

Lecture Notes in Electrical Engineering 347

Prabin Kumar Bora
S.R. Mahadeva Prasanna
Kandarpa Kumar Sarma
Navajit Saikia
Editors

Advances in Communication and Computing

 Springer

Lecture Notes in Electrical Engineering

Volume 347

Board of Series editors

Leopoldo Angrisani, Napoli, Italy
Marco Arteaga, Coyoacán, México
Samarjit Chakraborty, München, Germany
Jiming Chen, Hangzhou, P.R. China
Tan Kay Chen, Singapore, Singapore
Rüdiger Dillmann, Karlsruhe, Germany
Haibin Duan, Beijing, China
Gianluigi Ferrari, Parma, Italy
Manuel Ferre, Madrid, Spain
Sandra Hirche, München, Germany
Faryar Jabbari, Irvine, USA
Janusz Kacprzyk, Warsaw, Poland
Alaa Khamis, New Cairo City, Egypt
Torsten Kroeger, Stanford, USA
Tan Cher Ming, Singapore, Singapore
Wolfgang Minker, Ulm, Germany
Pradeep Misra, Dayton, USA
Sebastian Möller, Berlin, Germany
Subhas Mukhopadhyay, Palmerston, New Zealand
Cun-Zheng Ning, Tempe, USA
Toyoaki Nishida, Sakyo-ku, Japan
Bijaya Ketan Panigrahi, New Delhi, India
Federica Pascucci, Roma, Italy
Tariq Samad, Minneapolis, USA
Gan Woon Seng, Nanyang Avenue, Singapore
Germano Veiga, Porto, Portugal
Haitao Wu, Beijing, China
Junjie James Zhang, Charlotte, USA

About this Series

“Lecture Notes in Electrical Engineering (LNEE)” is a book series which reports the latest research and developments in Electrical Engineering, namely:

- Communication, Networks, and Information Theory
- Computer Engineering
- Signal, Image, Speech and Information Processing
- Circuits and Systems
- Bioengineering

LNEE publishes authored monographs and contributed volumes which present cutting edge research information as well as new perspectives on classical fields, while maintaining Springer’s high standards of academic excellence. Also considered for publication are lecture materials, proceedings, and other related materials of exceptionally high quality and interest. The subject matter should be original and timely, reporting the latest research and developments in all areas of electrical engineering.

The audience for the books in LNEE consists of advanced level students, researchers, and industry professionals working at the forefront of their fields. Much like Springer’s other Lecture Notes series, LNEE will be distributed through Springer’s print and electronic publishing channels.

More information about this series at <http://www.springer.com/series/7818>

Prabin Kumar Bora · S.R. Mahadeva Prasanna
Kandarpa Kumar Sarma · Navajit Saikia
Editors

Advances in Communication and Computing

 Springer

Editors

Prabin Kumar Bora
Department of Electronics and Electrical
Engineering
Indian Institute of Technology Guwahati
Guwahati, Assam
India

Kandarpa Kumar Sarma
Department of Electronics and
Communication Technology
Gauhati University
Guwahati, Assam
India

S.R. Mahadeva Prasanna
Department of Electronics and Electrical
Engineering
Indian Institute of Technology Guwahati
Guwahati, Assam
India

Navajit Saikia
Department of Electronics and
Telecommunication Engineering
Assam Engineering College
Guwahati, Assam
India

ISSN 1876-1100

ISSN 1876-1119 (electronic)

Lecture Notes in Electrical Engineering

ISBN 978-81-322-2463-1

ISBN 978-81-322-2464-8 (eBook)

DOI 10.1007/978-81-322-2464-8

Library of Congress Control Number: 2015939666

Springer New Delhi Heidelberg New York Dordrecht London

© Springer India 2015

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

Springer (India) Pvt. Ltd. is part of Springer Science+Business Media (www.springer.com)

*The volume is dedicated to everyone
associated with communication and
computing.*

Preface

The volume *Advances in Communication and Computing* is a compilation of papers presented in the 2014 National Workshop on Advances in Communication and Computing (WACC) held at Assam Engineering College, Guwahati, India during September 26–27, 2014. This volume presents three invited chapters, one each on Network Security, Biomedical Signal Processing, and Nanotechnology. These areas have great significance in the current engineering practices. The rest of the chapters comprise of papers in the areas of Communication System Engineering, Signal Processing, VLSI & Embedded System, and Computer Engineering & Information Theory.

The editors offer sincere gratitude to IEEE Kolkata Section, The Institution of Engineers (India) Assam State Centre, IETE Guwahati Centre, Software Technology Parks of India, Centre for Development of Advanced Computing, IBM, National Instruments, and SpeechWareNet for their technical support. We also extend our gratitude to Oil India Limited, Assam Electronics Development Corporation Limited, North Eastern Electric Power Corporation Limited, Indian Oil Corporation Limited, Numaligarh Refinery Limited, State Bank of India, and DS Systems for sponsoring the workshop. The editors are thankful to Mr. Manasjyoti Bhuyan and Mr. Manash Pratim Bhuyan, and also to the editing and printing support staff of Springer for making the compilation possible.

We sincerely hope that this volume will inspire researchers.

Guwahati, India
September 2014

Prabin Kumar Bora
S.R. Mahadeva Prasanna
Kandarpa Kumar Sarma
Navajit Saikia

Acknowledgments

The editors acknowledge the contributions by the authors, reviewers, and everyone who has contributed directly or indirectly to this book volume.

Committees

Technical Committee

Chair

Prof. Prabin Kumar Bora, IIT Guwahati

Area Chairs

Prof. Diganta Goswami, IIT Guwahati

Area: Computer Science & Engineering

Prof. Ratnajit Bhattacharjee, IIT Guwahati

Area: Electronics & Communication Engineering

Members

Dr. A. Rajesh, IIT Guwahati

Dr. Aditya B. Kandali, JEC

Dr. Amrita Ganguly, AEC

Dr. Amarendra K. Sarma, IIT Guwahati

Dr. Amitabh Chatterjee, IIT Guwahati

Prof. Anjana Kakoti Mahanta, Gauhati University

Dr. Arijit Sur, IIT Guwahati

Dr. Bani Kanta Talukdar, AEC

Dr. Bhabesh Deka, Tezpur University

Mr. Bijuphukan Bhagabati, AMTRON

Dr. Bubu Bhuyan, NEHU

Prof. Chitralkha Mahanta, IIT Guwahati

Dr. Debadatta Pati, BCET

Mr. Diganta Kumar Gogoi, Oil India Ltd.

Mr. Dipak Das, Gauhati University

Mr. Dipjyoti Sharma, DBCET Guwahati

Prof. Dhruva K. Bhattacharyya, Tezpur University

Prof. Durlav Hazarika, AEC
Mr. Gautam Chakraborty, NERIM Guwahati
Mr. Gayadhar Pradhan, NIT Patna
Dr. Guruprasad Khataniar, KKHSOU
Prof. Harshal B. Nemade, IIT Guwahati
Dr. Hemanta Kumar Kalita, NEHU
Dr. Hitesh Tahbildar, AEI
Dr. Indrani Dakua, DBCET
Prof. Indrajit Chakraborty, IIT Kharagpur
Prof. Indranil Sengupta, IIT Kharagpur
Prof. Kanak Ch. Sarma, Gauhati University
Dr. Kandarpa Kumar Sarma, Gauhati University
Dr. Kannan Karthik, IIT Guwahati
Mr. Kaustav Bhattacharya, DBCET Guwahati
Dr. K. Manglem Singh, NIT Manipur
Dr. K. Rakesh Singh, IIT Guwahati
Mr. Kumaresh Sarma, Gauhati University
Dr. L. Joyprakash Singh, NEHU
Dr. L.N. Sharma, IIT Guwahati
Dr. Manas K. Bhuyan, IIT Guwahati
Mr. Manasjyoti Bhuyan, Gauhati University
Prof. M. Bhuyan, Tezpur University
Ms. Moushumi Baruah, AEC
Ms. Mousmita Sarma, SpeechWareNet
Dr. Nalin Behari Dev Choudhury, NIT Silchar
Ms. Nandini Bhattacharya
Prof. Nityananda Sarma, Tezpur University
Dr. Prithwijit Guha, IIT Guwahati
Mr. Prabir Kr. Das, STPI
Dr. Prasanta Gogoi, NMIT Bangalore
Dr. Rohit Sinha, IIT Guwahati
Prof. Roy P. Paily, IIT Guwahati
Dr. Rubul Hussain Laskar, NIT Silchar
Dr. Rupaban Subadar, NEHU
Prof. Samarendra Dandapat, IIT Guwahati
Mr. Samir Kumar Borgohain, NIT Silchar
Dr. Sarat Saharia, Tezpur University
Dr. Shaik Rafi Ahamed, IIT Guwahati
Prof. Shakuntala Laskar, DBCET Guwahati
Prof. Shyamanta M. Hazarika, Tezpur University
Dr. Subhash C. Arya, NEHU
Prof. Sukumar Nandi, IIT Guwahati
Dr. Sunandan Baruah, DBCET Guwahati
Dr. Swarup Roy, NEHU
Prof. S.R. Mahadeva Prasanna, IIT Guwahati

Dr. S.R. Nirmala, Gauhati University
Dr. Tony Jacob, IIT Guwahati
Mr. Vaskar Deka, Gauhati University

Editorial Committee

Advisors

Prof. Indrajit Chakraborty, IIT Kharagpur
Prof. Indranil Sengupta, IIT Kharagpur
Dr. Jatindra Kr. Deka, IIT Gauhati
Prof. Kanak Ch. Sarma, Gauhati University
Mr. Prabir Kr. Das, Software Technology Parks of India
Dr. Pradeep Kumar Sinha, C-DAC
Prof. Samarendra Dandapat, IIT Guwahati
Prof. Sukumar Nandi, IIT Guwahati
Prof. Anup Kr. Gogoi, IIT Guwahati

Chief Editor

Prof. Prabin Kumar Bora, IIT Guwahati

Deputy Chief Editor

Prof. S.R. Mahadeva Prasanna, IIT Guwahati

Editors

Dr. Kandarpa Kumar Sarma, Gauhati University
Dr. Navajit Saikia, Assam Engineering College

Members

Mr. Manasjyoti Bhuyan, Gauhati University
Mr. Manash Pratim Bhuyan, Assam Engineering College

Organizing Committee

Chief Patron

Dr. Atul Bora
Director of Technical Education, Assam
Principal, Assam Engineering College

Patron

Prof. Apurba Kr. Kalita, Assam Engineering College

Chief Convener

Dr. Navajit Saikia, Assam Engineering College

Conveners

Mr. Gunajit Kalita, Assam Engineering College

Ms. Rashi Borgohain, Assam Engineering College

Advisors

Prof. Ashok Baruah, Assam Engineering College

Dr. Bichitra Kalita, Assam Engineering College

Dr. Damodar Agarwal, Assam Engineering College

Dr. Palash Jyoti Hazarika, Assam Engineering College

Dr. Ranjit Kumar Dutta, Assam Engineering College

Organizing Secretaries

Mr. Bijoy Goswami, Assam Engineering College

Mr. Partha Bhowmik, Assam Engineering College

Technical Secretaries

Mr. Mridul Jyoti Roy, Assam Engineering College

Ms. Riju Kalita, Assam Engineering College

Publicity Secretaries

Mr. Manash Pratim Bhuyan, Assam Engineering College

Ms. Rajashree Konwar, Assam Engineering College

Members

Mr. Apurba Kalita, Assam Engineering College

Mr. Gagan Kalita, Assam Engineering College

Mr. Gokul Ch. Sarma, Assam Engineering College

Mr. Jogananda Goswami, Assam Engineering College

Ms. Dharitri Talukdar, Assam Engineering College

Ms. Pankhi Das, Assam Engineering College

Mr. Pradip Baishya, Assam Engineering College

Ms. Ruchira Mazumder, Assam Engineering College

Mr. Suresh Ch. Haloi, Assam Engineering College

Mr. Utpal Sarma, Assam Engineering College

Student Members

Mr. Pallab Nath, Assam Engineering College

Mr. Pulak Jyoti Roy, Assam Engineering College

Mr. Pallav Jyoti Bora, Assam Engineering College

Contents

Part I Invited Papers

1	A Discrete Event System Based Approach for Obfuscated Malware Detection	3
	Chinmaya K. Patanaik, Ferdous A. Barbhuiya, Santosh Biswas and Sukumar Nandi	
2	Quantification of Diagnostic Information from Electrocardiogram Signal: A Review	17
	S. Dandapat, L.N. Sharma and R.K. Tripathy	
3	ZnO Nanostructures for Alternate Energy Generation	41
	Sunandan Baruah	

Part II Accepted Papers

4	Wavelet and Learning Based Image Compression Systems	61
	Mayuri Kalita and Kandarpa Kumar Sarma	
5	Quantifying Clinical Information in MECG Using Sample and Channel Convolution Matrices	73
	R.K. Tripathy and S. Dandapat	
6	Application of SHAZAM-Based Audio Fingerprinting for Multilingual Indian Song Retrieval	81
	S. Sri Ranjani, V. Abdulkareem, K. Karthik and P.K. Bora	
7	Multilevel-DWT-Based Image Denoising Using Adaptive Neuro-Fuzzy Inference System	93
	Torali Saikia and Kandarpa Kumar Sarma	

8	Active Learning Using Fuzzy k-NN for Cancer Classification from Microarray Gene Expression Data	103
	Anindya Halder, Samrat Dey and Ansuman Kumar	
9	A Practical Under-Sampling Pattern for Compressed Sensing MRI	115
	Bhabesh Deka and Sumit Datta	
10	Semi-automatic Segmentation and Marking of Pitch Contours for Prosodic Analysis	127
	Biswajit Dev Sarma, Meghamallika Sarma and S.R.M. Prasanna	
11	Semi-automatic Syllable Labelling for Assamese Language Using HMM and Vowel Onset-Offset Points.	139
	Biswajit Dev Sarma, Mousmita Sarma and S.R.M. Prasanna	
12	Block Matching Algorithms for Motion Estimation: A Performance-Based Study	149
	Hussain Ahmed Choudhury and Monjul Saikia	
13	Stressed Speech Recognition Using Similarity Measurement on Inner Product Space	161
	Bhanu Priya and S. Dandapat	
14	An Analysis of Indoor Power Line Network as a Communication Medium Using ABCD Matrices Effect of Loads on the Transfer Function of Power Line	171
	Banty Tiru, Rubi Baishya and Utpal Sarma	
15	Online and Offline Handwriting Recognition of Assamese Language—A Preliminary Idea	183
	Keshab Nath and Subhash Basishtha	
16	Speaker Verification for Variable Duration Segments and the Effect of Session Variability	193
	Rohan Kumar Das and S.R.M. Prasanna	
17	Two-Dimensional Processing of Multichannel ECG Signals for Efficient Exploitation of Inter and Intra-Channel Correlation	201
	Anurag Singh and S. Dandapat	

18 DCT-Based Linear Regression Approach for 12-Lead ECG Synthesis 211
Jiss J. Nallikuzhy and S. Dandapat

19 Design, Simulation, and Performance Evaluation of a High Temperature and Low Power Consumption Microheater Structure for MOS Gas Sensors 221
Kaushik Das and Priyanka Kakoty

20 Experimental Analysis on the Performance of a New Preprocessing Method Used in Data Compression 231
P. Khanikar, M.P. Bhuyan, R.R. Baruah and H. Sarma

21 Modeling a Nano Cylindrical MOSFET Considering Parabolic Potential Well Approximation 241
Jyotisikha Deka and Santanu Sharma

22 SVD Analysis on Reduced 3-Lead ECG Data 253
Sibasankar Padhy and S. Dandapat

23 Large Library-Based Regression Test Cases Minimization for Object-Oriented Programs 261
Swapan Kumar Mondal and Hitesh Tahbaldar

Author Index 277

Subject Index 279

Editors and Contributors

About the Editors

Prof. Prabin Kumar Bora (Ph.D., Indian Institute of Science, Bangalore) is currently a Professor in the Department of Electronics and Electrical Engineering, IIT Guwahati, India. His research interests include image processing, computer vision, and application of signal processing techniques to process physiological and communication signals. He has guided 7 Ph.D. students and over 50 Master's students. He has co-authored over 100 research papers in journals and conference proceedings.

Dr. S.R. Mahadeva Prasanna is currently a Professor in the Department of Electronics and Electrical Engineering, Indian Institute of Technology (IIT) Guwahati, India. He was Associate Professor from August 2007 to April 2012 and Assistant Professor from August 2004 to August 2007 in the same department. He obtained Ph.D. from IIT Madras, M.Tech. from NITK Surathkal, and B.E. from SSIT Tumkur (Bangalore University) in 2004, 1997, and 1994, respectively. His teaching interests include signal processing and speech processing, and research interests include speech and handwriting processing. He has supervised 5 Ph.D. theses in the speech processing area. He has over 100 publications to his credit in different conferences and journals. He successfully executed several funded projects in the speech processing area.

Dr. Kandarpa Kumar Sarma currently an Associate Professor in the Department of Electronics and Communication Technology, Gauhati University, Guwahati, Assam, India, has over 17 years of professional experience. He has been into teaching UG/PG level electronics courses including soft computing, mobile communication, pattern recognition, digital signal, and image processing. He obtained M.Tech. in Signal Processing in 2005 and subsequently completed Ph.D. in the area of Mobile Communication both from Indian Institute of Technology Guwahati, Guwahati, Assam, India. He has authored seven books, several book chapters, and several peer-reviewed research papers in international conference proceedings and

journals. His areas of interest are soft-computation and its applications, knowledge-aided system design, MIMO, mobile communication, antenna design, speech processing, document image analysis and signal processing applications in high energy physics, neurocomputing and computational models for social-science applications. He is senior member IEEE (USA), Fellow IETE (India), Member International Neural Network Society (INNS, USA), Life Member ISTE (India), and Life Member CSI (India). He serves as an Editor-in-Chief of International Journal of Intelligent System Design and Computing (IJISDC, UK), guest editor of several international journals, reviewer of over 30 international journals and reviewer/TPC member of over 100 international conferences.

Dr. Navajit Saikia is presently an Assistant Professor in the Department of Electronics and Telecommunication Engineering at Assam Engineering College, Guwahati, India. He completed his M.Tech. and Ph.D. from IIT Guwahati. Communications and signal processing are the two areas of his teaching interests. He is also looking after the training and placement activities of the college. His research interests include signal processing, image processing, speech processing, information security, etc. He has published several research papers in journals and conference proceedings.

Contributors

V. Abdulkareem College of Engineering, Cherthala, Malappuram, India

Rubi Baishya Gauhati University, Guwahati, India

Ferdous A. Barbhuiya Indian Institute of Technology Guwahati, Guwahati, India

R.R. Baruah Gauhati University, Guwahati, India

Sunandan Baruah Assam Don Bosco University, Guwahati, India

Subhash Basishtha Assam University, Silchar, India

M.P. Bhuyan Assam Engineering College, Guwahati, India

Santosh Biswas Indian Institute of Technology Guwahati, Guwahati, India

P.K. Bora Indian Institute of Technology, Guwahati, India

Hussain Ahmed Choudhury NERIST, Itanagar, Arunachal Pradesh, India

S. Dandapat Indian Institute of Technology Guwahati, Guwahati, India

Kaushik Das Tezpur University, Tezpur, India

Rohan Kumar Das Indian Institute of Technology Guwahati, Guwahati, India

Sumit Datta Tezpur University, Tezpur, India

Bhabesh Deka Tezpur University, Tezpur, India

- Jyotisikha Deka** Assam Down Town University, Guwahati, India
- Samrat Dey** Department of Computer Science, Don Bosco College, Tura, India
- Anindya Halder** Department of Computer Application, North-Eastern Hill University, Tura Campus, Tura, India
- Priyanka Kakoty** Tezpur University, Tezpur, India
- Mayuri Kalita** Gauhati University, Guwahati, India
- K. Karthik** Indian Institute of Technology, Guwahati, India
- P. Khanikar** Assam Engineering College, Guwahati, India
- Ansuman Kumar** Department of Computer Application, North-Eastern Hill University, Tura Campus, Tura, India
- Swapan Kumar Mondal** University of Science and Technology, Meghalaya, India
- Jiss J. Nallikuzhy** Indian Institute of Technology Guwahati, Guwahati, India
- Sukumar Nandi** Indian Institute of Technology Guwahati, Guwahati, India
- Keshab Nath** Assam University, Silchar, India
- Sibasankar Padhy** Indian Institute of Technology Guwahati, Guwahati, India
- Chinmaya K. Patanaik** Indian Institute of Technology Guwahati, Guwahati, India
- S.R.M. Prasanna** Indian Institute of Technology, Guwahati, India
- Bhanu Priya** Indian Institute of Technology Guwahati, Guwahati, India
- Monjul Saikia** NERIST, Itanagar, Arunachal Pradesh, India
- Torali Saikia** Gauhati University, Guwahati, India
- Biswajit Dev Sarma** Indian Institute of Technology, Guwahati, India
- H. Sarma** Assam Engineering College, Guwahati, India
- Kandarpa Kumar Sarma** Gauhati University, Guwahati, India
- Meghamallika Sarma** Indian Institute of Technology, Guwahati, India
- Mousmita Sarma** Indian Institute of Technology Guwahati, Guwahati, India
- Utpal Sarma** Gauhati University, Guwahati, India
- L.N. Sharma** Indian Institute of Technology Guwahati, Guwahati, India
- Santanu Sharma** Tezpur University, Tezpur, India
- Anurag Singh** Indian Institute of Technology Guwahati, Guwahati, India

S. Sri Ranjani Indian Institute of Technology, Guwahati, India

Hitesh Tahbaldar Assam Engineering Institute, Guwahati, India

Banty Tiru Gauhati University, Guwahati, India

R.K. Tripathy Indian Institute of Technology Guwahati, Guwahati, India

Acronyms

ANN	Artificial Neural Network: Artificial neural networks are non-parametric models inspired by animal central nervous systems. These are particularly the brain that are capable of machine learning and pattern recognition and usually presented as systems of interconnected “neurons” that can compute values from inputs by feeding information through the network
ANFIS	Adaptive Neuro Fuzzy Inference System: Adaptive neuro fuzzy inference system is a kind of fuzzy and artificial neural network that is based on Takagi-Sugeno fuzzy inference system
ASR	Automatic Speech Recognition: Automatic speech recognition is a technology by which a computer or a machine is made to recognize the speech of a human being
AWGN	Additive White Gaussian Noise: Additive white Gaussian noise is a channel impairment to communication which is additive and shows a constant spectral density and a Gaussian distribution of amplitude
BPA	Back Propagation Algorithm: Back propagation algorithm is a common supervised method of training ANN where from a desired output, the network learns from many inputs
DCT	Discrete Cosine Transform: Discrete cosine transform expresses a finite sequence of data points in terms of a sum of cosine functions oscillating at different frequencies
DWT	Discrete Wavelet Transform: Discrete wavelet transform is any wavelet transform for which the wavelets are discretely sampled
ECG	Electrocardiography: Electrocardiography is the recording of the electrical activity of the heart
GDALMBP	Gradient Descent with Adaptive Learning Rate and Momentum Back Propagation
GDALRBP	Gradient Descent with Adaptive Learning Rate Back Propagation
GDBP	Gradient Descent Back Propagation: Gradient descent back propagation is form of ANN training method
GDLMBP	Gradient Descent with Levenberg-Marquardt Backpropagation

GDMBP	Gradient Descent with Momentum Back Propagation
GMM	Gaussian Mixture Model: Gaussian mixture model is a probabilistic model for representing the presence of subpopulations within an overall population, without requiring that an observed data-set should identify the sub-population to which an individual observation belongs where mixture distribution is Gaussian
GPPC	General-Purpose Personal Computer: General-purpose personal computers ia a computer built around a microprocessor for use by an individual having capabilities to perform wide range of tasks
HMM	Hidden Markov Model: Hidden Markov models is a statistical Markov model in which the system being modeled is assumed to be a Markov process with unobserved or hidden states. It can be considered to be the simplest dynamic Bayesian network
HTK	Hidden Markov Model Toolkit: Hidden Markov model toolkit is a portable toolkit for building and manipulating hidden Markov models. HTK is primarily used for speech recognition research although it has been used for numerous other applications including research into speech synthesis, character recognition and DNA sequencing. HTK is in use at hundreds of sites worldwide
IC	Integrated Circuit: Integrated circuit is a set of electronic circuits embedded on one small chip of semiconductor material, normally silicon
IEEE	Institute of Electronics and Electrical Engineering: Institute of Electronics and Electrical Engineering is a professional association headquartered in New York City that is dedicated to advancing technological innovation and excellence
LPC	Linear Predictive Coding: Linear predictive coding is a tool used mostly in audio signal processing and speech processing for representing the spectral envelope of digital signal of speech in compressed form, using the information of a linear predictive model
LPCC	Linear Prediction Cepstral Coefficients: Linear prediction cepstral coefficients are the coefficients that can be found by converting the Linear Prediction coefficients into cepstral coefficients
LTI	Linear Time Invariant: Linear time invariant refers to the behavior of a system whose response varies linearly with time
LVQ	Learning Vector Quantization: Learning vector quantization is a prototype-based supervised classification algorithm and is the supervised counterpart of vector quantization systems
MFCC	Mel Frequency Cepstral Coefficient: Mel frequency cepstral coefficients are coefficients that collectively make up an Mel Frequency Cepstrum which is a representation of the short-term power spectrum of a sound, based on a linear cosine transform of a log power spectrum on a nonlinear mel scale of frequency

MIDI	Musical Instrument Digital Interface: Musical instrument digital interface is a technical standard that describes a protocol, digital interface and connectors and allows a wide variety of electronic musical instruments, computers and other related devices to connect and communicate with one another
MLP	Multi-Layer Perceptron: Multi-layer perceptron is a feed-forward artificial neural network model that maps sets of input data onto a set of appropriate outputs
MP3	MPEG-1 or MPEG-2 Audio Layer III: MPEG-1 or MPEG-2 Audio Layer III, more commonly referred to as MP3, is an audio coding format for digital audio which uses a form of lossy data compression. It is a common audio format for consumer audio streaming or storage, as well as a de facto standard of digital audio compression for the transfer and playback of music on most digital audio players
MPEG	Moving Pictures Experts Group: Moving Pictures Experts Group is a working group of ISO/IEC with the mission to develop standards for coded representation of digital audio and video and related data
MRI	Magnetic Resonance Imaging: Magnetic resonance imaging is a medical imaging technique used in radiology to investigate the anatomy and physiology of the body in both health and disease. MRI scanners use strong magnetic fields and radiowaves to form images of the body
MSB	Most Significant Bit: Most significant bit is the bit position in a binary number having the greatest value
MSE	Mean Square Error: Mean square error is one of many ways to quantify the difference between values implied by an estimator and the true values of the quantity being estimated
NOC	Network on Chip: Network on chip is a communication subsystem on an integrated circuit (commonly called a “chip”), typically between IP cores in a system on a chip (SoC)
PCA	Principal Component Analysis: Principal component analysis is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components
PLP	Perceptual Linear Prediction
PNN	Probabilistic Neural Network: Probabilistic neural network is a feed-forward neural network, which was derived from the Bayesian network and a statistical algorithm called Kernel Fisher discriminant analysis

PSNR	Peak Signal-to-Noise Ratio: Peak signal-to-noise ratio is an engineering term for the ratio between the maximum possible power of a signal and the power of corrupting noise that affects the fidelity of its representation. Because many signals have a very wide dynamic range, PSNR is usually expressed in terms of the logarithmic decibel scale
RF	Radio Frequency
RLE	Run-Length Encoding: Run-length encoding is a very simple form of data compression in which runs of data (that is, sequences in which the same data value occurs in many consecutive data elements) are stored as a single data value and count, rather than as the original run
RLS	Recursive Least Square: Recursive least square is an algorithm used in adaptive filter which recursively finds the filter coefficients that minimize a weighted linear least squares cost function relating to the input signals
ROI	Region of Interest: Region of interest is a selected subset of samples within a dataset identified for a particular purpose
SINR	Signal to Interference and Noise Ratio: Signal to interference and noise Ratio is the ratio of signal power to noise power plus interference power
SNR	Signal to Noise Ratio: Signal to noise ratio is the ratio of received signal power to the noise power
SOM	Self-Organizing Map: Self-Organizing Map is a type of ANN that is trained using unsupervised learning to produce a low-dimensional (typically two-dimensional), discretized representation of the input space of the training samples, called a map
SSE	Sum of Squares of Errors: Sum of squares of error is the sum of the squared differences between each observation and its group's mean
STFT	Short-Time Fourier Transform: Short-Time Fourier Transform or alternatively Short-Term Fourier Transform, is a Fourier-related transform used to determine the sinusoidal frequency and phase content of local sections of a signal as it changes over time
SVD	Singular Value Decomposition: Singular value decomposition is a factorization of a real or complex matrix, with many useful applications in signal processing and statistics
SVM	Support Vector Machine: Support vector machine are supervised learning models with associated learning algorithms that analyze data and recognize patterns, used for classification and regression analysis
UML	Unified Modeling Language: Unified modeling language is a general-purpose modeling language in the field of software engineering, which is designed to provide a standard way to visualize the design of a system

VLSI	Very Large Scale Integration: Very large scale integration is the process of creating integrated circuits by combining thousands of transistors into a single chip
WPD	Wavelet Packet Decomposition: Wavelet packet decomposition is a wavelet transform where the discrete-time (sampled) signal is passed through more filters than the discrete wavelet transform (DWT)
WT	Wavelet Transform: Wavelet transform is a form of spectral transformation method used for analysis of signals with time and frequency content

Part I
Invited Papers

Chapter 1

A Discrete Event System Based Approach for Obfuscated Malware Detection

Chinmaya K. Patanaik, Ferdous A. Barbhuiya, Santosh Biswas and Sukumar Nandi

Abstract With the growing use and popularity of Internet among people, security threats such as viruses, worms etc., are also rapidly increasing. In order to detect and prevent such threats, many antivirus softwares have been created. Signature matching approach used to detect malwares can be easily thwarted by using code obfuscation techniques. In this paper, we propose a discrete event systems-based approach to detect obfuscated malwares in a system, taking Bagle. A as our test virus. Commonly used obfuscation techniques have been applied to bagle. We built DES models for a process under attack and normal conditions with system calls as events. Based on the system calls evoked by any process, our detector will determine its maliciousness by comparing it with both the models.

Keywords Discrete event systems · ClamAV · DDoS · Malwares

1.1 Introduction

Malicious codes can infiltrate a system using a varieties of techniques such as software vulnerability, social engineering skills etc., for the purpose of stealing valuable information or using the compromised host as a bot to launch further attacks like distributed denial of service (DDoS) etc. To counter these malwares such as viruses, worms, botnets, rootkits, spywares etc., many host- and network-based intrusion detection system (IDS) or antivirus softwares have been designed.

C.K. Patanaik (✉) · F.A. Barbhuiya · S. Biswas · S. Nandi
Indian Institute of Technology Guwahati, Guwahati 781039, India
e-mail: chinmayapatanaik@yahoo.com

F.A. Barbhuiya
e-mail: ferdousa@gmail.com

S. Biswas
e-mail: santosh_biswas@iitg.ernet.in

S. Nandi
e-mail: sukumar@iitg.ernet.in

All the traditional antivirus softwares use signature matching approach to detect maliciousness of an executable. They compare specific byte sequences, strings or hashes of files of an executable and treat it as its signature. A database of signatures representing different viruses is maintained. While scanning different executables, they search for those specific malicious byte sequences. Though this method is quite fast and efficient, it is not reliable. It is useless against zero-day malwares. Again to defeat this signature matching approach, virus writers have deployed many techniques like encryption, polymorphism, metamorphism etc. Code obfuscation techniques, which were primarily used by software vendors to protect their softwares, are now being used by metamorphic and polymorphic viruses to avoid antivirus detection [1].

IDS can be divided into two broad categories as host based and network based. Both these IDSs can be further categorized into either signature- or anomaly-based intrusion detection systems. In case of signature-based IDS, a knowledge base of signatures is maintained. Models of well-known attacks are generated and fed into the IDS as their signatures. Finite-state machines, extended FSMs, discrete event system are already widely used in IDSs to generate models. Also DES have widely been used for failure detection and diagnosis of large systems like chemical reaction chambers, nuclear reactors etc. [2]. Previous signature-based HIDSs were prone to various kinds of evasion attacks. The reason was the signature models representing attacks that did not use the system call feature perfectly. Some of the earlier researches such as Markovian algorithms, finite-state automata, wavelets, N-grams etc., considered only the sequences of system calls evoked completely ignoring their arguments. This motivated us to use discrete event systems to model the behavior of an obfuscated malware using interdependent system calls. In the software-related domain, DES has been used in: analysis of database transactions, protocol verification, feature interaction, execution of work flows and concurrent software [3]. In this paper, we will show a discrete event systems-based approach for modeling well-known viruses.

1.2 Previous Work

The proposed approach monitors system calls and uses DES-based techniques to verify security properties of a program. Various techniques have been proposed to detect obfuscated executables with their own merits and demerits. By using static analysis, Christodorescu and Jha [4] used some common obfuscation techniques on certain malware samples and tested those samples against popular antivirus scanners.

They took the instruction sections of a virus which are used by different antivirus as its signature. Then they made a code automaton of it.

By taking the section of viruses which is used as signature, a code automaton is made. From the obfuscated version of the same, an annotated control flow graph was made. Code section having the malicious code inside the control flow graph of the

program is searched. But because of the complexity and high implementation cost involved with static analysis, we went for dynamic analysis approach.

Lee et al. [5] designed a call graph by monitoring system calls issued by a malicious program where API calls are used as nodes and jump instructions as edges. System calls are then organized in groups as per their properties and signature is depicted by building code graphs. The similarity of code graphs are determined to ascertain malware per process basis. As this method does not take into account the arguments of a system call, so it can be shown that it is prone to independent system call reordering, garbage system call insertion techniques etc.

Christodorescu et al. [6], analyzed the common part of different variants of a malware family and created a template made of instructions signifying their character. A block of instructions in an obfuscated sample that is semantically equivalent to the sequence of instructions in the signature of the original virus is analyzed to detect the malware.

Kolbitsch et al. [7] used behavioral method for detection of malwares. The sample malware is executed in an emulated environment and a model is built having its behavior using interdependent system calls called behavior graph and the arguments of every system call from the graph are noted. Each argument was back traced to find the source and instructions which affected the value of that argument and represented those instructions by a function. For any executable, the scanner tries to match the system calls invoked by the program with the behavior graphs and consider it malicious if a match is found.

Our proposed method is based on interdependent system call monitoring approach same as the previous case. But instead of using call graph or behavior graph to represent malware characteristic, we used an event-driven model to represent a malware behavior.

Neminath et al. [8] have used discrete event systems approach to detect ARP spoofing attacks in a LAN. They produced two DES models of a network under normal and ARP spoofing attack condition. Then they designed a detector which is nothing but a state estimator. At last an attack is detected when the detector comes to a state whose estimate comprises of states only from the attack model.

In our paper, we introduced the concept of observing certain system calls evoked by the malware which are interdependent. A preliminary version of the work is published in [?]. This paper reports the extension of the earlier work on obfuscated malware detection using the theory of discrete event system. The model adds a formal approach to detect obfuscated malware. A normal model of a program and a failure model under attack condition are made. Now a detector or diagnoser is created whose purpose is to identify whether a sequence of events generated by any process correspond to an attack or normal model. We tested these obfuscated malwares and some benign programs against our detector and open source antivirus *ClamAV*. Though many papers exist in system call sequence monitoring approach in case of a HIDS, not many of them use the dependency approach.

In Sect. 1.3, our proposed scheme is elaborated consisting of our DES model of test case Bagle.A. Then Sect. 1.4 gives a brief working of the detector followed by conclusion in Sect. 1.5.

1.3 Proposed Scheme

Our problem statement comes down to detect maliciousness of an obfuscated version of a known virus. In this paper, we use system call monitoring approach as every processes communicate with the operating system via system calls. We basically used the different obfuscation techniques described in [4, 9, 10]. Here, we used the code obfuscation techniques in the disassembly phase of bagle. The details of various obfuscation techniques can be found in our preliminary work [?]. We took the help of the tool *StraceNT*. We executed bagle in *StraceNT*, took the system call trace of it, and built a DES model known as attack model. Then we made another model of system call events under normal conditions. Now these were fed as signatures to the detector. The detector estimates the maliciousness of any process by checking it against the normal and attack model.

It can be noticed that because of the obfuscation techniques, only byte sequence or the syntactic structure of an executable is changed keeping the semantics of the program same. So only the arguments of API calls will change whereas their interdependency will remain same. In our paper, we will take API call interdependency approach to build DES models. Among many of its functions, bagle calls *sub_4010DD* to make a base 64 encoded copy of itself [11] which it further uses as email attachments. Some of the APIs called prior and after to this function are shown in the following subsection and is represented by Table 1.1. In case of other viruses, different properties can also be considered. For e.g., APIs issued for registry change operations, file creation etc. can also be taken into account.

Table 1.1 API call sequence of Bagle.A

Sequence number	API	Dependent APIs
1	GlobalAlloc	GetModuleFileName, CreateFile, UnmapViewOfFile GlobalFree
2	GetModuleFileName	
3	CreateFile	GetFileSize, CreateFileMapping, UnmapViewOfFile, CloseHandle
4	GetFileSize	UnmapViewOfFile
5	CreateFileMapping	MapViewOfFile
6	MapViewOfFile	GlobalAlloc, UnmapViewOfFile
7	GlobalAlloc	Istrlen
8	Istrlen	
9	UnmapViewOfFile	
10	CloseHandle	
11	GlobalFree	

1.3.1 Behavior of Bagle

Analysis of Bagle was done using StraceNt and the function responsible for its base 64-encoded version. The first API is *GlobalAlloc* which allocates certain memory from heap. Now *GetModuleFileName* is called which retrieves the path of the currently executing process. Next when *CreateFile* is called, it opens the specified process (the virus). Then *GetFileSize* returns the size of the file while *CreateFileMapping* creates a file mapping from the file opened by *CreateFile* API. Next it loads the file mapping into the currently executing process by calling *MapViewOfFile*. Next it calls its function *sub_4010DD* which basically makes a base64-encoded copy of the virus for use in email. Now *lstrlen* is called to get the length of the string. Then *UnmapViewOfFile* unmaps the file mapping from memory while *CloseHandle* closes the open handle returned from *CreateFile*. At last *GlobalFree* is called to release the allocated memory. MSDN library was used to find appropriate APIs [12].

1.3.2 Assumptions

Christodorescu and Jha have described in their paper [13] that a malicious behavior should follow some particular specifications. Based on those rules, we have the following assumptions.

1. Independent API calls are not considered while building the call graph.
2. A specification must relate dependent operations including call graph having data dependency among them.
3. A specification must capture only security-related operations.

This sequence of API calls is used to create a DES model for attack scenario. An API call *API1* is said to be dependent on another API call *API2* if any of its arguments uses the return value of *API2*. A graphical representation of it is shown in (Fig. 1.1) where nodes are API calls and transitions are made according to the data dependency between them. This is known as call graph or behavioral graph of the worm. Now according to the sequence and data dependency among API calls, a DES model is created and subsequently another model was built for normal conditions.

1.3.3 Des Model

A DES model G can be represented by a 6 tuple $\langle \Sigma, S, S_0, S_F, V, \mathfrak{S} \rangle$, where,

1. Σ is the set of events,
2. S is the set of states,
3. $S_0 \subseteq S$ is the set of initial states,
4. $S_F \subseteq S$ is the set of final states,

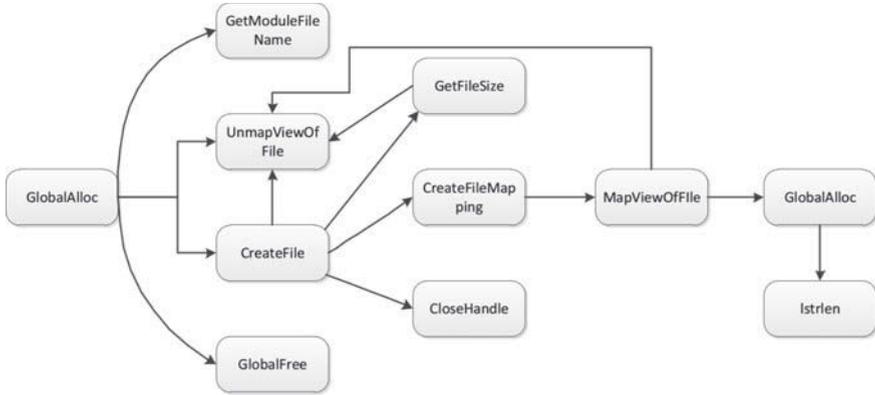


Fig. 1.1 Call graph representation of Bagle.A

5. V represents the set of model variables and
6. \mathfrak{S} is the set of transitions.

A state represents the renewal of a process if there is a transition from it to some initial state(s), and is defined as a renewal state. Each element in V is represented by v_i , where $i = 1, 2, 3, \dots, n$ and has a corresponding domain D_i .

A transition $\tau \in \mathfrak{S}$ is a five tuple $\langle s, s', \sigma, \phi(v), Assign(v) \rangle$, where

1. s is the source state,
2. s' is the destination state,
3. σ is an event based on which different transitions take place,
4. $\phi(v)$ represents a boolean conjunction of equalities of a subset of variables in V and
5. $Assign(v)$ is a subset of model variables and assignments with values from their corresponding domains.

A detailed description of DES terms, required for modeling the behavior of a process under normal and bagle-infected condition, is given below. Figure 1.3 illustrates the behavior of the process under normal conditions and Fig. 1.2 is a representation under infected conditions. Σ consists of the following APIs. $\{GlobalAlloc, GetModuleFileName, CreateFile, GetFileSize, CreateFileMapping, lstrlen, MapViewOfFile, UnmapViewOfFile, CloseHandle, GlobalFree\}$. The set of states S is shown in Figs. 1.2 and 1.3. States with no primes correspond to the normal condition and states with primes represent the attack condition. The initial state S_0 is also shown in Figs. 1.2 and 1.3. The model variable set V is $V = \{y, ret, para\}$, where y is an integer variable which is used to represent the number of independent APIs occurring in the model. ret is an array which is used to store the return values of each API and $para$ is another array which stores the parameters of APIs. The main purpose of ret and $para$ is to check for dependency relations between APIs.

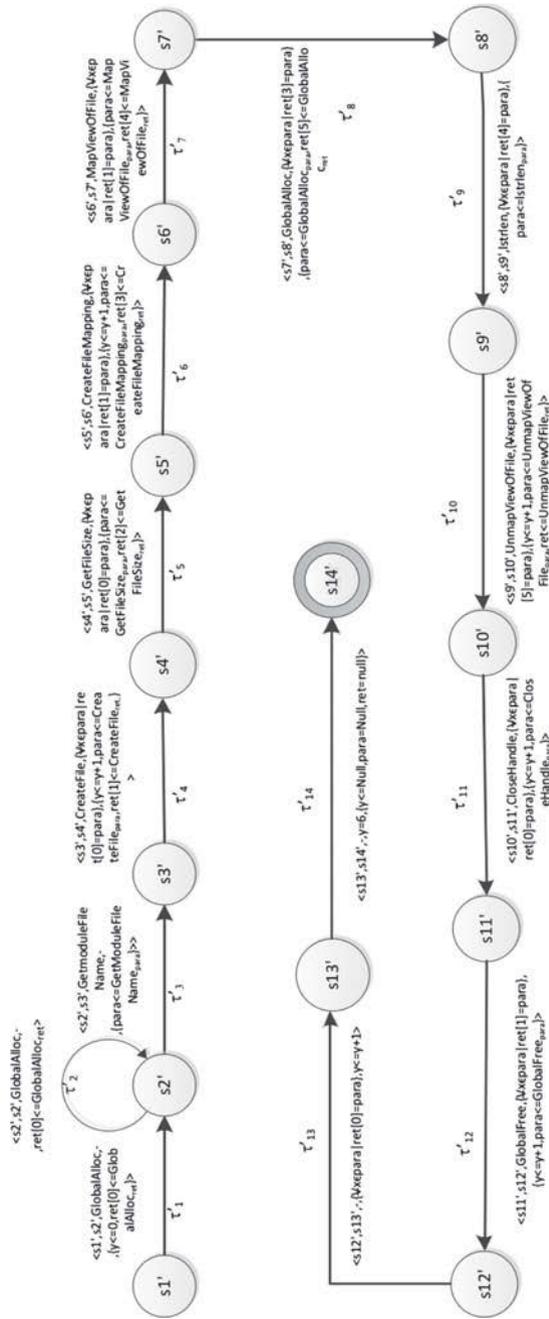


Fig. 1.2 Attack model

For example, $GlobalAlloc_{ret}$ will represent the return value of $GlobalAlloc$ whereas $CreateFile_{para}$ will denote the set of arguments of API $CreateFile$.

The transitions are shown in Figs. 1.2 and 1.3 with primed transitions representing the attack condition and transitions without primes representing the normal condition. It may be observed that a “-” for some fields in tuples of certain transitions. If “-” occurs in place of $\phi(v)$, it signifies that the condition is true. For $Assign(v)$, it means no action is taken, i.e., model variables is set or reset.

1.3.3.1 Attack Model

While building the attack model, both the call graph and trace file (Table 1.1) were taken into consideration. Trace file was helpful in keeping track of the sequence of APIs whereas call graph was used to check for dependency condition in the model for each of its edges. Now when $GlobalAlloc$ is called, a transition is made from state $s1'$ to $s2'$ while storing its return value in $ret[0]$. Now further if any other $GlobalAllocs$ are called and then it simply updates the value stored in $ret[0]$ with the current return value without changing its state. Same is done when $GetModuleFileName$ is called. But instead of storing its return value, its parameters set is stored in the model variable $para$. The reason for not storing the return value is because it does not have any further dependencies as seen from the call graph (Fig. 1.1). For all the rest of the APIs with no further dependencies, only their parameters are stored. As seen from API sequence from (Table 1.1), $CreateFile$ is called next. Now upon $CreateFile$ event, the dependency between $GlobalAlloc$ and $GetModuleFileName$ is checked, i.e., if any parameter of model variable $para$ contains the value stored in $ret[0]$ was checked. If found then the transition is made and the return value of $CreateFile$ is stored in $ret[1]$ and its parameters are stored in $para$. Again the value of y is increased to 1. For each independent API (i.e., APIs with no further dependency) such as $GetModuleFileName$, if its dependency with its previous API as per (Fig. 1.1) holds, then the value of y is increased by 1.

Now for all the rest of the APIs occurring in their previously observed sequence from (Table 1.1), their dependency is checked with their previous API as per (Fig. 1.1). If dependency holds then a state transition occurs; while if an independent API occurs and its dependency holds, then subsequently the value of y is increased by 1. In case of bagle, as we can see from the (Fig. 1.1) that there are six APIs with no further dependencies. So after the last API $GlobalFree$ is encountered and its dependency is checked, the value of y is noted. If this value is found equal to six, then a transition to the final state $s14'$ is made. Now it can be made sure that any trace satisfying this condition is infected by bagle.

1.3.3.2 Normal Model

Figure 1.3 represents the DES model of a process under normal condition. As stated above, if APIs issued by any process follows the attack model, i.e., calling the same

APIs as per the attack model in the same order with dependency condition holding true, can be considered as infected by bagle. Similarly, a process is not infected by bagle if it follows the normal model. We can say that any normal process can also issue the same APIs as shown in the attack model. Some of the APIs can also have similar dependencies w.r.t the attack model. But in order to be benign, there will be at least one state where either the specific event will not occur or dependency will fail. So it can be said that the number of independent APIs occurring in the normal model will always be less than the actual number of independent APIs occurring in bagle, i.e., six. This state distinguishes the attack model from the normal model. Suppose there is a normal process “A” whose system call trace contains all the APIs with the same execution sequence as per attack model. Again let us assume that it follows the same dependency upto *CloseHandle*. Now as per our model, it will not satisfy the last dependency condition, i.e., dependency between *GlobalAlloc* and *GlobalFree*.

In normal model, transition from s1 to s2 is made when *GlobalAlloc* occurs in the trace. Then the return value of it is stored in *ret[0]* whereas for all other APIs no state change occurs as shown by the self loop. Same is done when *GetModuleFileName* is called. If any *GlobalAlloc* appears, then a self transition is made in s2 but the value of *ret[0]* is updated. An additional transition is made from s2 to s1 if dependency fails or *GlobalFree* API appears because this will free the memory allocated by *GlobalAlloc*. For all the APIs constituting our event set, pairs of APIs which neutralize each other are given below.

1. *GlobalAlloc* and *GlobalFree*
2. *CreateFile* and *CloseHandle*
3. *MapViewOfFile* and *UnmapViewOfFile*

So for rest of the states if a pair of above APIs are found in the trace then a transition is made to s1 from that state. For example, from s7 a transition to s1 is made if an event belonging to the set *GlobalFree*, *CloseHandle*, *UnmapViewOfFile* occurs. Again, if on the occurrence of a particular event dependency does not hold, then also a transition is made to state s1 from the current state.

1.4 Detector

After creating the DES models, we designed a detector which can identify whether a given sequence of events belong to normal or attack scenario. The detector is basically a kind of state estimator of the model which updates the estimate using events of the system. Finally, it declares an attack when a state (of the detector) is reached whose estimate comprises of states only from the attack model. Before discussing the detector formally, let us introduce the following definitions.

1.4.1 Definition 1: Measurement of Equivalent Transitions and States

Two transitions, $\tau_1 = \langle s_1, s'_1, \sigma_1, \phi_1(V), Assign_1(V) \rangle$ and $\tau_2 = \langle s_2, s'_2, \sigma_2, \phi_2(V), Assign_2(V) \rangle$ are equivalent if $\sigma_1 = \sigma_2$ (same event), $\phi_1(V) \equiv \phi_2(V)$ (same equalities over the same subset of variables in V) and $Assign_1(V) \equiv Assign_2(V)$ (same subset of model variables with same assignment).

If $\tau_1 = \tau_2$, the source states of the transitions are equivalent and so are the destination states, i.e., $s_1 = s_2$ and $s'_1 = s'_2$. The detector is represented as a directed graph $O = \langle Z, A \rangle$, where Z is the set of detector states, called O -states, and A is the set of detector transitions, called O -transitions. Henceforth, terms like transitions, state etc. of the DES are termed as model transitions, model states etc. to distinguish them from those of the detector. Each O -state $z \in Z$ comprises a subset of equivalent model states representing the uncertainty about the actual state and each O -transition $a \in A$ is a set of equivalent model transitions representing the uncertainty about the actual transition that occurs. The initial O -state comprises all initial model states. Following that given any O -state z , the O -transitions emanating from z are obtained as follows. Let \mathfrak{S}_z denote the set of model transitions from model states $s \in z$. Let A_z be the set of all equivalence classes of \mathfrak{S}_z . For each $a \in A_z$, an O -transition is created comprising of all model transitions in a . Then the successor O -state for a is constructed as $z^+ = \{s | s \text{ is the destination model state of } \tau \in a\}$. This process is repeated until no O -transition can be created.

1.4.2 Definition 2: Attack-Certain O -state

An O -state, which contains only model states corresponding to attacks is called attack-certain O -state. attack-certain O -states denote that the current (model) state estimate comprises only of abnormal states, thereby detecting an attack.

Figure 1.4 illustrates the detector for DES models shown in Figs. 1.2 and 1.3.

1. The initial states of both models are s_1 and s_1' . In the detector, the initial O -state z_1 comprises of both s_1 and s_1' .
2. Now as seen from both the DES models, transitions τ_2 and τ_1 are equivalent. So an O -transition is made from z_1 to z_2 . Here z_2 contains the equivalent states s_2 and s_2' .
3. Next transitions emanating from s_2 and s_2' are checked for equivalence. Similarly, τ_6 and τ_3 are found equivalent and another transition is made from z_2 to z_3 .

The same procedure is repeated until it reaches the O -state z_{12} . z_{12} consists of equivalent states s_{12} and s_{12}' . Now if we refer to the attack model, we can observe that a transition from s_{12}' to s_{13}' is made confirming the final dependency between *GlobalAlloc* and *GlobalFree* whereas in case of normal model (Fig. 1.3), this condition does not hold. In case of normal model after state s_{12} , a transition from s_{12} to

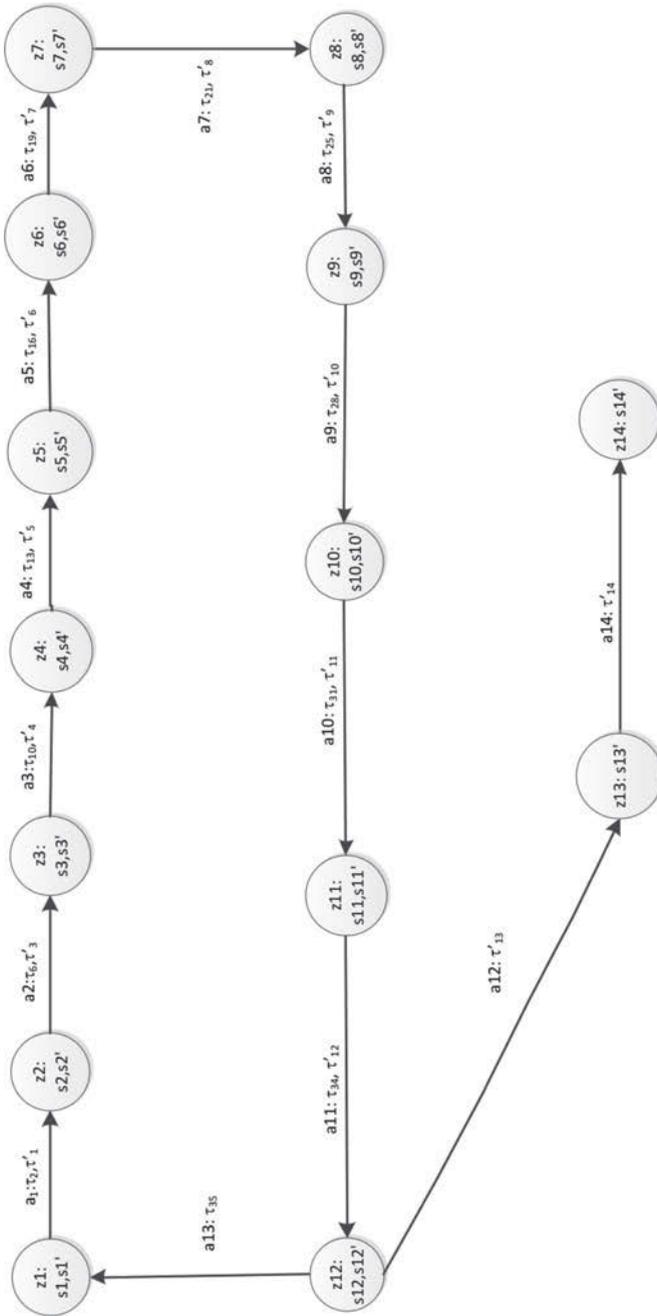


Fig. 1.4 Detector model

s1 is made as dependency will not be followed in normal case. This is illustrated in Fig. 1.4 by transition a13 representing the normal condition and transitions a12 and a14 showing attack condition. So z13 and z14 in the detector are the attack-certain *O-states*.

Now let a detector reach state z13 by sequence z1, z2, . . . , z12. In all of the above states, we cannot differentiate between attack and normal condition as each one consists of one state from normal DES models and another state from attack DES model, e.g., z7 has s7 and s7'. z13 being an attack-certain *O-state*, reached by occurrence of a12 declares an attack. It can be observed from (Fig. 1.4) that a12 and a14 correspond to the section where last dependency holds and number of independent APIs are checked against a certain number. It should be noted that the detector does not declare a normal condition. It only alerts when an attack-certain *O-state* is reached. So it can be said that unless an attack-certain *O-state* is reached, the process is normal.

1.5 Conclusion

With obfuscating tools available online, it is possible for anyone with very less knowledge to create an obfuscated version of a known malware. In this paper, we showed an event-driven DES model to represent the email worm Bagle.A with the help of interdependent system calls. Now around 20 obfuscated samples of bagle were created and tested against open source antivirus *ClamAV*. *ClamAV* was not able to detect any of the obfuscated samples (not even with only one NOP instruction inserted) whereas our detector successfully detected each of the variants. Here we demonstrated how the properties of a malware can be modeled efficiently taking only one virus. We are currently working on automating the process of DES model generation for a large set of malwares.

