23. Cross-language Spanish categorization of the natural tokens of Scottish /i/ and /ɪ/.

Here we see that, as predicted, the majority of SBE tokens were categorized as Spanish /i/. This results in a token-based category boundary that is 14 to 16 tokens lower than the SBE equal-likelihood line in production as established by the speaker's intended vowel. The reason is that only 8 tokens were perceived as /e/ instead of 24, i.e., 17% instead of 50%, thus suggesting that these listeners do not have a two-category assimilation similar to that of the optimal SBE listener. Table 5.11 reveals that although some tokens of both SBE /i/ (3 tokens) and /ɪ/ (5 tokens) were categorized as Spanish /e/, the predicted categorization of typical SBE tokens of both /i/ and /ɪ/ as Spanish /i/ was borne out. This means that Spanish listeners use their optimal Spanish perception when categorizing foreign sounds as their own vowels.

Table 5.11. Token averages and standard deviations (SD) for the cross-language categorization of the SBE 24 /i/ and 24 /ɪ/ tokens.

	Sp. /i/ Mean	SD	Sp. /e/ Mean	SD	Tokens
English /i/	20.55	3.43	2.59	2.69	24
English /ɪ/	16.09	4.49	5.42	3.80	24
Total tokens	36.64/48		8.01/48		48

With respect to the synthetic stimuli, Figure 5.24 shows how Spanish listeners categorized the stimulus square with values that matched the SBE production distributions for /i/ and /ɪ/.

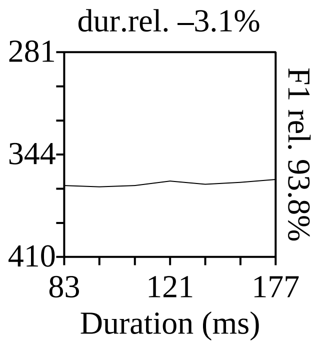

Fig. 5.24. Peruvian perception of the 49 synthetic stimuli.

As we can see, the Spanish listeners rely exclusively on duration and have a boundary line that is located higher (at 366 Hz instead of 408 Hz) than the optimal Spanish boundary. Nevertheless, most stimuli are categorized as Spanish /i/, and this is in striking contrast with the categorization of the square with higher F1 values wherein the same Spanish listeners classify half of the tokens as Spanish /i/ and the other half as /e/. From the monolingual Spanish categorization of the natural SBE stimuli, it can be concluded that Spanish learners classify typical SBE /i/ and /ɪ/ tokens as Spanish /i/. With respect to the synthetic stimuli, the monolingual Spanish listeners have a horizontal boundary with a non-existent or negative reliance on duration. This suggests that the Spanish listeners use their optimal cue reliance to categorize the SBE vowels. However, it has also been found that category boundaries can shift depending on the extent of the edges of the stimulus square. This entails that the Spanish listeners adjust their boundaries in relation to the auditory characteristics of the stimulus set that they are presented with, which in this case had unusually low F1 values.[52] The question now is whether Spanish learners of SBE can develop target-like L2 category boundary and perceptual cue use, and if so, how this can be achieved.

---

[52] The effect of the stimulus set in categorization will be discussed in § 5.3.4.

5.3.3 Spanish learners' development and end state

Escudero & Boersma (2004b) confirmed that the Spanish learners of English mentioned in Escudero (2001) had two different target L2 dialects, namely SBE and SE. In this section, I report on the perception of the 18 learners who turned out to have SBE (or a similar dialect) as their target L2. The learners' target dialect was determined on the basis of their length of residence in the south of England and the education they received in SBE, which was established through a language background questionnaire. The learners had had 1 to 15 years of formal English instruction in their home countries and had spent 2 months to 27.5 years in the south of England. Also, they reported using Spanish and English with native speakers of each language. The learners that had received higher education in the English language were studying towards a university degree in either English philology or diplomacy, or they had received an English proficiency degree such as the Cambridge University First Certificate or Advanced Certificate in English. Crucially, the variety of English used in those education settings in Spain is primarily SBE. Table 5.12 shows the background information for the 18 non-beginning learners of SBE.

Table 5.12. Eighteen Spanish learners of SBE. FCE = First Certificate in English, EP = student of English philology, CAE = Certificate in Advanced English, Dip = Diplomacy, - = no higher English education reported). Adapted from Escudero & Boersma (2004b: 561).[53]

Subject	Origin	Higher education in English	Months in Scotland	Months in England	Months in Ireland	Months in Zimbabwe
mao	Spain	Dip	0	1	0	0
ef	Spain	EP	0	0	2	0
manl	Spain	EP	46	0	0	0
of	Colombia	-	1	108	0	0
mvl	Colombia	-	48	36	0	0
pn	Spain	-	5	13	0	0
mc	Spain	EP	0	0	2	0
af	Spain	EP	0	0	0	0
ct	Spain	FCE	0	0	6	0
arg	Spain	EP	0	0	0	0
snd	Spain	EP	0	0	0	0
jtn	Spain	CAE	0	1	0	0
fjrg	Spain	EP	0	0	0	0
lj	Spain	EP	0	6	0	0
abg	Spain	-	0	2	0	0
mcsc	Spain	EP	0	1	0	0
jr	Venezuela	-	6	312	0	0

These L2 learners were subjected to exactly the same experimental procedure as the one used for the beginning learners described in § 5.3.2 and for the SBE listeners of § 5.3.1. They categorized the 37 synthetic tokens in Figure 5.18, which were presented 10 times each, as either English /i/ or /ɪ/. The prediction was that the

[53] Subject mvl had lived in the South of England before moving to Scotland, and had not used English to any large extent during her time in Scotland because she worked as a Spanish language teacher and spoke Spanish at home. Therefore, it seemed that her target L2 had not changed from SBE to Scottish English.

advanced learners with an SBE target would classify the vowels either on the basis of their duration differences alone or on a duration and F1 reliance, just like optimal SBE listeners. Table 5.13 shows their duration and spectral reliance, their boundary slope, and their perceptual cue use compared to the results for the SBE listeners.

Table 5.13. Cue reliance and boundary slope in the non-beginning learners of SBE.

Subject	Duration reliance (%)	Spectral reliance (%)	Slope of the boundary	Perceptual cue use
Southern (Edinburgh)	31.6	72.4	0.19	Cue integration
Southern (Reading)	37.3	58.6	0.28	Cue integration
Non-beginning learners:				
mc	79	-3	11.58	single-cue: duration
pn	99	-4	10.89	single-cue: duration
mvl	86	-4	9.46	single-cue: duration
of	99	-7	6.22	single-cue: duration
manl	90	-7	5.65	single-cue: duration
ef	86	-10	3.78	single-cue: duration
af	87	11	3.48	single-cue: duration
ct	67	9	3.28	single-cue: duration
mao	96	-17	2.48	single-cue: duration
arg	94	20	2.06	single-cue: duration
snd	89	20	1.96	single-cue: duration
jtn	99	27	1.61	single-cue: duration
fjrg	94	27	1.53	single-cue: duration
lj	70	36	0.86	Cue integration
abg	40	47	0.37	Cue integration
mcsc	29	81	0.16	Cue integration
*jr	13	-100*	0.06	single-cue: F1

As we can see, the majority of these advanced learners manifested a 'single-cue: duration' perceptual cue use, which is associated with a vertical boundary line. This is a very different strategy from that used by SBE listeners, which shows a diagonal shape. However, it was predicted that learners would use the duration differences in the stimuli to categorize the SBE vowels because they formed a length contrast as a first step in their L2 acquisition. The use of duration as the single cue for categorizing the vowels can be interpreted as the representation of the SBE contrast as one of length. On the other hand, three of the learners showed an integration of F1 and duration when perceiving the L2 vowels. This could be interpreted as the further L2 perceptual development predicted in § 5.2.4, which is the integration of F1 differences in the classification of the new length categories. However, other explanations for this apparent cue integration can be offered, as will be discussed in the next section.

5.3.4 Discussion

With respect to the monolingual perception, we have seen that the SBE listeners manifested a larger F1 reliance and a smaller boundary slope than those of the predicted optimal SBE perception. Escudero & Boersma (2003) gave a number of reasons for this discrepancy, the most important of which will be discussed here. First, the stimulus square presented to these listeners did not match the F1 distributions of their production environment but rather those of the F1 ranges of SE. In other words, the F1 ranges for high front vowels were much larger than those of the native SBE speakers. This may have enhanced the listeners' awareness of the F1 cue, and may therefore have either reduced their category slope or enhanced their F1 reliance. One could solve this problem by testing listeners with stimulus sets that match their distributions such as the stimuli shown in Figure 22. Second, the perception experiment had spectral information as the first available cue, which may have contributed to a smaller boundary slope. Also, the stimuli used were isolated vowels which are known to prevent listeners from normalizing away the influence of speaking rate on duration.

Third, the optimal cue reliance is sensitive to the standard deviations used for computing the equal-likelihood line which matches the optimal boundary. It is not known, however, to what extent listeners compensate for consonant environment, number of syllables, stress, or speaking rate. Escudero & Boersma (2003) suggested that if the standard deviation for duration is doubled to 0.80 instead of 0.4 duration doublings, or if the standard deviation for F1 is halved to 0.10 octaves, the optimal listener will have a category slope of about 0.25 oct/dur.doubling, which is equal to that of the real listeners. Finally, real listeners normally have contact with multiple dialects and can normalize for dialectal differences, whereas the predicted optimal perception does not incorporate dialectal normalization. For instance, the SBE listeners who had knowledge of the SE dialect may have adjusted their perception to this stimulus square. It is telling that the SBE listeners tested in Edinburgh had a less steep boundary slope (or, equivalently, more F1 reliance) than those tested in Reading, which suggests that the more contact listeners have with a particular dialect, the more they are likely to normalize for dialectal differences.

With respect to the stimulus set effect in the monolingual Spanish listeners, it was found that the F1 range of the stimulus set affected the location of their boundary. This is because this boundary was at the optimal location (410 Hz) with the 344 Hz to 480 Hz range shown in Figure 20, but it was at about 40 Hz higher with the 281 to 410 Hz F1 range shown in Figure 24. This seems to be a *stimulus range* which has been reported before for VOT categorization (cf. Keating et al. 1981). Boersma (2004) showed that this kind of effect could be modelled with Stochastic OT. However, the exact formal modelling of this effect in connection with the optimal listener and the exact threshold of the boundary shift are not yet available.

As for L2 development and the L2 end state, three of the learners showed the target L2 perceptual cue use, i.e., cue integration which seems to suggest that it is possible to reach the optimal target L2 perception. This finding clearly matches the L2LP model's prediction for L2 development, though a number of factors that may affect this interpretation must be taken into account. For instance, the cue integration and diagonal boundaries observed in three learners may be explained by assuming that they had a parallel activation of their two languages during the perception experiment, which would be the predicted bilingual perception shown in § 5.2.5. In other words, these learners may have had a greater L2 duration reliance and a greater category boundary slope than the one attested, but due to the parallel activation of their L1, their L2 perception may appear to be intermediate between the L1 and L2.

Also, these learners could have been affected by the stimulus set in the same way as the SBE listeners because the unusual F1 range may have triggered an unusual reliance on F1 differences in the learners' L2 perception. However, neither explanation would seem likely when we consider the possibility of the auditory perception of duration differences. That is, these learners may have sometimes paid attention to such differences through their psychoacoustic perception and not through their linguistic perception. Although the learners performed vowel identification, a task that triggers linguistic sound perception, the isolated nature of the stimuli may have triggered non-speech (or psychoacoustic) perception. These possible explanations would need to be tested with a more controlled longitudinal L2 development study.

5.4 Learning new sounds: L2LP predictions versus the evidence

At this point, it is worth reviewing the model's predictions for the new scenario and of the findings that support them. Table 5.14 summarizes the predictions that follow from the five ingredients of the L2LP model. These ingredients refer to the three logical states in language acquisition, to the optimal perception in the two languages, and to the L2 learning task. In this table, each of the results that were shown in the previous section is presented next to its specific prediction.

Table 5.14. The five predictions for a new L2 sound perception scenario and the evidence to support them.

L2LP ingredients	Predictions for new	Finding
Optimal L1 & L2	SBE and Spanish listeners will exhibit optimal L1 perception	Partially borne out
Initial state	Beginning Spanish learners will be equal to monolingual Spanish when listening to SBE	Borne out
Learning task	Create new categories along the duration continuum + learn cue integration	Borne out
Development	L1-like auditory-guided learning followed by lexicon-driven learning	Indirectly borne out
End state	Spanish learners will attain optimal L2 perception and, at the same time, will maintain their optimal L1 perception	No data available

The optimal perception hypothesis turns out to be partially borne out because stimulus set effects were at play for both SBE and Spanish listeners. This hypothesis will have to incorporate listeners' normalization for different ranges in the dimensions with which sound categories are produced. Regarding the L2 initial state, the beginning L2 learners behaved just like the monolingual Spanish listeners, a finding that confirms the model's prediction for this L2 state.

With respect to L2 development, we saw that L2 learners went through a stage in which they categorized the SBE vowels according to their duration differences, i.e., through a length distinction which they must have created to cope with this new L2 sound perception scenario. This indirectly suggests that they created the new duration categories through auditory-driven learning. However, to test whether this was the case, we would need to make sure that monolingual Spanish listeners do not have a single duration category. This could be done by collecting goodness of fit values for different duration values because category centres or most prototypical tokens normally receive higher ratings. Crucially, the majority of learners used a single cue namely duration, to categorize the L2 vowels, which shows that the integration of F1 to the length distinction is a difficult L2 learning task to handle. However, this does not mean that L2 learners cannot perform the cue integration task at all in order to become optimal target L2 listeners. As suggested in § 5.3.4, more data on L2 development, i.e., more subjects with higher levels L2 proficiency, as well as better control for proficiency levels would be required in order to rigorously test the model's developmental predictions. Also, data for the learners' L1 perception would be required to test the end state hypothesis for this scenario. These data could be collected following the experimental setting that includes perception modes.

All in all, these series of findings give enough reason to believe that the L2LP model can predict the cross-language and L2 perception in a new scenario. Crucially, the model's explanations for these predictions provide the linguistic knowledge that underlies the attested perceptual behaviour. The next two chapters will present an application of the model to two other scenarios in L2 perception, viz., the subset and the similar scenarios, together with empirical evidence to test the specific predictions.

6 Learning SUBSET L2 sounds

As shown in the previous chapter, a NEW L2 perception scenario presents a challenging learning task because it involves the formation of new perceptual mappings and sound categories that can later be used for word comprehension. Recall that in that scenario the learner is confronted with a larger number of sound categories than the ones that are found in her L1. According to the L2LP model, the opposite scenario, i.e., being faced with a language whose phonemic categories constitute a *subset* of the L1 ones, also poses an L2 learning problem. The reason is that this situation leads to an L2 initial state in which the learner perceives more categories than the optimal L2 listener. Figure 6.1 shows how the number of L1 and L2 categories compares in the NEW and in the SUBSET L2 sound scenarios.

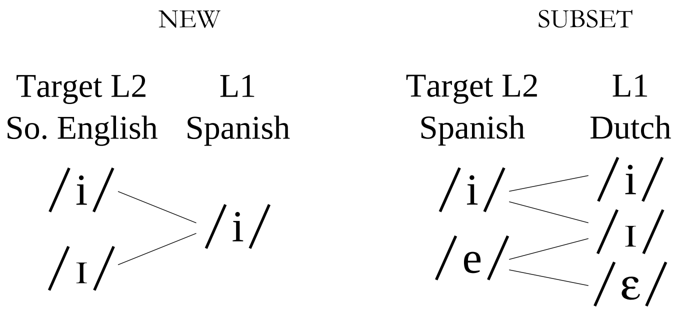

Fig. 6.1. L1 and L2 phoneme categories in the NEW versus the SUBSET L2 scenario.

Interestingly, this second scenario has not been addressed within the other L2 sound perception models reviewed in § 4, though Schmidt (1996) and Strange et al. (2001) have mentioned its possibility. The reason why this scenario has been neglected in the literature may be that at first sight it does not seem to pose a *lexical* learning problem for the L2 learner. Hypothetically, learners could apply the advantageous strategy of using part of their existing L1 categories in their L2. For instance, if Southern British English (SBE) listeners want to learn the Spanish vowels /i/ and /e/, they may use their own /i/ and /ɛ/ and avoid /ɪ/ to cope with their L2 environment. The same thing would apply to Dutch listeners learning Spanish.

In this chapter, it is shown that using a subset of the L1 vowels is not an automatic cross-language or L2 perception strategy but that it is a consequence of accomplishing the *representational* learning task in this novel L2 scenario. This is because entering an L2 environment with fewer categories will result in both the perception of 'too many' sound categories and the representation and access of 'too many' words in the lexicon. Thus, the learner will face two tasks, one lexical or representational and the other perceptual. In § 6.1, I describe the nature of the learning problem in a SUBSET scenario, and I present a summary of the predicted developmental path for this scenario. The application of the L2LP model to this novel scenario is demonstrated in § 6.2, and the L2LP predictions for this scenario are tested in § 6.3. Finally, § 6.4 presents an evaluation of the model given the L2 perception data.

6.1 Is there a learning task in a SUBSET L2 perception scenario?

The first study to attempt an explanation of the SUBSET scenario was Escudero & Boersma (2002). They reported on Dutch learners of Spanish /i/ and /e/ who have three vowels, viz., /i/, /ɪ/, and /ɛ/, that can be used to perceive the two Spanish ones. It was argued that listeners with an L1 that has more sounds than the target language face a number of L2 problems that are specific to this scenario, which they called *multiple-category assimilation*. That is, Dutch learners of Spanish first need to learn that the Dutch categories /i/ and /ɛ/ exist in Spanish as well, and that the acoustically intermediate Dutch category /ɪ/ does *not* exist in Spanish. Otherwise, they would assume the existence of phonological contrasts that do not exist in the target L2, i.e., /i/-/ɪ/ or /ɪ/-/e/. This situation is a typical example of the *subset problem* in language acquisition, which in this case refers to how the non-existence of an L1 property in the L2 environment can be discovered on the basis of positive evidence alone. Even if the learner were to resolve the lexical subset problem, she would be faced with a *perceptual* subset problem because she would perceive many /i/ or /e/ tokens as /ɪ/. Following the nature of the relation between L1 and L2 sounds and that of the resulting learning tasks, this novel L2 scenario will be referred to here as the SUBSET L2 perception scenario.

Crucially, the L2LP Full Copying hypothesis, together with the distinction between linguistic perceptual mappings and phonemic categories, allows for a more explicit account of the two learning tasks in this scenario. For instance, in the case of Dutch learners of Spanish, it is proposed that the copy of the L1 perception grammar will map Spanish tokens of /i/ and /e/ onto three Dutch categories, as shown in Figure 6.2.

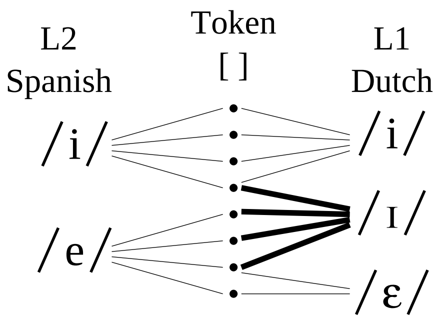

Fig. 6.2. The multiple-category-mapping in the SUBSET scenario.

With respect to L1 and L2 phonemes, Figure 3 shows the predicted phonemic equation for the SUBSET scenario. It is proposed that the perceptual mapping in Figure 2 will lead to the initial use of three L1 categories to store words that contain the two L2 phonemes shown in Figure 6.3.

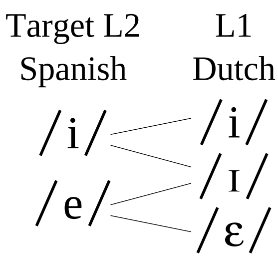

Fig. 6.3. Multiple phoneme equation in the SUBSET L2 perception scenario.

With respect to perceptual development, the L2LP model proposes that L2 learners have access to the GLA. This means that they will optimize their perception guided by the same learning mechanisms that are attested in L1 perceptual learning (cf. § 3.2.3). Also, the model's proposal of two separate perception systems for L1 and L2 predicts the maximum level of L2 perceptual success. That is, the L1 and L2 perceptions of advanced learners will match those of optimal perceivers in the two languages because L2 development can be achieved without affecting the already optimal L1 perception, which will remain stable. The model explains possible intermediate L1-L2 perception through its language activation mode hypothesis, which predicts that L2 learners can simultaneously activate their two perception grammars to different degrees depending on the language setting with which they are faced. In other words, learners will have a different perceptual behaviour depending on the language they think they hear, and therefore a bilingual language setting will lead them to use both their L1 and L2 during online sound perception. The L2LP model's ingredients for L2 sound perception are summarized in Table 6.1 together with the proposed predictions, explanations, and descriptions.

Table 6.1. The L2LP theoretical proposal.

L2LP	Prediction	Explanation	Description
Optimal L1 & L2	Human beings are optimal listeners	Optimal listeners handle the environment maximally well	L1 and L2 optimal category boundaries: Location & shape
Initial state	= Cross-language perception	Full Copying	L1 boundary location and shape
Learning task	= Reach the optimal target L2 perception	L2 learners want to reach target	Bridging mismatches between L1 and target optimal perception
Development	= L1-like	Full GLA Access	Category formation and boundary shifts
End state	Optimal L2 perception and optimal L2 perception	Input overrules plasticity Separate grammars	Language activation modes, through language setting variables

According to the L2LP's Full Copying hypothesis, the NEW and SUBSET scenarios will have different initial states in that the former will start with 'too few' categories and the latter with 'too many' categories. It is predicted that these differential initial states will lead to differential learning tasks and degrees of learning difficulty, as shown in Table 6.2.

Table 6.2. the initial state and learning tasks in a SUBSET scenario compared to a NEW scenario.

L2LP proposal	**Prediction for SUBSET**	**Prediction for NEW**
Initial state	Too many categories	Too few categories
Perceptual task	Adjust category boundaries	1. *Create* perceptual mappings 2. *Integrate* auditory cues
Representational task	1. *Reduce* lexical categories 2. *Reduce* perceptual categories	1. *Create* phonetic categories 2. *Create* segments
Degree of difficulty	Mildly difficult	Very difficult

Here we see that although both scenarios involve lexical and perceptual tasks, the nature of the two tasks in each scenario is different. That is, the representational task in a SUBSET scenario involves reducing L1 categories whereas in a NEW scenario it involves creating new L2 categories. With respect to the perceptual task, the former involves adjusting existing perceptual mappings whereas the latter involves creating new mappings and integrating new auditory dimensions to the ones that are already categorized. Given this differential nature of the learning tasks, the L2LP model predicts that attaining optimal target L2 perception and recognition will be easier in the SUBSET scenario than in the NEW. This is because perceptual development in the former will only involve category boundary shifting but not auditory-driven category formation whereas the latter scenario will require these two different learning mechanisms. Consequently, the developmental paths for these two scenarios will also be different, as shown in Table 6.3. Thus, unlike the NEW scenario in which perceptual learning will precede lexical learning, the

SUBSET scenario will show lexical development before the required perceptual development occurs.

Table 6.3. NEW versus SUBSET predicted developmental paths.

L2 scenarios:		NEW	SUBSET
S 1	Full Copying	Copied L1 categories: Too few	Copied L1 categories: Too many
S 2	Full GLA Access: Auditory-guided learning	Creating phonetic categories and rankings	------
S 3		Creating phonological categories and rankings	------
S 4		Using discrete categories for lexicalization	Reducing lexical categories
S 5	Full GLA Access: Lexicon-guided learning	Constraint reranking: One-dimensional boundary shifting	Constraint reranking: Boundary approximation (one-dimensional or multi-dimensional)
S 6		Low level integration of acoustic dimensions	
S 7		Using integrated categories in the lexicon	
S 8		Multidimensional boundary shifting	

In the next section, I will discuss the specific predictions of the L2LP model for this novel scenario in concordance with its five ingredients for explaining L2 sound perception.

6.2 Ingredients of L2 linguistic perception in a SUBSET scenario

The L2LP model aims at providing a comprehensive description, explanation, and prediction of L2 perceptual learning through the consideration of five ingredients whose purpose is to cover all the components of the L2 process. The first of these ingredients is the description of the adult optimal perception in the L1 and target L2. This ingredient provides a window to the initial state and the learning task of L2 learners.

In § 6.2.1, the optimal perception hypothesis proposed within the general LP model is applied to the Dutch and Spanish vowels that are likely to represent a case of the SUBSET scenario. The model's predictions for the initial L2 state are examined in § 6.2.2, the predicted solutions for this scenario's learning tasks are presented in § 6.2.3, and the predicted L2 developmental path is treated in § 6.2.4. Finally, the L2LP hypothesis of separate grammars for L1 and L2 perception as applied to the SUBSET scenario is discussed in § 6.2.5.

6.2.1 Ingredient 1: predicting optimal perception from environmental production

The L2LP model proposes that the first step toward explaining L2 perception is to establish the optimal perception in the languages at hand. Recall that the general LP model advances the hypothesis of 'optimal perception' which states that adult listeners are optimal perceivers of their specific production environment. This hypothesis relies on a particular definition of optimal speech perception which, for sound perception, entails that an optimal listener will classify auditory events into the sound categories that best match the speaker's intention. Thus, it is proposed that the production distributions of the relevant speech sounds in the languages involved will determine the optimal perception.

Consider the F1 values of the Dutch and Spanish vowels that constitute a case of the SUBSET perception scenario. Figure 6.4 shows the F1 distributions for Dutch /i/, /ɪ/, and /ɛ/ and Spanish /i/ and /e/, as shown in Boersma & Escudero (2004).[54] The tokens of these vowels were normally distributed along a base-10 logarithmic scale and the vowel centres were given by the median F1 values meas-

[54] Dutch listeners are unlikely to use /e/ to classify any Spanish vowel because they do not match in duration. Also, this Dutch vowel tends to be diphthongized, which may add to its perceived dissimilarity from Spanish /i/ or /e/.

ured in a production experiment. In the figure, we see the median F1 values of 305, 438, and 733 Hz for the Dutch vowels, and 332 and 500 Hz for the Spanish vowels. The vowels in both environments were produced with a standard deviation of approximately 0.166 octaves.

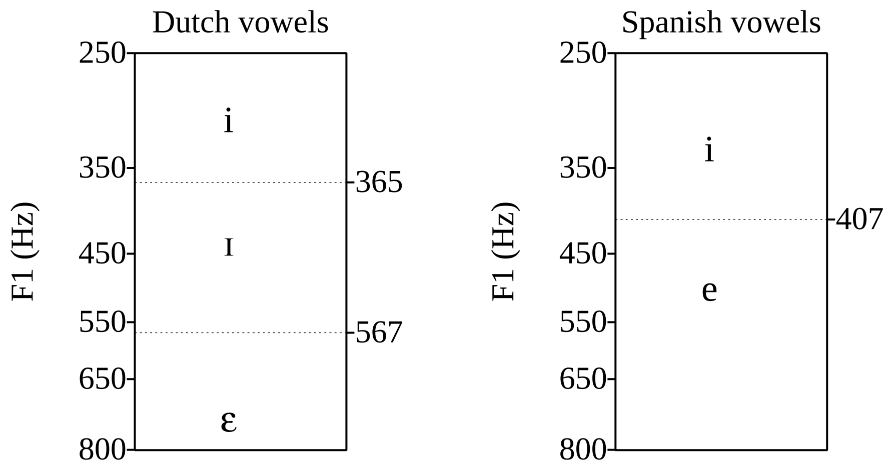

Fig. 6.4. Token distributions along the F1 dimension for the Dutch and Spanish vowels.

The dotted lines represent the equal-likelihood points in production which, as described in Chapters 2, 3, and 5, represent the acoustic values that have a 50-50% probability of having being intended as either vowel.[55] For instance, vowels produced with an F1 of 365 Hz are intended as either /i/ or /ɪ/ in the Dutch environment. We can see that, along the same F1 continuum, the Dutch environment produces three vowels as opposed to the two of the Spanish environment. Therefore, two equal-likelihood points are found in Dutch and only one in Spanish.

Following the optimal perception hypothesis (cf. § 2.1.2), it is predicted that the optimal perception of the Dutch and Spanish environments will resemble the production distributions in Figure 6.4 so that the optimal category boundaries will coincide with the equal-likelihood points. Thus, the Dutch optimal perceptual boundary between /i/ and /ɪ/ lies at 365 Hz, and the one between /ɪ/ and /ɛ/

[55] In this one-dimensional case, the equal likelihood in production is a point rather than a line. In Figures 4 and 5, they appear as lines solely to show that they constitute a category boundary, but the value of such lines is really a point, e.g., 407 Hz, along an acoustic dimension, e.g., F1.

lies at 567 Hz, while the optimal perceptual boundary between Spanish /i/ and /e/ lies at 407 Hz. Figure 6.5 shows the optimal category boundaries for the two languages when tokens are categorized along the same F1 continuum which ranges from 250 to 800 Hz.

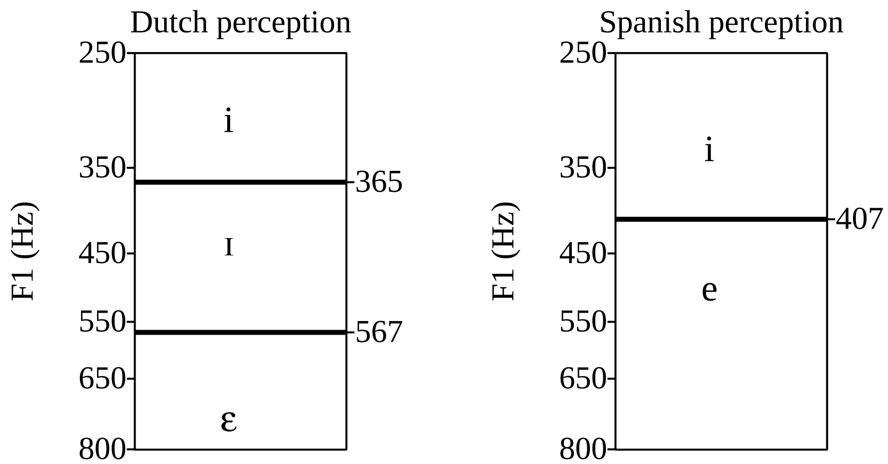

Fig. 6.5. Category boundaries for an optimal Dutch and an optimal Spanish listener, which coincide with the equal-likelihood points in production.

What this means is that an optimal Dutch listener, who has optimal perceptual boundaries, will categorize F1 values lower than 365 Hz as /i/, values between 365 and 567 Hz as /ɪ/, and values higher than 567 Hz as /ɛ/. As for the Spanish optimal listener, her perceptual boundaries will lead her to perceive F1 values lower than 407 Hz as /i/ and F1 values higher than 407 as /e/.[56] Furthermore, an optimal Dutch listener will correctly categorize 83.7% of all possible tokens drawn from the Dutch distributions and the remaining 16.3 % will constitute cases of perception errors caused by the overlap in the distributions. On the other hand, the Spanish optimal listener will, on the other hand, have a correctness percentage of 95.5, as computed from the Spanish distributions in Figure 6.4.

---

56 Boersma & Escudero (2004) proposed that an optimal Dutch listener perceives values up to 824 Hz as /ɛ/ and any higher values as /ɑ/. For the Spanish optimal listener, F1 values higher than 662 Hz are perceived as /a/. Given that this chapter concentrates on the perception of front vowels, the perception of Dutch /ɑ/ and Spanish /a/ will not be further considered.

The LP model proposes that an OT perception grammar represents the linguistic knowledge that underlies sound categorization. This means that the optimal Dutch and Spanish perceptual boundaries shown in Figure 6.5 result from the constraint rankings in the optimal perception grammars shown in Figures 6.6 and 6.7 below. Thus, Figure 6.6 shows the proposed continuous ranking of the *cue constraints* against perceiving F1 values as /i/, /ɪ/, and /ɛ/ in the optimal Dutch perception grammar. Note that the crossing points of the continuous constraint curves represent the optimal category boundaries. For instance, it can be read off the figure that a token with an F1 value of 380 Hz, which is represented by a vertical dotted line, will be mostly perceived as Dutch /ɪ/. This is due to the fact that the constraint against perceiving an F1 value of 380 Hz as /ɪ/, which is the point in the continuous dashed curve located at 380 Hz, is ranked lower than the cue constraints against perceiving the same F1 value as /i/ or /ɛ/.

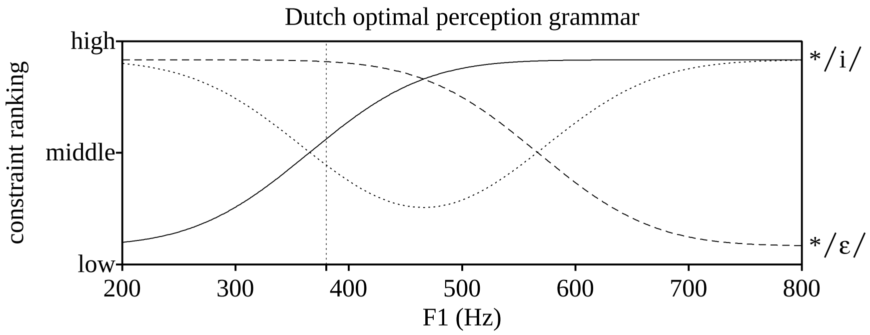

Fig. 6.6. The continuous constraint rankings in the optimal Dutch perception grammar.

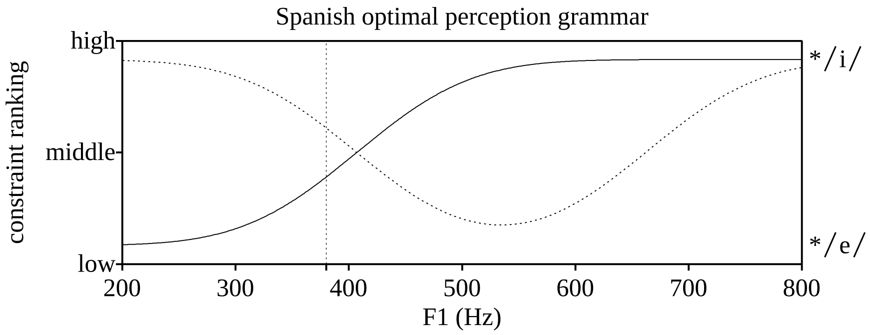

Fig. 6.7. The continuous constraint rankings in the optimal Spanish perception grammar.

Figure 6.7 shows the continuous constraint rankings that result in the optimal categorization of F1 values as the Spanish vowels /i/ and /e/. It can be read off the figure that, for instance, a token with an F1 value of 380 Hz will be mostly perceived as Spanish /i/ because the constraint against perceiving this value as Spanish /i/ is ranked lower than the one against perceiving it as /e/ in the Spanish optimal perception grammar.

In a computer simulation, Boersma & Escudero (2004) tested whether simulated GLA listeners could actually achieve the predicted optimal perception. To that end, a virtual L1 Dutch learner was immersed in the Dutch environment shown in Figure 6.4. The virtual learner was exposed to 10,000 F1 values per virtual month and learned through the GLA lexicon-driven mechanism.[57] It was shown that after 18 virtual years, she achieved the predicted optimal perception as well as the predicted optimal constraint ranking. She thus reached a correct Dutch categorization of 83.7%, and the crossing points in her perception grammar were located at the optimal boundaries of 365 and 567 Hz.

In sum, the optimal perception hypothesis predicts that Dutch and Spanish listeners will perceive F1 values in accordance with the production of vowels in their respective environments. This means that if real Dutch and Spanish listeners

[57] The virtual Dutch learner developed through the GLA lexicon-driven learning described in §2.1.3 and § 5.2.1. In the simulations, this learner was assumed to have correct lexical categories which enabled her to compare the result of her perception with that of her recognition. Thus, every time there was a mismatch between the perceived vowel and the one intended by the speaker, her perception grammar changed to a small degree.

are optimal perceivers, they will exhibit perceptual behaviour that reflects their language-specific production environment. In § 6.3.1, perceptual data will be presented as evidence for the optimal perception hypothesis in these languages. With respect to the implications of optimal perception for the acquisition of L2 vowels, the L2LP model predicts that Dutch listeners will use their L1-optimal perception grammar when faced with foreign sounds. The next section will show how the optimal Dutch perception handles Spanish vowel tokens, and what the predicted consequences of this cross-language perception are for Dutch learners of Spanish front vowels.

6.2.2 Ingredient 2: predicting cross-language and initial L2 perception

According to the L2LP Full Copying hypothesis (cf. § 3.3), beginning Dutch learners of Spanish will perceive Spanish vowels in the same way as Dutch listeners who are not learning this language. Thus, it is predicted that the Dutch cross-language perception of Spanish will provide a measure of the starting point for a Dutch learner of Spanish. Figure 6.8 shows how the optimal Dutch listener will categorize the Spanish vowel distributions shown in Figure 4.

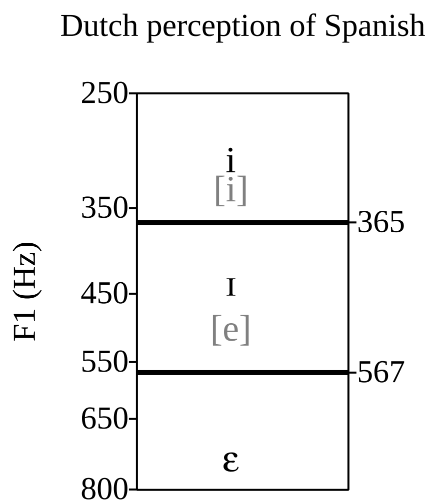

Fig. 6.8. How the Dutch optimal boundaries cope with Spanish vowel productions.

We can see that the Spanish vowel distributions fall in the perceptual space for three Dutch vowels. That is, some Spanish /i/ tokens fall above the optimal boundary between Dutch /i/ and /ɪ/ but others fall below it. A similar situation can be observed for Spanish /e/ tokens because some of them fall below the optimal boundary between Dutch /ɪ/ and /ɛ/ while others fall above it.

To compute which percentage of Spanish /i/ tokens will be perceived as /i/ and which as /ɪ/, we can compare the location of the Dutch perceptual boundaries to the mean productions and standard deviations of Spanish /i/ and /e/. If we assume that the Spanish vowel distributions have gaussian shapes, 50% of the Spanish /i/ tokens are produced below the mean and 50% above it. To measure the Dutch *perception percentages*, we first compute the logarithmic distance between the Dutch boundary location (i.e., 365 Hz) and the mean value for Spanish /i/ (i.e., 332 Hz) which is log2(365/332) = 0.137 octaves. The standard deviation for the Spanish tokens is 0.166 octaves, and therefore we can express the measured distance in standard deviations, which is 0.137/0.166 = 0.824 above the mean. Finally, we can compute the probability that a Spanish token will be more than 0.824 standard deviations above the Spanish mean as gaussQ(0.824) = 0.205, or 20.5% of all tokens. This will give us the percentage of tokens that are produced above 365 Hz. The percentage of Spanish /e/ tokens produced below the Dutch boundary between /ɪ/ and /ɛ/ which is located at 567 Hz can be computed with the same procedure, in this case yielding 13.7%.

Thus, 20.5 % of the Spanish /i/ tokens are produced above 365 Hz, while 13.7% of the Spanish /e/ tokens are produced below 567 Hz. With this percentage of Spanish token frequencies, we can predict the relative percentage of Spanish /i/ and /e/ tokens that will be perceived as each of two Dutch vowels by an optimal Dutch listener. Recall that a maximum-likelihood listener perceives every value that falls above her boundary as one category and everything that falls below as the other category. This means that an optimal Dutch listener will perceive 79.5% of the Spanish /i/ tokens as Dutch /i/ and 20.5% as Dutch /ɪ/. Likewise, she will perceive 83.7% of the Spanish /e/ tokens as /ɪ/ and 13.7% as /ɛ/. These cross-language perception percentages will also be found in beginning learners of Spanish, whose perception mappings and percentages are shown in Figure 6.9.