References

1. Sze, P., Ferrie, P.: Hunting for metamorphic. In: Virus Bulletin Conference, pp. 123–144 (2001)
2. Cassandras, C., Lafortune, S.: Introduction to discrete event systems, series Kluwer International Series on Discrete Event Dynamic Systems. Kluwer Academic (1999) [Online]. Available: <http://books.google.co.in/books?id=IuGo4aRyb00C>
3. Dingel, J., Rudie, K., Dragert, C.: Bridging the gap: discrete-event systems for software engineering (short position paper). In: Proceedings of the 2nd Canadian Conference on Computer Science and Software Engineering, series C3S2E'09, pp. 67–71. New York, NY, USA: ACM (2009) [Online]. Available: <http://doi.acm.org/10.1145/1557626.1557637>
4. Christodorescu, M., Jha, S.: Static analysis of executables to detect malicious patterns. In: Proceedings of the 12th USENIX Security Symposium, pp. 169–186 (2003)
5. Lee, J., Jeong, K., Lee, H.: Detecting metamorphic malwares using code graphs. In: Proceedings of the 2010 ACM Symposium on Applied Computing, series SAC'10, pp. 1970–1977.

- New York, NY, USA: ACM (2010) [Online]. Available: <http://doi.acm.org/10.1145/1774088.1774505>
6. Christodorescu, M., Jha, S., Seshia, S.A., Song, D.X., Bryant, R.E.: Semantics-aware malware detection. In: IEEE Symposium on Security and Privacy, pp. 32–46 (2005)
 7. Kolbitsch, C., Comparetti, P.M., Kruegel, C., Kirda, E., Zhou, X., Wang, X.: Effective and efficient malware detection at the end host
 8. Neminath, H., Biswas, S., Roopa, S., Ratti, R., Nandi, S., Barbhuiya, F., Sur, A., Ramachandran, V.: A des approach to intrusion detection system for ARP spoofing attacks. In: 2010 18th Mediterranean Conference on Control Automation (MED), June 2010, pp. 695–700
 9. Lakhotia, A., Kumar, E.U.: Abstract stack graph to detect obfuscated calls in binaries. In: Proceedings 4th IEEE International Workshop on Source Code Analysis and Manipulation, pp. 17–26, IEEE Computer Society (2004)
 10. You, I., Yim, K.: Malware obfuscation techniques: a brief survey. In: 2010 International Conference on Broadband, Wireless Computing, Communication and Applications (BWCCA), november, pp. 297–300 (2010)
 11. Rozinov, K.: Reverse code engineering: an in-depth analysis of the bagle virus. Technical Report, Bell Labs (2004)
 12. <http://msdn.microsoft.com/en-us/library/ms123401.aspx>
 13. Christodorescu, M., Jha, S., Kruegel, C.: Mining specifications of malicious behavior. In: Proceedings of the 6th joint meeting of the European Software Engineering Conference and the ACM SIGSOFT International Symposium on Foundations of Software Engineering, pp. 5–14 (2007)
 14. <http://www.clamav.net/>
 15. <http://www.ollydbg.de/>

Chapter 2

Quantification of Diagnostic Information from Electrocardiogram Signal: A Review

S. Dandapat, L.N. Sharma and R.K. Tripathy

Abstract Electrocardiogram (ECG) contains the information about the contraction and relaxation of heart chambers. This diagnostic information will change due to various cardiovascular diseases. This information is used by a cardiologist for accurate detection of various life-threatening cardiac disorders. ECG signals are subjected to number of processing, for computer aided detection and localization of cardiovascular diseases. These processing schemes are categorized as filtering, synthesis, compression and transmission. Quantifying diagnostic information from an ECG signal in an efficient way, is always a challenging task in the area of signal processing. This paper presents a review on state-of-art diagnostic information extraction approaches and their applications in various ECG signal processing schemes such as quality assessment and cardiac disease detection. Then, a new diagnostic measure for multilead ECG (MECG) is proposed. The proposed diagnostic measure (MSD) is defined as the difference between multivariate sample entropy values for original and processed MECG signals. The MSD measure is evaluated over MECG compression framework. Experiments are conducted over both normal and pathological MECG from PTB database. The results demonstrate that the proposed MSD measure is effective in quantifying diagnostic information in MECG. The MSD measure is also compare with other measures such as WEDD, PRD and RMSE.

Keywords Diagnostic measure · Multichannel ECG · Multivariate sample entropy · MSD · PRD · WEDD

S. Dandapat (✉) · L.N. Sharma · R.K. Tripathy
Indian Institute of Technology Guwahati, Guwahati 781039, India
e-mail: samaren@iitg.ernet.in

L.N. Sharma
e-mail: lns@iitg.ernet.in

R.K. Tripathy
e-mail: r.tripathy@iitg.ernet.in

2.1 Introduction

The cardiovascular system (CVS) consists of heart and blood vessels [1]. The function of heart is to provide oxygenated blood to whole body. The electrical activity of the heart is coordinated by pacemaker cells. Pacemaker cells are specialized cardiac myocytes [2]. These cells are having its own intrinsic peaks or firing rate. In heart, each cell can act as pacemaker. Due to firing rate, there are four major kind of pacemaker cells, that are responsible for electrical conduction in heart [2, 3]. These are categorized as sino-atrial (SA) node, atrio-ventricular (AV) node, HIS bundle and purkinje fibers. The firing of SA node, leads to atrial depolarization. After a delay, the AV node fires and these intrinsic peaks act as intermediate to bring the electrical activity from the upper chamber to lower chamber of heart. The HIS bundle and purkinje fibers responsible for depolarization of the septum and the ventricles. The re-polarization of myocardium is occurred after depolarization in the opposite direction. The bio-electrical activity of entire heart is measured through ECG [1, 2, 4]. Each heart beats in ECG consist of diagnostic features such as waveform amplitude and interval. Figure 2.1 shows the synthetic ECG marked with the diagnostic features. These features are examined by a cardiologist for accurate detection and localization of cardiac disorders. The detail description of each of these morphological diagnostic features are given as [1, 2].

- P Wave:** It is a low amplitude clinical component present in ECG. The P-wave gives the information about both left and right atrial contraction. Due the atrial flutter and atrial fibrillation kind of cardiac arrhythmia the unordered nature of P-wave occurs with a very high rate [2].

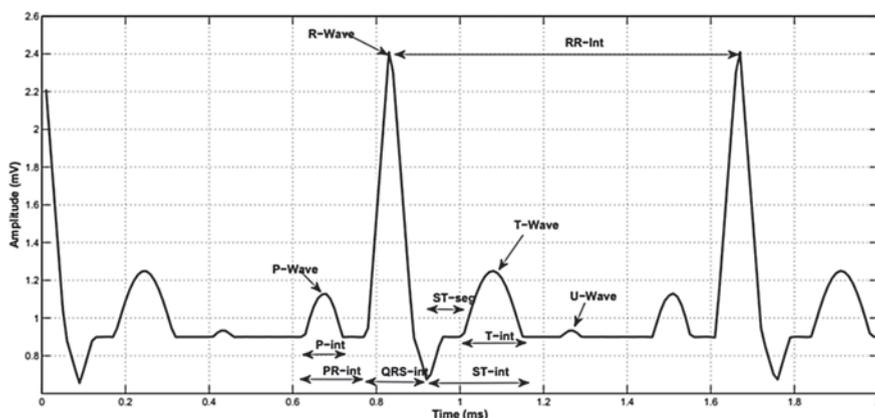


Fig. 2.1 Shows the synthetic ECG signal marked with morphological diagnostic features as waveform amplitude and interval [5]

- **QRS Complex:** This clinical component present in ECG is due to ventricular contraction. It is one of the important clinical feature for the evaluation of heart rate [2].
- **T Wave:** This waveform occurred in ECG due to ventricular relaxation. The shape of this waveform is one of the important characteristics to diagnose various cardiac arrhythmia. The shape of T-wave will change due to hyperkalemia, hypokalemia, hypercalcemia and myocardial infarction type of heart diseases [2].
- **PR Segment:** It is the time between the end of P-wave and the beginning of QRS complex. This component is a iso-electric line and mainly corresponds to the delay, after which the second pacemaker of heart (atrio-ventricular node) fires.
- **RR Interval:** In ECG the R-wave has highest amplitude among other clinical components. The difference between R-wave locations of consecutive beats is called as RR-interval. The RR-interval variation is one of the important feature for diagnosis respiratory disorder, sleep apnea etc. [2].
- **ST Segment:** This clinical component is due to the end of ventricular contraction and beginning of ventricular relaxation. The depression and elevation of ST-segment is the indicator of cardiac arrhythmia [2].
- **QT Interval:** This clinical component present in ECG is mainly due to the beginning of ventricular contraction and end of ventricular relaxation. The relationship between QT interval and RR interval is they are directly proportional. Due to tachycardia the heart rate is more this implies the QT interval shortens. Similarly, due to bradycardia the heart rate is less, and this results a longer QT interval [2].
- **U Wave:** This type of waveform seen in lead V2 and V3 and it is due to the delayed relaxation of purkinje fibers [1, 2]. It is present in ECG, when the heart rate falls below 60bpm [2].

The 12-lead or Multilead ECG (MECG) provides the spatial as well as temporal information about cardiac rhythm. It is widely used in hospitals for diagnosis of cardiac ailments [2]. These signals are recorded by placing 10 number of electrodes on the surface of body. The MECG consists of uni-polar limb lead, bipolar limb leads and precordial leads [6]. The unipolar and bipolar limb leads are derived from the potentials recorded through the electrodes placed at right leg (RL), left arm (LA) and right arm (RA) [2, 7]. The Anterior View (V1, V2, V3 and V4), Left lateral View (I, aVL, V5 and V6) and Inferior View (II, III and aVF) of heart is seen from MECG [2]. The cardiac ailments such as myocardial infarction, valvular disease, hypertrophy, myocarditis, bundle branch block, atrial flutter, atrial fibrillation, ventricular fibrillation, ventricular tachycardia and cardiomyopathy are detected from MECG [1, 2, 8]. The description of each of these cardiac arrhythmia are given below [2].

- **Sinus Bradycardia**—In this case all the beats in ECG are normal, RR-interval is more and heart rate (HR) < 60beats/min.
- **Sinus Tachycardia**—It is occurred due physiological stress or exercise. Here, all the beats are normal, RR-interval is less and HR > 100beats/min.
- **Sinus Arrhythmia**—The HR increases due to inspiration and decreases due to expiration. This type of arrhythmia particularly found in children [1].

- **Wandering Pacemaker**—In this type of arrhythmia the impulse is originated from various points of atria [1]. In ECG due to Wandering Pacemaker there is a variation in P-wave polarity, PR-interval, PP-interval and RR-interval.
- **Atrial Flutter**—In this case the end of T-wave and beginning of P-wave disappears, P-wave results in a circus movement inside atria with high rate of activation between 160 to 200 beats/min [1].
- **Atrial Fibrillation**—Impulse is having chaotic and random pathways in atria, Baseline and ventricular depolarization are irregular with high rate of activation between 250 to 300 beats/min [1].
- **Junctional Rhythm**—In this arrhythmia the impulse is originated at AV node, P-wave often inverted and may be under or after QRS complex [1].
- **Premature Ventricular Contraction**—In this case the origin of electrical conduction is in ventricular Muscle not the AV node, QRS-complex is abnormal and longer than 0.1 s [1].
- **Ventricular Tachycardia**—Slower conduction in ischemic ventricular muscle which leads to circular activation. this may cause a rapid, bizarre and wide QRS-complexes with a rate of 120 beats/min [2].
- **Ventricular Fibrillation**—Ventricular muscle contraction is irregular and ineffective of pumping blood. The ventricular fibrillation (VF) can be stops using external defibrillator [2].
- **Right Atrial Hypertrophy**—It is due to the consequence of right atrial overload, tricuspid valve disease and pulmonary valve disease [2]. Due to this arrhythmia the amplitude of P-wave is greater than 0.25 mv in Lead II, III and aVF.
- **Left Atrial Hypertrophy**—It is due to the consequence of right atrial overload and mitral valve disease [2]. A notched P-wave and biphasic p-wave seen in lead II lead V1 respectively with a negative amplitude >0.1 mv.
- **Right Ventricular Hypertrophy**—It is the consequence of pulmonary valve stenosis and pulmonary hypertension [2]. In ECG, a tall R-wave seen in Lead V1 and V2 with amplitude >0.7 mv, wide S-wave seen in lead V1 and V2, wide R-wave seen in V5 and V6.
- **Left Ventricular Hypertrophy**—It is the consequence of left atrial overload, mitral valve disease and aortic valve disease [2]. In MECCG a tall R-wave in lead I, V5, V6 and tall S-wave in lead III, V1 and V2 occurs.
- **Anterior Myocardial Infarction**—This type of MI is due to the occlusion in left anterior descending artery [2]. Due to anterior infarction, the R-wave progression occurred in precordial leads.
- **Inferior Myocardial Infarction**—This type of MI is due to the occlusion in right coronary artery [2]. Due to inferior infarction, the Q-wave variation occurred in lead II, III and aVF.
- **Posterior Myocardial Infarction**—This type of MI can be diagnosed by observing the reciprocal changes in the ECG at lead V1 [2].
- **Lateral Myocardial Infarction**—This type of MI is due to the occlusion in left circumflex artery [2]. The ST-segment elevation occurs in lead I, aVL, V5 and V6.

The pathological Q-wave formation in lead V6 is due to lateral myocardial Infarction.

- **Hyperkalemia**—This type of cardiac arrhythmia occurred due to the increase of potassium ion concentration in blood [2]. In ECG due to hyperkalemia a peaked T-wave and small p-wave occurs.
- **Hypokalemia**—Due to the decrease of potassium ion concentration in blood the flattened or inverted T waves, ST depression and a wide PR interval occurs [2]. These changes are due to the consequence of hypokalemia.

2.2 ECG Quality Assessment Techniques: A Review

ECG signals are subjected to different processing for storage, transmission and retrieval. The goal of distortion measures in ECG signal processing is to quantify the loss of clinical information. There are two kind of distortion measures are used to evaluate the quality of ECG signals [5]. These are subjective distortion measure and objective distortion measure. In the following subsections the state-of-art distortion measures and their limitations are briefly discussed.

2.2.1 Subjective Assessment

The subjective quality of ECG signals for clinical practice is evaluated by visual inspection of morphological diagnostic features [8, 9]. The researchers and medical experts judges the diagnostic quality of processed ECG signal by inspecting these morphological diagnostic features. After inspection they assigns a score to each of the morphological diagnostic feature. This score is defined as

$$MOS_l = \frac{\sum_{c=1}^{N_c} R_c}{N_c} \quad (2.1)$$

where N_c corresponds to number of experts involved for investigating the morphological diagnostic features. R_c is the quality rating or score for l th morphological diagnostic feature. The values of R_c are defined as 1-bad, 2-not bad, 3-good, 4-very good and 5-excellent. For a ECG signal the mean opinion score (MOS) is defined as the average of individual scores for l th morphological diagnostic feature. The MOS rating is defined by

$$MOS = \frac{\sum_{l=1}^{N_f} MOS_l}{N_f} \quad (2.2)$$

The percentage of difference between the MOS scores for original and processed ECG signal is defined as the subjective distortion measure. The error measures for

each morphological features are defined as

$$MOS_l^{err} = \frac{5 - MOS_l}{5} \times 100 \quad (2.3)$$

The MOS_l^{err} corresponds to the subjective error measure for l th morphological diagnostic feature. For MOS_l^{err} evaluation the MOS score for original ECG diagnostic features are assigned as excellent. The subjective distortion measures for ECG signal is defined as

$$MOS^{err} = \frac{5 - MOS}{5} \times 100 \quad (2.4)$$

The subjective tests are time taking and cumbersome processes. Due to this drawback, the researchers develop objective distortion measures. The objective measures are evaluated by comparing the diagnostic features in temporal and transformed domain for original and processed ECG signals. The objective measures are validated through subjective tests. There are two kinds of objective measures reported from literature. These are objective non-diagnostic and objective diagnostic measures.

2.2.2 Objective Non-diagnostic Distortion Measures

In this subsection, the non-diagnostic distortion measures, which are used to evaluate quality of processed ECG signals are discussed. The mean square error (MSE) is evaluated as the difference between the amplitude of original and the processed ECG signals [10]. The processed signals are obtained due to various ECG processing schemes as compression, transmission, enhancement and super-resolution ECG. Considering a discrete time signal given by $x(n) = \{x(1), x(2), x(3), \dots, x(N)\}$, which consists of N number of samples. The processed signal is given as $\tilde{x}(n) = \{\tilde{x}(1), \tilde{x}(2), \tilde{x}(3), \dots, \tilde{x}(N)\}$. The MSE between original and processed signal is expressed as

$$MSE = \frac{1}{N} \sum_{n=1}^N [x(n) - \tilde{x}(n)]^2 \quad (2.5)$$

In geometrical meaning, the MSE is defined as the euclidean distance between original and processed ECG signals. There are number of morphological diagnostic features embedded in ECG signal. Each of these features is having diagnostic importance. The MSE is only evaluated by considering the amplitudes, so for non-diagnosis regions (baseline in ECG) it exploits a large error. This large error make the processed ECG falls behind the clinical acceptable range. To overcome the drawbacks of MSE, the normalized mean square error (NMSE) is used [4]. This measure is given as

$$NMSE = \frac{\sum_{n=1}^N [x(n) - \tilde{x}(n)]^2}{\sum_{n=1}^N [x(n)]^2} \quad (2.6)$$

The normalization is used to make the error measure independent of the amplitudes of original ECG signal. As in ECG signal, the amplitudes are different for different leads and subjects. As NMSE measure is independent of the amplitude, so it is normally used as a good objective non-diagnostic distortion measures for various applications.

The root mean square error (RMSE) has been used to evaluate quality of ECG signals [11]. This measure is given as

$$RMSE = \sqrt{\frac{1}{N} \sum_{n=1}^N [x(n) - \tilde{x}(n)]^2} \quad (2.7)$$

The RMSE fails to quantify the distortion in local diagnostic regions. The normalized root mean square error (NRMSE), is also used as objective diagnostic error measures in different applications. It is given as

$$NRMSE = \sqrt{\frac{\sum_{n=1}^N [x(n) - \tilde{x}(n)]^2}{\sum_{n=1}^N [x(n)]^2}} \quad (2.8)$$

The percentage root mean square differences (PRD) is used to evaluate quality of processed signal in almost all ECG compression and enhancement applications [12]. It is defined by

$$PRD1 = \sqrt{\frac{\sum_{n=1}^N [x(n) - \tilde{x}(n)]^2}{\sum_{n=1}^N [x(n)]^2}} \times 100 \quad (2.9)$$

where $x(n)$, $\tilde{x}(n)$ are the original and processed ECG signals. PRD2 is also used in various compression and enhancement techniques for evaluating the quality of processed signals. This measure is evaluated after subtracting the baseline of 1024 and mean value from the original ECG signal. The PRD2 is given as

$$PRD2 = \sqrt{\frac{\sum_{n=1}^N [x(n) - \tilde{x}(n)]^2}{\sum_{n=1}^N [x(n) - \mu_0 - 1024]^2}} \times 100 \quad (2.10)$$

where μ_0 corresponds to the mean value of the original ECG signal. In some applications, the PRD3 is also used to evaluate the objective quality of processed ECG signals. This PRD3 measure is given as

$$PRD3 = \sqrt{\frac{\sum_{n=1}^N [x(n) - \tilde{x}(n)]^2}{\sum_{n=1}^N [x(n) - 1024]^2}} \times 100 \quad (2.11)$$

The PRD1 value is found to lowest from PRD2 and PRD3 for evaluating the quality of processed ECG signal. The ECG signal with low PRD value not necessarily provide better diagnostic quality. For the ECG signal $x(n)$ with fluctuating baseline, and high standard deviation, the PRD will be artificially lower [8]. The PRD and other similar type of error measures have the limitations to evaluate the diagnostic quality of ECG signals [9].

Signal to noise ratio (SNR) has been used as a objective non-diagnostic distortion measure to evaluate the performance of ECG compression technique [13]. This measure is given as

$$SNR = 10 \log_{10} \left(\frac{\sum_{n=1}^N [x(n) - \mu_0]^2}{\sum_{n=1}^N [x(n) - \tilde{x}(n)]^2} \right) \quad (2.12)$$

The SNR value is high at the high activity regions of interest as compared to other regions [9].

The normalized cross-correlation (NCC) measure has been used to evaluate the objective quality of ECG signals [14]. The NCC measure defines the similarity between original and processed ECG signals and it is given by

$$NCC = \frac{\frac{1}{N} \sum_{n=1}^N [x(n) - \mu_0] \sum_{n=1}^N [\tilde{x}(n) - \mu_r]}{\sqrt{\frac{1}{N} \sum_{n=1}^N [x(n) - \mu_0]^2} \sqrt{\frac{1}{N} \sum_{n=1}^N [\tilde{x}(n) - \mu_r]^2}} \quad (2.13)$$

The μ_0 and μ_r corresponds to the mean values of both original signal and processed ECG signal. The NCC measure is used to find the similarity in local waves between original and processed ECG signal.

The percentage area difference (PAD) based objective non-diagnostic distortion measure has been used to evaluate the quality of ECG signal in [15]. This distortion measure is computed by considering the difference in the area enclosed between the original and the processed ECG signals. The PAD measure is given as

$$PAD = \frac{\left| \int_{t_i}^{t_f} y(t) - \int_{t_i}^{t_f} y_r(t) \right|}{(t_i - t_f)(y_{max} - y_{min})} \times 100 \quad (2.14)$$

The t_i and t_f corresponds to the initial and final time instants of the segment, y_{max} and y_{min} are the maximum and minimum values in the original ECG signal. The numerator term correspond to the absolute error in terms of area difference between the original and processed ECG signal. The denominator is used as a normalization factor for making the distortion measure independent of area.

The maximum amplitude error (MAX) has been used as a distortion measure in ECG compression technique [16]. This measure can quantify the local distortion in ECG signal. This measure is given as

$$\text{MAX}_i = \max_{n=1}^{N_{ci}} \{|x(n) - \tilde{x}(n)|\} \quad (2.15)$$

where, N_{ci} corresponds to number of samples in i th cycle.

Similarly, the normalized maximum amplitude error (NMAX) was used for quantifying distortion in [9]. This measure is given as

$$\text{NMAX}_i = \frac{\max_{n=1}^{N_{ci}} \{|x(n) - \tilde{x}(n)|\}}{\max_{n=1}^{N_{ci}} \{x(n)\} - \min_{n=1}^{N_{ci}} \{x(n)\}} \quad (2.16)$$

The normalization is done to make error independent of amplitude value. The average of NMAX measure for entire ECG signal is evaluated by averaging over all the cycles.

The objective non-diagnostic distortion measures have limitations to quantify the local distortions in the ECG signals. The Standard error (StdErr) has been used as objective non-diagnostic measure for evaluating the quality ECG signals [9]. The StdErr is given by

$$\text{StdErr} = \sqrt{\frac{1}{N_c - 1} \sum_{n=1}^{N_c} [x(n) - \tilde{x}(n)]^2} \quad (2.17)$$

StdErr is similar to RMSE, where the denominator term is N_c instead of $N_c - 1$.

2.2.3 Objective Diagnostic Distortion Measures

Due to the drawbacks of non-diagnostic distortion measures for evaluating the quality of ECG signal, the objective diagnostic distortion measures are used. In this subsection, the objective diagnostic measures for ECG signals reported from literature are also briefly discussed.

Chen and Itoh proposed an objective diagnostic distortion measure for evaluating the quality of ECG signal [14]. The measure is the weighted PRD between clinical diagnostic features of original and processed ECG signals. It is given as

$$\text{WPRD} = \sqrt{\frac{\sum_{k=1}^M w_k \gamma_k}{\sigma}} \quad (2.18)$$

where w_k are the weights, γ_k is the MSE of clinical diagnostic features as P-wave, Q-wave, QRS-wave and ST-wave for original and processed ECG signals. The σ corresponds to the power of original ECG signal. The weights are assigned according to the clinical importance of diagnostic features. As per the example, the higher weight is given to ST-segment and T-wave and QRS amplitudes due to their importance in the diagnosis of cardiac arrhythmia.

Zigel and his group [17] proposed a objective diagnostic distortion measure for evaluating the quality of processed ECG signal. They extracted sixteen clinical morphological features from both original and processed ECG signals and define the distortion measure as:

$$WDD(\beta, \tilde{\beta}) = \Delta\beta^T \cdot \frac{w}{tr[w]} \cdot \Delta\beta \times 100 \quad (2.19)$$

where $\beta, \tilde{\beta}$ corresponds to the clinical feature vectors corresponding to original and processed signal. The ‘ w ’ corresponds to the weight value for the respective morphological features. The weights are assigned based on the diagnostic importance of the morphological features. The limitations of both WPRD and WDD measures are selection of optimal weights and identification of ‘*PQRST*’ points for evaluating the diagnostic features in ECG signal.

The wavelet based distortion measures overcome the limitations in WPRD and WDD by assigning the weights through different parameters as energy and entropy. Al-Fahoum proposed an objective diagnostic distortion measure for evaluating the quality of compressed ECG signal [18]. In the wavelet domain, the clinical diagnostic features are captured through approximation and detail coefficients. The distortion measure is defined as the weighted percentage root mean square difference between the wavelet coefficients of original and processed ECG Signals.

$$WWPRD = \sum_{j=1}^{L+1} w_j WPRD_j \quad (2.20)$$

where, L corresponds to the number of decomposition levels, w_j corresponds to the weight of the j th subband and $WPRD_j$ is the PRD value of the j th wavelet coefficients in the subband. In this measure the weights are computed as the ratio of sum of the absolute value of wavelet coefficients within that sub-band to the sum of absolute value of wavelet coefficients in all sub-bands.

Wavelet Energy-based Diagnostic Distortion (WEDD) measure is proposed in [9] to assess the quality of compressed signal in ECG data compression. WEDD was evaluated from the wavelet coefficients of the original and processed ECG signal. The WEDD measure is given as

$$WEDD = \sum_{j=1}^{M+1} w'_j WPRD_j \quad (2.21)$$

where w'_j is the weight calculated based on energy due to wavelet coefficients in sub-bands defined by

$$w'_j = \frac{\sum_{k=1}^{N_j} w_{(j,k)}^2}{\sum_{j=1}^{M+1} \sum_{k=1}^{N_j} w_{(j,k)}^2} \quad (2.22)$$

The errors in the wavelet coefficients is given as:

$$WPRD = \sqrt{\frac{\sum_{k=1}^{N_j} (w_{j,k} - \tilde{w}_{j,k})^2}{\sum_{k=1}^{N_j} w_{j,k}^2}} \times 100 \quad (2.23)$$

where $w_{j,k}$ and $\tilde{w}_{j,k}$ are the wavelet coefficients for original and processed ECG signals and N_j is the number of sub-band wavelet coefficients at j level of decomposition.

The multiscale entropy based PRD objective diagnostic distortion measure is proposed in [19]. The PRD between the approximation and detail subband wavelet coefficients of original and processed ECG signals is evaluated and the weights are assigned as the multiscale entropy values of each sub-bands. The MSEPRD measure is given as

$$MSEWPRD = w_{AL} \times \left(\sqrt{\frac{\sum_{k=1}^{N_{AL}} (A_L(k) - \tilde{A}_L(k))^2}{\sum_{k=1}^{N_{AL}} A_L(k)^2}} \times 100 \right) + \sum_{j=1}^L w_{Dj} \times \left(\sqrt{\frac{\sum_{k=1}^{N_{Dj}} (D_j(k) - \tilde{D}_j(k))^2}{\sum_{k=1}^{N_{Dj}} D_j(k)^2}} \times 100 \right) \quad (2.24)$$

where w_{AL} and w_{Dj} are the weights as multiscale entropy values for Lth approximation and Jth detail wavelet sub-bands. The multiscale entropy values are defined by $H_j = -\tilde{Y}_j \log(\tilde{Y}_j)$ and $\tilde{Y}_j = \frac{\bar{E}_j}{\bar{E}_{tot}}$. The multiscale subband energy at j th resolution level and total sub-band energy is given as $\bar{E}_j = \frac{1}{K_j} \sum_{k=1}^{K_j} |C_j(k)|^2$, $C_j(k)$ is the wavelet coefficients, $j \in (1, 2, \dots, J+1)$ and $\bar{E}_{tot} = \sum_{j=1}^{J+1} \bar{E}_j$.

The wavelet energy weighted PRD (WEWPRD) measure is proposed in [20] to evaluate the quality of ECG signal. It is computed as the sum of weighted error energies due to wavelet coefficients of original and processed ECG signals. The weights for the approximation and detail sub-bands are given as

$$w = \left[\frac{E_{A_j}}{E_t}, \frac{E_{D_j}}{E_t}, \dots, \frac{E_{A_1}}{E_t} \right] \quad (2.25)$$

The wavelet energy PRD is given as

$$WEPRD = \left[\sqrt{\frac{E_{A_j}^{err}}{E_{A_j}}} \times 100, \sqrt{\frac{E_{D_j}^{err}}{E_{D_j}}} \times 100, \dots, \sqrt{\frac{E_{D_1}^{err}}{E_{D_1}}} \times 100 \right] \quad (2.26)$$

where $E_{A_j}^{err} = \sum_{k=1}^{N_{A_j}} (A_j(k) - \tilde{A}_j(k))^2$ and $E_{D_j}^{err} = \sum_{k=1}^{N_{D_j}} (D_j(k) - \tilde{D}_j(k))^2$ are wavelet error energies in approximation and detail sub-bands. The WEWPRD measure is given as

$$WEWPRD = w^T WEPRD \quad (2.27)$$

The wavelet based methods are found to be effective measures in different applications and well correlated with subjective assessment.

2.3 ECG Feature Extraction for Disease Detection: A Review

Extracting diagnostic information from normal and pathological ECG signals for detection and classification of cardiovascular diseases have been an active area of research from decades. There are two type of methods used for quantifying diagnostic information from ECG signals. These are direct and indirect methods. The direct method corresponds to the visual inspection of local wave-forms as P-wave, QRS-complex and T-wave amplitudes and duration from ECG. The cardiologist investigate these features for detection and localization of cardiac arrhythmia. On the other hand, the indirect method corresponds to the use of various signal processing and machine learning techniques for detection and localization of cardiovascular diseases. There are number of signal processing techniques like heart rate variability analysis [21–25], discrete and continuous wavelet transform based analysis [26, 27], auto-regressive model coefficients [28], discrete cosine transform coefficients [29], principal component analysis [30, 31], linear discriminant analysis [32], independent component analysis [33], polynomial regression coefficients [34] etc. are reported in literature for extracting clinical diagnostic features from ECG for arrhythmia detection and classification. The detail description of each of these diagnostic information extraction methods are given below.

The heart rate variability features are widely used in applications like cardiovascular disease detection [21], diabetes diagnosis from ECG [22], prognosis of cardiac risk [23], identifying fatigue in elite athletes [24], monitoring sleep apnea from ECG [25] etc. The trace of RR interval with respect to number of beats is termed as heart rate signal. This heart rate signal is subjected to statistical, frequency domain and non-linear analysis for getting the HRV based features. The time domain HRV features are mean of RR-intervals (RR_m), standard deviation of RR-interval (RR_{std}), mean of heart rates (HR_m) and standard deviation of heart rates (HR_{std}), RMSSD, SDDSD, NN50, PNN50 and HRV triangular index. These features are given as

$$RR_m = \frac{\sum_{i=1}^N RR(i)}{N} \quad (2.28)$$

$$RR_{std} = \sqrt{\frac{\sum_{i=1}^N (RR(i) - RR_m)^2}{N}} \quad (2.29)$$

$$HR_m = \frac{\sum_{i=1}^N HR(i)}{N} \quad (2.30)$$

$$HR_{std} = \sqrt{\frac{\sum_{i=1}^N (HR(i) - HR_m)^2}{N}} \quad (2.31)$$

where $HR(i) = \frac{60}{RR(i)}$ called as heart rate time series. RMSSD is the square root of mean of square difference between adjacent RR-intervals and given by

$$RMSSD = \sqrt{\frac{\sum_{i=1}^N (RR(i+1) - RR(i))^2}{N}} \quad (2.32)$$

Similarly SDDSD feature is the standard deviation of difference between adjacent RR-intervals and given by

$$SDDSD = \frac{\sum_{i=1}^N (RR_{diff}(i) - \tilde{RR}_{diff})}{N} \quad (2.33)$$

where $RR_{diff}(i) = (RR(i+1) - RR(i))$ and $\tilde{RR}_{diff} = \frac{\sum_{i=1}^N (RR(i+1) - RR(i))}{N}$. The PNN50 feature is evaluated by identifying the RR-interval difference more than 50 ms. The PNN50 is given as

$$PNN50 = \frac{\sum_{i=1}^N ((RR(i+1) - RR(i)) > 50 \text{ ms})}{\sum_{i=1}^N RR_{diff}(i)} \quad (2.34)$$

The frequency domain information of heart rate signals are evaluated using autoregressive model based spectral analysis, discrete Fourier transform (DFT), short time fourier transform and wavelet transform based analysis. The high frequency band (0.15–0.4 Hz) of heart rate signal is related to parasympathetic activity and low frequency band (0.04–0.15 Hz) corresponds to sympathetic activity. The total power in low frequency (LF) band, high frequency (HF) band along with the power spectral density ratio in LF to HF band (LF/HF) are used as frequency domain HRV features. The non-linear HRV features are obtained from the Poincare plot. The $RR(i+1)$ are plotted as a function of $RR(i)$. The non-linear features SD1 and SD2 are obtained from this plot as the standard deviation of the distance of $RR(i)$ points to the lines $y = x$ and $y = -x + 2 \times RR_m$. Where y and x are termed as $RR(i+1)$ and $RR(i)$. RR_m is the mean value of heart rate signal. Although HRV analysis is found to be a better information extraction technique from ECG signal for detection of cardiac diseases and other applications, but it has some limitations. HRV analysis doesn't not give any information about local waveform variations like chaotic nature of P-wave in atrial fibrillation, rather it only evaluate the RR-interval variations.

Banerjee and Mitra proposed a cross wavelet based technique for extracting diagnostic information from ECG signals [26]. The wavelet cross spectrum (WCS) and wavelet coherence (WCOH) metrics between template ECG (Normal beat) and abnormal ECG beats are evaluated. Further, these metrics are used to investigate the variations in both QRS-complex and T-wave regions for pathological ECG beats. The cross wavelet coefficients are computed as

$$W^{XY} = W^X W^{Y*} \quad (2.35)$$

where W^{XY} corresponds to cross wavelet coefficients. The W^X and W^Y are wavelet coefficients for $x(n)$ and $y(n)$ respectively. The $x(n)$ and $y(n)$ are normal and abnormal beats segmented from ECG time series. The WCS is evaluated as the square of cross wavelet coefficients and it is given as

$$WCS(s, t) = |W^{XY}|^2 \quad (2.36)$$

The wavelet coherence (WCOH) of $x(n)$ and $y(n)$ are given as

$$WCOH(s, t) = \frac{|W^{XY}|^2}{|W^X|^2 \cdot |W^Y|^2} \quad (2.37)$$

The sum of WCS and WCOH for different morphological features (QT, QRS, ST) are given by

$$sum\ WCS(s, t) = \sum_{s=s1}^{s2} \sum_{t=t1}^{t2} WCS(s, t) \quad (2.38)$$

$$sum\ WCOH(s, t) = \sum_{s=s1}^{s2} \sum_{t=t1}^{t2} WCOH(s, t) \quad (2.39)$$

The wavelet based feature extraction method for classification of myocardial infarction (MI) was proposed in [27]. The discrete wavelet transform grossly segments the clinical component present in ECG into both approximation and detail coefficients. The energy and entropy due to these wavelet coefficients are evaluated as

$$E_m = \frac{\sum_{j=1}^{N_m} W_{m,n}^2}{N_m} \quad (2.40)$$

$$E_m^a = \frac{\sum_{j=1}^{N_M} S_{M,n}^2}{N_M} \quad (2.41)$$

The $W_{m,n}$, $m \in (1, 2, \dots, M)$ and $S_{M,n}$ are detail and approximation wavelet coefficients at level M . The probability of detail p_m , $m \in (1, 2, \dots, M)$ and approximation p_M wavelet sub-bands are given as

$$p_m = \frac{E_m^d}{\sum_{m=1}^M E_m^d + E_m^a} \quad (2.42)$$

$$p_M = \frac{E_M^a}{\sum_{m=1}^M E_m^d + E_m^a} \quad (2.43)$$

The entropy is given as

$$H = - \sum_{i=1}^M p_i \log_2 p_i \quad (2.44)$$

The entropy and energy based features are extracted from 2282 normal and 718 myocardial infarction pathology based ECG beats. A threshold based classifier has been used for detection of myocardial Infarction.

The DWT+PCA, DWT+LDA and DWT+ICA based diagnostic feature extraction and beat classification approaches from ECG signals have been proposed in [32, 33]. Initially, the QRS-complex is detected using pan-tomkin's algorithm [35]. After R-point detection, each ECG beats are decomposed into approximation and detail coefficients by using discrete wavelet transform. The PCA, LDA and ICA are used for dimension reduction. The first six components of both approximation and detail coefficients in PCA, ICA and LDA domain are used as diagnostic feature vectors. Further these features are given to both probabilistic neural network and least square support vector machine (LS-SVM) classifier for detection of different cardiac arrhythmia. The combination of discrete cosine transform (DCT) and PCA have been used to extract the diagnostic features from ECG beats [29]. The DCT is applied over ECG beats and the first few coefficients are selected. The co-variance matrix is evaluated for DCT domain signal and then the PCA is used to select first 12 principal components as feature vector. Both probabilistic neural network (PNN) and support vector machine are used for detection and classification of different cardiac arrhythmia. Sun et al. proposed a method for detection of myocardial infarction based on ST-segment analysis [34]. First, the R, S and T points are detected based on derivative based algorithm. Then, the 200 samples from each ST-segment along each lead heart beats of MEKG are segmented. A fifth order polynomial curve fitting is used over those 200 samples along each beats and the six polynomial coefficients are used as feature vectors. A total of 72 dimensional polynomial coefficient along each lead of MEKG, RR-intervals and ST-segment width to height ratio as features and multi instance learning as classifier is used for MI detection. The principal component analysis and kernel principal component analysis based analysis of ECG

signal for detection of respiratory disorder were proposed in [30, 31]. The principal components are selected based on statistical test and these values are used as diagnostic features for detection of respiratory disorder and other cardiac arrhythmia. Dingfei et al. used the Auto-regressive (AR) model coefficients as diagnostic features to classify normal sinus rhythm (NSR) and various cardiac arrhythmia [28]. AR model coefficients are evaluated from five abnormal ECG beats such as ventricular tachycardia (VT) ventricular fibrillation (VF), atrial premature contraction (APC), premature ventricular contraction (PVC), supraventricular tachycardia (SVT) and normal ECG beat. The decision tree based classifier has been to classify both normal and arrhythmia beats from the AR model coefficient based diagnostic feature vector.

The sample entropy quantifies the randomness in a signal [36]. If the signal is regular (not changing) then the sample entropy value is low. Similarly, if the signal is continuously changing and having irregularity then the sample entropy is more. Recently, number of approaches based on sample entropy are reported in literature for cardiac arrhythmia detection [37, 38] and EEG based seizure detection and classification [39]. The evaluation of sample entropy for a time series data is given below. Let's consider a N-point time series ECG data $x(n) = [x(1), x(2), \dots, x(N)]$. The two parameters used to derive the sample entropy are coefficients of tolerance (r) and dimension of template vector ' m '. The k th template vector for ECG time series is given as $x_m = [x(k), x(k + 1), \dots, x(k + 1 - m)]$. Each of the k th template vector is having dimension ' m '. The distance between ' i ' and ' j 'th template vector is given as

$$d[x_m(i), x_m(j)] = \max[|x(i + k) - x(j + k)|] \tag{2.45}$$

where $k \in [0, m - 1], i \neq j$ and $i, j \in 1, 2, \dots, N - m$. The number of distances are counted for which the value of $d \leq R$, where $R = r \times sd$ and ' sd ' is the standard deviation of ECG time series. The ' i 'th total number of distance is denoted as $B_i^m(r)$. The $B^m(r)$ is evaluated as

$$B^m(r) = \frac{\sum_{i=1}^{N-m} B_i^m(r)}{(N - m - 1)(N - m)} \tag{2.46}$$

The $B^{m+1}(r)$ is evaluated in similar way by replacing the dimension template vector to $m + 1$. The sample entropy is evaluated as the negative logarithmic ratio of $B^{m+1}(r)$ and $B^m(r)$ and given by

$$sampEN(m, r, N) = -\ln \left[\frac{B^{m+1}(r)}{B^m(r)} \right] \tag{2.47}$$

2.4 MSD Diagnostic Measure for MECG

The state-of-art objective diagnostic and non-diagnostic measures such as PRD, MSE, MAE, WDD, WEDD, WWPRD, MSEWPRD, WEWPRD etc. are limited to single channel ECG signals. There are numbers of MECG data processing techniques such as multiscale PCA (MSPCA) [40], multivariate empirical mode decomposition [41], compressive sensing based MECG compression [42], DCT and karhunen loeve transform (KLT) based MECG compression [43] etc. are reported from literature. For these MECG processing schemes, the aforementioned objective diagnostic and non-diagnostic measures are computed individually along each channel. Multivariate sample entropy (MSampEn) is the natural extension of sample entropy for multichannel signals [44]. This measure quantifies the confusionness or irregularity of signal along each channel. ECG signal with atrial fibrillation pathology has higher value of sample entropy than that normal sinus rythem [45]. In this work the MSampEn(e) is evaluated for original and compressed MECG signal and the diagnostic measure is defined as

$$MSD = e_o - e_r \quad (2.48)$$

where e_o and e_r are the multivariate sample entropy values for original and processed MECG signals.

2.5 Results and Discussions

For evaluation of the proposed diagnostic measure, the combination of PCA and discrete cosine transform based MECG compression framework is used. The DCT based MECG compression was proposed in [43]. In this work, first the PCA is applied to MECG signal and according to diagnostic importance the first six principal components (PCs) are retained [40] and others PCs are discarded. The DCT is applied to the PCA domain multivariate signal and due to energy compaction property the first few samples are retained. The transformed multivariate signal is uniformly quantized and huffman encoded for compression. Both 8-bit and 10-bit quantizer are used in this work. The compressed (reconstructed) signal is obtained by huffman decoding, inverse quantization followed by inverse DCT and PCA based reconstruction. The MECG signals are taken from PTB diagnostic [46] database. In this database, each of the MECG signals has sampled frequency and number of bits of 1000Hz and 16 respectively. Figures 2.2 and 2.3 shows the compressed signals along-with the original signals for four channels (lead-I, lead-aVL, lead-V2 and lead-V5) for healthy control (HC) and myocardial infarction (MI) pathology based MECG. It is observed that the compressed signals are denoised and the clinical components are preserved in both the cases (HC and MI). The PRD, WEDD and RMSE distortion measures are evaluated along each lead and the average of these values are defined as

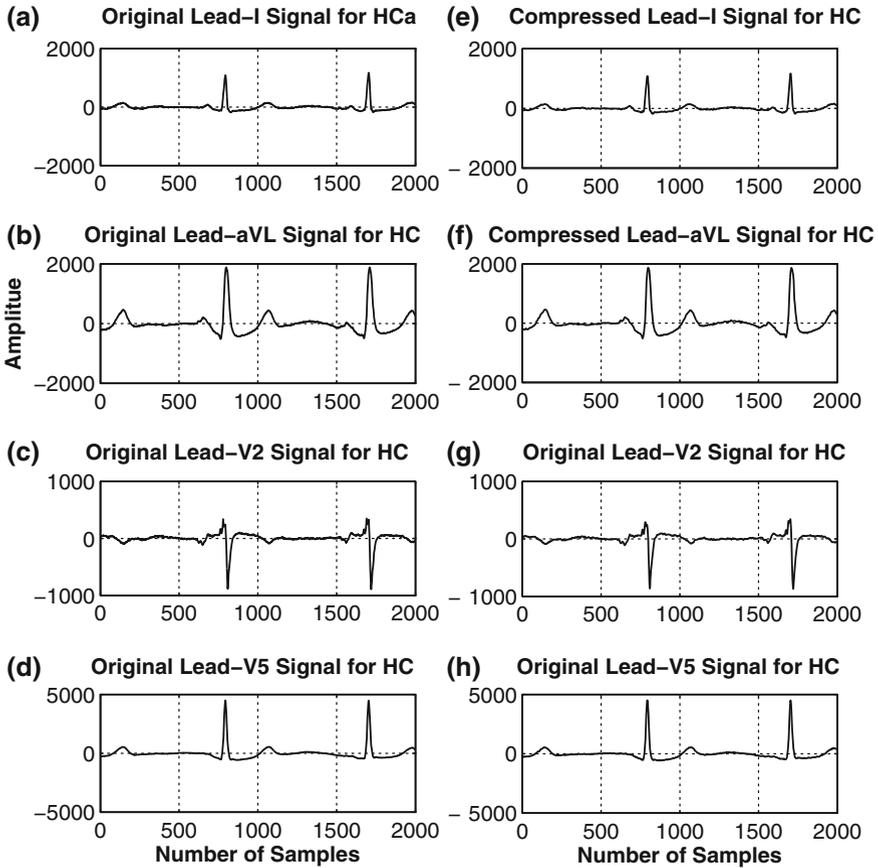


Fig. 2.2 Original signals (a), (b), (c) and (d) and Compressed signals (e), (f), (g) and (h) for Lead-I, aVL, V2 and V5 for HC based MECCG respectively

multichannel PRD (MPRD), multichannel WEDD (MWEDD) and multichannel RMSE (MRMSE).

The proposed MSD measure is also evaluated and compared with the MPRD, MWEDD and MRMSE measures. As the proposed measure is based on multivariate sample entropy, the irregularity in clinical components present in processed MECCG signal results a negative value of MSD. The MSD, e_o , e_r , MPRD, MWEDD and MRMSE values for HC and MI pathology based MECCG signals for different compression ratios (CRs) are shown in Table 2.1. Results are shown for 10 multichannel ECG signals with different compression ratios, which include normal and pathological cases. In HC1 dataset Compression ratio (CR) value of 13.16:1 produces MPRD, MWEDD and MRMSE values of 4.36 %, 1.98 % and 8.51 for processed MECCG signal. It is observed that for high values of CR, the corresponding MPRD and MWEDD, MRMSE values are also high. This signifies the higher distortion in clinical infor-

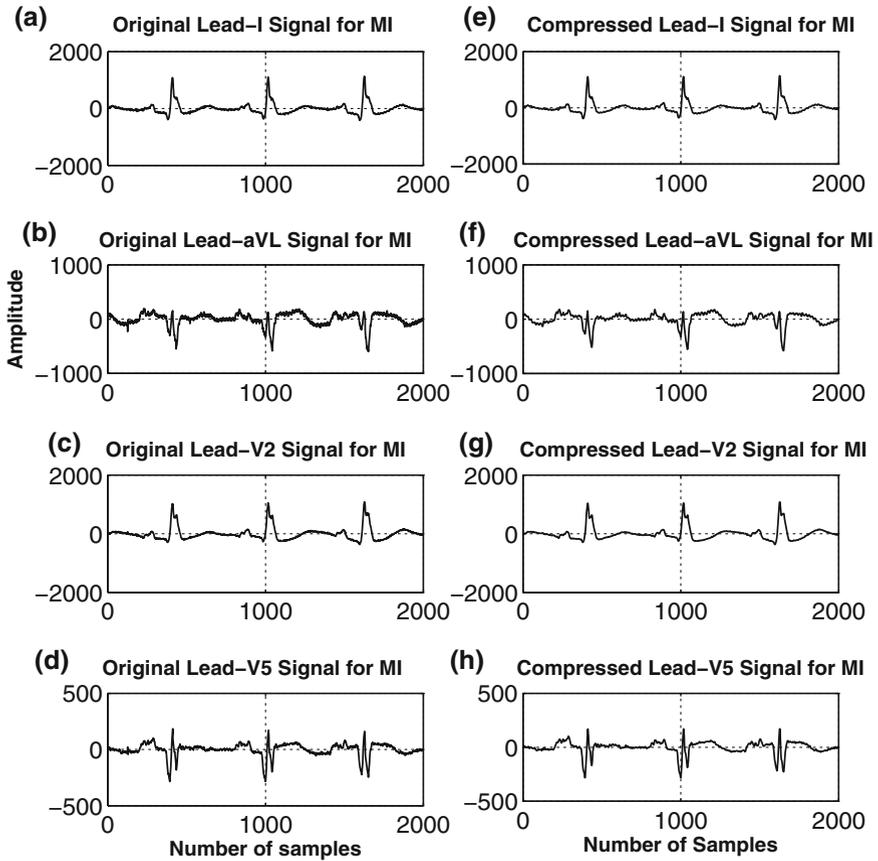


Fig. 2.3 Original signals (a), (b), (c) and (d) and Compressed signals (e), (f), (g) and (h) for Lead-I, aVL, V2 and V5 for MI pathology based MECG respectively

mation for the processed MECG signals with higher CR values. For HC1 dataset the multivariate sample entropy values for original and processed MECG signals are found to be 0.69 and 0.68. The MSD value is very small (0.01). This signifies that the processed MECG signal is regular (not changing) and the clinical components are also preserved. Similarly, for HC3 dataset with 8 bit quantization level, the CR is found to be 15.70:1. This CR value produces MPRD, MWEDD and MRMSE values of 4.08 %, 2.85 % and 14.64. The multivariate sample entropy values for original and processed MECG signals are found to be 1.01 and 1.03. The MSD measure shows a negative value of -0.02 , this signifies there may be the slight change in the regularity of clinical components ('PQRST' morphologies) present in processed MECG signals. For MI2 dataset with 8-bit quantizer the CR value of 14.44:1 produces a MPRD, MWEDD and MRMSE values of 6.37 %, 3.41 % and 15.00. The multivariate sample entropy for original and processed MECG signals are found to be 1.82 and 1.85.

Table 2.1 Performance of MSD, MWEDD, MPRD and MRMSE over MECG data compression

Record	Q. Bit	CR	MPRD (%)	MWEDD (%)	MRMSE	e_o	e_r	MSD
HC1	8	13.16	4.36	1.98	8.51	0.69	0.68	0.01
HC2	8	15.11	4.75	2.73	9.72	1.70	1.71	-0.01
HC3	8	15.70	4.08	2.85	14.64	1.01	1.03	-0.02
HC4	8	14.80	4.36	2.71	11.51	1.35	1.34	0.01
HC5	8	14.73	5.27	2.85	13.26	1.01	0.94	0.06
HC1	10	10.66	3.62	1.04	6.66	0.69	0.66	0.02
HC2	10	12.03	2.78	0.95	5.42	1.70	1.69	0.01
HC3	10	12.73	1.72	0.78	5.58	1.01	1.02	-0.01
HC4	10	11.78	3.12	0.83	7.65	1.35	1.34	0.01
HC5	10	11.58	4.01	1.09	8.69	1.01	0.96	0.04
MI1	8	13.44	6.78	3.57	11.41	1.51	1.53	-0.02
MI2	8	14.44	6.37	3.41	15.00	1.82	1.85	-0.02
MI3	8	14.70	6.18	2.59	13.80	1.53	1.49	0.04
MI4	8	13.01	9.76	3.99	13.98	1.31	1.34	-0.02
MI5	8	14.71	5.62	3.25	16.62	2.03	2.01	0.02
MI1	10	10.81	5.76	2.06	8.95	1.51	1.48	0.02
MI2	10	11.52	5.26	2.07	11.24	1.82	1.85	-0.03
MI3	10	11.62	5.35	1.50	10.92	1.53	1.49	0.04
MI4	10	10.55	9.11	2.95	12.23	1.31	1.28	0.03
MI5	10	11.58	4.38	1.71	12.15	2.03	2.04	-0.01

From this observation it is clear that the low MPRD, MWEDD and MRMSE does not guarantee the processed MECG signals to be clinically significant. The deviation of multivariate sample entropy from original value in MI2 dataset, shows that there is a small variation in the clinical components present in the processed MECG signal. This change is correctly quantified with the help of proposed MSD based diagnostic measure. The results for all other MECG dataset shows that the MSD measure is effective to quantify the regularity in the clinical component present in original and processed MECG signals.

2.6 Conclusion

In this paper, a review on state-of-art diagnostic information extraction approaches and their applications in various ECG signal processing such as filtering, compression and disease detection are presented. Then, a new diagnostic measure is proposed for evaluating the quality of multi lead ECG (MECG) signals. This measure (MSD) is

defined as the difference between multivariate sample entropy values for original and processed MCEG signals. The effectiveness of proposed MSD measure alongwith MPRD, MWEDD, MRMSE are tested over MCEG data compression framework. Comparison shows that the proposed measure is effective in quantifying clinical information in normal and pathological MCEG signals.