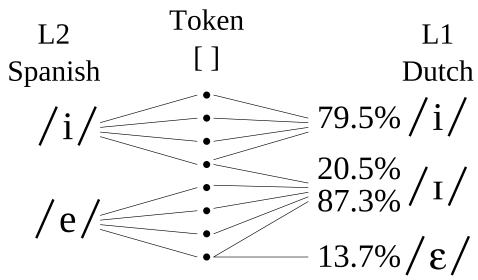

Fig.6.9. Predicted perception mappings and percentages for beginning Dutch learners of Spanish.

With respect to lexical categories, it is predicted that the patterns of lexicalization will match those of perception. This is because learners will use the categories that result from perceiving L2 tokens to lexicalize L2 words, as was proposed for L1 learning in Chapter 2. This means that, initially, the two Spanish vowels will be categorized and lexicalized as the three Dutch vowels shown in (6.1) below. Following the proposed distinction between the output of perception and that of recognition (cf. Chapter 2), perceived categories are represented between slashes and lexical categories between pipes. However, it is important to mention that the two types of categories are considered to be commensurable in that they convey the same level of abstraction from the acoustic signal (cf. §2.1.1). All the arrows represent perceptual mappings and the connecting lines represent *phonemic equation*. The double arrows, on the other hand, depict the largest percentage of perceptual mapping so that, for instance, /i/ is mapped onto /i/ more frequently than to /ɪ/.

(6.1) Dutch learners' initial categorization (left) and lexicalization (right) of Spanish front vowels. Left: Spanish vowels, right: Spanish vowel for a Dutch learner.

Spanish		Dutch	Spanish		Dutch
/i/	⇒	/i/	|i|	—	|i|
/i/	→	/ɪ/	|i|	—	|ɪ|
/e/	⇒	/ɪ/	|e|	—	|ɪ|
/e/	→	/ɛ/	|e|	—	|ɛ|

The predicted phonemic equation shown in (6.1) differs from the one proposed in Boersma & Escudero (2004) where it was assumed that, at the lexical level, Spanish |i| is exclusively identified with Dutch |i| while Spanish |e| is exclusively identified with Dutch |ɛ|. As they noted, their proposal was not one that referred to a learner's knowledge but rather to a linguist's observation of the perceptual behaviour found in Dutch learners of Spanish (p. 3). However, such pattern can only be proposed for the advanced learners they examined given that their beginning learners showed no signs of using only two lexical categories. If lexical access and later storage are driven by perceptual mappings (as proposed in § 2.1.3), the initial lexical knowledge that Dutch learners have of words containing Spanish |i| and |e| will exhibit two phonological forms. For instance, a beginning Dutch learner will have |tʃika| and |tʃɪka| as her initial lexical representations of the Spanish word *chica* 'girl' and |tʃɪka| and |tʃɛka| as those for *checa* 'Czech (fem.)', as shown in Figure 6.10.

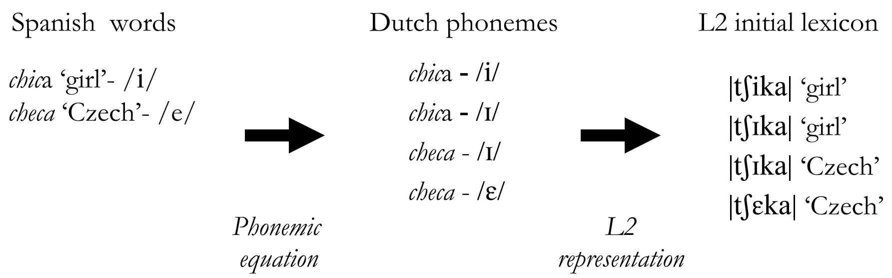

Fig. 6.10. Dutch learners initial lexical knowledge for Spanish words containing |i| and |e|.

It is argued here that the two-way lexicalization of the Spanish words results from their two-way perception, as shown by the perception percentages in Figure 9. That is, the reuse of the Dutch optimal perception grammar will lead to both the double perception and storage of L2 words containing each of the Spanish vowels. In addition, it is claimed that the perception percentages will also be used to store words in the developing L2 lexicon, and this will be shown in § 6.2.4.

In sum, it is predicted that the L2 sounds in a SUBSET scenario will yield the perception and lexicalization of 'too many' categories. As a result, the learner will be faced with a SUBSET problem for both perception and lexical access. The question now is how she can resolve these two different but related problems, that is to say, how she can realize that Spanish has only two vowels instead of the three that she perceives and recognizes at the onset of her L2 learning process. The next section provides a further consideration of the lexical and perceptual tasks in a SUBSET L2 perception scenario, and it presents the predicted solutions that L2 learners will entertain.

6.2.3 Ingredient 3: predicting the L2 learning task

To further investigate the nature of the learning tasks that result from a two-to-three perceptual mapping and lexical storage, we need to consider the linguistic processes involved in word comprehension. This procedure is defined as the linguistic mapping from the auditory properties of the speech signal onto lexical representations that are linked to word meanings. Recall that the LP model as well as most psycholinguistic models (cf. Chapter 1) propose that word comprehension involves two sequential mappings which are usually called speech perception and speech recognition. Figure 6.11 shows how a Dutch learner of Spanish accesses the two forms with which she has stored the L2 word *chica*.

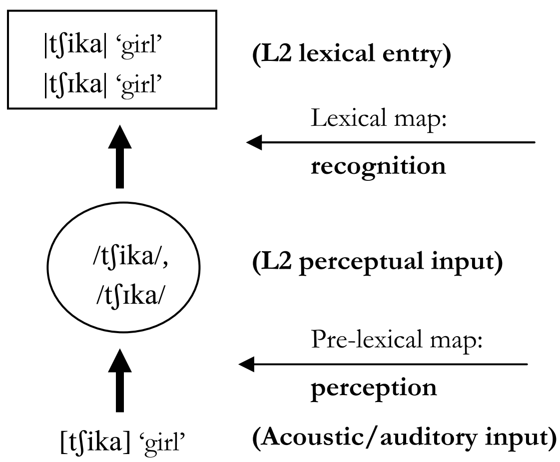

Fig. 6.11. The two processes involved in the comprehension of Spanish words by a beginning Dutch learner.

As can be seen, if a Spanish speaker produces the first vowel in the word [tʃika] ('girl'), a beginning Dutch learner will perceive /tʃika/ 79.5% of the time and store |tʃika| 'girl' in her L2 lexicon, but she will also perceive /tʃɪka/ 20.5% of the time and store |tʃɪka| 'girl' for the same Spanish word. A similar situation will occur for Spanish [tʃeka] ('Czech, female'), which will lead to the perception and recognition of two phonological forms. This is because following the L1 model described in § 2.3.2, we can assume that the L2 perceived forms will be copied into the lexicon, just as when L1 learners use their perceived categories to store new words in their L1 lexicon (cf. § 2.1.3). Thus, Dutch learners of Spanish will store in their L2 lexicon two identical forms with two different meanings as well as two different forms with identical meanings for every minimal pair containing Spanish /i/ and /e/. As a consequence, these learners will be faced with a representational or lexical task because their initial L2 perception and recognition will output more categories than the optimal ones. It is proposed that this task will be solved by somehow 'disfavouring' the extra lexical items.

However, even if learners are able to solve the representational task in an L2 SUBSET scenario, they will still be faced with a perceptual task because in order to become an optimal L2 listener, the learner would have to stop perceiving those categories which are not produced in the L2 environment. It is therefore proposed that the perceptual task will be performed by shifting boundaries to reduce extra perceptual categories. This category reduction will be manifested as the gradual approximation of the perceptual boundary between Dutch /i/ and /ɪ/ as well as the one between /ɪ/ and /ɛ/, as shown by the arrows in Figure 6.12.

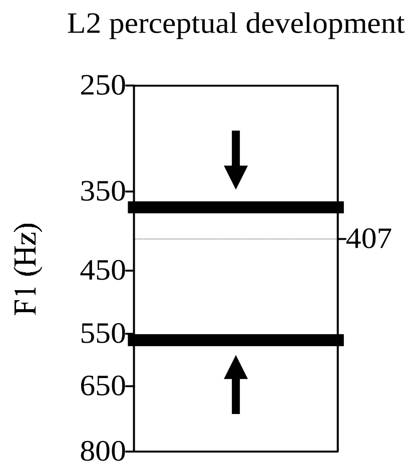

Fig. 6.12. Proposed solution for the perceptual task in a SUBSET scenario: Category boundary approximation that results in the reduction of the extra category.

In a SUBSET scenario, perceptual development depends on lexical development because the perceptual boundary shift that leads to optimal L2 perception can only be performed if the lexicon contains the optimal number of L2 lexical categories. This means that if performing the representational task in this scenario leads to a reduction in the recognition of lexical items such as |tʃɪka| 'girl' and |tʃɪka| 'Czech, female', the perception of the form that matches these items, viz., /tʃɪka/, will also be reduced. Specifically, it is predicted that Dutch learners of Spanish will reduce the number of perceived categories for Spanish /i/ and /e/ as a result of reducing the lexical entries for words containing the Spanish front vowels. The next section shows a formalization of the development that occurs in the initial L2 recognition and perception grammars.

6.2.4 Ingredient 4: predicting L2 development

In this section, I present a novel proposal for the formal modelling of the learning mechanisms that underlie the performance of the L2 representational and perceptual tasks introduced in the previous section.[58] Figure 6.13 illustrates the L2LP proposal for the developmental path in a SUBSET scenario. The squares in the figure show the initial state and learning mechanisms that result from Full Copying of L1 perception and the Full Access to the GLA (cf. § 3.2 and § 3.4). The first circle represents the predicted learning tasks that result from the reuse of L1 categories and perceptual mappings. The second circle represents the predicted L2 end state that results from applying GLA learning mechanisms to the lexical and perceptual tasks. Finally, the bidirectional arrow depicts the interactive application of the two GLA learning mechanisms because, in this scenario, recognition and perceptual learning trigger one another, as will be explained below.

---

[58] This L2 proposal is novel for two reasons. First, it is not found in previous accounts of the SUBSET or 'multiple category assimilation' scenario, viz. Escudero & Boersma (2002) and Boersma & Escudero (2004), because they assumed that learners somehow use the advantageous strategy of identifying the L2 vowels with two of their three perceived categories. Second, it is the first time that Boersma's (2001) proposal for the learning of recognition grammars is applied to L2 acquisition.

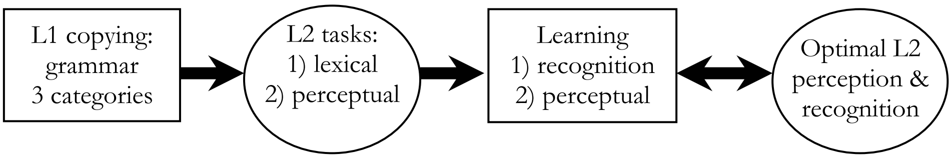

Fig. 6.13. Predicted developmental path for a SUBSET L2 sound perception scenario.

What we observe here is that the learner initially reuses three L1 categories and the perceptual mappings that output more categories than the ones produced in the target L2 environment. As a result, she will store 'too many' items in the L2 lexicon. These perception and recognition problems lead to the two learning tasks mentioned in the previous section, viz., perceptual and representational. In this section, I first explain the initial state in a subset L2 scenario. I then go on to describe how GLA recognition learning functions and how it generates perceptual learning. As we will see, these two types of learning interact so that recognition learning first leads to perceptual learning with perception then feeding recognition.

As a result, both optimal L2 perception and recognition can be attained.

The proposal is illustrated by Dutch learners of Spanish /i/ and /e/ who initially perceive and recognize three vowels instead of the optimal two. That is, these learners start with an L2 lexicon that contains four lexical items instead of the two optimal ones for every minimal pair containing these Spanish vowels, as shown in (6.2). [59]

(6.2) Phonological form-meaning pairs in Dutch learners' L2 lexicon

identical meaning { |tʃika| 'girl'
|tʃɪka| 'girl' }

identical meaning { |tʃɪka| 'Czech, female'
|tʃɛka| 'Czech, female'

} *identical form* (|tʃɪka| 'girl' and |tʃɪka| 'Czech, female')

[59] Crucially, although the proposal is illustrated by *chica* ~ *checa* in (6.2), it applies to any minimal pair in the target L2, e.g., *pisa* 'steps on' ~ *pesa* 'weight', *Lima* 'capital of Peru' ~ *lema* 'motto', *visa* 'visa' ~ *besa* 'kisses', etc.

As we can see, these four initial lexical items include two *identical* phonological forms with *different* meanings and two different forms with identical meanings. Importantly, this proposal of storing more lexical forms than the ones intended in the target L2 also applies to any word containing Spanish /i/ and /e/ that is not a member of a minimal pair, e.g., *mes* 'month', *cima* 'peak', *dedo* 'finger', *rico* 'tasty', etc, because these words will lead to the storage of two lexical items, i.e. two phonological forms with identical meanings, instead of the optimal one.

Thus, every time a word is perceived, it is copied as such in the lexicon and paired with the new meaning in the L2 environment. Following Boersma's (2001), it is proposed that the L2 recognition grammar contains *LEX and FAITH constraints, which limit lexical access and ensure that the perceived category is preserved in the chosen lexical representation (cf. § 2.4). Thus, constraints against the lexicalization of form-meaning pairs such as the ones shown in (6.3) are created in the initial grammar for every perceived word.

(6.3) *LEX constraints in the L2 recognition grammar of Dutch learners of Spanish
*LEX |tʃika| 'girl' or its shorter version *LEX |i| 'girl'
*LEX |tʃɪka| 'girl' or *LEX |ɪ| 'girl'
*LEX |tʃɪka| 'Czech, female' or *LEX |ɪ| 'Czech'
*LEX |tʃɛka| 'Czech, female' or *LEX |ɛ| 'Czech'

These *LEX constraints are ranked by *perceived frequency*, i.e., the perceived percentages of the vowels shown in Figure 6.9. This ranking is achieved after some of the frequency learning described in § 2.4 has occurred. Thus, a beginning Dutch learner of Spanish will have *LEX |tʃɪka| 'girl' ranked higher than *LEX |tʃika| 'girl' because tokens of the Spanish word *chica* are more frequently perceived as /tʃika/ than /tʃɪka/ in a proportion of 79.5% to 20.5%, as was computed in § 6.2.2. Given that the ranking of *LEX constraints matches the frequency of perceptual inputs, lexical representations that match common perceptual inputs will be most likely to be accessed. The production distributions of Spanish *checa* ('Czech') yield an initial L2 perception of 86.3% for /tʃɪka/ and 13.7% for /tʃɛka/, and this entails that *LEX |tʃɪka| 'Czech, female' will be ranked lower than *LEX |tʃɛka| 'Czech, female'. If we take into account the perceived frequency percentages of all four forms, i.e., 13.7%, 20.5%, 79.5% and 86.3% respectively, the constraints in the L2 recognition grammar of Dutch learners will have the ranking shown in (6.4).[60]

(6.4) L2 *LEX constraints ranked by L2 perceived frequency
*LEX |ɛ| 'Czech' >> *LEX |ɪ| 'girl' >> *LEX |i| 'girl' >> *LEX |ɪ| 'Czech'

Tableau 6.1 below shows the ranking in (6.4) together with the relevant lexical candidates, which are the ones listed in (6.2). With these *LEX constraints alone, the recognized candidate is |tʃɪka| 'Czech' no matter what phonological form for the Spanish words *chica* or *checa* the learner perceives because it has the lowest ranked *LEX constraint. Consequently, two types of errors will arise when recognizing these Spanish words, viz., *semantic errors* and *phonological errors*. The former occur when, given the context in which a word was produced, the recognized candidate has a different meaning than that of the perceptual input. Thus, the perceptual input /tʃɪka/ (meaning: 'girl') will incur a semantic error because, given the ranking of the lexical constraints, it will be recognized as |tʃɪka| 'Czech'. Phonological errors occur when the form of the perceptual input is not preserved in the recognized category. For instance, with the constraints and constraint ranking in Tableau 1, the input /tʃɛka/ (meaning 'Czech') will incur a phonological error because it will be recognized as |tʃɪka| 'Czech'. In addition, a perceptual input such as /tʃɪka/ ('girl') will incur both a semantic and a phonological error because the recognized lexical item, which is always |tʃɪka| 'Czech', has a different meaning and a different phonological form.

[60] *LEX constraints can also be ranked with respect to the frequency of the words in a minimal pair. For instance, the Spanish word *chica* is more frequent than the word *checa*, which means that the L2 *LEX constraints that refer to *checa* will be higher ranked than the ones for *chica*. Thus, word frequency should interact with perceived frequency in the ranking of the *LEX constraints. According to Boersma (2001), the ranking of *LEX constraints through token frequency is an automatic result of the GLA which indirectly and automatically takes into account both perceptual input frequencies and word frequency. This implies that a more realistic ranking for the constraints referring to the Spanish minimal pair *chica* ~ *checa* is *LEX |ɛ| 'Czech' >> *LEX |ɪ| 'Czech' >> *LEX |ɪ| 'girl' >> *LEX |i| 'girl'. Given that minimal pairs may not have differential word frequencies, the ranking in (6.4) can be assumed.

Perceptual input *chica* or *checa*	*LEX |tʃɛka| 'Czech'	*LEX |tʃɪka| 'girl'	*LEX |tʃika| 'girl'	*LEX |tʃɪka| 'Czech'
|tʃika| 'girl'			*!	
|tʃɪka| 'girl'		*!		
☞ | tʃɪka| 'Czech'				*
|tʃɛka| 'Czech'	*!			

Tableau 6.1. L2 ranking of *LEX constraints and L2 lexical candidates for a minimal pair of the /i/-/e/ Spanish contrast.

In order to at least remedy the phonological errors, FAITH constraints need to interact with the *LEX constraints. Thus, constraints such as 'do not recognize /i/ as |ɛ|' (or '*/i/ → |ɛ|') are also present in the recognition grammar and are introduced with a high ranking. Tableau 6.2 shows the FAITH constraints relevant to the four lexical candidates for *chica* and *checa* in the L2 lexicon of Dutch learners of Spanish. The ranking of these constraints may be driven by the perceptual similarities between vowels, so that */i/ → |ɛ| is ranked higher than */i/ → |ɪ| because /i/ is more similar to /ɪ/ than to /ɛ/. In this example, we observe that these FAITH constraints will ensure that the phonological form of the perceptual input is preserved in the recognized lexical category so that a perceptual input containing the vowel /i/ will be recognized as a lexical item containing |i|.

/tʃika/ Meaning: 'girl'	FAITH */i/→|ɛ|	FAITH */ɛ/→|i|	FAITH */i/→|ɪ|	FAITH */ɪ/→|i|	FAITH */ɛ/→|ɪ|	FAITH */ɪ/→|ɛ|
☞ |tʃika| 'girl'						
|tʃɪka| 'girl'			*!			
|tʃɪka| 'Czech'			*!			
|tʃɛka| 'Czech'	*!					

Tableau 6.2. FAITH constraints in the L2 recognition grammar.

Here we see that the FAITH constraints ensure that /tʃika/ is recognized as |tʃika| 'girl' because this is the only lexical candidate that preserves the perceived phonological form. Importantly, FAITH constraints need to be ranked high so that they can prevent, for instance, the recognition of a perceptual input containing the vowel /ɪ/ as a lexical item containing some other vowel, as shown in Tableau 6.3. Importantly, FAITH constraints apply to every L2 word with the corresponding perceived vowel. This means that, for instance, FAITH */ɪ/→|i| ensures the recognition of *any* perceptual input containing /ɪ/ as a lexical item containing |ɪ|. By contrast, *LEX constraints *only* apply to the L2 word to which they refer, e.g. *LEX |tʃɪka| 'girl' only applies to the perceptual input /tʃɪka/ 'girl'.

/tʃɪka/ Meaning: 'girl'	FAITH */ɪ/→|i|	FAITH */ɪ/→|ɛ|	*LEX |tʃɛka| 'Czech'	*LEX |tʃɪka| 'girl'	*LEX |tʃika| 'girl'	*LEX |tʃɪka| 'Czech'
|tʃika| 'girl'	*!				*	
|tʃɛka| 'Czech'		*!	*			
|tʃɪka| 'girl'				*!		
☞ |tʃɪka| 'Czech'						*

Tableau 6.3. L2 recognition with the interaction of phonological and lexical constraints.

In the tableau, the top left cell shows the perceptual input /tʃɪka/ with the speaker's intended meaning of 'girl'. This means that a Spanish speaker has produced the word *chica* with the F1 values that lead to an /ɪ/ perception by a Dutch learner, a situation that occurs 20.5 % of the time when the word *chica* is produced. The high ranking of FAITH preserves the phonological form of the perceptual input regardless of its meaning in the semantic context in which it was produced, and therefore the candidates containing a different vowel are not chosen. At that point, the grammar needs to choose from the third and fourth candidates. In the end, the winning candidate is |tʃɪka| 'Czech' because *perceived frequency* had it ranked lower that the constraint against |tʃɪka| 'girl'. This means that the output of recognition matches the phonological form of the perceptual input, viz. /tʃɪka/, but fails to match the intended L2 meaning, viz. 'girl', a situation that leads to a semantic error.

It is proposed that this semantic error will automatically result in *message-driven recognition learning* which applies when the learner notices that, given the speaker's intended message, she should have recognized the meaning 'girl' rather than 'Czech'.[61] Tableau 6.4 shows the automatic GLA recognition learning that is triggered by the learner's noticing of the semantic error, which is depicted by the "*☞*" symbol, and also by her noticing that two of the other lexical candidates have the correct meaning for the perceptual input, as depicted by the check mark (√). Crucially, one of the candidates with the correct meaning has a form that violates the FAITH constraint against changing perceived /ɪ/ for a recognized |ɪ|, which results in the lowering of this constraint in order to solve the semantic error.

/tʃɪka/ Meaning: 'girl'	FAITH */ɪ/→|i|	FAITH */ɪ/→|ɛ|	*LEX |tʃɛka| 'Czech'	*LEX |tʃɪka| 'girl'	*LEX |tʃika| 'girl'	*LEX |tʃɪka| 'Czech'
√ |tʃika| 'girl'	*!→				*→	
|tʃɛka| 'Czech'		*!	←*			
√ |tʃɪka| 'girl'				*!→		
☞ |tʃɪka| 'Czech'						←*

Tableau 6.4. GLA recognition learning in the L2 Spanish recognition grammar.

Here we see how the semantic error incurred by the winning candidate, i.e., 'Czech', and the phonological form of one of the semantically correct candidates, i.e., |tʃika|, result in GLA message-driven recognition learning. That is, the *LEX constraints of the three candidates involved will move so that *LEX |tʃɪka|

---

[61] This 'noticing' occurs perhaps as a result of the realization that the semantic or pragmatic context does not refer to the nationality of a female person but simply to a young woman. The mechanisms that allow such a noticing of semantic errors and correct forms are beyond the scope of this study. However, it can be said that they apply at a higher and more conscious level than that of the recognition grammar (or even the lexicon) and that they are drive by a functional principle of confusion minimization that applies to the message intended by the speaker.

'Czech' will be promoted because the candidate that refers to it incurred the semantic error, and *LEX |tʃika| 'girl' and *LEX |tʃɪka| 'girl' will be demoted because they both have the correct meaning. Also, the constraint 'do not recognize a perceived form /ɪ/ as a lexical form containing |i|' will be demoted because of the mismatch between the winning candidate |tʃika| 'girl' and the perceptual input /tʃɪka/. Crucially, this FAITH constraint will move much faster than any *LEX constraint, as illustrated by the larger size of the arrow that demotes FAITH in Tableau 6.4. This is because FAITH constraints apply to *every* perceptual input containing the same vowel while *LEX constraints apply to *only one* perceptual input. In other words, FAITH */ɪ/→|i| will move with every Spanish word containing /i/ or /e/ while the *LEX constraints will move with their specific minimal pair only.

However, the story does not end here because even when FAITH */ɪ/→|i| is demoted to the lowest position, the same semantic error caused by the low ranking of *LEX |tʃɪka| 'Czech' will still occur, as shown in Tableau 6.5.

/tʃɪka/ Meaning: 'girl'	FAITH */ɪ/→|ɛ|	*LEX |tʃɛka| 'Czech'	*LEX |tʃɪka| 'girl'	*LEX |tʃika| 'girl'	*LEX |tʃɪka| 'Czech'	FAITH */ɪ/→|i|
√ |tʃika| 'girl'				*!→		*
|tʃɛka| 'Czech'	*!	←*				
√ |tʃɪka| 'girl'			*!→			
☞ |tʃɪka| 'Czech'					←*	

Tableau 6.5. Reranking of *LEX constraints as a result of a semantic error.

After many instances of this semantic error, lexical constraints will be gradually reranked until they reach the ranking shown in (6.5), which leads to the recognition of /tʃɪka/ 'girl' as |tʃika| 'girl' because the *LEX constraint against its recognition is now ranked lowest. This recognized category would incur a phonological error if FAITH */ɪ/→|i| had its original high ranking. Crucially, FAITH */ɪ/→|i| was demoted to a very low position so that the recognition grammar now allows changing the phonological form /ɪ/ for |i|. Thus, no further error occurs when the recognition grammar faces the perceptual input /tʃɪka/ ('girl').

(6.5) L2 recognition grammar that optimally handles /tʃɪka/ 'girl'
FAITH */ɪ/→|ɛ| >> *LEX |tʃɛka| 'Czech' >> *LEX |tʃɪka| 'girl' >> *LEX |tʃɪka| 'Czech' >> *LEX |tʃika| 'girl' >> FAITH */ɪ/→|i|

However, a new semantic error will occur because now a perceptual input /tʃɪka/ 'Czech', which occurs in 86.3% of the time when the L2 word *checa* is produced, will lead to recognize |tʃika| 'girl', as shown in Tableau 6.6. This error not only occurs because the constraint FAITH */ɪ/→|i| is ranked low, so that perceptual inputs containing /ɪ/ can now be recognized as |i|, but also because *LEX |tʃɪka| 'Czech' is ranked higher than *LEX |tʃika| 'girl'. Thus, the learner will notice that her recognized candidate incurs a semantic violation and that two other lexical candidates have the correct meaning, which will lead to the promotion of the *LEX constraint of the incorrect winning candidate and the demotion of the *LEX constraints of the candidates with the correct meaning will fall. Also, FAITH */ɪ/→|ɛ| will be demoted because one of the correct candidates, viz., |tʃɛka| 'Czech', has a phonological form that violates this constraint.

/tʃɪka/ Meaning: 'Czech'	FAITH */ɪ/→|ɛ|	*LEX |tʃɛka| 'Czech'	*LEX |tʃɪka| 'girl'	*LEX |tʃɪka| 'Czech'	*LEX |tʃika| 'girl'	FAITH */ɪ/→|i|
☞ |tʃika| 'girl'					←*	*
√ |tʃɛka| 'Czech'	*!→	*→				
|tʃɪka| 'girl'			←*!			
√ |tʃɪka| 'Czech'				*!→		

Tableau 6.6. Reranking of FAITH as a result of a semantic error.

As for the perceptual consequence of solving these semantic errors, if we assume that with every evaluation one of the two semantically correct candidates is the winner, choosing the candidate that violates FAITH will incur a phonological error that leads to the mismatch between the outputs of recognition and perception. In the example of Tableau 6. 6, this means that if |tʃɛka| 'Czech' is chosen as the recognized word, it will be different from the phonological form of the perceptual input. This situation will generate perceptual learning because this sort of perceptual development occurs whenever there is mismatch between the output of recognition and the perceptual input, a conflict that the learner will notice, as demonstrated in § 2.2, 3.4, and 5.4.

Consequently, GLA lexicon-driven learning will occur every time Spanish words containing /i/ or /e/ are perceived as /ɪ/ and later recognized as a lexical item containing a different vowel, i.e. if there is a phonological error that leads to a perception-recognition mismatch. This type of learning leads to the reranking of the relevant F1-to-vowel cue constraints in the developing L2 perception grammar. Thus, in the case of the recognition of /tʃɪka/ 'girl' as |tʃika| 'girl', which occurs 50% of the time that the recognition grammar solves the semantic errors in Tableaux 4 and 5, the cue constraints against perceiving F1 values as /ɪ/ will be promoted and those against perceiving F1 values as /i/ will be demoted, as shown in Tableau 6.7.[62]

[F1 = 380 Hz] Recognition: |tʃika| 'girl'	380 Hz not /ɛ/	380 Hz not /i/	380 ms not /ɪ/
√ /tʃika/		*!→	
* ☞ * /tʃɪka/			←*
/tʃɛka/	*!		

Tableau 6.7. Lexicon-driven constraint reranking that results from recognizing /ɪ/ as |i|.

Here we see that a token of *chica* with the first vowel produced with an F1 value of [380 Hz], which was initially perceived as /tʃɪka/, will be more likely to be perceived as /tʃika/. This lexicon-driven GLA reranking will lead to the gradual shifting of the perceptual category boundary between a Dutch learner's L2 categories /i/ and /ɪ/ towards the location of the Spanish boundary between /i/ and /e/,

62 Importantly, this reranking also occurs in many other cases where *LEX |tʃika| 'girl' is low enough to warrant correct recognition.

which is at 407 Hz. In the perception grammar, this means that the cue constraints that map F1 values between 365 Hz (the L2 boundary between /i/ and /ɪ/) and 407 Hz (the Spanish boundary between /i/ and /e/) will be promoted, so that the probability of perceiving /ɪ/ when faced with tokens of Spanish *chica* will decrease. Crucially, this perceptual learning also applies to all the other Spanish words that do not constitute minimal pairs, e.g., if a perceived input /pɪta/ 'rope', a Spanish word that has no /e/ counterpart, is recognized as |pita| 'rope'. Figure 6.14 illustrates the lowering of the boundary between the L2 categories /i/ and /ɪ/ which results from lexicon-driven perceptual learning.

Original perceptual input:
/pɪta/
Meaning: 'rope'

Recognition after semantic error:
|pita| 'rope'

Result: Gradual boundary shift
in perception →

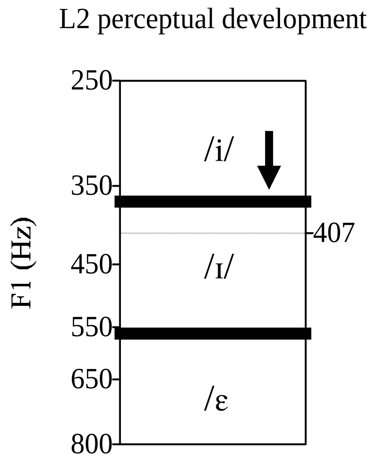

Fig. 6.14. Recognized candidate and resulting lexicon-driven boundary shift between the L2 perceptual categories /i/ and /ɪ/ in the SUBSET scenario.

Similarly, the phonological error that occurs when *checa* is perceived as /tʃɪka/ but recognized as |tʃɛka| 'Czech', which occurs 50% of the time that the semantic error in Tableau 6.6 is solved, will lead to the reranking of the relevant cue constraints.[63] As a result of this lexicon-driven constraint reranking, the L2 cue constraints in the developing L2 perception grammar that map the F1 values between 407 Hz (the Spanish boundary between /i/ and /e/) and 567 Hz (the L2 boundary

[63] This reranking also occurs in many other cases where *LEX |tʃɛka| 'Czech' is low enough to warrant correct recognition.

between /ɪ/ and /ɛ/) onto /ɪ/ will be promoted, so that the category /ɪ/ will be reduced in the perceptual outputs for tokens of the Spanish word *checa*. Tableau 6.8 shows the lexicon-driven perceptual learning that reduces the perception of *checa* as /tʃɪka/.

[F1 = 500 Hz] Recognition: |tʃɛka| 'Czech'	500 Hz not /i/	500 Hz not /ɛ/	500 ms not /ɪ/
/tʃika/	*!		
* ☞* /tʃɪka/			←*
√ /tʃɛka/		*!→	

Tableau 6.8: Lexicon-driven constraint reranking that results from recognizing /ɪ/ as |ɛ|.

Here we can see that a token of *checa* with the first vowel produced with an F1 value of [500 Hz], which was initially perceived as /tʃɪka/, will be more likely to be interpreted as /tʃɛka/. The lexicon-driven GLA reranking shown in Tableau 7 will lead to the shifting of the perceptual category boundary between Dutch learners' L2 categories /ɪ/ and /ɛ/ towards the location of the Spanish boundary between /i/ and /e/. Importantly, this perceptual development also applies to Spanish words containing /e/ that do not constitute a minimal pair, e.g., if the Spanish word *mes* 'month' is perceived as /mɪs/ but recognized as |mɛs| 'month'. Figure 6.15 illustrates the lowering of the boundary between the L2 categories /ɪ/ and /ɛ/ and the rising of the boundary between /i/ and /e/, which are a result of lexicon-driven perceptual learning.

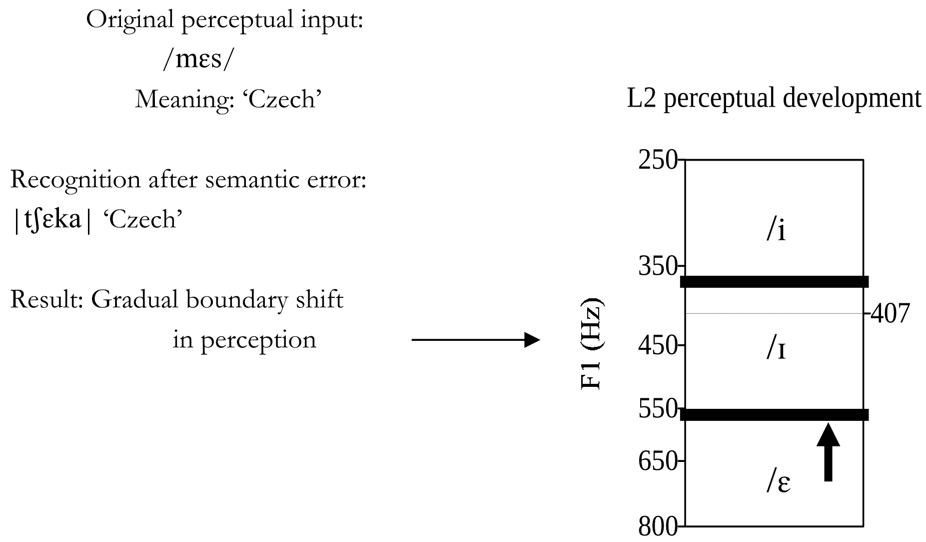

Fig. 6.15. Recognized candidate and resulting lexicon-driven boundary shift between the L2 perceptual categories /ɪ/ and /ɛ/ in the SUBSET scenario.

However, if the recognition grammar has *LEX |tʃɪka| 'Czech' as the lowest ranked *LEX constraint, as shown in (6.6), the semantic error that allows for the shifting of the boundary between /ɪ/ and /ɛ/ will no longer occur. This is because a perceptual input /tʃɪka/ 'Czech' will always be correctly recognized as |tʃɪka| 'Czech', a situation that matches the perceptual input in form and meaning.

(6.6) Constraint ranking that leads to a correct recognition of /tʃɪka/ 'Czech'
*LEX |tʃɛka| 'Czech' >> *LEX |tʃɪka| 'girl' >> *LEX |tʃika| 'girl' >>
*LEX |tʃɪka| 'Czech' >> FAITH */ɪ/→|ɛ| >> FAITH */ɪ/→|i|

Consequently, this correct recognition will have an effect on both L2 recognition and L2 perception because the learner will always recognize tokens of the Spanish word *checa* as both |tʃɛka| 'Czech' and |tʃɪka| 'Czech' as well as perceiving them as both /tʃɪka/ and /tʃɛka/. This non-optimal situation arises because the recognition-perception mismatch that leads to the boundary shift in Figure 15 can cease before the L2 boundaries come to match optimal L2 perception. Fortunately, this new constraint ranking induces a different semantic error, namely the recognition of the perceptual input /tʃɪka/ 'girl' as |tʃɪka| 'Czech' because *LEX

|tʃɪka| 'Czech' is the lowest ranked *LEX constraint. Tableau 6.9 shows the reranking that will occur as a result of this new semantic error.

/tʃɪka/ Meaning: 'girl'	*LEX |tʃɛka| 'Czech'	*LEX |tʃɪka| 'girl'	*LEX |tʃika| 'girl'	*LEX |tʃɪka| 'Czech'	FAITH */ɪ/→|ɛ|	FAITH */ɪ/→|i|
√ |tʃika| 'girl'			*!→			*
|tʃɛka| 'Czech'	←*!				*	
√ |tʃɪka| 'girl'		*!→				
☞|tʃɪka| 'Czech'				←*		

Tableau 6.9. Reranking of *LEX |tʃika| 'girl' >> *LEX |tʃɪka| 'Czech' due to a semantic error when recognizing /tʃɪka/('girl') as |tʃɪka| 'Czech'.

In order for this semantic error to be corrected, *LEX |tʃɪka| 'Czech' and |tʃika| 'girl' will need to exchange rankings to prevent the error from occurring again. Once these two *LEX constraints have changed places, they will go back to the ranking in Tableau 6.8 and again incur the semantic error that was previously described, viz., recognizing /tʃɪka/ 'Czech' as |tʃika| 'girl'. This misrecognition is shown in Tableau 6.10.

/tʃɪka/ Meaning: 'Czech'	*LEX |tʃɛka| 'Czech'	*LEX |tʃɪka| 'girl'	*LEX |tʃɪka| 'Czech'	*LEX |tʃika| 'girl'	FAITH */ɪ/→|ɛ|	FAITH */ɪ/→|i|
☞ |tʃika| 'girl'				←*		*
√ |tʃɛka| 'Czech'	*!→				*	
|tʃɪka| 'girl'		←*!				
√ |tʃɪka| 'Czech'			*!→			

Tableau 6.10. Reranking of *LEX |tʃɪka| 'Czech' >> *LEX |tʃika| 'girl' due to a semantic-error when recognizing /tʃɪka/('Czech') as |tʃika| 'girl'.

Thus, a *continuous* promotion and demotion of lexical constraints will occur,[64] and the semantic errors that cause the demoting and promoting of the *LEX |tʃɪka| 'Czech' and *LEX |tʃika| 'girl' constraints will lead to the recognition of |tʃɛka| 'Czech' and |tʃika| 'girl'.

As for the consequence of these continuous rerankings on the L2 perception grammar, the semantic errors shown in Tableaux 6.9 and 6.10 will result in phonological errors that generate GLA perceptual learning. Therefore, the learner's perceptual boundaries will continue to move towards the optimal location. Crucially, the semantic errors in the recognition grammar will stop when they have resulted in 0% of perceptual inputs containing /ɪ/. In other words, these errors will end when the perceptual category /ɪ/ is reduced as a result of the approximation of the L2 boundaries shown in Figures 6.14 and 6.15.

Going back to the learners' L2 recognition grammar, we can propose that the complete reduction of the perceived category /ɪ/ will result in the lexical forms |tʃɪka| 'girl' and |tʃɪka| 'Czech' no longer being recognized as such but as |tʃika| 'girl' and |tʃɛka| 'Czech' respectively. Importantly, the FAITH constraints against changing perceptual inputs containing either /i/ or /ɛ/, i.e., FAITH */i/→|ɪ|, FAITH */ɛ/→|ɪ|, FAITH */ɛ/→|i|, FAITH */i/→|ɛ|, will not move from their original high-ranked position because these perceptual inputs will never lead to semantic errors. Consequently, /tʃika/ and /tʃɛka/ will never result in the recognition of |tʃɪka| 'girl' or |tʃɪka| 'Czech'.[65]

This proposal for L2 lexical and perceptual development relies heavily on the existence of minimal pairs which, fortunately for the learners, are abundant in Spanish. However, as mentioned above, the perceptual reduction of /ɪ/ applies to all Spanish words containing /i/ and /e/ because the reduction of /ɪ/ via the semantic errors of minimal pairs first results in the disfavour and subsequent reduc-

---

[64] Due to the properties of the GLA, this continuous reranking of constraints will not lead to any special situation within the L2 recognition grammar other than the emergence of semantic errors. That is, no special action will be taken in order to solve the continuous reranking of *LEX |tʃika| 'girl' and *LEX |tʃɪka| 'Czech' because every constraint in this L2 recognition grammar will undergo a similar gradual reranking during message-driven recognition learning. For instance, *LEX |tʃɪka|'girl' and *LEX |tʃɛka|'Czech' will also be continuously promoted and demoted. This means that the learner's GLA will have no way of discriminating between different types of continuous rerankings so that they will continue to occur until an external force such as perception stops them.

[65] Notice that this ban on the recognition of lexical candidates containing |ɪ| will occur regardless of the lower ranking of *LEX |tʃɪka| 'Czech' with respect to *LEX |tʃɛka| 'Czech'.

tion of the entire perceived /ɪ/ category. That is, minimal pairs will initiate the reranking in the recognition grammar that generates reranking in perception. Crucially, the reduction of /ɪ/ will later apply to every Spanish word that contains a vowel produced with the F1 values that fall within the distributions of Spanish /i/ and /e/ because the perception grammar contains F1-to-vowel cue constraints.

However, although this learning proposal seems to succeed in reducing the middle L2 category in a SUBSET scenario, the mechanisms for recognition and perceptual learning as well as their interaction would surely benefit from a validation by a GLA computer simulation similar to the one demonstrated in Boersma & Escudero (2004). Nevertheless, there is no doubt that this is a promising avenue for a comprehensive explanation of L2 development in the SUBSET scenario as well as the development of lexical access in both L1 and L2 acquisition.

At this point in the discussion, it may prove helpful to summarize the combined lexical and perceptual proposal advanced in this section. Figure 6.16 shows the predicted interactive development of recognition and perception in the SUBSET scenario.

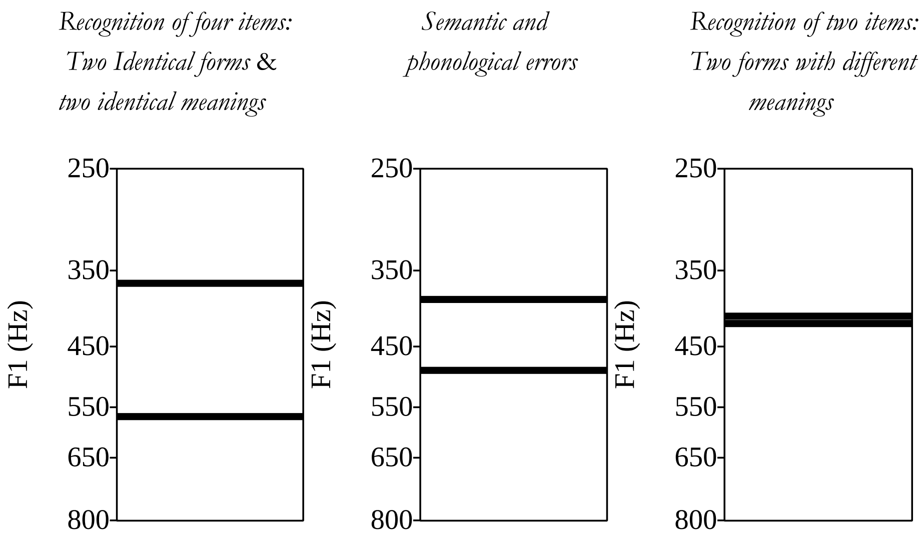

Fig. 6.16. Perceptual and recognition development (from left to right) for Dutch learners of Spanish.

Here we can see that a beginning Dutch learner of Spanish initially perceives three different vowels when confronted with Spanish /i/ and /e/. This situation will lead to the lexical storage of four lexical items, two of them having identical phonological forms and the other two having identical meanings. Thus, the existence of identical meanings and forms in the lexicon will automatically bring about semantic errors that will reduce the perception of /ɪ/ and the recognition of lexical items containing |ɪ|. This is because these errors will lead to a phonological error, which in turn will generate the recognition-perception mismatch that triggers lexicon-driven perceptual learning. In the case at hand, lexicon-driven learning will result in the reduction of the middle category through the approximation of the perceptual category boundary between /i/ and /ɪ/, on the one hand, and the boundary between /ɪ/ and /ɛ/, on the other hand. Finally, the complete reduction of the perception of the middle category through the merger of the two boundaries, as shown on the right side of Figure 16, will result in the recognition of only two lexical items, thereby matching the optimal number of perceived and recognized categories in the L2.

6.2.5 Ingredient 5: predicting the L2 end state

The L2LP's interpretation of Grosjean's (2001) language mode continuum can be referred to as the *language mode activation hypothesis* according to which the parallel activation of separate systems can lead to intermediate responses because the output of the two perception grammars can be merged when given a categorization response. That is, intermediate L1-L2 sound perception is a consequence of the activation of the learner's two perception grammars during on-line perception, rather than being a consequence of having a single set of categories and perceptual mappings for the two languages. The L2LP model thus predicts that L2 learners have different perception grammars for their two languages and that any result of intermediate perception depends on parallel activation during online speech perception. Thus, three different language modes are predicted to be found in L2 learners, namely a monolingual L1 mode, a completely bilingual mode, and monolingual L2 mode. This proposal for the relation between L1 and L2 perception is summarized in Table 6.4.

Table 6.4. L2LP proposal for the relation between L1 and L2 perception.

L2LP end state	L2 Proposal	Prediction for perceptual behaviour
L1-L2 relation: Separate grammars & and language mode activation	Separate grammar for L2	Optimal L2 perception under monolingual L2 setting
	Separate grammar for L1	Optimal L1 perception under monolingual L1 setting
	Parallel on-line activation of two separate grammars	Intermediate L1-L2 perception under bilingual setting

In the case of Dutch learners of Spanish /i/ and /e/, the reduction of the middle category will not transfer to the Dutch recognition and/or perception of vowels because this category is reduced in their L2 to optimally cope with their new production environment. Therefore, it is predicted that Dutch learners of Spanish will maintain the optimal number of perceived and recognized categories for Dutch, viz., /i/, /ɪ/, and /ɛ/. Figure 6.17 shows the three predicted modes for advanced Dutch learners of Spanish.

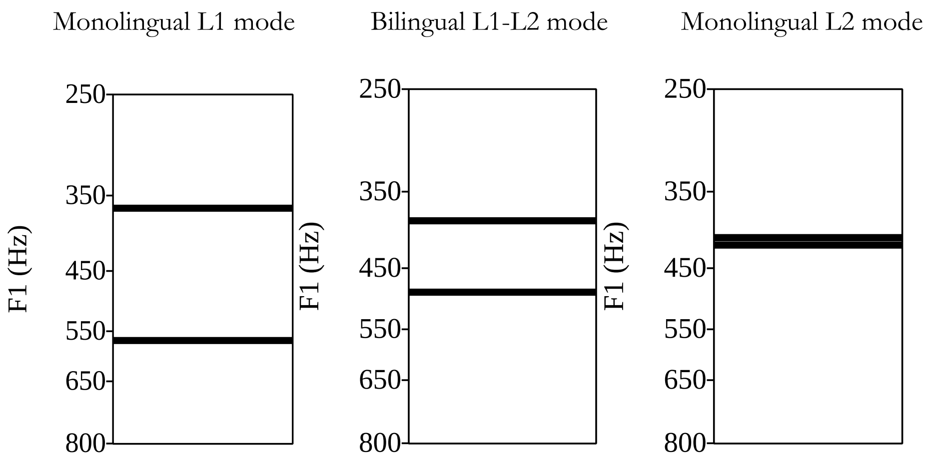

Fig. 6.17. Predicted perception modes in advanced Dutch learners of Spanish confronted with three different language settings.

Here we can see that Dutch learners will perceive three vowels when listening to their L1, i.e., when their monolingual L1 mode is activated. If they are primed to activate both of their languages at the same time but asked to respond with Dutch categories, they will automatically and unconsciously reduce the use of the vowel that is not found in Spanish, depending on the degree and manner in which the languages are primed. Finally, when advanced Dutch learners of Spanish activate their L2 only, i.e., when the use of their L1 is inhibited as much as possible, they will have a boundary location that matches that of the optimal L2 perception.

6.3 Evidence: Dutch learners of Spanish

Table 6.5 summarizes the L2LP predictions for a SUBSET L2 sound perception scenario. These predictions refer to the model's five ingredients, viz., the optimal perception in the two languages, the three logical states in language acquisition, and the L2 learning task.

Table 6.5. Summary of the L2LP predictions for the SUBSET scenario.

L2LP ingredients	Predictions for SUBSET	Description
Optimal L1 & L2	Spanish and Dutch listeners will exhibit optimal L1 perception	Optimal category boundaries
Initial state	Beginning Dutch learners will be equal to monolingual Dutch listeners	L1 category boundaries and L1 categories
Learning task	Reduce lexical and perceptual categories	
Development	GLA lexical learning and GLA lexicon-driven perceptual learning	Lexical reduction and perceptual boundary shift

End state	Dutch learners will **attain** optimal L2 perception and will **maintain** their optimal L1 perception	Two categories and optimal L2 boundary in their L2 mode, three categories in their L1 mode, intermediate boundaries in their bilingual mode

To test these specific predictions, I will now report on the results of a series of experiments that investigated the perception of Spanish /i/ and /e/ by Dutch learners. As mentioned in Chapter 3, the five ingredients that are part of the L2LP are also intended to be methodological phases for conducting L2 sound perception studies. Thus, Escudero & Boersma's (2002) perceptual data will be described here in accordance with the L2LP's methodology and sequence of predictions (cf. § 6.2). Their first objective was to show that the SUBSET L2 perception scenario does indeed exist and that it poses a problem for L2 learning. Their second objective was to provide evidence for the notion of 'language mode' in that their experimental design aimed at finding out whether speakers of two languages perceive them differently. Finally, the third objective of their study was to discover whether learners are able to solve the perceptual problem that results from a multiple perceptual mapping of L2 sound categories. In order to answer their research questions, the authors devised the experimental design shown in Table 6.6.

Table 6.6. Escudero & Boersma (2002)'s experimental design.

	Experiment 1: L1 mode only	**Experiment 2: L1&L2 modes**	**Experiment 3: L2-only**
Target stimuli:	50 Spanish CVC	50 Spanish CVC	50 Spanish CVC
Primed perception mode:	**L1**	**L1 + L2**	**L2**
The subjects were told that the stimuli were...	'Dutch'	'Spanish'	'Spanish'
Carrier phrase:	Dutch	Spanish	Spanish
Filler stimuli:	Dutch	Spanish	Spanish
Requested report:	**L1**	**L1**	**L2**
Explicit task: 'listen with your'	-	'Dutch ears'	'Spanish ears'
Response categories:	12 Dutch vowels	12 Dutch vowels	5 Spanish vowels

As we can see, the same target stimuli were presented in the three experiments. In this section, I will only report on the 50 Spanish CVC tokens that contained the vowels /i/ and /e/. It is important to mention, however, that their target stimuli consisted of 125 CVC tokens, i.e., 25 for each of the 5 Spanish vowel monophthongs. With respect to the language set conditions, a plausible interpretation of the authors' three experiments is that they refer to three different types of language mode activation. Thus, in Experiment 1, listeners were put in a monolingual L1 mode because the language setting induced them to only activate their L1. Experiment 2 can be interpreted as priming a bilingual mode because, although the listeners were instructed to respond in their L1, they were told that the language they were listening to was their L2. Given that the relative power of language activation that is found in explicit instruction versus stimuli information is not known, it can only be concluded that both languages could have been activated simultaneously to different degrees. Finally, Experiment 3 primed a monolingual L2 perception because the listeners were led to believe that they could only use their L2 to perform the task. However, since this experiment was performed after the one in which the learners had their L1 active, it is not clear a priori if they were able to completely inhibit it.

The perceptual data in Escudero & Boersma (2002) are not reported here following the authors' experimental sequence but rather in accordance with the five L2LP phases. Thus, their monolingual Dutch and Spanish results (Experiments 1 and 3) will be reported in § 6.3.1. Then, the beginning learners' L2 data (Experiment 3) will be described in § 6.3.2 so as to test the Full Copying hypothesis. Finally, in § 6.3.3, I will discuss the perception of all three groups of learners, showing that this directly bears on the L2 development (Experiment 3), the separate perception grammar hypothesis (Experiments 1 and 2), and the language mode activation hypothesis (Experiments 1, 2, and 3).

6.3.1 Model ingredient 1: Dutch and Spanish perception data

In this section, I report on the native perception of the 40 Spanish listeners who had no knowledge of Dutch and the 11 Dutch listeners who had no knowledge of the Spanish examined in Escudero & Boersma (2002). These subjects were tested at the Institute of Phonetic Sciences of the University of Amsterdam, and both groups were presented with the same 50 Spanish CVCs containing /i/ and /e/. The Spanish target tokens were extracted from a text read by a bilingual Spanish-Dutch speaker who was judged to have no foreign accent in either language, and were presented together with 25 tokens of each of the other three Spanish vowels /a, u, o/. These were CVC tokens containing consonants found in both Dutch and Spanish, which means that certain rhotics, approximants, and dentals were avoided. This was done so that both types of listeners could assume that they were listening to tokens of their own language. The Dutch listeners also heard 55 Dutch filler CVCs containing consonants and vowels typical of Dutch only, e.g., [høs], while the Spanish listeners also heard 55 Spanish filler CVCs with consonants typical of Spanish only, e.g., [roʀ]. Figure 6.18 shows the F1 values of the 50 target tokens. The Dutch mean F1 values for /i/, /ɪ/, and /ɛ/ (cf. § 6.2.1) are presented next to the Spanish target stimuli for comparison.

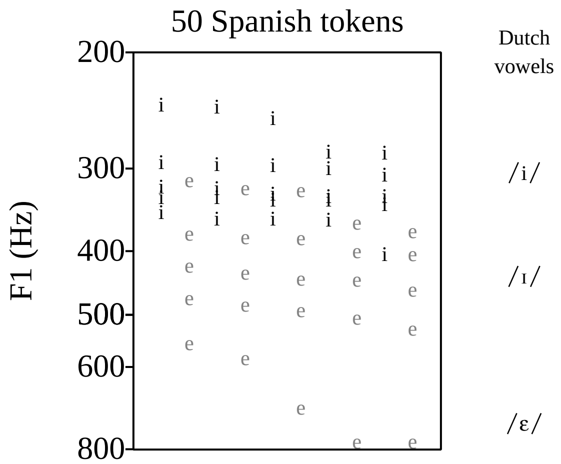

Fig. 6.18. F1 values for the target stimuli (25 /i/, 25 /e/) and for the three Dutch vowels (between slashes).

Escudero & Boersma (2002) assumed that the optimal perception boundary between Spanish /i/ and /e/ was the boundary of the intended vowel production. However, it seems that due to consonantal context, speech rate, or speaker idiosyncrasy, the F1 values of some tokens of one vowel fell in the F1 area of the other vowel. Given that the listeners did not have access to any semantic context, the characteristics of the target stimuli might have affected their vowel categorization if they relied on the F1 values to categorize the vowels. Thus, judging from the overlap in the productions shown in the figure, it can be predicted that at least 5 /e/ tokens should be perceived as /i/ by an optimal Spanish listener because those tokens have F1 values that fall in the /i/ region. Likewise, the Spanish listener should perceive one /i/ token as /e/ because its F1 value falls in the /e/ region. Thus, Spanish listeners were expected to have a category boundary that was approximately four tokens lower than the production boundary.

With respect to the Dutch listeners, these productions of Spanish /i/ and /e/ might lead to a higher perception of /ɪ/. In addition, given that there were at least 7 Spanish /e/ tokens with values higher than 567 Hz which was where the boundary between Dutch /ɪ/ and /ɛ/ was situated, these listeners might increase the use of /ɛ/ for Spanish /e/ tokens. In sum, given the nature of the 50 target tokens, the Dutch perception was predicted to resemble the perception percentages computed in § 6.2.1, whereas the Spanish listeners were predicted to have an /i/ bias in their perception.

The Dutch listeners performed Experiment 1 and the Spanish listeners performed Experiment 3 (cf. Table 6). This means that they were asked to categorize the 50 Spanish tokens as one of their native vowel categories, i.e., the former had the 12 Dutch vowel monophthongs /ɑ, a, ɛ, ɪ, e, i, ʏ, ø, y, ɔ, o, u / while the latter had the five Spanish monophthongs /a, e, i, o, u/ as response options. Both groups of listeners were asked to select the vowel they heard by clicking on a computer display showing the language-specific response options. Thus, the Dutch listeners were asked to classify the vowels in the stimuli as one of the 12 Dutch vowels, which were presented orthographically between <h> and <k> to avoid the ambiguity problems of Dutch orthography. On the other hand, given that Spanish has a transparent orthography in that written vowel symbols are always pronounced the same, the options for Spanish listeners were the graphemes <i, e, a, o, u>. To enhance the monolingual mode, the Dutch listeners heard the Spanish stimuli embedded in the Dutch carrier phrase *luister naar [kes]* 'listen to [kes]' and were told that all the stimuli had been cut from a Dutch text.

Figure 6.19 shows the extent to which Dutch listeners used their three closest vowels to categorize the 50 Spanish front vowel tokens (25 /e/ and the 25 /i/), and how optimal the perception of the Spanish listeners was. In the figure, the vertical dimension represents an F1 continuum because the majority of the /i/ tokens had lower F1 values than the majority of /e/ tokens, and therefore it was assumed that the 50 tokens formed a near continuum along the vowel height dimension. Thus, the top and bottom dotted lines demarcate the tokens, and the middle dotted line represents the production boundary between /i/ and /e/ tokens. On average, the Dutch monolinguals classified 6.1 /i/ tokens as /ɪ/ so that it was simply assumed that the remaining 18.9 were classified as /i/. This is represented as an /ɪ/-/i/ category boundary with a height of 6.1/25 between the dotted line in the middle and the one on top. On the other hand, they classified 18.1 /e/ tokens as /ɪ/ and here again it was assumed that the remaining 6.9 were classified as /ɛ/, thus leading to the /ɛ/-/ɪ/ boundary location of 6.9/25 between the dotted line on the bottom and the one in the middle. This means that the category /ɪ/ was chosen to classify 24 out of the 50 Spanish tokens.

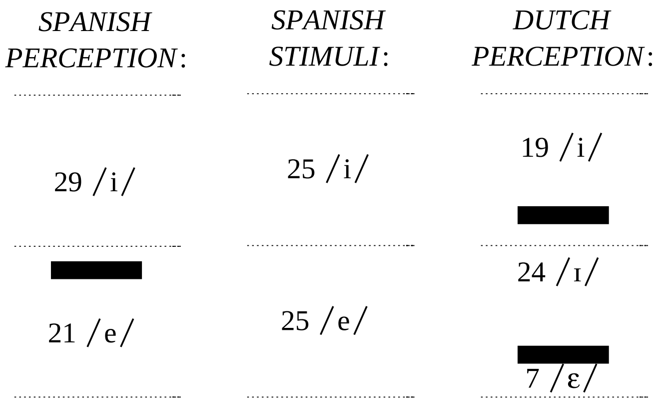

Fig. 6.19. Native Spanish and Dutch perception of the 25 /i/ and 25 /e/ Spanish tokens.

Given that the majority of the Spanish /i/ and /e/ tokens were perceived as such by the Spanish listeners and as /i/, /ɪ/ and /ɛ/ by the Dutch listeners, I will only discuss the use of these particular vowel categories for each group. Table 6.7 shows the average Dutch and Spanish categorization results in tokens and in percentages.

Table 6.7. Average token and percentage categorization for native Spanish and Dutch listeners.

	/i/	/e/	/ɛ/	/i/→ /ɪ/	/e/→ /ɪ/	/i/→ /e/	/e/→ /i/
Dutch	18.9 (37.8%)		6.9 (13.8%)	6.1 (12.2%)	18.1 (36.2%)		
Spanish	24 (48%)	19.5 (39%)				1 (2%)	5.5 (11%)

As predicted by the F1 token distributions, the Spanish listeners misclassified some of the 25 /e/ tokens and only a few of the 25 /i/ tokens. More specifically, they misclassified 5.5 /e/ tokens and only 1 /i/ token so that their boundary mismatch was 4.5 tokens. On the other hand, the Dutch listeners perceived 75.6% (18.9 out of 25) of the Spanish /i/ tokens as Dutch /i/ and the rest (24.4% or 6.1

tokens of 25) as Dutch /ɪ/. The Spanish /e/ tokens yielded a 72.4% Dutch /ɪ/ perception and a 28.6% /ɛ/ perception, which was predicted from the high F1 values of some Spanish /e/ tokens. Again as predicted, the total use of /ɪ/ in the perception of these Spanish /i/ and /e/ tokens was 48.4%, and this was close to the predicted use of Dutch /ɪ/ for the Spanish /i/ and /e/ tokens, which was 53.4%.

6.3.2 Model ingredient 2: cross-language and initial L2 perception data

In this section, I report on the perception of Spanish /i/ and /e/ by Dutch listeners with no knowledge of Spanish and by beginning Dutch learners of Spanish when performing Escudero & Boersma's (2002) Experiment 2. In this experiment, the 11 Dutch listeners from the previous section and 11 beginning learners of Spanish were tested. The latter were enrolled in the BA Spanish program at the University of Amsterdam and were classified as beginners on the basis of a language background questionnaire which revealed that they had had less than a year of instruction in the Spanish language. The two groups were tested at the Institute of Phonetic Sciences at the University of Amsterdam. They were presented with the same 50 /i/ and /e/ target stimuli except that this time they were embedded in the Spanish carrier phrase *la palabra [kes]* 'the word [kes]'. The subjects also heard the 55 Spanish fillers (cf. Experiment 2 in Table 6), and as part of the instructions, they were told that the language they would be hearing was Spanish. However, they were still asked to listen to the stimuli with 'Dutch ears' and to classify them as Dutch vowels.

Given that the Dutch listeners did not have any knowledge of Spanish, it was predicted that their performance in this task would be very similar to the one in Experiment 1. Crucially, according to the Full Copying hypothesis, the beginning learners should have exhibited a performance similar to that of the Dutch listeners if they had copied their L1 Dutch perception and if they had been absolute L2 beginners. As shown in Figure 6.20, both predictions were borne out. Note that the token categorization was computed in the same way as in § 6.3.1.

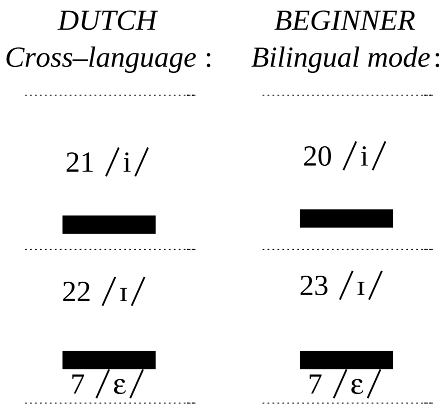

Fig. 6.20. Dutch and beginning L2 perception of Spanish in a Dutch-Spanish mode.

Importantly, this experiment explicitly tapped into the Dutch listeners' cross-language perception because they were aware that they were listening to a foreign language. In the case of beginning learners, it triggered the activation of their L2 because the stimuli were in Spanish but also of their L1 because they were requested to respond in Dutch. When comparing the boundaries on the left of the figure with those on the right, we see that the two groups of listeners have almost identical categorization. Crucially, if we compare the Dutch listeners' results here with those of Experiment 1 (cf. Figure 19), it is clear that the three Dutch categories, viz., /i/, /ɪ/, and /ɛ/, were used to the same extent when classifying the Spanish /i/ and /e/ tokens. From these results, we can deduce that both the perception of Dutch listeners with no knowledge of Spanish and the perception of beginning Dutch learners of Spanish are not affected by the bilingual Dutch-Spanish priming prompted in Experiment 2. That is, the cross-language and beginning L2 perception of the Spanish /i/ and /e/ tokens exhibit the same characteristics as the native Dutch perception of the same tokens, as predicted in § 6.2.2.

6.3.3 Dutch learners' L2 perception data

At this point, the question arises as to whether the bilingual mode conditioning in Experiment 2 had an effect on the perception of Dutch learners with more L2 experience in Spanish. Escudero & Boersma (2002) tested a total of 38 Dutch learners of Spanish enrolled in the BA Spanish degree programme at the University of Amsterdam. The authors divided the learners into three groups on the basis of a language background questionnaire with 11 of the subjects being the beginning learners reported in the previous section. In addition, 18 subjects were judged to be intermediate learners (second-year students) and 9 were deemed to be advanced learners (third and fourth year students). Figure 6.21 shows the categorization results of these three groups of learners when performing Experiment 2.

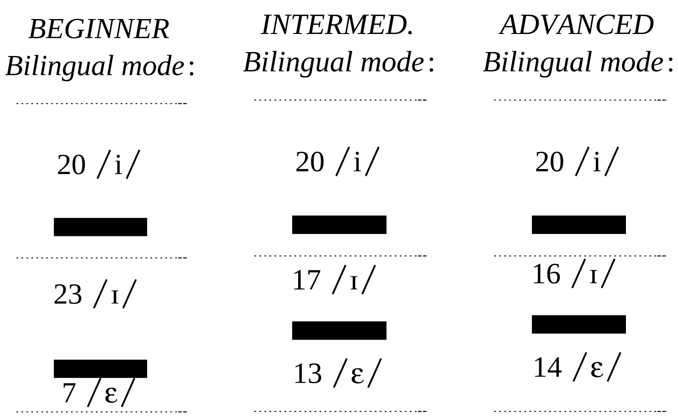

Fig. 6.21. Developmental reduction in the perception of /ɪ/ under the parallel activation of Dutch and Spanish (Experiment 2).

As can be seen, the intermediate and advanced learners used the /ɪ/ category less often than the beginners. Apparently, these two groups of learners could not simply listen to the tokens as Dutch. Rather, their different behaviour suggests that their Spanish perception was active. This learner group difference yielded a reliable effect of L2 experience on the extent of the use of the /ɪ/ category (for the three learner groups: $r = -0.38$, $p < 0.005$). That is, the more Spanish experience the learners had, the less often they used the /ɪ/ category when they listened to the Spanish tokens under a bilingual Spanish-Dutch mode.

Another measure of L2 development can be found in the learners' results for Experiment 3 which had the subjects classify the same 50 Spanish tokens under a monolingual L2 setting. Here, they performed the same task as the monolingual Spanish listeners described in § 6.3.1. Figure 6.22 shows the results of their perception under this monolingual L2 mode.

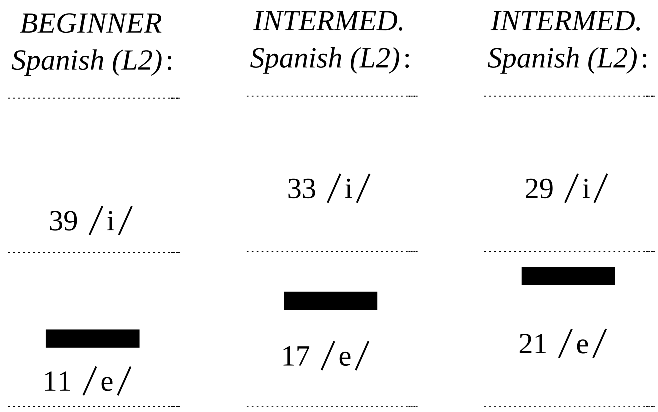

Fig. 6.22. Dutch learners' perception under a monolingual L2 (Spanish) mode.

Here we observe that the location of the boundary becomes more native-like as the learners' experience level increases. When taking into account all three groups, there was found to be a correlation between their L2 boundary location and their L2 experience level (r = –0.57, $p < 0.0001$). That is, the perceptual data reliably show that the more experience learners have with their L2, the more their category boundary between Spanish /i/ and /e/ will resemble the optimal boundary. What we observe is that the location of the boundary for the beginning learners exhibits a large difference in comparison with the L2 production distributions. Crucially, this highly inaccurate L2 perception seems to be related to their result in the bilingual Dutch-Spanish mode because their boundary between Dutch /ɪ/ and /ɛ/ corresponds to their L2 boundary between /i/ and /e/. This suggests that the beginning learners' use of the Dutch category /ɪ/ is also attested in their L2 perception.

With respect to the other two learner groups, recall that the optimal boundary was predicted to be approximately four tokens below the L2 production boundary because of the particular F1 values of the target Spanish tokens. This means that the intermediate and advanced learners approximated the optimal L2 perceptual boundary. In fact, the advanced learners matched the optimal L2 perception because their boundary was only four tokens away from the production boundary. Thus, the data from Experiments 2 and 3 show L2 development under bilingual Dutch-Spanish perception and under monolingual L2 perception.

To test the hypothesis of separate systems for L1 and L2 perception, the learners' monolingual L1 perception needed to be examined. Thus, the learners were also presented with Experiment 1, i.e., the same experiment reported for Dutch-only listeners in § 6.3.1. In other words, they also listened to the 50 target Spanish tokens as if they had been Dutch because the language setting in Experiment 1 induced them to think that the tokens had been drawn from a Dutch text. Figure 6.23 shows the learners' perception under this monolingual Dutch condition.

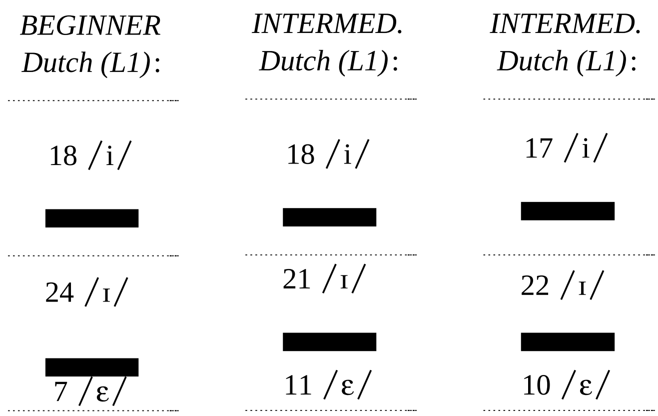

Fig. 6.23. Dutch learners' perception under a monolingual L1 (Dutch) mode.

Although the intermediate group seemed to use the /ɪ/ category to a lesser degree than any of the other four groups, a correlation test did not reveal a reliable pattern of development for the three learner groups. In addition, their use of the three Dutch categories was very similar to that of the Dutch listeners with no knowledge of Spanish. This means that the learners' L1 perception was not affected by their L2 development, as predicted by the separate grammars hypothesis.

Finally, to test whether the learning mechanism that leads to the attested L2 development is indeed the reduction of the extra category /ɪ/ through the approximation of two perceptual boundaries, we can compare the /ɪ/ categorization of the Spanish vowels in Experiments 1 and 2, as shown in Table 6.8.

Table 6.8. Average reduction of /ɪ/ perceptions versus the number of L2 /i/ and /e/ responses.

	/ɪ/ (Exp. 1)	/ɪ/ (Exp. 2)	L2 boundary
Beginners	24	23	39 /i/-11 /e/
Intermediate	21	17	33 /i/-17 /e/
Advanced	22	16	29 /i/-21 /e/

We can see that the learners' use of the /ɪ/ category is similar in Experiment 1 (L1 monolingual mode) but that it diminishes with L2 experience in Experiment 2 (bilingual Dutch-Spanish mode). The difference between the results of the two experiments is significantly different from zero for the intermediate and advanced learners. In addition, the reduction of the /ɪ/ category between the two experiments is developmental because it correlates with experience level ($r = 0.39$, greater than zero, $p < 0.005$). Crucially, the fact that these learners made less use of the /ɪ/ in Experiment 2 suggests that they have different language-specific *perception modes* for Dutch and Spanish because they performed differently depending on the language they thought they heard.

To confirm whether perceptual boundary shifting is the mechanism underlying the L2 development observed in Figures 6.20 and 6.21, we can compare the reduction in the use of /ɪ/ between Experiments 1 and 2 with the location of the learners' L2 category boundary. It turns out that the location of the L2 category boundary between Spanish /i/ and /e/ correlates strongly with the difference between the use of /ɪ/ in Experiment 1 and its use in Experiment 2 ($r = –0.62$, less than zero, $p < 0.00001$; 95% = –0.78...–0.37). That is, beginning learners exhibited either a small or unreliable reduction in the use of the extra category while having a non-optimal boundary. Conversely, intermediate and advanced learners manifested a significant reduction of the extra category along with an optimal or near optimal perceptual boundary.

6.3.4 Discussion

With respect to optimal perception, it was found that Spanish and Dutch listeners interpreted the stimuli in a language-specific optimal way. Also, as predicted, beginning learners were found to perceive 'too many categories' in Spanish when they listened to tokens either under a monolingual Dutch or bilingual Dutch-Spanish setting, just like the Dutch listeners with no knowledge of Spanish did. This could mean that their L2 perception system had failed to develop sufficiently to avoid the extra Dutch category in Spanish. This Dutch-like behaviour in the beginning learners is confirmed by their highly inaccurate L2 category boundary location under a monolingual L2 Spanish conditioning.

Regarding the L2 acquisition process, the learners were seen to undergo the predicted perceptual development. That is, under a bilingual Dutch-Spanish mode, they reduced their perception of the extra Dutch category. Crucially, this reduction correlated with the approximation of their L2 perceptual boundary to that of the Spanish perception. In fact, the advanced learners had an L2 boundary that matched the production distributions of the tokens more closely than the boundary exhibited by the native Spanish listeners. Therefore, it can be argued that the advanced learners had achieved optimal Spanish perception by disfavouring the use of the extra category.

As for the L1 perception of the Dutch learners of Spanish, there was no developmental correlation in the perceptual results when the three groups of learners listened to the Spanish stimuli in the L1 monolingual mode. This means that there was little or no change in their L1 perception and that the somewhat lesser use of /ɪ/ by the intermediate and advanced learners could be explained by the activation of their Spanish perception system. In other words, some intermediate and advanced learners might have perceived some of the tokens presented in Experiment 1 as non-Dutch or even as Spanish. This is because the target stimuli were naturally produced tokens which may have had cues that revealed their Spanish origin. Nevertheless, this effect was much smaller than the one found in the overtly bilingual Spanish-Dutch conditioning, which indicates that the subjects did maintain their L1 optimal perception.

6.4 Learning SUBSET sounds: the predictions versus the evidence

In this final section, I return to the L2LP predictions for the SUBSET scenario as well as to the findings that support them. Table 6.9 summarizes the predictions that follow from the five ingredients of the L2LP model which, as we have seen, refer to the three logical states in language acquisition, to the optimal perception in the two languages, and to the L2 learning task. In this table, each of the results outlined in the previous section is presented next to its specific prediction.

Table 6.9. The five predictions for a SUBSET L2 sound perception scenario and the evidence to support them.

L2LP ingredients	Predictions for "truly new"	Finding
Optimal L1 & L2	Dutch and Spanish listeners will exhibit optimal L1 perception	Borne out
Initial state	Beginning learners will be equal to monolingual Dutch listeners	Borne out
Learning task	Reducing the number of perceived categories	Borne out
Development	GLA lexicon-driven boundary shifting	Indirectly borne out
End state	Dutch learners will **attain** optimal L2 perception and will **maintain** their optimal L1 perception	Borne out

These findings confirm the Full Copying hypothesis because the beginning learners did start as optimal perceivers of their L1. This is because these learners used three L1 categories to perceptually map two L2 categories, just like the native listeners of their L1 did. These results also confirm that the SUBSET scenario indeed poses learning problems which are not attested in the NEW scenario. In other words, learners start with 'too many' categories in a SUBSET scenario, whereas they start with 'too few' categories in a NEW scenario. It was also shown that although both the SUBSET and NEW scenarios involve a representational task, its nature is different because in the former case it involves creating categories whereas in the latter case it involves reducing or disfavouring categories.

Importantly, the findings reported here clearly show the existence of a perceptual learning task. On the other hand, the potential existence of a lexical learning task was not directly tested because this would have required the examination of the initial L2 word recognition. What the attested L2 perception problem in beginning Dutch learners of Spanish suggests is that the initial L2 lexicon cannot help them avoid perceptual errors.

As predicted by the hypothesis of Full Access to the GLA, the present findings suggest that learners faced with a SUBSET perception scenario develop by adjusting their grammars to reduce the perception of L1 categories that do not exist in the L2. This perceptual category reduction is an instance of lexicon-driven boundary shifting which is an L1-like learning mechanism. The L2LP model provides the formal mechanism that underlies such a development. In addition, the occurrence of the predicted type of perceptual development suggests that the learners' lexical representations have developed because the type of perceptual development attested in the learners can only occurred if the lexicon contains the optimal number of L2 lexical items. However, this would need to be tested in a word recognition task, which would also provide more insight into the patterns of lexical development.

Finally, with respect to the separate grammars hypothesis, Escudero & Boersma's (2002) experimental design aimed at testing the relation between L2 experience and the possibility of language-specific perception modes. Therefore, they claimed that the results could be taken as evidence for the existence of two separate language-dependent and low-level perception systems in L2 learners. Crucially, it was found that under the bilingual Spanish-Dutch condition (Experiment 2), the learners' perception strongly correlated with L2 experience, whereas under the monolingual Dutch condition, no such correlation was found. The results of the monolingual L1 experiment showed that the learners continued using the optimal number of Dutch categories in their L1 perception. That is, they reduced the perception of a category in their L2 while the perception of the same category in their L1 remained stable. Thus, given the weight of the evidence, it can be concluded that in the case of Dutch learners of Spanish, the hypothesis of separate perception systems for L2 learners was borne out.

7 Learning SIMILAR L2 sounds

The goal of this chapter is to provide evidence for the acquisition of L2 perception when it involves sounds or sound contrasts that already exist in the L1. Unlike the NEW scenario, L2 learners faced with a SIMILAR scenario equate two L2 phonemes with two L1 phonemes. In other words, there is a ‘two-to-two’ identification of the L2 sounds with L1 categories for purposes of lexical storage, as shown in Figure 7.1.

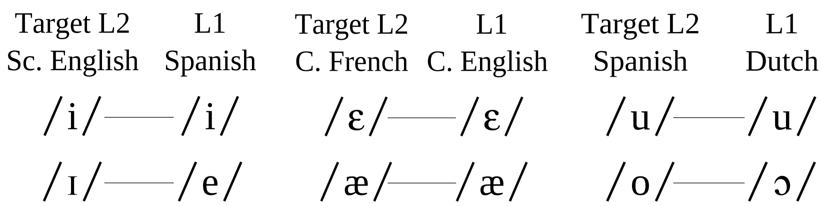

Fig. 7.1. Phonemic equation in the SIMILAR L2 perception scenario. SE = Scottish English, CE = Canadian English, and CF = Canadian French

Despite the equation of L1 and L2 categories at the abstract phonemic level, the main characteristic of the SIMILAR L2 perception scenario is a mismatch in the mapping from auditory events to phonological categories. Thus, some tokens (i.e., phonetic realizations) of one of the L2 categories are not mapped onto their L1 counterpart but rather onto the other L1 category, as shown in Figure 7.2.