References

1. Opie, L.H.: *Heart Physiology: From Cell to Circulation*. Lippincott Williams & Wilkins, Philadelphia (2004)
2. Goldberger, A.L.: *Clinical Electrocardiography: A Simplified Approach*. Elsevier Health Sciences, Philadelphia (2012)
3. De Luna, A.B.: *Clinical Electrocardiography: A Textbook*. Wiley, Chichester (2012)
4. Sundnes, J., Lines, G.T., Cai, X., Nielsen, B.F., Mardal, K.-A., Tveito, A.: *Computing the Electrical Activity in the Heart*, vol. 1. Springer, New York (2007)
5. McSharry, P.E., Clifford, G.D., Tarassenko, L., Smith, L.A.: A dynamical model for generating synthetic electrocardiogram signals. *IEEE Trans. Biomed. Eng.* **50**(3), 289–294 (2003)
6. Malmivuo, J., Plonsey, R.: *Bioelectromagnetism: Principles and Applications of Bioelectric and Biomagnetic Fields*. Oxford University Press, New York (1995)
7. Fereniec, M., Maniewski, R., Karpinski, G., Opolski, G., Rix, H.: High-resolution multichannel measurement and analysis of cardiac repolarization. *Biocybern. Biomed. Eng.* **28**(3), 61–69 (2008)
8. Thaler, M.S.: *The Only EKG Book You'll Ever Need*, vol. 365. Lippincott Williams & Wilkins, Philadelphia (2010)
9. Manikandan, M.S., Dandapat, S.: Wavelet energy based diagnostic distortion measure for ECG. *Biomed. Signal Process. Control* **2**(2), 80–96 (2007)
10. Bradie, B.: Wavelet packet-based compression of single lead ECG. *IEEE Trans. Biomed. Eng.* **43**(5), 493–501 (1996)
11. de Filho, E.B.L., da Silva, E.A.B., de Carvalho, M.B., da Silva Jr, W.S., Koiller, J.: Electrocardiographic signal compression using multiscale recurrent patterns. *IEEE Trans. Circuits Syst. I: Regul. Pap.* **52**(12), 2739–2753 (2005)
12. Miaou, S.-G., Yen, H.-L.: Quality driven gold washing adaptive vector quantization and its application to ECG data compression. *IEEE Trans. Biomed. Eng.* **47**(2), 209–218 (2000)
13. Tai, S.C.: ECG data compression by corner detection. *Med. Biol. Eng. Comput.* **30**(6), 584–590 (1992)
14. Chen, J., Itoh, S.: A wavelet transform-based ECG compression method guaranteeing desired signal quality. *IEEE Trans. Biomed. Eng.* **45**(12), 1414–1419 (1998)
15. D'Ambrosio, A.C., Ortiz-Conde, A., Sanchez, E.J.G.: Percentage area difference (PAD) as a measure of distortion and its use in maximum enclosed area (MEA), a new ECG signal compression algorithm. In: *Proceedings of the Fourth IEEE International Caracas Conference on Devices, Circuits and Systems*, pp. I035-1–I035-5 (2002)
16. Chou, H.-H., Chen, Y.-J., Shiau, Y.-C., Kuo, T.-S.: An effective and efficient compression algorithm for ECG signals with irregular periods. *IEEE Trans. Biomed. Eng.* **53**(6), 1198–1205 (2006)
17. Zigel, Y., Cohen, A., Katz, A.: The weighted diagnostic distortion (WDD) measure for ECG signal compression. *IEEE Trans. Biomed. Eng.* **47**(11), 1422–1430 (2000)
18. Al-Fahoum, A.S.: Quality assessment of ECG compression techniques using a wavelet-based diagnostic measure. *IEEE Trans. Inf. Technol. Biomed.* **10**(1), 182–191 (2006)
19. Manikandan, M.S., Dandapat, S.: Multiscale entropy-based weighted distortion measure for ECG coding. *IEEE Signal Process. Lett.* **15**, 829–832 (2008)

20. Manikandan, M.S., Dandapat, S.: Effective quality-controlled SPIHT-based ECG coding strategy under noise environments. *Electron. Lett.* **44**(20), 1182–1183 (2008)
21. Pecchia, L., Melillo, P., Bracale, M.: Remote health monitoring of heart failure with data mining via CART method on HRV features. *IEEE Trans. Biomed. Eng.* **58**(3), 800–804 (2011)
22. Acharya, U.R., Faust, O., Sree, S.V., Ghista, D.N., Dua, S., Joseph, P., Ahamed, V.I.T., Janarthanan, N., Tamura, T.: An integrated diabetic index using heart rate variability signal features for diagnosis of diabetes. *Comput. Methods Biomech. Biomed. Eng.* **16**(2), 222–234 (2013)
23. Ramirez-Villegas, J.F., Lam-Espinosa, E., Ramirez-Moreno, D.F., Calvo-Echeverry, P.C., Agredo-Rodriguez, W.: Heart rate variability dynamics for the prognosis of cardiovascular risk. *PLoS ONE* **6**(2), e17060 (2011)
24. Schmitt, L., Regnard, J., Desmarests, M., Mauny, F., Mourou, L., Fouillot, J.-P., Coulmy, N., Millet, G.: Fatigue shifts and scatters heart rate variability in elite endurance athletes. *PLoS ONE* **8**(8), e71588 (2013)
25. Chouchou, F., Pichot, V., Barthélémy, J.-C., Bastuji, H., Roche, F.: Cardiac sympathetic modulation in response to apneas/hypopneas through heart rate variability analysis. *PLoS ONE* **9**(1), e86434 (2014)
26. Banerjee, S., Mitra, M.: Application of cross wavelet transform for ECG pattern analysis and classification. *IEEE Trans. Instrum. Meas.* **63**(2), 326–333 (2014)
27. Jayachandran, E.S., et al.: Analysis of myocardial infarction using discrete wavelet transform. *J. Med. Syst.* **34**(6), 985–992 (2010)
28. Ge, D., Srinivasan, N., Krishnan, S.M.: Cardiac arrhythmia classification using autoregressive modeling. *Biomed. Eng. Online* **1**(1), 5 (2002)
29. Martis, R.J., Acharya, U.R., Lim, C.M., Suri, J.S.: Characterization of ECG beats from cardiac arrhythmia using discrete cosine transform in PCA framework. *Knowl.-Based Syst.* **45**, 76–82 (2013)
30. Langley, P., Bowers, E.J., Murray, A.: Principal component analysis as a tool for analyzing beat-to-beat changes in ECG features: application to ECG-derived respiration. *IEEE Trans. Biomed. Eng.* **57**(4), 821–829 (2010)
31. Widjaja, D., Varon, C., Dorado, A.C., Suykens, J.A.K., Van Huffel, S.: Application of kernel principal component analysis for single-lead-ECG-derived respiration. *IEEE Trans. Biomed. Eng.* **59**(4), 1169–1176 (2012)
32. Martis, R.J., Acharya, U.R., Min, L.C.: ECG beat classification using PCA, LDA, ICA and discrete wavelet transform. *Biomed. Signal Process. Control* **8**(5), 437–448 (2013)
33. Giri, D., Acharya, U.R., Martis, R.J., Sree, S.V., Lim, T.-C., Ahamed, T., Suri, J.S.: Automated diagnosis of coronary artery disease affected patients using LDA, PCA, ICA and discrete wavelet transform. *Knowl.-Based Syst.* **37**, 274–282 (2013)
34. Sun, L., Yanping, L., Yang, K., Li, S.: ECG analysis using multiple instance learning for myocardial infarction detection. *IEEE Trans. Biomed. Eng.* **59**(12), 3348–3356 (2012)
35. Pan, J., Tompkins, W.J.: A real-time QRS detection algorithm. *IEEE Trans. Biomed. Eng.* **32**(3), 230–236 (1985)
36. Richman, J.S., Moorman, J.R.: Physiological time-series analysis using approximate entropy and sample entropy. *Am. J. Physiol.-Heart Circ. Physiol.* **278**(6), H2039–H2049 (2000)
37. Alonso-Atienza, F., Morgado, E., Fernandez-Martinez, L., Garcia-Alberola, A., Rojo-Alvarez, J.L.: Detection of life-threatening arrhythmias using feature selection and support vector machines. *IEEE Trans. Biomed. Eng.* **61**(3), 832–840 (2014)
38. Li, Q., Rajagopalan, C., Clifford, G.D.: Ventricular fibrillation and tachycardia classification using a machine learning approach. *IEEE Trans. Biomed. Eng.* **61**(6), 1607–1613 (2014)
39. Huang, J.-R., Fan, S.-Z., Abbod, M.F., Jen, K.-K., Wu, J.-F., Shieh, J.-S.: Application of multivariate empirical mode decomposition and sample entropy in EEG signals via artificial neural networks for interpreting depth of anesthesia. *Entropy* **15**(9), 3325–3339 (2013)
40. Sharma, L.N., Dandapat, S., Mahanta, A.: Multichannel ECG data compression based on multiscale principal component analysis. *IEEE Trans. Inf. Technol. Biomed.* **16**(4), 730–736 (2012)

41. Fleureau, J., Kachenoura, A., Albera, L., Nunes, J.-C., Senhadji, L.: Multivariate empirical mode decomposition and application to multichannel filtering. *Signal Process.* **91**(12), 2783–2792 (2011)
42. Sharma, L.N., Dandapat, S.: Compressed sensing for multi-lead electrocardiogram signals. In: 2012 World Congress on Information and Communication Technologies (WICT), pp. 812–816, October 2012
43. Cetin, A.E., Koymen, H., Aydin, M.C.: Multichannel ECG data compression by multirate signal processing and transform domain coding techniques. *IEEE Trans. Biomed. Eng.* **40**(5), 495–499 (1993)
44. Ahmed, M.U., Mandic, D.P.: Multivariate multiscale entropy analysis. *IEEE Signal Process. Lett.* **19**(2), 91–94 (2012)
45. Alcaraz, R., Rieta, J.J.: A review on sample entropy applications for the non-invasive analysis of atrial fibrillation electrocardiograms. *Biomed. Signal Process. Control* **5**(1), 1–14 (2010)
46. Oeff, M., Koch, H., Bousseljot, R., Kreiseler, D.: The PTB diagnostic ECG database. National Metrology Institute of Germany. <http://www.physionet.org/physiobank/database/ptbdb> (2012)

Chapter 3

ZnO Nanostructures for Alternate Energy Generation

Sunandan Baruah

Abstract Extensive use of fossil fuel in industries and automobiles has severely polluted the environment, adversely affecting the ecosystem. The fossil fuel reserves are also dwindling, creating a serious concern in the area of energy generation. With rapid advances in nanotechnology, researchers are putting in their efforts to exploit unique properties of nanomaterials to come up with environmentally friendly energy solutions. The abundantly freely available solar energy is undoubtedly the least utilized form of natural energy. Efficient tapping of solar energy can resolve the energy crisis that our world is currently going through. Solar cells developed using nanomaterials, though still at the infancy stage, will be able to harness solar energy quite efficiently and most importantly, will be able to do it very cheaply. Piezoenergy resulting from physical deformation of near-elastic crystals shows promise as energy source for self-powering of low energy consuming devices. This article discusses the possibility of using nanostructures of a very promising material, zinc oxide (ZnO), for energy generation. ZnO is a wide bandgap semiconductor (3.37 eV) and the absence of a central symmetry in its crystal endows it with piezoelectric property. This material has been successfully used for energy generation and tapping schemes like solar cells, hydrogen generators and piezogenerators, among others.

Keywords ZnO · Solar energy · Wide bandgap semiconductor · Piezogenerators · Lowest unoccupied molecular orbitals (LUMO)

3.1 Introduction

The amount of energy showered on the earth's surface by the sun in 1 h is more than the energy consumed by humans in an entire year. The sun is nothing but an energy fountain providing earth with an unlimited supply of energy. The best part is that the sun is doing this without any carbon emission! The challenge to tap the enormous amount of freely available energy has been urging researchers to develop

S. Baruah (✉)
Assam Don Bosco University, Guwahati 781017, India
e-mail: sunandan.baruah@dbuniversity.ac.in

© Springer India 2015
P.K. Bora et al. (eds.), *Advances in Communication and Computing*,
Lecture Notes in Electrical Engineering 347, DOI 10.1007/978-81-322-2464-8_3

novel materials and strategies for designing photovoltaic systems to convert sunlight into electricity. Research efforts for efficient tapping of solar energy is concentrated in two broad areas: improving the efficiency of conventional solar cells made from silicon and development of new semiconductors, inorganic nanocrystals, organic polymers, and many other light-sensitive materials. Inorganic colloidal dispersions, like organic materials, can be deposited onto solid substrates as thin films through standard techniques like spin-coating or dip-coating. These films act as active layers for applications like gas sensors and solar cells. Compared with typical high-temperature and high-vacuum vapor-deposition methods used for fabricating silicon solar cells, solution-phase procedures can offer significant cost savings. Simple low cost techniques can be successfully utilized to form nearly defect-free crystals ideal for the development of solar cells. On the other hand, the same techniques can be used to tailor crystal defects, which are needed for applications involving surface reactions like photocatalysis and environmental sensors.

Nanostructured ZnO materials have received considerable interest from scientists due to their remarkable performance in electronics, optics and photonics. As early as the 1960s, synthesis of ZnO thin films was an active area because of possible applications in sensors, transducers, and as photocatalysts. In the past few decades, study of one-dimensional material has gained importance in nanoscience and nanotechnology. With reduction in size, novel electrical, mechanical, chemical and optical properties are introduced, which are understood to be the result of surface and quantum confinement effects. ZnO is a significant technological material. The absence of a center of symmetry in its wurtzite structure, along with large electromechanical coupling, results in strong piezoelectric and pyroelectric properties. ZnO is therefore widely used in mechanical actuators and piezoelectric sensors. In addition, ZnO is a wide bandgap (3.37 eV) compound semiconductor that is appropriate for short wavelength optoelectronic applications. The high exciton binding energy (60 meV) in ZnO crystal can result in efficient excitonic emission at room temperature. ZnO is transparent to visible light and its conductivity can be tailored. A wide range of high technology applications ranging from surface acoustic wave filters [1], photonic crystals [2], photodetectors [3], light emitting diodes [4], photodiodes [5], gas sensors [6], optical modulator waveguides [7], solar cells [8, 9] to varistors [10]. ZnO is also receiving a lot of attention because of its antibacterial property and its bactericidal efficacy has been reported to increase as the particle size decreases [11].

Different nanostructures of ZnO have been reported in the literature, such as nanowires and nanorods [12, 13], nanocombs [14], nanorings [15], nanoloops and nanohelices [16], nanobows [17], nanobelts [18], nanocages [19] amongst others. These structures have been successfully synthesized under specific growth conditions [13]. The synthesis of ZnO nanostructures can be carried out both in solution and gaseous environments. The gas phase synthesis methods are expensive and require expensive setups. The solution phase synthesis is normally done in water or other mild solvents and is normally referred to as the hydrothermal process. This process of growing ZnO nanostructures has gained immense popularity due to its simplicity and tolerable growth conditions. As synthesis of the nanostructures is done in aqueous solution, the growth conditions are mild with growth temperatures less than

the boiling point of water. Researchers are concentrating on inorganic nanocrystalline materials as they can be processed through easily scalable techniques. The growth parameters like temperature, pressure, pH, etc., can be controlled to fabricate various morphologies. These advantages of colloidal chemistry open up possibilities for low-cost manufacturing of devices. A computer-controlled setup for the growth of ZnO nanorods of various dimensions on any type of substrates (up to A4 size) was developed at the Asian Institute of Technology in Thailand aimed at scaling-up devices like dye-sensitized solar cells and gas sensors.

Dye-sensitized solar cells (DSSC) have been widely studied over the past decade and have been considered as a technically and economically plausible alternative concept to the contemporary p-n junction solar cells [23][228–230]. Dye-sensitized solar cells (DSSC) are composed of nanostructured metal oxide semiconductor films, of which titanium dioxide (TiO_2), zinc oxide (ZnO), and tin dioxide (SnO_2) gaining important status [23][228–230]. In a conventional photovoltaic system, the semiconductor serves the dual role of light absorption and charge carrier transport. In a DSSC, light is absorbed by a sensitizer dye adsorbed onto the surface of a wideband semiconductor. Charge separation takes place at the interface through the injection of photo-excited electrons from the dye into the conduction band of the semiconductor. Carriers are transported in the conduction band of the semiconductor and move to the external circuit [20].

Advances in embedded technology have led to the downscaling of MEMS/NEMS devices to unimaginable limits. These devices, in spite of consuming very low power, still require independent energy sources. Self-powering of these devices has remained an issue and is an emerging field of research in state-of-the-art technologies. Self-powering devices, especially those that consume very low power, are being developed relying on the voltage generated by piezoelectric crystals when they are mechanically deformed. Piezoelectric energy harvesters in sub-micron scale have tremendous possibilities for application in futuristic technologies and are catching the fancy of researchers. Life support electronic devices implanted inside human bodies like pacemakers and real-time diagnostic tools like halter monitors can be powered by the vibrations picked up from heartbeats. Vibrations of the diaphragm of a wireless microphone can be translated to bend piezotronic crystals and tapping the voltage generated can make it a self-powered device. Nanowires put on radiators of cars will be bent when the car is moving and this can be used to power small gadgets like music systems and GPS. These are just a few examples and the possibilities are endless!

3.2 Dye/Quantum Dot Sensitized Solar Cell

Dye- and Quantum Dot sensitized solar cells (DSSC/QDSSC) are receiving a lot of attention from researchers because of their promise of achieving high efficiency at low manufacturing cost. The fabrication procedure is simple and does not require sophisticated facilities. The best part is that the entire fabrication process is

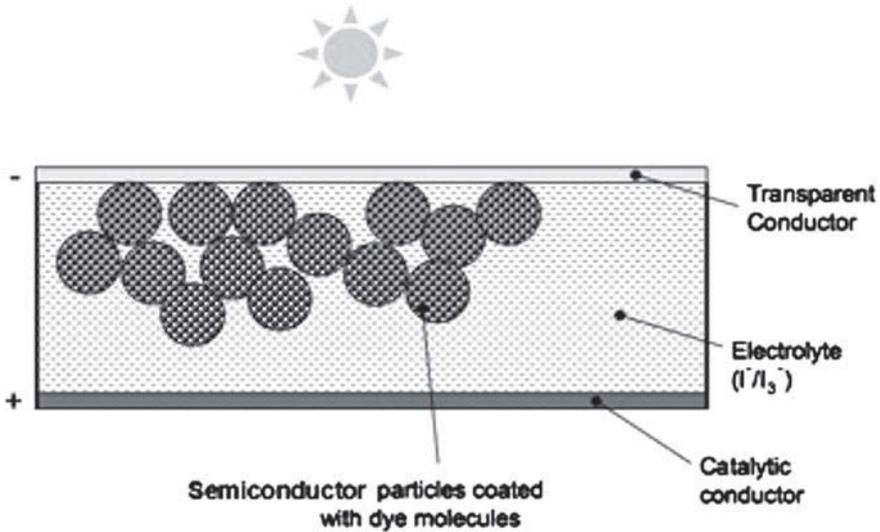


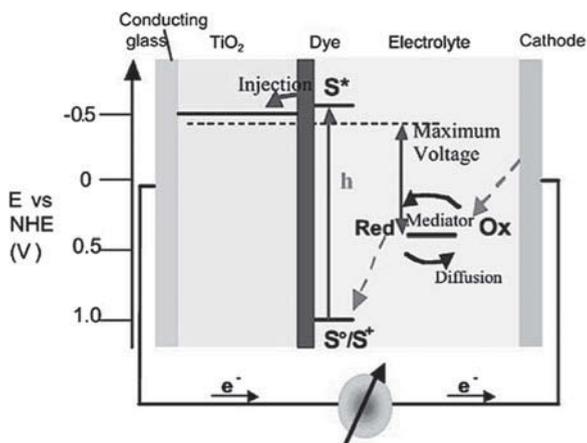
Fig. 3.1 A typical structure of Dye-Sensitized Solar Cell (DSSC) [21]

environmentally friendly, unlike Si solar cells. In normal p–n junction solar cells, charge creation (the process of generation of electron-hole pairs) and charge transfer (diffusion of electrons and holes) occurs in the same material (typically silicon). However, in DSSC/QDSSC, these two main processes occur in two different materials.

In order to properly comprehend the operation of a DSSC/QDSSC, it is important to first understand its construction. A typical DSSC has four main components in its structure: a photoelectrode, the sensitizer (dye or QD), the electrolyte, and the counter electrode. Figure 3.1 shows the different components of a typical DSSC; the replacement of the dye molecules with QDs will make it a QDSSC.

The photoelectrode is a transparent substrate; typically glass coated with a thin layer of doped tin oxide (SnO_2) for conductivity with sheet resistance as low as 12–18. A thin porous film of metal oxide semiconductor (TiO_2 or ZnO) is deposited onto the conducting side of the substrate through techniques like chemical vapour deposition (CVD), hydrothermal process, plasma deposition, spin coating, etc. The metal oxides, being wide bandgap semiconductors with bandgaps above 3 eV extract the electrons from the excited dye molecules/QDs adsorbed on to their surface. The extracted electrons then diffuse through the semiconductor to generate the current in the cell. Thickness of the semiconducting thin film in a DSSC is typically about 10 μm . The dye/QD are sensitizers that are adsorbed on the surface of the metal oxide semiconductor thin film. On light irradiation, the dye molecules/QDs absorb the photons and generate electron–hole pairs. The electrons from the lowest unoccupied molecular orbital (LUMO) of the dye get injected into the conduction band of the semiconductor as its energy is slightly lower than the LUMO of the dye. The

Fig. 3.2 Principle of operation showing the energy levels of the dye-sensitized solar cell [20]



electrolyte is a solution that contains redox couples (both oxidized and reduced forms) like I^-/I_3^- , Br^-/Br_3^- , etc. The function of the electrolyte in a DSSC is to reduce the oxidized dye molecules by supplying electrons to replace the ones being extracted by the semiconductor. The counter electrode is similar to the photoelectrode with the semiconducting layer on the latter being replaced by a metallic layer of nanoparticles (mostly Pt or Pd). The electrolyte, which gets oxidized by transferring electrons to the dye gets reduced at the counter electrode interface. The metallic thin film is deposited on the counter electrode to act as a catalyst to accelerate the reaction at the electrolyte/counter electrode interface.

Figure 3.2 shows a schematic representation of the operation of a DSSC [20] and Fig. 3.3, a comparison of the process of charge creation and transfer in a P–N junction solar cell and a DSSC [22]. As shown in Fig. 3.2, photoexcitation of the sensitizer (S) results in electron injection into the conduction band of semiconducting metal oxides prepared as a thin film. The dye molecule is regenerated by the redox system, which undergoes a similar process at the counter electrode by the electrons passing through the external load. The open circuit voltage of the DSSC corresponds to the

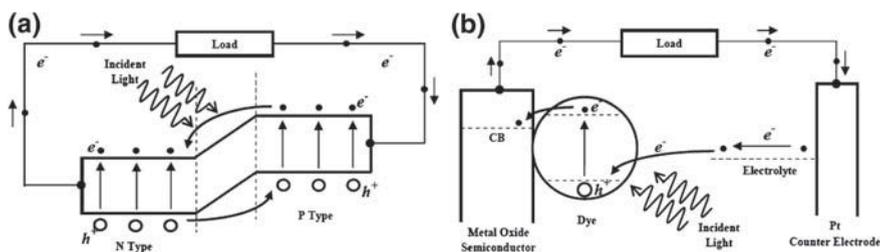


Fig. 3.3 Process of charge creation and charge transfer in **a** P–N junction and **b** Dye-sensitized solar cell (DSSC) [22]

difference between the redox potential of the mediator and the Fermi level of the nanocrystalline film (shown with a dashed line in Fig. 3.2). The energy levels shown for the sensitizer and the redox mediator match the redox potentials of the ground state of the N3 sensitizer, which is deprotonated twice and the iodide/triiodide couple [20].

The process of charge creation in a DSSC occurs in the sensitizer (dye or quantum dot) molecules adsorbed on the surface of the metal oxide semiconductor film, on exposure to light. Excited photoelectrons get injected into the conduction band of the metal oxide semiconductor. Finally, the injected electrons diffuse through the semiconductor layer thereby generating current.

On illumination of the photoelectrode of a DSSC/QDSSC with light, the dye molecules/QDs adsorbed on the nanostructured semiconductor film absorb energy from the impinging photons, thereby creating electron–hole pairs. The electrons get energized to the lowest unoccupied molecular orbitals (LUMO) from the highest occupied molecular orbitals (HOMO) of the dye molecules or to the valence band of the QDs from their conduction bands. The common dyes used by researchers for DSSC include ruthenium (Ru)II complexes, like N3 [23], N719 [24], and black dye (N749) [25]. Use of organic dyes like D5 [26] and D149 [27] has also been reported in the literature. The light harvesting capability of these inorganic dyes is comparable to their organic counterparts. For efficient photoabsorption, the dye must have an intense and flat absorption band in the visible range of the solar spectrum.

The presence of carboxylic or phosphonic acid anchoring groups in the dye molecules results in electronically favorable bonding of the dye with the semiconductor through the contribution of a proton. The excited electrons move from the quasi-stable LUMO of the dye to the conduction band of the semiconductor as it is at an energy level below it. The energy level positions of the conduction band of the semiconductor and the LUMO of the dye are parameters worth considering for efficient electron injection. It is imperative that there is an overlap of the LUMO of the dye overlaps with the conduction band of the semiconductor [22] for getting that near perfect bonding.

How does the charge transport in a dye/QD sensitized solar cell actually take place? The photo-excited electrons injected into the semiconductor from the dye molecules diffuse to the base of the semiconductor thin film. These electrons then drift toward the counter electrode through the external circuit. The oxidized dye molecules reduce back to their original state by accepting electrons from the redox couple (I^-/I_3^-) in the electrolyte and the process repeats over a million times every second [20]. The oxidized electrolyte then quickly reduces its redox couple back the original state by accepting electrons from the counter electrode. This successive oxidation and reduction of dye and electrolyte continues as long as the excitation source is present. The redox potential of the electrolyte should be above the HOMO level of dye for efficient dye regeneration [20, 28]. The difference in the Fermi level of the semiconductor and the redox potential of the electrolyte gives a measure of the maximum open circuit voltage of a DSSC [20]. For optimum performance of the DSSC/QDSSC, the electron injection should take place at a much faster rate compared to the time needed for regeneration of the dye/QD; typically they are in the picosecond and

nanosecond ranges respectively. The probability of recombination of the electrons is very high at interfaces between different materials. A DSSC has many interfaces like dye/semiconductor/electrolyte interface, electrolyte/counter electrode interface, among others. For efficient performance of a cell, it is important that the charge transport through the cell should be faster than the interfacial recombination [20].

3.3 Zinc Oxide in DSSC Photoelectrodes

TiO_2 has been holding an important position as the semiconductor of choice for the photoelectrode of DSSCs for quite some time. However, of late, another wide bandgap semiconductor, zinc oxide (ZnO) is showing potential to replace TiO_2 in DSSCs [29]. The advantages that ZnO has over TiO_2 are a direct bandgap (3.37 eV) and a high exciton binding energy of 60 mV [30, 31]. The higher electron mobility in ZnO ($200 \text{ cm}^2/\text{V/s}$) compared to TiO_2 ($30 \text{ cm}^2/\text{V/s}$) makes it a better choice for use in DSSCs [32]. ZnO is perhaps the only material that has the largest reported nanostructures like nanoparticles, nanowires, nanorods, nanosheets, nanobelts, nanoneedles, nanotubes hierarchical nanostructures, etc. Researchers use this varied plethora of different nanostructures as the semiconducting layer to study DSSC/QDSSC performance [33–36]. Electron hopping over crystallite boundaries can be reduced by controlling the structure of single crystalline ZnO resulting in an enhancement of the electron mobility. After the first report on the use of single crystal ZnO as dye sensitized electrodes in 1969 [37], there was practically no research on its use for over a decade. The interest in ZnO for DSSC use was rekindled after Matsumura et al. [38] reported the use of ZnO porous disks in DSSC electrodes. During the past decade, escalating fossil fuel prices and the depletion of its reserves has led to a renewed interest in solar cells in general and DSSCs in particular considering economical power generation. Numerous reports are available on the use of ZnO nanostructures in DSSC.

Research in the use of ZnO as a photoelectrode material of DSSC took a new turn after Law et al. [29] reported a DSSC using hydrothermally grown ZnO nanowires with high possibility for direct electron conduction. The small cell with an active area of only 0.2 cm^2 displayed a higher open circuit voltage (VOC) and short circuit density (JSC) than a larger version (0.8 cm^2). The fill factor and efficiency then reported were 0.37 and 1.51 %\$ and 0.38 and 1.26 %\$, respectively; definitely are not attractive numbers to consider. A schematic diagram of the DSSC and a scanning electron micrograph of the ZnO nanowires are shown in Fig. 3.4. Figure 3.5a shows the IV characteristics for the DSSCs with active areas 0.2 and 0.8 cm^2 . The open-circuit voltages and fill factors plotted as a function of light intensity are shown in Fig. 3.5b.

Akhtar et al. [33] used ZnO with different morphologies like flowers, spheres, etc., as DSSC photoelectrodes. They observed that sheet-sphere ZnO were highly crystalline and had a large surface area to volume ratio. This was reflected in the photovoltaic performance with the overall conversion efficiency of 2.61 % in

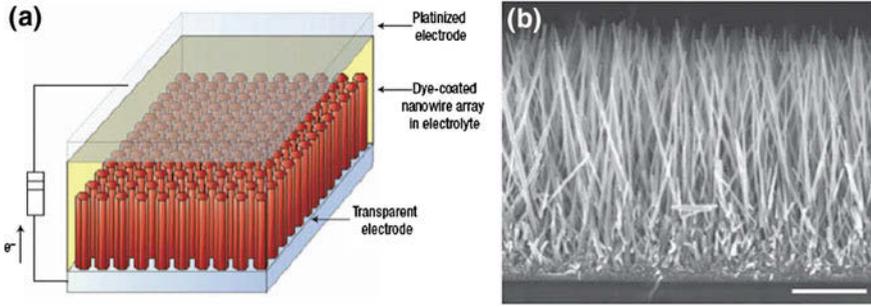


Fig. 3.4 a Schematic diagram showing the ZnO nanowire DSSC b scanning electron micrograph of the ZnO nanowires. Scale bar: 5 μ m [29]

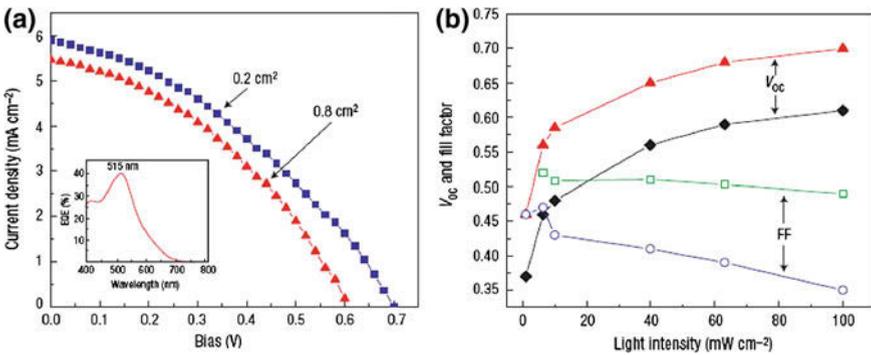


Fig. 3.5 a IV characteristic curves for ZnO nanowire DSSCs with active areas of 0.2 and 0.8 cm^2 b open circuit voltage and fill factor variations of the two cells with light intensity [29]

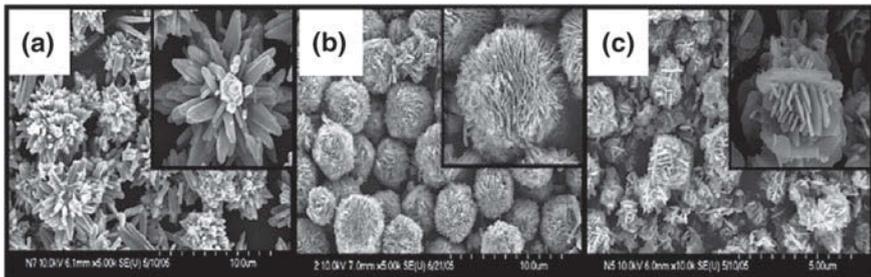


Fig. 3.6 Different ZnO nanostructures used as the semiconducting layer in DSSCs [33]. a flowers b porous balls c platelets

dye-sensitized solar cell (DSSC) fabricated with sheet-sphere ZnO as compared to the others. The scanning electron micrographs of the ZnO nanostructures used as the active material in the DSSCs are shown in Fig. 3.6.

It has been observed that the performance efficiency of ZnO-based DSSCs are less compared to TiO₂ based ones which can be attributed to better dye adsorption on TiO₂. The N719 dye solution used for DSSC is mildly acidic (pH 5.5) and as ZnO has a higher point of zero charge (8–9 pH), dissolution starts at the surface of the ZnO crystal resulting in the formation of $Zn^{2+}/dye\text{complex}$. This reduces the electron injection efficiency of the dye as a result of which fewer electrons get injected into the ZnO compared to TiO₂ [39, 40]. The dissolution rate from the ZnO surface gets enhanced due to the presence of formic acid molecules in the dye solution which are responsible for increase in the bond length between the Zn and O atoms at the surface of the ZnO crystal.

Keis et al. [41] reported the highest efficiency of 5 % for a ZnO nanoparticle porous thin film DSSC sensitized with the inorganic dye (N719) while a higher efficiency of 5.6 was achieved by Yoshida et al. [41] using an organic dye (D149). Tailoring the surface morphology of the photoelectrode can vary the effective surface area for dye adsorption/QD deposition. The increase in surface area may enhance the light harvesting capabilities of the DSSC/QDSSC. The ZnO nanorod is by far the most widely used structure for fabricating DSSC as electron recombinations that take place due to electron hopping is minimized. The single crystalline nanorods provide a direct path for the diffusion of electrons through them. Further, densely grown ZnO nanorod arrays can provide a very high surface area compared to volume. ZnO nanorods can be grown using simple hydrothermal methods [42].

Apart from having higher electron mobility, ZnO has other advantages over TiO₂ to make it a better semiconducting material for solar cells. In order to comprehend the charge transfer mechanisms in a ZnO nanoparticle/N719 complexation, the photoluminescence (PL) from the ZnO nanoparticles was taken to observe if the PL overlaps with the absorption spectrum of the N719 dye [43]. Figure 3.7 shows a HRTEM micrograph of the ZnO nanoparticles and their UV/Vis absorption spectra and that of dye N719. The photoluminescence (PL) spectrum of the ZnO NPs upon excitation with a 375 nm laser (Fig. 3.7) shows a significant spectral overlap with

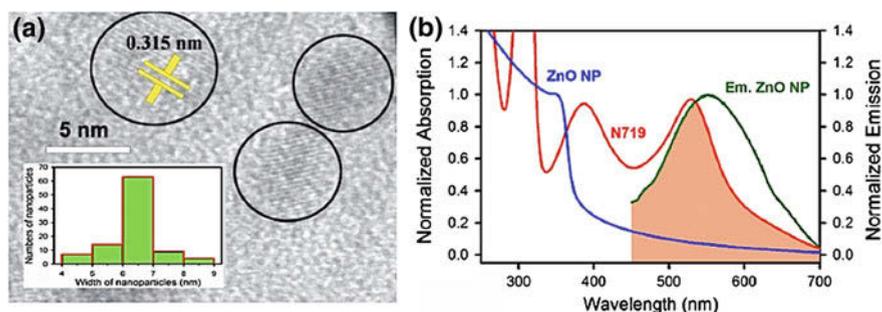


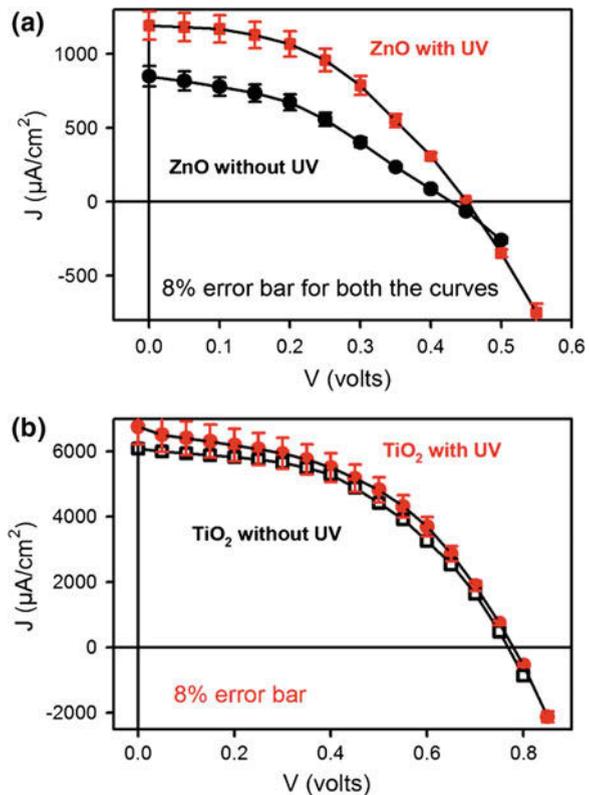
Fig. 3.7 a High resolution transmission electron micrograph (HRTEM) showing the ZnO nanoparticles with 6nm diameter with size distribution shown in *inset* b Steady state absorption spectra of ZnO NP, N719 and emission spectra of ZnO NP are shown. The overlapping zone between emission of ZnO NP and absorption of N719 is shaded [43]

that of the absorption of the N719. This indicates the possibility of transfer of energy from ZnO to the N719 molecules on photoexcitation.

3.4 ZnO as Piezoelectric Energy Generators

Single crystalline ZnO exhibits piezoelectric property due to its wurtzite structure, which lacks a central symmetry and the bonds are ionic. The nonsymmetric distribution of positive and negative charges starts at a unit cell and repeats through the whole crystal. A deformation in the crystal results in net polarization on the surface. Nanostructures of ZnO are capable of maintaining the versatility of their parent form and can undergo considerable deformation [44]. Anisotropic ZnO nanowires and nanorods have been of interest for building up self-powered devices, thereby opening up the path for size-tuned sources of power for devices undergoing miniaturization [45] (Fig. 3.8).

Fig. 3.8 a I–V curves of DSSC with photoelectrodes having **a** ZnO nanoparticles **b** TiO₂ nanoparticles, in presence and absence of UV light [43]



3.4.1 Principle of Piezoelectric Generation in ZnO Nanorods

If a vertical ZnO nanorod is deflected, the outer surface is stretched while the inner one is compressed resulting in an electric field along the c -axis of the hexagonal nanorod. The measure of the generated electric field is given by, $E_z = z/d$, where d is the piezoelectric coefficient along the length of the nanorod. Just below the tip, the electric potential approximately varies from V_s^+ at the stretched side to V_s^- at the compressed side. As a result of the strain due to applied pressure, there is a relative displacement of the Zn^{2+} ions with respect to O^{2-} ions in the ionic crystal lattice. This builds up a potential difference, which will persist till the strain is released. As a result, the differences in potential between the two edges of the nanorods remain as long as the deformation exists and the strain field is directly proportional to the amount of bending [46]. Figure 3.9 shows the distribution of the longitudinal strain and electric field along with the generated voltage across the length of the ZnO nanorod when subjected to deformation. On applying an external load during a scanning of vertically aligned ZnO nanowires in AFM contact mode sharp output voltage peaks could be observed [47].

ZnO is extremely piezoelectric [48] and has the largest piezoelectric coefficient considering tetrahedral crystals and exhibit high electro-mechanical coupling [49]. Nanostructures of ZnO, apart from having a high piezoelectric coefficient, also have high elasticity and can therefore be bent considerably [50]. Based on piezo-response force microscopy done using an AFM tip on ZnO nanobelt surface, Zhao et al. [51] opined that the piezopotential was higher at lower frequency of vibration. In the ZnO crystal, every cation is enclosed tetrahedrally by four anions. The negative

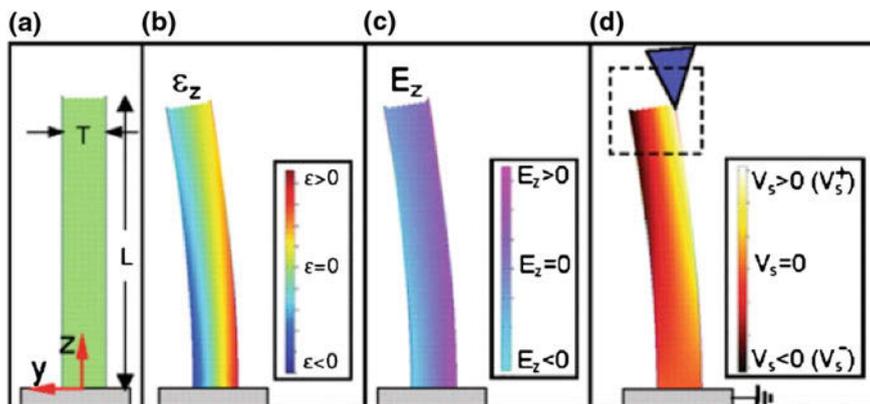


Fig. 3.9 Principle of power generation in a ZnO nanorod (a) schematic of the nanorod and coordinate system (b) longitudinal strain z distribution in the nanorod after deflection by an AFM tip (simulation for a ZnO NW of length 1 μ m and aspect ratio of 10) (c) corresponding longitudinal electric field E_z distribution in the nanorod induced by the piezoelectric effect. **d** distribution of the generated in the nanorod [47]

charges are localized at the tetrahedron center counterbalancing the positive charge to achieve crystal neutrality. On applying mechanical stress, the center of gravity of the anions gets shifted thereby creating an electric dipole. For pure crystals, with all unit cells having the same orientation, the electric charge separation (hence the potential difference) would be multiplied to an appreciable level. The final outcome will be two opposite charges accumulated at two opposite faces of the strained macrostructure.

The piezo-potential developed across a ZnO nanorod can be considerably controlled by controlling parameters like nanorod dimensions, magnitude of applied stress and carrier concentration. Mantini et al. [52] theoretically estimated these effects by modeling a ZnO nanorod as a cylindrical structure using finite element analysis, applying a uniform horizontal force on the top surface of the nanorod. They speculated an increase in the generated potential difference with increase in applied bending force which can be attributed to increased polarization resulting from increased strain. The magnitude of the piezo-potential is inversely proportional to the thickness of the nanorod, and beyond a certain height the nanorod dimension was found not affecting the voltage generated [53].

When a ZnO nanorod is laterally bent, positive and negative charges get accumulated on the elongated and compressed sides respectively. Apart from these charges, bandgap reduction also compels electrons to accumulate near the positive side. A screening effect results because of partial charge cancellations, which however is significant only on the elongated surface (positive charge) because of low donor concentration on the n-type nanowires [54]. An observation by Gao et al. [55] is that an n-type ZnO nanorod with a diameter of 50 nm and length 600 nm long having a carrier concentration of $1 \times 10^{17} \text{ cm}^{-3}$ generates a positive voltage of 0.05 V and a negative voltage of -0.3 V on application of a bending force of 80 nN.

In order to design robust energy generators, the mechanical properties of the piezo-electric nanostructures are of importance. Elasticity and tensile strength are parameters that need serious consideration for device fabrication. The nanorod morphology of ZnO exhibits higher resilience than bulk structures, and is therefore attracting a lot of attention. Stretching and bending capabilities were observed to decrease with decreasing diameters and were highest in the diameter range of 80–20 nm [56]. In an experiment, a ZnO nanorod with a diameter of 20 nm was subjected to tensile test and it was found to sustain a force up to 3.05 N. Nanorods are therefore favorable for applications involving piezo-potential generation as they can be subjected to high stress and can be repetitively used for energy generation.

3.4.2 ZnO Nanorod-Based Piezotronic Energy Generators

ZnO nanorods are semiconducting in nature and act as diodes when in contact with conducting electrodes. Under mechanical deformation, the charges generated can be transferred to an external circuit. The nanorods conduct in one direction only as metal/ZnO interfaces mostly result in schottky contacts [57, 58]. Arrays of nanorods have been studied for use as energy generators as the amount of current in a

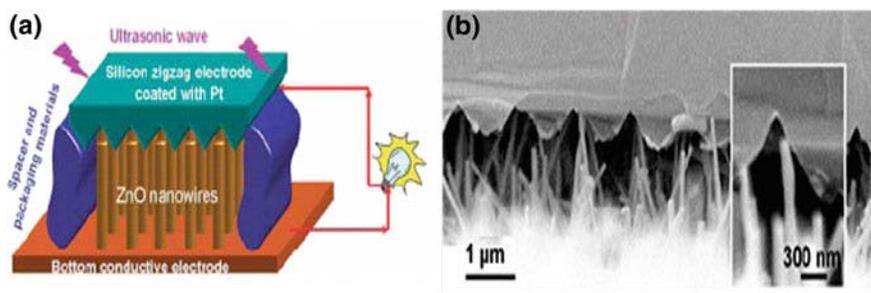


Fig. 3.10 **a** Schematic showing the structure of the nanogenerator built using ZnO nanorods using a platinum coated Si zig zag electrode (vi) **(b)** Scanning electron micrograph showing the aligned ZnO nanorods and the platinum coated Si zig zag electrode [59]

single nanorod is very less. The question is, how to deform each and every nanorod array together so that the output power is considerable. The major issues related to piezoelectric energy harvesting include the simultaneous generation of voltage in an array of nanorods, the efficient tapping of that energy and the output power. The piezo-potential is highest at the edges just below the top surface of the nanorod and persists as long as the nanorod is in the deformed state. In order to efficiently extract the energy from the deformed nanorods, the metal electrode needs to contact the nanorods at the maximum potential positions just below the tips. Wang et al. [59] used a zigzag silicon electrode with a metal layer can serve this purpose, a schematic arrangement of which is shown in Fig. 3.10a and a SEM micrograph in Fig. 3.10b. Wang and his team used ultrasonic vibrations of 41 kHz to deform the ZnO nanorods and they managed to detect output current in the range of a few nano amperes [59]. However, this setup displayed a steady current when energized using an external power supply. The magnitude of this steady current showed an abrupt increase the moment it was subjected to ultrasonic waves and this current was found to be constant for about one hour.

An important consideration is the energy generators frequency of vibration that generates an appreciable potential that can be effectively used to power devices. Vibrations in natural systems are quite low in frequency and in order to tap natural vibrations like heartbeats, other physical motions of the human body, wind power, etc., the device needs to be usable at low frequency of vibration of the nanorods. Qin et al. [60] demonstrated that ZnO nanorods could act as piezoelectric energy generators at frequencies as low as 10Hz. They grew ZnO nanorods on fibers and used them both as the generator as well as the electrode. Gold was deposited on the ZnO nanorods used as electrode for tapping the piezopotential. The two fibers were then entangled as shown in Fig. 3.11 and pulled so that the nanorods physically brushed with each other and piezoelectric energy was generated. The Au coated nanorods picked up the generated voltage as they are in contact with the bent nanorods. According to the authors, the Au coating rendered stiffness to the nanorods and as such they did not generate any potential. Moreover, as the Au was completely covering the

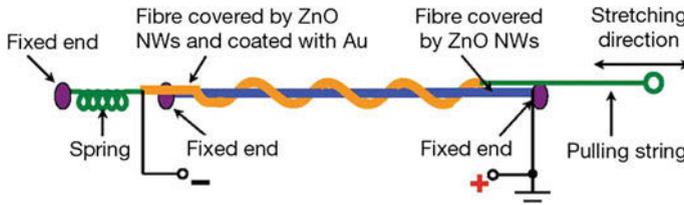


Fig. 3.11 Schematic depiction of energy generation using ZnO nanorods grown on polyamide fibers [60]

rods, the generated potential, if any, get shorted resulting in the lack of potential difference between any two points on the surface. The amount of current measured in the external circuit was 5 pA which reversed direction when the stretching was introduced in the opposite direction. It was observed that the short circuit current resulting from the open circuit voltage of about 1–3 mV could be increased using multiple fibers.

Integration of numerous nanogenerators or using big arrays of them is crucial to develop viable energy generators. The integrated system needs to be properly aligned and their charging and discharging processes synchronized. Structured arrays of vertical nanorods or systematically arranged horizontal ones could be used to produce adequate power to drive actual working devices. Xu et al. [61] reported a voltage generator involving 700 rows of nanorods. The peak voltage reported is 1.26 V, a considerable value for real applications. Vertical integration of three layers of nanorods generated a voltage of 0.243 V and an output power density of 2.7 mW/cm³. This vertically integrated system was reportedly used to power different sensors [61]. Lateral integration allowed ac voltages to be increased to noticeable levels. Seven hundred aligned nanorods placed horizontally and integrated with a schottky contact on one side generated an output voltage of 1.26 V with a maximum current of 28.8 nA. Proper control on the size and alignment of ZnO nanorods and integration with accurately designed electrodes will soon pave the way for self-powered devices reaching the hands of consumers.

There is no doubt now on the possibility of using ZnO nanorods for harvesting piezo-potential. The challenge for researchers is to design electrodes that can efficiently tap the potential generated when the nanorods are deformed. Different research groups have already reported a few designs; however, they can be improved upon in terms of efficiency of extraction. In the coming years a lot more reports of efficient nanoenergy generators will be available in the literature.

3.5 Conclusions

ZnO has inherent properties that make it attractive for use in alternative and novel energy generation techniques. Nanostructures of ZnO can be grown under mild hydrothermal conditions at temperatures below 100°C and as such are ideal for

large-scale production. Nanorods of ZnO can be grown on any type of substrates and they remain firmly affixed at different orientations. The nanorod morphology has been proved to be a good semiconducting film for dye and quantum dot sensitized solar cells as they are single crystals offering large surface to volume ratios and the electrons can freely flow through, which does not happen in nanoparticulate thin films. ZnO has better photon conversion efficiency and higher electron conductivity compared to contemporary metal oxide semiconductors like TiO₂, WO₃, etc., and ideal for solar cell applications. The piezoelectric property of ZnO nanocrystals can be utilized to generate power for driving self-powered devices and can set a new paradigm in the area of powered medical implants like pacemakers.

References

1. Emanetoglu, N.W., Gorla, C., Liu, Y., Liang, S., Lu, Y.: Epitaxial ZnO piezoelectric thin films for saw filters. *Mater. Sci. Semicond. Process.* **2**, 247–252 (1999)
2. Chen, Y., Bagnall, D., Yao, T.: ZnO as a novel photonic material for the UV region. *Mater. Sci. Eng. B: Solid-State Mater. Adv. Technol.* **75**, 190–198 (2000)
3. Liang, S., Sheng, H., Liu, Y., Huo, Z., Lu, Y., Shen, H.: ZnO Schottky ultraviolet photodetectors. *J. Cryst. Growth* **225**, 110–113 (2001)
4. Saito, N., Haneda, H., Sekiguchi, T., Ohashi, N., Sakaguchi, I., Koumoto, K.: Low-temperature fabrication of light-emitting zinc oxide micropatterns using self-assembled monolayers. *Adv. Mater.* **14**, 418–420 (2002)
5. Lee, J.Y., Choi, Y.S., Kim, J.H., Park, M.O., Im, S.: Optimizing n-ZnO/p-Si heterojunctions for photodiode applications. *Thin Solid Films* **403–404**, 553–557 (2002)
6. Mitra, P., Chatterjee, A.P., Maiti, H.S.: ZnO thin film sensor. *Mater. Lett.* **35**, 33–38 (1998)
7. Koch, M.H., Timbrell, P.Y., Lamb, R.N.: The influence of film crystallinity on the coupling efficiency of ZnO optical modulator waveguides. *Semicond. Sci. Technol.* **10**, 1523–1527 (1995)
8. Gratzel, M.: Dye-sensitized solid-state heterojunction solar cells. *MRS Bull.* **30**, 23–27 (2005)
9. Baxter, J.B., Walker, A.M., Van Ommering, K., Aydil, E.S.: Synthesis and characterization of ZnO nanowires and their integration into dye-sensitized solar cells. *Nanotechnology* **17**, (2006)
10. Lin, Y., Zhang, Z., Tang, Z., Yuan, F., Li, J.: Characterisation of ZnO-based varistors prepared from nanometre precursor powders. *Adv. Mater. Opt. Electron.* **9**, 205–209 (1999)
11. Padmavathy, N., Vijayaraghavan, R.: Enhanced bioactivity of ZnO nanoparticles—an antimicrobial study. *Sci. Technol. Adv. Mater.* **9**, (2008)
12. Baruah, S., Thanachayanont, C., Dutta, J.: Growth of ZnO nanowires on nonwoven polyethylene fibers. *Sci. Technol. Adv. Mater.* **9**, (2008)
13. Wang, Z.L.: Zinc oxide nanostructures: growth, properties and applications. *J. Phys. Condens. Matter* **16**, (2004)
14. Huang, Y., Zhang, Y., Bai, X., He, J., Liu, J., Zhang, X.: Bicrystalline zinc oxide nanocombs. *J. Nanosci. Nanotechnol.* **6**, 2566–2570 (2006)
15. Hughes, W.L., Wang, Z.L.: Controlled synthesis and manipulation of ZnO nanorings and nanobows. *Appl. Phys. Lett.* **86**, (2005)
16. Kong, X.Y., Wang, Z.L.: Spontaneous polarization-induced nanohelices, nanosprings, and nanorings of piezoelectric nanobelts. *Nano Lett.* **3**, 1625–1631 (2003)
17. Hughes, W.L., Wang, Z.L.: Formation of piezoelectric single-crystal nanorings and nanobows. *J. Am. Chem. Soc.* **126**, 6703–6709 (2004)
18. Sun, T., Qui, J., Liang, C.: Controllable fabrication and photocatalytic activity of ZnO nanobelt arrays. *J. Phys. Chem. C* **112**, 715–721 (2008)

19. Snure, M., Tiwari, A.: Synthesis, characterization, and green luminescence in ZnO nanocages. *J. Nanosci. Nanotechnol.* **7**, 481–485 (2007)
20. Gratzel, M.: Conversion of sunlight to electric power by nanocrystalline dye-sensitized solar cells. *J. Photochem. Photobiol. A: Chem.* **164**, 3–14 (2004)
21. <http://www.observatorynano.eu/project/document/2013/> (2013)
22. Bora, T.: Studies on zinc oxide nanorod dye sensitized solar cell. Master thesis AIT (2009)
23. Nazeeruddin, M.K., Kay, A., Rodicio, I., Humphry-Baker, R., Mller, E., Liska, P., Vlachopoulos, N., Grtzel, M.: Conversion of light to electricity by cis-X2bis(2,2-bipyridyl)-4,4-dicarboxylate)ruthenium(II) charge-transfer sensitizers (X = Cl-, Br-, I-, CN-, and SCN-) on nanocrystalline TiO₂ electrodes. *J. Am. Chem. Soc.* **115**, 6382–6390 (1993)
24. Nazeeruddin, M.K., Humphry-Baker, R., Liska, P., Grtzel, M.: Investigation of sensitizer adsorption and the influence of protons on current and voltage of a dye-sensitized nanocrystalline TiO₂ solar cell. *J. Phys. Chem. B* **107**, 8981–8987 (2003)
25. Aiga, F., Tada, T.: Molecular and electronic structures of black dye; an efficient sensitizing dye for nanocrystalline TiO₂ solar cells. *J. Mol. Struct.* **658**, 25–32 (2003)
26. Hagberg, D.P., Edvinsson, T., Marinado, T., Boschloo, G., Hagfeldt, A., Sun, L.: A novel organic chromophore for dye-sensitized nanostructured solar cells. In: *Chemical Communications*, pp. 2245–2247 (2006)
27. Ito, S., Zakeeruddin, S.M., Humphry-Baker, R., Liska, P., Charvet, R., Comte, P., Nazeeruddin, M.K., Pchy, P., Takata, M., Miura, H., Uchida, S., Grtzel, M.: High-efficiency organic-dye-sensitized solar cells controlled by nanocrystalline-TiO₂ electrode thickness. *Adv. Mater.* **18**, 1202–1205 (2006)
28. Suri, P., Mehra, R.M.: Effect of electrolytes on the photovoltaic performance of a hybrid dye sensitized ZnO solar cell. *Sol. Ener. Mater. Sol. Cells* **91**, 518–524 (2007)
29. Law, M., Greene, L.E., Johnson, J.C., Saykally, R., Yang, P.: Nanowire dye-sensitized solar cells. *Nat. Mater.* **4**, 455–459 (2005)
30. Yang, P., Yan, H., Mao, S., Russo, R., Johnson, J., Saykally, R., Morris, N., Pham, J., He, R., Choi, H.J.: Controlled growth of ZnO nanowires and their optical properties. *Adv. Funct. Mater.* **12**, 323–331 (2002)
31. Xia, J.B., Zhang, X.W.: Electronic structure of ZnO Wurtzite quantum wires. *Eur. Phys. J. B* **49**, 415–420 (2006)
32. Quintana, M., Edvinsson, T., Hagfeldt, A., Boschloo, G.: Comparison of dye-sensitized ZnO and TiO₂ solar cells: studies of charge transport and carrier lifetime. *J. Phys. Chem. C* **111**, 1035–1041 (2007)
33. Akhtar, M.S., Khan, M.A., Jeon, M.S., Yang, O.B.: Controlled synthesis of various ZnO nanostructured materials by capping agents-assisted hydrothermal method for dye-sensitized solar cells. *Electrochim. Acta* **53**, 7869–7874 (2008)
34. Pradhan, B., Batabyal, S.K., Pal, A.J.: Vertically aligned ZnO nanowire arrays in Rose Bengal-based dye-sensitized solar cells. *Sol. Energy Mater. Sol. Cells* **91**, 769–773 (2007)
35. Chen, J., Li, C., Song, J.L., Sun, X.W., Lei, W., Deng, W.Q.: Bilayer ZnO nanostructure fabricated by chemical bath and its application in quantum dot sensitized solar cell. *Appl. Surf. Sci.* **255**, 7508–7511 (2009)
36. Baxter, J.B., Aydil, E.S.: Dye-sensitized solar cells based on semiconductor morphologies with ZnO nanowires. *Sol. Energy Mater. Sol. Cells* **90**, 607–622 (2006)
37. Gerischer, H., Tributsch, H.: *Ber. Bunsen-Ges. Phys. Chem.* **73**, 850–854 (1969)
38. Matsumura, M., Matsudaira, S., Tsubomura, H., Takata, M., Yanagida, H.: Dye sensitization and surface structures of semiconductor electrodes. *Ind. Eng. Chem. Prod. Res. Dev.* **19**, 415–421 (1980)
39. Boschloo, G., Edvinsson, T., Hagfeldt, A.: Dye-sensitized nanostructured ZnO Electrodes for solar cell applications. In: *Nanostructured Materials for Solar Energy Conversion*, pp. 227–254 (2007)
40. Ku, C.H., Wu, J.J.: Chemical bath deposition of ZnO nanowire-nanoparticle composite electrodes for use in dye-sensitized solar cells. *Nanotechnology* **18** (2007)

41. Keis, K., Bauer, C., Boschloo, G., Hagfeldt, A., Westermark, K., Rensmo, H., Siegbahn, H.: Nanostructured ZnO electrodes for dye-sensitized solar cell applications. *J. Photochem. Photobiol. A: Chem.* **148**, 57–64 (2002)
42. Baruah, S., Dutta, J.: Hydrothermal growth of ZnO nanostructures. *Sci. Technol. Adv. Mater.* **10**, 013001 (2009)
43. Makhal, A., Sarkar, S., Bora, T., Baruah, S., Dutta, J., Raychaudhuri, A.K., Pal, S.K.: Role of resonance energy transfer in light harvesting of zinc oxide-based dye-sensitized solar cells. *J. Phys. Chem. C* **114**, 10390–10395
44. Desai, A.V., Haque, M.A.: Mechanical properties of ZnO nanowires. *Sens. Actuators* **134**, 169–174 (2007)
45. Kou L.Z., Guo, W.L., Li, C.: In: *Proceedings of IEEE Symposium on Piezoelectricity, Acoustic Waves, and Device Applications*, pp. 354–359 (2008)
46. Song, J.H., Zhou, J., Liu, J., Wang, Z.L.: *Nano Lett.* **6**, 1652–1656 (2006)
47. Wang, Z.L., Song, J.H.: *Science* **312**, 242–246 (2006)
48. Hutson, A.R., White, D.L.: Elastic wave propagation in piezoelectric semiconductors. *J. Appl. Phys.* **33**, 40–47 (1962)
49. Corso, A.D., Pastemak, M., Resta, R., Balderschi, A.: *Phys. Rev. B* **50**, 10715–10721 (1994)
50. Fan, Z., Lu, J.G.: *J. Nanosci. Nanotechnol.* **5**, 1561–1573 (2005)
51. Zhao, M.H., Wang, Z.L., Mao, S.X.: *Nano Lett.* **4**, 587–590 (2004)
52. Mantini, G., Gao, Y., D'Amico, A., Falconi, C., Wang, Z.L.: *Nano Res.* **2**, 624–629 (2009)
53. Mantini, G., Gao, Y., Wang, D.Z.L., Falconi, A.C.: *Procedia Chem.* **1**, 1403–1406 (2009)
54. Wang, Z.L.: *Adv. Mater.* **21**, 1311–1315 (2009)
55. Gao, Y., Wang, Z.L.: *Nano Lett.* **9**, 1103–1110 (2009)
56. Xu, F., Qin, Q., Mishra, A., Gu, Y., Zhu, Y.N.R.: *Nano Res.* **3**, 271–280 (2010)
57. Park, W.I., Yi, G.C., Kim, J.W., Park, S.M.: *Appl. Phys. Lett.* **82**, 4358–4360 (2003)
58. Park, W.I., Yi, G.C.: Ohmic and Schottky nanocontacts. In: *IEEE Proceedings*, pp. 410–414 (2003)
59. Wang, X., Song, J., Liu, J., Wang, Z.L.S.: *Science* **316**, 102–105 (2007)
60. Qin, Y., Wang, X., Wang, Z.L.: *Nature* **451**, 809–813 (2008)
61. Xu, Q.Y.S., Xu, C., Wei, Y., Yang, R., Wang, Z.L.N.N.: *Nat. Nanotechnol.* **5**, 366–373 (2010)

Part II
Accepted Papers

Chapter 4

Wavelet and Learning Based Image Compression Systems

Mayuri Kalita and Kandarpa Kumar Sarma

Abstract Image compression is a critical element in storage, retrieval and transmission applications. The list of traditional approaches to image compression has already been expanded by wavelet and learning based systems. Here, we report a few techniques which are based on discrete wavelet transform (DWT), Artificial Neural Network (ANN) in feedforward and unsupervised form. The experiments are repeated with images mixed with salt and pepper noise and the outcomes are compared. The quality of the image compression systems is determined by finding the mean square error (MSE), Peak Signal to Noise Ratio (PSNR) and Compression Ratio (CR).

Keywords ANN · MSE · CR · PSNR · SOM · DWT · DCT

4.1 Introduction

The use of data has increased significantly with the expansion of communication and multimedia applications. Uncompressed multimedia data require considerable storage capacity and transmission bandwidth [1]. Therefore, image compression is an essential requirement and has always been a growing area of research. The goal of image compression is to reduce memory requirements and bandwidth to transmit the image over a communication channel [2]. Image compression technique deals with reduction of redundancies present in an image. The situation is more critical with medical imaging. While for diagnostic applications, imagery volume is always increasing, the compression methods should never compromise the quality after decompression. Hence, there is a constant requirement to explore new methods of image compression which enhances quality. Learning based attributes in compression systems can provide better performance in this respect. Learning based

M. Kalita (✉) · K.K. Sarma
Gauhati University, Guwahati 781014, India
e-mail: mayuri29kalita@gmail.com

K.K. Sarma
e-mail: kandarpaks@gmail.com

tools like Artificial Neural Networks (ANN) have already received attention in this respect. Hence, the design of an efficient medical image compression system is a challenging task. ANNs have been applied to various image compression problems because of their ability to learn, retain the learning and use it subsequently which is not observed in traditional methods. They have the ability to preprocess input patterns to produce reduced data sizes with fewer components [3]. In such cases, ANN based systems don't require preprocessing blocks and reduce system complexity. Further, as ANNs are robust to noise variations, such systems show high level of resilience against sudden variations mostly essential for medical images. In order to perform image compression using ANNs, which is based on different learning algorithms, some topologies that have already been used are Multi-Layer Perceptron (MLP), Self-Organizing Maps (SOM), Learning Vector Quantization (LVQ) and Principal Component Analysis (PCA). The back-propagation learning algorithm used by MLP is a widely used method suitable for image processing applications. This feed-forward (FF) architecture of an ANN is capable of approximating most problems with high accuracy and generalization [4]. This ability of the FF ANN enables the user to use it for compressing images. Self-Organizing Feature Map (SOFM) is a kind of unsupervised ANN which consists of components called nodes (neurons) and is based on competitive learning. It gives a low dimensional and discretized representation of the input space of the training samples which is known as map [5]. The SOFM or SOM network is inherently a data clustering technique which can also be used for scaling data. The key point behind using SOM in image compression is its ability to make a proper approximation for a large set of data by detecting a smaller set of weights in the network. A large number of image compression techniques have been designed using conventional and ANN based methods. Some of the methods are spatial, time and frequency domain. But very few works are reported which have explored the possibility of deriving performance improvement by using the above techniques with learning based systems. Here, we present a comparative analysis between a learning aided image compression technique using MLP and another technique dependent on SOM and Discrete Wavelet Transform (DWT). The work describes a comparison between two lossy image compression schemes. The image to be compressed is decomposed into smaller sized blocks and then applied to either an MLP or a SOM network followed by DWT for compression. The compressed image is reconstructed using an MLP in the first approach. In the second method, the image after compression is recovered by a composite block formed using a FF ANN and a Discrete Cosine Transform (DCT) based compression-decompression system. The original image is degraded to some extent by adding salt and pepper noise to it in order to check whether the image compression systems are robust to degradations like noise. The performance of the systems is evaluated by determining the Mean Square Error (MSE), Compression Ratio (CR) and Peak Signal-to-Noise Ratio (PSNR). The remaining part of the paper is divided into the following sections. Sections 4.2 and 4.3 describes literature review and background theory. The presented system models and comparison analysis is explained in Sects. 4.3 and 4.4. Conclusion part of the proposed work is discussed in Sect. 4.5.

4.2 Literature Review

1. In [3], the authors proposed a method on improved image compression approach with SOFM Network using Cumulative Distribution Function. This work provided improved CR and network convergence.
2. In [4], digital image compression using neural networks has been reported. The ANN algorithm used here is mainly the back-propagation of multilayer perceptrons. The algorithm preserves most of the characteristics of the image and maintains the compression performance.
3. In the work [5], the authors proposed a work on image compression technique based on DWT and Self Organizing Map. In this work, a better CR and PSNR is attained compared to existing techniques.

4.3 Certain Theoretical Concepts

Here, we discuss the related theoretical concepts.

4.3.1 ANN in FF Mode

An ANN is a highly interconnected network of a large number of processing elements called neurons in an architecture inspired by the brain [6, 7]. The network consists of three layers, the input layer, the hidden layer and the output layer which are interconnected to each other layers as shown in Fig. 4.1. The network can be trained with input patterns to achieve the desired output.

4.3.2 SOM

SOM network is a kind of unsupervised ANN based on competitive learning algorithm. Here, the output neurons compete with each other to get activated. The architecture of SOM network is shown in Fig. 4.2. The SOFM algorithm is based on two basic expressions shown in (4.1) and (4.2). For a network with n number of neurons and at time t , if x is the input then

$$\|x(t) - w_c(t)\| = \min \|x(t) - w_i(t)\| \quad (4.1)$$

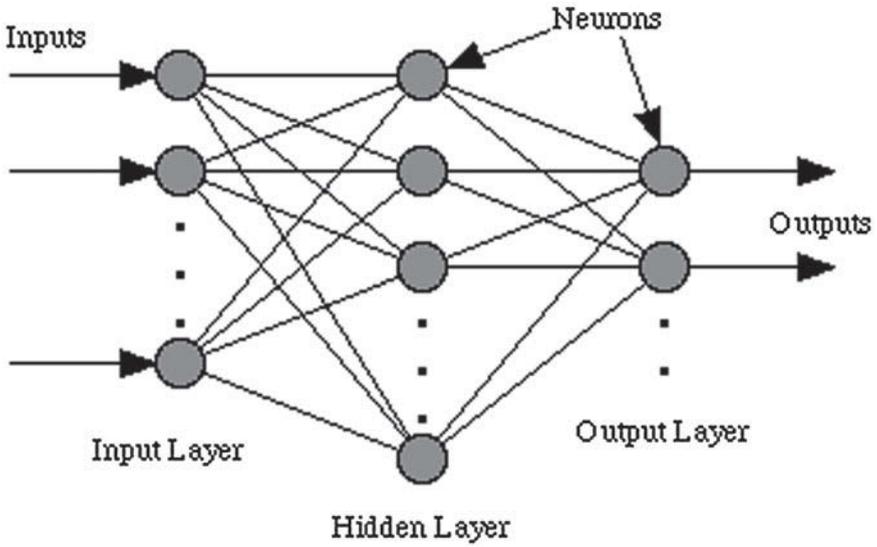


Fig. 4.1 Structure of an ANN

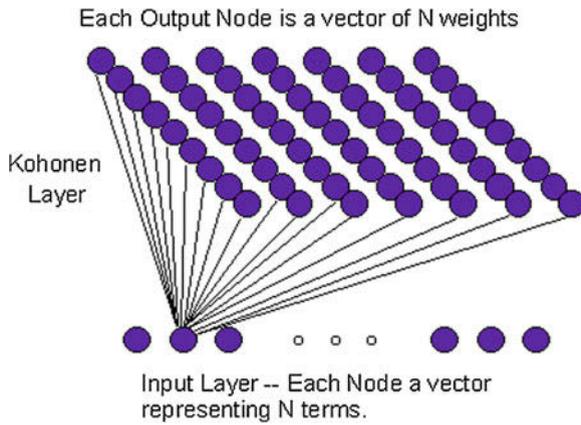


Fig. 4.2 Architecture of SOFM network

$$w_i(t + 1) = w_i(t) + \alpha(t)[x_i(t) - w_i(t)] \tag{4.2}$$

where, $1 < i < n$, w_i is the node, w_c is the winner and α is the learning rate, $0 < \alpha < 1$.

4.3.3 DWT

DWT decomposes the input data into a set of wavelets, orthogonal to its translations and scaling. When we apply single level DWT to the input data, four sub-bands, LL, LH, HL and HH are formed as shown in Fig. 4.3. Only the LL sub-band containing useful information regarding the image quality is considered and processed for reconstruction. In the proposed work, single level decomposition using Haar DWT is performed.



Fig. 4.3 Decomposition using DWT

4.3.4 DCT

DCT is one of the most popular compression scheme for still image compression. The property of energy compaction found in DCT and availability of fast algorithms applicable for real operations makes it suitable for use in image compression. DCT helps in separating an image into spectral sub-bands of differing importance with respect to image’s visual quality [8]. For an image block of size 4×4 , the ordering in frequency is shown in Fig. 4.4. The lower frequencies are denoted by smaller numbers.

0	1	5	6
2	4	7	12
3	8	11	13
9	10	14	15

Fig. 4.4 Zigzag scan of DCT coefficients

4.4 Image Compression Models

The proposed learning based image compression systems have certain constituent block which are discussed in the following sections.

4.4.1 Image Compression System Using FF ANN

The ANN used for compression in the proposed method is of the type MLP trained with back-propagation algorithm. The parameters of the ANN is shown in Table 4.1. A sample image of size 256×256 is considered initially. The system model is shown in Fig. 4.5. The image is sent to two ANNs for compression and decompression. The quality of the reconstructed image is found to be poor when the whole image is given as input to ANN. In order to improve the quality of the reconstructed image, multi-level decomposition of the image is done. Decomposition of an image into sub-blocks reduces computational complexity of the system [9]. We describe a 4-level and 9-level decomposition done to perform image compression using ANN. Different patterns are presented to the ANN which it learns during training. The learning is preserved and used subsequently. In 4-level decomposition, the image to

Table 4.1 ANN parameters

ANN type	Training algorithm	Training function	Epochs	MSE goal
<i>Feed-forward MLP with 2 hidden layers</i>	Backpropagation <i>Levenberg–Marquardt</i> Optimization	<i>Trainlm</i>	1–100	10^{-6}

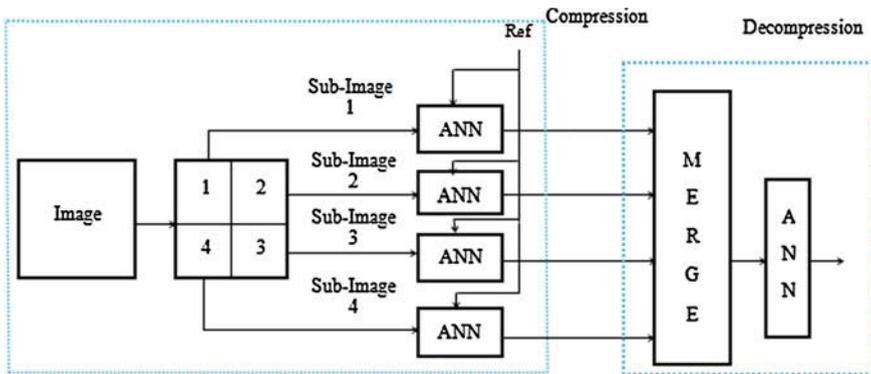


Fig. 4.5 Block diagram of compression system using ANN

Table 4.2 Simulation parameters

Item	Description
<i>Data dimension</i>	128 × 128
	256 × 256
	512 × 512
<i>Data size</i>	10 samples each of the above dimensions both in noise free and noise mixed form
<i>Tools</i>	ANN in FF form, SOM, DWT and DCT

be compressed is partitioned into four equal sized blocks. Each block is then sent to an ANN for compression, by selecting a suitable reference. The compressed sub-images are combined to achieve a CR of 128:1 for an original image of size 256 × 256. Although, the compression rate achieved in this compression system is high, yet the image quality after reconstruction is not up to the level of an acceptable limit. Hence, higher level decomposition is done and it is seen that 9-level decomposition gives better results. The simulation parameters are given in Table 4.2.

4.4.2 Image Compression Based on SOM and DWT and Reconstruction Using FF ANN

The decomposed parts of the original image of size $m \times n$ are sent to a SOM network individually for generating the weight matrices, also known as the codebook. The size of the codebook generated by SOM is kept fixed at $64 \times n/2$. There will be four weight matrices or codebooks with respect to each sub-image. The weight matrix is decomposed into four sub-bands by applying single-level DWT to it. The four components formed after single-level decomposition are the approximation coefficient (LL), the horizontal coefficient (LH), the vertical coefficient (HL) and the diagonal coefficient (HH). Among the four components, only the LL part contains useful information required for preserving the image quality. The other coefficients (LH, HL and HH) contain very less information details and if they are discarded, the quality of the reconstructed image is not hampered. The LL part is retained to get the compressed sub-image. For each sub-image, the above compression procedure is repeated which gives a compressed image of size $64 \times n/2$ with a CR of 8:1. The block diagram of the just described compression system is shown in Fig. 4.6. The decompression process uses an ANN in multi-layer feed-forward form containing a single hidden layer. The compressed sub-image is fed as input to the ANN along with a reference. The reconstructed matrix generated on application of DCT based compression and decompression to the original image sub-blocks is taken as the reference. The decompression scheme for the regeneration process is shown in Fig. 4.7. The ANN parameters are shown in Table 4.3. The size of layers will vary with respect to four and nine level decompositions.

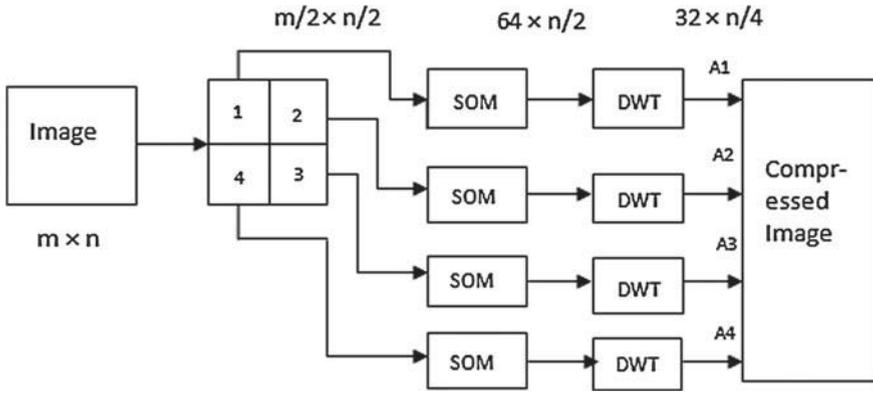


Fig. 4.6 Block diagram of compression system using SOM and DWT

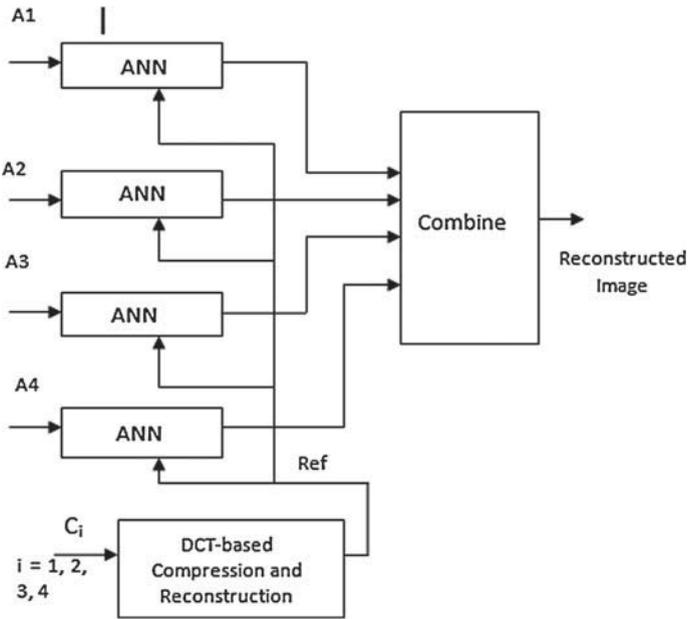


Fig. 4.7 Block diagram of decompression scheme

Table 4.3 ANN Parameters

Network topology	SOM grid, one layer	MLP, one input, one hidden, one combined output
Size		$32 \times n/4$
	$64 \times n/2$	$1.5 \times$ input layer size
		$64 \times n/2$
Training type	<i>Competitive</i> learning	Back-propagation with <i>Levenberg–Marquardt</i> optimization
Epochs	100–500	50–200
MSE goal	–	10^{-3}

4.5 Comparative Analysis

The performance of the two image compression systems are evaluated by finding parameters like MSE, CR and PSNR. The computational time required for the two systems are also calculated. Table 4.4 shows the experimental results for three different sample images of MRI, CT scan and X-ray. In the table, system 1 refers to image compression technique using FF ANN whereas system 2 represents image compression system using SOM and DWT. The two systems are analyzed based on these results. From Table 4.4, it is seen that the proposed image compression schemes differ from each other in terms of CR, PSNR and computational time. System 2 undergoes image compression without requirement of a reference unlike system 1. The image quality after reconstruction is maintained in both the cases which is judged by finding out the MSE. The system is experimented for various medical images and the performance is found out to be satisfactory. The experimental results for a sample image of X-ray is shown in Fig. 4.8. The recovery of the original image for the two proposed systems is satisfactory which can be viewed from Fig. 4.8. This enables the users to use the system for storage and transmission purposes with medical images. The PSNR values that are evaluated experimentally, shows the robustness of the systems. The compressed images can be transmitted through communication channels irrespective of degradations like noise that may be present in an image due to false switching of a device and transmission media. Thus with the FF form described in Sect. 4.1, the CR is found to be high but computational requirement are found to be more and decompression quality is lower making it unsuitable for medical applications. Next, the SOM-DWT-FF ANN combination provides lower compression but has superior decompression quality, shows less computational requirements and is also robust against noise like variations. Thus, this approach is suitable for medical applications even when images are transmitted using certain communication systems. The method when compared with Run-length Encoding (RLE) also provides better quality but is computationally demanding. Tables 4.5 and 4.6 shows the comparison of MSE and PSNR values of the proposed systems with existing RLE.

Table 4.4 Performance evaluation

Image	Original size	CR for system 1	CR for system 2	MSE, system 1	MSE, system 2	PSNR, system 1	PSNR, system 2	Computational time, system 1 (s)	Computational time, system 2 (s)
MRI	240 × 240	80:1	15:2	0.015	0.007	65.52	69.41	3465.15	153.00
CT scan	240 × 240	80:1	15:2	0.018	0.010	64.97	67.30	2877.50	138.20
X-ray	240 × 240	80:1	15:2	0.040	0.007	65.57	69.60	3251.44	133.37

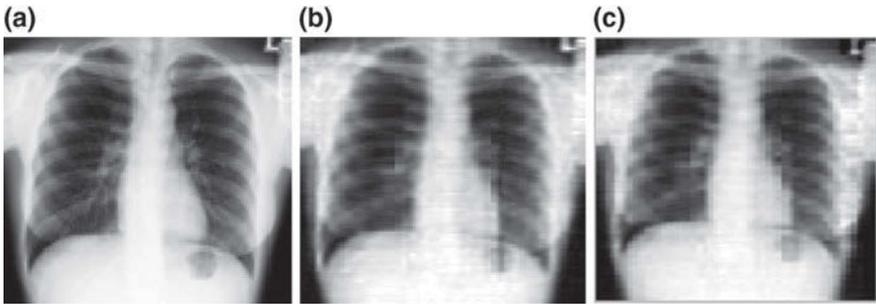


Fig. 4.8 Original and reconstructed images for the proposed systems. **a** Original image. **b** Reconstruction (system 1). **c** Reconstruction (system 2)

Table 4.5 MSE comparison of proposed systems with existing RLE

Image	MSE, RLE	MSE, system 1	MSE, system 2
MRI	0.36	0.015	0.007
CT Scan	0.34	0.018	0.010
X-ray	0.32	0.040	0.007

Table 4.6 PSNR comparison of proposed systems with existing RLE

Image	PSNR, RLE (dB)	PSNR, system 1 (dB)	PSNR, system 2 (dB)
MRI	52.56	65.52	69.41
CT Scan	52.79	64.97	67.30
X-ray	52.98	65.57	69.60

4.6 Conclusion

Here, we discussed certain approaches for image compression with the objective of making it suitable for medical applications. We proposed a FF ANN and a SOM-DWT-FF ANN based approaches out of which the later is found to be more suitable.

References

1. Ettaouil, M., Ghanou, Y., Moutaouakil, K., Lazaar, M.: Image medical compression by a new architecture optimization model for the Kohonen networks. *Int. J. Comput. Theory Eng.* **3**(2), 1793–8201 (2011)
2. Salmoon, D.: *Data Compression*. Springer, New York (1998)
3. Durai, S.A., Saro, E.A.: An improved image compression approach with SOFM network using cumulative distribution function. In: *International Conference on Advanced Computing and Communications*, pp. 304–307. Surathkal (2006)

4. Dutta, D.P., Choudhury, S.D., Hussain, M.A., Majumder, S.: Digital image compression using neural network. In: IEEE International Conference on Advances in Computing, Control, and Telecommunication Technologies (2009)
5. Sharma, M., Kuamri, R.: A technique of image compression based on discrete wavelet image decomposition and self organizing map. *Int. J. Comput. Eng. Sci.* **3**(8) (2013)
6. Krishnamurthy, A.K., Ahalt, S.C., Melton, D.E., Chen, P.: Neural networks for vector quantisation of speech and images. *IEEE J. Sel. Areas Commun.* **8**(8), 1449–1457 (1990)
7. Sicuranzi, G.L., Ramponi, G., Marsi, S.: Artificial neural network for image compression. *Electron. Lett.* **26**(7), 477–479 (1990)
8. Thota, N.R., Devireddy, S.K.: Image compression using discrete cosine transform. *Georgian Electron. Sci. J. Comput. Sci. Telecommun.* **3**(17) (2008)
9. Sreekumar, N., Baboo, S.S.: Neural network based complex image compression using modified Levenberg-Marquardt method for learning. *Int. J. Comput. Sci. Technol.* **2**(2) (2011)

Chapter 5

Quantifying Clinical Information in MEEG Using Sample and Channel Convolution Matrices

R.K. Tripathy and S. Dandapat

Abstract In this paper, a novel distortion measure is presented for quantifying loss of clinical information in multichannel electrocardiogram (MECG) signals. The proposed measure (SCPRD) is defined as the sum of percentage root mean square difference between magnitudes of convolution response of original and processed MEEG signals. The convolution operation is performed with the help of proposed sample and channel convolution matrices. The SCPRD measure is compared with average wavelet energy diagnostic distortion (AWEDD) and multichannel PRD (MPRD) measures over different processing schemes such as multiscale principal component analysis (MSPCA) and multichannel empirical mode decomposition (MEMD)-based MEEG compression and filtering. The normal and pathological MEEG signals from the Physikalisch Technische Bundesanstalt (PTB) database is used in this work. The result shows that the proposed diagnostic distortion measure is effective to quantify the loss of clinical information in MEEG signals.

Keywords Diagnostic measure · Multichannel ECG · SCPRD · MEMD · MSPCA · PRD · WEDD

5.1 Introduction

Multilead ECG views the heart from different angles and it is widely used in clinical practice for diagnosis of cardiovascular diseases [1, 2]. The clinical information present in MEEG are the waveform amplitudes and the duration. There are a number of MEEG data compression and enhancement techniques are reported in the literature [3, 4]. The quality of processed MEEG signals is essential for clinical applications. Subjective and objective distortion measures are used to assess the quality of ECG signals [5–8]. Subjective measures are evaluated by medical experts.

R.K. Tripathy (✉) · S. Dandapat
Indian Institute of Technology, Guwahati 781039, India
e-mail: r.tripathy@iitg.ernet.in

S. Dandapat
e-mail: samaren@iitg.ernet.in

The subjective tests are expensive and time consuming. To overcome the limitations, various objective measures are used [9]. The wavelet-based distortion measures and percentage root mean square difference (PRD) are popular objective measures for ECG signal [3]. The PRD, WEDD etc. are local measures. They quantify the loss of diagnostic information along each lead in MECG. In this paper, a new global objective diagnostic measure based on sample and channel convolution matrix for MECG signals is proposed. Section 5.2 describes the proposed SCPRD diagnostic measure. In Sect. 5.3, the results and discussion are presented. Conclusion are drawn in Sect. 5.4.

5.2 Convolution Matrix-Based Distortion Measure

The proposed sample and channel convolution matrix-based objective diagnostic distortion measure for MECG signals is evaluated in four stages.

(i) The convolution of MECG ($\mathbf{X} \in \mathbf{R}^{n \times m}$) with both the proposed sample and the channel convolution matrices is defined as:

$$\mathbf{SC}_j = \mathbf{X} \otimes \mathbf{SM} \quad (5.1)$$

$$\mathbf{CC}_j = \mathbf{X} \otimes \mathbf{CM} \quad (5.2)$$

where \mathbf{SM} and \mathbf{CM} are sample and channel convolution matrices defined as

$$\mathbf{SM} = \begin{bmatrix} 1 & 1 & 1 \\ -1 & -1 & -1 \\ 1 & 1 & 1 \end{bmatrix}$$

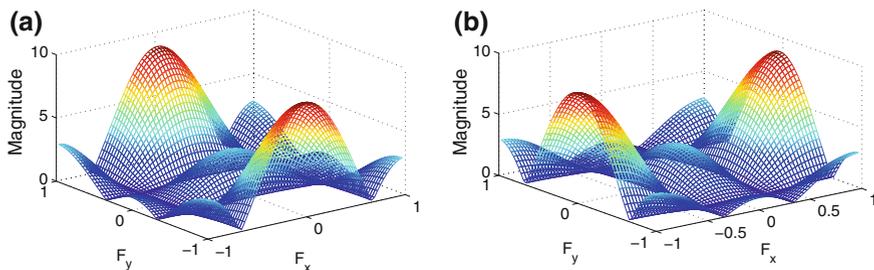


Fig. 5.1 Shows the frequency response of both the sample and channel convolution matrix. **a** Frequency response of SM. **b** Frequency response of CM

and $\mathbf{CM} = \mathbf{SM}^T$. The \mathbf{SM} and \mathbf{CM} metrics which figure out the intrabeat and interlead differences in MECG signals. Figure 5.1 shows the frequency response of both the sample convolution matrix (\mathbf{SM}) and channel convolution matrix (\mathbf{CM}). It is observed that \mathbf{SM} acts as a filter bank to capture the diagnostic features for MECG along each lead. The normalized frequency along sample direction (F_X) varies from -1 to 1 . The diagnostic features for ECG signals is present between 0.5 to 70 Hz. These features are well captured through \mathbf{SM} by observing the spectral peaks as shown in Fig. 5.1a. The \mathbf{CM} matrix is used to measure the interlead difference.

The ' \mathbf{n} ' and ' $\mathbf{m} = 12$ ' are the number of samples and number of leads, respectively. \mathbf{SC}_j and \mathbf{CC}_j correspond to the convolution output and $\mathbf{j} \in (1, 2, \dots, \mathbf{g})$. The value of ' \mathbf{g} ' is the number of columns present in both \mathbf{SC} and \mathbf{CC} matrices after convolution operation. As the proposed sample and channel convolution matrices (\mathbf{SM} and \mathbf{CM}) are of 3×3 size, after convolution MECG signals the value of ' \mathbf{g} ' which is found to be 4 .

(ii) The magnitude of convolution response is defined as

$$A_j = \sqrt{(\mathbf{SC}_j)^2 + (\mathbf{CC}_j)^2} \quad (5.3)$$

(iii) The PRD between magnitudes of convolution response of original and processed MECG signals is defined as

$$\mathbf{CMPRD}_j = \sqrt{\frac{\sum_{k=1}^{\mathbf{K}} (A_j(\mathbf{k}) - \tilde{A}_j(\mathbf{k}))^2}{\sum_{k=1}^{\mathbf{K}} (A_j(\mathbf{k}))^2}} \times 100 \quad (5.4)$$

where the value of \mathbf{K} corresponds to the number of samples (rows) present in both \mathbf{SC} and \mathbf{CC} matrices after convolution operation.

(iv) The proposed SCPRD measure is defined by the sum of \mathbf{CMPRD}_j values as

$$\mathbf{SCPRD} = \sum_{j=1}^{\mathbf{g}} \mathbf{CMPRD}_j \quad (5.5)$$

For testing of proposed SCPRD measure, the multivariate empirical mode decomposition (MEMD) [10, 11] and multiscale PCA (MSPCA) [3]-based filtering and compression for MECG signals are used. The MEMD segregates the clinical information embedded in MECG signals into different scales. These scales are called as intrinsic mode functions (IMF). There are various noises such as baseline wandering, powerline interference, Gaussian noise are present in MECG signals during recording. Filtering of these noises from MECG is done by reconstruction of signals using different IMF subsets. After filtering and compression/reconstruction, the proposed SCPRD, AWEDD, and MPRD measures are evaluated by considering both the original and processed MECG signals. The WEDD and PRD are computed along each lead of MECG signals and the average of these values are given as AWEDD and MPRD measures.

5.3 Results and Discussion

The MEECG signals used in this work are taken from the PTB diagnostic database [12]. Both normal and pathological MEECG are used over MEECG filtering and compression platform. The sampling frequency for each of the MEECG signals is 1000 Hz. The SCPRD, AWEDD, and MPRD measures are evaluated and compared over aforementioned MEECG processing frameworks.

5.3.1 MEMD-Based Filtering and Distortion Measure

MEMD decomposes the MEECG data matrix into 13 IMFs along each leads. The IMFs for lead I ECG signal are shown in Fig. 5.3. It is noted that the high-frequency components are present in both the first and second IMF, while the low-frequency components are present in the last three IMFs. For baseline wandering removal, the last three IMFs (IMF11, IMF12, and IMF13) and for high frequency noise (HFN)

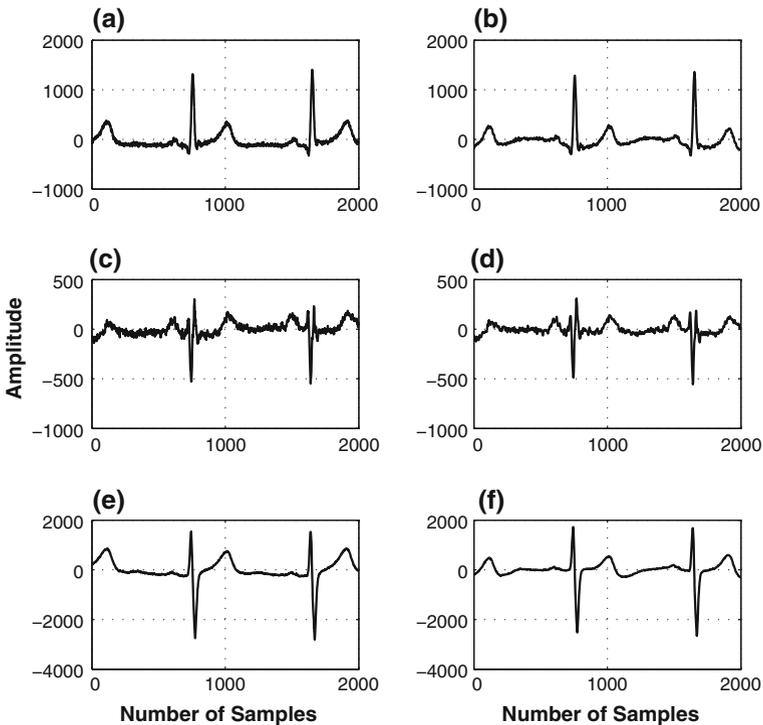


Fig. 5.2 a, c, e Shows the original ECG signal at lead I, II and V2. b, d and f Shows the proceed ECG signal after baseline wandering noise and high frequency noise filtering

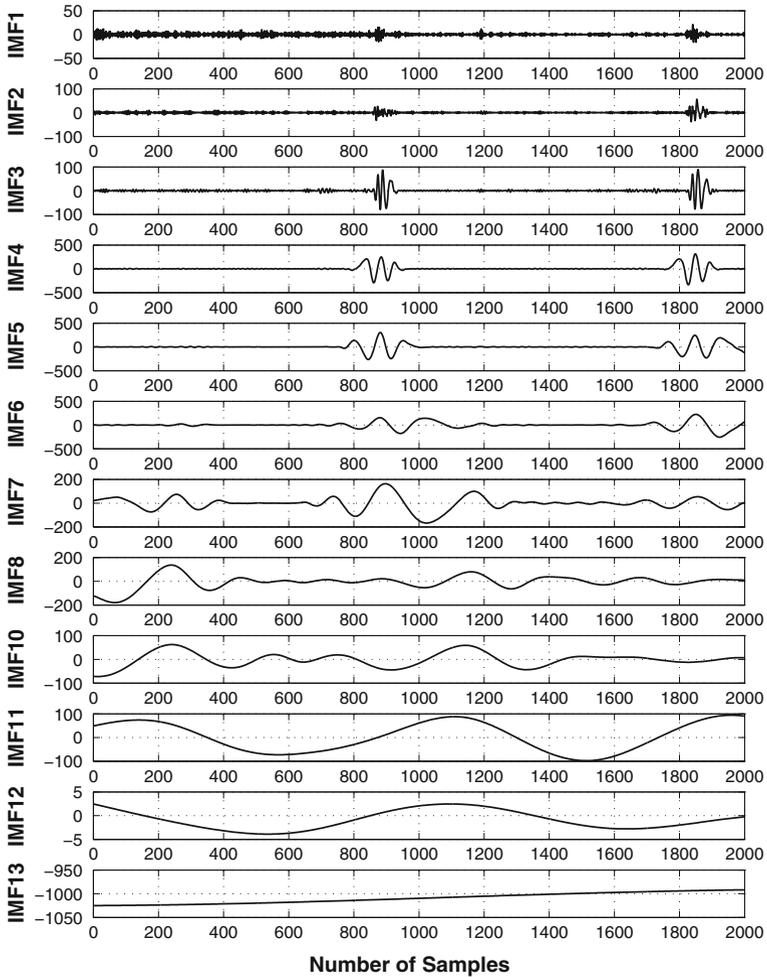


Fig. 5.3 Shows the IMFs of lead I ECG signal

removal, the first two IMFs (IMF1 and IMF2) from each lead signals are eliminated. Then, the MECG signals is reconstructed from the remaining IMFs along each lead. The original and processed ECG signals for lead I, III, and V2 after removal of both baseline wandering (BW) and high-frequency noise (HFN) are shown in Fig. 5.2. The proposed SCPRD along with AWEDD and MPRD objective diagnostic distortion measures are evaluated and shown in Table 5.1. It is observed that the SCPRD, AWEDD, and MPRD values for HC1 MECG data are found to be 37.73, 31.14, and 28.79%. Similarly for HC4, MECG data, the MPRD and SCPRD values are 22.21 and 19.73%.

Table 5.1 Comparisons of SCPRD with AWEDD and MPRD over BW noise and HFN filtering from MECG

MECG filtering	Record	MPRD (%)	AWEDD (%)	SCPRD (%)
BW and HFN removal	HC1	37.73	31.14	28.79
	HC2	53.50	47.74	37.03
	HC3	36.49	35.39	28.47
	HC4	22.21	18.17	19.73
	HC5	39.68	39.49	31.90

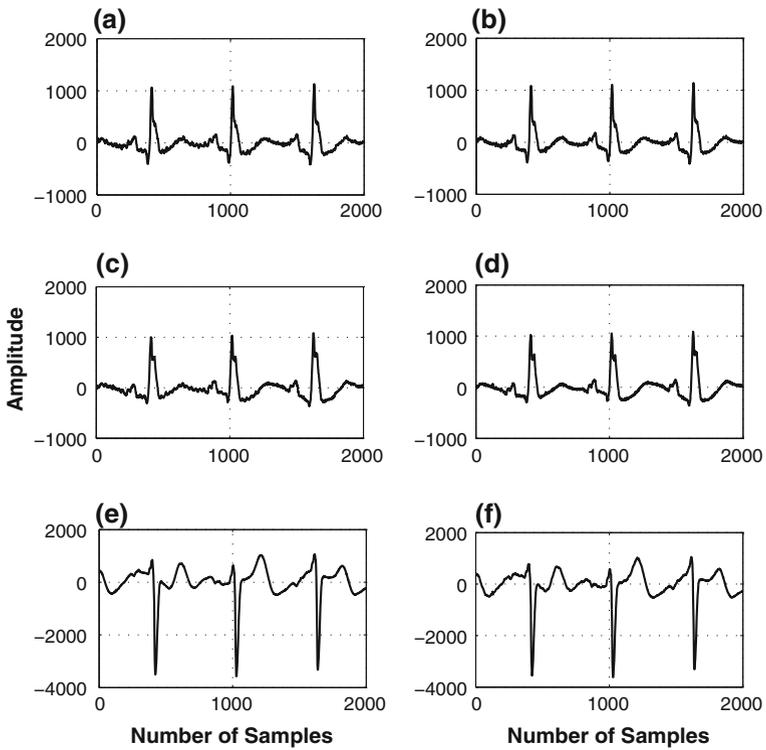


Fig. 5.4 a, c, e Shows the original ECG signal at lead I, II and V2. b, d and f Shows the processed ECG signal after MECG compression and reconstruction

Table 5.2 Comparisons of SCPRD with AWEDD and MPRD over MSPCA-based MEGC Compression

MEGC compression	Record	CR	MPRD (%)	AWEDD (%)	SCPRD (%)
MSPCA	HC1	8.87	10.47	10.59	9.53
	HC2	11.73	20.57	19.90	17.40
	HC3	7.81	8.91	11.58	7.05
	HC4	10.54	16.39	21.35	15.10
	HC5	11.28	16.43	18.15	13.53
MSPCA	MI1	9.03	10.12	11.34	8.13
	MI2	5.12	4.75	5.16	4.71
	MI3	8.75	10.82	11.11	7.43
	MI4	9.27	12.98	11.66	8.54
	MI5	10.49	13.60	13.49	8.98

5.3.2 MSPCA-Based MEGC Compression and Distortion Measure

In this subsection, the proposed SCPRD, and existing AWEDD and MPRD measures are tested over MSPCA-based MEGC compression and reconstruction scheme. The original and compressed signal for lead I, II, and V2 are depicted in Fig. 5.4. The SCPRD, AWEDD, and MPRD values for normal and MI pathology-based MEGC signals for different compression ratio (CR) values are shown in Table 5.2. It is noted that for CR value of 8.87:1 the values of SCPRD, AWEDD, and MPRD are found to be 9.53, 10.59, and 10.47%. Similarly, with increase in the values of CR as 11.28:1, the corresponding SCPRD, AWEDD, and MPRD values 16.43, 18.15, and 13.53% also increase.

The clinical acceptable range of MPRD, AWEDD for compressed ECG signal is given as [5, 9]:

- Excellent category: MPRD (0 – 4.33%), MWEDD (0 – 4.517%)
- Very good category: MPRD (4.33 – 11.59%), MWEDD (4.517 – 6.914%)
- Good category: MPRD (11.59 – 22.57%), MWEDD (6.914 – 11.125%)

As the proposed diagnostic measure (SCPRD) is evaluated only for MEGC signal, the clinically acceptable range can be found out through subjective test.

5.4 Conclusion

In this paper, a new diagnostic distortion measure is presented for MEEG signals. This measure (SCPRD) is defined as the sum of percentage root mean square difference between magnitudes of convolution response of original and processed MEEG signals. The proposed SCPRD measure is tested over MEMD- and MSPCA-based MEEG enhancement and compression schemes. The values of SCPRD are compared with existing distortion measures as AWEDD and MPRD. The future direction of our work is to find out the clinically acceptable range for SCPRD measure through subjective test.

References

1. Malmivuo, J., Plonsey, R.: Bioelectromagnetism: principles and applications of bioelectric and biomagnetic fields. Oxford University Press, Oxford (1995)
2. Thaler, M.S.: The Only EKG Book You'll Ever Need, vol. 365. Lippincott Williams and Wilkins (2010)
3. Sharma, L.N., Dandapat, S., Mahanta, A.: Multichannel ECG data compression based on multiscale principal component analysis. *IEEE Trans. Inf. Technol. Biomed.* **16**(4), 730–736 (2012)
4. Sharma, L.N., Dandapat, S., Mahanta, A.: Multiscale principal component analysis to denoise multichannel ECG signals. In: 2010 5th Cairo International Biomedical Engineering Conference (CIBEC), pp. 17–20. IEEE (2010)
5. Zigel, Y., Cohen, A., Katz, A.: The weighted diagnostic distortion (WDD) measure for ECG signal compression. *IEEE Trans. Biomed. Eng.* **47**(11), 1422–1430 (2000)
6. Al-Fahoum, A.S.: Quality assessment of ECG compression techniques using a wavelet-based diagnostic measure. *IEEE Trans. Inf. Technol. Biomed.* **10**(1), 182–191 (2006)
7. Manikandan, M.S., Dandapat, S.: Multiscale entropy-based weighted distortion measure for ECG coding. *IEEE Signal Process. Lett.* **15**, 829–832 (2008)
8. Manikandan, M.S., Dandapat, S.: Effective quality-controlled SPIHT-based ECG coding strategy under noise environments. *Electron. Lett.* **44**(20), 1182–1183 (2008)
9. Manikandan, M.S., Dandapat, S.: Wavelet energy based diagnostic distortion measure for ECG. *Biomed. Signal Process. Control* **2**(2), 80–96 (2007)
10. Fleureau, J., Kachenoura, A., Albera, L., Nunes, J.-C., Senhadji, L.: Multivariate empirical mode decomposition and application to multichannel filtering. *Signal Process.* **91**(12), 2783–2792 (2011)
11. Mandic, D.P., et al.: Empirical mode decomposition for trivariate signals. *IEEE Trans. Signal Process.* **58**(3), 1059–1068 (2010)
12. Oeff, M., Koch, H., Bousseljot, R., Kreiseler, D.: The PTB diagnostic ECG database. National Metrology Institute of Germany <http://www.physionet.org/physiobank/database/ptbdb> (2012)

Chapter 6

Application of SHAZAM-Based Audio Fingerprinting for Multilingual Indian Song Retrieval

S. Sri Ranjani, V. Abdulkareem, K. Karthik and P.K. Bora

Abstract Extracting film songs from a multilingual database based on a query clip is a challenging task. The challenge stems from the subtle variations in pitch and rhythm, which accompany the change in the singer's voice, style, and orchestration, change in language and even a change in gender. The fingerprinting algorithm must be designed to capture the base tune in the composition and not the adaptations (or variations which include lyrical modifications and changes in the singer's voice). The SHAZAM system was developed for capturing cover audio pieces from millions of Western songs stored in the database, with the objective of tapping into the melodic construct of the song (devoid of other forms of embellishments). When applied to the Indian database the system was found less effective, due to subtle changes in both rhythm and melody mainly due to the semiclassical nature of Indian film songs. The retrieval accuracy was found to be 85 %. Potential reasons for the failure of this SHAZAM system have been discussed with examples.

Keywords Audio fingerprinting · Hash · Multilingual song retrieval · Indian songs · Spectrogram sum

6.1 Introduction

Audio fingerprinting or content-based music retrieval systems extract a *simulated* perceptual digest from an audio file called a fingerprint (or a hash) and store it in a database. When presented with unlabeled audio, its fingerprint is calculated

S. Sri Ranjani (✉) · K. Karthik · P.K. Bora
Indian Institute of Technology, Guwahati 781039, India
e-mail: s.ranjani@iitg.ernet.in

K. Karthik
e-mail: k.karthik@iitg.ernet.in

P.K. Bora
e-mail: prabin@iitg.ernet.in

V. Abdulkareem
College of Engineering, Cherthala, Malappuram 673641, India
e-mail: vabdulkareemv@gmail.com

and matched against those stored in the database. Using fingerprints and matching algorithms, distorted versions of a recording can be identified as the same audio content [1].

Due to the increase in the amount of music production in the film industry, there is a huge corpus of audio files available for the audience to select from. Specifically, the South Indian film industry collaterally makes over 500 films every year. The increase in film budgets has made bilingual and trilingual viable options [2]. By providing an excerpt of a song in a particular language, it should be possible to retrieve all the songs that belong to the same composition.

In the literature, the problem of identifying alternate versions of a song, differing from the original song in timbre, tempo, key, arrangements, or language of the vocals is called the *cover version* problem [3]. The main challenge from a signal processing perspective is to automate the distinct perceptual reading of a song, by identifying the right choice of features. In [4], Ellis and Poliner have used *beat synchronous chroma*-based features, which represent the signal energy's distribution across a predefined set of pitch classes. To compare any two song pieces, a cross-correlation is performed between the chroma features and presence of sharp peaks indicate a good local alignment between the pieces. In [5], Kim and Narayan proposed a delta chroma as the feature for cover song detection. This delta chroma feature along with the chroma feature provide a better retrieval rate but at the cost of an increase in the feature size. In [6] by Tsai et al., the tune from the vocal rendition part only is extracted and the underlying note sequence is estimated and the tune is saved as a MIDI file. The MIDI sequence of the query is compared with the already stored note sequences in the database using the dynamic time warping (DTW) algorithm. While this algorithm was found to be moderately effective despite tempo, singer, and language variations, the process was computationally intensive since heavy pre- and postprocessing is required to extract and match the note sequences.

Singer pitch tracking techniques have been shown to be effective for music retrieval, provided the pitch trajectory is tracked accurately [7, 8]. The pitch trajectory carries significant information regarding the melodic structure. If the base pitch is normalized across several singers, then the trajectory which carries the relative pitch variations, can be used to discriminate between melodies. Furthermore, to match songs with different temporal scale variations, approaches such as DTW [9] are used to align the pitch trajectories. Such a complex temporal alignment based on DTW may not be required, if the onset of a note can be reliably detected from a vocal rendition [10, 11]. Multiband pitch tracking approaches have also been proposed [12], but these do not offer a significant improvement over plain pitch tracking.

Fingerprinting-inspired set of features for music retrieval was first proposed in Shazam fingerprinting. In Shazam, Wang [13] identify peaks in the spectral domain of the audio signal, and record the distances between them. For cover songs, which are similar in composition, but differ in the language and singer, where the melody and the orchestration are preserved often, the sequences of peaks and the time intervals between them can be used as hash codes. The overall feature-linking structure resembled a constellation map. The positive side to the Shazam system was that there

was no attempt to temporally align the song pieces or according to pitch, which made the implementation fast and computationally simple.

In this paper, we propose to analyze the MATLAB implemented version of the Shazam system by Ellis which is available in [14] for retrieval of songs from a multilingual database. This code is originally designed to find, from a snippet of a song as a query, the original song which contains the query available in the database. Even if the query is corrupted with noise, the system is able to extract the song. This code is adapted for Indian film songs and analyzed how it works and retrieves songs from a multilingual database.

The rest of this paper is organized as follows: The Shazam algorithm including the selection of features and the hash construction process is discussed in Sect. 6.2. In Sect. 6.3, robustness of the Shazam [15] feature set, for same compositions, with changes in language is analyzed and the experimental results are discussed in Sect. 6.4.

6.2 Shazam Fingerprinting

Shazam is a commercially available audio search engine. The algorithm is robust to noise and computationally efficient, capable of identifying a very short snippet of song captured through a mobile device in the presence of background noises, from a database of over a million tracks [13]. Our system is based on Shazam fingerprinting algorithm [13] and Ellis's implementation of the algorithm in MATLAB [14]. This code is adapted for Indian film songs and analyzed for how it works and retrieves songs from a multilingual database.

6.2.1 Feature Selection

In [13] the audio file is first converted into time-frequency spectrogram representation. The spectrogram of the audio signal is calculated as

$$X(\tau, k) = \psi\{x(n)\} = \sum_{k=0}^{N-1} x(n)w(n - \tau)e^{-jnk/N} \quad (6.1)$$

where $\psi\{\cdot\}$ is the short time Fourier transform (STFT) operator. $x(n)$, is the signal to be transformed and $w(n)$ is the window function of length N . Its absolute values are stored. Then the amplitude peaks from the spectrogram are selected locally and taken for further processing. The spectrogram peaks are selected as the primary feature, because they are most likely to survive noise. A trimming of the spectrogram peaks is done based on a threshold criterion. Once the shortlisting is done, only those points are chosen as the constellation points (Fig. 6.1).

6.2.2 Associative Hash Formation

Each constellation point P_i is a vector comprising of three different pieces of information as shown in Fig. 6.1a (i) Timing information t_i (i.e. temporal location), (ii) Frequency f_i , (iii) Magnitude of the spectral peak A_i . Of these pieces, Shazam uses only the first two to represent a particular constellation point. Thus, $P_i \equiv (t_i, f_i)$. For computing a sequential hash, the constellation point P_i is chosen as an anchor and the relative positions of m neighboring points $P_{i+1}, P_{i+2}, \dots, P_{i+m}$ are used to generate the hash. Thus, every hash comprises of a cluster of $m + 1$ points with one reference point. As the hashing window is slid over the constellation, such an associative hash is computed for every constellation point, which is shown in Fig. 6.1b. If N_{anchor} is the number of anchor points per second of a particular music clip and m is the number of points associated with each anchor point, the total storage complexity, is proportional to,

$$Storage = O(m \cdot N_{anchor}) \tag{6.2}$$

And the computational complexity is related to the total number of anchor points per second,

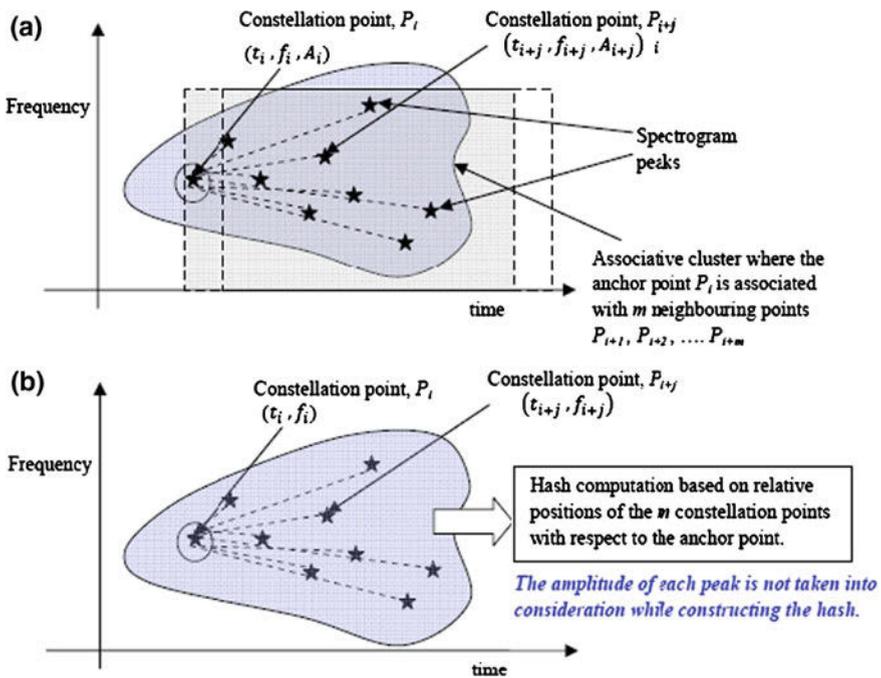


Fig. 6.1 Selection of features as constellation points in the Shazam fingerprinting scheme. **a** Constellation description in SHAZAM. **b** Features for hash computation with reference to each anchor point

$$\text{Computation} = O(N_{\text{anchor}}) \quad (6.3)$$

In Shazam, matching is done by aligning constellations through a hash matching process. However, the feature extraction process is heavily dependent on the robust detection of the spectrogram peaks, while the hash generation and matching is dependent on the relative positions of the constellation points. Only if the hash is invariant to shift and scale can the matching be robust. We have found based on a feature analysis that spectrogram peak detection is not very robust to lyrical changes in the song, as this results in a drift of constellation points and also possibly the insertion/deletion of several new points. Since the hashing does not depend on the absolute time positions, the shift in the peak positions in time axis is taken care of.

Hash strings are calculated from the constellation maps, in which pairs of spectrogram peaks are associated combinatorially. Each point in the constellation map is taken as an anchor point and a target zone is fixed with it. All the points in the target zone are sequentially paired with the anchor point, each pair giving two frequency values and the time difference between the points. A database index is created using the hash obtained from all the files in a database and their corresponding offset times, with a 64-bit struct, where first 32 bits are for the hash and the second 32 bits are for the time offset and song ID combined together. The number of hashes per second of the audio depends on the number of spectrogram peaks selected and also the size of the target zone. As the size of the target zone increases, the number of combinatorial explosion of song increases.