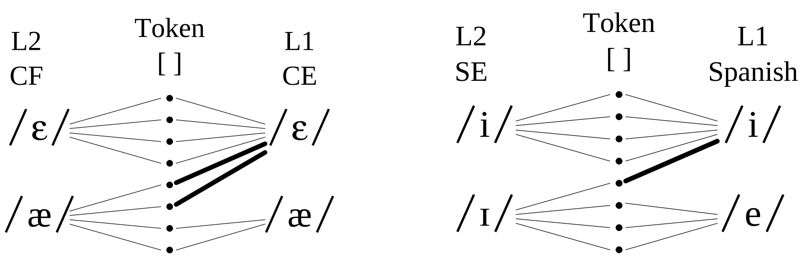

Fig. 7.2. Perceptual mapping in the SIMILAR L2 scenario. SE = Scottish English, CE = Canadian English, and CF = Canadian French

Recall that the L2LP model provides a general framework that explains the development of L2 sound perception. Table 7.1 provides a summary of its three main tenets. This model interprets the well-known hypothesis of Full Transfer to predict that, initially, the L1 perception grammar and the L1 categories will be copied to the L2 perception. It also interprets the hypothesis of Full Access as Full Access to the GLA in order to predict that L2 development will occur and that it will be guided by the same learning mechanisms that the algorithm performs in L1 perceptual learning (cf. § 2.1.3). Also, the proposal of two separate perception systems for the L1 and L2 presupposes that L2 perception will become optimal without affecting the already optimal L1 perception, which will remain stable provided that the learner continues to be exposed to her L1. Finally, the L2LP hypothesis of language mode activation predicts that the L2 learners' two languages can be activated at the same time when they are in a bilingual setting.

Table 7.1. Summary of L2LP five ingredients for L2 sound perception.

L2LP	Prediction	Explanation	Description
Optimal L1 & L2	Human beings are optimal listeners	Optimal listeners handle the environment maximally well	L1 and L2 optimal category boundaries: Location & shape
Initial state	= Cross-language perception	Full Copying	L1 boundary location and shape
Learning task	= Reach the optimal target L2 perception	L2 learners want to reach target	Bridging mismatches between L1 and target optimal perception
Development	= L1-like	Full GLA Access	Category formation and boundary shifts
End state	Optimal L2 perception and optimal L2 perception	Input overrules plasticity Separate grammars	Language activation modes, through language setting variables

According to this model, then, if two L2 sounds are equated to a single sound in the L1, the learner faces the common NEW learning scenario discussed in Chapter 5. However, if two L2 sounds are equated to two L1 sounds, the learner faces a SIMILAR scenario. Figure 7.3 illustrates the main difference between these two situations.

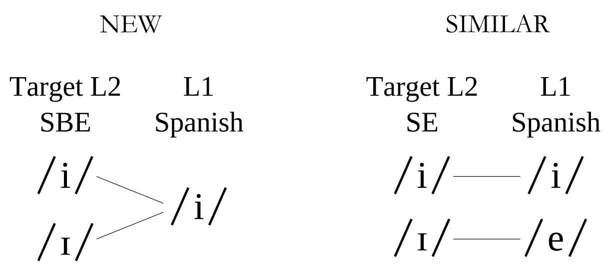

Fig. 7.3. Comparison between a case of NEW L2 sounds versus a case of SIMILAR L2 sounds. Southern British English (SBE).

As shown in Chapter 5, using a single category to represent two L2 sounds often presents a challenging learning problem because it involves the formation of new categories and the perceptual integration of new acoustic dimensions. In this chapter, it is argued that learning SIMILAR sounds also poses a learning problem because there will often be a mismatch between the L1 and the L2 perception of the two sounds in question. However, this idea seems to be controversial judging by previous L2 perception proposals. In the next section, I discuss the two opposing approaches to the learning task in this scenario and I present a summary of the L2LP proposal for the exact nature of this learning task.

7.1 Is there an L2 learning task in a SIMILAR scenario?

In a SIMILAR scenario, the L1 and L2 categories share acoustic properties that lead to an overlap in production distributions. However, it is often the case that such a production overlap is not complete, so that the L1 and L2 categories display fine phonetic production differences. Such differences may in turn lead to perceptual mismatches such as the ones illustrated by the mapping lines in Figure 7.2.

Despite the clear possibility of such a learning problem, the current proposals that explain L2 perception, which were reviewed in chapter 4, do not concur with the degree of difficulty that L2 acquisition poses in this scenario. In fact, they are divided into two major contradictory approaches, one suggesting that a SIMILAR scenario involves *no* L2 learning challenge, and the other suggesting that it poses the *greatest* L2 challenge. The first approach is represented by Brown's (1998, 2000) Phonological Interference model, by Best's (1995) Perceptual Assimilation Model, and by Kuhl's (2000) Native Language Magnet model. All three of these have in common the idea that the presence of L2 sounds, features, or phonetic dimensions in the L1 guarantees the absence of an L2 perceptual learning problem so that no need for an explanation of L2 perceptual development is required. In contrast, the second approach leads to the assumption that this scenario will be the most difficult one because L2 perception will never, or very rarely, resemble native-like perception. This is the one that can be found in Major's (2001) Ontogeny Phylogeny Model and in Flege's (1995, 2003) Speech Learning Model.

The L2LP model predicts that in a SIMILAR scenario, the number of phonological representations will already match the target L2 categories so that no category formation will need to occur. Thus, it is proposed that the learning task involves the adjustment of perception which can be performed without forming new categories but rather by reusing already existing ones. The model also predicts that learning SIMILAR sounds will be easier than learning NEW sounds. Table 7.2 summarizes the comparative predicted initial states, learning tasks, and degree of difficulty in these two scenarios.

Table 7.2. Comparative initial states and learning tasks in the NEW and SIMILAR scenarios.

L2LP proposal	**Prediction for NEW**	**Prediction for SIMILAR**
Initial state	Too few categories	Same number of categories
Perceptual task	1. *Create* perceptual mappings 2. *Integrate* auditory cues	Adjust category boundaries
Representational task	1. *Create* phonetic categories 2. *Create* segments	None
Degree of difficulty	Very difficult	Not difficult

The different degree of difficulty predicted for these two scenarios resides in the number and nature of the learning tasks that need to be performed to attain optimal perception. The L2LP model also explains these different degrees of difficulty by proposing that more learning mechanisms are at play in a NEW scenario than in a SIMILAR one. That is, the NEW scenario in which the learning task is to create new perceptual mappings and new categories, will involve both category formation and boundary shifting, which are the L1 learning mechanisms described in §2.1.3. In contrast, it is predicted that the SIMILAR scenario will *only* involve the boundary shifting mechanism that changes perception. As a consequence, the NEW scenario will require many more learning steps than the SIMILAR scenario, as shown in Table 7.3.

Table 7.3. Comparative predicted developmental paths for the NEW and SIMILAR scenarios.

Steps		**NEW**	**SIMILAR**
S 1	Full Copying	Copied L1 categories are too few	Copied L1 categories are the same in number
S 2	Full Access: Auditory-guided learning	Creating phonetic categories and rankings	------
S 3		Creating phonological categories and rankings	------
S 4		Using discrete categories for lexicalization	------
S 5	Full Access: Lexicon-guided learning	Category boundary adjustment	Category boundary adjustment
S 6		Low level integration of acoustic dimensions	------
S 7		Using integrated categories in the lexicon	------
S 8		Further boundary adjustments/abstraction	------

In the following section, I discuss how the predicted L2 learning steps for a SIMILAR scenario come about and how development is predicted to occur. Crucially, I show how the principled separation of sound categories and the grammar that performs sound categorization is able to adequately describe and explain the learning task as well as the perceptual behaviour of L2 learners faced with a SIMILAR scenario. In order to make specific predictions, in § 7.2 I present an application of the ingredients of the L2LP model to two cases of the SIMILAR L2 sound perception scenario, namely that of Spanish learners of Scottish English (SE) and that of Canadian English (CE) learners of Canadian French (CF). In § 7.3 and § 7.4, I produce empirical evidence to substantiate the predictions for each one of these cases. Finally, in § 7.5 I evaluate the L2LP's SIMILAR proposal in accordance with this empirical evidence.

7.2 Ingredients of L2 linguistic perception in a SIMILAR scenario

This section presents an application of the L2LP model to the prediction of L2 development in a scenario in which learners need to adjust their perception of L2 sound categories. The model's specific predictions for this L2 scenario will be illustrated with two examples, viz., the learning of SE /i/ and /ɪ/ by Spanish listeners and the acquisition of CF /æ/ and /ɛ/ by CE listeners, and it will be argued that such predictions can be extended to any case of L2 perception that evinces the same characteristics.

In the following sections, I will apply the five ingredients that constitute the L2LP model to the two cases that illustrate a similar scenario. While four of the ingredients are clearly related to L2 acquisition, the first ingredient refers to the optimal L1 perception of the languages involved. According to the model's proposal, describing the optimal perception in the L2 will allow us to determine the learning task and the target of perceptual learning. Also, according to the model's Full Copying hypothesis, measuring the optimal L1 perception will allow us to determine the system that the learner brings to the L2 task, i.e., the system she will initially use to cope with the L2 environment. Crucially, the L1 optimal perception will provide a reliable estimate of the system that the learner needs to maintain in order to be able to optimally cope with her L1 environment while she develops into an optimal L2 listener.

7.2.1 Ingredient 1: predicting optimal perception from environmental production

The L2LP model proposes that the first step in predicting L2 perception is to describe the production of the languages involved. For the similar case of Spanish learners of SE /i/ and /ɪ/, this involves measuring the production of SE /i/and /ɪ/ and that of the closest Spanish vowels, namely /i/ and /e/. This is because the model's Optimal Perception hypothesis states that perception strongly depends on the specific production environment, so that the optimal way of perceiving the sounds of a language depends on how such sounds are produced. Following the descriptions used in Chapter 5, Table 7.4 shows the production averages and standard deviations for the F1 and duration values of the SE and Spanish vowels.

Table 7.4. F1 and duration average values for Southern British English (SBE) and Spanish vowels. s.d. = Standard deviation

SE		Average	s.d.	Spanish		Average	s.d.
/ɪ/	duration	84.8 ms	0.4 dou.	/e/	duration	78 ms	0.4 dou
	F1	485 Hz	0.2 oct.		F1	502 Hz	0.2 oct.
/i/	duration	94 ms	0.4 dou.	/i/	duration	81 ms	0.4 dou.
	F1	343 Hz	0.2 oct.		F1	331 Hz	0.2 oct.

The F1 and duration geometric averages are taken from Escudero & Boersma (2003) for the English vowels, and from Cervera, Miralles & Gonzalez-Alvarez (2001) for the Spanish vowels. Just as in Chapter 5, the standard deviations are set at the same values in the two languages to ensure that all possible variation is taken into account. Figure 7.4 shows the production distributions in an F1-duration plane.

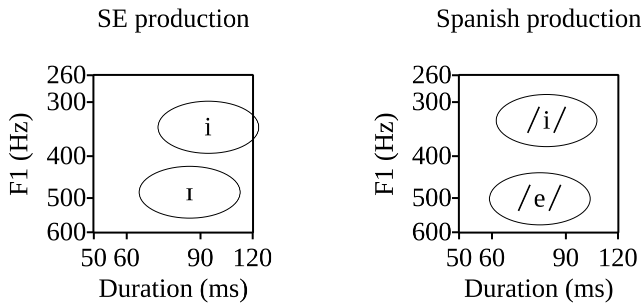

Fig. 7.4. SE and Spanish production distributions

To establish a production-perception relationship, we need to compute three production values, as was done in Chapter 5. First, we compute the location of the midpoint of the line that connects the two vowels in each language, a location that in this case has an F1 and a duration value. The SE midpoint has an F1 location of log2(485) - 0.500/2 octaves and a duration location of log2(94) - 0.149/2 duration doublings. This means that the SE midpoint for /i/ and /ɪ/ is located at [408 Hz, 89 ms].[66] For Spanish /i/ and /e/, the midpoint is located at [408 Hz, 80ms].[67]

The second production value we need to compute is the slope of the equal-likelihood line which is the line that connects all the F1-duration pairs that are intended 50% of the time as one vowel and 50% of the time as the other one. This slope is calculated as the ratio of the duration and F1 acoustic distances between the two vowels (as expressed in logarithmic units) multiplied by the squared ratio of the F1 and duration standard deviations (which are 0.2 octaves and 0.4 duration doublings respectively). Thus, the SE equal-likelihood line has a slope of $(0.149/0.500) \cdot (0.2/0.4)^2 = 0.075$ octaves/dur. doublings while the Spanish one has a slope of $(0.055/0.601) \cdot (0.2/0.4)^2 = 0.023$ octaves/dur. doublings. Figure 7.5 shows the midpoint and the equal-likelihood line for the two languages.

66 The conversion from octaves to Hertz is $2^{(\log2(485) - 0.500/2)} = 407.84$ Hz and the conversion from duration doublings to milliseconds is $2^{(\log2(94) - 0.149/2)} = 89.27$ ms.

67 That is, $2^{(\log2(502)-0.601/2)} = 407.61$ Hz and $2^{(\log2(81)-0.055/2)} = 79.47$ ms.

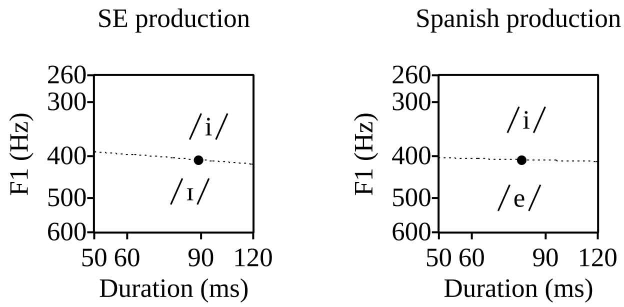

Fig. 7.5. Midpoint and equal-likelihood line (represented as a dotted line) for the SBE and Spanish vowels.

The third and last computation is that of the relative use of the F1 and duration dimensions for signalling the differences between the two vowels in each language, which can be arrived at from the slope of the equal-likelihood line. For SE, the F1 difference (as expressed in standard deviations of F1) is 7 times larger than the duration difference, i.e., (0.500/0.2)/ (0.149/0.4) = 6.71, whereas in the Spanish environment the F1 difference (as expressed in standard deviations of F1) is 23 times larger than the duration difference, i.e., (0.601/0.2)/(0.055/0.4) = 21.9. Table 7.5 shows the production measures for the two languages where the slope and the edge points refer to the same calculation, i.e., to the equal-likelihood line in production.

Table 7.5. Production values for the SE and Spanish production distributions (oct. = octaves, durdou = duration doublings).[68]

Language	Vowels	Midpoint	Equal-likelihood slope	Equal-likelihood edge points	F1/dur use
SE	/i/-/ɪ/	[408 Hz, 89ms]	0.075 oct./durdou.	[391 Hz, 50 ms]; [417 Hz, 120 ms]	7/1
Spanish	/i/-/e/	[408Hz, 80ms]	0.021 oct./durdou.	[404 Hz, 50 ms]; [412 Hz, 120 ms]	22/1

[68] The equal-likelihood points at the edges were obtained following the computations in § 2.2.2 and § 5.2.2.

In previous chapters, it was hypothesized that an optimal listener will have a perception that matches the production distributions of her language environment. As described in Chapter 5, optimal perception implies that three perceptual values will match their production counterparts, viz., the perceptual use of the auditory dimensions, the location of the category boundary, and the slope of this boundary. In an optimal listener, then, the slope of the perceptual boundary will match that of the equal-likelihood line in production. Also, the location of the perceptual category boundary will coincide with the location of the equal-likelihood line in production. Finally, cue reliance in perception will match the use of the cues in production. Figure 7.6 shows the predicted optimal boundary and optimal cue reliance for the SE and Spanish vowels. Note that the token division as well as the computation of cue reliance follow the ones shown in § 3.1.2 and § 5.2.2.

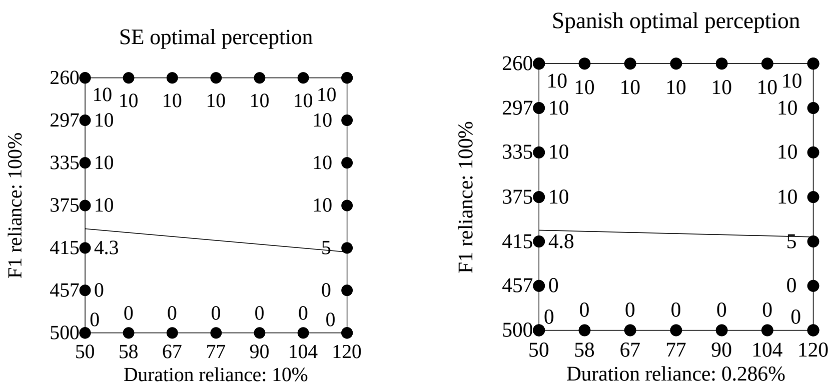

Fig. 7.6. Hypothesized optimal category boundary and optimal cue reliance for SE and Spanish listeners. The numbers inside the figures denote the /i/ responses given by an optimal listener presented with the stimulus point next to the number.

As we can see from the slightly diagonal shape of the SE optimal perceptual boundary, an optimal SE listener will categorize the vowels /i/ and /ɪ/ mainly on the basis of their F1 differences and, to a much smaller extent, by means of their durational differences. This is because SE /i/ is produced on average with a slightly longer vowel duration than /ɪ/. In contrast, an optimal Spanish perceiver will have a horizontal perceptual boundary between Spanish /i/ and /e/ because these vowels are produced with very similar duration values. As a result, she will categorize the vowels on the basis of their F1 values only. In addition, we can observe that optimal SE and Spanish listeners both have a 100% F1 reliance. However, their duration reliances differ (10% versus 3%), a variation that results in different category slopes for the two languages (0.075 versus 0.021). Finally, we can see that the category boundaries of the two languages cross the left and right edges of the square at different locations.

Following the LP model, then, it is posited that a language-specific perception grammar underlies the SE and Spanish optimal perception shown above, i.e., the optimal category boundaries and cue reliance. Such a perception grammar contains constraints ranked in language-specific ways that have been learned during L1 acquisition (cf. § 2.1.3). Thus, the SE and Spanish perception grammars contain cue constraints that optimally place both F1 and vowel duration into the vowel categories of each language. The similarities in the SE and Spanish categorization of F1 values shown above can be described as an almost equal ranking of F1-to-vowel cue constraints. Figures 7.7 and 7.8 show the adult SE and Spanish perception grammars that optimally categorize F1 values into /i/ and /ɪ/ and /i/ and /e/ respectively.

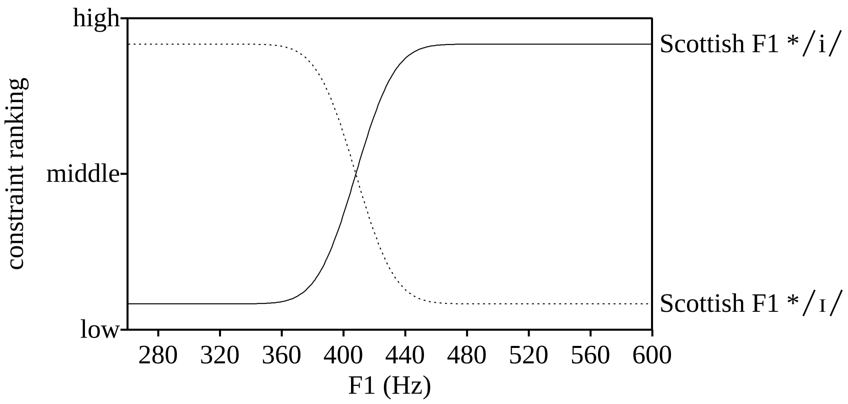

Fig. 7.7. The constraints in the optimal SE perception grammar for the categorization of F1 values (260-600 Hz) as /i/ (solid) or /ɪ/ (dotted).

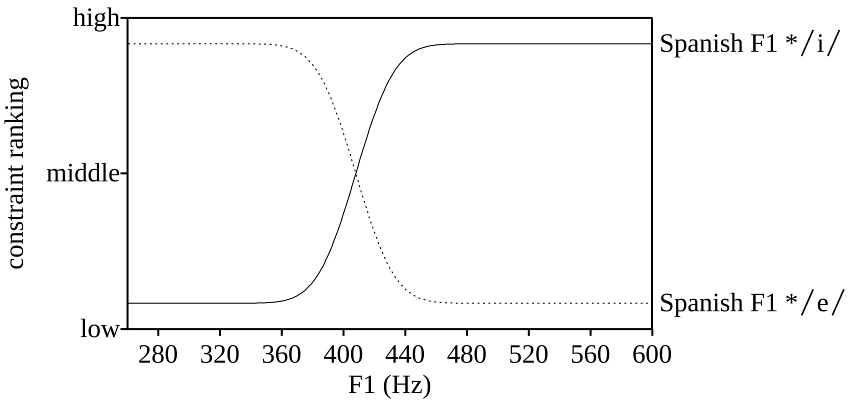

Fig. 7.8. The constraints in the optimal Spanish perception grammar for the categorization of F1 values (260-600 Hz) as /i/ (solid) or /e/ (dotted).

When these figures are compared, we observe that the two perception grammars are the same (or very similar) with respect to the categorization of two vowels on the basis of F1 values. For instance, for an auditory input of 300 Hz, the constraint that says 'do not perceive 300 Hz as /i/' is the lowest ranked in both languages so that both types of listeners will categorize the input as /i/. The cue constraints that refer to /ɪ/ and /e/ in SE and Spanish respectively are also ranked in the same way. That is, whenever an SE listener categorizes an F1 value as /ɪ/, an optimal Spanish listener will categorize the same value as /e/. Thus, although the arbitrary labels for the vowels in the two languages are not the same, they are the targets of the same auditory tokens as a result of the constraint rankings in the respective perception grammars. Crucially, the steep slope of the curves that represent the continuous ranking of 'do not perceive an F1 value *x* as /i/ or /ɪ/' in SE and 'do not perceive an F1 value x as /i/ or /e/' in Spanish conveys the listener's equally strong reliance on the F1 dimension when categorizing vowels. This means that L1 acquisition results in the ranking of F1-to-vowel cue constraints that lead to the determining role of the F1 dimension, and hence to the relatively high ranges of the constraint curves.

Vowel duration, on the other hand, does not play an important role in the categorization of these vowels. This is because unlike the optimal L1 SBE grammar discussed in § 2.2.2 and in § 5.2.1, the SE and Spanish L1 perception grammars never contain auditory-to-phonological cue constraints involving vowel length. In other words, the distributions of vowel duration values in SE and Spanish do not allow for the creation of auditory categories along the vowel duration continuum (cf. § 2.3.3). As a result, auditory constraints involving duration, or one-dimensional cue constraints involving length, or /length/ categories were never introduced in the developing L1 SE and Spanish grammars.

On the other hand, following the L1 acquisition proposal in § 2.3, these listeners do introduce constraints that connect duration to height (/height/). This is because once auditory categories are turned into abstract phonological representations, any auditory dimension can be related to them in one-dimensional cue constraints, such as 'F1-to-/height/', 'F2-to-/height/', 'duration-to-/height/', etc. Thus, both the SE and Spanish developing grammars contain cue constraints such as 'do not perceive a vowel duration value x as /high/', which allows for the use of vowel duration even though this dimension has not been categorized. These constraints are turned into duration-to-vowel cue constraints when adult cue integration is attained, as described in § 2.3.3. Figures 7.9 and 7.10 show the continuous ranking of duration-to-vowel cue constraints in the optimal SE and Spanish adult grammars.

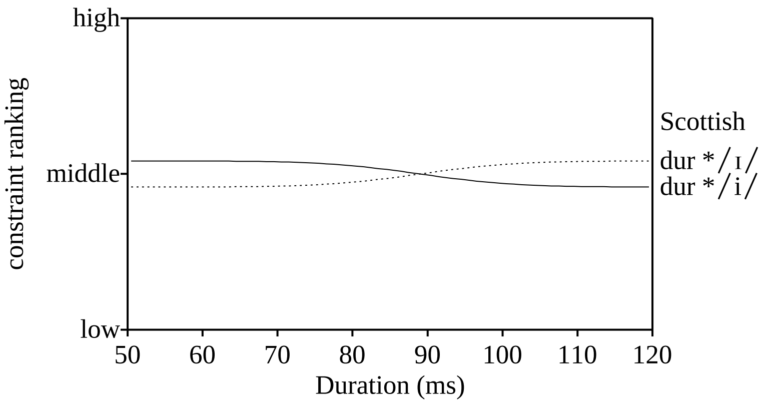

Fig. 7.9. The constraints in the optimal SE perception grammar for the categorization of vowel duration values (50-120 ms) as /i/ (solid) or /ɪ/ (dotted).

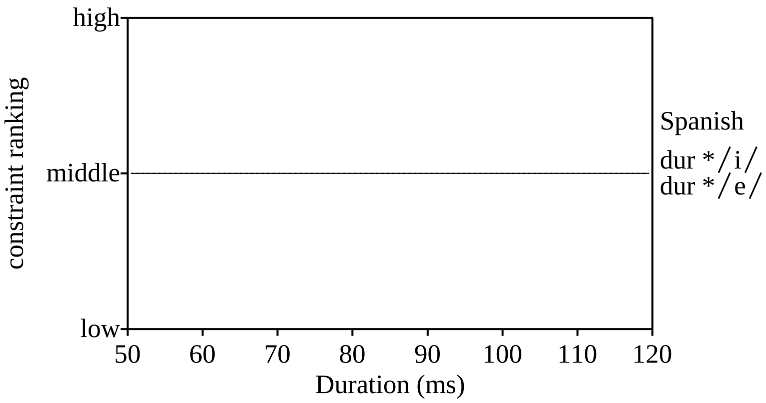

Fig. 7.10. The constraints in the optimal Spanish perception grammar for the categorization of vowel duration values (50-120 ms) as /i/ (solid) or /ɪ/ (dotted).

When comparing Figures 7.9 and 7.10, we observe that these constraint curves have a very different ranking from those of F1-to-vowel cue constraints in Figures 7 and 8. A middle ranking of the duration-to-vowel constraints implies that they hardly contribute to vowel categorization, and this is due to the rather small vowel duration differences in the two languages. This being the case, the figures also show a difference between the ranking in SE and Spanish because although neither language has vowel length categories, the use of vowel duration is different in each language. Thus, the SE optimal listener has a 10% duration reliance whereas the Spanish optimal listener has a very low duration reliance of 3%, which may not be significantly different from zero. This cue reliance difference between the SE and Spanish optimal listeners is conveyed in the grammars by the differential ranking of the duration-to-vowel constraints so that the SE grammar shows a shallow high and low ranking for these constraints whereas the Spanish grammar has them all at middle ranking. This means that, unlike in Spanish, vowel duration contributes slightly to optimal vowel categorization in SE.

The second case discussed in this chapter involves CF and CE /æ/ and /ɛ/, and Figure 7.11 shows the production distributions of these vowels. The average productions for the vowels and their variation are taken from Escudero & Boersma (2004a) and from Escudero (in progress b). Although other auditory dimensions are involved in the production of these vowels in each language, it has been assumed in these studies that F1 and duration are the main reasons for the difference in production in the two languages.

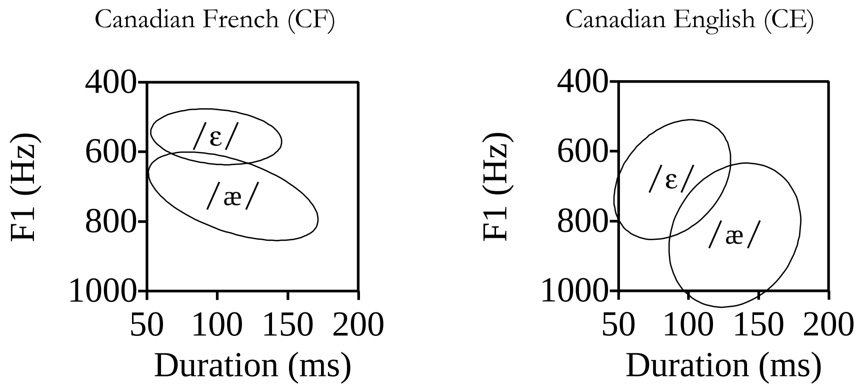

Fig. 7.11. Average productions of CF and CE /æ/ and /ɛ/. The ellipses show one standard deviation from the mean.

As we can see, despite a similarity in the productions of the vowels be in that they occupy similar acoustic regions, it is clear that the use of F1 and duration is rather different because the CE vowels differ in both F1 and duration values while the CF vowels have mainly F1 differences. If we observe the direction of the ellipses, we can see a language-specific correlation of F1 and duration. This is because the CF vowels are produced with a positive correlation between the dimensions, i.e., the longer the vowel, the higher its F1 value, whereas CE has a correlation that results in the use of vowel duration as a cue to vowel identity, i.e., the longer the vowel, the more likely it is to have been intended as /æ/. In addition, intended CE /æ/ tokens usually have a high F1 value together with a long duration while intended CE /ɛ/ tokens have a lower F1 value together with a short duration. For instance, the great majority of vowel tokens produced with an F1 value of 700 Hz are intended as CE /æ/ if they are longer than 110 ms but as CE /ɛ/ if they are shorter than that.

In contrast, although intended CF /æ/ tokens also have higher F1 values than /ɛ/, these two vowels vary freely between short and long. For instance, CF tokens produced with values around 700 Hz are almost always intended as /æ/ and almost never as /ɛ/, regardless of their duration values. In addition, apart from the cross-linguistic difference in the integration of acoustic dimensions, a further difference in F1 distributions can be readily observed. That is, the mean F1 productions of CE /æ/ and /ɛ/ (which are 840 and 681 Hz respectively) are higher than their CF counterparts (which are 728 and 557 Hz respectively).

Following the Optimal Perception hypothesis, the optimal perception of the vowels in each language will exhibit the attested production differences, namely the difference found in the relative use of the acoustic dimensions as well as the difference in F1 distributions, as shown in Figure 7.12. That is, the shape and location of the optimal language-specific perceptions will be exactly the same as the equal-likelihood curves in production, as shown in Figure 7.5. Here we can also see that the same acoustic event [750 Hz, 85 ms.], which is depicted by the diamonds in Figure 7.12, will likely be perceived as /æ/ by an optimal CF listener but as /ɛ/ by an optimal CE listener.

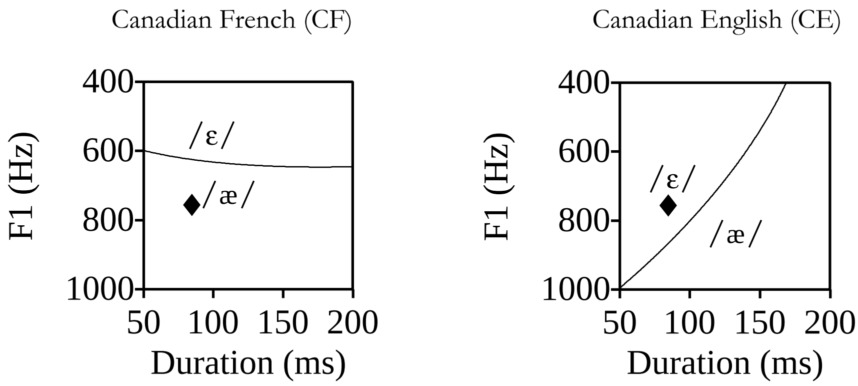

Fig. 7.12. Predicted CF and CE native category boundaries for the vowels /æ/ and /ɛ/. Diamond: [750 Hz, 85 ms].

The proposed perception grammars that underlie the predicted cross-language difference in optimal perception are shown in Tableaux 7.1 and 7.2. Given that the highest ranked constraint in the CF optimal perception grammar is 'do not perceive 750 Hz as /ɛ/', probably because such a value is highly unlikely to have been intended as /ɛ/ in the CF environment, an optimal CF listener will perceive [750 Hz, 85 ms] as /æ/, as shown in Tableau 7.1. This optimal CF grammar contains vowel duration constraints that are ranked in the middle and which therefore do not play an important role in vowel categorization.

[750 Hz, 85 ms]	750 Hz not /ɛ/	85 ms not /æ/	85 ms not /ɛ/	750 Hz not /æ/
/bɛk/	*!		*	
☞ /bæk/		*		*

Tableau 7.1. The constraints and constraint rankings relevant for the categorization of [750 Hz, 85 ms] in the optimal CF perception grammar.

Tableau 7.2 shows that the constraint 'do not perceive 750 Hz as /æ/' is also found in the CE optimal perception grammar but is ranked lower than in the CF grammar. Crucially, this constraint is ranked higher than the one saying that the same value should not be perceived as /ɛ/ because such an F1 value is unlikely to have been intended as /ɛ/ in the CE environment. Interestingly enough, the optimal CE grammar has lower and higher rankings for constraints involving vowel duration because this dimension plays an important role in identifying the perceived vowel so that the constraint 'do not perceive 85 ms as /æ/' is ranked highest. Therefore, a token with a short vowel duration such as [750 Hz, 85 ms] is categorized as CE /ɛ/, irrespective of its F1 value.

[750 Hz, 85 ms]	85 ms not /æ/	750 Hz not /æ/	750 Hz not /ɛ/	85 ms not /ɛ/
☞ /bɛk/			*	*
/bæk/	*!	*		

Tableau 7.2. The constraints and constraint rankings relevant for the categorization of [750 Hz, 85 ms] in the optimal CE perception grammar.

7.2.2 Ingredient 2: predicting cross-language perception and initial L2 perception

A similar scenario will first manifest itself as a cross-language 'two-category assimilation' (cf. Best 1995) of foreign sounds to already existing native sound categories. This is because of the phonemic equation that exists between the L1 and L2 categories, and also because of the great similarity between these two perception grammars. Recall that the L2LP's Full Copying hypothesis proposes that cross-language categorization is the initial state of L2 perception. This entails that that both foreign language listeners and beginning L2 learners will automatically and unconsciously reuse their L1 categories and perception grammars when coping with a similar scenario. Therefore, it is predicted that cross-language perception will determine how beginning learners initially perform in their L2.

In general, cross-language perception is governed by the listeners' optimal L1 perception. In a similar scenario, where two non-native sounds are mapped onto two L1 sounds, an L2 initial state that matches cross-language perception will turn out to be highly successful because it will lead to a near optimal L2 perception. That is, the number of reused L1 categories is predicted to match the number of categories perceived by an optimal listener of the target L2. This is because the values of the L2 tokens allow the copied L1 perception grammar to output the optimal number of perceived categories in the new environment. This initial copying of the L1 perception grammar and categories, which leads to near-optimal cross-language perception and beginning L2 sound perception, provides the linguistic mechanism underlying Best's (1995) two-category assimilation and Flege's (1995) equivalence classification hypotheses, as was mentioned in Chapter 4.

To illustrate the cross-language and L2 categorization in a similar scenario, two cases will be considered, namely the Spanish perception of SE /i/ and /ɪ/ and the CE perception of CF /æ/ and /ɛ/. For the first case, the left side of Figure 13 shows that given the SE acoustic values (as found in Escudero & Boersma 2003) and the Spanish optimal perception, it is predicted that Spanish listeners and beginning Spanish learners will reuse their vowels /i/ and /e/ to categorize the SE vowels /i/ and /ɪ/. In the second case, the CE listeners will use the phonologically equivalent CE /æ/ and /ɛ/ to categorize the CF vowels, as shown on the right side of Figure 7.13.

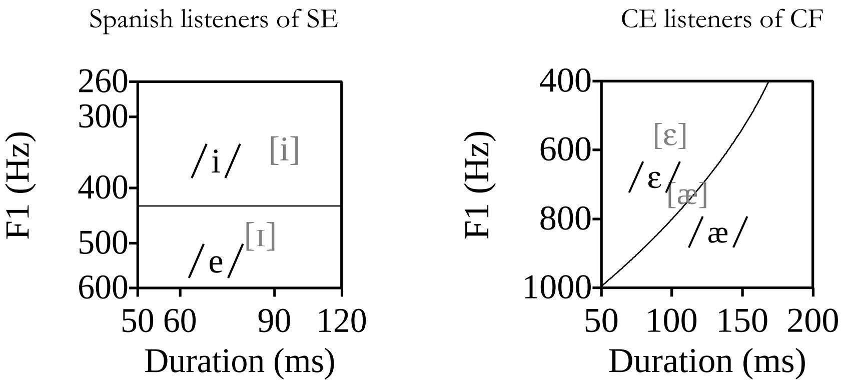

Fig. 7.13. Cross-language perception and predicted initial stage for Spanish listeners of SE and CE listeners of CF. In grey: Average target L2 productions. In black: Average L1 productions. Dotted line: L1 category boundary.

7.2.3 Ingredient 3: predicting the L2 learning task

Given the predicted initial state, learners will have problems with optimally categorizing all vowel tokens of the target L2. In general, it is predicted that beginning learners will be able to differentiate between similar L2 sounds because the number of reused L1 categories will match the number of target sounds to be learned. However, there will be a degree of mismatch between the L1 optimal perception grammar and that of the target language because of the differences in the production of the two categories involved. This degree of perceptual mismatch will determine the level of accuracy in the initial L2 categorization of a learner who faces a similar scenario. In addition, the degree of mismatch between the perception grammars will constitute the L2 learning task in this scenario.

A non-optimal L2 perception grammar is manifested as a boundary mismatch in the learner's categorization of L2 sounds (cf. Escudero & Boersma 2004b). In other words, depending on the cross-language mismatch between perception grammars, it is predicted that L2 beginning learners will categorize target tokens in a non-optimal manner because the copy of their L1 grammar will fail to handle the produced token distributions optimally. Thus, we find that there is a cross-language perceptual mismatch between the Spanish and the SE equal-likelihood lines, as depicted by the grey region in Figure 7.14. This means that a beginning Spanish learner of SE will categorize the SE tokens that fall in the grey region solely as

Spanish /i/, irrespective of their duration values. In contrast, the optimal SE listener will perceive two different vowels in the same region.

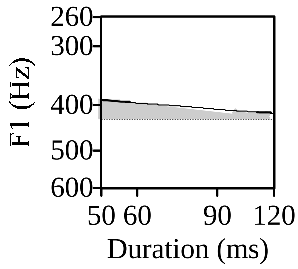

Fig. 7.14. Region of cross-language perceptual mismatch (in grey) for Spanish learners of SE. Horizontal line: L2 initial state. Diagonal line: target L2 boundary.

Because of this rather small cross-language difference in production, it is predicted that the degree of perceptual mismatch will also be rather small. In fact, Escudero & Boersma's (2004b) simulation of the perception of SE tokens resulted in a categorization that was correct in 78.7% of the cases for an optimal Spanish listener and 81.5% for an optimal SE listener.[69] Note that these percentages were computed assuming Bradlow's (1995) values for these Spanish vowels. However, whether one assumes the Spanish values given in § 7.2.1 or the ones in Escudero & Boersma (2004b), an optimal Spanish listener will make only a few errors when categorizing SE /i/ and /ɪ/ as Spanish /i/ and /e/, their two closest counterparts. Therefore, it can be predicted that she will be a near-optimal, or even optimal, listener of the SE vowels because the Spanish grammar almost matches the SE grammar. Crucially, the exact level of correspondence between Spanish and SE will depend on a listener's specific production environment, which could exhibit fine-grained dialectal differences.

The cross-language mismatch between the CE and CF production distributions of /æ/ and /ɛ/ is much larger than that of SE and Spanish, as can be computed

69 Recall that optimal perception matches the probability of occurrence in the specific production environment. Therefore, it should not be surprising that an optimal perceiver does not have 100% correct perception given that most environments exhibit overlapping distributions. As described in § 2.1.2, the overlap in production determines the probability of a token being perceived as more than one vowel. In the SE case, 18.5% of all the possible F1-duration pairs have values that fall in the area of overlap. Thus, an optimal SE listener has an 81.5% probability of having a correct perception, i.e. of perceiving a vowel and subsequently finding it in her lexicon.

from the tokens produced in the grey regions found in Figures 7.14 and 7.15. That is, given that 50% of the CF /æ/ tokens are produced in the region where only CE /ɛ/ tokens are found, we can safely assume that CE listeners perceive half of the CF /æ/ tokens as CE /ɛ/. Moreover, it is important to notice that the CF productions overlap, as was shown in Figure 7.11, thus entailing that CF listeners also have some errors. However, given the fact that only a few CF /ɛ/ tokens are produced in the grey region, their percentage of correct categorization is expected to be higher than 90%. Thus, assuming the L2LP's proposal, CE listeners as well as CE beginning learners of CF are predicted to have at least 40% more errors than CF listeners when categorizing the CF /æ/ tokens that fall in the region of acoustic/auditory mismatch.

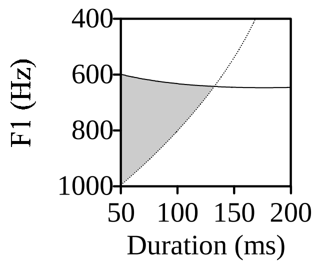

Fig. 7.15. Region of cross-language perceptual mismatch (in grey) for CE learners of CF. Dashed line: L2 initial state. Horizontal line: Target L2 boundary.

As we can see, the learning task for beginning CE learners of CF will involve a two-dimensional boundary shift because they will have to learn that duration differences do not signal vowel identity, which should enable them to categorize short tokens as their L2 /æ/. In addition, they will have to adjust their F1 boundaries so that they can categorize vowel tokens produced with F1 values between 600 and 780 Hz as their L2 /ɛ/ category. In sum, they will have to turn their diagonal line into the horizontal line shown in Figure 7.15.

Despite the difference in degree of perceptual mismatch, both the Spanish-to-SE and CE-to-CF cases will have implications for word recognition because these learners will sometimes access words in their L2 lexicon that were not intended by the target language speakers. This situation will arise as a result of an inaccurate perception that will lead them to access an incorrect lexical item. The question is how learners can resolve the perceptual mismatch that causes a lexical access problem. It is proposed that given a certain cross-language perceptual mismatch, the learning task in this L2 perception scenario will be to adjust or shift the copied L1 perceptual category boundaries and perceptual cue weightings. In the next section, I will spell out the specific proposal and predictions for L2 development in a similar sound perception scenario.

7.2.4 Ingredient 4: predicting L2 development

It is predicted that L2 perception will develop following the same learning mechanisms observed in babies learning to perceive their L1, viz., category formation and perceptual boundary shifting. For a similar scenario, any learner will need to recategorize the new L2 tokens by adjusting her perception grammar, which is initially a copy of her L1 grammar, in order to optimally cope with the new production environment. Such an L2 perception adjustment will manifest itself as a shift in the category boundary location for one-dimensional or multi-dimensional boundaries. As was described in § 2.1.3, children undergo perceptual category boundary shifts and cue weighting adjustments in their L1 perception development (cf. Nittrouer 1996, Gerrits 2001). According to the LP model, these L1 cue weighting and boundary location shifts can be formally modelled by Stochastic Optimality Theory (cf. Boersma 1998) and the GLA (cf. Boersma & Hayes 2001), which is the learning algorithm associated with it. Recall that this type of learning is guided by the lexicon in that semantic knowledge supervises perception learning. Therefore, if a child accesses a lexical item whose phonological form does not correspond to the perceived form, perceptual learning will occur.

The L2LP model postulates that this lexicon-driven L1-like learning will also be observed in similar sound perception. As an example, Spanish learners need to shift their initial L2 boundary location slightly in order for F1 to match that of the optimal SE perceiver. This occurs through lexicon-driven constraint reranking in the L2 perception grammar. Thus, if an L2 learner of SE perceives /ʃip/ and later recognizes a word |ʃip| when the speaker intended |ʃɪp|, her recognition system will tell her that she has made a perception error. This is because, given the semantic context, the perceived token cannot refer to the concept attached to |ʃip| but only to the one attached to |ʃep|. As a result, the learner realizes that she should have perceived /e/ and not /i/. Thus, this recognition-perception mismatch will result in a change in the learner's L2 perception grammar, as shown in Tableau 7.3 where the arrows depict the GLA constraint reranking procedure.

[400 Hz, 50 ms]	400 Hz not /e/	50 ms not /e/	50 ms not /i/	400 Hz not /i/
☞ /i/			← *	← *
√ /e/	*! →	* →		

Tableau 7.3. Predicted lexicon-driven constraint reranking in the L2 perception of Spanish learners of SE.

Therefore, in order to solve the problem of perceiving a few SE tokens as Spanish /i/ instead of /e/, these learners need to shift their perceptual category boundary by 30 Hz, i.e., from 430 Hz to 400 Hz, and they also need to start using durational differences to categorize vowels. However, the perceptual mismatch seems to be rather small so that learning will occur rapidly, provided that tokens in the mismatch region do occur in the L2 environment. This lexicon-driven L2 acquisition is manifested as a shift of perceptual category boundaries or as a change of perceptual cue-weighting in order to gradually reduce the mismatch between the perceived distributions of L2 vowels and those of the L2 production environment, as shown in Figure 7.16.

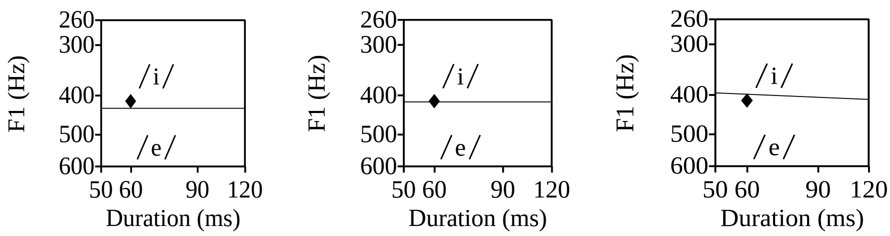

Fig. 7.16. Predicted category boundary shift for Spanish learners of SE /i/ and /ɪ/. Between slashes: Copied Spanish vowels. Line: L2 category boundary development.

Here we see the predicted developmental path of a Spanish learner of SE. Assuming optimal perception and L1 copying, the boundary shift for F1 will be gradual but fast, given that the distance of the shift is rather small (i.e., Δ = 30 Hz). The learning steps in the figure compare well with the ones shown in the simulations of Escudero & Boersma (2004b) where after 10 virtual months, the simulated learner was found to reach the second level of proficiency, and where after 90 months she attained the third level. In the simulations, after only 10 months of virtual input, she had already approximated the optimal boundary by about 20 Hz and had 81.2 % of correct categorizations, which is close to optimal SE perception. The same simulations revealed a small reliance on duration after 90 virtual months, which is also shown in Figure 7.16. However, duration may not be a reliable cue for vowel categorization in SE given that Escudero & Boersma (2003: 72) suggested that SE speakers do not use this dimension for signalling differences between vowels but rather to distinguish between the number of syllables or the voicing value of the final consonant.

As mentioned above, L2 development in the case of the CE learner implies a rather large and multi-dimensional boundary shift because she will start with a grammar that categorizes vowels by integrating duration and F1 values. Thus, CE learners will have to move their F1 boundary by 100 Hz and their vowel duration boundary by at least 60 Hz in order to perceive as their L2 category /æ/ tokens that fall in the grey region if they wish to reduce the predicted 40% error percentage. It is proposed that two-dimensional category boundary shifts are also the result of GLA lexicon-driven learning (cf. § 2.3.3). For these learners, this type of L1-like learning will demote the constraints against perceiving certain F1 and duration values as the L2 /æ/ category, as shown in Tableau 7.4 where the arrows depict the lexicon-driven constraint reranking.

[750 Hz, 85 ms]	85 ms not /æ/	750 Hz not /æ/	750 Hz not /ɛ/	85 ms not /ɛ/
☞ /bɛk/			← *	← *
Target /bæk/	*! →	* →		

Tableau 7.4. Predicted lexicon-driven constraint reranking for CE learners of CF.

This will lead to the duration and boundary shift required to improve their L2 categorization. Figure 7.17 shows the predicted gradual boundary shift, which involves a two-dimensional shift of perceptual category boundaries that leads to a change of perceptual cue integration in the L2 perception

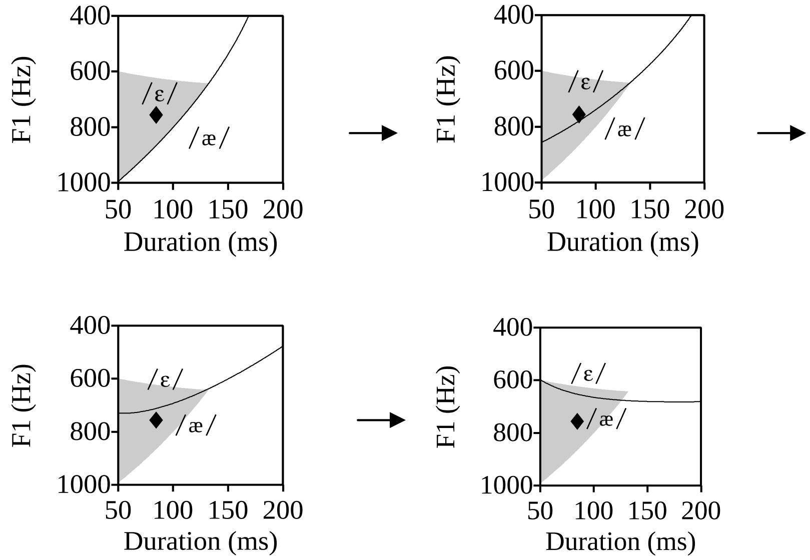

Fig. 7.17. Predicted category boundary shift for CE learners of CF /æ/ and /ɛ/.

At this point, we can summarize the predicted developmental path for any case of the similar scenario as follows. In the first stage, the learner is faced with L2 categories that are phonemically equated to the same number of L2 categories with which they share similar production distributions. Because of Full Copying, she starts out with L1 categories and their categorization constraint rankings. This starting point guarantees that she will have the optimal number of perceived categories and a near-optimal perception, the latter occurring as a result of a slight difference in the constraint rankings of the L1 and L2 perception grammars due to discrepancies in production distributions. In the second stage, the learner's recognition grammar notices a semantic mismatch between the word accessed through perception and the speaker's intended word, and this leads to GLA lexicon-driven learning.[70] Thus, the GLA gradually reranks the constraints in the perception grammar in order to reduce the probability of recognition-perception mismatches occurring again.

---

[70] It should be mentioned that the existence of minimal pairs in the lexicon is not a requirement for lexicon-guided learning to occur because perceiving a word intended to be *shift* as the non-word *sheeft* also results in the semantic mismatch that will lead to learning, as was proposed by Escudero & Boersma (2004b).

With respect to the speed with which the L2 constraint rerankings will take place, it is assumed that development is constrained by the properties of the GLA which, in this case, will implement gradual learning depending on the frequency of perception-recognition mismatches, which in turn depends on the size of the cross-language difference. For instance, in performing the predicted L2 boundary shift of SE /i/ and /ɪ/ by Spanish speakers, the GLA can reach the optimal percentage of correct categorizations in 10 virtual months, though cue reliance may not be entirely optimal before 90 months of virtual L2 exposure. However, such a simulation is heavily influenced by the production distributions, which are likely to vary, as illustrated by the case of the Spanish environment in § 7.2.1 and the one reported in Bradlow (1995). Thus, in general, the speed and end result of development in a similar scenario will depend on how L1 and L2 production distributions as well as their optimal perceptions compare to one another.

7.2.5 Ingredient 5: predicting the L2 end state

Following the L2LP separate perception grammars hypothesis, it is predicted that the L1 and L2 perceptions of L2 learners will be optimal in that they will match the production distributions of their respective environments. As discussed in § 3.5, the possible external limitations of L2 development are also acknowledged to exist in the L2LP model, i.e., the type of L2 input and the decreased cognitive plasticity which are characteristic of adult learners.

With respect to the role of the L2 input in determining the end state in a similar scenario, the simulation of the development of a virtual Spanish learner of SE described in Escudero & Boersma (2004b: 577) assumed that L2 learners can be exposed to rich, L1-like input environments. However, such a rich input is rarely attested in L2 development. Furthermore, very few learners have access to the enhancement of acoustic properties in their L2 environment, whereas such enhancement is a core feature of infant-directed speech or motherese (cf. § 3.5.3). Thus, complete optimal perception may not be guaranteed unless the frequency of the input can be enhanced in the same way motherese enhances the crucial properties of the sounds of a language during L1 acquisition. On the other hand, it may be the case that in the similar sound perception scenario, the ambient input is enough to trigger ultimate L2 boundary shift given that the required shift is rather small.

Another type of L2 learning limitation resides in cognitive plasticity. Recall that the L2 model incorporates this property through the size of the learning step that the GLA takes when changing the ranking of the constraints in a perception grammar. That is, the higher the plasticity, the larger the GLA learning step will be. This means that learners with a high level of cognitive plasticity can learn faster than those with a lower plasticity. In the case of adult L2 learners faced with a similar perception scenario, the simulation mentioned above assumed that the adult L2 virtual learner had a low plasticity which was kept constant throughout her L2 exposure of 800 virtual months. With this plasticity, she seemed to reach optimal perception faster than an L1 listener because her initial L1 grammar was close to the optimal SE grammar. It is worth mentioning that, even after only ten months of virtual exposure to the L2, the learner's category boundary shifted close to the optimal location. Therefore, a low plasticity can slow down the L2 learning process when compared to that of an L1 learner, but in a similar scenario, the copied L1 grammar guarantees that the L2 learner has a head start that even allows her to outperform an L1 learner, at least with respect to one-dimensional boundary shifts.

With respect to exhibiting optimal perception in the L1 and L2, the L2LP's interpretation of the relation between the learner's two languages supports optimal L1 and L2 end states (cf. § 3.5). Recall that Flege's SLM, which was described in Chapter 4, explicitly assumes that L1 and L2 categories are stored in the same phonological space, and that therefore L2 boundary shifts will inevitably affect L1 perception. This implies, for instance, that if two categories have different boundaries, they must be two separate categories. By contrast, the L2LP model proposes that sound categories or phonological representations may be shared in the L1 and L2 lexicons so that, for example, SE /ɪ/ might continue to be equated with Spanish /e/. However, vowel categorization will differ in the two languages because L2 learners possess two separate perception grammars for each language rather than a single system as in Flege's proposal.

For a similar perception scenario, this means that learners should be able to adjust their L2 category boundaries through lexicon-guided constraint reranking without affecting their already optimal L1 category boundaries. That is, it is predicted that L1 perception will remain stable while L2 perception develops, so that an advanced learner faced with a similar scenario will exhibit L1 and L2 optimal perception. As mentioned in § 3.5.3, the hypothesis of separate grammars can be empirically tested as a difference in perceptual behaviour, depending on which language setting a learner finds herself in. In the similar scenario, this language setting effect is particularly important because we can test whether learners reuse the same number of categories while having different perceptions in their L1 and L2 settings. In addition, following the L2LP language mode activation hypothesis (cf. § 3.5), L2 learners and bilinguals can have different degrees of language activation depending on the situation they find themselves in. Figure 7.18 shows the three possible types of perception that result from different degrees in the activation of two languages.

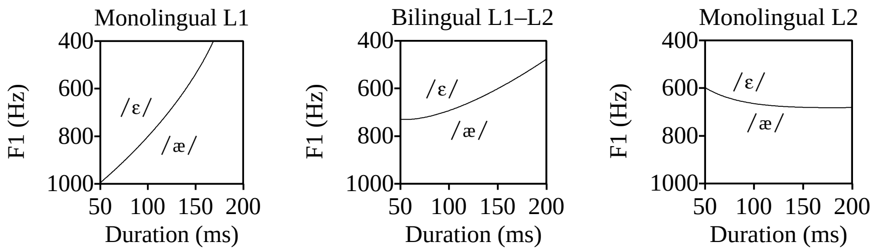

Fig. 7.18. Predicted different types of perception in advanced CE learners of CF.

In the next section, each of the L2LP model's ingredients will be put to the test with the aid of empirical data. The evidence will show that these data mostly confirm the model's predictions. Table 7.6 summarizes the predictions for each of the five ingredients involved in L2 sound perception, namely the optimal perception in the two languages, the three logical states in language acquisition, and the L2 learning task. How the model describes the sound perception in these components is also shown.

Table 7.6. Summary of the L2LP predictions for the new scenario.

L2LP ingredients	Predictions for similar	Description
Optimal L1 & L2	SE and Spanish listeners will exhibit optimal L1 perception	Optimal category boundaries and optimal cue reliance
Initial state	Beginning Spanish and CE learners will be equal to monolingual Spanish and CE listeners	Both types of listeners will exhibit L1 category boundaries, L1 cue reliance, and L1 categories
Learning task	Boundary shift for CE learners, no action for Spanish learners	
Development	Lexicon-driven learning	Category boundary shifts
End state	Spanish & CE learners will attain optimal L2 perception and will maintain their optimal L1 perception	Optimal category boundaries and cue reliance for L1 and L2 in the respective monolingual modes

7.3 Empirical evidence A: Spanish learners of Scottish English (SE)

This section reports on the Spanish perceptual learning of SE /i/ and /ɪ/. In § 7.3.1 to § 7.3.3, a series of perception experiments is described in accordance with the model's sequential ingredients for investigating L2 perception. Then, in § 7.3.4, I present a discussion of the experimental methodology and findings for this case of L2 similar perception.

7.3.1 Scottish English (SE) and Spanish perception

This subsection examines the monolingual perception of SE and Spanish listeners when they hear the same synthetic stimuli. The results of the studies reported here directly bear on the Optimal Perception Hypothesis which predicts a small difference in the perception of SE and Spanish listeners (cf. § 6.2.1). Escudero (2001) tested the perception of /i/ and /ɪ/ in 20 (10 female, 10 male) SE listeners who had lived in Edinburgh for most of their lives, who were between 23 and 35 years of age, and who had little knowledge of other languages. For Spanish, Escudero (in progress a) tested the perception of 32 (16 male, 16 female) monolingual Peruvian Spanish listeners who had lived in Lima for most of their lives, who were between 18 and 28 years of age, and who had little or no knowledge of English. The SE listeners were tested at the University of Edinburgh while the Spanish listeners were tested in Lima at the Pontificia Universidad Católica del Perú. Since the subjects were tested in their country of origin, it was thought that this would provide a good measure of monolingual perception, as suggested by Beddor & Gotfried (1995).

The SE listeners were presented with the 37 synthetic stimuli shown in Figure 7.19. These are the same tokens with which SBE perception was measured (cf. § 5.3.1). In the square, the token with the lowest F1 value shown in the top right hand corner represents the most /i/-like stimulus with a typical SE value of 344 Hz, while the token with the highest F1 value and the shortest duration shown in the bottom left hand corner represents the most /ɪ/-like stimulus with a typical SE value of 480 Hz. Although SE /i/ and /ɪ/ do not exhibit large durational differences, a larger than usual difference between the duration values of the end points was used in order to enhance the listeners' possibility of perceptually relying on vowel duration. The Peruvian Spanish listeners were also presented with synthetic stimuli but instead heard 49 tokens, i.e., the tokens shown in the figure plus the missing 12 tokens that together form a 7 x 7 matrix.

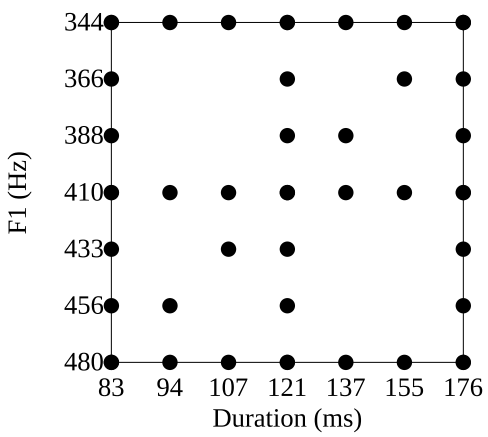

Fig. 7.19. Thirty-seven synthetic stimuli used to test the SE and Spanish perception.

Every SE listener was presented with the 37 stimuli 10 times in different randomized orders. The Spanish listeners heard the 49 stimuli only three times because of the larger number of stimuli and the larger number of subjects. The listeners were all asked to categorize the stimuli they heard as either of two vowel categories, i.e., /i/ or /ɪ/ for the SE listeners, and /i/ or /e/ for the Spanish listeners. If the Optimal Perception Hypothesis is correct, the 20 SE listeners and the 32 Spanish listeners would have been expected to exhibit category boundaries and a cue reliance equal to the ones that were hypothesized to be optimal (cf. §7.2.1).

In order to compute the SE cue reliance and boundary slope as continuous representations, the 7 x 7 matrix was completed by interpolating the values of the 12 missing cells from neighbouring values. Figure 7.20 shows the average results for the SE and Spanish listeners. Recall that the boundary is measured as the 50-50% categorization, and that cue reliance is computed following the procedure described in § 2.2.2 and § 5.2.1.

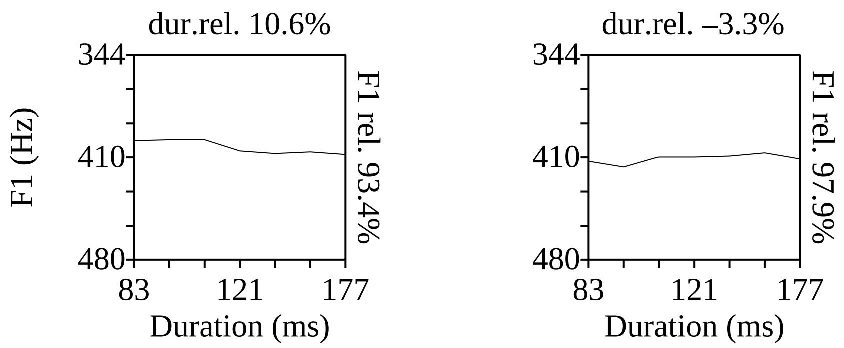

Fig. 7.20. Average SE categorization (left) and monolingual Spanish categorization (right) of the synthetic stimuli (left).

When the average boundary lines in the figure are compared with the SE and Spanish optimal perception, it seems evident that listeners of these languages were optimal perceivers. This is because they exhibited a category boundary at the expected language-specific optimal location and also because they had the expected optimal cue reliance, which was to rely mainly on F1 and very little on duration in both languages. We can also observe that, as expected, the duration reliance of SE listeners was higher than that of their Spanish counterparts.

7.3.2 Cross-language and initial L2 perception

What will be reported here are the results of a study by Escudero (in progress a) to test whether the size of the cross-language region of perceptual dissimilarity for Spanish listeners of SE has the same characteristics as those that were proposed in § 7.2.2. This is because this study considered a cross-language perception task in which Spanish monolingual listeners were presented with natural tokens of SE. As far as can be determined, the perception of that variety of English by monolingual Spanish listeners had never been documented before. Escudero (in progress a) also tested the cross-language perception of 64 (32 female, 32 male) monolingual Peruvian Spanish listeners which included the 32 subjects presented with the synthetic stimuli described in the previous section. The 32 additional listeners had the same characteristics and were tested at the same location.

To measure cross-language perception, the Spanish listeners were presented with 24 natural tokens of the SE vowels /i/ and /ɪ/ drawn from the corpus described in Escudero & Boersma (2003). Figure 7.21 shows how these natural SE tokens compare to the Spanish vowel averages reported in Cervera et al. (2001).

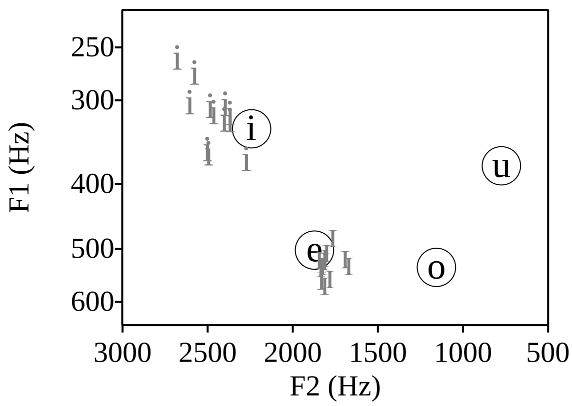

Fig. 7.21. Spanish average productions and the SE tokens. In grey: SE tokens; circled: Spanish averages in Boersma & Escudero (2004).

The listeners were presented with two repetitions of the SE tokens, half of them intended as /i/ and the other half as /ɪ/, as produced by a native speaker of Standard SE. They were asked to click on the Spanish vowel they thought they had heard, and for this purpose they were provided with a computer screen that showed the five Spanish vowel monophthongs <i, e, a, u, o>. The listeners had a practice session with ten tokens before performing the experiment. Iit was expected that if the Spanish listeners' production environment matched that of SE, most SE /i/ tokens would be categorized as Spanish /i/ and most SE /ɪ/ tokens as Spanish /e/. Figure 7.22 shows that the experiment prediction was borne out.

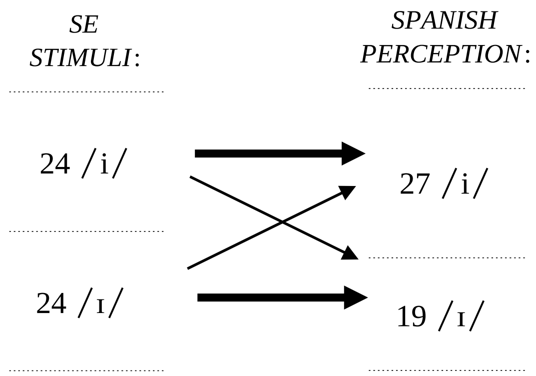

Fig. 7.22. Cross-language Spanish categorization of the natural tokens of SE /i/ and /ɪ/.

As we can see, most tokens of each SE vowel were categorized as their Spanish counterpart. However, there is a slight mismatch in the number of tokens assigned to each Spanish category. Table 7.7 reveals that a greater mismatch is observed for SE /ɪ/ which is mainly categorized as Spanish /e/ but also as Spanish /i/ in a few cases. This results in the token boundary mismatch observed in the figure which, on average, is three to four tokens lower than the optimal SE boundary. In other words, 10 % of the tokens were misclassified. However, the standard deviations shown in the table reveal a large variation in the listeners, with some of them optimally perceiving the SE vowels as two different L1 vowels.

Table 7.7. Token averages and standard deviations (s.d.) for the cross-language categorization of the SE 24 /i/ and 24 /ɪ/ tokens.

	Sp. /i/ Mean	s.d.	Sp. /e/ Mean	s.d.	Tokens
SE /i/	22.34	2.76	1.22	2.18	24
SE /ɪ/	5.23	5.06	17.36	5.15	24
Total tokens	27.57/48		18.58/48		48

7.3.3 L2 development in Spanish learners of Scottish English (SE)

The Spanish learners of English reported in Escudero (2001) turned out to have different target L2 dialects, as mentioned in § 5.3.3. In Escudero & Boersma's (2004b) reanalysis of these data, it was shown that 6 of the 30 Spanish learners clearly had SE as their target L2 dialect, and that another 6 were absolute beginning learners of English, as determined by a language background questionnaire. Only these 12 learners will be considered in this section. They were thus presented with the 37 synthetic stimuli shown in Figure 19, and they were subjected to exactly the same experimental procedure as that of the SE listeners whose performance was analyzed in § 7.3.1. That is, they had to classify the synthetic tokens, which were presented 10 times each, as either English /i/ or /ɪ/. If the L2LP model is correct, the six beginning learners of English would have needed to perform like monolingual Spanish listeners in order for the perceptual boundary line between /i/ and /ɪ/ to resemble the equal-likelihood line between Spanish /i/ and /e/, thereby differing only slightly from the optimal SE perception. The L2LP model also predicts that the six non-beginning learners with a SE target language should have matched the optimal SE perception. Figures 7.23 and 7.24 show the perception of the beginning and the non-beginning learners respectively.

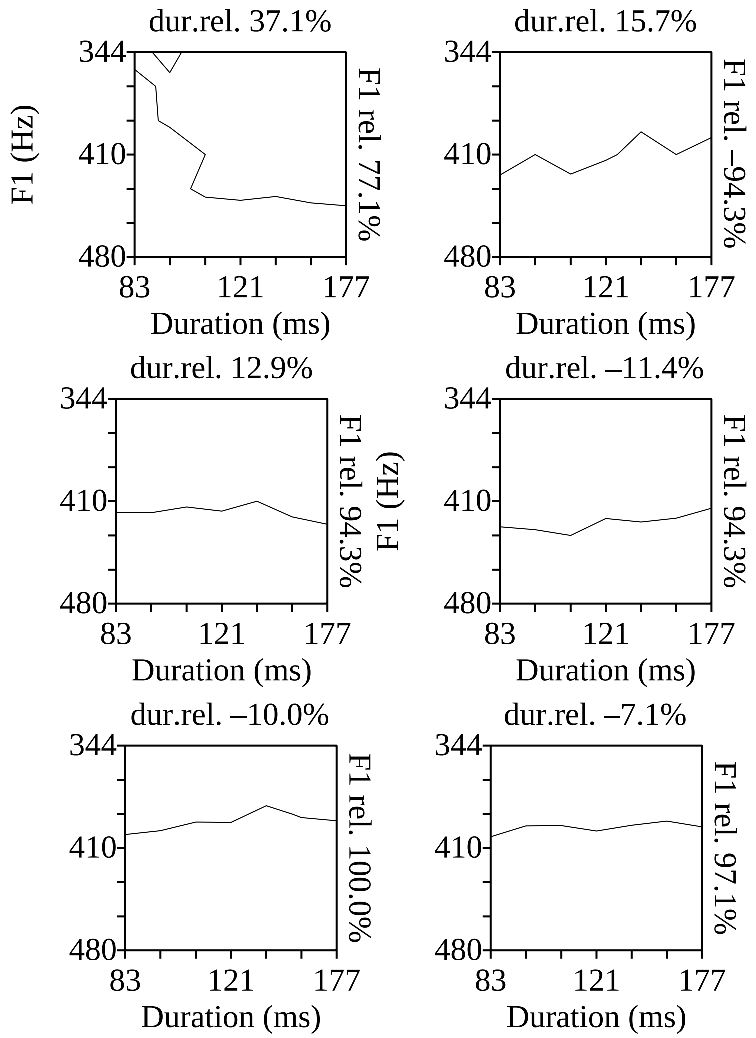

Fig. 7.23. Category boundaries and cue reliances for the 6 Spanish beginning learners.

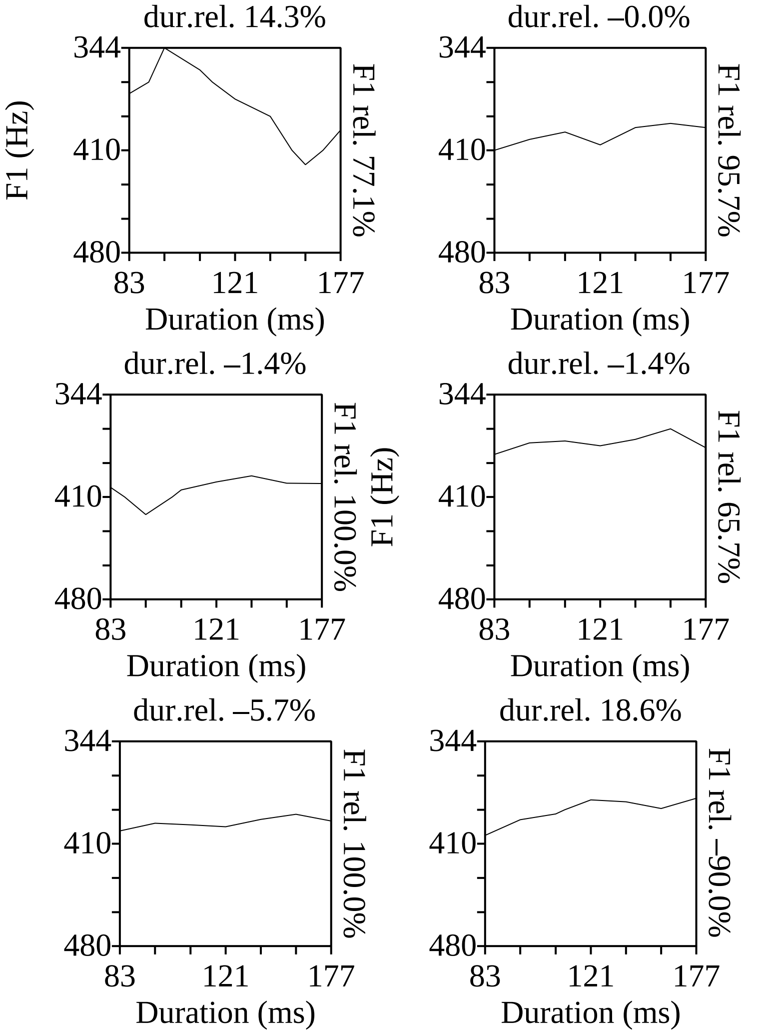

Fig. 7.24. Category boundaries and cue reliance for the 6 non-beginning learners with SE (or a similar dialect) as their target L2.

As we can see, two of the Spanish learners had F1 boundary locations that matched optimal Spanish perception and two of them had boundaries that matched optimal SE perception. This means that their Spanish production environment could have been different from the one used to predict Spanish optimal perception. As for the advanced learners, they all exhibited an F1 boundary line that matched the SE optimal perception. However, it cannot be concluded whether this pattern represents L2 development or Spanish optimal perception, given that a similar pattern was observed in two of the beginners.

It is worth mentioning that almost all learners relied on the F1 cue only, and this reveals a further difference between them and the SE optimal perception. However, if we consider that 16 out of 20 SE listeners did not rely on duration at all, this would imply that they were typically non-optimal perceivers. Escudero & Boersma (2003) suggested that the use of duration in SE seems to depend on differences in final consonant voicing , e.g., /lid/ versus /lɪd/, or on the number of syllables, e.g., /lid/ versus /filliŋ/, and not on differences in vowel identity s. Since the experiment had response choices (i.e., sheep and ship) that did not differ in the final consonant or in their number of syllables, the optimal listener would not be expected to use vowel duration when categorizing the two vowels. What this means is that the 16 SE listeners who did not use duration were optimal perceivers, from which it can be concluded that Spanish learners may not need to rely on duration to become optimal perceivers of SE /i/ and /ɪ/.

7.3.4 Discussion

In this study, it was found that SE and Spanish listeners were indeed optimal listeners. However, the data suggest that a more accurate sampling of specific Spanish and SE vowel production is needed. First, the actual production environment of the Spanish listeners who took part in the perception experiment should be measured to adequately predict cross-language and initial L2 perception which equals optimal L1 perception. In terms of the latter, it is important to mention that the beginning learners had a perception that resembled the Spanish monolingual perception.

With respect to L2 development, no clear sign of change between the beginning and non-beginning learners was found. This could mean that, depending on their specific production distributions, many Spanish learners may already be optimal listeners of SE or that the small difference between environments does not trigger a real task for them. Therefore, it may be the case that no learning needs to occur because the Spanish copied L1 categorization is already optimal for the L2 environment. If the learners' environment does not match that of SE, they will need to take action. However, the findings reported in this section do not clearly show whether there is a learning problem in this case, nor do they show whether learners develop to solve such a problem. Essentially, the small number and great individual variability of the L2 learners do not allow strong conclusions about the initial state and development in this case of L2 similar perception. However, the analysis of this small sample of subjects together with the large number of monolingual Spanish listeners strongly suggests that Spanish learners can, from the beginning, be optimal SE listeners. That is, this case may not involve similar but 'equal' L1-L2 perception, which is an important result in itself.

7.4 Empirical evidence B: Canadian English (CE) learners of Canadian French (CF)

The second case to be presented here is the perceptual learning of CF /æ/ and /ɛ/ by CE learners. Importantly, the methodology that was used to test this case of a similar scenario aimed at directly assessing the L2LP model's ingredients for explaining L2 perception. That is, unlike what was done in the previous section, the series of experiments that were conducted to investigate this case followed the same sequence as the model's ingredients. This means that the production environments were first sampled to determine the optimal perception, then monolingual and cross-language perception were tested, and finally an L2 perception study was conducted. In addition, the 'language mode' or 'language activation' variable was subjected to a more rigorous procedure than the one reported in § 6.3 because two separate listening sessions were incorporated into the experimental design, as will be described in § 7.4.3. This experimental procedure directly addressed the model's hypothesized separate perceptions for L2 learners by resorting to a monolingual setting for each language. Finally, a thorough background questionnaire was administered to establish the learners' lifetime French and English hours of exposure, given that input factors are known to play an important role in the attainment of optimal L2 perception, as was discussed in § 3.6.1.

7.4.1 Canadian English (CE) and Canadian French (CF) perception

To test whether the production difference between CE and CF /æ/ and /ɛ/ results in a larger perception difference than the one reported in the Spanish and SE case, the native perception of eight monolingual adult CF listeners and eight monolingual adult CE listeners was tested. The subjects for this cross-language study were tested at the speech perception laboratory at McGill University's School of

Communication Sciences and Disorders.[71] The listeners were presented with the same natural tokens described in § 1.1.2 and § 7.1. Thus, the target stimuli were 60 CF and 60 CE tokens, i.e., 30 of /æ/ and 30 of /ɛ/ for each language or, in other terms, five different CVCs times 6 speakers per vowel. Given that the figures in § 7.2.1 only show the average productions for the two languages, Figure 7.25 below illustrates the acoustic values of the 60 tokens produced in each language.

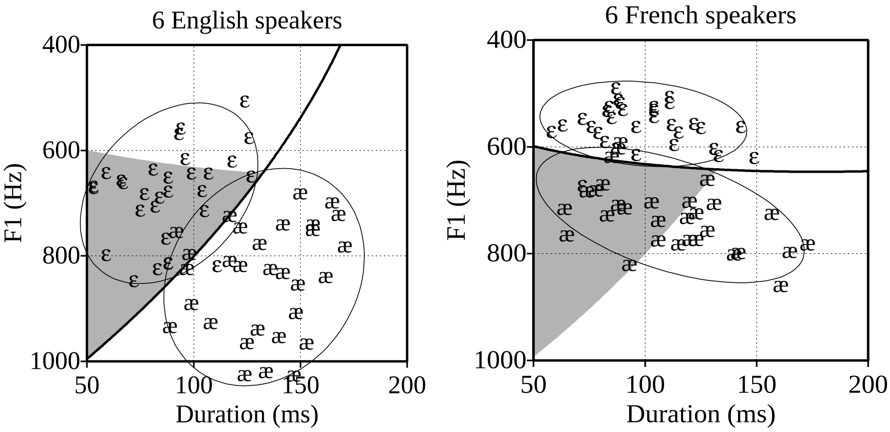

Fig. 7.25. F1 and duration values of the 60 CE (left) and the 60 CF (right) stimuli tokens.

If we look at individual tokens in the figure, then the grey region, which was predicted to be the region of cross-language mismatch in § 7.2.3, seems to represent the cross-language production variation of the two languages. This is because out of the 26 CE tokens that fall in that F1-duration acoustic region, only 3 were intended as /æ/ and the remainder as /ɛ/ whereas out of the 15 CF tokens produced in the same region, only one was intended as /ɛ/ and the rest as /æ/. Due to this clear production difference, the categorization of tokens in this region was

[71] The monolingual study reported in this section and the cross-language study reported in the next section were conducted in collaboration with Linda Polka at McGill University, and they were partially reported in Escudero & Polka (2003). The analysis of the acoustic and perceptual data presented here was conducted in collaboration with Paul Boersma and presented at the Ninth Laboratory Phonology conference (cf. Escudero & Boersma 2004a). Special thanks go to Linda Polka for her collaboration on this project and for her financial support during my first months in Montreal. Special thanks also go to Stephanie Blue, Pascale Tremblay, and Eva Villalba who assisted in the recruitment and testing of subjects.

used to test the hypothesis of a differential optimal perception in the two languages.

In this experiment, the stimuli were extracted from the carrier sentences, and each target CVC was repeated twice in each trial. The CE and CF listeners performed a perception task in their native language, and they were asked to choose between one of five English or French vowels. Thus, the English listeners had five keyword options, namely see, say, at, it, and pet, while the French listeners had to choose from bise, bisse, besse, be, and bace.[72] Following the optimal perception hypothesis, the native CE and CF categorizations were predicted to resemble their specific production distributions so that the same large difference in production was expected in perception. The subjects were only addressed in their native language to avoid any possible language mode effects despite the fact that they reported having only a basic knowledge of the other language.