6.2.3 Matching

Once the hash strings are computed, matching is done. The hash value for the query is extracted using the same procedure described above. Each hash thus generated is used to search the stored database hash table. If a match is found the corresponding offset time (t_k) and song_ID is paired with query offset time (t'_k). This matching procedure is carried out for all hashes in the query. Then the difference between offset times ($\delta t_k = t_k - t'_k$) of all matches is calculated and a histogram of this δt_k is computed for every song in the database. The peak in the histogram gives the matching score and position of match with the song. To take care of small changes in time of events in original and alternate version songs hash matches with ± 1 tolerance are also considered. The matching score (MS) with a specific song r is given as

$$MS_r = \frac{\text{Matching Score of song } r}{\text{Total no. of hashes in the query}} \times 100 \% \quad (6.4)$$

Alternate versions of the query song will have sufficient percentage match. By setting a suitable threshold in percentage match, similar songs can be retrieved from the database. The threshold is dependent on the parameters used for hashing and also the application.

6.3 Experimental Setup

6.3.1 Database

To check how the Shazam system works in finding the tune similarity across different Indian languages, we focus on a set of songs which are explicitly covered in other languages. Here we use parallel songs in three Indian languages, viz. Hindi, Tamil, and Telugu. Since there are no publicly available multilingual database, songs are collected totalling upto 178 songs with an average duration of 5 min each. All songs are in MP3 format and are downloaded from various MP3 downloadable websites. The split in the database based on different languages is shown in Table 6.1. The database is arranged such that the assumed original version songs are taken and fingerprints are extracted only for those songs and stored as hash tables. Queries are taken from the remaining cover versions in different languages and compared with the hash table and the corresponding match is obtained. All the songs in the database are preprocessed by resampling to 8 KHz and the mean of the two channels of stereo songs is computed before finding the fingerprints.

Songs in one language, viz., Hindi are taken as the original version and stored as the database. Hash table is created for the songs in the database. The other language songs, Telugu and Tamil, are considered as the cover version and are searched in the database. The songs which matches more than a preset threshold value, which is selected empirically, will be retrieved as a ranked list. Retrieval rate of Top-N means the percentage of queries whose original version is in the Top-N results.

6.3.2 Robustness to Language Change

Hash formation in Shazam is based on the absolute values of time difference and frequency of combinatorial pairs of spectrogram peaks. Peak points in the spectrogram may be from the instrumental or vocal. Hence the hashes created from these peak points depend on both the instrumental and vocal part. When there is a change in the instrumental or vocal segment only a small percent of hashes will not match. Vocal changes include changes in the singer, lyrics, and language. Figure 6.2 shows the Shazam features extracted from a musical piece sung in two different languages, Hindi and Tamil. The constellation points for the Shazam show a considerable

Table 6.1 Database—
different languages

Language	Number of songs
Hindi	53
Tamil	85
Telugu	40
Total	178

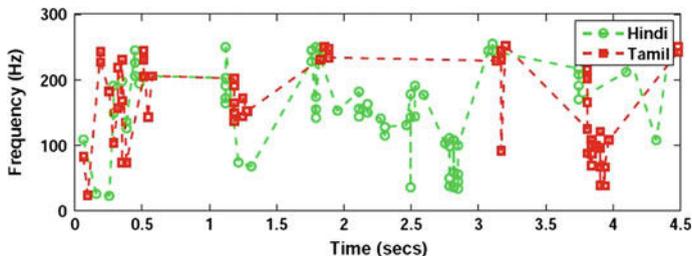


Fig. 6.2 Analysis of Shazam features for songs in different languages with same composition

irregularity for both the language versions. There is not only a misalignment of points, but several insertions and deletions can be observed per frame. Even with these disadvantages, the retrieval rate is high.

6.3.3 Results

Spectrogram of each preprocessed song is calculated using a window length of $N = 1024$ (128 ms for a sampling rate of 8000 Hz) is taken with 75% overlap with the next frame to reduce artifacts at the boundary. The resulting spectrogram is a complex number which is then converted into its absolute value and stored. A hash table is created for the original songs (one version is taken as original and others as alternatives) in the database. A part of the alternative song is chosen as the query and searched in the database. The songs that match more than the threshold value will be retrieved as a ranked list based on the matching percentage. From different languages, a total of 178 songs were selected, with 72 of them being distinct. A segment from each song was chosen as the query and matched with all other songs. Retrieval rate and false positive rate were calculated using Eqs. (6.5) and (6.6),

$$\text{Retrieval Rate} = \frac{NR}{N_q} \times 100 \quad (6.5)$$

where NR —number of queries whose original version is retrieved, N_q —number of queries.

$$\text{False Positive Rate} = \frac{\sum FP_q}{\sum (FP_q + TN_q)} \times 100 \quad (6.6)$$

where, FP_q —number of false retrievals for each query, TN_q —number of true negatives for each query.

The retrieval rate of Top-N means the percentage of queries whose original version is in the Top-N results. For small threshold values, the retrieval rate is expected to be high but accuracy will be extremely poor due to several false positives. When the

Table 6.2 Retrieval rate for different languages (with query length = 30 s)

Top1	Top3	With FPR			
		0.3 %	0.2 %	0.1 %	0 %
91.56	97.59	87.95	86.74	81.92	74.69

Table 6.3 Retrieval rate for different languages (with query length = 20 s)

Top1	Top3	With FPR			
		0.3 %	0.2 %	0.1 %	0 %
84.33	91.56	75.90	72.28	68.87	62.65

threshold is increased the selection process is made more stringent, which will lead to a reduction in the false positive rate and also decrease in retrieval rate. Tables 6.2 and 6.3 show the retrieval results when queries are submitted to a multilingual database. For a fixed false positive rate ($FPR = 0.3\%$, Query length = 30s), the accuracy of the Shazam system is 87.95 % (Table 6.2). For $FPR = 0\%$, the accuracy is 74.69 %. Changes in language create fluctuation in pitch which disturbs the spectrogram peaks used in Shazam, which leads to poor performance for lesser FPR. The performance degradation is slightly severe for shorter queries (20s) as can be seen in Table 6.3, where the accuracies corresponding to $FPR = 0\%$ is 62.65 %.

6.4 Challenges and Scope for Future Work

One of the main problems with SHAZAM is that the rhythms are not used to adapt the window size for computing the STFT. The STFT uses a fixed window size of 64 ms and the duration of every fundamental note *Sa, Re, Ga, Ma, Pa, Dha, Ni* varies according the style of the singer and the lyric used based on the language (Hindi, Tamil, Malayalam, etc.). Hence, synchronicity of the spectral feature with respect to the underlying rhythm is not maintained.

Because of the loss of synchronous timing information, the hash values change when there are style and lyrical changes. Furthermore, when there is a change in the voice of the singer, such as a change from a male voice to a female voice, the peaks are not only out of temporal sync because of a change in style, the spectral peaks also shift in the frequency scale since the base note frequency will be different for different singers. For instance, the *Sa* note will be higher for a female singer as compared to a male singer. Consequently, the spacings between the harmonics will also change leading to a shift in the peaks. Unless normalization in the frequency scale and timescale is done, the hash will remain sensitive to changes in the voice/lyrics and style. Two examples have been given to illustrate the effects of lyrical changes and also changes in the singer's voice. Figure 6.3 shows the spectrogram peaks of a

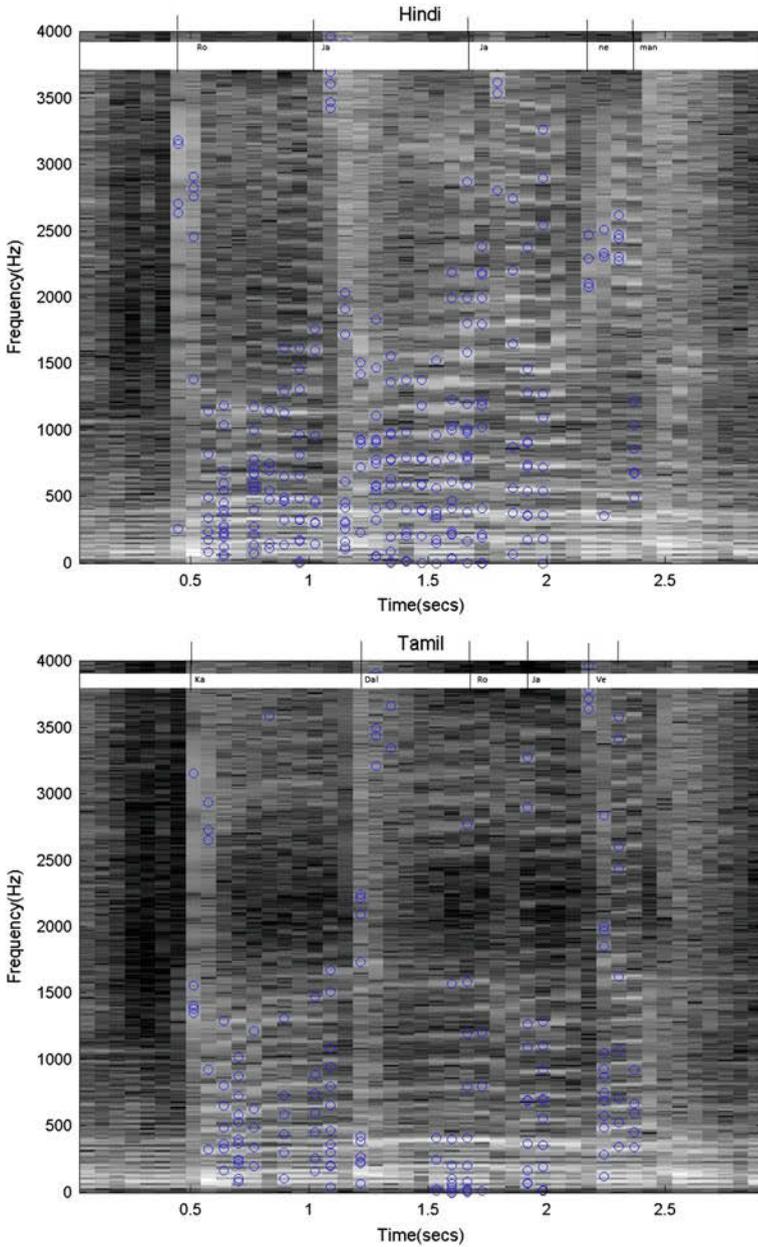


Fig. 6.3 Spectrogram and its peaks of Hindi song (Film: Roja, Song: Roja janeman) (Top), Tamil song (Film: Roja, Song: Kadhal Rojave) (Bottom) sung by the same singer S.P. Balasubramaniam

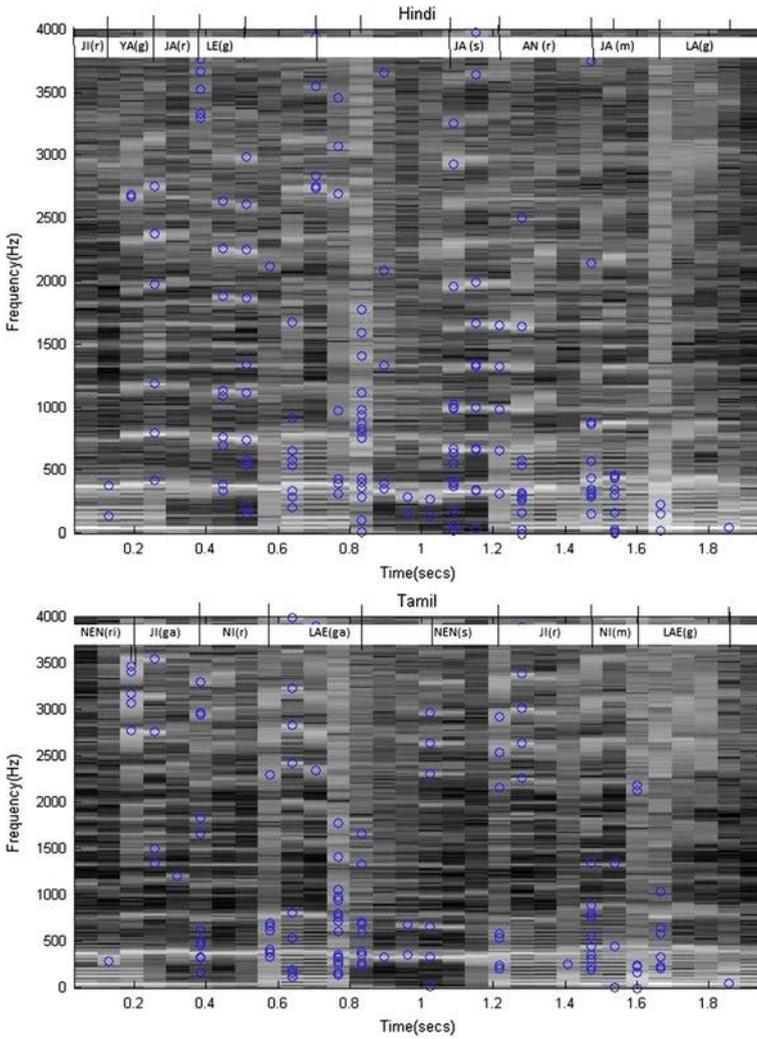


Fig. 6.4 Spectrogram and its peaks of Hindi song (Film: Dil se, Song: Jiya Jale) (Top), Tamil song (Film: Uyire, Song: Nenjinile) (Bottom) sung by Lata Mangeshkar and Janaki

musical piece sung in two different languages by the same singer S.P. Balasubramanium. The differences in the locations of the constellation points for the two versions are considerable, mainly because of the significant lyrical change introduced when one language is translated to another. For instance, the lyric used in Tamil is “Kadhal Rojave” while the one used in Hindi is “Roja Janeman”. These may be termed as “musical anagrams” where the order of the notes are rearranged mainly because of the lyrics. Figure 6.4 shows the spectrogram peaks of a musical piece sung in two

different languages (Hindi and Tamil) by different singers (Lata Mangeshkar and S. Janaki). In contrast to the earlier example where the change in lyrics introduced a musical anagram, here the notes and the orderings are better matched. However, because of subtle differences in the base pitch of both the singers the constellations points are not spaced identically in the frequency space. The differences in the two constellation maps are in the frequency spacings and less in the temporal segment.

6.5 Conclusions

In this paper we have applied the SHAZAM system which was developed for Western songs to a multilingual Indian database. The retrieval accuracy was found to be 85%, indicating that the system is not designed for the variations in semiclassical Indian songs. Synchronicity in rhythm and normalization across the frequency scale is necessary to absorb the variations in the base composition due to changes in the style, language, and voice of the singer. This will be a subject for future work.

References

1. Cano, P., Batlle, E., Kalker, T., Haitsma, J.: A review of audio fingerprinting. *J. VLSI Signal Process. Syst.* **41**, 271–284 (2005)
2. South film gets bigger. Article, *Times of India*, 31 May 2011
3. Serra, J., Gomez, E., Herrera, P.: Audio cover song identification and similarity: background, approaches, evaluation, and beyond. *Adv. Music Inf. Retr.* **274**, 307–332 (2010)
4. Ellis, D.P.W., Poliner, G.E.: Identifying ‘cover songs’ with chroma features and dynamic programming beat tracking. In: *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 4, pp. 1429–1432 (2007)
5. Kim, S., Narayanan, S.: Dynamic chroma feature vectors with applications to cover song identification. In: *IEEE Workshop on Multimedia Signal Processing (MMSp)*, pp. 984–987 (2008)
6. Tsai, W., Yu, H., Wang, H.: Using the similarity of main melodies to identify cover versions of popular songs for music document retrieval. *J. Inf. Sci. Eng.* **24**, 1669–1687 (2008)
7. Pollastri, E.: Melody-retrieval based on pitch-tracking and string-matching methods. In: *Proceedings of the XIIth Colloquium on Musical Informatics* (1999)
8. Zhu, Y., Kankanhalli, M., Tian, Q.: Similarity matching of continuous melody contours for humming querying of melody databases. In: *IEEE Workshop on Multimedia Signal Processing*, pp. 249–252 (2002)
9. Lijffijt, J., Papapetrou, P., Hollmén, J., Athitsos, V.: Benchmarking dynamic time warping for music retrieval. In: *Proceedings of the 3rd International Conference on Pervasive Technologies Related to Assistive Environments, PETRA’10* (2010)
10. Kumar, P.P., Rao, P., Roy, S.D.: Note onset detection in natural humming. In: *Proceedings of the International Conference on Computational Intelligence and Multimedia Applications, ICCIMA’07*, pp. 176–180 (2007)
11. Thoshkahna, B., Ramakrishnan, K.R.: An onset detection algorithm for query by humming (QBH) applications using psychoacoustic knowledge. In: *Proceedings of 17th European Signal Processing Conference (EUSIPCO), EURASIPCO’09* (2009)

12. Kumar, N., Tsiartas, A., Narayanan, S.: Features for comparing tune similarity of songs across different languages. In: International Workshop on Multimedia Signal Processing, pp. 331–336 (2012)
13. Wang, A.L.: An industrial-strength audio search algorithm. In: Proceedings of International Conference on Music Information Retrieval, Baltimore, Maryland (2003)
14. Ellis, D.: Robust landmark-based audio fingerprinting (2007)
15. Wang, A.: The Shazam music recognition service. *Commun. ACM* **49**, 44–48 (2006)

Chapter 7

Multilevel-DWT-Based Image Denoising Using Adaptive Neuro-Fuzzy Inference System

Torali Saikia and Kandarpa Kumar Sarma

Abstract Images corrupted by noise requires enhancement for subsequent processing. Traditional approaches of denoising rely upon spatial, statistical, and spectral properties of image which at times fails to capture the finite details. Discrete wavelet transform (DWT) is a commonly adopted method for image processing applications. Fuzzy-based systems are suitable for modeling uncertainty. In the proposed work, we present a hybrid approach which combines multilevel DWT and adaptive neuro-fuzzy inference system (ANFIS) to capture the benefits of two different domains into a single framework. We apply our algorithm to denoise the images corrupted by multiplicative noise like speckle noise. The results obtained shows that the proposed method proves effective for denoising of images.

Keywords Noise · Denoising · Multilevel discrete wavelet transform (DWT) · Adaptive neuro-fuzzy inference system (ANFIS)

7.1 Introduction

Digital images are a part of present day devices ranging from hand held mobile devices to diagnostic equipments. It is seen that during image acquisition, storage, retrieval, or transmission, due to various unavoidable reasons images get contaminated by noise. The presence of noise highly degrades the quality of image. Here, lies the importance of denoising, which is a form of image restoration operation dealing with removal of noise. The main criteria for the different denoising approaches is to remove these noise and achieve quality enhancement taking into account that

T. Saikia (✉) · K.K. Sarma
Gauhati University, Guwahati 781014, India
e-mail: torali493@gmail.com

K.K. Sarma
e-mail: kandarpaks@gmail.com

the details of the images are preserved [1] which are to be used for subsequent processing. With different varieties of noise and its extent, denoising becomes challenging. Additive noise (e.g., Gaussian and impulse) and multiplicative noise (e.g., speckle) are considered as the two most commonly occurring types of noise. It is seen that the task of denoising is dependent on the type of noise that contaminates the image. Various works related to image denoising have appeared in the literature. Most of them deal with removal of a particular type of noise like Gaussian [2], impulse [3], speckle [4–6] etc. In the literature due to energy compaction property and multiresolution capabilities, wavelet transform (WT) [7, 8] is widely used for image denoising. Also it is seen that due to numerous advantages like robustness and human-like learning capabilities, different soft computing techniques are being increasingly used as image denoising tools. Traditional approaches of denoising rely upon spatial, statistical, and spectral properties of image which at times fail to capture the finite details. Discrete wavelet transform (DWT) is a commonly adopted method for image processing applications. Fuzzy-based systems are suitable for modeling uncertainty. In the proposed work, we present a hybrid approach which combines multilevel DWT and adaptive neuro-fuzzy inference system (ANFIS) to capture the benefits of both in a single framework. We apply our algorithm to denoise the images corrupted by a kind of multiplicative noise known as speckle noise. The results show that the proposed method proves effective for a range of variations and is suitable for critical applications.

The rest of this paper is divided among the following sections. In Sect. 7.2, we provide the relevant theoretical concepts. In Sect. 7.3, we describe the proposed model whereas in Sect. 7.4, experimental details and results are included. In Sect. 7.5 we draw the conclusion regarding the proposed model.

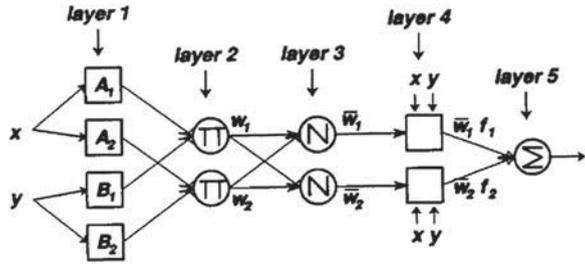
7.2 Related Concepts

Here, we briefly discuss about ANFIS and DWT.

7.2.1 Adaptive Neuro-Fuzzy Inference System

ANFIS is a hybrid soft computational technique which can be considered as an integration of both artificial neural network (ANN) and fuzzy inference system (FIS). ANFIS is a kind of ANN that is based on Takagi–Sugeno FIS with learning capabilities. The inference system corresponds to a set of fuzzy IF-THEN rules that have learning capability to approximate nonlinear functions. ANFIS is an adaptive network trained with a learning rule like backpropagation. It is known by the name of adaptive network because some or all the nodes have parameters which affect the output of the nodes [9]. Neuro-Fuzzy Systems (NFS) can combine the parallel computation and learning abilities of ANNs with human-like knowledge representation

Fig. 7.1 Architecture of ANFIS



and explanation abilities of fuzzy systems. As a result of which, NFSs become more transparent and with fuzzy system attribute become capable of learning finite details.

Figure 7.1 represents the architecture of an ANFIS. Here, the circular nodes represent the nodes that are fixed whereas the square nodes are nodes that have parameters to be learnt. Layer 1 performs the fuzzification process, layer 2 is the rule layer whereas layer 3 is the normalization layer. Defuzzification process is carried out in layer 4. Layer 5 calculates the sum of outputs of all defuzzification neurons and produces the overall output.

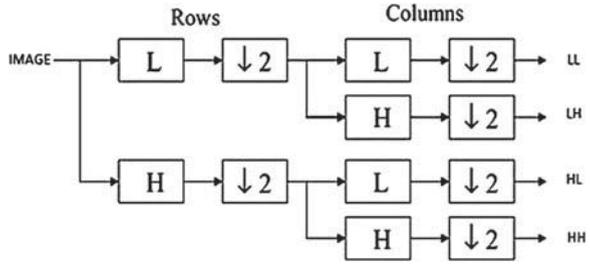
7.2.2 Discrete Wavelet Transform

When the wavelets are discretely sampled it is referred to as DWT, which is a multiresolution decomposition structure. It reduces the computation time and is easier to implement. It captures both, frequency and location information. Using DWT, the original image is decomposed into two subspaces viz. high frequency subband and a low frequency subband [10]. At every iteration of the DWT, the lines of the input image (obtained at the end of the previous iteration) are low-pass filtered with filter L and high-pass filtered with filter H. Then the lines of the two images obtained at the output of the two filters are decimated with a factor of 2. Next, the columns of the two images obtained are low-pass filtered with L and high-pass filtered with H. The columns of those four images are also decimated with a factor of 2. This entire algorithm is shown in Fig. 7.2.

This results in four new subimages. The first subimage is named LL image or approximation subimage. Rest of the three are called detail subimages: LH, HL, and HH. For the next iteration, input is represented by LL image. In the following coefficients of DWT noted with iD_m^l , i represents the image whose DWT is computed, the resolution level or iteration represented by m and $l = 1$ for HH image, $l = 2$ for HL, $l = 3$ for LH, and $l = 4$ for LL image. These coefficients are calculated using Eqs. (7.1) and (7.2).

$$iD_m^l [n, p] = \langle i(\tau_1, \tau_2), \psi_{m,n,p}^l(\tau_1, \tau_2) \rangle \tag{7.1}$$

Fig. 7.2 Wavelet decomposition for two-dimensional pictures



where we can factorize the wavelets as:

$$\psi_{m,n,p}^l(\tau_1, \tau_2) = \alpha_{m,n,p}^l(\tau_1) \cdot \beta_{m,n,p}^l(\tau_2) \tag{7.2}$$

Using the following relations and the scale function $\varphi(\tau)$ and mother wavelet $\psi(\tau)$, the two factors can be calculated using Eqs. (7.3) and (7.4).

$$\alpha_{m,n,p}^l(\tau) = \begin{cases} \varphi_{m,n}(\tau), & l = 1, 4 \\ \psi_{m,n}(\tau), & l = 2, 3 \end{cases} \tag{7.3}$$

$$\beta_{m,n,p}^l(\tau) = \begin{cases} \varphi_{m,n}(\tau), & l = 2, 4 \\ \psi_{m,n}(\tau), & l = 1, 3 \end{cases} \tag{7.4}$$

where,

$$\varphi_{m,n}(\tau) = 2^{-m/2} \varphi(2^{-m} \tau - n) \tag{7.5}$$

$$\psi_{m,n}(\tau) = 2^{-m/2} \psi(2^{-m} \tau - n) \tag{7.6}$$

7.3 Proposed Model

The proposed approach is summarized in Fig. 7.3. Here, we combine the features of multilevel DWT with a hybrid soft computing system like ANFIS. The proposed model is an amalgamation of advantages of both DWT and ANFIS in a single framework. The basic concepts of ANFIS and DWT have already been discussed in Sect. 7.2. A three-level DWT is applied to a noise-free image to obtain 10 subbands named as LL1, LH1, HL1, HH1, LH2, HL2, HH2, LH3, HL3, and HH3. We partition each of these subbands into eight different blocks. Each of these blocks serve as reference for the ANFIS. Here, we use the bell-shaped membership function (MF) which defines all the information necessary to make fuzzy inferences based on certain rules. While testing, we provide the noise-mixed pixel and repeat a similar process as described for clean image. In case of the corrupted image, the different

Fig. 7.3 Block diagram of the proposed model

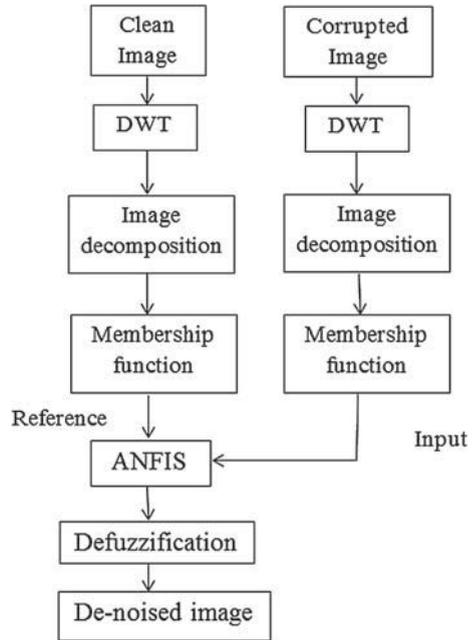
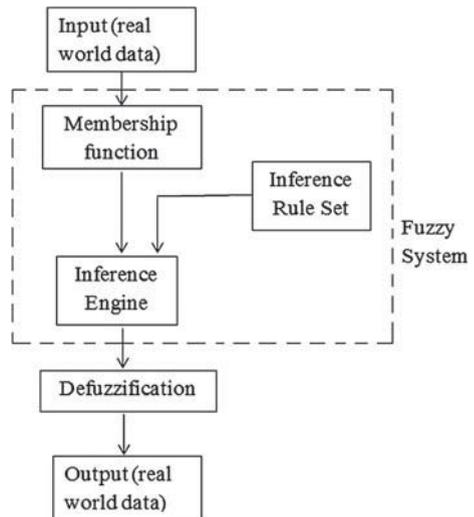


Fig. 7.4 Fuzzification and defuzzification process of the ANFIS



blocks of the 10 subbands serve as input for the ANFIS. In this way, we train the network to learn the characteristics of speckle and eliminate the effect of speckle. The output of the ANFIS gives the denoised image.

The ANFIS system involves fuzzification and defuzzification process as described in Fig. 7.4. The uncertainty observed in pixels of an image corrupted by noise is learnt

during training by the ANFIS system. It does so by carrying out comparisons with a reference image from which inference rules are derived. The real world to fuzzy and reverse mapping is executed by ANFIS itself but it involves certain critical steps. The MFs play a significant role during fuzzification and inference making. The Bell MF is the most commonly used type as it offers several advantages like closely approximating real-world data, invariance to transforms, and flexibility of parameter selection. As seen in Fig. 7.4, the following steps are integral to an ANFIS.

- Fuzzification is the process of transforming crisp quantities to fuzzy quantities using MF.
- Inference engine is mainly involved with the decision-making process. It uses “IF-THEN” rules along with connectors “OR” or “AND” for making necessary decision rules from the inference rule set. The output from the fuzzy inference engine is always crisp set.
- Defuzzification is the process of converting the fuzzy results into crisp results suitable for various real-world applications.

7.4 Experimental Details and Results

The configuration of ANFIS used is summarized in Table 7.1. For experimental purpose and for a range of variations, we consider images of different sizes. By applying the proposed algorithm for different noise levels for all the different image types and sizes, we obtain the denoised image. A few samples considering four test images viz. Cameraman, Lena, Mandril, and medical (Ultrasound) of sizes 256×256 are shown in Fig. 7.5(a–h).

Table 7.1 Configuration of ANFIS

No. of layers of ANFIS	5
Membership function	Bell curve
Training method	Fuzzified back propagation
Data volume for training	20 images each of sizes (128×128) , (256×256) , (512×512) with noise-free and noise-mixed forms
Data volume for testing	20 images each of sizes (128×128) , (256×256) , (512×512) with noise-free and noise-mixed forms
No. of iterations	min-50 max-200

We then calculate the value for mean square error (MSE) and peak signal-to-noise ratio (PSNR) for both corrupted image and the denoised image using Eqs. (7.7) and (7.8).

$$PSNR = 10\log_{10} \left(\frac{\max^2}{MSE} \right) \quad (7.7)$$

where,

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i, j) - K(i, j)]^2 \quad (7.8)$$

Here, I and K are the original and noisy/denoised images, respectively, m and n represents the dimensions of the images and \max represents the maximum pixel value. The experimental results are tabulated from Tables 7.2, 7.3, 7.4 and 7.5.

We also compared the effectiveness of the proposed work with three conventional denoising techniques viz. Median filtering (MF), Bilateral filtering (BF), and Wiener filtering (WF). Table 7.6 gives the comparative analysis of PSNR values for the mentioned conventional denoising techniques with the proposed method. For this, we consider four test images viz. Cameraman, Lena, Mandril, and Ultrasound image corrupted by speckle noise of variance 0.02. Lower the value of MSE and higher the value of PSNR, better will be the image quality. From the experimental results, we find that there has been an improvement in PSNR value of the denoised image as compared to the noisy image. The visual quality of the denoised image is also better than that

Table 7.2 PSNR of Cameraman image with different noise levels

Variance	PSNR of noisy image (dB)	PSNR of denoised image (dB)	% increase in PSNR
0.01	73.78	80.44	9.03
0.02	70.79	78.93	11.49
0.03	69.01	77.57	12.40
0.04	67.81	76.90	13.41

Table 7.3 PSNR of Lena image with different noise levels

Variance	PSNR of noisy image (dB)	PSNR of denoised image (dB)	% increase in PSNR
0.01	74.21	80.39	8.33
0.02	71.24	78.54	10.25
0.03	69.54	77.73	11.78
0.04	68.33	77.05	12.76

Table 7.4 PSNR of Mandril image with different noise levels

Variance	PSNR of noisy image (dB)	PSNR of denoised image (dB)	% increase in PSNR
0.01	72.67	76.28	4.97
0.02	69.71	74.48	6.84
0.03	68.02	73.61	8.22
0.04	66.82	72.99	9.23

Table 7.5 PSNR of ultrasound image with different noise levels

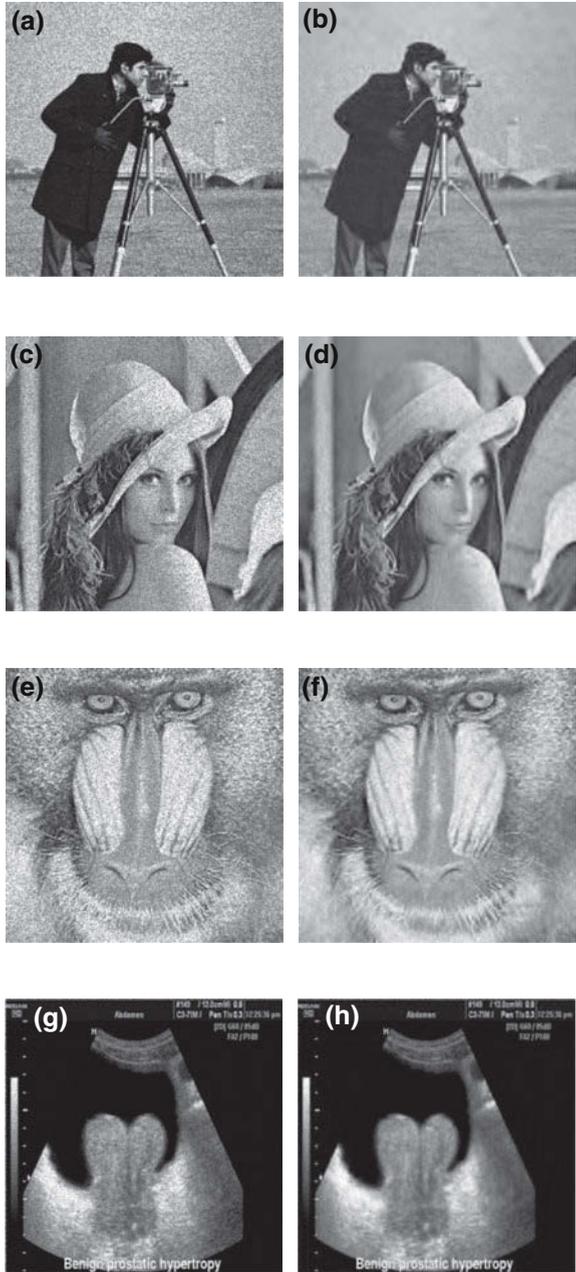
Variance	PSNR of noisy image (dB)	PSNR of denoised image (dB)	% increase in PSNR
0.01	78.24	82.37	5.27
0.02	75.29	80.35	6.72
0.03	73.53	78.62	6.92
0.04	72.38	78.13	7.94

Table 7.6 Comparative analysis of psnr values for different denoising methods

Image	Noisy image (dB)	Denoised image (dB) (MF)	Denoised image (dB) (BF)	Denoised image (dB) (WF)	Denoised image (dB) (Proposed method)
Cameraman	70.79	74.22	74.73	76.27	78.93
Lena	71.24	74.78	74.66	75.88	78.54
Mandril	69.71	70.93	72.13	71.01	74.48
Ultrasound	75.29	70.87	77.15	78.45	80.35

of the corrupted image. The comparative analysis of different denoising methods shows that the proposed denoising algorithm has superior features as compared to the conventional techniques. Some of the related works have been reported in [11].

Fig. 7.5 A sequence of denoised images obtained from the proposed approach.
a Corrupted image.
b De-noised image.
c Corrupted image.
d De-noised image.
e Corrupted image.
f De-noised image.
g Corrupted image.
h DE-noised image



7.5 Conclusion

Considering the importance of image denoising in various image processing applications here, we present one such approach to denoise the images. In this work, we proposed a hybrid approach combining the features of multilevel DWT and ANFIS to formulate an efficient image denoising algorithm. Here, for experimental purpose, we use different test images, for a range of variations and noise levels. We consider the speckle noise-affected images. Our evaluation is based on PSNR and visual methods. From the above experimental results, we find that the proposed method proves to be effective for image denoising and enhances the image quality. The proposed method requires no prior knowledge of the noise which proves to be more advantageous than many of the current techniques that assume noise model to be known. The method also preserves the image details. The effectiveness of our proposed work as compared to the conventional methods makes it suitable for critical applications.

References

1. Sbou, Y.A.AI.: Artificial neural networks evaluation as an image denoising tool. *World Appl. Sci. J.* **17**(2), 218–227 (2012)
2. Krishnan, M.H., Viswanathan, R.: A new concept of reduction of Gaussian noise in images based on fuzzy logic. *Appl. Math. Sci.* **7**(12), 595–602 (2013)
3. Sau, K., Chanda, A., Karmakar, P.: An unique edge preserving noise filtering technique for impulsive noise removal. *Int. J. Signal Image Process.* **2**(3), 33–46 (2011)
4. Kaur, M., Mandal, D.: Speckle noise reduction in medical ultrasound images using particle swarm optimization with artificial neural networks. *Int. J. Electron. Commun. Technol.* **2**(3), 147–150 (2011)
5. Ragesh, N.K., Anil, A.R., Rajesh, R.: Digital image denoising in medical ultrasound images: a survey. *Int. J. Artif. Intell. Mach. Learn.* 67–73 (2011)
6. Vishwa, A., Sharma, S.: Speckle noise reduction in ultrasound images by wavelet thresholding. *Int. J. Adv. Res. Comput. Sci. Softw. Eng.* **2**(2) (2012)
7. Ufade, A.S., Khadse, B.K., Suralkar, S.R.: Restoration of blur image using wavelet based image fusion. *Int. J. Eng. Adv. Tech.* **2**(2), 159–161 (2012)
8. Ruikar, S.D., Doye, D.D.: Wavelet based image denoising technique. *Int. J. Adv. Comput. Sci. Appl.* **2**(3), 49–53 (2011)
9. Jang, J.S.R.: ANFIS: Adaptive-network-based fuzzy inference system. *IEEE Trans. Sys. Man Cybern.* **23**(3), 665–685 (1993)
10. Kanika, D.N., Sharma, K.: Comparative analysis of discrete wavelet transform and fast wavelet transform on image compression. *Int. J. Eng. Res. Tech.* **1**(5), 1–7 (2012)
11. Saikia, T., Sarma, K.K.: Multilevel-DWT based image de-noising using feed forward artificial neural network. In: *IEEE International Conference on Signal Processing and Integrated Networks*, pp. 791–794 (2014)

Chapter 8

Active Learning Using Fuzzy k-NN for Cancer Classification from Microarray Gene Expression Data

Anindya Halder, Samrat Dey and Ansuman Kumar

Abstract Classification of cancer from microarray gene expression data is an important area of research in the field of bioinformatics and biomedical engineering as large amounts of microarray gene expression data are available but the cost of correctly labeling it prohibits its use. In such cases, active learning may be used. In this context, we propose active learning using fuzzy k-nearest neighbor (ALFKNN) for cancer classification. Active Learning technique is used to select most confusing or informative microarray gene expression patterns from the unlabeled microarray genes, so that labeling on the confusing data maximizes the classification accuracy. The selected most confusing/informative genes are manually labeled by the experts. The proposed method is evaluated with a number of microarray gene expression cancer datasets. Experimental results suggest that in comparison with traditional supervised k-nearest neighbor (k-NN) and fuzzy k-nearest neighbor (fuzzy k-NN) methods, proposed active learning method (ALFKNN) provides more accurate result for cancer prediction from microarray gene expression data.

Keywords Active learning · Cancer classification · Microarray gene expression data · (Fuzzy) k-nearest neighbor

8.1 Introduction

Traditional methods for cancer classification are based on the clinical tests and the morphological appearances of the tumor. These methods are expensive as well as time consuming. Latest technologies like microarray gene expression data address

A. Halder (✉) · A. Kumar
Department of Computer Application, North-Eastern Hill University, Tura Campus,
Tura 794002, India
e-mail: anindya.halder@gmail.com

S. Dey
Department of Computer Science, Don Bosco College, Tura 794002, India
e-mail: dey.samrat@gmail.com

A. Kumar
e-mail: ansuman.kumar@gmail.com

those issues in order to produce relatively inexpensive, primary diagnosis and prediction of cancer using computational approaches. Microarray is a technology that is capable of determining the expression levels of thousands of genes simultaneously. The limitations of the traditional methods are the strong bias in identifying the tumor by the observation of expert and also the difficulties in differentiating between cancer subtypes. This is due to the fact that most cancers are highly related to the specific biological insights such as responses to different clinical treatments. It therefore makes biological sense to perform cancer classification at the genotype level compared to the phenotypic observation. Due to the large amount of microarray gene expression data available on various cancerous samples, it is important to construct classifiers that have high predictive accuracy in classifying cancerous samples from the microarray gene expression data [1].

Several machine-learning algorithms have been applied to classify cancer from microarray gene expression data. Some of the methods are based on similarity like k -nearest neighbor (k -NN) [2]. Some methods rely on decision-tree and recursive partitioning [3]. Neural networks [4] and probabilistic induction [5] like Naive Bayes method can also be used for cancer classification.

Generally, traditional computational methods fail to achieve desired accuracy in class prediction using gene expression data due to low availability of clinically labeled samples. In these methods, the test/unlabeled data are not used for class prediction. Although unlabelled microarray gene expression patterns are easily available, the use of active learning method in this context aims to choose the most informative/confusing genes so that expert labeling of those unlabeled genes are possible to include in the training set, resulting in better prediction accuracy of the test samples. Usually, in the microarray gene expression data, the classes have overlapping nature due to which fuzzy classifiers tend to give better accuracy in class prediction in cancer classification problem [6]. Motivated from the above-mentioned issues, in this paper, we propose a novel active learning method using fuzzy k -NN (ALFKNN) for cancer classification from microarray gene expression data. The proposed method selects a set of most confusing microarray gene expression data from the unlabeled microarray genes such that labeling on the confusing genes maximizes the prediction accuracy. The selected most confusing genes are manually labeled by the experts.

Our proposed method is tested with six cancer datasets of microarray gene expression data. The accuracy of the proposed method is compared with that of the conventional state-of-the-art nearest neighbor-based supervised nonfuzzy (k -NN) and fuzzy classifier (fuzzy k -NN).

The organization of the rest of the article is as follows. In Sect. 8.2, the background of our work is given. In Sect. 8.3, detail description of the proposed active learning using fuzzy k -NN (ALFKNN) is presented. In Sect. 8.4, details of the experiments and analysis of the results are given, and finally, the conclusions are drawn in Sect. 8.5.

8.2 Background

As our proposed method active learning using fuzzy k-NN (ALFKNN) is an extension of fuzzy k-NN and k-NN classifiers, the brief outline of those classifiers are provided below.

8.2.1 *k-NN Classifier*

The k-nearest neighbor (k-NN) is one of the most popular and simplest nonparametric classification algorithm for pattern classification. This algorithm predicts the class level of test set based on the class levels represented by the k-nearest neighbors from the training set, where k is the (usually small) positive number. The distances are calculated between test data and all training data with the help of any distance measure (e.g., Euclidean distance, Mahalanobis distance etc.). After all the distances are calculated, they are sorted and the nearest k samples are determined by using the majority of the class of nearest neighbors; the class level of the test data is assigned. But if more than one class has the highest frequency, then tie-breaking algorithm can be used like the labeled class of the first nearest neighbor can be used as the predicted class [6].

8.2.2 *Fuzzy k-NN Classifier*

Fuzzy k-NN is an extension of the k-NN classifier. In k-NN algorithm, equal weightage is given to all the k-nearest neighbors to calculate the predicted class of a test data. Fuzzy k-NN algorithm assigns fuzzy membership of a test pattern in each class. That class is taken to be the predicted class (of that test pattern) for which the fuzzy-membership is maximum. Microarray gene expression data have a very high dimension and may contain thousands of genes. But number of classes present in the microarray gene expression data is usually low and sometimes these classes have overlapping nature. In these cases, the fuzzy k-NN algorithm can give a better result than k-NN [6].

8.3 Active Learning Using Fuzzy k-NN Classifier

The proposed active learning using fuzzy k-NN classifier (ALFKNN) consists of two steps. The first step uses “active learning” strategy. Objective of the active learning is to find out the “most informative” unlabeled samples to get the labels from the experts, so that the labeled “informative samples” can be iteratively added to the

training set to improve the prediction accuracy. In the second step, the test samples are predicted to assign a particular class based on the final enlarged training set (a small number of initial training patterns along with added informative samples with labels) obtained from the active learning phase.

Let, $L = \{\vec{l}_1, \vec{l}_2, \dots\}$ be the finite set of labeled data patterns distributed in C number of classes present in the dataset. The proposed algorithm is initially trained with small number of labeled data. The algorithm is iteratively tested with a finite pool of unlabeled set $U = \{\vec{u}_1, \vec{u}_2, \dots\}$ from which the most informative data/confusing samples $I^t = \{\vec{i}_1^t, \vec{i}_2^t, \dots, \vec{i}_p^t\}$ at iteration t are selected for labeling by the expert. The informative unlabeled samples along with their labels are added to the labeled/training data, and those informative samples are subtracted from the unlabeled set pool. This process is repeated until convergence. The block label view of the proposed active learning process is shown in Fig. 8.1 and the details are provided below.

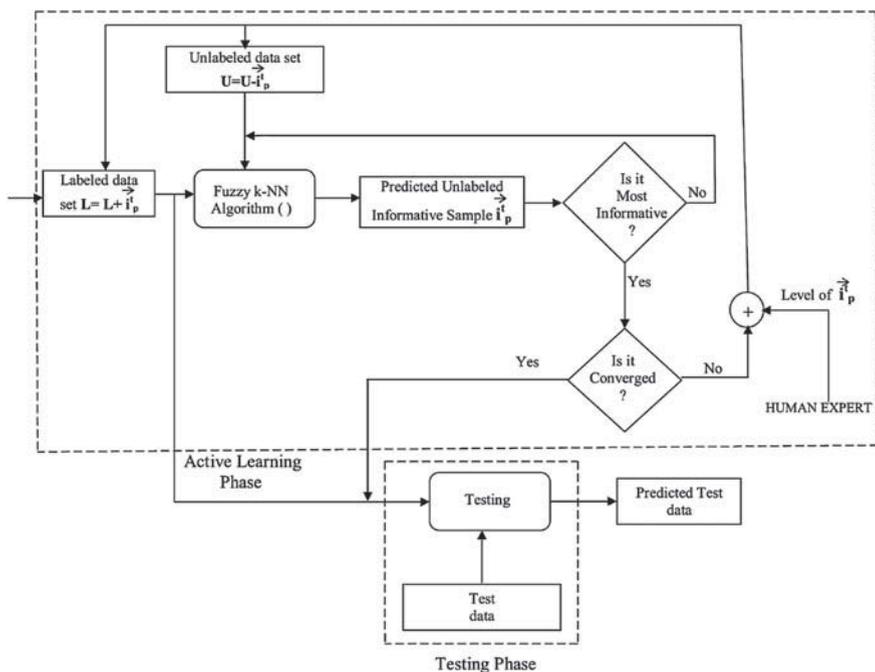


Fig. 8.1 Block diagram of the proposed active learning using fuzzy k-NN (ALFKNN) method

8.3.1 Active Learning Phase

Active learning phase of the proposed ALFKNN method is described in details here.

8.3.1.1 Calculate the Membership Degree of Training Pattern

Initially, the class center of each class is calculated based on the labeled patterns $\{\vec{l}_1, \vec{l}_2, \vec{l}_3, \dots\}$. Let, \vec{l}_i be the i th labeled pattern belonging to k th class (C_k), and $|C_k|$ be the number of labeled patterns in k th class (C_k). The class center (\vec{c}_k) of the k th class is computed using Eq. (8.1).

$$\vec{c}_k = \frac{1}{|C_k|} \sum_{\vec{l}_i \in C_k} \vec{l}_i; \quad \forall k = 1, 2, \dots, C. \quad (8.1)$$

Thereafter, the membership degree (μ_{ij}) of i th labeled pattern for belonging to j th class (C_j) is calculated as follows:

$$\mu_{ij} = \frac{1}{\sum_{k=1}^C \left(\frac{\|\vec{l}_i - \vec{c}_j\|}{\|\vec{l}_i - \vec{c}_k\|} \right)^{\frac{2}{m-1}}}; \quad \forall i, \quad \forall j = 1, 2, \dots, C; \quad (8.2)$$

where, $\|\vec{l}_i - \vec{c}_j\|$ is the distance of the i th labeled pattern (\vec{l}_i) from the class center (\vec{c}_j) of the j th class; $\|\vec{l}_i - \vec{c}_k\|$ is the distance of the i th labeled pattern (\vec{l}_i) from the class center (\vec{c}_k) of the k th class; and m is the fuzzifier ($1 < m < \infty$).

8.3.1.2 Calculate the Membership Degree of Unlabeled Pattern

The k nearest neighbors of each unlabeled pattern is calculated based on the distances of each labeled pattern (from that unlabeled pattern). Thereafter, the membership degree $\mu_c(\vec{u})$ of unlabeled pattern \vec{u} for belonging to each class c is calculated based on the k -nearest neighbors of that particular unlabeled pattern according to the following equation.

$$\mu_c(\vec{u}) = \frac{\sum_{j=1}^k \mu_{jc} \left(\frac{1}{\|\vec{u} - \vec{l}_j\|} \right)^{\frac{2}{m-1}}}{\sum_{j=1}^k \left(\frac{1}{\|\vec{u} - \vec{l}_j\|} \right)^{\frac{2}{m-1}}}; \quad \forall c = 1, 2, \dots, C, \quad \forall \vec{u}; \quad (8.3)$$

where, μ_{jc} is the membership degree of j th labeled pattern for belonging to c th class and $\|\vec{u} - \vec{l}_j\|$ is the distance of the j th labeled pattern from the unlabeled pattern \vec{u} whose membership value is to be calculated.

8.3.1.3 Find the Most Informative Unlabeled Pattern

The set of most informative/confusing unlabeled patterns $I^t = \{\vec{i}_1^t, \vec{i}_2^t, \dots\}$ in the t th iterations are selected based on the following. If the difference of the highest and the second highest membership degree of an unlabeled pattern is very less (than a predefined threshold value), then the unlabeled pattern is considered to be an informative pattern. That pattern is sent for class labeling by the expert. Thereafter, the labeled informative pattern is added to the existing labeled set L , and simultaneously the pattern is removed from the set of unlabeled pattern U . In this way, the labeled informative patterns which are added to the labeled set are now considered as the training patterns and thus the size of the training set is increased with the help of the unlabeled patterns.

8.3.1.4 Convergence of the Active Learning Phase

All the above three steps are repeated iteratively until convergence. The process converges when class centers are not changing significantly and no newer informative patterns are being added to the labeled set.

8.3.2 Testing Phase

After convergence, the final class centers ($\vec{c}_j; \forall j = 1, 2, \dots, C$) are computed based on the enlarged labeled set using Eq. (8.1), and the test patterns are tested to allocate the class labels. The membership degree $\mu_j(\vec{t})$ of a test pattern \vec{t} for belonging to j th class C_j is calculated based on the final class center as follows.

$$\mu_j(\vec{t}) = \frac{1}{\sum_{k=1}^C \left(\frac{\|\vec{t} - \vec{c}_j\|}{\|\vec{t} - \vec{c}_k\|} \right)^{\frac{2}{m-1}}}; \quad \forall \vec{t}, \quad \forall j = 1, 2, \dots, C; \quad (8.4)$$

where, $\|\vec{t} - \vec{c}_j\|$ is the distance of the test pattern (\vec{t}) from the class center of the j th class and $\|\vec{t} - \vec{c}_k\|$ is the distance of the test pattern (\vec{t}) from the class center (\vec{c}_k) of the k th class, and m is the fuzzifier ($1 < m < \infty$). Thereafter, the highest membership value of a test pattern (for belonging to a class) is obtained as follows:

$$class_label(\vec{t}) = \arg \max_j (\mu_j(\vec{t})); \quad \forall \vec{t}. \quad (8.5)$$

In this way the test pattern \vec{t} is assigned to that particular class for which the membership value is highest. The complete procedure of the proposed method is summarized in Algorithm 8.1.

Algorithm 8.1 : Active learning using fuzzy k-NN (ALFKNN)*Active learning phase*

- 1: Compute the class centers ($\vec{c}_k, \forall k = 1, 2, \dots, C$) based on the labeled patterns $\{\vec{l}_1, \vec{l}_2, \dots\}$ using Eq. (8.1).
- 2: Calculate the membership degree (μ_{ij}) of each labeled pattern (\vec{l}_i) for belonging to each class (\vec{c}_j) based on the class center (already computed in Step 1) using Eq. (8.2).
- 3: Compute the k-nearest neighbors of each unlabeled pattern \vec{u} based on the distances of each labeled pattern.
- 4: Calculate the membership degree ($\mu_c(\vec{u})$) of each unlabeled pattern \vec{u} for belonging to each class c based on the k-nearest neighbors using Eq. (8.3).
- 5: Compute the set of most informative/confusing patterns I' as described in Sect. 8.3.1
- 6: The informative/confusing patterns I' thus obtained from Step 5 is sent for class labeling by the expert.
- 7: The labeled informative patterns are added to the existing labeled set L and removed from the unlabeled set U .
- 8: Repeat Step 1 through Step 7 until converges (there are no informative/confusing points in the unlabeled set U).

Testing phase

- 1: Compute the final class centers for each class ($\vec{c}_k, \forall k = 1, 2, \dots, C$) using Eq. (8.1).
- 2: Calculate the membership degree ($\mu_j(\vec{l})$) of each test pattern \vec{l} for belonging to each class ($\vec{c}_j, \forall j = 1, 2, \dots, C$) based on the class center using Eq. (8.4).
- 3: Assign the class label for each test pattern \vec{l} using Eq. (8.5).

8.4 Experimental Evaluation

In this section, we provide the details of the experimental setup followed by the description of microarray gene expression datasets used and finally, analyze the experimental results.

8.4.1 Datasets Used

In this work, we have used six microarray gene expression datasets for the experiments. These are Brain Tumor dataset [7], Colon Cancer dataset [8], Leukemia dataset [9], Lymphoma dataset [10], Prostate Cancer dataset [11], and Small Round Blue-Cell Tumors (SRBCT) dataset [4]. The preprocessed form of these data are publicly available at www.stat.ethz.ch/dettling/bagboost.html. The datasets consist of the gene expression values and class information corresponding to each observation. In our experiments, we have taken very small amount of these data as labeled data and used their class information for classification. The remaining part of the set is used as test data. As the number of samples are very few for these microarray gene expression datasets, the entire test dataset is used as unlabeled data in our experiment. Though, any subset of the test data or data different from the test data may also be considered.

Brain tumor dataset [7] contains 42 microarray gene expression profiles. Each observation consists of 5597 genes. The data represents five classes of brain tumor. The five types of brain tumors are medulloblastomas, malignant gliomas, atypical teratoid/rhabdoid tumors, primitive neuroectodermal tumors, and human cerebella. Number of observations for these classes are 10, 10, 10, 8, and 4, respectively.

Colon cancer dataset [8] consists of expression levels of 62 (40 tumor and 22 normal) colon tissues for 6500 human genes. A selection of 2000 genes with highest minimal intensity across the samples has been made.

Leukemia dataset [9] consists of microarray gene expression data of 72 patients distributed in two classes namely, lymphoblastic leukemia and myeloid leukemia. Number of observations corresponding to lymphoblastic leukemia is 47 and 25 observations correspond to myeloid leukemia. Each observation contains expression values of 3571 genes.

Lymphoma dataset [10] is having a microarray gene expression data of 62 patients. Number of subtype classes present is three. First class corresponds to diffuse large B-cell lymphoma. The second class corresponds to follicular lymphoma, and the third type corresponds to chronic lymphocytic leukemia. In total, 4026 number of genes are present in the data.

Prostate cancer dataset [11] contains 102 samples. Total number of classes is 2. 52 observations are from prostate cancer tissues and 50 are from normal patients. Each observation contains expression values for 6033 genes.

Small round blue-cell tumors dataset [4] contains gene expression profiles for classifying small round blue-cell tumors of childhood (SRBCT) into four classes (neuroblastoma, rhabdomyosarcoma, non-Hodgkin lymphoma, Ewing family of tumors) and was obtained from cDNA microarrays. A training set comprising 63 SRBCT tissues, as well as a test set consisting of 20 SRBCT and 5 non-SRBCT samples are available. Each tissue sample is associated with a thoroughly preprocessed expression profile of 2308 genes, already standardized to zero mean and unit variance across genes. Here in this work, 63 SRBCT tissues are used as they have class labels (required for validation). The summary of the dataset used is provided in Table 8.1.