In Figure 7.26, the left and right squares show that the CE and CF listeners perceived the tokens drawn from their native production environments in a language-specific optimal way. That is, their perceptual category boundary lines resembled their respective equal-likelihood line in production, as can be observed when comparing Figure 7.26 with Figure 7.25. In the figures below, question marks, i.e. '?', represent tokens that were classified as both /æ/ and /ɛ/, while dots, i.e. '•', represent tokens that were classified as a vowel other than /æ/ and /ɛ/, which in most cases was /e/.

72 The majority of the French category responses are nonsense words. Given that these words are orthographically transparent, the subjects were simply provided with options whose pronunciations were all undoubtedly the expected vowels. An identical or very similar CVC context was also provided to ease the listeners' task.

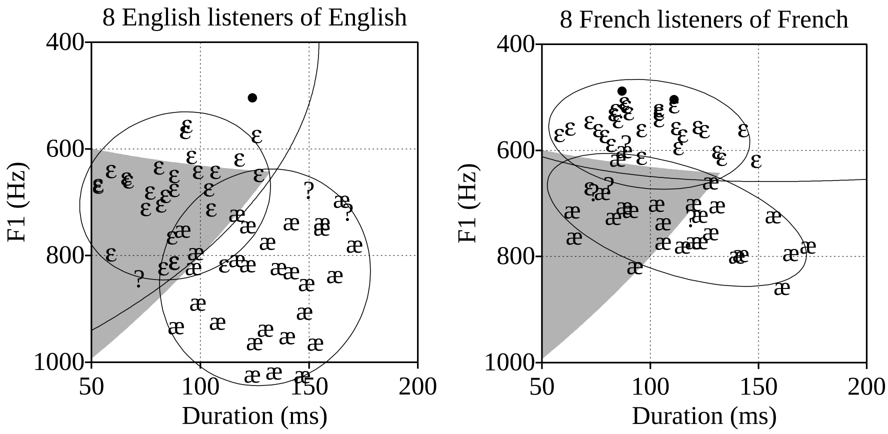

Fig. 7.26. Monolingual categorization of CE /æ/ and /ɛ/ versus monolingual categorization of CF /æ/ and /ɛ/.

7.4.2 Cross-language perceptual mismatch and L2 initial state

The eight monolingual CE listeners also categorized the CF tokens as their own English vowels during the listening experiment reported in the previous section. That is, they performed a cross-language categorization of the 60 CF tokens. The target CF tokens were presented together with the target CE tokens so that the procedure was exactly the same as the one reported in the previous section for the CE listeners. It was predicted that CE listeners would use their L1 perceptual strategies to categorize the CF vowels, i.e., that they would integrate duration and F1 acoustic properties when identifying vowels, just like optimal CE perceivers. Figure 7.27 shows the CE cross-language perception of CF /æ/ and /ɛ/. The question marks, i.e. '?', represent tokens that were classified as both /æ/ and /ɛ/, while the dots, i.e. '•', represent tokens that were classified as a vowel other than /æ/ and /ɛ/, which in most cases was /e/.

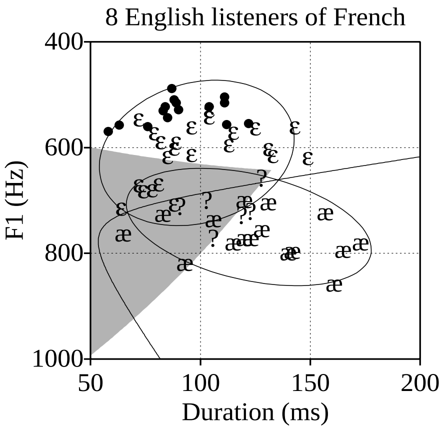

Fig. 7.27. Cross-language perception of CF /æ/ and /ɛ/ by CE monolingual listeners.

At first glance, it may seem that the CE cross-language perception of the CF vowels is closer to the CF category boundary than to the one in CE. However, the influence of the L1 perception is shown in the categorization of the CF tokens in the grey region. Recall that tokens in that region were mostly perceived as /æ/ by CF listeners but as /ɛ/ by CE listeners. When categorizing the CF tokens, the CE listeners identified most of the tokens in the same region as /ɛ/. This categorization results from the use of two L1 perception strategies. First, the CE listeners' relative use of durational and spectral cues for the identification of CF /æ/ was English-like because, unlike the native CF listeners who only used F1, they used both F1 and duration information for categorizing vowels. As a result, the CF tokens with relatively low F1 values of approximately 700 Hz that were produced with a short duration were most likely to be identified as /ɛ/ by these CE listeners but as /æ/ by the native CF listeners. Second, the CE cross-language F1 boundary in the grey region is 200 Hz lower than that of the native CF boundary, which means that the two groups of listeners categorize the F1 dimension differently.

7.4.3 L2 development in Canadian English (CE) learners of Canadian French (CF)

Escudero (in progress b) tested 21 CE learners of CF who were registered in a French language course at the McGill Language Centre.[73] All these learners were originally from non-French speaking regions of Canada and had come to Montreal to start university. They were aged between 18 and 25 and all had monolingual CE-speaking parents. A thorough language background questionnaire was administered to determine their life exposure to French in comparison with English, and they were divided into three exposure groups of seven each. Under the assumption that there was a positive correlation between their exposure and their perceptual proficiency, the three groups of learners were labelled as beginning, intermediate and advanced. The target stimuli were the same 60 CF /æ/ and /ɛ/ tokens presented to monolingual CE and CF listeners in the previous experiments, and they were asked to listen to the French stimuli twice in two different sessions. The procedure is summarized in Table 7.8.

Table 7.8. Summary of the experimental procedure.

Subjects	Session	Stimuli	Experimenter	Language used
21 L2 learners	French	French	French-speaking	Only French
The same 21	English	English & French	English-speaking	Only English

During their French session, they listened to the CF tokens only and were told that all stimuli were French. On a different day, they listened to the same CF stimuli within an English context and they were told that all the stimuli, including the CF tokens, were English. This two-session procedure was intended to measure one of the learners' languages with the least possible influence of the other language. That is, it was expected that the learners would be in a monolingual L2 mode during their French session and in a monolingual L1 mode during their English ses-

[73] The analysis of the L2 perceptual data in the French session was conducted in collaboration with Paul Boersma and presented at the Ninth Laboratory Phonology conference (cf. Escudero & Boersma 2004a).

sion. Furthermore, to enhance the monolingual environment, two different experimenters were involved in the subject testing, viz., a dominantly CF speaker for the French session and a dominantly CE speaker for the English session who only addressed the subjects in the language of the session. Consequently, it was predicted that the perceptual data would not be influenced by parallel language activation (cf. § 3.5.3).

Note that this experimental setting directly bears on the L2LP model's claim that L2 learners and bilinguals have separate perception grammars for their languages. However, if it turned out that there were no perceptual difference between listening to the CF stimuli as French (during the learners' L2 session) and listening to the same stimuli as English (during the learners' L1 session), the L2 model's proposal would not be supported. In such an event, the experimental setting in Table 7.8 might not be needed to measure L1 and L2 perception in L2 learners and, crucially, the alternative proposal that L1 and L2 sounds are handled within a common system would be confirmed.

Figure 7.28 shows the results of the three groups of learners in their L2 (the squares on the left of the figure) and L1 sessions (the squares on the right). The question marks, i.e. '?', in the figures represent tokens that were classified as both /æ/ and /ɛ/, while the dots, i.e. '•', represent tokens that were classified as a vowel other than /æ/ and /ɛ/, which in most cases was /e/. With respect to their L2 categorization, the most evident finding is that the learners needed little exposure to CF in order to decrease their use of duration for L2 vowel categorization. This is shown by the beginning learners' duration boundary, which had already shifted close to the L2 optimal location. However, the results also show that it may take some time to adjust the F1 perceptual boundary between the L2 vowels because beginning learners incorrectly categorized F1 values in the grey region. Nevertheless, it appears that with more L2 exposure, the optimal L2 boundary for vowel height can be achieved, as shown in the perception of the intermediate and advanced learners.

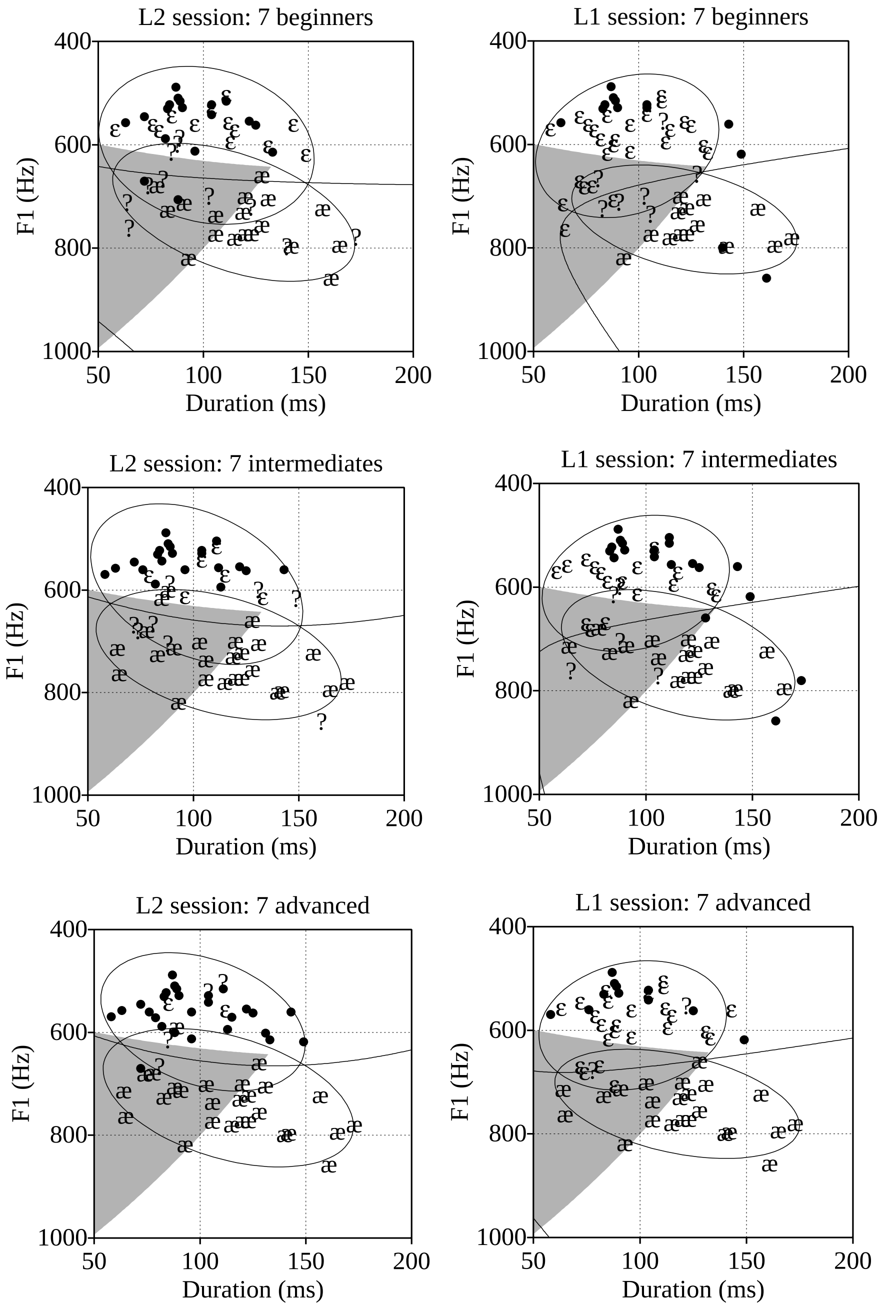

Fig. 7.28. L2 and L1 categorization for the three groups of CE learners of CF.

If the perceptions of the three groups are taken together, it can be observed that, depending on their level of exposure, learners increase the number of L2 /æ/ categorizations in the grey region to match the optimal one, as shown in Figure 7.29. This means that CF /æ/ tokens produced with low F1 values and short durations that were /ɛ/ in monolingual CE perception were in contrast categorized as /æ/ in the L2 perception of CF. A ranked correlation test yielded a significant positive interaction between the number of CF /æ/ categorizations in the grey region and the learners' exposure level (one-tailed Kendall's tau-b = 0.45, N = 21, p = 0.004, i.e., p from zero = 0.23%).

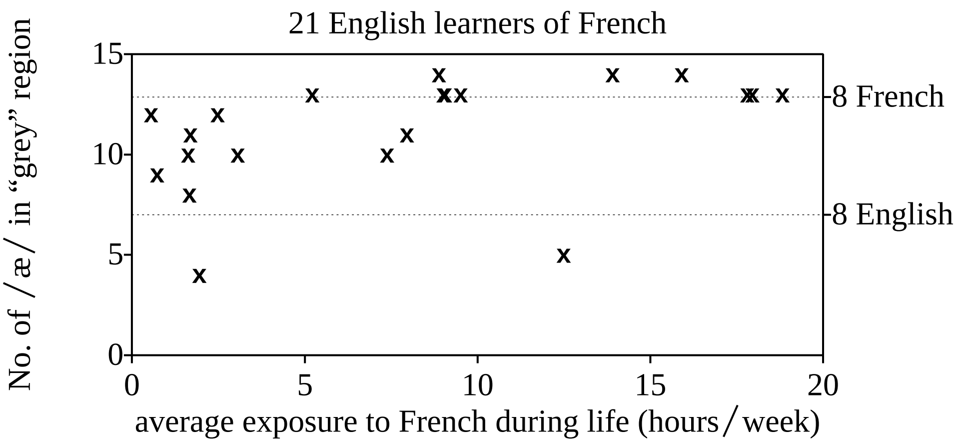

Fig. 7.29. Categorization of the 15 CF /æ/ tokens that fall on the grey region category during the learners' French session (L2 session).

Finally, in the analysis of the perception of the same CF /æ/ tokens presented to the learners during their English session, very different results were obtained. This difference can be seen when comparing the boundary shapes of the squares on the left of Figure 7.28 (i.e. French session) and those on the right (i.e. English session) for all three groups of learners. Thus, when listening to the target CF tokens in their English session, the learners' perception resulted in category boundaries that resembled the monolingual CE categorization of the same CF tokens. In addition, the learners' /æ/ categorization for the CF /æ/ tokens that were produced in the grey region was closer to the monolingual CE categorization when they listened to these tokens as English. If we compare the results of Figures 7.29

and 7.30, we can observe that the learners perceived the same tokens differently when they heard them as French than when they heard them as English. A paired-samples test confirms that the observed difference is significant ($t = 4.51$, $N = 21$, $p < 0.0001$).

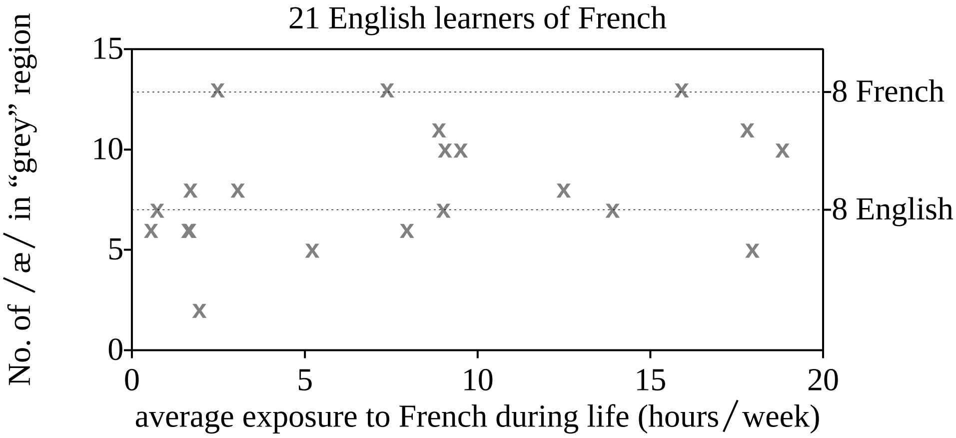

Fig. 7.30. Categorization of the 15 CF /æ/ tokens that fall on the grey region category during the learners’ English session (L1 session).

7.4.4 Discussion

With respect to cross-language perception, it is important to note that monolingual CE listeners might have relied on cues other than F1 and duration when they categorized the CF tokens. For instance, they could have relied on the F2 differences between the vowels that the acoustic analysis revealed to be small but real. As a consequence, a token with a low F1 value and a short duration could still have yielded an /æ/ native or cross-language categorization if its F2 value had been too low to support an /ɛ/ categorization. In fact, it has been shown (cf. Hillenbrand et al. 2000) that when confronted with ambiguous tokens, as were some of those of CF /æ/, English listeners may rely on cues that would only be secondary when categorizing unambiguous native tokens. Thus, using more than two dimensions when predicting and examining cross-language vowel categorization may provide a more reliable measure to predict perception.

Nevertheless, for purposes of L2 initial state and development, the two cues considered in the present chapter, namely F1 and duration, would seem to be extremely informative precisely because L2 learners were shown to have developmentally adjusted their perception of these two dimensions. Furthermore, this development clearly shows that L2 recategorization is possible, and that it is performed through the adjustment of perceptual boundaries and perceptual cue weighting, which is an instance of L1-like development.

7.5 Learning similar sounds: the L2LP predictions versus the evidence

In conclusion to this chapter, we can endeavour to evaluate the predictions of the L2LP model for a similar scenario with respect to the evidence presented in the two previous sections. Table 7.9 summarizes this evidence by showing the prediction for each of the five ingredients of the model and the specific finding that relates to it.

Table 7.9. The five L2LP predictions for a similar L2 sound perception scenario and the evidence to support them.

L2LP ingredients	Predictions for similar	Finding
Optimal L1 & L2	SE and Spanish listeners will exhibit optimal L1 perception	Borne out
Initial state	Beginning Spanish and CE learners will be equal to monolingual Spanish and CE listeners	Partially borne out
Learning task	Boundary shift for CE learners, no action for Spanish learners	Borne out
Development	Lexicon-driven learning	Indirectly borne out
End state	CE learners will attain optimal L2 perception and, at the same time, will maintain their optimal L1 perception	Borne out

Thus, the L2LP model's prediction for the L2 initial state claims that beginning L2 learners start with a copy of their L1 perception grammar and perceived categories. Both the Spanish-to-SE and the CE-to-CF cases support the reuse of L1 categories as well as that of L1 perception that results in perceptual mismatches.

Although the Spanish learners' perception did not show direct evidence of L2 development, it was predicted that they might not need to develop into optimal L2 perceivers, which was an important prediction in itself. That is, the L2LP can predict cases that will not pose any challenge to L2 learners as well as cases that will. One of the cases of a similar scenario that does pose a challenge is the learning of CF vowels by CE learners since the data clearly show that perceptual development occurs in such a situation. Thus, the evidence reported in this chapter confirms the L2LP model's proposal for the possibility of L2 development in a similar scenario. Furthermore, the learning mechanism has been shown to be boundary shift because the CE learners manifested a developmental shift for vowel duration and vowel height perceptual boundaries. Crucially, the model provides a formalization of this learning mechanism through the proposal of a perception grammar and GLA lexicon-driven constraint reranking which result in one-dimensional as well as multi-dimensional category boundary shifts.

In addition, the L2LP hypothesizes that both L1 and L2 perceptions can be optimal because they are handled by two separate perception grammars. The question is whether the data reported here provide any evidence for separate perception systems in L2 learners. The study involving CE learners of CF shows that a more rigorous control of the 'language mode' variable results in a significant difference between the L1 and L2 perceptions of the same tokens. Thus, the L2LP hypothesis of separate perception grammars is borne out. However, an even more rigorous procedure, especially with respect to the nature of the stimuli presented, would be required to show whether L2 learners' L1 perception remains monolingual-like, and whether the difference between L1 and L2 perception does in fact increase with greater L2 exposure.

8 Evaluation and conclusion

What has been presented in this study is a theoretical and empirical contribution to the fields of second language acquisition, phonetics/phonology, and speech perception. With respect to the theoretical contribution, the L2LP model constitutes a linguistic proposal for the phenomenon of sound perception which has often been considered to lie outside the domain of formal linguistic theory, and to be the subject of disciplines such as phonetics and psycholinguistics. In § 8.1, I assess the linguistic nature of sound perception which has been put forth in this study as well as the adequacy of the L2LP theoretical framework to handle native and cross-language sound perception. The L2LP model is based on an emergentist view of language acquisition which proposes that sound perception becomes linguistic knowledge through an interaction of language experience and a general language learning device. In § 8.2, I restate the advantages of this L1 proposal over the other proposals that were discussed in Chapter 4.

In Chapters 3 and 4, it was argued that the L2LP model can describe, explain, and predict L2 sound perception more adequately than rival models. In § 8.3, I evaluate this claim in light of the L2 perceptual data presented in Chapters 5 to 7. This is done with respect to its adequacy in predicting the behaviour of learners in each of the three states of the L2 acquisition process. According to the L2LP model, the comprehensive handling of L2 sound perception involves five ingredients, namely the thorough description of the L1 and the target L2, the L2 learning tasks, and the three logical states in language acquisition, i.e., the initial, developmental, and end states. The evaluation of the model's predictions as well as their respective theoretical explanations and descriptions is divided into five steps which correspond to the L2LP's five ingredients. Finally, in § 8.4, I state the overall conclusions and implications that can be drawn from the theoretical and empirical elements of this study, and I touch on the issues that were not fully addressed in the present work and that can be considered topics for future research.

8.1 Why a linguistic model of sound perception?

In Chapter 1, it was argued that six properties must be considered in any adequate and comprehensive model of speech perception in general and sound perception in particular. First, speech perception refers to the mapping of the variable and continuous speech signal onto perceptual targets. Second, it involves two elements, viz., the perceptual targets, which are abstract phonological representations, and the perceptual mappings which connect the continuous signal with the abstract representations. Third, speech perception is language-specific and language-dependent because the mapping of the speech signal is developmentally shaped by a language environment and is therefore only appropriate to such an environment. Fourth, the degree of abstraction of perceptual targets depends on the properties of the signal and how they are encoded in the perceptual mappings. Fifth, according to psycholinguistic research, listeners map the signal onto pre-lexical representations which precede the access to meaning. And sixth, speech perception is a pre-lexical mapping because it occurs without the intervention of lexical knowledge, which means that it constitutes a bottom-up procedure that receives no feedback from the lexicon. All of these properties have been evidenced in phonetic and psycholinguistic modelling. Thus, to incorporate phonology into the modelling of speech perception, the perceptual mapping of the speech signal needs to represent linguistic knowledge in the form of a set of linguistic rules or a linguistic grammar.

Both the LP and the L2LP models incorporate the six properties that are deemed necessary for a comprehensive modelling of sound perception. This is because the LP proposes that adult sound perception is handled by a linguistic perception grammar that contains auditory-to-phonological representation constraints which connect the signal with abstract phonological categories. Importantly, the perception grammar works without the intervention of the lexicon, thus mapping the speech signal at a pre-lexical level. Fig. 8.1 shows where speech perception in general and sound perception in particular occur within the LP view of speech comprehension. In this figure, we observe that this faculty is viewed as a process that involves two consecutive mappings: the mapping of auditory dimensions onto abstract phonological categories performed by the perception grammar, and the mapping of perceived phonological categories onto lexical representations performed by the recognition grammar, otherwise known as perceptual input.

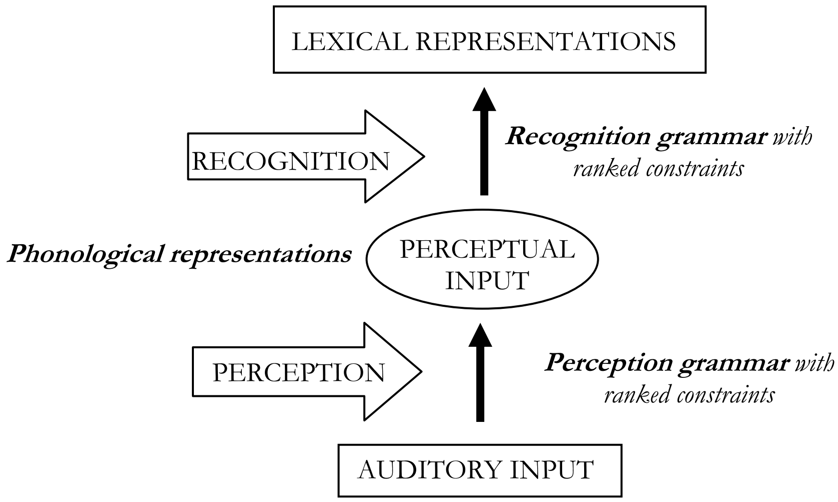

Fig. 8.1. The LP elements for speech comprehension: Auditory input, two mappings, and two levels of abstract representations.

In Chapter 4, it was argued that in comparison to five other models, the LP proposal handles sound perception in the most explanatorily adequate way. This is because this model makes a principled distinction between the two elements of speech perception, namely perceptual mappings and sound categories, whereas four of the five other models fail to distinguish them. This distinction can also capture small and large variations in the perception of the sounds of different languages and language varieties by native and non-native listeners, as shown by the cross-language findings reported throughout this study. What comes closest to an adequate proposal for speech perception is Kuhl's Native Language Magnet (NLM) model described in Chapter 4 which posits neural mappings and a set of abstract phonetic categories that result from such mappings. However, the LP model constitutes the linguistic version of the NLM, and therefore brings speech perception into the domain of phonological theory. Given that sound perception is language-specific, its linguistic modelling may not be just simply a reasonable choice but perhaps the most adequate one.

With respect to the acquisition of sound perception, it was shown that the L1 proposal based on the LP model found in Boersma, Escudero & Hayes (2003) provides a phonological formalization of the mechanisms underlying the acquisition of sound perception. That is, the learning mechanisms that are performed by the model's GLA (cf. § 2.3.2), viz., auditory-driven learning and learning guided by lexical knowledge, provide a formal phonological account of the formation of phonetic and phonological sound categories because the GLA's mechanisms apply in the context of a perception grammar. Three of the other models reviewed in Chapter 4 do not give an explicit L1 learning proposal. The other two do provide an L1 proposal that considers learning mechanisms such as categorization, distributional learning, the perceptual warping of acoustic dimensions, category formation, and equivalence classification. However, the LP advantageously synthesizes all these learning mechanisms in an explicit formalization of the steps through which infants and children obtain optimal sound perception.

8.2 What does the L2LP model provide?

In Chapter 4, five models that aim at explaining cross-language or L2 sound perception were compared to the L2LP model, and the advantages of the latter were shown to lie in its scope and level of explanation. With respect to its scope, the L2LP model has the most ambitious objective because it provides an explicit proposal for each of the three states of L2 sound perception. Also, the other L2LP ingredients provide thorough descriptions of the learner's native language, target L2, and L2 learning task, which the other models allude to but do not elaborate on. All in all, then, it is argued that the L2LP model constitutes a comprehensive proposal that integrates, synthesizes, and improves on the other L2 sound perception models. That is, the theoretical ingredients of the model provide an explicit theory of the entire acquisition process by offering the most detailed proposal for explaining each of the five components of L2 sound perception. In the following subsections, I restate the explanations for L2 sound perception in each ingredient and the predictions that follow from them. In light of the L2 perceptual data presented in Chapters 5 to 7, I evaluate the model's explanatory and predictive power in comparison to the other models of L2 sound perception.

8.2.1 A thorough description of the learner's L1 and target L2

The L2LP model provides a rigorous phonetic and phonological description of L1 and target L2 perception. The phonetic measures for sound perception are connected to the optimal perception hypothesis which states that an optimal listener matches the distributions of her production environment. That is, the location and shape of optimal perceptual boundaries match the equal-likelihood boundaries in production. Likewise, an optimal listener's perceptual use of auditory dimensions matches the use of the same dimensions in production.

Within the L2LP model, optimal perception is described with the aid of linguistic perception grammars and abstract phonological categories. One of its basic claims is that a separation between perceptual mappings, which are performed by the perception grammar, and sound representations, which are constructed by the perception grammar, allows us to adequately compare the languages involved. This principled separation between mappings and categories is crucial to the explanation, description, and prediction of L2 sound perception. Thus, when dealing with two languages, the separate comparison of sound categories and perceptual mappings allows for the prediction of the two L2 learning tasks, i.e., the perceptual one and the representational one. Crucially, when perceptual mappings constitute the only difference between the L1 perception and the target L2 perception, as in the case of /æ/ and /ɛ/ for Canadian English and Canadian French listeners, the L2 learner will have only a perceptual learning task.

The L2LP model proposes that the description of optimal L1 perception leads to predicting the initial state for L2 acquisition, i.e., the perceptual system that learners will initially use in their L2. Computing both the L1 and target L2 optimal perceptions also allows us to determine cross-language mismatches that lead to L2 learning tasks. Moreover, the L1 optimal perception will give a reliable estimate of the system that the learner needs to maintain to be able to optimally cope with her L1 environment as she develops into an optimal L2 listener. The other models for L2 sound perception that were reviewed in Chapter 4 do not have this descriptive and predictive power because they lack a clear comparative measure of sound perception and the knowledge that underlies it. They also fail to explicitly assume and use the separation between mappings and categories.

8.2.2 A linguistic model for the L2 initial state

According to the L2LP Full Copying hypothesis, the learner automatically uses her entire L1 perception grammar and categories when starting to learn her L2. Crucially, the L2LP model suggests that the Full Copying hypothesis can only be tested if the cross-language or non-native perception of the target L2 is compared to the L2 perception of absolute beginners. For instance, the cross-language perception of Southern English vowels by Spanish listeners must be compared to the L2 perception of beginning Spanish learners of Southern British English. This cross-language connection is not found in the models that fail to take into account the perception of beginning learners or the non-native perception of monolingual listeners. In contradistinction, the L2LP proposal for the initial state in L2 sound perception proposes an the explicit linguistic mechanism of Full Copying to account for the effects of linguistic experience in the categorization of non-native or L2 sounds. In addition, not only does this explicit proposal describe the L2 initial state but it can also adequately predict it because it postulates a direct correlation between cross-language perception and the initial L2 state. In sum, the prediction that follows from the L2LP model is that beginning L2 learners will have a perception that matches the cross-language perception of monolingual speakers in their L1. Thus, as was shown in Chapters 5 to 7, beginning Spanish learners of English and beginning Dutch learners of Spanish perceive target L2 sounds in a manner that resembles the monolingual Spanish and Dutch perception of the same sounds.

8.2.3 A thorough description of the L2 learning task

The L2LP model proposes that depending on how the initial L2 perception compares to the target L2 perception, the learner can encounter tasks that differ in both number and type. Establishing the proper learning task is crucial to predicting what sort of learning needs to occur in order to attain optimal L2 sound perception. Crucially, the number and nature of the tasks that the learner needs to perform determine the level of difficulty that a particular L2 sound perception scenario poses. The comparative level of difficulty for the L2LP's three scenarios is summarized below in § 8.2.4.

The L2LP model assumes a distinction between the two elements of sound perception, namely the phonological categories and the perceptual mappings that connect the signal to those categories. This distinction is particularly useful in predicting that the learning task in a SIMILAR scenario in which learners start with the same number of sound categories as the target L2 will only be perceptual. This prediction goes against some of the other models of sound perception that predict that there is no learning task in this scenario. In Chapter 7, it was shown that advanced Canadian English learners of Canadian French are closer to the optimal target L2 perception than beginning learners, and that it can therefore be concluded that learning has taken place in this case.

The L2LP model also predicts that a SUBSET scenario will have both a perceptual and a representational task. In this case, learners start with more categories than the target L2 has so that they need to reduce the number of lexical and perceptual categories. Crucially, other L2 sound perception models fail to establish the learning tasks for this scenario, and therefore do not consider its existence. However, it was shown in Chapter 6 that beginning Dutch learners of Spanish perceive more sound categories than are required to optimally categorize Spanish sounds, and this strongly suggests that they need to perform a learning task in order to become optimal target L2 listeners. In addition, the L2LP model makes a further distinction between the perceptual and representational tasks that are involved in creating new categories or splitting existing L1 categories. That is, the creation of sound categories occurs when the L2 sound perception involves non-previously-categorized dimensions whereas the splitting of sound categories takes place when already-categorized-dimensions are found. However, these cases belong to the same NEW scenario because in both instances the learner starts with fewer categories than the target L2 has.

8.2.4 An explicit and comprehensive proposal for L2 development

The L2LP model provides an explicit and comprehensive account of how L2 learners develop the linguistic knowledge that will turn them into optimal L2 listeners. The model's Full Access to the GLA hypothesis states that L2 learning will be governed by the same mechanisms that are present in the acquisition of L1 sound categories. That is, the L1 learning device or GLA, which is responsible for the perception and recognition learning in L1 acquisition, also applies to L2 acquisition. This means that L2 learners will create new categories and new perceptual mappings through auditory-driven GLA perceptual learning. Later on, their perception will become optimal through lexicon-driven GLA perceptual learning. In addition, optimal L2 storage will be achieved through the GLA recognition learning proposed in § 6.2.4. As was shown, these three types of mechanisms can be applied to predict L2 sound perception in the NEW, SIMILAR, and SUBSET scenarios.

For instance, Spanish learners of Southern British English will create new categories along a non-previously-categorized continuum, which in this case vowel duration. Although other explanations of this phenomenon, such as the one found in Bohn (1995), predict the same result, the L2LP explanation combines the Full Copying of L1 non-previously-categorized continua with the L1-like mechanism of auditory-driven learning. This L2LP Full Copying and category creation contrasts with Bohn's non-transfer and category split hypothesis, though both proposals lead to the prediction that Spanish learners will distinguish English high front vowels by means of duration. This duration use was confirmed by the L2 perceptual data shown in Chapter 5. Thus, given that the L2LP proposes a formal and explicit modelling of the initial knowledge that triggers L2 development as well as the specific learning mechanism in this NEW sound perception scenario, it should be considered a better explanation for L2 acquisition than Bohn's.

The L2LP lexicon-driven GLA perceptual learning results in the adjustment of the perception grammar. The observable consequence of this in L2 perception is the one-dimensional or multi-dimensional shift of L2 category boundaries which entails that they will gradually match the optimal location and shape of target L2 perception. Learners facing a SIMILAR scenario will show an adjustment in L2 category boundaries as a function of their proficiency. This L2LP prediction contradicts the one in Flege's (1995) Speech Learning Model (SLM) whereby the acquisition of SIMILAR sounds will never result in native-like L2 perception. The findings reported in Chapter 7 were seen to confirm the L2LP predictions of the availability of L2 development and the attainment of optimal L2 sound perception in a SIMILAR scenario. This is because the advanced Canadian English learners of Canadian French manifested a perception that matched that of monolingual Canadian French listeners. Furthermore, it was shown that the learning mechanism was the predicted category boundary shift, as shown by the Canadian English learners' developmental adjustment of their multi-dimensional L2 perceptual boundaries.

For the SUBSET scenario, the L2LP proposes that lexicon-driven perceptual learning will be initiated by recognition learning. That is, the learning task is to reduce the number of perceptual categories, and this can only occur if lexical categories are also reduced. The novel proposal described in § 6.2.4 posits that L2 acquisition in this scenario occurs through the parallel adjustment of the recognition and perception grammars. Learning starts when recognition has to change due to a semantic-driven error, and this in turn triggers the recognition-perception mismatch that is needed to get lexicon-driven perceptual learning started. In the case of Dutch learners of Spanish, both perceptual and recognition learning were shown to feed each other, thereby resulting in the gradual reduction of perceptual and lexical categories. In sum, the perceptual data reported in Chapter 6 show that Dutch learners of Spanish reduce the perception of a copied L1 category when listening to Spanish front vowels, thus confirming the L2LP prediction.

8.2.5 An explanation for the attainment of optimal L2 sound perception

The L2LP model puts forth the separate perception grammars hypothesis which states that L2 learners and bilingual speakers have separate systems for perceiving each of their languages. In contrast, Flege's SLM proposes that the perception of L1 and L2 sounds is performed by a phonological system that is composed of a single set of sound categories. These two positions yield completely different predictions for the L2 end state. That is, the L2LP predicts that an L2 learner can attain optimal L2 perception and maintain her optimal L1 perception because the two languages have separate perception grammars. Conversely, the SLM presupposes that L2 development will inevitably affect L1 perception.

The L2LP's interpretation of language modes is that two separate systems can be activated at the same time during online perception, thus leading to intermediate L1-L2 categorization. Unlike the SLM, the L2LP predicts that experienced L2 learners can have optimal L1 and L2 perceptions in the monolingual setting of each language. Thus, the studies that were conducted in Chapters 6 and 7 had experimental designs that made the language-mode variable operative by controlling the sub-variables involved in the conditioning of a language mode, e.g., stimuli, language of instruction, language of responses, etc. In Chapter 6, for example, it was shown that intermediate and advanced Dutch learners' L1 perception was not affected by their L2 development when they were conditioned to activate their L1 only. These same learners had an intermediate L1-L2 perception when they were conditioned to activate both of their languages, and they had an optimal or near optimal L2 perception while performing a monolingual L2 task. Similarly, Canadian English listeners had different results when perceiving Canadian French tokens under a monolingual L1 conditioning than when perceiving them under a monolingual L2 conditioning. In sum, these results in two different types of learners in the face of different learning scenarios would certainly appear to confirm the L2LP's hypotheses on language mode activation and separate perception grammars.

8.2.6 Three different scenarios and their comparative learning paths

The L2LP model predicts that L2 learners will be confronted with learning tasks that depend on the cross-language difference between their L1 and target L2 optimal perceptions. The number and nature of the tasks will determine the learner's

L2 sound perception scenario and the level of difficulty in the pursuit of optimal L2 perception. Table 1 shows how learners start in each of the L2LP scenarios, what their learning tasks are, and what the comparative level of difficulty between the different scenarios is.

Table 8.1. Predicted initial states and learning tasks for the three L2LP scenarios.

	NEW	SUBSET	SIMILAR
Initial state	Too few categories	Too many categories	Non-optimal mappings
Perceptual task	yes	yes	yes
Representational task	yes	yes	no
Relative difficulty	Most difficult	Medium difficulty	Less difficult

The L2LP predicts that the NEW scenario in which the representational learning task is either to create new categories or to split already existing ones will be more difficult than the SIMILAR scenario because the latter only requires a perceptual task whereas the former involves both a representational and a perceptual task. This contradicts the SLM prediction that the SIMILAR scenario will be the most difficult because it will be unlikely for L2 learners to attain native-like or optimal perception. In Chapter 7, it was shown that Canadian English learners of Canadian French who faced a SIMILAR scenario achieved optimal L2 perception because the ones who were at an advanced level manifested the same perception as that of monolingual Canadian French listeners. Crucially, the beginning Canadian English learners already showed an L2 perception that was close to the optimal target L2. In contrast, the majority of the experienced Spanish learners of Southern British English who faced a NEW scenario (cf. Chapter 5) did not reach optimal performance because they showed clear signs of having created new categories without having the necessary cue integration to optimally perceive the target L2. This does not mean that these learners could not attain optimal L2 perception but only that they would do so with more difficulty than learners who were faced with a SIMILAR scenario. Thus, it can be concluded that the L2LP prediction of the relative level of difficulty between these two scenarios is borne out, and that it is more adequate than the SLM prediction.

In addition, the L2LP predicts that the level of difficulty for the SUBSET scenario will be higher than for the SIMILAR scenario but lower than for the NEW scenario. This prediction contradicts all other models of L2 sound perception since they fail to consider the possibility of a learning task in the SUBSET scenario and to foresee any level of difficulty. When comparing the results of the Dutch learners of Spanish that were reported in Chapter 6 to those of the two previous types of learners, it is clear that the L2LP predictions for this scenario are borne out. That is, beginning Dutch learners of Spanish had a non-optimal L2 perception which perfectly matched their L1 perception, whereas advanced Dutch learners of Spanish had an L2 perception that matched that of monolingual Spanish listeners.