Table 8.1 Summary of the six microarray gene expression cancer datasets used for the experiments

Dataset	Subtype (classes)	Genes (dimension)	Samples (patterns)
Brain tumor	5	5597	42
Colon cancer	2	2000	62
Leukemia	2	3571	72
Lymphoma	3	4026	62
Prostate cancer	2	6033	102
SRBCT	4	2308	63

8.4.2 Experimental Results and Analysis

The average of 10 best simulations (taken from 20 simulations with 20 different labeled, unlabeled and test sets) of all the algorithms are reported to compare the performance of the proposed active learning using fuzzy k-NN (ALFKNN), with those of k-NN and fuzzy k-NN classifiers. All the algorithms used in this article are implemented in C language. Two separate sets of readings are measured (and summarized in Table 8.2) to test all the algorithms with three and five nearest neighbors (shown as $k = 3$ and $k = 5$ in Table 8.2) for all the datasets.

For brain cancer dataset, eight labeled patterns have been randomly selected such that they represent all the five classes. From the experimental results it is seen that in case of Brain cancer dataset the performance increase in accuracy obtained by proposed ALFKNN in comparison to other supervised counterpart fuzzy k-NN is 13.73 % (for $k = 3$), 6.87 % (for $k = 5$) and k-NN is 29.32 % (for $k = 3$) and 38.05 % (for $k = 5$), respectively. In case of colon cancer dataset, six patterns have been taken as labeled patterns selected randomly such that these labeled patterns represent all 2 classes. It is seen from the experimental outcome that the proposed ALFKNN algorithm (84.35 and 85.33 % accuracy) outperforms the other two algorithms k-NN

Table 8.2 Summary of the experimental results performed on six microarray gene expression cancer datasets

Data sets (trained with)	Methods	Accuracy (%)	
		$k = 3$	$k = 5$
Brain tumor (L = 8)	k-NN	47.94	34.41
	Fuzzy k-NN	63.53	65.59
	Proposed ALFKNN	77.26	72.46
Colon cancer (L = 6)	k-NN	68.04	63.75
	Fuzzy k-NN	78.57	74.99
	Proposed ALFKNN	84.35	85.33
Leukemia (L = 7)	k-NN	89.54	81.08
	Fuzzy k-NN	95.23	95.45
	Proposed ALFKNN	96.18	96.27
Lymphoma (L = 6)	k-NN	82.14	73.93
	Fuzzy k-NN	98.93	98.98
	Proposed ALFKNN	99.10	99.25
Prostate cancer (L = 10)	k-NN	68.70	65.43
	Fuzzy k-NN	65.76	63.15
	Proposed ALFKNN	73.19	72.60
SRBCT (L = 6)	k-NN	58.25	39.23
	Fuzzy k-NN	70.88	68.60
	Proposed ALFKNN	88.60	90.48

(for $k = 3$, 68.04 %, for $k = 5$, 63.75 %) and fuzzy k-NN (for $k = 3$, 78.57 %, for $k = 5$, 74.99 %) in terms of classification accuracy. Similarly, for leukemia dataset, seven patterns have been taken as labeled patterns selected randomly such that these labeled patterns represent all 2 classes. It is seen from the experimental outcome that the proposed ALFKNN algorithm outperforms the other two algorithms (k-NN and fuzzy k-NN) in terms of classification accuracy. Though the accuracy of k-NN (89.54 % and 81.08 % for $k = 3$ and $k = 5$) and that of fuzzy k-NN (95.23 and 95.45 % for $k = 3$ and $k = 5$) is quite high, still the accuracy of proposed ALFKNN method (96.18 and 96.27 % for $k = 3$ and $k = 5$) is even better than those two methods. From lymphoma dataset, six patterns have been randomly taken as labeled patterns such that these labeled patterns represent all three classes. It is seen from the experimental outcome that the proposed ALFKNN algorithm outperforms the other two algorithms (k-NN and fuzzy k-NN) in terms of classification accuracy. In this case, the prediction accuracy of the proposed method is as high as 99.1 and 99.25 % for $k = 3$ and $k = 5$, respectively. In prostate cancer dataset, 10 patterns have been taken as labeled patterns selected randomly ensuring representation from all 2 classes. It is seen from the Table 8.2 that the proposed ALFKNN algorithm produces significantly better accuracy (73.19 % for $k = 3$ and 72.6 % for $k = 5$) than the other two algorithms (k-NN 68.7 and 65.43 % and fuzzy k-NN 65.76 and 63.15 % for $k = 3$ and $k = 5$, respectively). For SRBCT (small round blue-cell tumors) dataset, six patterns have been taken randomly such that these labeled patterns represent all four classes. The result summary (Table 8.2) shows a high percentage of accuracy in case of ALFKNN algorithm (88.6 % for $k = 3$ and 90.48 % for $k = 5$) in comparison to other methods (k-NN 58.25 and 39.23 % and fuzzy k-NN 70.88 and 68.6 % for $k = 3$ and $k = 5$, respectively). In all the results, the number of labeled pattern is denoted by L in the Table 8.2.

In all the experiments, the value of fuzzifier (for fuzzy k-NN, and ALFKNN) m was fixed at 1.5. In summary, it is obvious that, for all the cases, our proposed method ALFKNN gives significantly better accuracy than k-NN and fuzzy k-NN.

8.5 Conclusions

Scarcity of the clinically labeled data in cancer classification from microarray gene expression data limits the use of traditional supervised classifiers due to their low prediction accuracies. In this context, the use of active learning method (to select only few most informative unlabeled sample's labels from biological expert) may be done to handle the above said problem in order to improve the cancer prediction accuracy. In this article, we propose a novel active learning method using fuzzy k-NN (ALFKNN) for cancer classification from microarray gene expression data. Active learning strategy used here handles the scarcity of training samples, whereas fuzzy technique takes care of the overlapping nature of the cancer subtype classes commonly present in the gene expression datasets. We have demonstrated the comparative results of the three algorithms, namely supervised k-NN, fuzzy k-NN, and

proposed active learning method (ALFKNN). Experimental results obtained from six cancer datasets suggest that proposed ALFKNN method significantly outperforms its existing supervised counterparts (k-NN and fuzzy k-NN). In near future, the proposed algorithm may be compared with other active learning-based methods on more microarray gene expression cancerous samples. Inspiring results of the proposed ALFKNN motivate us to deploy the active learning strategy to other classifiers.

References

1. Liu, Y.: Active learning with support vector machine applied to gene expression data for cancer classification. *J. Chem. Inf. Comput. Sci.* **44**(6), 1936–1941 (2004)
2. Keller, J.M., Gray, M.R., Givens, J.A.: A fuzzy k-nearest neighbor algorithm. *IEEE Trans. Syst., Man Cybern.* **15**(4), 580–585 (1985)
3. Gehrke, J., Ganti, V., Ramakrishnan, R., Loh, W.-Y.: BOAT-optimistic decision tree construction. *Proc. SIGMOD Rec.* **28**, 169–180 (1999)
4. Khan, J., et al.: Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nat. Med.* **7**(6), 673–679 (2001)
5. Keller, A.D., Schummer, M., Hood, L., Ruzzo, W.L.: Bayesian classification of DNA array expression data. Technical report UW-CSE-2000-08-01, University of Washington, Washington, DC (2000)
6. Halder, A., Misra, S.: Semi-supervised fuzzy K-NN for cancer classification from microarray gene expression data. In: *Proceedings of 1st International Conference on Automation, Control, Energy and Systems (ACES 2014)*, pp. 1–5. IEEE Computer Society Press (2014)
7. Pomeroy, S.L., et al.: Prediction of central nervous system embryonal tumour outcome based on gene expression. *Nature* **415**(6870), 436–442 (2002)
8. Alon, U., et al.: Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proc. Nat. Acad. Sci.* **96**, 6745–6750 (1999)
9. Golub, T.R., et al.: Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* **286**(5439), 531–537 (1999)
10. Alizadeh, A.A., et al.: Distinct types of diffuse large b-cell lymphoma identified by gene expression profiling. *Nature* **403**(6769), 503–511 (2000)
11. Singh, D., et al.: Gene expression correlates of clinical prostate cancer behavior. *Cancer Cell* **1**(2), 203–209 (2002)

Chapter 9

A Practical Under-Sampling Pattern for Compressed Sensing MRI

Bhabesh Deka and Sumit Datta

Abstract Typically, magnetic resonance (MR) images are stored in k-space where the higher energy samples, i.e., the samples with maximum information are concentrated near the center only; whereas, relatively lower energy samples are present near the outer periphery. Recently, variable density (VD) random under-sampling patterns have been increasingly popular and a topic of active research in compressed sensing (CS)-based MR image reconstruction. In this paper, we demonstrate a simple approach to design an efficient k-space under-sampling pattern, namely, the *VD Poisson Disk* (VD-PD) for sampling MR images in k-space and then implementing the same for CS-MRI reconstruction. Results are also compared with those obtained from some of the most prominent and commonly used sampling patterns, including the *VD random with estimated PDF* (VD-PDF), the *VD Gaussian density* (VD-Gaus), the *VD uniform random* (VD-Rnd), and the *Radial Type* in the CS-MRI literature.

Keywords Magnetic resonance imaging · k-space · Compressed sensing · Variable density under-sampling pattern

9.1 Introduction

According to the compressed sensing (CS) theory, we can reconstruct a signal or an image from just a few samples or measurements than that required by the Shannon–Nyquist sampling theorem. For perfect reconstruction, a signal must be *sparse* over a selected basis set, called the *representation bases* which must be *incoherent* with the *measurement bases* used to acquire the signal. Magnetic resonance imaging (MRI) is the most commonly used imaging technology for imaging soft-tissues and

B. Deka (✉) · S. Datta
Tezpur University, Tezpur 784028, India
e-mail: bdeka@tezu.ernet.in

S. Datta
e-mail: sumitdatta96@yahoo.in

organs present in the human body. This imaging technique also implicitly fulfills the key requirements for image reconstruction using the CS technique [1]. The MRI data are stored in the k-space or frequency domain containing the spatial frequency information of the target in two dimensions. Since in k-space, the center region (*zero frequency region*) contains more information than that of the periphery or outer region, more number of samples are usually collected from the center.

Nayak and Nishimura [2] have experimentally demonstrated that aliasing due to under-sampling in k-space can be minimized if the point spread function (PSF) possesses a very large approximate zero region around its main lobe, and becomes very similar to the white noise elsewhere. They also reported that the *poisson disk distribution* satisfies the requirements of an ideal sampling pattern as mentioned above.

However, the MRI data in k-space must be collected densely near the center only and sparsely near the outer periphery. Therefore, an uniform random sampling pattern would not be quite effective for sampling in k-space. In [1], the authors have introduced the concept of variable density (VD) for more effective and intuitive sampling of MRI data in k-space.

In this paper, we have explored the design of a most promising class of sampling patterns, namely, the *variable density poisson disk* (VD-PD) for sampling MRI data in k-space. It satisfies the requirements of both the optimal and the intuitive sampling patterns as discussed above. Figure 9.1a, c shows that the PSF of variable density

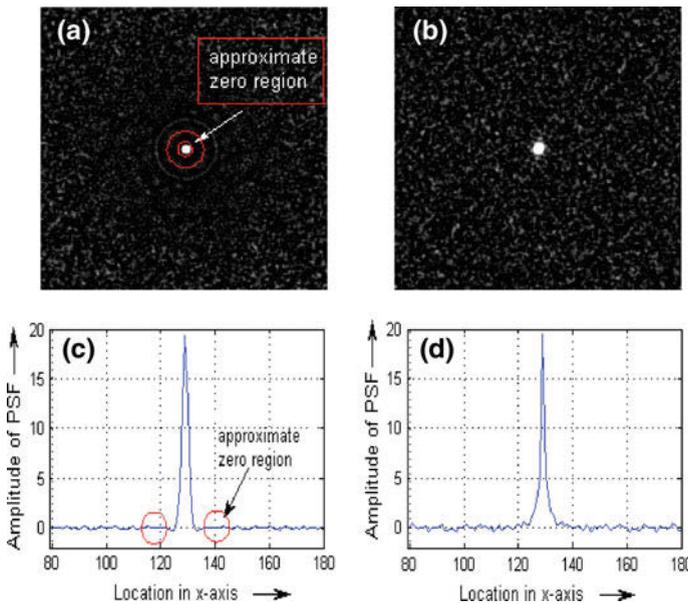


Fig. 9.1 2D PSFs for **a** the VD-PD and **b** the VD-PDF. **c** The middle row of the 2D VD-PD and **d** the middle row of the 2D VD-PDF

poisson disk have a large approximate zero region around its main lobe and becomes similar to the white noise when observed away from the main lobe. From these figures, we can conclude that the corresponding sampling patterns would also very likely to be incoherent with respect to the representation bases such as the wavelets. The 2D k-space data is the result of the simultaneous application of phase and frequency encoding gradient pulses along with the RF pulse in an MRI system. It is also noted that a practical sampling pattern for under-sampling k-space data is the one which would give the lowest gradient pulse variation for a *particular trajectory*, i.e., the path taken for collecting the most effective samples in the k-space. This is because of both the hardware (such as gradient amplitude, slew rate, etc.) and physiological (nerve stimulation) limitations in a typical MR imaging system. Figure 9.2a–d show the X-gradient variations across the different phase encoding lines for various k-space

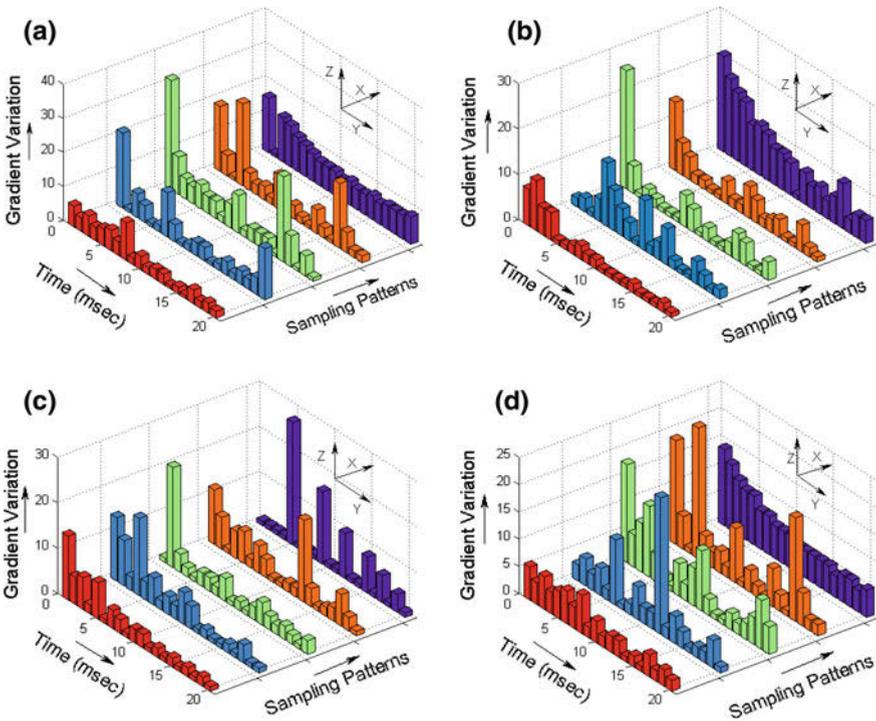


Fig. 9.2 X-gradient pulses variation for different under-sampling patterns. **a** The gradient variation to acquire the 50th phase encoding line of the k-space, **b** the gradient variation to acquire the 100th phase encoding line of the k-space, **c** the gradient variation to acquire the 150th phase encoding line of the k-space, and **d** the gradient variation to acquire the 200th phase encoding line of the k-space [4]. From *left to right* of the X-axis: the variable density poisson disk (VD-PD) under-sampling pattern, the variable density estimated PDF (VD-PDF) under-sampling pattern, the variable density gaussian (VD-Gaus) under-sampling pattern, the variable density random (VD-Rnd) under-sampling pattern, and the radial under-sampling pattern, respectively

under-sampling patterns, namely, the VD random (VD-RND), the VD Gaussian (VD-Gaus), the VD random with estimated PDF (VD-PDF), the radial type (Radial), and the VD poisson disk (VD-PD). It is noted that the VD-PD gives the least gradient pulse variation irrespective of the locations of the samples in k-space compared to other sampling patterns as mentioned above.

We aim here to design an optimal under-sampling pattern which would not only be efficient but also yield the least incoherent aliasing compared to other commonly used sampling patterns for the reconstruction of MR images using the CS. The results are detailed in Sect. 9.5.

9.2 Main Contributions

1. Investigated gradient variations for different random under-sampling patterns, namely, the Radial, the VD-Rnd, the VD-GD, and the VD-PDF for designing an efficient random under-sampling pattern for the CS-MRI.
2. Proposed a variable density poisson disk (VD-PD) under-sampling pattern and also computed the gradient variations for three different k-space trajectories [3], namely the *line*, the *spiral*, and the *zigzag* trajectories *to find the best trajectory* for the proposed random under-sampling pattern.

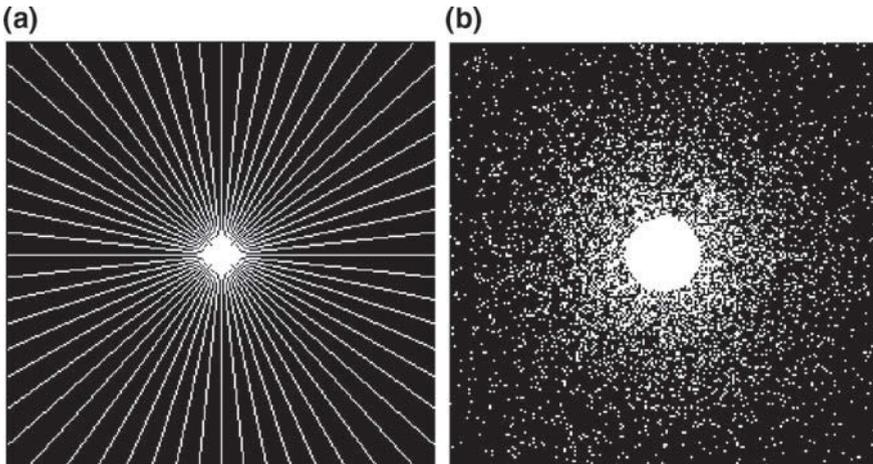


Fig. 9.3 Some well-known under-sampling patterns. **a** The 256×256 radial type with 10% sampling ratio. **b** The 256×256 VD-PDF with 10% sampling ratio

3. Evaluated CS reconstruction quality with the proposed under-sampling pattern by measuring aliasing present in the reconstructed output and compared to some of the most widely used under-sampling strategies in the CS-MRI (Fig. 9.3).

9.3 Practical Under-Sampling Pattern for the CS-MRI

The proposed VD-PD sampling pattern is shown in Fig. 9.4a for 10% sampling ratio. In this pattern, N random points are generated in such a way that no two neighboring points are too close and always maintain a minimum distance (MD) between them. For effective k-space sampling, we consider here the variable density poisson disk sampling. To implement variable density sampling, an ensemble of concentric poisson disks with varying MD are considered. A poisson disk at the center would have the *smallest* MD (SMD) which may be fixed as follows. For example, considering a rectangular grid representation for the image and each unit square denoting a pixel, the SMD for generating a poisson disk at the center would be only 0.5. The variation of MD from the center to the periphery is approximately exponential as shown in Fig. 9.4b.

It is further mentioned here that the minimum distance for the outer disks also varies inversely with the sampling ratio, i.e., the ratio of the number of selected sample points to the total number of samples present in the k-space data of an MR image. For simulation of poisson disk sampling, we use the MATLAB codes publicly available at [5]. As mentioned earlier, it is experimentally found that a practical under-sampling pattern should have the lowest gradient pulse variation or gradient load [3].

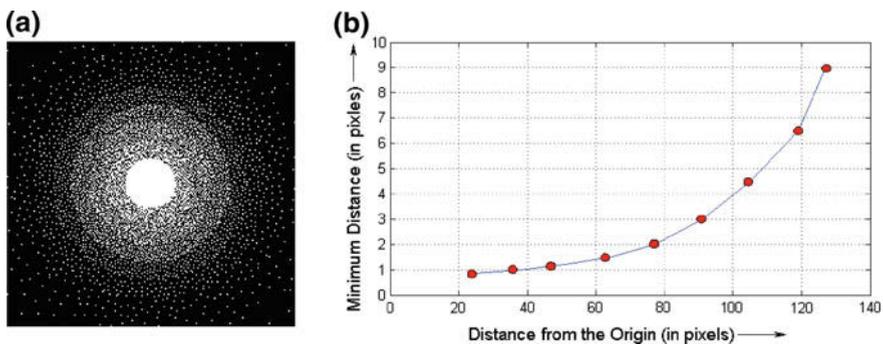


Fig. 9.4 Proposed under-sampling pattern. **a** The 256×256 VD-PD with 10% sampling ratio. **b** The variation of the “Minimum distance” in poisson disks with the increasing distance from the center

9.4 Computation of Gradient Load

The gradient load E is the normalized vector sum of the applied gradients G^x and G^y , in X and Y directions, respectively, in an MRI data acquisition system. It is defined by:

$$E = \frac{1}{NG_o^2} \sum_{l=1}^N [(G_l^x)^2 + (G_l^y)^2] \quad (9.1)$$

where G_o is the magnitude of gradient step and N is the number of selected samples in k-space [3].

The k-space under-sampling pattern with the lowest gradient load would not only give the lowest gradient pulse variation but also minimizes the limitations imposed by physiological constraints, such as the *peripheral nerve stimulation* due to high gradient amplitudes and slew rates. We have experimentally verified that the VD-PD sampling pattern is an efficient k-space under-sampling pattern with the lowest gradient load compared to other sampling patterns, namely, the VD-Rnd, the VD-Gaus, Radial, and the VD-PDF. The experimental results are detailed in the next section.

9.5 Experimental Results

In the first set of experiments, we calculated the penalty function or gradient load with the help of Eq.(9.1). For simplicity of the calculation, we take the magnitude of gradient step $G_o = 1$ and the duration of the gradient step $\Delta t = 1$. The resultant gradient loads for various under-sampling patterns, namely, the VD-Rnd, the VD-Gaus, the VD-PDF, Radial, and the proposed VD-PD are shown in Fig.9.5. It is noted that the VD-PD have the least gradient load amongst all other under-sampling patterns. We have also considered three different trajectories (shown in Fig.9.6), namely, the line, the spiral, and the zigzag trajectories to see the influence of k-space trajectory on the gradient load. Although, the intergradient pulse variations for the proposed sampling pattern are different for different k-space trajectories, the intragradient variations show that the proposed VD-PD pattern has the lowest gradient load or variations irrespective of the type of trajectory used.

In the next set of experiments, the performance of the proposed under-sampling pattern is evaluated for reconstruction of MR image using the CS. We use different ℓ^1 optimization algorithms such as the primal-dual interior-point method (PDIPM) [6], the truncated newton interior-point method (TNIPM) [7], the iterative shrinkage-thresholding (IST) [8], the nonlinear conjugate gradient method (NCGM) [1], the gradient projection for sparse reconstruction (GPSR) [9], and the two-step IST (TwIST) [10] to solve the CS-MRI problem.

The PDIPM solves the problem in a classical way as the result requires only $O(n^3)$ operations at each iteration. But in the TNIP method, a preconditioner is used

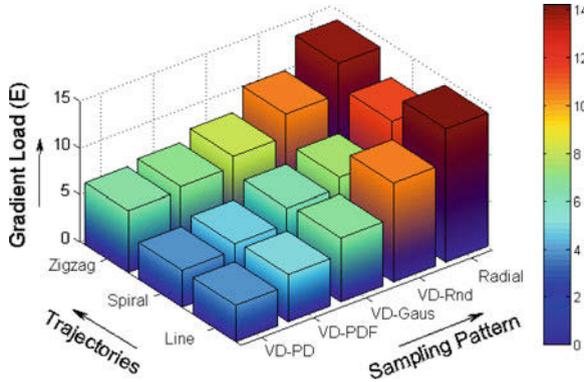


Fig. 9.5 Penalty function E for the proposed under-sampling patterns with different trajectories, namely, the line, the spiral, and the zigzag trajectories, respectively

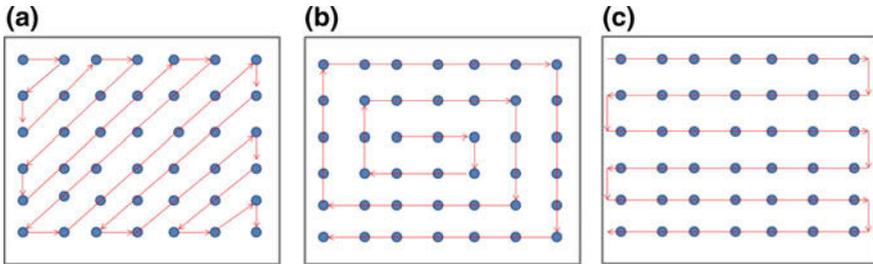


Fig. 9.6 Different k-space trajectories. **a** The Zigzag trajectory, **b** the Spiral trajectory, and **c** the Line trajectory [3]

to reduce the number of operations. From the Fig. 9.7, we can say that the required CPU time (in seconds) for the TNIPM and the IST are continuously reduced with increasing sampling ratios, but in case of the nonlinear conjugate gradient algorithm, the required CPU time is continuously increased with increasing sampling ratios. In case of the TwIST, we have a different scenario. That is, the required CPU time does not vary with sampling ratios. Therefore, from the above analysis, we can conclude that the TwIST is the fastest algorithm. We choose the TwIST for CS-MRI reconstruction. Figures 9.8 and 9.9 show the Brain MR images, the corresponding CS reconstruction results with 20 % sampling ratio for different under-sampling patterns. From the results, it is seen that the proposed VD-PD gives better reconstruction results in terms of higher contrast and better preservation of edges. Also, the proposed method gives very less visual aliasing, meaning that the Fourier under-sampling with the proposed sampling pattern is significantly incoherent with respect to the wavelet basis. We also compute the PSNR (peak signal-to-noise ratio) to compare the CS reconstruction results quantitatively for different sampling patterns. The results are shown in Fig. 9.10. It is clearly observed the proposed method gives higher PSNR

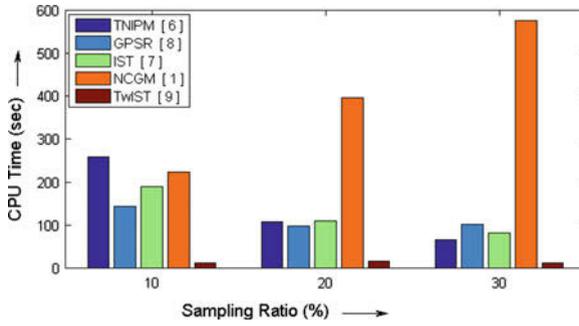


Fig. 9.7 The variation of CPU time for different algorithms with increasing sampling ratio

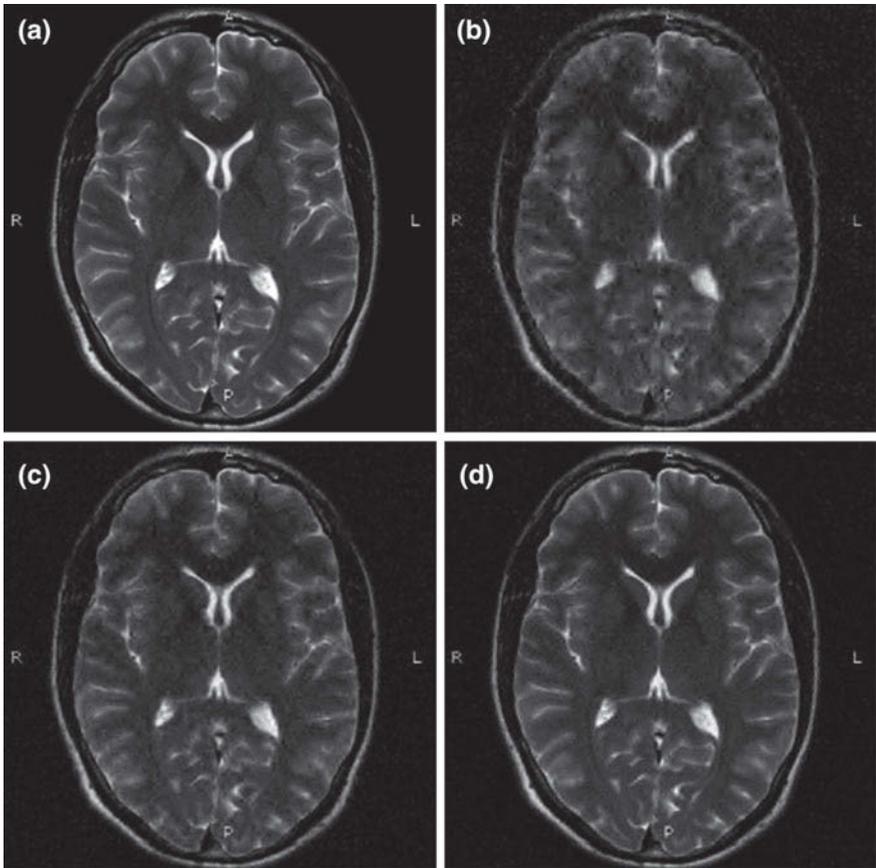


Fig. 9.8 a Original brain MR image, b the reconstructed image by radial with 20 % sampling ratio, c the reconstructed image by VD-PD with 20 % sampling ratio, and d the reconstructed image by VD-PDF with 20 % sampling ratio

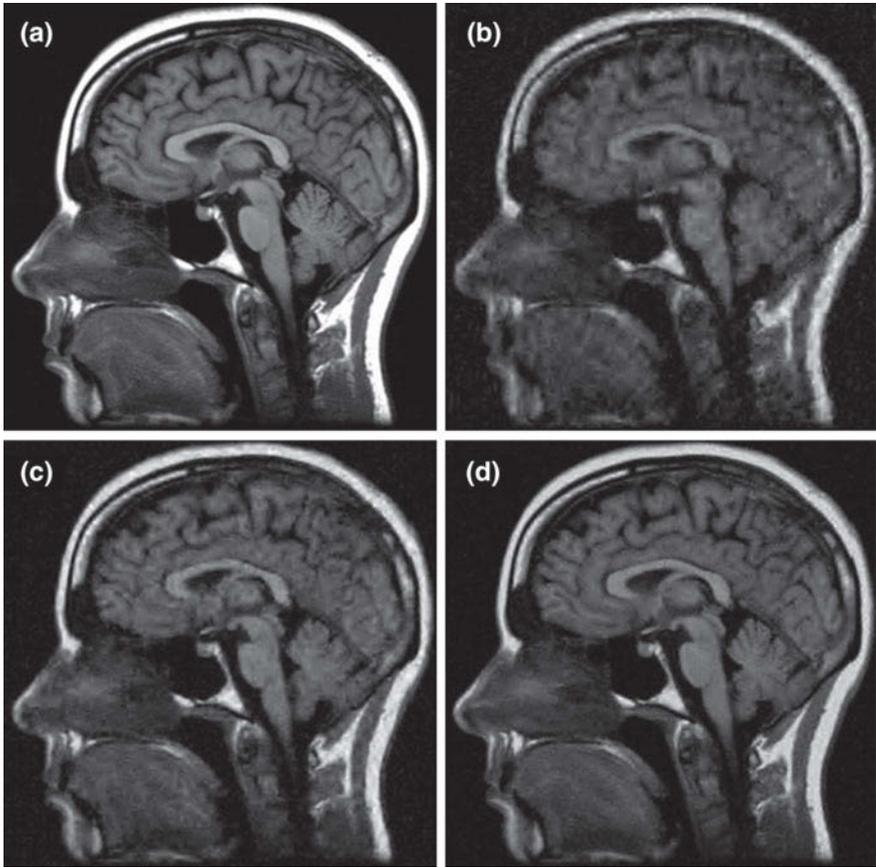


Fig. 9.9 **a** Original brain MR image, **b** the reconstructed image by radial with 20 % sampling ratio, **c** the reconstructed image by VD-PD with 20 % sampling ratio, and **d** the reconstructed image by VD-PDF with 20 % sampling ratio

values for different sampling ratios. We also measure the mean structural similarity (MSSIM) index for the reconstructed MR images. Figure 9.11 shows that the proposed sampling pattern has the best MSSIM values compared to other sampling patterns for different sampling ratios.

Fig. 9.10 **a** The changes of MSE with increasing sampling ratio for different types of sampling patterns, and **b** The changes of PSNR (in dB) with increasing sampling ratio for different types of sampling patterns

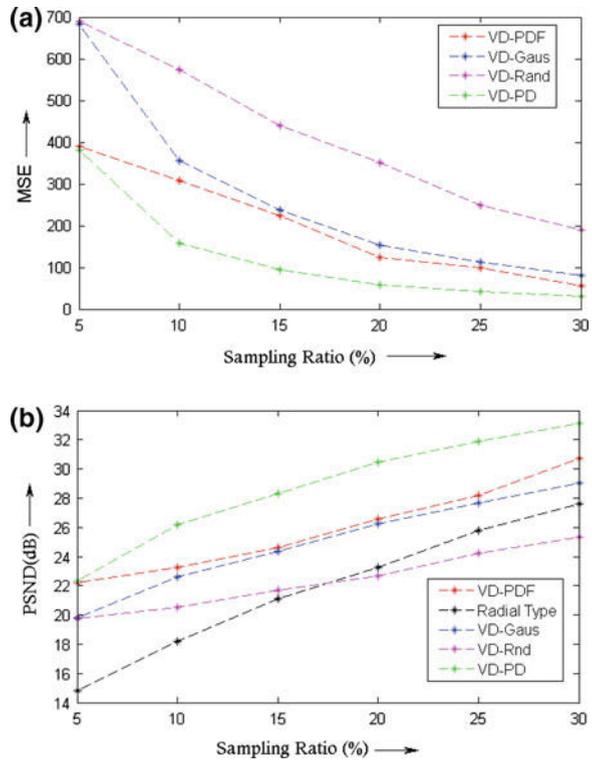
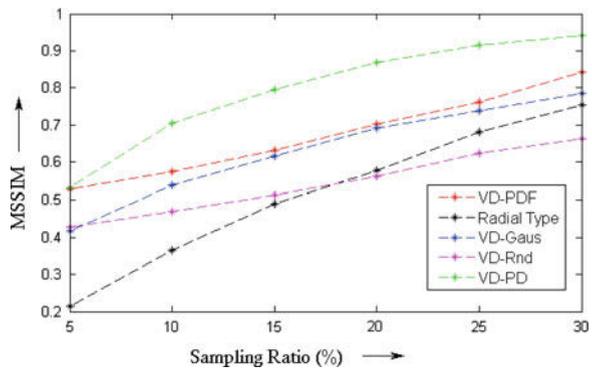


Fig. 9.11 The changes of MSSIM with increasing sampling ratio for different types of sampling patterns



9.6 Conclusion

We have showed that the proposed under-sampling pattern, VD-PD, not only gives the least gradient pulse variation or gradient load among different under-sampling patterns but also yields better CS reconstruction of MR images with visually no aliasing artifacts even at sampling ratios as low as 20 % when compared to one of the most widely studied reconstruction technique in the CS-MRI. The proposed under-sampling pattern may be used to design an efficient and practical under-sampling hardware for the CS-MRI.

Acknowledgments The authors want to thank UGC for the financial support of the above project No. 41-603/2012 (SR) dated 16th July, 2012 and the Department of ECE, Tezpur University for giving us the opportunity and necessary infrastructure to carry out the project work.

References

1. Lustig, M., Donoho, D., Pauly, J.M.: Sparse MRI: the application of compressed sensing for rapid MR imaging. *Magn. Reson. Med.* **58**, 1182–1195 (2007)
2. Nayak, K.S., Nishimura, D.G.: Randomized trajectories for reduced aliasing artifact. In: *Proceedings of the Scientific Meeting and Exhibition of ISMRM*. Sydney (1998)
3. Sersa, I., Macura, S.: Excitation of arbitrary shapes by gradient optimized random walk in discrete k-space. *Magn. Reson. Med.* **37**, 920–931 (1997)
4. Usman, M., Batchelor, P.G.: Optimized sampling patterns for practical compressed MRI. In: *International Conference on Sampling Theory and Applications (SAMPTA'09)*, Marseille (2009)
5. Cusack, R.: `Generate_Poisson_2d.m`. https://www.github.com/gpeyre/2014-JMIV-SlicedTransport/blob/master/sliced/generate_poisson_2d.m (2013)
6. Boyd, S., Vandenberghe, L.: *Convex Optimization*. Cambridge University Press (2004)
7. Koh, K., Kim, S.J., Boyd, S., Lin, Yi.: An interior-point method for large-scale l_1 -regularized logistic regression. *J. Mach. Learn. Res.* **8**(22), 1519–1555 (2007)
8. Daubechies, I., Defrise, M., De Mol, C.: An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Commun. Pure Appl. Math.* **57**(11), 1413–1457 (2004)
9. Figueiredo, M.A.T., Nowak, R.D., Wright, S.J.: Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems. *IEEE J. Sel. Top. Signal Process.* **1**(4), 586–597 (2007)
10. Bioucas-Dias, J.M., Figueiredo, M.A.T.: A new TwIST: two-step iterative shrinkage/thresholding algorithms for image restoration. *IEEE Trans. Image Process.* **16**(12), 2992–3004 (2007)

Chapter 10

Semi-automatic Segmentation and Marking of Pitch Contours for Prosodic Analysis

Biswajit Dev Sarma, Meghamallika Sarma and S.R.M. Prasanna

Abstract Prosody is used in both phonetics and speech synthesis systems in the literature. Pitch information is one of the extensively used prosodic information. This paper aims at semiautomatizing the process of pitch marking for prosodic analysis. Prosody is a suprasegmental information; therefore, it may be wiser to analyze the change in pitch over a segment of voiced speech instead of directly using the pitch calculated from a small window. In a particular voiced segment of speech, pitch may vary from low-to- high, high-to-low, or it may not vary at all. This work describes a method for automatically segmenting speech into certain regions having a continuous pitch contour and marking the nature of pitch change within those regions. Zero frequency filtering is used to segment the speech into voiced and unvoiced segments. This segment is further divided into small segments depending on a discontinuity present in the pitch contour. A height value of the pitch contour in the final segment is measured and accordingly marking is done. Now, automatic segmentation and markings are manually corrected by deleting, inserting, or shifting the segmentation boundaries and substituting the wrong markings. Automatic process is evaluated in terms of these four parameters.

Keywords Pitch contour · Segmentation · Marking · Manual error correction · Voiced region

10.1 Introduction

In phonetics, the use of pitch, loudness, tempo, and rhythm in speech to convey information about the structure and meaning of an utterance is called prosody. These aspects of an utterance can span regions larger than a single phonemic segment

B.D. Sarma (✉) · M. Sarma · S.R.M. Prasanna
Indian Institute of Technology, Guwahati 781039, India
e-mail: s.biswajit@iitg.ernet.in

M. Sarma
e-mail: megha.mallika@iitg.ernet.in

S.R.M. Prasanna
e-mail: prasanna@iitg.ernet.in

and hence are sometimes called suprasegmental [1]. These prosodic aspects of a language can convey distinctive semantic and syntactic facts. These can be useful for different tasks such as mapping prominence and grouping patterns to meaning differences, understanding the effects of prominence, and grouping on the pronunciation of words, synthesizing prosodically natural-sounding speech etc. [2]. Change of pitch in intonational phrases can bear different meanings. A high-to-low pitch can bear declarative meaning, while a low-to-high pitch can bear interrogative meaning. Speech scientists are interested in annotating the prosodic structure of large numbers of utterances in order to study these phenomena.

The ToBI annotation system (Tones and break) is a popular system used to annotate pitch and boundary information [1]. This system attempts to convert the continuous information in the speech signal present in the form of pitch, duration, loudness, breaks etc. into a set of categories, which may have some interpretation in the context. A ToBI transcription of an utterance at least consists of a recording of the speech, its fundamental frequency contour, and symbolic labels for other prosodic events. The symbolic labels are mainly transcription of tonal events, transcription of words, transcription of boundaries between words (called break index), and transcription of additional observations. Manual labeling is always a time-consuming process and effort had been made for automatizing the manual labeling of ToBI transcriptions [3].

We are developing a similar system for capturing prosodic information which includes syllable boundary marking, pitch contour segmentation and marking, and break index marking along with phonetic transcription using International Phonetic Alphabets. The objective is to prepare a database for Assamese language with the above-mentioned four-level transcriptions. In the manual transcription process, the transcriber listens to the speech signal and with the help of the waveform, spectrogram, pitch contour etc., tries to transcribe all the four levels of transcription. In [4], we developed a semiautomatic method for marking the syllable boundaries. In this work, we focus on semiautomatizing the process of pitch contour marking. In particular, the final objective is to segment the speech signal into some regions having continuous pitch contours and then mark the nature of the pitch contour in that region. Figure 10.1 describes the process. Figure 10.1a shows the speech signal and Fig. 10.2b shows the pitch contour along with the manually marked pitch contour boundaries (arrows) and markings (LH, HL etc.). In the manual process, the transcriber will listen to the speech and look into the pitch contour, and if a perceivable pitch change within a voiced segment of speech is noticed, the pitch change will be marked. In the figure, LH means a pitch transition from low-to-high, HL means pitch transition from high-to-low etc. (Different symbols for marking will be discussed in the subsequent sections). Semiautomatizing this process involves automatizing the process up to some extent and reducing the human labor.

The task is to segment the speech signal into some regions having a continuous pitch contour and then mark the regions according to the nature of the pitch contour in it. The segmentation task is carried out in two steps, first voiced–nonvoiced classification, then segmentation of voiced regions according to continuous pitch contours. There is no pitch in case of the unvoiced speech, silence, and background noise; so, they are first separated from the voiced regions. One voiced region itself can have

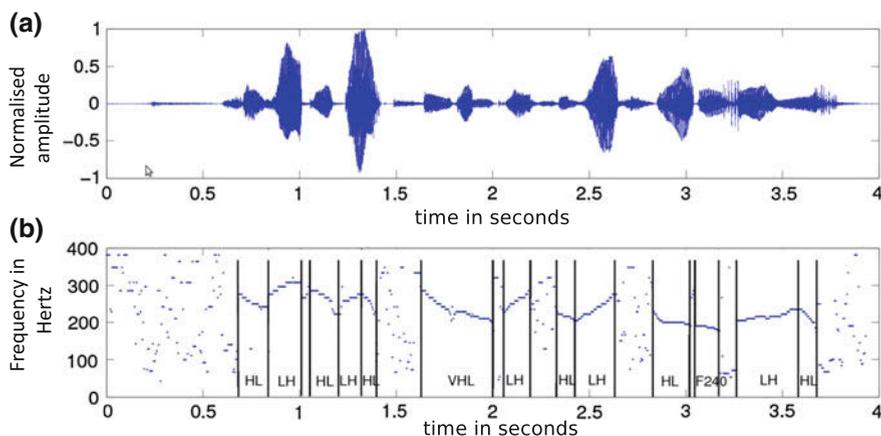


Fig. 10.1 Pitch contour segmentation and marking process. Plot **a** shows Speech signal and plot **b** shows Pitch contour with segmentation boundaries (*solid lines*) and markings

several pitch contours. There may be discontinuities in the pitch contour in a particular voiced region. Also there may be significant variation of pitch within a voiced region which may lead to possibility of more than one pitch contour marking for that region. Therefore, a voiced region is further segmented into some regions having a continuous and flat or rising or falling pitch contours. Then the segmented regions are marked according to the degree of change of pitch contour. If the contour is flat, then FR is used for marking where F tends for flat and R is the average pitch value. A rising pitch is marked by LH (low-to-high) or VLH (very low-to-high) or LVH (low-to-very high) according to the rising slope and position of the pitch contour edges. Similarly, a falling pitch is marked by HL (high-to-low) or VHL (very high-to-low) or HVL (high-to-very low). These automatic segmentations and markings are then corrected manually to get a complete error-free marking. This is done by deleting, inserting, or shifting the automatically segmented boundaries followed by a substitution of pitch contour marking (if necessary).

The whole procedure is described in the following sections. Section 10.2 describes voiced region detection. Section 10.3 describes segmentation of voiced speech into small segments. Automatic marking of pitch contour is described in Sect. 10.4. Manual correction of automatic segmentation and marking procedure is described in Sect. 10.5. Automatic process evaluation and discussion is done in Sect. 10.6. Section 10.7 summarizes the work and concludes.

10.2 Detection of Voiced Speech Regions

Voiced speech segment possesses significant glottal activity. Detection of voiced speech involves detection of the significant glottal activity in the speech. Rest of the regions include unvoiced speech, silence as well as background noise. Zero frequency filtering (ZFF) of speech is one of the best method to estimate the voiced epochs [5]. The effect of impulse like excitation of the glottal source is felt in all frequencies including the zero frequency. ZFF output contains energy around the zero frequency which is mainly the impulses due to excitation. The ZFFS are robust to various degradations, since most of the frequency components are attenuated. The ZFFS can be computed from the speech signal as follows [5]:

- Difference the speech signal $s(n)$

$$x(n) = s(n) - s(n - 1) \quad (10.1)$$

- Compute the output of a cascade of two ideal digital resonators at 0Hz

$$y(n) = - \sum_{k=1}^4 a_k y(n - k) + x(n) \quad (10.2)$$

where $a_1 = 4, a_2 = -6, a_3 = 4, a_4 = -1$

- Remove the trend i.e.,

$$\hat{y}(n) = y(n) - \bar{y}(n) \quad (10.3)$$

where $\bar{y}(n) = (1/(2N + 1)) \sum_{n=-N}^N y(n)$ and $2N + 1$ corresponds to the average pitch period computed over a longer segment of speech.

- The trend removed signal $\hat{y}(n)$ is the ZFFS.

The positive zero crossings of the ZFFS will give the location of epochs [5]. The first-order difference of the ZFFS can be treated as strength of excitation at the epochs [6]. A procedure for detection of voiced region using ZFF is described in [7]. The key idea is that addition of a small amount of noise to the speech signal does not affect the zero crossings of the ZF filtered signal in the voiced region, whereas it leads to zero crossings at random locations in all other regions. The epochs in voiced speech are produced due to glottal closure which impart the most significant impulse like excitation to the vocal tract system. These high SNR epochs are robust to noise and can be located using ZFF with high precision and accuracy even in presence of degradation. Unvoiced and silence region does not contain any significant excitation and result in zero crossings located at random instants. These locations can easily get affected by the addition of even a small amount of noise. Additional spurious epochs are eliminated using the knowledge of pitch and jitter information. The procedure of detection of voiced speech region is as follows:

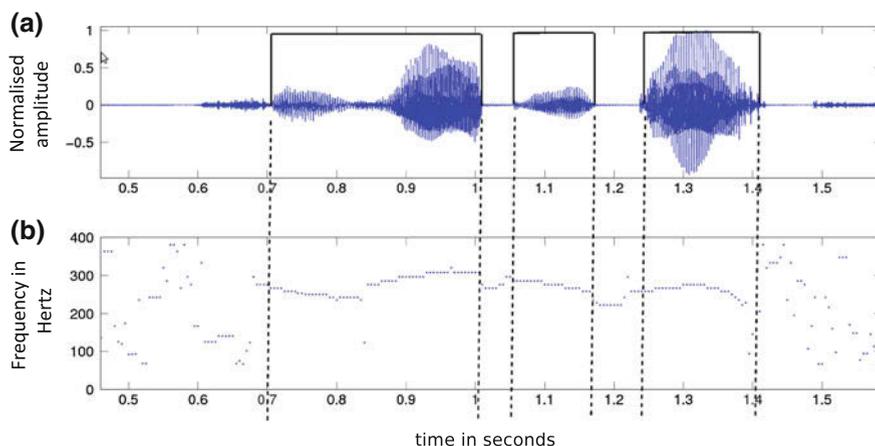


Fig. 10.2 Detection of voiced speech regions. Plot **a** shows Speech signal with detected voiced region (*solid lines*) and plot **b** shows pitch contour with voiced region boundaries (*dashed lines*)

- Extract the epoch locations of clean speech using ZFF.
- Add 20db of white gaussian noise to the speech signal.
- Extract the epoch locations of noisy speech using ZFF.
- If epoch difference between the locations obtained in the two cases are less than 2 ms, then retain it; otherwise, discard the epoch.
- Calculate the instantaneous pitch period using the epoch locations and eliminate the epoch if its pitch period is more than 15 ms.
- Calculate the instantaneous jitter and eliminate the epoch if jitter is more than 1 ms.

Figure 10.2 shows speech signal along with voiced region detected by the above method.

10.3 Segmentation of Voiced Speech into Small Segments

Final objective is to mark the regions with continuous pitch contour. There are several well-known methods available in the literature for determining pitch. Autocorrelation method is one of the classical methods [8]. Zero frequency filtering-based method is good for instantaneous pitch determination [9]. Here, we are interested to get a smooth pitch contour for analysis, so autocorrelation method is used to compute average pitch using a 20 ms frame size with 5 ms frame shift. Within a voiced segment, there can be discontinuity in the pitch contour. Whenever there exists a significant pitch variation in the voiced segment, more than one pitch marking is possible in the same voiced segment. The discontinuity is calculated by taking difference of the pitch contour. Difference of the pitch contour will have high value at the point of

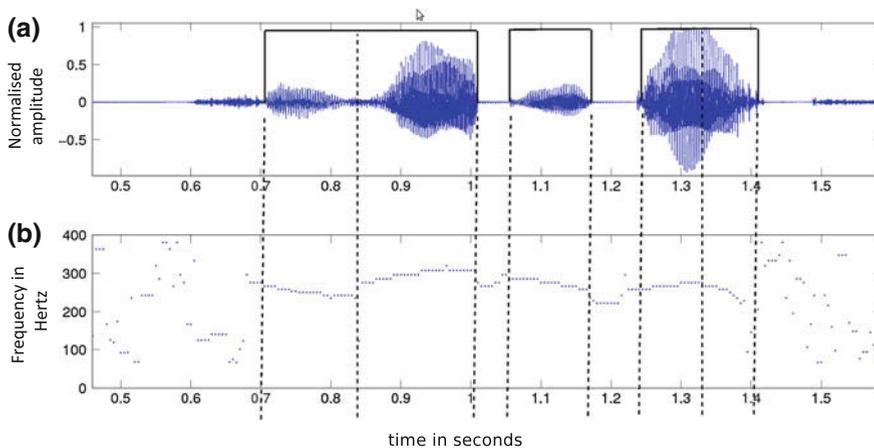


Fig. 10.3 Segmentation of voiced speech regions (*solid lines*) into small segments. Segmentation is shown in *dashed lines* with Speech signal in plot (a) and Pitch contour in plot (b)

discontinuity. A threshold is fixed to detect the point of discontinuity and marked a segment boundary at the discontinuity. Figure 10.3 describes segmentation of one voice segment to several segments. For example, in the first voiced region (*solid line*), the pitch contour is discontinuous. This discontinuity is detected and declared as a pitch contour boundary as shown by the dashed line.

10.4 Pitch Contour Marking

Once the segments are obtained, they are marked with one of the following marking: HL (high-to-low), VHL (very high-to-low), HVL (high-to-very low), LH (low-to-high), VLH (very low-to-high), LVH (low-to-very high), FR (F means flat and R is the average value of the pitch in that segment). One simple way to decide whether a contour is rising or falling can be to see the difference between average value of first and last few samples. But sometimes, spurious detection of pitch values especially at the edges, may lead to incorrect decision. That is why the pitch values are fitted with a line. Fitting is done using linear regression which estimates the coefficients (β_1, β_2) of a polynomial of degree one that fits the values best in a least-squares sense. The estimated line is

$$\hat{y} = \beta_1 x + \beta_2 \quad (10.4)$$

where, x is a vector = $(1, 2, \dots, N)$ and N is the length of the segment. Now the marking is obtained based on the measure of the height value. The height (\hat{h}) is calculated by taking the difference between the first and the last sample in the fitted line.

$$\hat{h} = \hat{y}(1) - \hat{y}(N) \quad (10.5)$$

Now if the height value is more than 20, then it is decided that the pitch contour is rising and if it is less than -20 , then it is a falling pitch contour. If height is more than 100 or less than -100 , then the pitch contour can have a very high or very low pitch value at the edge. Which edge will have the high or low value is determined by finding the closeness of the pitch values at the edges to the average pitch value (APV) of the speech utterance. If height is in between -20 and 20, the contour is decided to be flat. For a flat pitch contour, the average pitch value is calculated and written along with the flat marking. Based on these conditions, the pitch contours are marked as one of the following: HL, VHL, HVL, LH, VLH, LVH, and FR. The pseudocode for marking the pitch contour in a given segment based on the conditions described below.

```

If  $\hat{h} > 20$ 
then
  if  $\hat{h} > 100$ 
  then
    if  $(|\hat{y}(1) - APV| - |APV - \hat{y}(N)|) > 0$ 
    then mark VHL
    else mark HVL
  else mark HL
elseif  $\hat{h} < -20$ 
then
  if  $\hat{h} < -100$ 
  then
    if  $(|\hat{y}(1) - APV| - |APV - \hat{y}(N)|) > 0$ 
    then mark VLH
    else mark LVH
  else mark LH
else mark FR
    
```

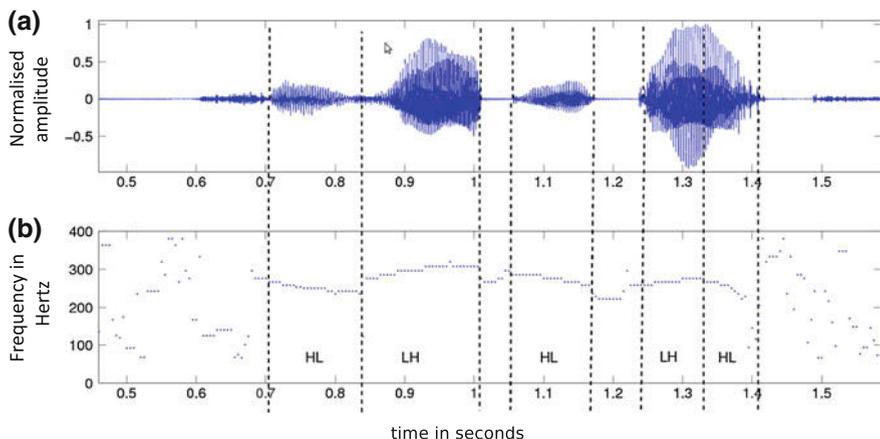


Fig. 10.4 Automatic pitch contour marking. Plot **a** shows Speech signal and plot **b** shows its pitch contour along with automatic segmentation (*dashed lines*) and markings

Figure 10.4 shows a speech waveform along with automatic pitch contour markings.

10.5 Manual Error Correction

Automatically segmented and marked speech segments are then manually corrected by looking at the output of the automatic marking in the wavesurfer. An error can be in the segmentation boundary or in the pitch contour marking. An error in the segmentation boundary can be of the following types.

- There should not be any segment boundary, but one is detected. In this case, an error correction will be to delete the segment boundary.
- There should be a segment boundary, but no segment boundary is detected. In this case, an error correction will be to insert a segment boundary.
- The segment boundary is not detected at appropriate position. In this case, an error correction will be to shift the position of the segment boundary.

An error in the pitch contour marking is a wrong marking of a particular segment. The marking process is dependent on the height value computed from the first and last samples of the estimated line in the segment. Anything wrong in the detection of segment boundary may lead to spurious pitch values at the edges which may in turn lead to wrong computation of the height value. Wrong computation of height values may result in a wrong pitch contour marking. In this case, an error correction can be to manually change the marking.

Figure 10.5 illustrates the manual correction process. Two different errors are illustrated in the figure. The automatic segmentation and marking is loaded in the transcription pane of the wavesurfer along with the waveform and pitch contour pane.

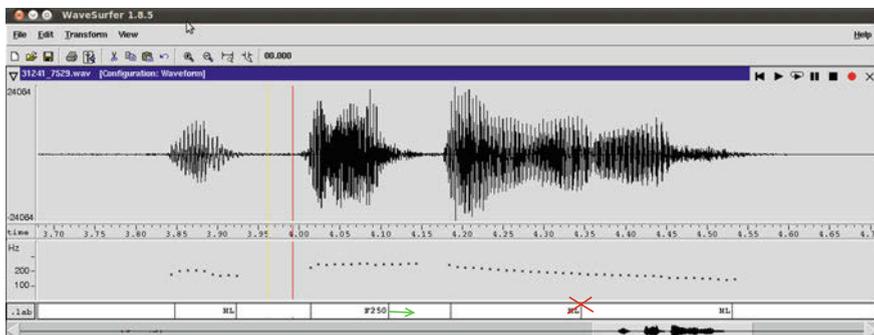


Fig. 10.5 Manual error correction in wavesurfer. Automatic process detects seven segmentation boundaries and four markings. Fourth boundary is misplaced, hence shifted toward right (shown by an *arrow*); and sixth boundary is a spurious one and hence is deleted (shown by the *cross mark*)

Table 10.1 Performance evaluation

DEL (%)	INS (%)	SFT (%)	SUB (%)
9.38	7.18	6.89	18.91

Seven segmentation boundaries and four markings are detected by the automatic segmentation and marking process. The fourth boundary detected by the automatic method is to be shifted toward right as shown by the arrow mark. Again the sixth boundary is a spurious one, hence to be deleted as described by the cross mark. The last marking should now be corrected as HVL (Table 10.1).

10.6 Results and Discussion

100 speech sentences containing around 700 pitch contour segments collected from 30 different speakers are used for the evaluation of automatic segmentation and marking of pitch contours. Performance is evaluated in terms of percentage of manual error correction.

- DEL: Percentage of deleted segment boundaries in the manual correction process out of total number of actual boundaries.
- INS: Percentage of inserted segment boundaries in the manual correction process out of total number of actual boundaries.
- SFT: Percentage of shifted segment boundaries in the manual correction process out of total number of actual boundaries.
- SUB: Percentage of substituted pitch contour markings in the manual correction process out of total number of actual pitch markings.

Percentage of substituted marking is quite high which is expected because manual marking is totally a subjective procedure and any wrong pitch value in the segment boundaries may lead to a wrong automatic marking. It is observed that the significant amount of work is reduced by introducing the semiautomatic procedure, instead of doing it fully manually. In a complete manual process, the transcriber has to first decide about voiced/unvoiced region. Sometimes, one has to zoom-in the waveform again and again to check if there is periodicity which is a time-consuming process. But in the semiautomatic process, the voiced/unvoiced decision is made fully automatic with high accuracy. Apart from that writing everything in the complete manual transcription procedure is a highly time-consuming process compared to inserting, deleting, or shifting a segment boundary wherever required.

10.7 Summary and Conclusion

This work describes a semiautomatic procedure for segmentation and marking of pitch contour useful for prosodic analysis of speech. The procedure is automatic segmentation and marking followed by manual error correction. Automatic segmentation is carried out using a voiced–nonvoiced classification method followed by segmentation of continuous pitch contour within voiced regions. Automatic marking is done by calculating a height value from the line fitted to the pitch contour. Depending on height value it is decided whether a pitch contour is rising, falling, or flat. Seven different markings as discussed in Sect. 10.4 are used for pitch contour marking. Automatic segmentation and markings are then corrected manually. Manual correction is done by loading the speech files and labels in wavesurfer. A correction in segmentation is made by either deleting a spurious segment boundary or inserting a missing segment boundary or shifting the position of a detected segment boundary. A correction in marking is done by substituting a wrong marking. Performance is evaluated in terms of these four parameters. It is observed that semiautomatization have reduced significant amount of manual work. The semiautomatically marked pitch contours will be used for prosodic analysis of speech required for different purposes such as speech synthesis, emotion detection etc.

Acknowledgments This work is part of the ongoing project on the development of Prosodically guided phonetic Engine for Assamese language funded by the Technology Development for Indian Languages (TDIL) Programme initiated by the Department of Electronics & Information Technology (DeitY), Ministry of Communication & Information Technology (MC&IT), Govt. of India under the consortium mode headed by IIIT Hyderabad.

References

1. Beckman, M.E., Elam, G.A.: Guidelines for ToBI labeling. Guidelines version 3.0, The Ohio State University Research Foundation (1997)
2. Veilleux, N., Shattuck-Hufnagel, S., Brugos, A.: 6.911 Transcribing prosodic structure of spoken utterances with ToBI, January IAP 2006. (MIT OpenCourseWare: Massachusetts Institute of Technology), <http://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-911-transcribing-prosodic-structure-of-spoken-utterances-with-tobi-january-iap-2006> (Accessed 29 Sep, 2013). License: Creative Commons BY-NC-SA
3. Syrdal, A.K., Hirschberg, J., McGory, J., Beckman, M.: Automatic ToBI prediction and alignment to speed manual labeling of prosody. *Speech Commun.* **33**, 135–151 (2001)
4. Sarma, B.D., Prasanna, S.R.M., Sarma, M.: Semi-Automatic Syllable Labelling for Assamese Language Using HMM and Vowel Onset-offset Points. Communicated to Indicon-2013, IIT Bombay
5. Murthy, K.S.R., Yegnanarayana, B.: Epoch extraction from speech signals. *IEEE Trans. Audio, Speech, Lang. Process.* **16**(8), 1602–1613 (2008)
6. Murty, K.S.R., Yegnanarayana, B., Joseph, M.A.: Characterization of glottal activity from speech signals. *IEEE Signal Process. Lett.* **16**(6), 469–472 (2009)
7. Dhananjaya, N., Yegnanarayana, B.: Voiced/nonvoiced detection based on robustness of voiced epochs. *IEEE Signal Process. Lett.* **17**(3), 273–276 (2010)

8. Rabiner, L.R.: On the use of autocorrelation analysis for pitch detection. *EEE Trans. Acoust., Speech, Signal Process.* **ASSP-25**(1), 24–33 (1977)
9. Yegnanarayana, B., Murty, K.S.R.: Event-based instantaneous fundamental frequency estimation from speech signals. *IEEE Trans. Audio, Speech, Lang. Process.* **17**(4), 1558–7916 (2009)

Chapter 11

Semi-automatic Syllable Labelling for Assamese Language Using HMM and Vowel Onset-Offset Points

Biswajit Dev Sarma, Mousmita Sarma and S.R.M. Prasanna

Abstract Syllables play an important role in speech synthesis and recognition. Prosodic information is embedded into syllable units of speech. Here we present a method for semi-automatic syllable labelling of Assamese speech utterances using Hidden Markov Models (HMMs) and vowel onset-offset points. Semi-automatic syllable labelling means syllable labelling of the speech signal when transcription or the text corresponding to the speech file is provided. HMM models for 15 broad classes of phone is built. Time label of the transcription is obtained by the forced alignment procedure using the HMM models. A parser is used to convert the word transcription to syllable transcription using certain syllabification rules. This syllable transcription and the time label of the phones are used to get the time label of the syllables. Now the syllable labelling output is refined using the knowledge of vowel onset point and vowel offset point derived from the speech signal using different signal processing techniques. This refinement gives improvement in terms of both syllable detection as well as average deviation in the syllable onset and offset.

Keywords Syllable, Syllabification · Hidden markov model · Vowel onset point · Vowel offset point

11.1 Introduction

Syllables are viewed as an important unit of prosody because many phonological rules and constraints apply within syllables or at syllable boundaries [1]. In addition to the linguistic significance, syllable boundary detection is very important in speech synthesis and automatic speech recognition [2]. Most of the Indian languages are

B.D. Sarma (✉) · M. Sarma · S.R.M. Prasanna
Indian Institute of Technology Guwahati, Guwahati 781039, India
e-mail: s.biswajit@iitg.ernet.in

M. Sarma
e-mail: mousmitas@iitg.ernet.in

S.R.M. Prasanna
e-mail: prasanna@iitg.ernet.in

syllabic in nature. So syllable-based text-to-speech synthesis or speech recognition (SR) system is suitable for Indian languages. Syllables have better representational and durational stability relative to the phoneme and therefore considered as the basic unit for many SR systems in literature [3]. Syllable-like units are proposed as the basic unit of SR in 1975 [4]. In [5], it is shown that for Indian languages, syllable-like unit segmentation followed by isolated recognition of segmented syllables in continuous speech performs well. In text-to-speech synthesis system, different modules such as a letter-to-phoneme module, a prosody module or a synthesis module uses syllables. The pronunciation of a given phoneme tends to vary depending on its location within a syllable. In unit selection approach for speech synthesis, syllable units are labelled in the speech utterances and stored. At the time of synthesis, suitable syllable units, depending on the context, are extracted from the database and concatenated. In all these applications, syllable-marked speech database is required either for creating a model or for concatenation. This work presents an alternate way of manual syllable labelling of speech signal when transcription is provided.

Segmentation techniques are mostly based on hidden markov models and uses phonemes or syllable models [6]. To get the boundaries of the syllables from continuous speech, a temporal flow model network is used in [7]. Group delay-based techniques are used in literature for automatic syllable transcription in [3].

In our work, we use Hidden Markov Model (HMM) to get the syllable boundaries at first level. Later, these boundaries are refined by the knowledge of certain landmarks in the vowel region. In case of Indian languages more than 90% of syllables have C^nV and VC^n structure [8]. Where C denotes consonant, V denotes vowel and n is referring to number of consonants ($n = 0, 1, 2, 3$). For a syllable of type VC^n and C^nV , the onset or offset of the syllable is the onset or offset of the vowel in the unit. There are suitable techniques available in the literature for accurate detection of vowel onset and offset [9–11]. By using this knowledge, syllable boundary marking is obtained more accurately.

The paper is structured as follows. Section 11.2 describes the Assamese phoneset used in the syllabification process. Section 11.3 describes about the syllabification parser for converting word transcription to syllable transcription. HMM-based syllable labelling is discussed in Sect. 11.4. Refinement of the syllable boundary marking using vowel landmarks are described in Sect. 11.5. Section 11.6 demonstrates the results and Sect. 11.7 concludes.

11.2 Assamese Phoneset

In [12], a phoneset of Assamese is presented for isolated word recognition task. However, that phoneset is made for a limited vocabulary database (around 150 words) and does not contain all phones present in the Assamese language. 23 consonants,

Table 11.1 Table representing the Assamese script symbol, corresponding IPA symbol and symbol used in the syllabification parser

LETTERS	IPA	SYMBOL USED
অ	/ɔ/	a
	/o/	o
আ	/a/	aa
ই/ঈ	/i/	i
উ/ঊ	/u/	u
এ	/ɛ/	e
	/e/	ee
ও	/ɔ/	u
ঐ	/oi/	oi
ঔ	/ou/	ou
ক	/k/	k
খ	/kʰ/	kh
গ	/g/	g
ঘ	/gʰ/	gh
ঙ	/ŋ/	ng
চ/ছ	/s/	s
জ/ষ	/z/	j
ট/ত	/t/	t
ঠ/থ	/tʰ/	th
ড/দ	/d/	d
ঢ/ধ	/dʰ/	dh
ণ/ন	/n/	n
প	/p/	p
ফ	/pʰ/	ph
ব	/b/	b
ভ	/bʰ/	bh
ম	/m/	m
ৰ	/r/	r
ল	/l/	l
ৱ	/w/	w
শ/স/ষ	/x/	x
হ	/h/	h
য	/j/	y
	/ɔi/	ai
	/oi/	aa
	/au/	aa
	/ei/	ei
	/eu/	eu
	/ui/	ui
	/iu/	iu

It is to note that some entries under the Assamese script are blank as there is no symbol defined for those sounds

8 vowels and 5 diphthongs are used in the task. In continuous speech having large vocabulary, some additional diphthongs are observed to be present which are required for the syllabification parser in our work. The list of phones used in this work is shown in Table 11.1.

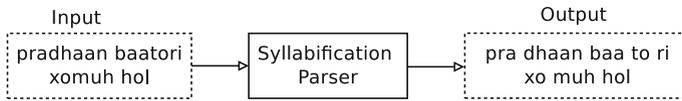


Fig. 11.1 Syllabification parser

11.3 Syllabification Parser

Syllabification parser converts phone-level transcriptions having word boundaries to a syllable-level transcription as shown in Fig. 11.1. Using the phoneset shown in Table 11.1, a phone-level transcription is made for the speech database. A space is inserted between the words to retain the word-level information. Now the transcription is processed one word at a time to get the syllables present in that word. To convert one word to syllables, certain general syllabification rules [13] along with some rules specific to the language are derived and applied. The rules are as follows:

- Each vowel or diphthong will produce a separate nucleus
- If a single consonant is present in the left of the nucleus, it will be an onset of the right syllable
- If two consonants are present in between two vowels, the first consonant will be part of coda of previous syllable and the second will be onset of the next syllable (e.g. gahbar -- > gah bar)
- In the previous case, when the second consonant is a member of the set /r/ /y/, both the consonants would be a part of onset of the next syllable (e.g., kritrim -- > kri trim)
- If there are three or more consonants between two consecutive vowels, the first consonant will be a part of the coda of the previous syllable while the remaining consonants will be onset of the next syllable. (e.g. jaantrik -- > jaan trik)
- A consonant cluster at the beginning and end of a word will be part of onset and coda of the first and last syllable, respectively.

11.4 HMM-Based Semi-automatic Syllabification

A hidden markov model-based classifier is developed to classify the broad sound categories. Three classes are formed for each of nasal, affricate and fricative, i.e. all sounds belonging to a particular class is merged to form a single class. Stops have many variations, so six separate classes are prepared for the stop sounds. Diphthongs are transcribed as a combination of two vowels and are not treated as separate classes; three classes are formed for the vowels by merging similar vowels into one class. Approximants are put into three different broad classes. Assamese phones described in Sect. 11.3 are used to form 14 such classes and Table 11.2 shows the phones corresponding to each class. The phones are merged because the purpose is not to

Table 11.2 Assamese phones under different classes

Stops	k, kh	g, gh	t, th	d, dh	p, ph	b, bh
Fricatives	s, j, x, h					
Nasals	m, n, ng					
Vowels	a, u, o	aa			i, e, ee	
Approximants	r		l		y	

build a phone recognizer but to get the phone boundaries. These phone boundaries will in turn be used to get the syllable boundaries for the syllables produced by the parser. HMM is used for this purpose because viterbi algorithm can give the best optimal path for a given model to produce a known observation sequence.

Context-independent monophone HMMs are built for each of these classes and one for the silence. Language information is not used in the system. Three states are used to model each class. Each state uses 32 mixtures Gaussian mixture model (GMMs). 9 hours of speech data collected from 35 native Assamese speakers are used for building the phone models. Five iterations of Embedded Baum–Welch training is carried out for reestimation of model parameters. The open-source HMM tool kit (HTK) [14] is used for building the phone recognition system. The same system when built for the 39 phone labels listed in [15] using TIMIT training set (10 utterances of duration 3–5 s from each of 462 speakers) and tested on the TIMIT test set (10 utterances from each of 168 speakers), the performance accuracy of phoneme recognition is found to be 66.33 %.

The transcription with the reduced phoneset is used for forced-alignment using the HMM models to get the phone boundaries. Now a syllable boundary is obtained from the phone boundary of the first and the last phone of that syllable. A perl script is used to do this operation.

11.5 Refinement of the Syllabification Output Using Vowel Onset and Offset Point

The vowel contains the major portion of the syllable unit. For syllables of type CV and VC, accurate detection of a syllable boundary is dependent on accurate detection of onset and offset of the vowel unit. For Indian languages, 90 % of the syllables are CV units [8]. For continuous CV utterances, accurate detection of vowel offset will in turn make accurate detection of the next consonant onset. In one line, we can say that accurate detection of vowel onset and offset will make overall accurate detection of syllable onset and offset. There are several signal processing methods available in the literature to detect the vowel onset point (VOP) and vowel offset or end point (VEP) accurately. These techniques can be used to refine the boundaries of the syllable units. In [10], a VOP detection technique is proposed which is one of the recent

VOP detection technique. The VOP is detected by combining three evidences—1. excitation source, 2. sum of spectral peaks and 3. modulation spectrum energy. The most recent paper on VEP detection proposes two techniques for VEP detection [11]. First method uses spectral energy computed in the glottal closure region and the second method uses the combination of three evidences. In our work, the method that uses the combined evidence is used for VOP and VEP detection as described in [10, 11].

Evidence from excitation source information is extracted in following sequence of steps: First the Hilbert envelope (HE) of linear prediction (LP) residual of speech signal is determined. HE of LP residual is smoothed by convolving the signal with a Hamming window of size 50 ms. The change of the smoothed signal at the VOP and VEP present is further enhanced by computing its slope by taking first-order difference (FOD). The signal is convolved with the first-order Gaussian differentiator (FOGD). The output of the convolution is the evidence using excitation source.

Evidence from spectral peaks energy is obtained using the following sequence of operations. 20 ms block processing is performed in the speech signal with a shift of 10 ms. A 256-point DFT is computed for each block and 10 largest peaks are obtained from the frequency spectrum. The sum of energies of these spectral components is computed. The change at the VOP and VEP present is further enhanced by computing its slope by taking the first-order difference (FOD). The enhanced signal is convolved with FOGD operator. The output of the convolution is the evidence using spectral peaks energy.

Modulation spectrum represents the slowly varying temporal envelope of speech signal. VOP and VEP detection using modulation spectrum energy is carried out in the following sequence of steps. Low-frequency components dominate the temporal envelope of speech signal. Speech signal is passed through a set of critical bandpass filters and the components corresponding to 4–16 Hz is summed up. The output of this summation is the modulation spectrum-based VOP/VEP evidence. The modulation spectrum energy-based evidence is further enhanced by determining its slope using FOD. This enhanced signal is convolved with FOGD operator. The output of the convolution is the evidence using modulation spectrum energies.

All three evidences are combined to get final evidence for VOP and VEP. The positive and negative peaks in the combined evidence signal are marked as the VOPs and VEPs, respectively.

Figure 11.2a shows syllable marking by HMM-based technique described in the previous section along with actual marking. In case of the C^1V unit around 400 ms it is seen that although the C^1V unit is spotted, the end point of the unit is much deviated from the actual endpoint. Similarly in case of the VC^0 unit around 100 ms the end point of the unit is deviated from the actual endpoint. Endpoint of the C^1V and VC^0 unit is the endpoint of the vowel in it. Similarly, start point of the VC^0 unit is the onset point of the vowel. Using accurate detection of VOP and VEP, segmentation accuracy of these two units can be improved. Figure 11.2b shows syllable marking after refinement along with actual marking.

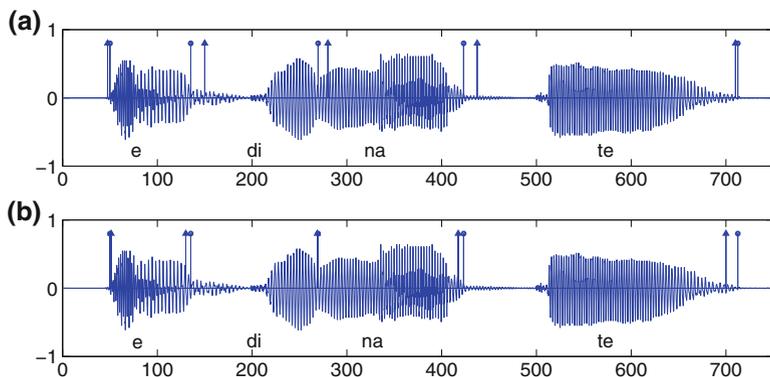


Fig. 11.2 Syllable labelling of the word ‘edinate’: **a** HMM-based marking with *arrow* and manual marking with *circle* **b** Marking after refinement with *arrow* and manual marking with *circle*

The refinement procedure is as follows:

- Determine the syllable boundaries using the HMM-based technique described in the previous section.
- Now, if a syllable is having $C^n V$ structure, detect the VEP of the vowel present in the unit and if any VEP is found within 20 ms of the syllable end point, declare the detected VEP as the refined syllable end point.
- Similarly, if a syllable is having $V C^n$ structure, detect the VOP of the vowel present in the unit and if any VOP is found within 20 ms of the syllable onset point, declare the detected VOP as the refined syllable onset point.

11.6 Results and Discussion

50 sentences from the Assamese database which are not part of the training data are taken and are manually labelled the syllables using wave-surfer. Pitch contour, spectrogram etc. are used for accurate labelling. Now the same sentences are tested with the proposed system. The performance is evaluated using the parameters shown below.

Detection Rate (DR): Percentage of syllable onset and offset detected within 40 ms deviation of actual syllable onset and offset.

Average Deviation (AD): Average of the deviations of all the detected syllable onsets and offsets.

Spurious detection (SD): Syllable Onset or Offset that is detected beyond 40 ms of an actual Onset or Offset.

System performance before and after refinement are shown in Table 11.3. It is seen that using only HMM-based method gives 92.81 % detection rate which is improved to 93.35 % after refining with the knowledge of VOP and VEP. In case of average

Table 11.3 Syllable labelling performance

Method	DR (%)	AD (ms)	SD (%)
HMM	92.81	14	7.19
HMM+VOP/VEP	93.35	12	6.65

deviation, HMM gives around 14 ms of average deviation, whereas, the refined output gives 12 ms of average deviation. Thus, after refinement, improvement is observed in terms of detection rate and average deviation.

11.7 Summary and Conclusion

This work describes a semi-automatic syllable labelling procedure using HMM and the knowledge of VOP and VEP. A syllabification parser which uses a set of syllabification rules to convert the phone transcriptions to syllable transcription is built. A 15 class broad phone classifier is trained to get forced aligned boundary of the phones. Syllable boundaries are obtained from the phone boundaries. These syllable boundaries are refined using the knowledge of VOP and VEP detected by different signal processing techniques. Refinement gives 0.54 % improvement in the detection rate and 2 ms improvement in average deviation of the syllable onset and offset. As a part of future work, this syllable labelling procedure can be used in Assamese text-to-speech synthesis systems as well as in speech recognition systems to incorporate prosodic knowledge.

Acknowledgments This work is part of the ongoing project on the development of prosodically-guided phonetic engine for Assamese language funded by the TDIL Programme initiated by the DeitY, MC&IT, Govt. of India under the consortium mode headed by IIT Hyderabad.

References

1. Bartlett, S., Kondrak, G., Cherry, C.: On the syllabification of phonemes. In: Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the ACL, pp. 308–316. Boulder, Colorado, June 2009
2. Bartlett, S., Kondrak, G., Cherry, C.: Automatic syllabification with structured SVMs for letter-to-phoneme conversion. In: Proceedings of Association for Computational Linguistics-08: HLT, pp. 568–576, Columbus, Ohio, USA, June 2008
3. Nagarajan, T., Hema, A., Murthy., Hegde, R.M.: Segmentation of speech into syllable-like units. In: Eurospeech (2003)
4. Fujimura, O.: Syllable as a unit of speech recognition. *IEEE Trans. Acoust. Speech Signal Process.* **23**(1), 82–87 (1975)

5. Prasad, K.V.: Segmentation and recognition of continuous speech, Ph.D dissertation, Indian Institute of Technology. Department of Computer Science and Engineering, Madras, India (2002)
6. Nakagawa, S., Hashimoto, Y.: A method for continuous speech segmentation using hmm. In: Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 960–962 (1988)
7. shastri, L., chang, S., Greenberg, S.: Syllable detection and segmentation using temporal flow neural networks. In: ICPhS, pp. 1721–1724. San Francisco (1999)
8. Gangashetty, S.V.: “Neural Network Models For Recognition of Consonant-Vowel units of speech in multiple languages. Ph.D thesis, Department of Computer Science and Engineering, IIT Madras October 2004
9. Vappula, A.K., Yadav, J., Chakrabarti, S., Rao, K.S.: Vowel onset point detection for low bit rate coded speech. *IEEE Trans. Audio, Speech Lang. Process.* **20**(6), 2552–2565 (2012)
10. Prasanna, S.R.M., Reddy, B.V.S., Krishnamoorthy, P.: Vowel onset point detection using source, spectral peaks, and modulation spectrum energies. *IEEE Trans. Audio, Speech, Lang. Process.*, **17**(4), pp. 556–565 (2009)
11. Yadav, J., Rao, K.S.: Detection of vowel offset point from speech signal. *IEEE Signal Process. Lett.*, **20**(4), (2013)
12. Shahnawazuddin, S., Thotappa, D., Sarma, B.D., Deka, A., Prasanna, S.R.M., Sinha, R.: Assamese spoken query system to access the price of agricultural commodities. In: Proceedings of the National Conference on Communication, (February 2013)
13. Prahallad, K.S., Kumar, R., Sangal, R.: A data-driven synthesis approach for Indian languages using syllable as basic unit. In: Proceedings of the International Conference Natural Language Processing (ICON-2002) (2002)
14. Young, S., et al.: The HTK Book (for HTK Version 3.4). Cambridge University Engineering Department, Cambridge (2009)
15. Lee, Kai-Fu, F.: Speaker-independent phone recognition using hidden Markov models. *IEEE Trans. Acoust. Speech Signal Process.*, vol. 37, pp. 1641–1648 (November 1989)

Chapter 12

Block Matching Algorithms for Motion Estimation: A Performance-Based Study

Hussain Ahmed Choudhury and Monjul Saikia

Abstract Motion estimation (ME) process is the most crucial and time-consuming part of video compression technique. So many block-based motion estimation techniques were developed to make ME easy and fast. In this paper we have reviewed almost all the existing BMA from very old Full Search (FS) to recently developed Reduced Three-Step Logarithmic Search (RTSLS) and Cross Three-Step Logarithmic Search (CTSLS), and so on. We have also compared them based on the computations needed per macroblock and the PSNR value of compensated image. Earlier Adaptive Rood Pattern Search (ARPS) was found to be most computationally efficient but during the review we have applied the old algorithms along with recently developed Zero Motion preadjusted RTSLS (ZMRTSLS) and Zero Motion preadjusted CTSLS (ZMCTSLS) are found to be more computationally efficient than even ARPS.

Keywords BMA · RTSLS · ZMRTSLS · CTSLS · ZMCTSLS

12.1 Introduction

In the recent years, the **video communication** became an integral part of communication channel and most popular medium of entertainment. So we need to process the videos fast and in better way and for doing that we will have to store them efficiently and encode them properly with the help of hybrid video codec. The process of encoding requires finding out the difference in position of the moving object from one frame to another. This requires a great deal of computation and this is similar to search technique which finds the best match of current block in reference frame. This process is the most important aspect of video compression called **motion estimation**. The motion estimation can be done by comparing pixel to pixel values of two consecutive or so frames. But it requires a great deal of computations and time.

H.A. Choudhury (✉) · M. Saikia
NERIST, Itanagar 791109, Arunachal Pradesh, India
e-mail: hachoudhury10@gmail.com

M. Saikia
e-mail: monjuls@gmail.com

So we divide the image frames into nonoverlapping blocks called macroblocks and do block by block matching to get the best match. This process is called as *block-based motion estimation* and it reduces the number of computations by many folds. The motion estimation process gives us the coordinate difference between the positions between two frames referred as current frame and reference frame. We denote it as motion vectors (MV).

In this paper we have reviewed almost all the major existing techniques for motion estimation and compared their performances based on the computations taken per macroblock and PSNR value of compensated image. In the paper Sect. 12.2 explains the mode of macroblocks (MB), Sect. 12.3 explains different existing BMA, Sect. 12.4 explains simulation results, Sect. 12.5 explains comparative study results and followed by conclusion and references.

12.2 Mode of Macroblock

There are certain macroblocks which do not undergo significant amount of motion and other undergoing very significant motion. Suppose that we have to encode a macroblock which has no correspondence in past frame then have to encode them directly with *intra-frame* coding. Intra frame means there is no past reference. There are some macroblocks which have correspondence in past frame, can be encoded with *inter-frame* coding.

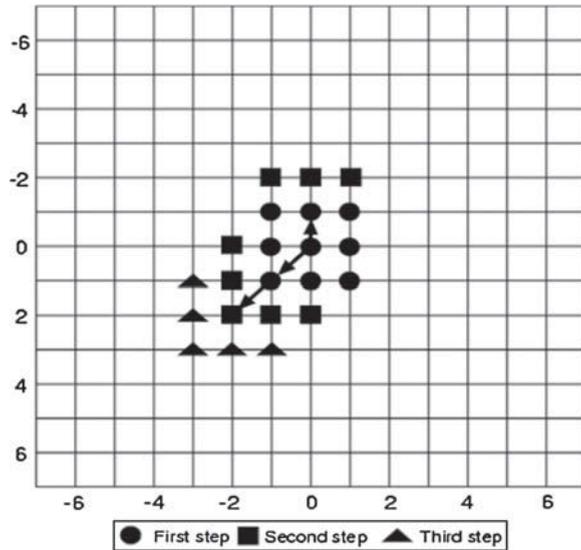
12.3 Existing Block Matching Algorithm

The following are the existing algorithms used in block matching of videos.

12.3.1 Full Search (FS)

It is the earliest and slowest block-based motion estimation technique for motion estimation [1–3]. But as it checks all positions within a search, it gives us optimal result, i.e., best match with maximum PSNR value for compensated image. For a search range of five pixels in all directions, we need to compare and check $(2 * 5 + 1)^2 = 121$ positions for minimum MAD value corresponding to best MV.

Fig. 12.1 Search pattern of block-based gradient descent search algorithm



12.3.2 Block-Based Gradient Descent Search (BBGDS)

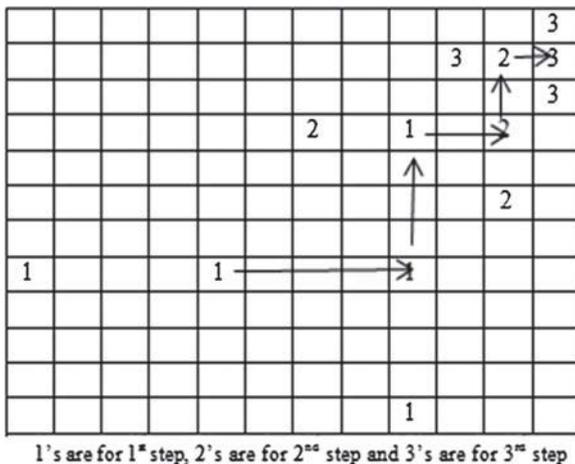
BBGDS [4] starts searching nine locations with step size 1 in an area of 3×3 for best motion vector (MV). The search pattern follows the steps mentioned below (Fig. 12.1):

1. With the initial step size 1, BBGDS searches the all nine positions around the center (0, 0) positions for best match corresponding to minimum Mean Absolute Difference.
2. If the minimum MAD is found to be at center of earlier step, we will stop the search and the (0, 0) will be our position of best match and (0, 0) becomes the MV. This is known as halfway stop policy. If we do not find the best match at center, we will shift the center to the new minimum and will follow the first step until we find the best match at the center.

12.3.3 Three-Step Search (TSS)

This is the combination of *2DLS* and *Cross Search* (CS) which completes in three steps [2, 3, 5–9]. The search starts with a step size 4 and searches nine points in all direction in first step. In the second and third step, this algorithm searches 8 and 8 locations for best match with step sizes 2 and 1, respectively. When the step size becomes one, the position corresponding to minimum MAD will give the best MV. So a total of 25 locations are needed to be searched for each macroblock for finding best-matching MV.

Fig. 12.2 OSA search pattern



12.3.4 Orthogonal Search Algorithm (OSA)

This algorithm [10] has combined the properties of *Three-Step Search* (TSS) and *2-Dimensional Logarithmic Search* (2DLS). It starts with search in horizontal direction and then in vertical direction. If we consider the maximum displacement to be +8 then the search follows the steps mentioned below: (Fig. 12.2)

1. The search starts with step size equal to the half of the maximum displacement, i.e., 4. In this step, it searches three points in the horizontal direction including left and right points at a distance of 4 from center and the center position itself for minimum MAD. Then we shift the center to the new minimum point.
2. Next we search the points in vertical direction of the new minimum position, i.e., one point above and one point below at a distance of 4 for minimum MAD. We shift the center to the new minimum.
3. Reduce the step size by half, i.e., step size becomes 2. Then we follow the first and second step with step size 2 until step size becomes 1. When step size becomes 1, we follow the earlier steps and the coordinate of minimum MAD will give us the best MV.

12.3.5 One at a Time Search (OTS)

As the name suggests it searches the points in *horizontal direction* for best MV followed by vertical direction. In each step of vertical or horizontal search one extra point is need to be checked. This algorithm follows the steps as follows:

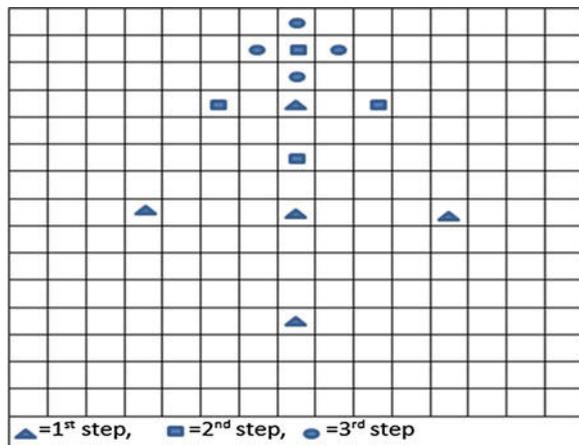
1. In this step, we check three positions for minimum MAD at center and two points around it in horizontal direction.
2. If the minimum point is the center of previous step, then we start searching 2 points around the center in vertical direction and the minimum MAD position will be the best MV else follow step 3.
3. Shift the center to the minimum position of the first step and continue search in the same direction till we find the minimum at center. In each step of horizontal search, we consider one extra point to search for best match. When we get the minimum at center we follow step 2.

12.3.6 Reduced Three-Step Logarithmic Search (RTSLS)

This algorithm [11] is the modified version of TSS. It is proposed in the year 2014 which is the newest of all [11]. In the first step, this algorithm computes five locations for minimum MAD with step size $S = 4$, including center at the end of '+'. In each step, it reduces the step size by half till it becomes 1. In the rest two steps, it computes 4 points each time with $S = 2$ and $S = 1$, respectively. When the step size is 1, the coordinate position corresponding to minimum MAD will give us best MV (Shown in Fig. 12.3).

Thus it computes 13 points for finding the best match and it completes in three steps.

Fig. 12.3 RTSLS search pattern



12.3.7 Zero Motion Preadjusted RTSLS (ZMRTSLS)

It is the RTSLS with halfway stop condition. In the first step, RTSLS computes 5 points for minimum MAD with step size $S = 4$. In this algorithm, one condition is added to RTSLS to check whether the minimum is at the center of the first step or not. If it is found that the minimum is at the center, there is no need to search for minimum MAD in the next steps. This stops the search in the first step and it is predicted that the block is not moving and give us the MV as (0, 0) with only 5 computations. As before doing many steps, we predict the zero motion of block so accordingly it is called zero motion preadjusted RTSLS (ZMRTSLS).

12.3.8 Cross Three-Step Logarithmic Search (CTSLS)

This [12] is also the modification of TSS and other form of RTSLS [12]. The difference between RTSLS and CTSLS is that RTSLS checks the points at the end of '+' (plus) whereas CTSLS checks points at the end points of 'X' (cross). This search also computes 5 points with $S = 4$, 4 points with $S = 2$ and 4 points with $S = 1$. So the difference in search pattern in CTSLS than that of RTSLS, effects the PSNR values of compensated image though the computations needed are same, i.e., with 13 checking points we can get the best MV.

12.3.9 Zero Motion Preadjusted CTSLS (ZMCTSLS)

This is also similar with that of ZMRTSLS. The added condition of preadjusted zero motion to CTSLS makes it computationally more efficient than to CTSLS. It takes an average of only 5 checking points for finding best match.

12.3.10 Reduced Diamond Search Algorithm (RDS)

This is basically the modification of the Diamond Search Algorithm [13, 14]. It is computationally more efficient than DS [13]. The search steps are:

1. It starts with the search of 5 points at end of + for best matching block which is the point of minimum MAD. If the minimum is located at the center, follow step 4 else step 2.
2. Now we check two more new points near the minimum location to form half diamond shape pattern. Then we calculate MAD at these points and compare for minimum MAD at those locations.

3. Shift the center to the new minimum and form new LDSP taking this as center. If the minimum occurs again at the same position, follow step 4 else keep repeating this step.
4. We follow SDSP. The position with minimum MAD will be the best MV.

12.4 Simulation Results

In our experiment we have taken the mean absolute difference (MAD) as a measure of matching criteria. We have applied the techniques by taking macroblock size of 8×8 and 16×16 . The maximum displacement in search area is taken as $+7$ and the search area as $(2 \times 7 + 1) \times (2 \times 7 + 1) = 225$.

The simulation is performed on different video sequences with different frame length as listed in Table 12.1. The results and outputs are obtained as average no. of searching points required per macroblock and the PSNR.

In the following Table 12.2 we have shown the results obtained by using various existing block-based ME techniques for average values of PSNR.

12.5 Comparative Study Results

12.5.1 Complexity Point of View

The Average number of computations required by different Block Based Motion Estimation techniques is shown in Fig. 12.4.

Table 12.1 Video Sequences chosen for analysis

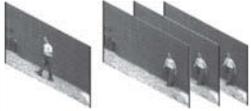
Frame format (Numbers of frames)	Sequences	Video frames
avi (240×360 , 80 frames)	Sample video	
ras (288×352 , 30 frames)	Missa	

Table 12.2 Avg. computations and avg. PSNR obtained using all techniques

Algorithm	Sample video (mb Size 16 × 16)		Sample video (mb Size 8 × 8)		Missa (mb Size 16 × 16)	
	Avg. PSNR	Avg. Comp	Avg. PSNR	Avg. Comp	Avg. PSNR	Avg. Comp
TSS	32.7325	23.0182	33.1574	24.0045	36.5393	23.26515
NTSS	32.4105	16.1636	33.5114	17.0003	36.53965	19.3586
4SS	32.7723	15.9076	33.4328	16.6053	36.4791	17.3434
SES	32.2074	16.8909	32.9832	17.3625	36.2906	16.5215
DSA	32.8564	12.4818	33.3539	12.9723	36.5338	14.4950
ARPS	32.8751	5.4016	33.164	5.5223	36.4833	7.09345
RTSLS	32.2018	12.3273	33.0153	12.6651	36.16315	12.4103
ZMRTSLS	37.1860	4.8727	35.2703	5.1126	35.4267	6.8131
CTSLS	31.4534	11.5679	32.0182	12.3379	36.1407	11.8485
ZMCTSLS	29.9193	4.6848	30.7020	4.9667	35.1350	5.6768

Fig. 12.4 Average computations required by different techniques (X-axis: Frame Number, Y-axis: Computations)

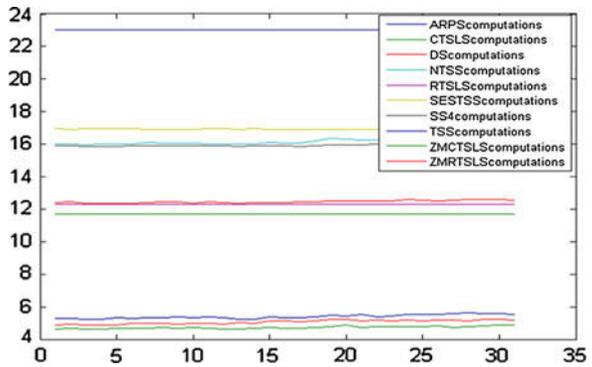
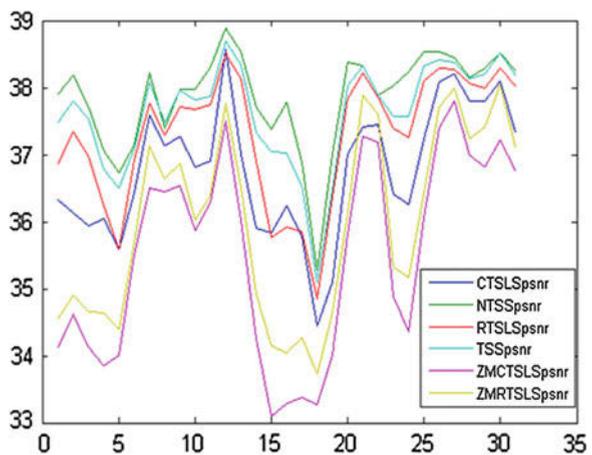


Fig. 12.5 PSNR values per macroblock of different techniques (X-axis: Frame Number, Y-axis: PSNR)



12.5.2 In Case of PSNR Values

We have applied the existing algorithms on different video sequences and the PSNR values are obtained for them. Following Fig. 12.5 shows the PSNR values obtained by using different techniques on SampleVideo.avi (Table 12.3).

Table 12.3 Summary table of Comparative study

Algorithm	Summary
FS:	The FS algorithm searches for the best matching within a large window $[-p:p][-p:p]$. So for a displacement of +7, we need to check almost 225 locations for best match
TSS:	This algorithm completes in three steps and takes an average of 25 computations per MB for finding beat match [2, 3, 5–9]. (shown in Fig. 12.4)
NTSS:	It is an improved version of TSS. It takes an average of 17 computations/MB in best case and 33 computations/MB in worst case [1, 2, 5, 7, 19]. (shown in Fig. 12.4)
4SS:	It completes in four steps. The computations/MB of 4SS varies from 17 to 27 [1–3, 7, 8, 15, 20]. (shown in Fig. 12.4)
2DLS:	This algorithm takes an average of 13 to 17 computations/MB to find the best MV [6, 9]
OSA:	If we consider a maximum displacement of +7, OSA takes 13 computations/MB [3]
CSA:	The computations taken by CSA [16] can be found by using the formula $5 + 4 * \log_2 S$ where S is the initial step size. If we choose $S = 4$ for a displacement of +7, the computations taken by CSA will become 13 per MB [3, 6, 7]
OTS:	The number of checking points required by the OTS algorithm varies from 5 to 17 [3]
HXSP:	It can be easily derived that the total number of search points per block will be $N(x, y) = 7 + 3 * n + 4$ where (x, y) is the final motion vector found and n is the number of iterations [8]
RTSLS:	This takes an average of 13 computations per macroblock for finding MV [11]. (shown in Fig. 12.4)
ZMRTSLS:	This technique reduces the number of comparisons per macroblock to almost 5 if it finds the least cost at the center of first step. (shown in Fig. 12.4)
CTSLS:	This technique also has the advantage over TSS and NTSS but takes same 13 checking points to give the best motion vector. (shown in Fig. 12.4)
ZMCTSLS:	This technique reduces the computations required per block in CTSLS to almost half. (shown in Fig. 12.4)

12.6 Conclusion

From the study results, it is seen that FS has both the **highest computational complexity and Peak Signal to Noise Ratio (PSNR)** value and guarantees the optimal solution. But still to make faster search many fast block-based motion estimation techniques are developed, though they give suboptimal solutions only. From the experimental results shown in Table 12.2, we can say that out of the techniques, RTSLS and CTSLS take almost half of the computations for finding MV required by TSS whereas

produce compensated image of similar quality. These algorithms are even computationally more efficient than NTSS, 4SS, 2DLS, SES [17], and many others. Results also show that ZMRTSLS and ZMCTSLS are found to be most computationally efficient even better than Adaptive Rood pattern search (ARPS) [2, 18].

References

1. MS. Chauhan, A.P., Parmar, R.R., Parmar, S.K., Chauhan, S.G.: Comparative analysis on the performance of block matching motion estimation algorithm. *J. Inf. Knowl. Res. Comput. Eng.* **2**(2), 366–370 (2012). ISSN 0975-6760
2. Barjatya, A.: Block matching algorithms for motion estimation. DIP 6620 Spring 2004 Final Project Paper
3. Pandian, S.I.A., Bala, G.J., George, B.A.: A study on block matching algorithms for motion estimation. *Int. J. Comput. Sci. Eng. (IJCSE)* **3**(1) (2011). ISSN: 0975-3397
4. Lurug, L.K., Ephraim, F.: A block-based gradient descent search algorithm for block motion estimation in video coding. *IEEE Trans. Circuit Syst.* **6**(4), 419–422 (1996)
5. Manjunatha, D.V., Sainarayanan Dr.: Comparison and implementation of fast block matching motion estimation algorithms for video compression. *Int. J. Eng. Sci. Technol. (IJEST)* **3**, 7608–7613 (2011). ISSN 0975-5462
6. Choudhury, H.A., Saikia M.: Comparative study of block matching algorithm for motion estimation. In: International Conference on Advanced Computer Science and Information Technology, Bhubaneswar, India, 17 November 2013. ISBN: 978-93-82702-39-9
7. Choudhury, H.A., Saikia M.: Survey on block matching algorithms for motion estimation. In: IEEE International Conference on Communication and Signal Processing-ICCSP' 14, 3–5 April 2014. pp. 633–637. ISBN 978-1-4799-3357-0
8. Manap, R.A., Sarban Singh, R.S.: Performance analysis of hexagon-diamond search algorithm for motion estimation using Matlab. *J. Telecommun, Electron. Comput. Eng.* **2**(2), 13 (2010). ISSN: 2180-1843
9. Shenolikar, P.C., Narote, S.P.: Different approaches for motion estimation. In: International Conference on Control, Automation, Communication and Energy Conservation, 4–6 June 2009
10. Usama, S., Montaser, M., Ahmed, O.: A complexity and quality evaluation of block based motion estimation algorithms. *Acta Polytech.* **45**(1), 29 (2005)
11. Choudhury, H.A., Saikia M.: Reduced three steps logarithmic search for motion estimation. In: IEEE 4th International Conference on Information Communication and Embedded Systems—ICICES 2014. IEEE 27, 28 February 2014. ISBN 978-1-4799-3834-6/14/31.002014
12. Choudhury, H.A., Saikia M.: Cross three step logarithmic search for motion estimation. In: IEEE International Conferences for Convergence of Technology, Pune, India, 5–7 April 2014, IEEE. 978-1-4799-3759-2/14/31.002014
13. Atalla, I.H., Rowayda, A.S., Sara, K.M.: A novel reduced diamond search (RDS) algorithm with early termination for fast motion estimation. *Int. J. Video Image Process. Netw. Secur. IJVIPNS-IJENS* **10**(4), 14 (2010)
14. Zhu, S., Ma, K.K.: A new diamond search algorithm for fast block matching motion estimation. *IEEE Trans. Image Process.* **9**(2), 287–290 (2000)
15. Yusop, N.B.: Implementation of four step search (4SS) algorithm for motion estimation in MATLAB, final year Degree report, April 2007
16. Lam, C.W., Po, L.M., Cheung, C.H.: A new cross-diamond search algorithm for fast block matching motion estimation. In: Proceeding of 2003 IEEE International Conference on Neural Networks and Signal Processing, Nanjing, China, pp. 1262–1265, December 2003
17. Jianhua, L., Liou, M.L.: A simple and efficient search algorithm for block-matching motion estimation. *IEEE Trans. Circuits Syst. Video Technol.* **7**(2), 429–433 (1997)

18. Nie, Y., Ma, K.-K.: Adaptive rood pattern search for fast block-matching motion estimation. *IEEE Trans. Image Process.* **11**(12), 1442–1448 (2002)
19. Li, R., Zeng, B., Liou, M.L.: A new three step search algorithm for block motion estimation. *IEEE Trans. Circuits Syst. Video Technol.* **4**(4), 438–442 (1994)
20. Po, L.M., Ma, W.C.: A novel four-step search algorithm for fast block motion estimation. *IEEE Trans. Circuits Syst. Video Technol.* **6**(3), 313–317 (1996)

Chapter 13

Stressed Speech Recognition Using Similarity Measurement on Inner Product Space

Bhanu Priya and S. Dandapat

Abstract In this paper, similarity measurement on different inner product space approach is proposed for analysis of stressed speech. The similarity is measured between neutral speech subspace and stressed speech subspace. Cosine between neutral speech and stressed speech is taken as similarity measurement parameter. It is assumed that, speech and stress components of stressed speech are linearly related to each other. Cosine between neutral and stressed speech multiples of stressed speech contains speech information of stressed speech. Complement cosine (1-cosine) multiples of stressed speech is taken as stress component of stressed speech. Neutral speech subspace is created by all neutral speech of the training database and stressed speech subspace contain stressed (angry, sad, lombard, happy) speech. From experiment, it is observed that, stress information of stressed speech is not present in the complement cosine (1-cosine) times of stressed speech on different inner product space. The linear relationship between speech and stress component of stressed speech exists only for some specific inner product space. All the experiments are done using nonlinear (TEO-CB-Auto-Env) feature.

Keywords TEO-CB-Auto-Env · Dot product · Unbiased covariance · Cosine · Inner product

13.1 Introduction

Speech is produced by articulator movement and airflow from the respiratory system. Stressed speech means speech produced by some alteration of speech production system. Therefore, stressed speech is produced by speaker, under some environmental factor, emotional state, emergency condition, sleep deprivation, multitasking condition, etc. Neutral speech is produced by speaker in stress-free condition. Hansen

B. Priya (✉) · S. Dandapat
Indian Institute of Technology Guwahati, Guwahati 781039, India
e-mail: b.priya@iitg.ernet.in

S. Dandapat
e-mail: samaren@iitg.ernet.in

and Patil [1] shown that moment (mean) analysis of fundamental frequency for loud, angry, and lombard condition speech is different from neutral condition speech. Phoneme duration changes in stressed speech as compared to neutral speech. Duration of consonant is an important factor for perceiving speaker's information [1].

Speech recognition performance reduces due to the presence of stress component in speech. Number of works are done for separation of stress component from stressed speech. Cepstral domain talker stress compensation is used to remove the stress from stressed speech in cepstral domain [2]. In sinusoidal model-based analysis, sinusoidal model is used to model the variation in speech properties under stressed condition [3]. Ramamohan and Dandapat [3] shown that amplitude, frequency, and phase features of sinusoidal model can be used as features for stressed speech recognition. Concentration of energy within subband on logarithmic scale [4] and euclidean distance [5] is also chosen as stress classification parameters. In an other work, orthogonal projection technique is used to remove the stress component from stressed speech [6–9]. Shukla et al. [8] assumed that the speech and the stressed subspaces are orthogonal to each other when dot product is used as inner product [8]. MFCC is taken as the feature vector. Stressed speech is orthogonally projected onto some specific individual basis vector (basis corresponding to minimum norm of orthogonal complement component of stressed speech) of neutral subspace [8].

In this work, cosine angle between neutral speech and stressed speech is used as similarity measurement parameter. The basis vectors for neutral speech subspace are created by *k-mean* clustering [10, 11]. It is assumed that, speech and stress components of stressed speech are linearly related. Cosine between neutral and stressed speech is taken as coefficient of linear combination between speech and stress component of stressed speech. Also, this linearity is tested on different inner product space [12]. Experimentally, it is found that, this linearity between speech and stress component of stressed speech does not hold for all inner product spaces. All the analyses in this paper are done using speech recognition.

Speech signal is parameterized into TEO-CB-Auto-Env feature [13–15]. TEO-CB-Auto-Env feature is nonlinear. This feature, based on human speech production process is modeled by air flow pattern in the vocal tract. Teager et al. [13] emphasized that, while the vocal tract articulators do move to configure the vocal tract shape, resulting airflow property excites those models which a listener will perceive as a particular phoneme. Teager energy depends on both amplitude and oscillation frequency. Thus in case of stressed speech, this energy measure is capable to measure instantaneous changes in formant frequency.

This paper is organized as follows: Methods of similarity measurement between speech and stressed subspace is described in Sect. 13.2. Section 13.3 contains the method of extraction of nonlinear (TEO-CB-Auto-Env) feature and experimental setup. In Sect. 13.4 stressed speech recognition using similarity measurement on different inner product space is analyzed. Finally, conclusion of this paper is presented in Sect. 13.5.

13.2 Method of Similarity Measurement Between Speech and Stressed Subspace

Stressed speech is produced by any alteration of speech production system. This alteration induced additional stress component in speech produced under neutral condition. It is assumed that, additional stress component is linearly related to the neutral speech as shown in Eqs. (13.1) and (13.2);

$$\mathbf{w} = \mathbf{w}_{ne} + \mathbf{w}_{st} \quad (13.1)$$

$$\mathbf{w}_{ne} = \alpha \mathbf{w} \quad (13.2)$$

where, \mathbf{w} is the stressed speech vector. \mathbf{w}_{ne} and \mathbf{w}_{st} represent the speech and stress component of stressed speech vector, respectively. Coefficient of linear combination is represented by α . Method of determining the value of coefficient is described in subsection A. In subsection B, various inner products used for cosine similarity measurement are discussed.

13.2.1 Procedure of Determining the Coefficient of Linear Combination

The neutral speech subspace for all the words in the vocabulary of neutral speech training database is created. The basis vectors for all the neutral speech subspace are created by *k-mean* clustering. In this way, neutral speech subspace for every word in the vocabulary of neutral speech training database is created as given in Eq. (13.3);

$$\mathbf{U}^i = \{\mathbf{u}_j^i\} \quad 1 \leq j \leq k, 1 \leq i \leq m \quad (13.3)$$

where m and k denote total number of words in the vocabulary of neutral speech training database and total number of basis vectors used to span the neutral speech subspace. \mathbf{u}_j^i represents the basis vector used to span the neutral speech subspace \mathbf{U}^i . After that cosine between stressed speech vector and basis vectors of all the neutral speech subspaces is calculated as given in Eq. (13.4);

$$\cos \theta_j^i = \frac{\langle \mathbf{u}_j^i, \mathbf{w} \rangle}{\|\mathbf{u}_j^i\| \|\mathbf{w}\|} \quad (13.4)$$

θ_j^i is the angle between stressed speech vector \mathbf{w} and basis vector \mathbf{u}_j^i . The matrix \mathbf{A}^i contains the cosine between the stressed speech vector (belongs to stressed speech word containing total T number of frames) and basis vectors of neutral speech subspace \mathbf{U}^i as given in Eq. (13.5);

$$\mathbf{A}^i = \begin{bmatrix} \cos \theta_{11}^i & \cos \theta_{12}^i & \dots & \cos \theta_{1j}^i \\ \cos \theta_{21}^i & \cos \theta_{22}^i & \dots & \cos \theta_{2j}^i \\ \vdots & & & \\ \cos \theta_{T1}^i & \cos \theta_{T2}^i & \dots & \cos \theta_{Tj}^i \end{bmatrix} \quad 1 \leq i \leq m \quad (13.5)$$

Each element of vector \mathbf{b}^i contains the average value of cosine between stressed speech vector and basis vectors of neutral speech subspace \mathbf{U}^i (which is the summation of each row of \mathbf{A}^i) as given in Eq. (13.6);

$$\mathbf{b}^i = \begin{bmatrix} c_1^i \\ c_2^i \\ \vdots \\ c_T^i \end{bmatrix} \quad 1 \leq i \leq m \quad (13.6)$$

where, $c_1^i, c_2^i, \dots, c_T^i$ is the average value of cosine between stressed speech vector corresponds to first, second, and last frame (T) of stressed speech word and basis vectors of neutral speech subspace \mathbf{U}^i . The neutral speech subspace, similar to stressed speech subspace under cosine similarity is determined as,

$$i' = \underset{i}{\operatorname{argmax}} [c_1^i + c_2^i + \dots + c_T^i] \quad (13.7)$$

In this way similar neutral speech subspace $\mathbf{U}^{i'}$ is determined for stressed speech word. For any stressed speech vector, coefficient of linear combination ($\alpha = \alpha_t$ for speech frame at time instant t) between speech and stress component of stressed speech is taken as maximum value of cosine among all the cosines between stressed speech vector and basis vectors of neutral speech subspace $\mathbf{U}^{i'}$ as,

$$j' = \underset{j}{\operatorname{argmax}} [\cos \theta_{t1}^{i'} \cos \theta_{t2}^{i'} \dots \cos \theta_{tj}^{i'}] \quad (13.8)$$

$$\alpha_t = \cos \theta_{tj'}^{i'} \quad (13.9)$$

13.2.2 Overview of Inner Products Used for Orthogonal Projection

In this work, two types of inner product are tested which satisfy all conditions of inner product space. These are as follows (Fig. 13.1):

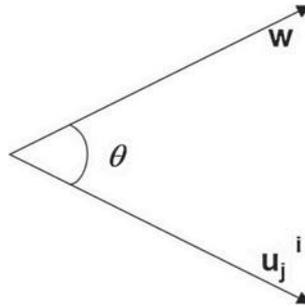


Fig. 13.1 Similarity measurement between neutral and stressed speech

- Dot product: The dot product or scalar product as inner product is an algebraic operation between two equal length vector, i.e.,

$$\langle \mathbf{u}, \mathbf{w} \rangle = \sum_{j=1}^k u_j w_j \quad (13.10)$$

where k is the length of feature vector. Dot product corresponds to euclidean norm. Algebraically, it is the sum of product of corresponding element of each vector. Geometrically, it is the product of norms of the two vectors and the cosine of angle between them.