8.3 Overall contribution

The present study has provided a linguistic model for L1 and L2 sound perception which synthesizes insights from phonetics, phonology, and psycholinguistics, and which therefore constitutes the most comprehensive proposal currently available. On the one hand, the L1 acquisition model puts forth an integrated proposal not only for the creation of phonetic categories from the input distributions of a language environment but also for the creation of abstract phonological sound categories that are used in adult perception. With respect to L2 acquisition, the L2LP model synthesizes previous models for L2 sound perception and improves on their explanatory and predictive power. Moreover, it accounts most adequately for the attainment of optimal L2 sound perception in different scenarios by predicting and explaining their different learning tasks, levels of difficulty, and developmental paths. The new proposal for modelling the parallel learning of perception and recognition learning (cf. § 6.2.4) constitutes a large step forward in achieving a realistic and comprehensive model of the acquisition of L2 segmental phonology.

The L2 perceptual data presented in this study have shown that L2 learners can develop. That is, regardless of the type of learning task, e.g., creating categories, reducing categories, or adjusting category boundaries, and notwithstanding the level of difficulty, L2 learners can attain optimal sound perception. Perhaps more interestingly, the data reported in this study show that the common assumption that L2 development has negative effects on the L1 does not hold. This provides more supporting evidence for the alternative assumption that bilingualism is not detrimental or problematic. On the contrary, bilingual speakers may well attain full proficiency in two languages. With respect to the key to L2 development, the data confirm the L2LP hypothesis that it is driven by rich input. In other words, learners benefit a great deal from a rich environment such as that found in the community of speakers of their L2, as observed, for example, in Montreal for the very successful Canadian English learners of Canadian French. What this strongly suggests is that the L2 environment enhances the critical auditory cues of sound categories, especially if the learners do not have the benefit of a rich natural environment, i.e., if they do not live in a country were the L2 is a native language.

8.4 Future research

In Chapter 2, it was hypothesized that human listeners are optimal perceivers because their perception matches the production distributions of their language environment. However, the Southern British English listeners reported on there were found to adapt their perception to the size of the stimulus square with which they were presented because their category boundary and perceptual cue weighting were different from the ones that were expected. Similarly, the monolingual Spanish listeners' perception of stimuli with Southern British English F1 values resulted in a non-expected category boundary. For instance, the size of the stimulus square might have triggered the use of a different sound category in Southern British English listeners, viz., /ɛ/ for tokens with high F1 values, or it might have triggered their knowledge of other English dialects. For the Spanish listeners, the unusually low F1 values for /i/ might have triggered *extremity* and/or *prototypicality* effects in their vowel categorization (cf. Nakai 1998). This type of stimulus-related boundary shift needs to be addressed in future work. More specifically, a measure should be provided for the size of the shift and the variable that causes it. This phenomenon could be incorporated into the LP model on the basis of the modelling of prototypicality described in Boersma (2003, 2005). Importantly, the extent of stimulus-related boundary shifts should be compared to the category boundary shifts that result from lexicon-driven learning in L1 and L2 acquisition.

With respect to L2 development, the L2LP predictions were tested with a comparison of different groups of learners who were examined in a cross-sectional manner, i.e., at a single juncture in their development. Although in most cases this comparison yielded the expected developmental differences, some groups of learners and some individuals in particular either did not exhibit the predicted L2 development or developed faster than expected. For instance, the beginning Spanish learners of English could not be differentiated from the non-beginning Spanish learners of Scottish English. Also, the beginning Canadian English learners of Canadian French already had a perception that was closer to the target L2 perception than to the cross-language perception of monolingual Canadian English listeners. These results suggest that we need to resort to more stringent criteria in the selection of learners and to consider more groups of learners, something that should allow us, inter alia, to find learners who are true beginners. Crucially, longitudinal studies are needed to show the relative degree of difficulty of all three L2 sound perception scenarios in the same group of learners. This approach should also allow us to cope with the large individual variation in L2 sound perception.

In Chapters 3 and 5, it was mentioned that two different learning tasks are attested in the NEW scenario. One is to *create* categories along non-previously-categorized auditory continua and the other is to *split* existing categories along already-categorized-continua. However, although the initial state and learning task in both types of NEW L2 sound perception scenarios were explained and predicted, this was done only for the first type in the case of the end state and L2 development. This is because category splitting does not result from either of the two L1-like perceptual learning mechanisms that the GLA instantiates, namely auditory-driven category formation and lexicon-driven boundary shift. Given that this type of NEW scenario is widely attested in the L2 sound perception literature, the L2LP model should be able to explain and predict how L2 development occurs. More specifically, the model should offer an explicit formal proposal of how L2 learners split already-categorized-continua in order to perceive the number of target L2 categories. In addition, the exact relative level of difficulty between creating and splitting categories needs to be formulated. Although a comparison of L2 category split with L2 category creation was beyond the scope of the present study, this should certainly be addressed in future work.

In § 6.2.4, a proposal for explaining the parallel acquisition of recognition and perception in second languages was advanced. This proposal could be combined with Boersma, Escudero & Hayes' (2003) L1 model in an attempt to provide a comprehensive account of the acquisition of the lexical and perceptual knowledge involved in optimal sound perception. However, before exploring these modelling possibilities, such a novel proposal would have to be validated in a computer simulation that could show whether the predicted learning does in fact occur. Although such simulation is also beyond the purview of the present study, it should certainly constitute one of the future directions that the linguistic modelling of the acquisition of sound categories needs to take. This more comprehensive approach could potentially allow us to model the acquisition of entire vowel and consonant systems in L1, L2, and bilingual learners.

In § 3.2.5, the language-mode activation hypothesis was formulated so as to explain how learners develop an optimal L2 sound perception without creating a negative effect in their already optimal L1 perception. This hypothesis was based on an experimental design whereby a differential perception mode could be triggered depending on the language setting with which the learner was faced. In future work, a rigorous and fully controlled experimental paradigm that uses language mode activation as the main variable will need to be devised. Such a paradigm should allow for a more valid replication of the separate perception grammar effects found in Dutch learners of Spanish as well as Canadian English learners of Canadian French. Preferably, this paradigm should be incorporated in a longitudinal study designed to discover whether the activation and inhibition of a bilingual mode is acquired as part of the learning process. Such a language-mode activation study could be used to test adult L2 learners who start learning an L2 after puberty, simultaneous bilingual children who are born in a bilingual environment, and sequential bilingual children who are born in a monolingual environment but who learn an L2 at four to five years of age. These are but a few examples of the future research that can be envisaged in order to further test the theoretical and methodological proposals that were put forth in this study.

Resumen

El objetivo principal de este estudio es proporcionar una descripción, explicación y predicción exhaustiva de cuán óptima es la adquisición de la percepción de sonidos en L2. Aunque la mayoría de las observaciones y explicaciones de la fonología segmental de L2 se han basado en datos de producción, enfoques basados en dificultades perceptuales también se han considerado, aunque la mayoría en el campo de la fonética. Por ejemplo, las investigaciones sobre percepción del discurso cruzado-lingüístico realizadas en los años 60, mostraron que los estudiantes de L2 también poseen 'acentos extranjeros perceptuales'. Esto sugiere que el origen de un acento extranjero es el uso de estrategias perceptuales específicas de la lengua que están afianzadas en el estudiante de L2 y que no se pueden evitar al encontrarnos con sonidos en L2. En otras palabras, los problemas al producir sonidos en L2 se originan en su mayoría de dificultades al percibir tales sonidos de manera tipo nativa. En la introducción a este estudio, argumento que dar prioridad al papel de la percepción al explicar la adquisición de sonidos en L2 parece ser válido y quizás sea la manera más propicia de acercarnos al fenómeno.

Este estudio propone un modelo *lingüístico* de percepción de sonidos en L2 que frecuentemente ha sido considerado que yace fuera del dominio de la misma teoría lingüística y que constituye el tema de importancia de disciplinas tales como la fonética y la psicolingüística. Con respecto a la contribución empírica, este estudio documenta tres escenarios diferentes en la adquisición de percepción de sonidos en L2 y los explica a raíz del modelo de L2 propuesto. Este estudio contiene tres partes principales: la Parte I discute el fenómeno de la percepción del discurso y como se adquiere en la primera lengua (L1) del hablante, la Parte II introduce un nuevo modelo de percepción de sonidos y examina los modelos que lo han precedido, y la Parte III presenta datos empíricos para evaluar la propuesta de L2. Luego prosigue un resumen del contenido de cada una de las tres partes.

La Parte I comprende dos capítulos que estimulan las suposiciones teóricas del modelo de L2 propuesto en la Parte II. En el Capítulo 1, hablo sobre la visión mas común de la percepción del discurso dentro de la teoría fonológica actual, que es, considerar el fenómeno como una propiedad física y no-lingüística que aplica a todos los oyentes humanos por igual. Luego presento evidencia empírica que contradice tal universalidad, demostrando que la percepción del discurso esta formada por la experiencia lingüística y que por lo tanto es sólo apropiada para un entorno de lengua específica. Dado el peso de esta evidencia, argumento a favor de aproxi-

mar la percepción del discurso al dominio de la teoría fonológica debido a su naturaleza de lengua específica y por lo tanto cognitiva. Este primer capítulo también expone el hecho de que casi todos los modelos fonéticos y psicolingüísticos de la percepción del discurso asumen la especificidad de lengua del fenómeno, y finaliza con una lista de criterios para un modelo exhaustivo de percepción de sonidos que incluye una propuesta para proyecciones cognitivas de fonética-a-fonología y representaciones perceptuales pre-lexicales.

En el Capítulo 2, presento un modelo fonológico exhaustivo de percepción de sonidos que se denomina el modelo de *Percepción Lingüística* (PL) y sostiene que es la propuesta más adecuada de tipo explicativo para percepción de sonidos y su adquisición. Primero hablo sobre la propuesta general de percepción de sonidos, que está basada en Boersma (1998) y Escudero & Boersma (2003). Luego, paso a examinar cómo explica la adquisición de percepción de sonidos en L1, una explicación que se basa en el trabajo que fue presentado por primera vez en Boersma, Escudero & Hayes (2003). Aquí presento mi interpretación y explicación personal de la propuesta de percepción de sonidos y adquisición de lenguas que se encuentran en estos tres artículos.

Con respecto a la percepción de sonidos, el modelo PL propone que los oyentes adultos clasifican las vocales y las consonantes de su lengua por medio de una *gramática de percepción*. Esta gramática contiene *restricciones clave* que permiten la proyección de la señal auditiva continua a vocales y consonantes, que dentro del modelo se llaman representaciones fonológicas o entradas perceptuales. De esta manera, las restricciones clave en la gramática perceptual adulta conectan cualquier evento auditivo a cualquier dimensión con vocales o consonantes. Tales restricciones se clasifican de acuerdo a la *hipótesis de percepción óptima* del modelo que establece que un oyente óptimo clasifica la señal del discurso en vocales y consonantes que probablemente sean más previstas por el hablante. Esto significa que la percepción óptima de sonidos de una lengua coincide con las distribuciones de producción de tales sonidos en esta lengua.

En lo que se refiere a la adquisición de L1, el Capítulo 2 describe los dos tipos de desarrollo, a saber, aprendizaje con impulso auditivo y con impulso de léxico, al que se somete una gramática de percepción infantil en su camino a convertirse en de tipo-adulto. Estos tipos de desarrollo son instados por el algoritmo de aprendizaje asociado al modelo, es decir, el GLA, pero en diferentes momentos durante el desarrollo de la percepción infantil. De esta manera, el capítulo demuestra el camino de desarrollo propuesto en el aprendizaje de percepción óptima de sonidos.

Asimismo, resume una cuenta lingüística de cómo los niños adquieren un almacenamiento de palabras de tipo-adulto, que es una habilidad resultado de adquirir una *gramática de reconocimiento* óptima.

La Parte II de este estudio trata con propuestas teóricas para percepción de sonidos en L2. De esta manera, en el Capítulo 3 realizo una propuesta *lingüística* teórica y metodológica para percepción de sonidos en L2 que se conoce como el modelo de *Percepción Lingüística de Segundas Lenguas* (PLL2). Este modelo posee cinco ingredientes teóricos que están dirigidos a describir, explicar y predecir el proceso de adquisición en su totalidad. Primero, propone que la óptima descripción de L1 y la percepción meta de L2 nos permiten predecir y explicar tres diferentes aspectos de la percepción de sonidos en L2, es decir, el estado inicial, la tarea de aprendizaje, y el estado final. Segundo, la hipótesis del modelo de *Copiar Completamente* constituye una explicación lingüística formal para la predicción de que los estudiantes de L2 manifestarán una percepción de L2 que se iguala a la percepción óptima de su L1. Tercero, predice que el *grado de disparidad* entre las gramáticas de percepción definirán el número y la naturaleza de las tareas de aprendizaje de L2 . Cuarto, propone que para cumplir con la tarea de aprendizaje de L2, el estudiante necesitará crear nuevas proyecciones y categorías perceptuales o ajustar sus proyecciones existentes a través de los mismos mecanismos observados en la adquisición de L1. Por último, la hipótesis del modelo de *gramáticas de percepción separadas* y *activación de lengua* predicen que los estudiantes de L2 alcanzarán la percepción óptima de L2 mientras conserven su percepción óptima de L1.

En el Capítulo 4, examino cinco modelos previos del modelo PLL2 con respecto a su percepción del discurso general y propuestas de adquisición de L2. Esta comparación se basa sólo en el campo teórico pero la predicción de los modelos para percepción de sonidos en L2 en diversos escenarios de aprendizaje están establecidos de tal manera que permiten que su validación sea fácilmente evaluada en términos de los datos de percepción de L2 presentados en la Parte III. La conclusión general de este capítulo es que el modelo PLL2 tiene la meta más ambiciosa y por lo tanto el ámbito más grande de los seis modelos debido a que proporciona una propuesta explícita para los tres estados de percepción de sonidos en L2. Asimismo, el primer ingrediente del PLL2, es decir, la descripción minuciosa de la percepción óptima de L1 y L2, no se considera en ninguno de los otros modelos. En resumen, se argumenta que el modelo PLL2 representa una propuesta exhaustiva que integra, sintetiza y mejora el resto de los modelos de percepción de sonidos en L2.

La Parte III constituye la porción empírica de este estudio. Presenta los datos de percepción de sonidos en L2 que documentan tres escenarios diferentes de aprendizaje. El primero en ser examinado es el bien avalado NUEVO escenario, donde los estudiantes se enfrentan con categorías fonológicas de L2 (es decir, fonemas) que no existen en su L1. De esta manera, en el Capítulo 5, la aplicación del modelo PLL2 en estos escenarios se ilustra con la adquisición de vocales /i/ y /ɪ/ de Inglés Británico del Sur (IBS) por estudiantes de Español. El modelo predice que los estudiantes de Español primero aprenderán a identificar las dos vocales de IBS como *larga* y *corta* puesto que así están distribuidas en IBS. Asimismo, se predice que los estudiantes de Español pueden someterse a un mayor desarrollo perceptual que les permitirá incorporar diferencias de calidad de vocales (por ejemplo, el primer formante de las vocales) a sus recién creadas categorías de duración de vocales. En apoyo a esto, se determinó que los oyentes de Español primero categorizan las vocales de IBS como una sola vocal de Español, es decir, /i/ y posteriormente crean nuevas proyecciones perceptuales y categorías perceptuales basadas en la duración de las vocales. En otras palabras, ellos clasifican las vocales IBS de acuerdo con sus distribuciones de duración, lo que sugiere que ellos aprenden a percibir dos categorías de duración de vocales. A pesar de que algunos estudiantes de Español manifiestan una integración de calidad y duración en su percepción de vocales, se necesitan más datos perceptuales, de oyentes con una destreza en L2 más elevada, para poder evaluar esta predicción rigurosamente.

El Capítulo 6 introduce el concepto de un escenario SUBCONJUNTO donde los estudiantes se enfrentan a categorías fonológicas de L2 que tienen más de un homólogo en su L1. A pesar de que los modelos previos no consideran que este escenario constituya un problema de aprendizaje, el modelo PLL2 predice que los estudiantes de L2 encontrarán dificultades si los sonidos de L2 forman un subconjunto de las categorías de su L1. Estos descubrimientos confirman de manera crucial la primera predicción del modelo, al punto que los estudiantes principiantes reutilizan su percepción óptima de L1. De esta manera, por ejemplo, los estudiantes holandeses de Español utilizan tres categorías de L1 para clasificar las vocales en Español /i/ y /e/. En lo que se refiere al desarrollo de L2, el capítulo muestra que los estudiantes holandeses reducen o desfavorecen la categoría extra de L1 en su percepción de L2, un patrón de desarrollo que se predice por el modelo PLL2. Con respecto a la relación entre la percepción de L1 y L2, se muestra que los estudiantes holandeses reducen gradualmente la percepción de una categoría en su L2 pero no en su L1, un descubrimiento que confirma la hipótesis del modelo de gramáticas de percepción separadas para L1 y L2. Esta diferencia entre la percepción de L1 y L2 también confirma crucialmente la predicción del modelo que establece que los estudiantes de L2 pueden alcanzar la óptima percepción en ambas de sus lenguas.

El Capítulo 7 presenta la aplicación del modelo PLL2 al aprendizaje de sonidos SIMILARES de L2. Este escenario surge cuando los estudiantes se enfrentan a dos sonidos en L2 que son fonémicamente equivalentes, pero diferentes fonéticamente de dos homólogos de L1. El caso de la adquisición de las vocales/æ/ y /ɛ/ del Francés Canadienses (FC) por parte de estudiantes de habla inglesa canadiense (IC) se utiliza para ilustrar este escenario. Para este caso, el modelo PLL2 predice que los estudiantes de IC clasificarán las dos vocales de L2 como dos homólogas de L1 y utilizarán estrategias de percepción de L1 al categorizarlas. De esta manera, los datos empíricos sirven para confirmar que los estudiantes de IC manifiestan un cambio gradual en su percepción de duración de vocales y calidad de vocales que les permite convertirse en oyentes óptimos de FC, un patrón de desarrollo que es predecido por el modelo. Asimismo, se muestra como los estudiantes perciben las mismas señales de manera diferente dependiendo de si creen que son Inglés o Francés. Esta diferencia tan significativa en la percepción de L1 y L2 sirve para substanciar la hipótesis de gramáticas de percepción separadas para estudiantes de L2.

Por último, el Capítulo 8 ofrece una discusión general de las conclusiones mientras se relacionan con el modelo PLL2 propuesto así como también con el resto de los modelos de percepción revisitados en este estudio. Asimismo, establece las conclusiones que se pudieron obtener de los temas empíricos y teóricos mencionados en este estudio, así como también el impacto potencial previsible en los campos de adquisición de lenguas, fonología, fonética y psicolingüística. Este último capítulo también trata algunos defectos potenciales del modelo y menciona las investigaciones que actualmente se prevén para mejorar y evaluar más adelante las propuestas metodológicas y teóricas del modelo PLL2.

Summary

The primary objective of the present study is to provide a comprehensive description, explanation, and prediction of how optimal second language (L2) sound perception is acquired. Although most observations and explanations of L2 segmental phonology have been based on production data, approaches based on perceptual difficulties have also been considered, though mainly in the field of phonetics. For instance, cross-linguistic speech perception research performed in the 1960s showed that L2 learners also have 'perceptual foreign accents'. This suggests that the origin of a foreign accent is the use of language-specific perceptual strategies that are entrenched in the L2 learner and that cannot be avoided when encountering L2 sounds. In other words, problems producing L2 sounds originate in large measure from difficulties in perceiving such sounds in a native-like fashion. The basic argument is that prioritizing the role of perception in explaining the acquisition of L2 sounds is a valid and most propitious approach to the problem.

This study advances a *linguistic* model of L2 sound perception, a phenomenon which has often been considered to lie outside the domain of linguistic theory proper and to constitute the subject matter of disciplines such as phonetics and psycholinguistics. In this regard, it documents three different scenarios in the acquisition of L2 sound perception and proposes a theoretical model to account for them. There are three main parts to this study. Part I discusses the phenomenon of speech perception and how it is acquired in a speaker's first language (L1), Part II introduces a new model of L2 sound perception and examines the models that have preceded it, and Part III presents empirical data to test and evaluate this L2 proposal. Following is a summary of the contents of each of these three parts.

Part I comprises two chapters which serve to motivate the theoretical assumptions of the L2 model to be presented in Part II. In Chapter 1, I first discuss the most common view of speech perception within current phonological theory, which is to consider the phenomenon as a non-linguistic and purely physical property that applies to all human listeners equally. I then present empirical evidence that contradicts such a concept of universality by demonstrating that speech perception is shaped by linguistic experience and is consequently only appropriate to a specific language environment. Given the weight of this evidence, I argue in favour of bringing speech perception into the domain of phonological theory because of its language-specific and therefore cognitive nature. In this first chapter I also discuss the fact that, in contradistinction to the phonological models, most phonetic

and psycholinguistic models of speech perception assume the language specificity of this phenomenon. Finally, I draw a list of criteria that are deemed to be essential for a comprehensive model of sound perception, one which includes a proposal for phonetic-to-phonological cognitive mappings and pre-lexical perceptual representations.

In Chapter 2, I introduce a comprehensive phonological model of sound perception called the *Linguistic Perception* (LP) model which is argued to be the most explanatorily adequate proposal for sound perception and its acquisition. I begin by discussing the basic assumptions of this model, which are drawn from Boersma (1998) and Escudero & Boersma (2003), and I then go on to examine how they serve to explain the acquisition of L1 sounds, an explanation that is based on work that was first presented in Boersma, Escudero & Hayes (2003). In sum, I present my personal interpretation and explanation of the sound perception and language acquisition proposals that are found in these three articles.

In a nutshell, the LP model proposes that adult listeners classify the vowels and consonants of their language by means of a *perception grammar*. This grammar contains *cue constraints* that allow for the mapping of the continuous auditory signal onto vowels and consonants which, within the model, are called phonological representations or perceptual input. Thus, the cue constraints in an adult perception grammar connect any auditory event produced along any auditory dimension with vowels or consonants. Such constraints are ranked following the model's *optimal perception hypothesis* which states that an optimal listener classifies the speech signal onto the vowels and consonants that are most likely to be intended by the speaker. This means that the optimal perception of the sounds of any language must necessarily match their production distributions.

In regard to L1 acquisition, Chapter 2 describes the two types of developments, viz., auditory-driven and lexicon-driven learning, that an infant perception grammar undergoes on its way to becoming adult-like. These are instantiated at different times by the learning algorithm which is associated with the model, viz. the Gradual Learning Algorithm (GLA). Overall, this chapter endeavours to demonstrate the proposed developmental path in the learning of optimal sound perception, and it provides a linguistic account of how children reach a stage of adult-like word storage, which is an ability that results from the acquisition of an optimal *recognition grammar*.

Part II of this study deals with theoretical proposals for L2 sound perception. Thus, in Chapter 3, I advance the *linguistic* model of *Second Language Linguistic Percep-*

tion (L2LP) which comprises five theoretical ingredients whose aim is to describe, explain, and predict the whole acquisition process. First, it proposes that the description of optimal L1 and L2 perception allows us to predict and explain three different aspects of L2 sound perception, viz., the initial state, the learning task, and the end state. Second, the model's hypothesis of *Full Copying* constitutes a formal linguistic explanation for the prediction that L2 learners will initially manifest an L2 perception that matches their optimal L1 perception. Third, it predicts that the *degree of mismatch* between perception grammars will define the number and nature of L2 learning tasks. Fourth, it posits that in order to accomplish the L2 learning task, the learner will either need to create new perceptual mappings and categories, or else adjust her existing mappings by means of the same learning mechanisms that operate in L1 acquisition. Finally, the model's hypotheses of *separate perception grammars* and *language activation* predict that learners will achieve optimal L2 perception while preserving their optimal L1 perception.

In Chapter 4, I review five previous models of L2 sound perception and compare them to the L2LP model with respect to their general speech perception and L2 acquisition proposals. This comparison is made on theoretical grounds alone but the models' predictions for L2 sound perception in diverse learning scenarios are stated in such a way as to allow their validity to be easily evaluated in terms of the L2 perception data to be presented in Part III. The overall conclusion of this chapter is that the L2LP model has the most ambitious objective and therefore the largest scope of all six models because it provides an explicit proposal for all three states of L2 sound perception. Furthermore, the first ingredient of this model, viz., the thorough description of L1 and L2 optimal perception, is not considered in any of the others. In sum, it is argued that the L2LP model represents a comprehensive proposal that integrates, synthesizes, and improves on the other L2 sound perception models.

Part III constitutes the empirical portion of this study. It presents L2 sound perception data which document three different learning scenarios. First to be examined is the well-attested NEW scenario wherein learners are faced with L2 phonological categories (i.e., phonemes) that do not exist in their L1. Thus, in Chapter 5, the application of the L2LP model to this scenario is illustrated with the acquisition of the Southern British English (SBE) vowels /i/ and /ɪ/ by Spanish learners. What is predicted to occur is that Spanish learners will first learn to identify the two SBE vowels as *long* and *short* because that is how these vowels are distributed in SBE. In addition, it is foreseen that they can undergo a further perceptual development that allows them to incorporate vowel quality differences into their newly created vowel duration categories. This is supported by the fact that Spanish listeners have been found to first categorize the SBE vowels as a single Spanish vowel, viz., /i/, and then to create new perceptual mappings and categories based on vowel duration. In other words, they seem to classify the SBE vowels according to their duration distributions, suggesting that they learn to perceive two vowel duration categories. Although some Spanish learners have been observed to manifest an integration of quality and duration in their vowel perception, more data of listeners with higher L2 proficiency would be needed to rigorously test this prediction.

Chapter 6 introduces the concept of a SUBSET scenario whereby learners are faced with L2 phonological categories that have more than one counterpart in their L1. Although previous models have not found this to constitute a learning problem, the L2LP model predicts that learners will encounter difficulties if the L2 sounds form a subset of their L1 categories. Crucially, these findings confirm the model's first prediction to the effect that beginning learners reuse their optimal L1 perception. Thus, for example, beginning Dutch learners of Spanish use three L1 categories to classify the Spanish vowels /i/ and /e/ but as they become more competent, they tend to reduce or disfavour the extra L1 category in their L2 perception, a developmental pattern that is predicted by the L2LP model. With respect to the relation between L1 and L2 perception, it is shown that Dutch learners gradually reduce the perception of a category in their L2 but not in their L1, a finding which confirms the model's hypothesis of separate perception grammars. Crucially, this difference between L1 and L2 perception also confirms the model's prediction that L2 learners can achieve optimal perception in both of their languages.

Chapter 7 presents the application of the L2LP model to the learning of SIMILAR L2 sounds. This scenario arises when learners are faced with two L2 sounds that are phonemically equivalent but phonetically different from their L1 counterparts. The acquisition of Canadian French (CF) /æ/ and /ɛ/ by Canadian English (CE) learners is used to illustrate this situation wherein the L2LP model predicts that CE learners will assimilate the two L2 vowels to their L1 counterparts and will use L1 perception strategies when categorizing them. Thus, empirical data serves to confirm that CE learners manifest a gradual shift in their perception of vowel duration and quality that allows them to become optimal CF listeners, a developmental pattern that is predicted by the model. In addition, it is shown how learners perceive the same CF tokens differently depending on whether they believe them to be English or French. This significant difference in L1 and L2 perception serves to substantiate the L2LP hypothesis of separate perception grammars for L2 learners.

Finally, Chapter 8 provides a general discussion of the findings as they relate to the proposed L2LP model as well as to the other L2 sound perception models reviewed in this study. It presents the general conclusions that can be drawn from the theoretical and empirical issues that have been raised as well as the foreseeable impact they might be expected to have on the fields of language acquisition, phonology, phonetics, and psycholinguistics. This final chapter also addresses some potential shortcomings of the L2LP model and touches on the research that is currently envisaged to improve and further test its theoretical and methodological proposals.

Samenvatting

Het doel van de huidige studie is het beschrijven, verklaren en voorspellen van de verwerving van een optimale perceptie (waarneming) van spraakklanken bij het leren van een tweede taal (T2). Hoewel de meeste observaties en verklaringen in de literatuur over de verwerving van het klanksysteem van een tweede taal gericht zijn op productieproblemen (d.w.z. problemen bij het uitspreken), is er ook al wel aandacht besteed, vooral in het fonetisch onderzoek, aan perceptieproblemen (d.w.z. problemen bij het horen). Het taalvergelijkende onderzoek naar spraakperceptie dat in de jaren zestig is uitgevoerd heeft bijvoorbeeld aangetoond dat tweedetaalleerders ook een 'perceptief buitenlands accent' hebben. Dit suggereert dat een buitenlands accent voortkomt uit het gebruik van taalspecifieke perceptieve strategieën, die de leerder niet kan uitschakelen bij het verwerven van klanken uit de tweede taal. Met andere woorden, de problemen met het produceren van klanken in een tweede taal komen grotendeels doordat de leerders deze klanken anders waarnemen dan de mensen voor wie deze taal de moedertaal is. In de inleiding tot deze studie pleit ik er dan ook voor dat de rol van perceptie bij het verklaren van de verwerving van T2-klanken prioriteit zou moeten krijgen omdat het de meest vruchtbare manier is om het fenomeen te benaderen. In dit onderzoek stel ik een *taalkundig* model van T2-spraakperceptie voor, terwijl dit fenomeen vaak is beschouwd als iets wat buiten het domein van de eigenlijke taalkunde ligt, en meer op het terrein van de fonetiek en psycholinguïstiek. Het proefschrift levert bovendien een empirische bijdrage door drie verschillende scenario's in de verwerving van T2-klankperceptie te verklaren op basis van het voorgestelde theoretische model.

De studie bevat drie delen; in deel I worden de spraakperceptie en de eerstetaalverwerving (verwerving van de moedertaal of 'T1') besproken. Deel II introduceert een nieuw model voor T2-klankperceptie, nadat eerder voorgestelde modellen zijn behandeld, en in deel III worden empirische data gepresenteerd om het voorgestelde T2-model te testen en evalueren. De inhoud van deze drie delen wordt hier kort samengevat.

Deel I bevat twee hoofdstukken die de theoretische aannames van het in deel II beschreven L2 model motiveren. In hoofdstuk 1 bespreek ik eerst de meest gangbare visie op spraakperceptie binnen de huidige fonologische theorie, die ervan uitgaat dat het niet een talig maar een fysisch fenomeen betreft, dat voor elke menselijke luisteraar gelijk is. Hierna presenteer ik empirische evidentie die deze universaliteit tegenspreekt, omdat het aantoont dat spraakperceptie wordt gevormd door talige ervaring, en dus wordt gevormd door de specifieke taalomgeving waarin een leerder zich bevindt. Ik pleit er dan ook voor spraakperceptie op te nemen in de fonologische theorie, gezien haar taalspecifieke en daardoor cognitieve aard, een argument dat nog weinig fonologische modellen van spraakperceptie hanteren.

Het hoofdstuk gaat hiernaast ook in op het feit dat bijna alle fonetische en psycholinguïstische modellen van spraakperceptie de taalspecificiteit van dit fenomeen aannemen. Dit eerste hoofdstuk eindigt met een aantal criteria voor een compleet model van klankperceptie, waaronder een voorstel voor fonetisch-cognitieve *afbeeldingen* en prelexicale perceptieve representaties. In hoofdstuk 2 stel ik een fonologisch model voor voor klankperceptie dat alle criteria bevat waar een compleet model aan moet voldoen. In dit proefschrift noem ik dit model *Linguïstische Perceptie* (LP) en beschouw ik dit als het meest adequate verklarende model voor (de verwerving van) klankperceptie. In dit hoofdstuk bespreek ik ook een voorstel voor algemene klankperceptie, gebaseerd op Boersma (1998) and Escudero & Boersma (2003). Hierna leg ik uit hoe het model de eerstetaalverwerving van klankperceptie verklaart, mij baserend op de theorie van Boersma, Escudero & Hayes (2003). In dit onderzoek presenteer ik mijn persoonlijke interpretatie van de voorstellen die in deze artikelen zijn gedaan voor klankperceptie en taalverwerving. Met betrekking tot klankperceptie neem ik binnen het LP-model aan dat volwassen luisteraars de klinkers en medeklinkers van hun taal classificeren met behulp van een *perceptiegrammatica*. Deze grammatica bevat *cue constraints* die zorgen voor de afbeelding van het continue auditieve signaal naar klinkers en medeklinkers, die binnen het model *fonologische oppervlakterepresentaties* of *perceptieve input* worden genoemd. De *cue constraints* in de volwassen perceptiegrammatica verbinden dus elke auditieve gebeurtenis langs elke auditieve dimensie met klinkers of medeklinkers.

Zulke *constraints* worden gerangschikt volgens de *optimale-perceptiehypothese* van het model, die zegt dat een optimale luisteraar het spraaksignaal classificeert in die klinkers en medeklinkers die het meest waarschijnlijk waren bedoeld door de spreker. Dit betekent dat de optimale perceptie van de klanken van een taal nauw correspondeert met de verdeling van deze klanken in de productie.

Met betrekking tot eerstetaalverwerving beschrijft Hoofdstuk 2 de twee typen van ontwikkeling die de perceptiegrammatica van een kind dient te ondergaan op de weg naar een volwassen grammatica, te weten auditief-gedreven en lexicongedreven leren. Deze twee typen van ontwikkeling worden in gang gezet door het leeralgoritme dat bij het model hoort, het *GLA*, maar dit gebeurt op verschillende momenten tijdens de ontwikkeling van perceptie door jonge kinderen.

In hetzelfde hoofdstuk wordt getoond hoe de voorgestelde ontwikkeling van het leren van optimale klankperceptie verloopt. Daarnaast geeft het een taalkundige beschrijving van de manier waarop kinderen de eerste woorden leren opslaan zoals volwassenen dat doen, iets waarvoor een optimale *herkenningsgrammatica* nodig is.

Deel II van deze studie behandelt theoretische voorstellen voor T2-klankperceptie. In hoofdstuk 3 wordt een *taalkundig* theoretisch en methodologisch voorstel voor T2-klankperceptie gedaan, het *Second-Language Linguistic Perception* (L2LP) model. Dit model heeft vijf theoretische ingrediënten om het verwervingsproces te beschrijven, verklaren en te voorspellen. Het L2LP-model stelt allereerst dat de beschrijving van optimale L1-perceptie en optimale T2-perceptie ons in staat stelt om drie verschillende aspecten van L2 klankperceptie te verklaren en voorspellen, namelijk de begintoestand, de leertaak en de eindtoestand. Ten tweede vormt de *Full Copying* hypothese een formele taalkundige verklaring voor de voorspelling dat T2-leerders aanvankelijk een T2-perceptie vertonen die correspondeert met hun optimale T1-perceptie. Ten derde voorspelt het model dat de *mate van mismatch* tussen de twee perceptiegrammatica's het aantal en de aard van de T2-leertaken bepaalt. Ten vierde wordt voorgesteld dat de leerder om de T2-leertaak te volbrengen ofwel nieuwe perceptieve afbeeldingen en categorieën moet creëren of de bestaande afbeeldingen aan moet passen via dezelfde leermechanismen als in eerstetaalverwerving. Tenslotte voorspellen de hypothesen van *afzonderlijke perceptiegrammatica's* en *taalactivatie* dat T2-leerders een optimale perceptie zullen bereiken in hun tweede taal en tegelijkertijd hun optimale perceptie van hun eerste taal zullen behouden.

In hoofdstuk 4 bespreek ik vijf modellen van T2-perceptie, en vergelijk ik ze met het L2LP-model met betrekking tot algemene spraakperceptie en T2-verwerving. In dit hoofdstuk wordt de vergelijking alleen op theoretische gronden gemaakt, maar de voorspellingen van het model voor T2-klankperceptie in diverse leerscenario's worden duidelijk beschreven, zodat de lezer hun geldigheid kan beoordelen aan de hand van de T2-perceptiedata die in het laatste deel van het proefschrift gepresenteerd worden. Tegen het eind van het hoofdstuk wordt geconcludeerd dat het L2LP-model het meest ambitieus is en daarom ook de grootste draagwijdte van alle zes modellen heeft, omdat het een expliciet voorstel doet voor de drie stadia van T2-klankperceptie. Bovendien wordt het eerste ingrediënt van het L2LP-model, de gedetailleerde beschrijving van optimale T1- en T2-perceptie, niet behandeld in de andere modellen. Samenvattend kan worden gesteld dat het

L2LP-model een veelomvattend voorstel behelst dat de overige modellen integreert en bovendien een verbetering oplevert ten opzichte van bestaande modellen.

Deel III bevat het empirische gedeelte van deze studie en presenteert T2-klankperceptiedata in drie verschillende leerscenario's. In elk van de empirische hoofdstukken worden gevallen die een specifiek leerscenario illustreren geproblematiseerd en getest. Ten eerste wordt het bekende *NEW* scenario onderzocht, waarin leerders geconfronteerd worden met fonologische categorieën in hun tweede taal (d.w.z. fonemen) die niet bestaan in hun moedertaal. In hoofdstuk 5 wordt het L2LP-model toegepast op dit scenario, en geïllustreerd aan de hand van de verwerving van de Zuidelijk Brits-Engelse (ZBE) klinkers /i/ and /ɪ/ door Spaanse leerders. Zoals voorspeld door het model categoriseren Spaanse luisteraars de ZBE-klinkers eerst als een enkele Spaanse klinker, t.w. /i/. Daarnaast voorspelt het model dat deze leerders zich zullen ontwikkelen door het creëren van nieuwe perceptieve afbeeldingen en nieuwe perceptieve categorieën. Dit komt omdat ZBE-klinkers verschillen laten zien langs een auditieve dimensie die niet wordt gebruikt bij het classificeren van Spaanse klinkers, namelijk klinkerduur. Daardoor zullen Spaanse leerders de twee ZBE-klinkers eerst leren identificeren als *lang* en *kort*, omdat ze op die manier verdeeld zijn in het ZBE. Wat betreft de empirische ondersteuning van deze voorspelling kunnen we stellen dat Spaanse leerders inderdaad door een stadium gaan waarin ze ZBE-klinkers classificeren naar gelang hun duratie, wat suggereert dat ze twee klinkercategorieën leren, gebaseerd op duur. Bovendien voorspelt het model dat Spaanse leerders een verdere perceptieve ontwikkeling kunnen ondergaan die hen in staat stelt kwaliteitsverschillen in de klinker (bijv. de eerste formant van de klinker) te combineren met hun net gecreëerde klinkerduurcategorieën. Hoewel sommige Spaanse leerders inderdaad een integratie van kwaliteit en duur lieten zien, zijn om deze voorspelling nauwkeurig te testen meer perceptieve gegevens nodig van luisteraars met een hogere mate van bekwaamheid in hun tweede taal.

In hoofdstuk 6 komt het leren van *SUBSET* T2-klanken aan bod, een scenario waarin leerders geconfronteerd worden met fonologische categorieën in hun tweede taal die meer dan één tegenhanger in hun moedertaal hebben, en dus een deelverzameling ('subset') van de T1-categorieën vormen. Hoewel eerdere modellen dit scenario niet beschouwen als een leerprobleem, voorspelt het L2LP-model dat T2-leerders problemen zullen ondervinden als de T2-klanken een deelverameling vormen van hun T1-klankcategorieën. De bevindingen bevestigen de voorspelling van het model dat beginnende T2-leerders hun optimale T1-perceptie zullen hergebruiken. Beginnende Nederlandse leerders van het Spaans gebruiken dus drie categorieën uit hun moedertaal om de Spaanse klinkers /i/ and /e/ te classificeren. Deze resultaten bevestigen ook dat het SUBSET-scenario inderdaad tot specifieke T2-leerproblemen leidt, die niet voorkomen in andere scenario's. Wat betreft T2-ontwikkeling laat hoofdstuk 6 zien dat Nederlandse leerders de extra T1-categorie in hun T2-perceptie laten 'inkrimpen', een ontwikkelingspatroon dat voorspeld wordt door het L2LP-model voor dit scenario. Het is wellicht nog interessanter dat in een tweetalige Spaans-Nederlandse omgeving de perceptie van de Nederlandse leerders sterk correleert met hun ervaring in de tweede taal, terwijl in een ééntalige omgeving deze correlatie niet gevonden wordt. Met andere woorden, deze leerders reduceren wel geleidelijk de perceptie van een categorie in hun tweede taal, maar niet in hun eerste taal. Dit divergerende perceptieve gedrag bevestigt de hypothese van het model dat de L2-leerders twee aparte perceptiegrammatica's voor de twee talen hebben gevormd. Bovendien ondersteunt het de voorspelling dat T2-leerders optimale perceptie kunnen bereiken in beide talen.

Hoofdstuk 7 beschrijft de toepassing van het L2LP model op het leren van *SIMILAR* T2-klanken. Dit scenario doet zich voor wanneer leerders geconfronteerd worden met twee T2-klanken die fonemisch equivalent maar fonetisch verschillend zijn van twee T1-tegenhangers. De verwerving van de Canadees-Franse (CF) klinkers /æ/ and /ɛ/ door Canadees-Engelse (CE) leerders gebruik ik om dit scenario te illustreren. Hier voorspelt het L2LP-model dat CE-leerders de twee T2-klinkers als twee T1-tegenhangers zullen classificeren, en dat ze hun L1-perceptiestrategieën zullen gebruiken bij het categoriseren van deze T2-klanken. Deze voorspelling kon worden bevestigd met empirische gegevens. Wat betreft T2-ontwikkeling laten CE-leerders in dit scenario een geleidelijke verschuiving zien in hun perceptie van klinkerduur en klinkerkwaliteit, die hen in staat stelt optimale CF-luisteraars te worden, een ontwikkeling die voorspeld wordt door het model. Verder toon ik ook aan hoe leerders dezelfde CF-geluiden anders waarnemen als zij denken dat deze Engels zijn dan als zij denken dat deze Frans zijn. Dit verschil in T1- en T2-perceptie bevestigt de L2LP-hypothese dat T2-leerders twee gescheiden perceptiegrammatica's hebben.

Tenslotte voorziet hoofdstuk 8 in een algemene bespreking van de bevindingen en hun relatie tot het voorgestelde L2LP-model en de andere T2-klankperceptiemodellen die in dit proefschrift besproken zijn. Bovendien bevat dit hoofdstuk de conclusies die getrokken kunnen worden uit de theoretische en empirische kwesties die naar voren zijn gekomen, alsook de mogelijke betekenis van mijn proefschrift op de gebieden van taalverwerving, fonologie, fonetiek en psycholinguïstiek. Dit laatste hoofdstuk gaat ook in op eventuele tekortkomingen van het model en nader onderzoek om het L2LP-model te verbeteren en de theoretische en methodologische voorspellingen verder te testen.

References

Abrahamson, A. & L. Lisker (1970). Discriminability along the voicing continuum: cross-language tests. In B. Hala, M. Romportl & P. Janota (eds.), *Proceedings of the 6th International Congress of Phonetic Sciences,* 569-753. Prague: Academia.

Archibald, J. & M. Young-Scholten (2003). The second language segment revisited. *Second Language Research* 19, 163-167.

Ashby, F. G. & R. Gott (1988). Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 14, 33-53.

Ashby F. G. & W. T. Maddox (1998). Stimulus categorization. In M. H. Birnbaum (ed.), *Measurements, Judgement, and Decision Making*, 251-301. San Diego: Academic Press.

Barry, W. (1989). Perception and production of English vowels by German learners: instrumental-phonetic support in language teaching. *Phonetica* 46, 155-168.

Beddor, P. S. & T. L. Gottfried (1995). Methodological issues in cross-language speech perception research with adults. In W. Strange (ed.), *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues in Cross-Language Speech Research,* 207-232. Timonium, MD: York Press.

Best, C. T. (1984). Discovering messages in the medium: speech and the prelinguistic infant. In H. E. Fitzgerald, B. Lester & M. Yogman (eds.), *Advances in Pediatric Psychology,* Vol. 2. New York: Plenum.

Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (ed.), *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues in Cross-Language Speech Research*, 171–203. Timonium, MD: York Press.

Best, C. T. (2003). Cross-language perception of nonnative vowels: phonological and phonetic effects of listeners' native languages. In M. J. Sole, D. Recansens & J. Romero (eds.), *Proceedings of the 15th International Congress of Phonetic Sciences*, 2889-2892. Barcelona: Causal Productions.

Best, C. T. & W. Strange (1992). Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics* 20, 305-331.

Boersma, P. (1998). *Functional phonology.* Doctoral Dissertation, University of Amsterdam. The Hague: Holland Academic Graphics.

Boersma, P. (1999). On the need for a separate perception grammar. *Rutgers Optimality Archive* 358.

Boersma, P. (2000). The OCP in the perception grammar. *Rutgers Optimality Archive* 435.

Boersma, P. (2001). Phonology-semantics interaction in OT, and its acquisition. In R. Kirchner, W. Wikeley & J. Pater (eds.), *Papers in Experimental and Theoretical Linguistics*, Vol. 6, 24-35. Edmonton: University of Alberta.

Boersma, P. (2004). A Stochastic OT account of paralinguistic tasks such as grammaticality and prototypicality judgments. *Rutgers Optimality Archive* 648.

Boersma, P. (2005). Some listener-oriented accounts of hache-aspiré in French. *Rutgers Optimality Archive* 730.

Boersma, P. & P. Escudero (2004). Learning to perceive a smaller L2 vowel inventory: an Optimality Theory account. *Rutgers Optimality Archive* 684.

Boersma, P., P. Escudero & R. Hayes (2003). Learning abstract phonological from auditory phonetic categories: an integrated model for the acquisition of language-specific sound categories. In M. J. Sole, D. Recansens & J. Romero (eds.), *Proceedings of the 15th International Congress of Phonetic Sciences*, 1013-1016. Barcelona: Causal Productions.

Boersma, P. & B. Hayes (2001). Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry* 32, 45-86.

Boersma, P. & C. Levelt (2003). Optimality Theory and phonological acquisition. *Annual Review of Language Acquisition* 3, 1-50.

Bohn, O.-S. (1995). Cross language speech perception in adults: first language transfer doesn't tell it all. In W. Strange (ed.), *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues in Cross-Language Speech Research*, 279–304. Timonium, MD: York Press.

Bohn, O.-S. & J. E. Flege (1990). Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied Psycholinguistics* 11, 303-328.

Bohn, O.-S. & L. Polka (2003). Asymmetries in vowel perception. *Journal of the Acoustical Society of America* 41, 221-231.

Borden, G., A. Gerber & G. Milsark (1983). Production and perception of the /r/-/l/ contrast in Korean adults learning English. *Language Learning* 33, 499-526.

Bradlow, A. (1995). A comparative study of English and Spanish vowels. *Journal of the Acoustical Society of America* 97, 1916–1924.

Bradlow, A. (1996). A perceptual comparison of the /i/-/e/ and /u/-/o/ contrasts in English and in Spanish: universal and language-specific aspects. *Phonetica* 53, 55–85.

Broselow, E. (to appear). Language contact phonology: richness of the stimulus, poverty of the base. In K. Moulton & M. Wolf (eds.), *Proceedings of the North Eastern Linguistics Society (NELS)* 34. Amherst, MA: GLSA.

Browman, C. P. & L. Goldstein (1989). Articulatory gestures as phonological units. *Phonology* 6, 201-251.

Brown, C. (1998). The role of the L1 grammar in the L2 acquisition of segmental structure. *Second Language Research* 14, 136-193.

Brown, C. (2000). The interrelation between speech perception and phonological acquisition from infant to adult. In J. Archibald (ed.), *Second Language Acquisition and Linguistic Theory*. Oxford: Blackwell.

Brown, C. & J. Matthews (1997). The role of feature geometry in the development of phonemic contrasts. In S. J. Hannahs & M. Young-Scholten (eds.), *Focus on Phonological Acquisition*, 67-112. Amsterdam: John Benjamins.

Bürki-Cohen, J., J. L. Miller & P. D. Eimas (2001). Perceiving non-native speech. *Language and Speech* 44, 149-169.

Caramazza, A., G. Yeni-Komshian, E. Zurif & E. Carbone (1973). The acquisition of a new phonological contrast: the case of stop consonants in French-English bilinguals. *Journal of the Acoustical Society of America* 5, 421–428.

Cervera, T., J. L. Miralles & J. Gonzalez-Alvarez (2001). Acoustical analysis of Spanish vowels produced by laryngectomized subjects. *Journal of Speech, Language, and Hearing Research* 44, 988-996.

Cho, T. & P. Ladefoged (1999). Variation and universals in VOT: evidence from 18 languages. *Journal of Phonetics* 27, 207-229.

Chomsky, N. (1957). *Syntactic Structures*. The Hague: Mouton.

Chomsky N. & M. Halle (1968). *The Sound Pattern of English*. New York: Harper and Row.

Clements, J. N. (1985). The geometry of phonological features. *Phonology* 2, 223-252.

Cook, V. J. (2002). Background to the L2 user. In V. J. Cook (ed.), *Portraits of the L2 User*, 1-28. Clevedon: Multilingual Matters.

Diehl, R. L., M. R. Molis & W. A. Castleman (2001). Adaptive design of sound systems: some auditory considerations. In E. Hume & K. Johnson (eds.), *The Role of Speech Perception in Phonology*, 123-139. New York: Academic Press.

Dijkstra, A., J. Grainger & W. J. B. Van Heuven (1999). Recognition of cognates and interlingual homographs: the neglected role of phonology. *Journal of Memory and Language* 41, 496-518.

Dupoux, E., C. Pallier, K. Kakehi & J. Mehler (2001). New evidence for prelexical phonological processing in word recognition. *Language and Cognitive Processes* 5, 491-505.

Eckman, F. R. (1977). Markedness and the contrastive analysis hypothesis. *Language Learning* 27, 315-330.

Eckman, F. R. (1981). On the naturalness of interlanguage phonological rules. *Language Learning* 31, 195–216.

Elman, J., R. Diehl & S. Buchwald (1977). Perceptual switching in bilinguals. *Journal of the Acoustical Society of America* 62, 971-974.

Escudero P. (2001). The role of the input in the development of L1 and L2 sound contrasts: language-specific cue weighting for vowels. In A. H.-J. Do, L. Dominguez & A. Johansen (eds.), *Proceedings of the 25th Annual Boston University Conference on Language Development*, 50-261. Somerville, MA: Cascadilla Press.

Escudero, P. (2002). The perception of English vowel contrasts: acoustic cue reliance in the development of new contrasts. In J. Leather & A. James (eds.), *New Sounds 2000: Proceedings of the 4th International Symposium on the Acquisition of Second-Language Speech*, 122-131 Amsterdam: Amsterdam University Press

Escudero, P. (2004). Revisiting the role of perception in segmental phonology. In K. Kerkhoff, J. de Lange & O. Sadeh Leicht (eds.), *UiL OTS Yearbook 2004*. Utrecht: Utrecht Institute of Linguistics.

Escudero, P. (in progress a). Cross-language perception of /i/ and /ɪ/ produced in different varieties of English.

Escudero, P. (in progress b). Evidence for gradual L2 re-categorization: boundary shifts in the L2 perception of Canadian French /æ/.

Escudero, P. (to appear). Prioritising the role of perception in L2 segmental phonology. In M. C. Pennington (ed.), *Palgrave Guide to Phonology*. Basingstoke: Palgrave Macmillan.

Escudero P. & P. Boersma (2002). The subset problem in L2 perceptual development: multiple-category assimilation by Dutch learners of Spanish. In B. Skarabela, S. Fish & A. H.-J. Do (eds.), *Proceedings of the 26th Annual Boston University Conference on Language Development*, 208-219. Somerville, MA: Cascadilla Press.

Escudero, P. & P. Boersma (2003). Modelling the perceptual development of phonological contrasts with Optimality Theory and the Gradual Learning Algorithm. In S. Arunachalam, E. Kaiser & A. Williams (eds.), *Proceedings of the 25th Annual Penn Linguistics Colloquium. Penn Working Papers in Linguistics* 8, 71-85.

Escudero, P. & P. Boersma (2004a). L2 re-categorization of an 'old' phonological contrast. Poster presented at the *9th Laboratory Phonology Conference*, University of Illinois at Urbana-Champaign.

Escudero, P. & P. Boersma (2004b). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition* 26, 551-585.

Escudero, P. & L. Polka (2003). A cross-language study of vowel categorization and vowel acoustics. In M. J. Sole, D. Recansens & J. Romero (eds.), *Proceedings of the 15th International Congress of Phonetic Sciences,* 861-864. Barcelona: Causal Productions.

Escudero, P. & M. Sharwood Smith (2001). Reinventing the native speaker: or what you never wanted to know about the native speaker so never dared to ask. In S. H. Foster-Cohen & A. Nizegorodcew (eds.), *EUROSLA Yearbook* 1, 275-286. Amsterdam: John Benjamins.

Flege, J. E. (1987). The production of 'new' and 'similar' phones in a foreign language: evidence for the effect of equivalence classification. *Journal of Phonetics* 15, 47-65.

Flege, J. E. (1988). Factors affecting degree of perceived foreign accent in English sentences. *Journal of the Acoustical Society of America* 84, 70-79.

Flege, J. E. (1991). The interlingual identification of Spanish and English vowels: orthographic evidence. *Journal of Experimental Psychology: Human Experimental Psychology* 43, 701–731.

Flege, J. E. (1993). Production and perception of a novel, second-language phonetic contrast. *Journal of the Acoustical Society of America* 93, 1589-1608.

Flege, J. E. (1995). Second language speech learning: theory, findings, and problems. In W. Strange (ed.), *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues in Cross-Language Speech Research*, 233–277. Timonium, MD: York Press.

Flege, J. E. (1998). The role of phonetic category formation in second-language speech learning. In J. Leather & A. James (eds.), *New Sounds 97: Proceedings of the Third International Symposium on the Acquisition of Second-Language Speech,* 79-88. Klagenfurt: University of Klagenfurt.

Flege, J. E. (2002). Interactions between the native and second-language phonetic systems. In P. Burmeister, T. Piske & A. Rohde (eds.), *An Integrated View of Language Development: Papers in Honor of Henning Wode*, 217-244. Trier: Wissenschaftlicher Verlag.

Flege, J. E. (2003). Assessing constraints on second-language segmental production and perception. In A. Meyer & N. Schiller (eds.), *Phonetics and Phonology in Language Comprehension and Production*. Berlin: Mouton de Gruyter.

Flege, J. E., O.-S. Bohn & S. Jang (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics* 25, 437–470.

Flege, J. E. & W. Eefting (1987). Cross-language switching in stop consonant production and perception by Dutch speakers of English. *Speech Communication* 6, 185–202.

Flege, J. E. & W. Eefting (1988). Imitation of a VOT continuum by native speakers of English and Spanish: evidence for phonetic category formation. *Journal of the Acoustical Society of America* 83, 729–740.

Flege J. E. & R. Hammond (1982). Mimicry of non-distinctive phonetic differences between language varieties. *Studies in Second Language Acquisition* 5, 1-18.

Flege, J. E. & I. R. A. MacKay (2004). Perceiving vowels in a second language. *Studies in Second Language Acquisition* 26, 1-34.

Flege, J. E., I. R. A. MacKay & D. Meador (1999). Native Italian speakers' production and perception of English vowels. *Journal of the Acoustical Society of America* 106, 2973-2987.

Flege, J. E., M. J. Munro & R. A. Fox (1994). Auditory and categorical effects on cross-language vowel perception. *Journal of the Acoustical Society of America* 95, 3623-3641.

Flege, J. E., C. Schirru & I. R. A. MacKay (2003). Interaction between the native and second language. *Speech Communication* 40, 467-491.

Flege, J. E., G. Yeni-Komshian & S. Liu (1999). Age constraint on second-language acquisition. *Journal of Memory and Language* 41, 78-104.

Fowler, C. A. (1986). An even approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics* 14, 1-38.

Fowler, C. A. (1989). Real objects of speech perception: a commentary on Diehl and Kluender. *Ecological Psychology* 1, 145-160.

Fox, E. (1996). Cross-language priming from ignored words: evidence for a common representational system in bilinguals. *Journal of Memory and Language* 35, 353-370.

Fox, R. A., J. E. Flege & M. J. Munro (1995). The perception of English and Spanish vowels by native English and Spanish listeners: a multidimensional scaling analysis. *Journal of the Acoustical Society of America* 97, 2540–2550.

Francis, W. (1999). Cognitive integration of language and memory in bilinguals: semantic representation. *Psychological Bulletin* 125, 193-222.

Frieda, E. M., C. W. Amanda, J. E. Flege & M. E. Sloane (1999). Adults' perception of native and nonnative vowels: implications for the perceptual magnet effect. *Perception & Psychophysics* 61, 561–577.

Gerrits, E. (2001). *The categorisation of speech sounds by adults and children.* Doctoral Dissertation, University of Utrecht.

Goodsitt, J. V., J. L. Morgan & P. K. Kuhl (1993). Perceptual strategies in prelingual speech segmentation. *Journal of Child Language* 20, 229-252.

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds "l" and "r". *Neuropsychologia* 9, 317-323.

Gottfried, T. L. (1984). Effects of consonant context on the perception of French vowels. *Journal of Phonetics* 12, 91-114.

Gottfried, T. L & P. S. Beddor (1988). Perception of temporal and spectral information in French vowels. *Language and Speech* 31, 57–75.

Grasseger, H. (1991). Perception and production of Italian plosives by Italian learners. *Actes du XIIème Congrès international des sciences phonétiques,* Vol. 5, 290-293. Aix-en-Provence: Université de Provence, Service des Publications.

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition* 1, 67-81.

Grosjean, F. (2000). The bilingual's language modes. In J. Nicol (ed.), *One Mind, Two Languages: Bilingual Language Processing*, 1-22. Oxford: Blackwell.

Guenther, F. H. & M. N. Gjaja (1996). The perceptual magnet effect as an emergent property of neural map formation. *Journal of the Acoustical Society of America* 100, 1111-1121.

Guenther, F. H., F. T. Husain, M. A. Cohen & B. G. Shinn-Cunningham (1999). Effects of categorization and discrimination training on auditory perceptual space. *Journal of the Acoustical Society of America* 106, 2900-2912.

Guion, S., J. E. Flege, R. Akahane-Yamada & J. Pruitt (2000). An investigation of current models of second language speech perception: the case of Japanese adults' perception of English consonants. *Journal of the Acoustical Society of America* 107, 2711–2724.

Hale, M. & C. Reiss (1998). Formal and empirical arguments concerning phonological acquisition. *Linguistic Inquiry* 29, 656-683.

Halle, P. A., C. T. Best & A. Levitt (1999). Phonetic versus phonological influences on French listeners' perception of American English approximants. *Journal of Phonetics* 27, 281-306.

Hammanberg, B. (1997). Conditions on transfer in phonology. In A. James & J. Leather (eds.), *Second Language Speech: Structure and Process,* 161-180. Berlin: Mouton de Gruyter.

Hancin-Bhatt, B. & R. Bhatt (1997). Optimal L2 syllables: interactions of transfer and development effects. *Studies in Second Language Acquisition* 19, 331-378.

Hayes, B. (1999). Phonetically driven phonology: the role of Optimality Theory and inductive grounding. In M. Darnell, E. Moravcsik, M. Noonan, F. Newmeyer & K. Wheatley (eds.), *Functionalism and Formalism in Linguistics*, 243–285. Amsterdam: John Benjamins.

Helmholtz, H. von (1910). *Handbuch der Physiologischen Optik*, Vol. 3. Hamburg: Leopold Voss.

Hillenbrand, J. M., M. J. Clark & R. A. Houde (2000). Some effects of duration on vowel recognition. *Journal of the Acoustical Society of America* 108, 3013-3022.

Hume, E. & K. Johnson (2001a). *The Role of Speech Perception in Phonology.* New York: Academic Press.

Hume, E. & K. Johnson (2001b). A model of the interplay of speech perception and phonology. In E. Hume & K. Johnson (eds.), *The Role of Speech Perception in Phonology*, 3-26. New York: Academic Press.

Hyltenstam K. & N. Abrahamsson (2003). Maturational constraints in SLA. In C. Doughty & M. Long (eds.), *Handbook of Second Language Acquisition*, 104-129. Cambridge, MA: Blackwell.

Hyman, L. M. (2001). The limits of phonetic determinism in phonology: *NC revisited. In E. Hume & K. Johnson (eds.), *The Role of Speech Perception in Phonology*, 141-186. New York: Academic Press.

Ingram, J. C. L. & S. G. Park (1997). Cross-language vowel perception and production by Japanese and Korean learners of English. *Journal of Phonetics* 25, 343–370.

Iverson, P. & P. K. Kuhl (1995). Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *Journal of the Acoustical Society of America* 97, 553–562.

Iverson, P. & P. K. Kuhl (1996). Influences of phonetic identification and category goodness on American listeners' perception of /r/ and /l/. *Journal of the Acoustical Society of America* 99, 1130–1140.

Jacquemot, C., P. Pallier, D. LeBihan, S. Dehaene & E. Dupoux (2003). Phonological grammar shapes the auditory cortex: a Functional Magnetic Resonance Imaging study. *Journal of Neuroscience* 23, 9541-9546.

Jakobson, R., G. M. Fant & M. Halle (1963). *Preliminaries to Speech Analysis*. Cambridge, MA: MIT Press.

Jared, D. & J. F. Kroll (2001). Do bilinguals activate phonological representations in one or both of their languages when naming words? *Journal of Memory and Language* 44*,* 2-31.

Johnson K. & J. W. Mullenix (1997). Complex representations used in speech perception. In K. Johnson & J. W. Mullenix (eds.), *Talker Variability in Speech Processing,* 145-165. San Diego: Academic Press.

Jones, C. (2003). *Development of phonological categories in children's perception of final voicing*. Doctoral Dissertation. University of Massachusetts at Amherst.

Jusczyk, P. W. (1986). Toward a model of the development of speech perception. In J. S. Perkell & D. H. Klatt (eds.), *Invariance and Variability in Speech Processes,* 1–35. Hillsdale, NJ: Lawrence Erlbaum.

Jusczyk, P. W. (1992). Developing phonological categories from the speech signal. In C. A. Ferguson, L. Menn & C. Stoel-Gammon (eds.), *Phonological Development: Models, Research, Implications*. Timonium, MD: York Press.

Jusczyk, P. W. (1997). *The Discovery of Spoken Language*. Cambridge, MA: MIT Press.

Jusczyk, P. W., A. Cutler and N. J. Redanz (1993). Infants' preference for the predominant stress patterns of English words. *Child Development* 64, 675-687.

Keating, P. A. (1984). Phonetic and phonological representation of stop consonant voicing. *Language* 60, 286-319.

Keating, P. A., M. J. Mikos & W. F. Ganong (1981). A cross-language study of range of voice onset time in the perception of initial stop voicing. *Journal of the Acoustical Society of America* 70, 1261-1271.

Kenstowicz, M. (2001). The role of perception in loanword phonology. *Studies in African Linguistics* 32, 95-112.

Kewly-Port, D. (1995). Thresholds for formant-frequency discrimination of vowels in consonantal context. *Journal of the Acoustical Society of America* 97, 3139-3146.

Kingston J. (2003a). Learning foreign vowels. *Language and Speech* 46, 295-349.

Kingston J. (2003b). What are phonetic categories? In M. J. Sole, D. Recansens & J. Romero (eds.), *Proceedings of the 15th International Congress of Phonetic Sciences*, 285-288. Barcelona: Causal Productions

Klein, W. (1986). *Second Language Acquisition*. Cambridge University Press.

Kroll, J. F. & G. Sunderman (2003). Cognitive processes in second language acquisition: the development of lexical and conceptual representations. In C. Doughty & M. Long (eds.), *Handbook of Second Language Acquisition*, 104-129. Cambridge, MA: Blackwell.

Kuhl, P. K. (1991). Human adults and human infants show a 'perceptual magnetic effect' for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics* 50, 93–107.

Kuhl, P. K. (2000). A new view of language acquisition. *Proceedings of the National Academy of Sciences USA* 97, 11850–11857.

Lado, R. (1957). Linguistics across cultures. In *Language Transfer in Language Learning*, 21-32. Rowley, MA: Newbury House.

Lahiri, A. & W. Marslen-Wilson (1991). The mental representation of lexical form: a phonological approach to the recognition lexicon. *Cognition* 38, 245-294.

Lambert, W. E. (1969). Psychological studies of the interdependencies of the bilingual's two languages. In: J. Puhvel (ed.), *Substance and Structure of Language*. Los Angeles: University of California Press

Leather, J. (1999). Second-language speech research: an introduction". In J. Leather (ed.), *Phonological Issues in Language Learning*, 1-58. Oxford: Basil Blackwell.

Lenneberg, E. (1967). *Biological Foundations of Language*. New York: Wiley.

Liberman, A. M., P. C. Delattre & F. S. Cooper (1958). Some cues for the distinction between voiced and voiceless stops in initial position. *Language and Speech* 1, 153–167.

Liberman, A. M. & Mattingley (1985). The motor theory of speech perception. *Cognition* 21, 1-36.

Lisker, L., A. S. Abramson, F. S. Cooper & M. H. Schvey (1969). Transillumination of the larynx in running speech. *Journal of the Acoustical Society of America* 45, 1544–1546.

Llisterri, J. (1995) Relationships between speech production and speech perception in a second language. In K. Elenius & P. Branderud (eds.), *Proceedings of the 13th International Congress of Phonetic Sciences,* Vol. 4, 92-99. Stockholm: KTH / Stockholm University.

Long, M. H. (1990). Maturational constraints on language development. *Studies in Second Language Acquisition* 12, 251-285.

Luce, P. A., S. D. Goldinger, E. T. Auer & M. S. Vitevitch (2000). Phonetic priming, neighborhood activation, and PARSYN. *Perception & Psychophysics* 62, 615-625.

Mack, M. (1989). Consonant and vowel perception and production: early English-French bilinguals and English monolinguals. *Perception & Psychophysics* 46, 187-200.

Major, R. C. (1987). Phonological similarity, markedness, and rate of L2 acquisition. *Studies in Second Language Acquisition* 9, 63-82

Major, R. C. (2001). *Foreign Accent*. Mahwah, NJ: Lawrence Erlbaum Associates.

Major, R. C. (2002). The ontogeny and phylogeny of second language phonology. In J. Leather & A. James (eds.), *New Sounds 2000: Proceedings of the 4th International Symposium on the Acquisition of Second-Language Speech,* 223–235. Klagenfurt: University of Klagenfurt.

Major, R. C. (2002). The phonology of the L2 user. In V. Cook (ed.) *Portraits of the L2 User*, 65-92. Clevedon: Multilingual Matters.

Massaro, D. W. (1987). *Perceiving Talking Faces: From Speech Perception to Behavioural Principle*. Cambridge, MA: MIT Press.

Maye, J., J. F. Werker & L. A. Gerken (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition 82*, B101–B111.

McQueen, J. M. (2004). Speech perception. In K. Lamberts and R. Goldstone (eds.), *The Handbook of Cognition*. London: Sage Publications.

McQueen, J. M. & A. Cutler (1997). Cognitive processes in speech perception. In W. J. Hardcastle and J. Laver (eds.), *The Handbook of Phonetic Sciences*, 566-585. Oxford: Blackwell.

Mehler, J. (1981). The role of syllables in speech processing: infant and adult data. *Philosophical Transactions of the Royal Society* 295, 333-352.

Miller, J. L. & E. R. Dexter (1988). Effects of speaking rate and lexical status on phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance* 14, 369-378.

Miller, J. L. & F. Grosjean (1997). Dialect effects in vowel perception: the role of temporal information in French. *Language and Speech* 40, 277-288.

Miller, J. L. & E. R. Dexter (1988). Effects of speaking rate and lexical status on phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance* 14, 369-378.

Miyawaki, K., W. Strange, R. R. Verbrugge, A. M. Liberman, J. J. Jenkins & O. Fujimura (1975). An effect of linguistic experience: the discrimination of [r] and [l] by native speakers of Japanese and English. *Perception & Psychophysics* 18, 331-340.

Morrison, G. S. (2002). Perception of English /i/ and /ɪ/ by Japanese and Spanish listeners: longitudinal results. In G. S. Morrison & L. Zsoldos (eds.), *Proceedings of the North West Linguistics Conference 2002*, 29-48. Burnaby, BC: Simon Fraser University Linguistics Graduate Student Association.

Nakai, S. (1998). The effect of vowel prototypicality and extremity on discrimination sensitivity. *Journal of the Acoustical Society America* 103, 2983.

Neary, T. M. (2001). Phoneme-like units and speech perception. *Language and Cognitive Processes* 16, 673-681.

Neufeld, G. G. (1988). Phonological asymmetry in second language learning and performance. *Language Learning* 38, 531-559.

Nittrouer, S. (1996). Discriminability and perceptual weighting of some acoustic cues to speech perception by 3-year-olds. *Journal of Speech and Hearing Research* 39, 278-297.

Nittrouer, S. (2001). Challenging the notion of innate phonetic boundaries. *Journal of the Acoustical Society of America* 110, 1598-1605.

Nittrouer, S. & M.E. Miller (1997). Developmental weighting shifts for noise components of fricative-vowel syllables. *Journal of the Acoustical Society of America* 102, 572-580.

Nittrouer, S., C. Manning & G. Meyer (1993). The perceptual weighting of acoustic cues changes with linguistic experience. *Journal of the Acoustical Society of America* 94, 1865.

Norris, D. (1994). Shortlist: a connectionist model of continuous speech recognition. *Cognition* 52, 189-234.

Norris, D., J. M. McQueen & A. Cutler (2000). Merging information in speech recognition: feedback is never necessary. *Behavioral and Brain Sciences* 23, 299-325.

Parker, E. M., R. L. Diehl & K. R. Kluender (1986). Trading relations in speech and nonspeech. *Perception & Psychophysics* 39, 129-142.

Pater, J. (2004). Bridging the gap between receptive and productive development with minimally violable constraints. In R. Kager, J. Pater & W. Zonneveld (eds.), *Constraints in Phonological Acquisition*, 219-244. Cambridge: Cambridge University Press.

Penfield, W. & L. Roberts (1959). *Speech and Brain Mechanisms*. Princeton: Princeton University Press.

Peterson, G. E. & H. L. Barney (1952). Control methods used in a study of vowels. *Journal of the Acoustical Society of America* 24, 175-184.

Peterson, G. E. & I. Lehiste (1960). Duration of syllable nuclei in English. *Journal of the Acoustical Society of America* 32, 693–703.

Picard, M. (1987). *An Introduction to the Comparative Phonetics of English and French in North America*. Amsterdam: John Benjamins.

Piske, T., I. Mackay & J. Flege (2001). Factors affecting degree of foreign accent in an L2: a review. *Journal of Phonetics* 29, 191-215.

Pisoni, D. B. & P. Luce (1987). Acoustic-phonetic representations in word recognition. *Cognition* 25, 21-52.

Pisoni, D. B., S. E. Lively & J. S. Logan (1994). Perceptual learning of nonnative speech contrasts: implications for theories of speech perception. In J. Goodman & H. C. Nusbaum (eds.), *Development of Speech Perception: The Transition from Recognizing Speech Sounds to Spoken Words*, 121-166. Cambridge, MA: MIT Press.

Pitt, M. A. & J. M. McQueen (1998). Is compensation for coarticulation mediated by the lexicon? *Journal of Memory and Language* 39, 347-370.

Polivanov, E. D. (1931). La perception des sons d'une langue étrangère. *Travaux du Cercle linguistique de Prague* 4, 79–96.

Polka, L. (1995). Linguistic influence in adult perception of non-native vowel contrasts. *Journal of the Acoustical Society of America* 97, 1286–1296.

Polka, L. & J. F. Werker (1994). Developmental changes in the perception of non-native vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance* 20, 421-435.

Prince, A. & P. Smolensky (1993). *Optimality Theory: Constraint Interaction in Generative Grammar*. New Brunswick, NJ: Rutgers University Center for Cognitive Science.

Repp, B. H. (1982). Phonetic trading relations and context effects: new experimental evidence for a phonetic mode of perception. *Psychological Bulletin* 92, 81–110.

Rice K. & P. Avery (1991). Segmental complexity and the structure of inventories. Paper presented at the GLOW workshop on the acquisition of phonology, Leiden, The Netherlands.

Rochet, B. (1995). Perception and production of second-language speech sounds by adults. In W. Strange (ed.), *Speech Perception and Linguistic Experience*. Timonium, MD: York Press.

Saffran, J. R., R. N. Aslin & E. L. Newport (1996). Statistical learning by 8-month old infants. *Science* 274, 1926-1928.

Schacter, D. L. & B. A. Church (1992). Auditory priming: implicit and explicit memory for words and voices. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 18, 521-533.

Schmidt, A. M. (1996). Cross-language identification of consonants. Part I. Korean perception of English. *Journal of the Acoustical Society of America* 99, 3201–3211.

Schouten B., E. Gerrits & A. van Hessen (2003). The end of categorical perception as we know it. *Speech Communication* 41, 71-80.

Schwartz, B. & R. Sprouse (1996). L2 cognitive states and the Full Transfer/Full Access model. *Second Language Research* 12, 40–72.

Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics* 10, 209-231.

Sheldon, A. & W. Strange (1982). The acquisition of /r/ and /l/ by Japanese learners of English: evidence that speech production can precede speech perception. *Applied Psycholinguistics* 3, 243-261.

Steriade, D. (2001). Directional asymmetries in place assimilation. In E. Hume & K. Johnson (eds.), *The Role of Speech Perception in Phonology*. New York: Academic Press.

Stevens, K. (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, S. S., J. Volkmann & E. B. Newman (1937). A scale measurement of the psychological magnitude pitch. *Journal of the Acoustical Society of America* 8, 185-190.

Strange, W. (1995). Cross-language study of speech perception: a historical review. In W. Strange (ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research* 3-45. Timonium, MD: York Press.

Tesar, B. & P. Smolensky (1998). Learnability in Optimality Theory. *Linguistic Inquiry* 29, 229-268.

Tesar, B. & P. Smolensky (2000). *Learnability in Optimality Theory*. Cambridge, MA: MIT Press.

Trubetzkoy, N. S. (1929). Zur algemeinen theorie der phonologischen vokalsysteme. *Travaux du Cercle linguistique de Prague* 1, 39-67.

Trubetzkoy, N. S. (1969). *Principles of Phonology*. Berkeley, CA: University of California Press.

Walley, A. C. & J. E. Flege (1999). Effect of lexical status on children's and adults' perception of native and non-native vowels. *Journal of Phonetics* 27, 307–332.

Wang, W. S.-Y. (1959). Transition and release as perceptual cues for final plosives. *Journal of Speech and Hearing Research* 2, 66–73.

Wardrip-Fruin, C. (1982). On the status of temporal cues to phonetic categories: preceding vowel duration as a cue to voicing in final stop consonants. *Journal of the Acoustical Society of America* 71, 187–195.

Wardrip-Fruin, C. & S. Peach (1984). Developmental aspects of the perception of acoustic cues in determining the voicing feature of final stop consonants. *Language and Speech* 27: 367–379.

Wells, J. C. (1982a). *Accents of English 2: The British Isles*. New York: Cambridge University Press.

Wells, J. C. (1982b) *Accents of English 3: Beyond the British Isles*. New York: Cambridge University Press.

Werker, J. F. & J. S. Logan (1985). Cross-language evidence for three factors in speech perception. *Perception & Psychophysics* 37: 35-44.

Werker, J. F. & R. C. Tees (1984). Cross-language speech perception: evidence for perceptual reorganization during the first year of life. *Infant Behaviour and Development* 7, 49-63.

White, L. (2000). Second language acquisition: from initial to final state. In J. Archibald (ed.), *Second Language Acquisition and Linguistic Theory*. Oxford: Blackwell.

White L. (2003). On the nature of interlanguage representation: universal grammar in the second language. In C. Doughty & M. Long (eds.), *Handbook of Second Language Acquisition*,104-129. Malden, MA: Blackwell.

Wickelgren, W. A. (1969). Context sensitive decoding, associative memory, and serial order in (speech) behavior. *Psychological Review* 76, 1-15.

Williams, L. (1977). The perception of consonant voicing by Spanish English bilinguals. *Perception & Psychophysics* 21, 289-297.

Young-Scholten, M. (1995). The negative effects of 'positive' evidence on L2 phonology. In L. Eubank, L. Selinker & M. Sharwood Smith (eds.), *The Current State of Interlanguage: Studies in Honor of William E. Rutherford*, 107-121. Amsterdam: John Benjamins.

AGRADECIMIENTOS
ACKNOWLEDGEMENTS
DANKWOORD

For their insightful comments, suggestions, questions, ideas, etc., on the manuscript.

Paul Boersma
Janet Grijzenhout
Rene Kager
Wim Zonneveld

F*or serving as members of the reading committee that approved the manuscript and provided helpful comments to improve it*

Anne Cutler
Janet Grijzenhout
Wolfgang Herrlitz
Bert Schouten
Mike Sharwood Smith

For editorial work above and beyond the call of duty

Marc Picard

For layout and formatting help

Andreia Rauber

For translating the English summary to Spanish and Dutch respectively

Jorge Iván Perez, Andreia Rauber, and María Verónica Rivas Zambrano
Paul Boersma, and Annemarie Kerkhoff

For making my research project financially possible

Jan Don
Netherlands Organization for Scientific Research (NWO), Travel Grants Section
Linda Polka
Utrecht Institute of Linguistics

For funding numerous conference presentations and research visits

Utrecht Institute of Linguistics

For their general interest, comments, and inspiring ideas

Ricardo Bion
Paul Boersma
Ivana Brasileiro
Ellen Broselow
Rachel Hayes-Harb
Silvia Lipski
Geoff Morrison
Denize Nobre Oliveira
Andreia Rauber

For trusting in my research potential

Paul Boersma
Rene Kager
Linda Polka

For moral support during the writing period

Marco Escudero Egúsquiza
Olaf Koeneman
Rosa Luz Neyra Castilla

¡Un millón de gracias!
Thanks a million!
Duizend maal dank!

Curriculum Vitae

Paola Escudero was born in Lima, Perú on December 5, 1976. She holds a BA in Linguistics and Literature from the Pontificia Universidad Católica del Perú (1999) and an MSc in Applied Linguistics from the University of Edinburgh (2001).

From 1999 to 2000, she worked as a Spanish lecturer, an editorial assistant for an academic journal, and a project director for a speech technology company in Edinburgh. She also obtained a research and teaching fellowship from the University of Reading, a fellowship from Dr. Linda Polka's Natural Sciences and Engineering Research Grant, and a Graduate Fellowship from the Faculty of Graduate Studies and Research at McGill University. From October 2001 to March 2005, she worked as a PhD researcher at the Utrecht Institute of Linguistics (UiL OTS) and conducted research that resulted in the present study (2005). As part of her PhD work, she carried out research projects at McGill University's School of Communication Sciences and Disorders. During her affiliation with the UiL OTS, she developed and taught a course entitled *Perception and Production in L2 Phonological Acquisition* and conducted research in Perú which was funded by the Netherlands Organisation for Scientific Research (NWO).

Since 2004, she has supervised international and Dutch MA and PhD students and currently holds a post-doctoral research position at the Institute of Phonetics Sciences of the University of Amsterdam, working on a project entitled *Explaining L2 Speech Perception* which is part of Professor Paul Boersma's VIDI project. She is also teaching Speech Perception/Production and Language Development in the MA program in General Linguistics at the University of Amsterdam.