- Unbiased covariance: unbiased covariance can be visualized as mean subtracted normalized dot product

$$\langle \mathbf{u}, \mathbf{w} \rangle = \frac{1}{k-1} \sum_{j=1}^k (u_j - \mu_u)(w_j - \mu_w) \quad (13.11)$$

where j varies from 1 to is the length of vector k .

μ_u and μ_w are the mean values of vectors \mathbf{u} and \mathbf{w} , respectively. Since the unbiased covariance between two vectors gives information about linearity between them, therefore this inner product measures the similarity between neutral subspace and stressed subspace.

13.3 Experimental Setup and Extraction Process of Nonlinear Feature

All the analyses, in this work are done by continuous speech recognition. It is important to parameterize stressed speech signal which is robust under stressed condition. In other words, it can be said that the feature which has minimum variation in prop-

Table 13.1 Content of the database

Vocabulary of the database				
Angoothi	Darwaja	Shakaahaari	Puraana	Namaste
Sahaayata	Saamaajik	Bataya	Parivartan	Dikhaayi
Kahaani	Maataji	Jaankari	Banaaya	Jaayega
Bhaavana	Kitane	Parampara	Ghatanaa	Videshi
Laagataar	Tumhaaari	Saamachar	Janvari	Majduri
Pareshaan	Saathiyo	Mariyaada	Nivedan	Nikaalkar
Mahaaraaja	Dakghar	Ticket	–	–

erties occurred between stressed and neutral condition. In this work, speech signal is parameterized into nonlinear (TEO-CB-Auto-Env) features. Subsection *A* contains description of experimental setup used for implementation. Subsection *B* described the process of extraction of nonlinear feature (Table 13.1).

13.3.1 Experimental Setup

In this work, basically the effect of stress over neutral speech is analyzed. Therefore, single word extracted from sentence spoken under stressed condition is sufficient to analyze the effect of stress over neutral speech. Four types of stressed condition are tested, namely: angry, sad, lombard, and happy, to analyze the effect of stress over neutral speech. Database contains 33 different keywords of hindi language. 15 speakers (5 females and 10 males) were used to generate the database in all stressed condition including neutral state. Neutral speech is collected when speaker is not under any emotional state. Recording is done in sound proof recording room. Stressed speech is produced by speaker under full blow of stress condition. Training database contains 495 words of neutral speech. Test database contains neutral speech and speech under stress (angry, sad, lombard, and happy) condition. Ratio of number of words in training and testing database is taken as 1/3.

Speech signal is sampled at sampling frequency 8000Hz. Small segmentation (also called framing) of speech signal is done using Hanning window with frame size 20 and 10ms overlapping of frame. Each frame is parameterized into 39 length feature vector. Speech recognizer is developed using Hidden Markov Model Toolkit (HTK). Neutral speech word in the vocabulary of training database is modelled as left-right or Bakis Hidden Markov Model (HMM) with no skip of any state. For every word, HMM has 10 states. The observation vector in each state of HMM is modeled by three Gaussian densities. Features extracted from software MATLAB are converted into HTK format. This HTK format features of speech are used in HTK to develop the HMM model. HMM model is trained by neutral speech and tested by speech under stressed as well as neutral condition.

13.3.2 Extraction Process of Nonlinear (TEO-CB-Auto-Env) Features

Nonlinear feature is derived on the assumption that speech production model is nonlinear. That is, air propagates within the vocal tract in nonlinear (combination of amplitude and frequency modulated wave) fashion. The extraction process of TEO-CB-Auto-Env feature is as follows.

The entire audible frequency range is divided into fixed number (length of feature vector) of critical band. In this work length of feature vector is taken as 39. The speech signal is filtered through a band-pass Gabor filterbank. This filterbank contains 39 band-pass Gabor filters. Gaussian modulated cosine pulse is taken as Gabor filter. The RMS bandwidth of each band-pass Gabor filter is individual critical band. In this way the bandwidth of all band-pass Gabor filters in filterbank cover entire audible frequency range. The impulse response $h_k(n)$ of the k th band-pass Gabor filter in Gabor filterbank is defined as,

$$h_k(n) = e^{\frac{-\sigma_k \pi n^2}{0.664}} \cos(\omega_k n) \quad (13.12)$$

where ω_k is the center frequency and σ_k is RMS bandwidth of the k th band-pass Gabor filter. The speech signal is filtered through Gabor filterbank in time domain as,

$$x_k(n) = s(n) * h_k(n) \quad 1 \leq k \leq 39 \quad (13.13)$$

where $*$ represents the convolution operator and $s(n)$ is input speech signal sample. $h_k(n)$ and $x_k(n)$ are the impulse response and filtered speech signal sample from the k th Gabor band-pass filter, respectively. Now there is 39 Gabor band-pass filtered speech signal. Teager Energy Operator (TEO) is applied to each Gabor band-pass filtered speech as given in Eq. (13.13),

$$\Psi(x_k(n)) = x_k^2(n) - x_k(n-1)x_k(n+1) \quad 1 \leq k \leq 39 \quad (13.14)$$

This TEO operated band-pass Gabor filtered speech signal reflects the nonlinear energy flow within the vocal tract of resonant frequency corresponding to center frequency of Gabor filter. Every Gabor band-pass filtered speech signal is segmented into small durations (framing). For every frame normalized autocorrelation is calculated. After that, area under each autocorrelation envelop from zero to maximum lag (due to symmetricity of autocorrelation function) is calculated. Normalized autocorrelation envelop area for every frame is further normalized by *frame length*/2. Normalized autocorrelation envelop area corresponding to first Gabor band-pass filter is the first element of nonlinear feature vector. Similarly, normalized autocorrelation envelop area corresponding to second Gabor band-pass filter is the second element of nonlinear feature vector, and so on. In this manner, 39 nonlinear features are obtained for each time frame (one for each frequency band).

Table 13.2 Speech recognition (in %) using nonlinear features

Stress class	Recognition (in %)
Neutral	100
Angry	48
Sad	56
Lombard	50.67
Happy	52
Average	61.334

13.4 Result Analysis

In this section, speech recognition under stressed condition is analyzed. Speech recognition using nonlinear feature (without any stress compensation) is summarized in Table 13.2. From this table, it is observed that, in neutral condition recognition performance goes to 100 %. For angry, sad, lombard, happy stress class speech, speech recognizer performance became 48, 56, 50.67, and 52 % respectively. Therefore, it can be said that, nonlinear (TEO-CB-Auto-Env) feature can extract the speech information from speech signal. So that, recognition performance of classifier reaches upto 100 % under neutral stressed condition. Also for stressed condition speech, nonlinear (TEO-CB-Auto-Env) feature is able to extract the speech information. On an average speech recognizer has 61.334 % recognition performance.

Now, speech recognition using cosine multiples of stressed speech is analyzed. This stress compensation technique is based on assumption of linear relationship between speech and stressed components of stressed speech. Speech component of stressed speech is presented in the cosine times of stressed speech vector. Recognition performance using cosine multiples of stressed speech on different inner product (dot product and unbiased covariance) spaces is summarized in paragraph in Fig. 13.2.

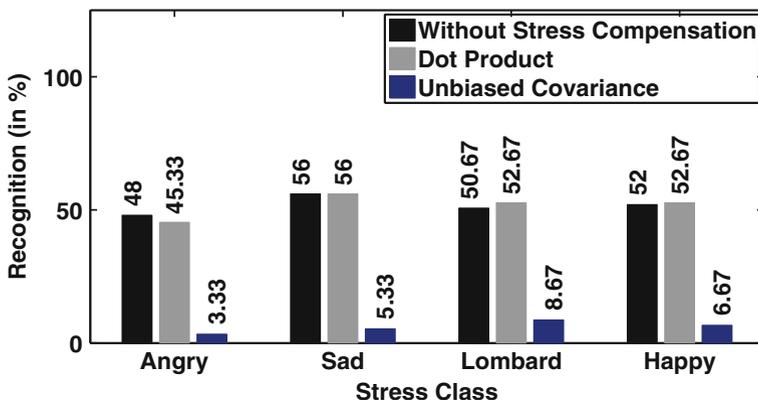


Fig. 13.2 Speech recognition using similarity measurement

- When inner product is used as dot product, speech recognition increases in lombard and happy stressed condition by 2 % (50.67–52.67 %) and 0.67 % (52–52.67 %), respectively. In sad stressed condition, percentage of speech recognition remains same 56 %. There is a slight reduction in performance of speech recognizer under angry stress class by 2.67 % (48–45.33 %).
- For unbiased covariance as inner product, recognition performance reduces for all stress class speech. In angry stress class speech, speech recognition reduces by 44.67 % (48–3.33 %). For sad stress class speech recognition percentage became very low by 50.67 (56–5.33 %). Also in lombard and happy stress condition speech, recognition percentage reduces by large amount 42 % (50.67–8.67 %) and 45.33 % (52–6.67 %), respectively.

Performance of speech recognizer increases (lombard and happy stress class), remains same (sad stress class), and slightly reduces (angry stress class) when inner product is used as dot product. Whereas, for unbiased covariance used as inner product, recognition performance reduces by large amount in all stress condition. This means that, cosine multiples of stressed speech does not contain speech information of stressed speech for all inner product spaces. The coefficient of linear combination (cosine between stressed speech and basis vector of neutral speech subspace) is varied as inner product space changes. When dot product is used as inner product, coefficient of linear combination (α) is able to extract the feature of speech properties from stressed speech. Whereas, the coefficient of linear combination (α) determined for inner product space (dot product), is not able to separate the speech component of stressed speech. In this case, cosine (between stressed speech and basis vector of neutral speech subspace), multiples of stressed speech does not contain speech information of stressed speech. Therefore, it can be concluded that, linear relationship between speech and stressed component of stressed speech exists for some specific inner product space, not for all inner product space.

13.5 Conclusion

The work presented in this paper is based on cosine similarity measurement between neutral speech subspace and stressed speech subspace. This cosine similarity measurement is tested on different inner product space. It is concluded from experiment that the speech and stressed components of stressed speech are not linearly related with respect to all inner product spaces. This linear relation between speech and stressed components of stressed speech holds for some specific inner product space. Also, it is shown that, nonlinear (TEO-CB-Auto-Env) feature is able to capture the speech information from speech signal, so that speech recognition percentage reaches upto 100 %.

References

1. Patil, S.A., Hansen, J.H.: *Speech Under Stress: Analysis, Modeling and Recognition*. Springer, Berlin (2007)
2. Chen, Y.: Cepstral domain stress compensation for robust speech recognition. In: *IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP '87*, vol. 12, pp. 717–720 (1987)
3. Ramamohan, S., Dandapat, S.: Sinusoidal model-based analysis and classification of stressed speech. *IEEE Trans. Audio Speech Lang. Process.* **14**(3), 737–746 (2006)
4. Nwe, T., Foo, S.W., De Silva, C.: Detection of stress and emotion in speech using traditional and fft based log energy features. In: *Proceedings of the 2003 Joint Conference of the Fourth International Conference on Information, Communications and Signal Processing, 2003 and Fourth Pacific Rim Conference on Multimedia*, vol. 3, pp. 1619–1623 (2003)
5. Ruzanski, E., Hansen, J., Meyerhoff, J., Saviolakis, G., Norris, W., Wollert, T.: Stress level classification of speech using euclidean distance metrics in a novel hybrid multi-dimensional feature space. In: *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2006*, vol. 1, pp. I-I (2006)
6. Loizou, P., Spanias, A.: Improved speech recognition using a subspace projection approach. *IEEE Trans. Speech Audio Process.* **7**(3), 343–345 (1999)
7. Pavan Kumar, D., Bilgi, R.R., Umesh, S.: Non-negative subspace projection during conventional mfcc feature extraction for noise robust speech recognition. In: *National Conference on Communications (NCC)*, pp. 1–5 (2013)
8. Shukla, S., Dandapat, S., Prasanna, S.: Subspace projection based analysis of speech under stressed condition. In: *World Congress on Information and Communication Technologies (WICT)*, pp. 831–834 (2012)
9. Vera, E.: The subspace approach as a first stage in speech enhancement, *Latin America transactions. IEEE (Revista IEEE America Latina)* **9**(5), 721–725 (2011)
10. Rabiner, L.R., Schafer, R.W.: *Digital Processing of Speech Signals*, vol. 19. IET, Stevenage (1979)
11. Rabiner, L., Juang, B.-H.: *Fundamentals of Speech Recognition*. Prentice-Hall, Inc., Upper Saddle River (1993)
12. Nakos, G., Joyner, D.: *Linear Algebra with Applications*. Brooks/Cole Publishing Company, Pacific Grove (1998)
13. Zhou, G., Hansen, J., Kaiser, J.: Methods for stress classification: nonlinear teo and linear speech based features. In: *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 4, pp. 2087–2090 (1999)
14. Nwe, T.L., Foo, S., De Silva, L.: Classification of stress in speech using linear and nonlinear features. In: *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '03)*, vol. 2, pp. II-9-12 (2003)
15. Zhou, G., Hansen, J., Kaiser, J.: Nonlinear feature based classification of speech under stress. *IEEE Trans. Speech Audio Process.* **9**(3), 201–216 (2001)

Chapter 14

An Analysis of Indoor Power Line Network as a Communication Medium Using ABCD Matrices Effect of Loads on the Transfer Function of Power Line

Banty Tiru, Rubi Baishya and Utpal Sarma

Abstract Power line communication is a technique that uses the available power lines as a communication media. In this paper, the power line is considered as a two-wire transmission line and modeled using the transmission, chain, or ABCD matrices. The line is simulated for different conditions commonly found in practical networks and the salient features discussed in detail. It is found that channel shows a deterministic behavior if the complete network is known a priori. However, in practical cases, discrepancies occur due to unavailability of complete information of the channel. This leads to decreased correlation or/and variable attenuation between the theoretical and experimental readings. The effects of these shortcomings on the efficiencies of discrete multitone system commonly used for power line communications are discussed.

Keywords Power line communication · Transmission matrices · Transmission line · Correlation · Attenuation

14.1 Introduction

Power Line Communication (PLC) is a technique that uses the available power line as a communication media for various purposes. Recently, the possibility to incorporate it with wireless and other wired systems like Ethernet and Multimedia over Co-Ax (MoCA) forming a heterogeneous/hybrid communication system has been studied [1]. For this, it is necessary that all the components of the system be modeled with maximum reliability. Modeling of wireless and other wired channels are

B. Tiru (✉) · R. Baishya · U. Sarma
Gauhati University, Guwahati 781014, India
e-mail: banty_tiru@rediffmail.com

R. Baishya
e-mail: arpitabaishya@gmail.com

U. Sarma
e-mail: utpalsarma@gauhati.ac.in

quite matured than their power line counterpart. This is because the channels are deterministic and nearly time invariant. However, power lines are characterized by time-variant channel conditions that offer a harsh media for data transfer [2, 3]. Researchers have tried to model the power line using different methodologies. In the top-down approach, the model parameters are estimated after the measurement of the actual transfer function. In the bottom-up approach, the same is obtained theoretically. The power line environment can be analyzed in the time domain (multipath dependent echo model) [4] or frequency domain using transmission [5] and scattering matrices [6].

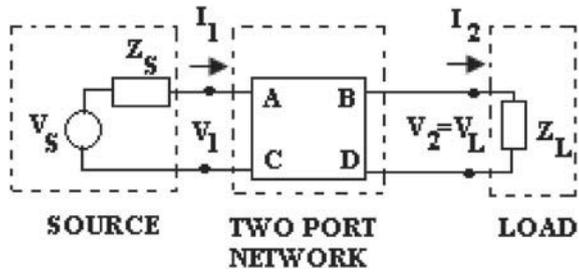
Transmission matrices provide a simple and straightforward method for simulation of complex networks compared to its time domain counterpart. Here, the power line can be considered as a two- or three-wire transmission lines (TL) depending on the absence or presence of the ground conductor (as in multiconductor theory or MTL [7]). TL models have been used extensively to study a number of variabilities in the transmission path of a power line channel [5, 8]. Increase of cable lengths and number of bridge taps increases the attenuation of the signal [5]. Decreasing the distributed inductance and capacitance decreases the number of ripples and signal attenuation in the transfer function keeping the position of the notches fixed [8]. However, in practical situations, the power line is also loaded with loads which have different characteristics. It would therefore be worthwhile to study this effect in the transfer characteristics. As such in this paper, we study the effect of loads that are capacitive, inductive, or complex in nature on the transfer function using the TL theory. As the grounding practices in India are not well defined up to the extent required by the MTL theory, the third return ground conductor is not considered and the power line is assumed to consist of only the phase and the neutral lines in a casing capping enclosure. The study is also limited to the magnitude of the transfer function (TR) only and the phase response is taken to be ideal as given by the model. Here the bottom-up methodology is used and the power line is modeled using transmission matrices which are very suitable for cascaded networks. In Sects. 14.2 and 14.3, an introduction of the transmission matrices with applicability to channel modeling and the methodology extended to power lines is given. In Sect. 14.4, the power line is loaded with different loads and the characteristics studied in detail. Section 14.5 compares the theoretical and experimental values and Sect. 14.6 tests the efficiency of the methodology for channel simulation when applied for practical system namely a discrete multitone transceiver (DMT) commonly used for PLC.

14.2 The Transmission, Chain, or ABCD Matrices

The transmission, chain, or ABCD matrix of any two port device (Fig. 14.1) is defined in terms of the input voltage (v_1), input current (i_1), output voltage (v_2), and output current (i_2) and is given by

$$\begin{bmatrix} v_1 \\ i_1 \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} v_2 \\ i_2 \end{bmatrix} \quad (14.1)$$

Fig. 14.1 The ABCD matrix of a two-port network



where,

$$A = \left. \frac{v_1}{v_2} \right|_{i_2=0} \quad B = \left. \frac{v_1}{-i_2} \right|_{v_2=0} \quad C = \left. \frac{i_1}{v_2} \right|_{i_2=0} \quad D = \left. \frac{i_1}{-i_2} \right|_{v_2=0}$$

The overall transmission matrix of a complex network is equal to the product of the matrices of the cascaded portions. Knowledge of the ABCD matrices enables the TR of complex networks to be obtained and is given by [10]

$$H(f) = \frac{Z_L}{AZ_L + B + CZ_S Z_L + DZ_S} \tag{14.2}$$

Here, Z_S and Z_L are the source and load impedances of the network. The ABCD matrices of sections of a TL are given in Table 14.1. Here, Z_0 and γ are the char-

Table 14.1 ABCD Matrices of sections of transmission line [9]

Sections	ABCD Matrices	Models
Transmission line	$\begin{bmatrix} \cosh \gamma l & Z_o \sinh \gamma l \\ 1/Z_o \sinh \gamma l & \cosh \gamma l \end{bmatrix}$	
Bridged taps	$\begin{bmatrix} 1 & 0 \\ 1/Z_{in} & 1 \end{bmatrix}$; $Z_{in} = Z_o \frac{Z_{br} + Z_o \tanh(\gamma l_{br})}{Z_o + Z_{br} \tanh(\gamma l_{br})}$	
Open Circuits	$\begin{bmatrix} 1 & 0 \\ \tanh \gamma l_{br} / Z_o & 1 \end{bmatrix}$	
Short Circuits	$\begin{bmatrix} 1 & 0 \\ \coth \gamma l_{br} / Z_o & 1 \end{bmatrix}$	
Series Impedance (Z)	$\begin{bmatrix} 1 & Z \\ 0 & 1 \end{bmatrix}$	
Shunt Impedance (Z)	$\begin{bmatrix} 1 & 0 \\ 1/Z & 1 \end{bmatrix}$	
Transformers N=no. of turns	$\begin{bmatrix} N & 0 \\ 0 & 1/N \end{bmatrix}$	

acteristics impedance and propagation constants of the line, respectively. Z_{in} is the input impedance of the bridged taps loaded with load Z_{br} through a length l_{br} .

14.3 Power Line Modeling of the Test Channels

The network where the simulation is done consists of a section of power line ‘AB’ of length $l(= 37.1 \text{ m})$ with a bridged tap ‘CD’ of length l_{br} placed at a distance l_1 from ‘A’ (Fig. 14.2). ‘A’ acts as an inlet plug for a transmitter sending signal to a receiver at ‘B’. Both are assumed to be connected to the plugs via couplers constituted of the capacitors and transformers ($c1, T1$) and ($c2, T2$). In actual PLC system, the couplers act as high-pass filters that prevent the 50/60 Hz, 230 V high voltage from harming the connected equipments. For simulation, a number of steps are taken. First, the primary line constants namely the resistance per unit length (R), inductance per unit length (L), capacitance per unit length (C), and conductance per unit length (G) of the wires used for practical household wiring [9] are found out using Eq. 14.3.

$$R = \frac{1}{\pi} \sqrt{\frac{\pi f \mu_c}{\sigma_c}} \Omega/\text{m}; L = \frac{\mu}{\pi} a \cosh\left(\frac{D}{2a}\right) \text{H}/\text{m};$$

$$C = \frac{\pi \epsilon}{a \cosh\left(\frac{D}{2a}\right)} \text{F}/\text{m}; G = \frac{\pi \sigma_{diel}}{a \cosh\left(\frac{D}{2a}\right)} \text{mho}/\text{m} \tag{14.3}$$

Here, a is the radius of the conductor of a single wire, D is the distance between the phase and neutral lines, μ_c and σ_{cond} are the permeability and conductivity of the conducting medium, respectively, and $\sigma_{diel}, \epsilon,$ and μ are the conductivity, permittivity, and permeability of the dielectric material between the conductors.

Next, to account for the decrease of effective area of current flow arising because of the multistranded nature of power lines, R is replaced by the corrected value $R_{corrected} = R/X_R$; where the ratio of the effective area to the total area is given by [10].

$$X_R = \frac{\cos^{-1}\left[\frac{r-\delta}{r}\right] r^2 - (r-\delta)\sqrt{r^2 - (r-\delta)^2}}{2r\delta} \tag{14.4}$$

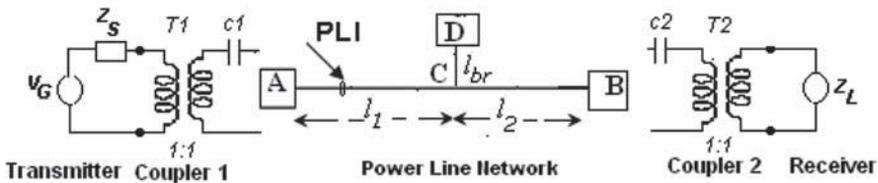


Fig. 14.2 The power line network with communication system

Here, δ is the skin depth of the conductor. C and G are also changed to C_{total} and G_{total} , respectively, as the communication signals are not only coupled between the phase and neutral lines, but also from the live cable to the earth cable and then to the neutral cable [11].

$$C_{total} = C + \frac{C}{2}; G_{total} = G \cdot \frac{C_{total}}{C} \quad (14.5)$$

The characteristics impedance and the propagation constants of the wires are evaluated using $Z_0 = \sqrt{(R + j\omega L)/(G + j\omega C)}$ and $\gamma = \sqrt{(R + j\omega L)(G + j\omega C)}$, respectively. It is found that $L = 0.69 \mu\text{H/m}$, $G = 0.018 \mu\text{mho/m}$, $C = 38 \text{ pf}$, and $Z_0 = 135 \Omega$ for the wire used. Finally, the TR of 'AB' is found out using Eq. 14.2 from the product of the ABCD matrices of the cascaded sections as given in Sect. 14.4 for different conditions using Z_s and Z_L both equal to 50Ω .

14.4 Analysis of Transfer Function of Power Line Channels

14.4.1 Open and Short Circuits

First, the branch CD is assumed to be open or short circuited. The TR shows deep notches dependent on the length of l_{br} (Table 14.2 and Fig. 14.3a). When an electromagnetic wave reaches an open circuit (or a short circuit), the signal is completely reflected with a voltage maximum or current minimum (or voltage minimum or current maximum) at the end. The reflected voltage will be in phase (or opposite phase) with the incident signal. Notches are found when the path difference between the direct ($A \rightarrow C \rightarrow B$) and the reflected signal ($A \rightarrow C \rightarrow D \rightarrow C \rightarrow B$) reaching the load is such that the phase difference completely nullify each other. For very small lengths of the open circuit, the path difference is very small and the minimum condition of notch formation is not satisfied. A change in l_1 or l keeps the position of the notches unchanged but the attenuation varies from a maximum to a minimum. Figure 14.3b shows the attenuation of a notch at 10.7 MHz for $l_{br} = 4.55 \text{ m}$ when l_1 or l is varied through 30m. The corresponding TR is shown in Fig. 14.3c. Ripples occur because the 135Ω is weakly resonant when terminated by a load of 50Ω .

14.4.2 Capacitive and Inductive Loads

Next, CD is loaded by a capacitive load C_{load} . For large values (Table 14.3, 1–3), the impedance is very small and the TR's are characteristics of short circuits. The additional notch occurs at the frequency of resonance (f_{res}) of the series LCR circuit constituted by the net inductance (L_{net}), capacitance (C_{net}), and resistance (R_{net}) of the wire as seen at the input of the branched circuit and given by $f_{res} = 1/2\pi\sqrt{L_{net}C_{net}}$. For small values (Table 14.3, 4–6), the circuit acts like open

Table 14.2 Position of notch for different lengths of line for an open circuit (1–10) and short circuit (11–16) with $l_1 = 6.35$ m, $l_2 = 30.75$ m, $l(= l_1 + l_2) = 37.1$ m

Sl no	l_{br} (m)	Position of Notches (MHz)			Path Length			
		1 st	2nd	3rd	A-B-C	A-B-D-B-C	Path difference for consecutive notches	
							m	
1	1.55	--	--	--	37.1	40.3	3.1	--
2	2.55	19.2	--	--	37.1	42.3	5.1	$\lambda/2$
3	3.55	13.8	--	--	37.1	44.3	7.1	$\lambda/2$
4	4.55	10.7	--	--	37.1	46.3	9.1	$\lambda/2$
5	5.55	8.8	26.4	--	37.1	48.3	11.1	$\lambda/2, 3\lambda/2$
6	6.55	7.5	22.4	--	37.1	50.3	13.1	$\lambda/2, 3\lambda/2$
7	7.55	6.5	19.5	--	37.1	52.3	15.1	$\lambda/2, 3\lambda/2$
8	8.55	5.7	17.1	28.6	37.1	54.3	17.1	$\lambda/2, 3\lambda/2, 5\lambda/2$
9	9.55	5.1	15.3	25.6	37.1	56.3	19.1	$\lambda/2, 3\lambda/2, 5\lambda/2$
10	10.55	4.6	13.9	23.3	37.1	58.3	21.1	$\lambda/2, 3\lambda/2, 5\lambda/2$
11	1.55	--	--	--	37.1	42.3	3.1	--
12	3.55	27.5	--	--	37.1	44.3	10.7	$3\lambda/2$
13	5.55	17.6	--	--	37.1	48.3	16.80	$3\lambda/2$
14	6.55	14.9	29.7		37.1	50.3	19.9, 16.6	$3\lambda/2, 5\lambda/2$
15	8.55	11.4	22.9		37.1	54.3	25.7, 21.42	$3\lambda/2, 5\lambda/2$
16	10.55	9.2	18.5	27.8	37.1	56.3	31.86, 26.4, 24.6	$3\lambda/2, 5\lambda/2, 7\lambda/2$

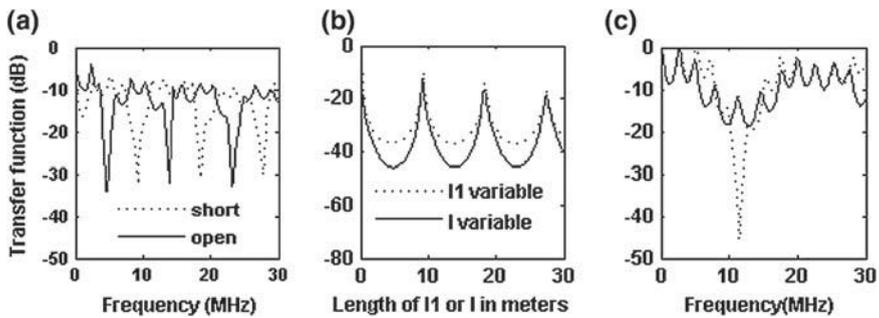


Fig. 14.3 TR of open and short circuits. **a** TR for open and short circuit for $l_{br} = 10.55$ m. **b** TR at 10.7 MHz for $l_{br} = 4.55$ m, for variables l_1 and l . **c** TR for $l_{br} = 4.55$ m showing maximum and minimum attenuation

Table 14.3 Position of the notch with capacitors and inductance as loads with $l_{br} = 5.55$ m

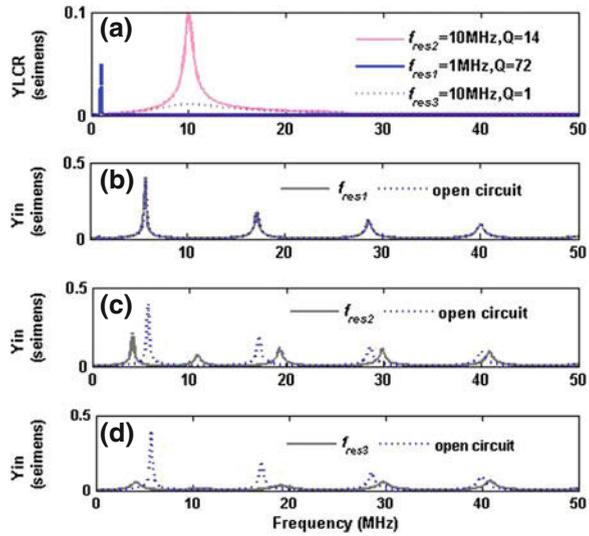
Sl no	C_{load} or L_{load}	Notches (MHz)			Characteristics of Resonant circuit			
		1st	2nd	3rd	L_{net} μ H	C_{net} $\times 10^{-8}$ F	f_{res} MHz	Inference
1	10^{-1} μ F	0.26	17.6	--	3.83	10.04	0.26	short
2	10^{-2} μ F	0.80	17.6	--	3.83	1.004	0.80	
3	10^{-3} μ F	2.5	17.9	--	3.82	0.142	2.16	
4	10^{-4} μ F	--	6.1	20.5	3.83	0.005	3.56	open
5	10^{-5} μ F	--	8.4	25.6	3.83	0.004	3.91	
6	10^{-6} μ F	--	8.8	26.2	3.83	0.042	3.96	
7	10^{-6} H	--	14.3	30.1	8.6	0.02	3.72	open
8	10^{-1} H	--	8.8	26.5	10^5	0.02	0.04	
9	1H	--	8.8	26.4	10^6	0.02	0.01	
10	10H	--	8.8	26.4	10^7	0.02	0.003	
11	100H	--	8.8	26.2	10^8	0.02	0.1	

circuit as expected. For inductive loads L_{load} , the impedance is high and the TR's are characteristics of open circuits. The minimum value of L_{net} is however limited by the primary line constant L and the circuit acts always like an open circuit (Table 14.3, 7–11) for various values of L_{load} .

14.4.3 Complex Loads

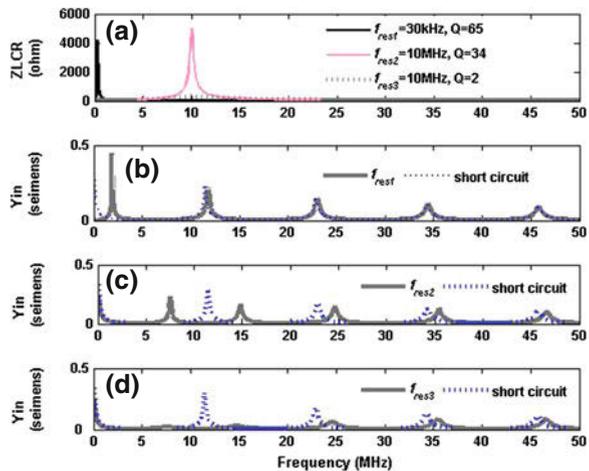
Impedance of practical loads may act as a series or parallel LCR circuits. Series LCR circuits give an impedance (Z_{LCR}) minima or admittance Y_{LCR} maxima at resonance. When the tapped bridge is loaded by a series LCR circuit, the positions of the notches are not only dependent on l_{br} , but also on the resonant frequency f_{res} and the Q value of the circuit. Figure 14.4 gives an example of the effect of series LCR circuit connected as a load on the bridge CD of length 8.55 m. In Fig. 14.4a, Y_{LCR} is plotted as a function of frequency for different values of f_{res} and Q. It must be noted that notches are found where the input admittance $Y_{in} = 1/Z_{in}$ of the bridged tap is a maximum so as to offer an easy way for the signal. For low resonant frequency f_{res1} and high Q, Y_{in} is same like that found for open circuits (Fig. 14.4b). This is because at off resonance, the impedance is very high and so the circuit acts like an open circuit. For high values like f_{res2} and medium Q, the values are shifted from those for open circuits (Fig. 14.4c). When Q is small as in Fig. 14.4d, the attenuation of the notches nearly disappear as the bridge tap offers uniform impedance over entire bandwidth.

Fig. 14.4 **a** Absolute value of Y_{LCR} of series resonant circuit with different resonant frequencies f_{res1} , f_{res2} , and f_{res3} , quality factor Q and **b, c, d** Absolute value of Y_{in} at the input of the branched tap for open circuits for f_{res1} , f_{res2} , and f_{res3} , respectively



Parallel LCR circuits give impedance maxima or current minima at resonance. For low resonant frequency f_{res1} and high Q , Y_{in} is same like that found for short circuits (Fig. 14.5b). For high f_{res2} and medium Q , the notches are shifted from those for open circuits (Fig. 14.5c). When Q is small as in Fig. 14.5d, the attenuation of the notches decreases as the bridge tap offers uniform impedance over entire bandwidth.

Fig. 14.5 **a** Absolute value of Z_{LCR} of parallel resonant circuit with different resonant frequencies f_{res1} , f_{res2} , and f_{res3} , quality factor Q and **b, c, d** Absolute value of Y_{in} at the input of the branched tap for short circuits for f_{res1} , f_{res2} , and f_{res3} , respectively



14.5 Comparison with Experimental Circuit

The modeling of power line channels is tested experimentally using a signal generator as a transmitter in plug A and a Digital Storage Oscilloscope (Tektronix TDS 2,022, 200MHz, 2 GS/s) as receiver at Plug B (Fig. 14.2). The couplers have cutoff frequencies of 1 kHz each. An arrangement is also made to connect/ disconnect the test network from the 230 V, 50Hz main line with the help of the power line inlet plug (PLI). In the PLI-OFF condition, the test network is disconnected from the main line of the building and the configuration of the network is completely specified in terms of known parameters. In the PLI-ON condition, the network is connected to the main line and the configuration is not known completely due to unknown loads via the distribution box. Figure 14.6 shows the TR as predicted theoretically and that obtained experimentally for open circuits and capacitive load. It is seen that for PLI-OFF state, there is 82–93 % correlation between the simulated and practical TR. However, in the PLI-ON state the correlation decreases as expected. It is also observed that the practical TR has more attenuation than the theoretical one. This is more so in the PLI-ON state. The model is tested experimentally for a maximum frequency of 10MHz.

14.6 Effect of Correlation and Attenuation on System Simulation

Discrepancies of the experimental TR from that predicted by theory are based on two variables: correlation factor and attenuation. Correlation is a measurement of the similarity between two sequences having a value between +1 and -1. A value of 1 depicts positive correlation and a -1 depicts negative correlation. Attenuation is the decrease of the measured values from those predicted. Simulation shows that a

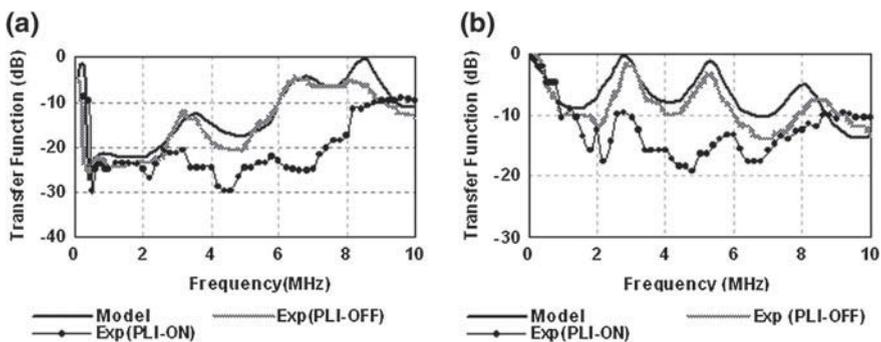


Fig. 14.6 Comparison between theoretical and experimental transfer function for **a** open circuit and **b** $C_{load} = 0.1\mu\text{F}$ and $l_{br} = 2.55\text{ m}$

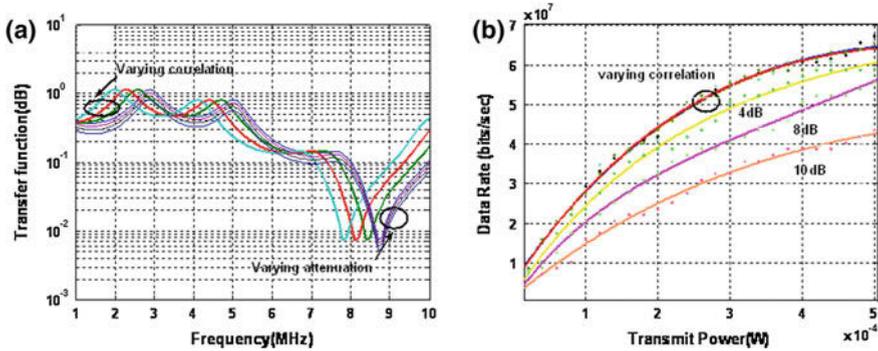


Fig. 14.7 Effect of correlation due to variations in L and C and attenuation due to variations in R on the **a** transfer function and **b** data rate of a DMT system

change in L and C leads to lack of correlation between the obtained TR (Fig. 14.7a). Similarly, change in R leads to variable attenuation. Thus the decreased correlation and increased attenuation in the experimental results may be accounted due to the error in the measurements in line parameters. The effects of these errors are studied on the efficiency of a DMT system in a simulated channel. It is found that the efficiency is more dependent on the attenuation than on the correlation (Fig. 14.7b). However, inclusion of additional notches than that predicted due to gross deviations of the parameters is expected to give drastic changes in the results of interest.

14.7 Conclusion

In this work, the power line network is modeled using the transmission matrix and the salient features studied in detail for different loaded conditions. The results are tested experimentally in a casing-capping wiring technique commonly used in India. Though this provides a most unideal condition with variable distances between the two wires, there is a high correlation between the theoretical and the practical transfer function. Increased attenuation of the later may be due to error in the measurement of resistance per unit length or improper wiring. However at certain frequencies, the experimental values are same to the modeled for different conditions (like that at 9 MHz in Fig. 14.6) which has to be studied in detail in future work. As the model requires the initial knowledge of the specifications of the wires used for practical wiring, it necessitates the values to be obtained with minimum error. Error in the measurements propagates to the variations of the simulated transfer function. Discrepancies between simulated and observed values lead to variation of predicted efficiencies of communication system tested. In the simulation of actual system, these discrepancies must be kept in mind and suitable steps must be taken into account for these errors.

References

1. Salvadori, F., Gehrke, C.S., de Campos, M., Oliveira, A.C.M., Sausen, P.S.: Smart grid infrastructure using a hybrid network architecture. *IEEE Trans. Smart Grid* **4**(3), 1630–1639 (2013)
2. Biglieri, E., Torino, P.: Coding and modulation for a horrible channel. *IEEE Commun. Mag.* **41**(5), 92–98 (2003)
3. Tiru, B., Boruah, P.K.: Multipath effects and adaptive transmission of indoor power line background noise. *Int. J. Commun. Syst.* **23**, 63–76 (2010)
4. Zimmermann, M., Dostert, K.: A multipath signal propagation model for power line channel in the high frequency range. In: 3rd International Symposium on Power-Line Communications and its Applications, pp. 45–51 (1999)
5. Khan, S., Salami, A.F., Lawal, W.A., Alam, Z., Hameed, S.A., Salami, M.J.E.: Characterization of indoor power lines as data communication channels experimental details and results. *World Acad. Sci. Eng. Tech.* **2**, 543–548 (2008)
6. Wei, S., Wang, H., Chen, F., Chen, Y., Liu, W.: Innovative model for underground coal mine power line communication based on wavelet scattering parameter matrix. In: 2nd International Symposium on Intelligent Information Technology Application, pp. 733–737 (2008)
7. Galli, S., Banwell, T.: A novel approach to the modeling of the indoor power line channel part I: circuit analysis and companion model, pp. 655–653 (2005)
8. Hossain, E., Khan, S., Ali, A.: Modeling low voltage power line as a data communication channel. *Int. J. Electr. Robot. Electron. Commun. Eng.* **2**(9), 79–83 (2008)
9. Ludwig, R., Bretchko, P.: *RF Circuit Design-Theory and Applications*. Pearson Education, New Delhi (2000)
10. Dickinson, J., Nicholson, P.J.: Calculating the high frequency transmission line parameters of power cables. In: 1st International Symposium on Power Line Communications and Applications, p. 127 (1997)
11. Meng, H., Chen, S., Guan, Y.L., Law, C., So, P.L., Gunawan, E., Lie, T.T.: Modeling the transfer characteristics for broadband power line communication channel. *IEEE Trans. Power Deliv.* **19**(3), 1057–1064 (2004)

Chapter 15

Online and Offline Handwriting Recognition of Assamese Language—A Preliminary Idea

Keshab Nath and Subhash Basishtha

Abstract Recognition of Online handwriting and machine-printed text is very hard and complex task. In this paper we are discussing some novel approaches for online handwriting and machine-printed text recognition in Assamese language. Assamese language is so much cursive unlike English and some other languages. So to recognize such cursive handwriting we have to follow several steps. In the twenty-first century everything is globalized, so Assamese language needs also to be, so that people can communicate or exchange their idea using Assamese language too. E.g., Sending email/SMS in Assamese, searching Assamese books content via search engine, writing diary using Assamese, etc. So keep in mind these things here in this paper we are discussing some approaches to succeed in the above-mentioned task. Here first we are discussing about the structure of Assamese language, then we discuss some idea for recognition of Online and Offline handwriting.

Keywords Online · Offline · Line segmentation · Word segmentation · HMM · SVM

15.1 Introduction

Assamese language structure is closely related to Bangla and Oriya. It is mostly used by the people of Eastern India (Assam).

Significant research work is done on optical character recognition (OCR) in the past few years both for online and offline. But very few research works are done for recognition of handwriting character on Indian script and for Assamese language it is almost negligible. Recognition of handwritten character is not an easy task due to enormous variabilities in people handwriting styles. So if it is possible to develop

K. Nath (✉) · S. Basishtha
Assam University, Silchar 788011, India
e-mail: keshabnath@live.com

S. Basishtha
e-mail: subhash.cse08@gmail.com

Fig. 15.3 Assamese composite characters



15.2 Recognition Phases of Online Handwriting

Online handwriting is considered as the most convenient way of communication. There are a number of techniques [1, 2] available for online handwriting recognition. However the most used technique includes basically four phases, they are:

- Data Collection
- Preprocessing
- Feature Extraction
- Recognition

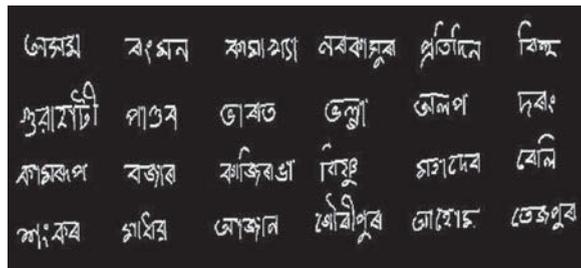
15.2.1 Data Collection

Data collection [3] is the initial stage of online handwriting recognition. In this stage handwritten samples are collected by using electronic tablets or digitizers which capture the sequences of consecutive points on the x-y plan. A special pen is used for writing on these devices. Figure 15.4 shows some collected Assamese word data. Most of the digital pen includes two actions namely pen down and pen up.

15.2.2 Preprocessing

The data collected by the digital pad need to be preprocessed [4] before recognition for better accuracy. Because at the time of user input through digicom or digitizer

Fig. 15.4 Data collection



tablets some unnecessary coordinates may be present which will be of no use toward recognition process. This may occur due to several reasons like limitations of the digital device or fault of the writer, which leads to decrease the recognition accuracy. Some examples of such distortions are variable size, random distance from each neighboring points, curvy handwriting, missing points due to fast handwriting of the writer. These noises which are found in the input data are removed in this phase. The preprocessing phase of online handwriting recognition includes commonly five steps, [4, 5] namely:

- Normalization of size
- Centering of the text
- Smoothing the points
- Duplicate Point Removal
- Equidistance Resampling of points

Size normalization is required to ensure that each character has the same height. The input stroke varies in size and depends upon how the writer moves the pen on the writing pad at the time of writing. So the handwriting sample may be of different size. This phase is necessary to rescale the sample data collected by the digitizer. In the centering step, the sample data text is reallocated at a constant from the origin of the x-y plane.

Smoothing of the collected stroke is necessary for the removal of the unwanted coordinate points which is also known as noise. This unwanted points are needed to be removed from the handwriting sample in order achieve higher recognition accuracy. Here we replace each point $x(i)$ by calculating the average of its preceding and succeeding points. Equation 15.1 shows the smoothing operation.

$$x(i) = \frac{(x(i-1) + x(i) + x(i+1))}{3} . \quad (15.1)$$

Here $x(i)$ is the i th point of input data, $x(i-1)$ is its preceding point and $x(i+1)$ is the succeeding point. Figure 15.5 shows an example of smoothing.

After the smoothing phase some points are generated having the same value; which are of no use. So these points need to be removed from the captured ink. Every writer writes in different speed. Different writer has different writing style, some may be fast some may be slow; due to this reason each collected sample may

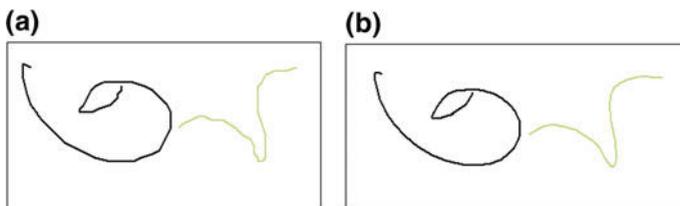


Fig. 15.5 Examples of smoothing. **a** Before. **b** After

contain different no. of points. That is why we need to obtain a constant number of points in hand writing samples. That is why we need to do resampling of point. This step is done to obtain a constant number of points in handwritten data.

15.2.3 Feature Extraction

Feature extraction [1, 4] is considered to be the most important phase of handwriting recognition process. A feature is the special characteristic of a thing which can distinguish it from others clearly. So selection of good and significant feature is very much necessary to build a better recognizer. We should be very careful of selecting the features of online data samples as based on this classifier will distinguish the character classes. However, there is no hard and fast rule for selecting features. A standard method for vertical position of a point, curvature, pen up/pen down, direction of the writing, aspect, slope are the various features extracted in [4]. The main aspect of handwriting recognition process is based on the good feature set selection, reasonably invariant with respect to shape variations due to various writing styles [6].

15.3 Offline Handwriting Recognition

The main task in offline handwriting recognition is word recognition and character recognition. We need to perform some necessary preliminary steps to recognize Assamese handwriting.

- Preprocessing
- Segmentation
- Representation
- Training and recognition

15.3.1 Preprocessing

For recognition purpose it is necessary to perform some preprocessing operation on several scanned documents. Preprocessing basically involved the following operations such as: Thresholding, gray scale image to binary image, removal of noise present in the document, segmentation of each line from the document, separation of individual word from text line, character segmentation, and separation of individual character.

15.3.2 Thresholding

In order to increase the processing speed and to reduce storage space thresholding is done on color image. The importance of thresholding is to separate areas from each other (extract the foreground (ink) from the background) by using the information present on a histogram of an image. The histogram of a gray scale image gives two peak values—high and low, high peak values represent the white background and smaller peak values represent foreground.

15.3.3 Noise Removal

The noise present in typed or machine printed document are generally created by those instrument itself through which documents are created (optical scanner, Writing instrument, etc.) and it may cause disconnected line segmentation, gaps in between line, etc. There are several noise removal techniques available and they are categorized as [7].

15.3.4 Line Segmentation

For segmentation purpose there are several approaches available for line [7], word, and character segmentation. Applying horizontal projection method we can compute corresponding horizontal histogram (Fig. 15.6 shows the horizontal histogram) of an image, based on the peak/low points of the histogram, each and every row of a document image can be separate individually. Figure 15.7 represents the line segmentation on a printed document. This similar operation is more difficult in case of handwritten document (Fig. 15.8 represents the horizontal histogram for handwritten document), because deviation may be present in lines and two lines may be close or overlap.

So to overcome this problem, one method [8] suggested an efficient technique. According to this method, while writing something most of the people follow an

Fig. 15.6 Horizontal histogram on printed document

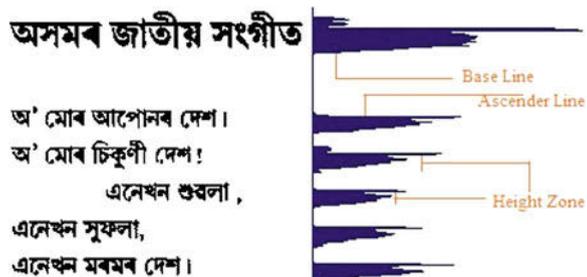
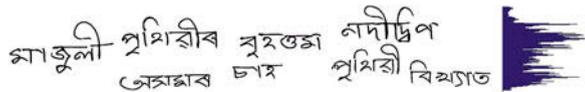


Fig. 15.7 Line segmentation on printed document



Fig. 15.8 Horizontal histogram for handwritten lines



imaginary line and upon which each word of the line resides. Based on local minima point from each and every component, this imaginary line is measured and using clustering technique all minima are grouped together to recognize the different handwriting lines.

15.3.5 Word and Character Segmentation

Word segmentation method [7, 9, 10] assume that the gaps between the word is always larger than the gaps between the characters. So if we perform vertical projection to compute corresponding vertical histogram (Fig. 15.9 shows the vertical histogram for a text line) and depending on the threshold distance (D) between two consecutive zero break of the histogram we can determine the words [11]. In case of character segmentation there are several techniques proposed, one of the method [12] gives an idea about headline or matra detection presented in Assamese or Bangla script. Figure 15.10 represents the head line or matra present in Assamese script. Once we have detected the headline than we can erase the head line and performing vertical projection we may separate character. Figure 15.11 gives the complete idea about character segmentation.

Fig. 15.9 Word segmentation

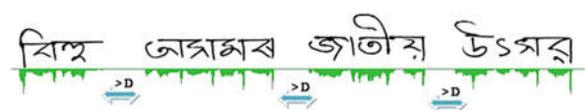


Fig. 15.10 Headline detection



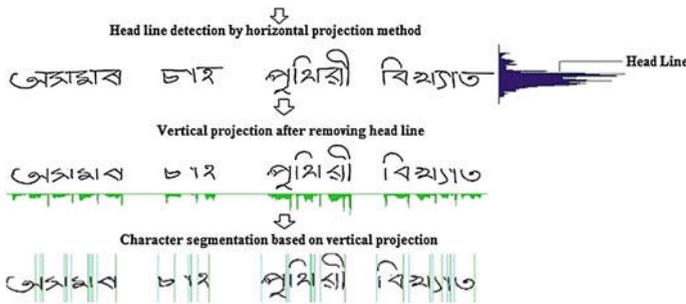


Fig. 15.11 Segmentation process

But sometimes head line detection is not so easy task specially when writing styles with leading and trailing ligatures. For this reason normalization of subsequent features value gets affected. To overcome this problem simple heuristics is used to detect headline first and after that feature is computed [6].

15.4 Recognition (Online and Offline)

Recognition process is the major phase of any character recognition system, where character recognition system uses the methodologies to assign some unknown sample into a previously defined class. Some of best-known methods are briefly described below.

15.4.1 Neural Network

In neural networks (NN) [13, 14] all interconnected “neural” processors are connected parallelly and due to this computing architecture it has a higher performance rate compared to the other approaches. Neural network can adapt the data changes easily and the various properties of the input signal.

Neural network has some limitation. The data or parameters used in NN must all be defined at the beginning. If at the time of processing there is extra additional data which is not present at the beginning time, then it becomes impossible to do, developers need to start again the whole process. This nature makes NN not suitable for those system where data are collected in time series.

15.4.1.1 Support Vector Machine

Support Vector Machine (SVM) is a powerful classifier with good generalization performance compared to NN also. SVM has been broadly used in recent years, almost for all pattern recognition system. SVM considers [15] experimental data as well as structural behavior to produce high generalization performance. In [16] a recognition technique based on SVM has been discussed.

Apart from using single classifiers for recognition, combination of two or three classifiers, i.e., hybrid classifiers are also used for high performance and accuracy. The hybrid HMM-MLP approach is the combination of Hidden Markov Model (HMM) and multilayer perceptron classifier.

15.5 Future Works

In this paper we have discussed the basic phases and approaches of online and offline handwriting recognition and how it could be implemented in Assamese script. However each approach has its own advantages and disadvantages. According to the best of our knowledge we believe that this paper will be the first attempt toward automatic recognition of online and offline handwriting recognition in Assamese language. In future we will try to develop one convenient recognition system in Assamese Script.

References

1. Parui, S.K., Guin, K., Bhattacharya U., Chaudhari, B.B.: Online handwritten Bangla character recognition using HMM. In: Proceedings of 19th International Conference on Pattern Recognition (2008)
2. Tagougui, N., Kherallah, M., Alimi, A.M.: Online Arabic handwriting recognition: a survey. *Int. J. Doc. Anal. Recognit.* **16**, 209–226 (2013)
3. Mondal, T., Bhattacharya, U., Parui, S.K., Das, K.: Online handwriting recognition of Indian scripts the first benchmark. In: Proceedings of 12th ICFHR, pp. 200–205 (2010)
4. Bharath, A., Madhavath, S.G.: HMM-based lexicon-driven and lexicon-free word recognition for online handwritten Indic scripts. *IEEE Trans. Pattern Anal. Mach. Intell.* **34**(4) (2012)
5. Bandyopadhyay, A., Chakrabati, B.: Development of online handwriting recognition system: a case study with handwritten Bangla character. In: IEEE Transactions World Congress on Nature and Biologically Inspired Computing, pp. 514–519 (2009)
6. Bhattacharya, U., Shridhar, M., Parui, S.K.: On recognition of handwritten Bangla characters. In: ICVGIP 2006. LNCS, vol. 4338, pp. 817–828. Springer, Berlin (2006)
7. Plamondon, R., Srihari, S.N.: On-line and off-line handwriting recognition: a comprehensive survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **22**(1) (2000)
8. Kim, G., Govindaraju, V., Srihari, S.N.: An architecture for handwritten text recognition systems. *Int. J. Doc. Anal. Recognit.* **2**, 37–44 (1999)
9. Mahadevan, U., Nagabhushanam, R.C.: Gap metrics for word separation in handwritten lines. In: Proceedings of Third International Conference Document Analysis and Recognition, pp. 124–127, Montreal, (ICDAR '95), August 1995

10. Seni, G., Cohen, E.: External word segmentation of off-line handwritten text lines. *Pattern Recognit.* **27**(1), 41–52 (1994)
11. Mashiyat, A. S., Hasan Talukder, M. K.: Bangla off-line handwritten character recognition using superimposed matrices. In: *Proceedings of 7th ICCIT (2004)*
12. Basu, S., Chaudhuri, C., Kundu, M., Nasipuri, M., Basu, D.K.: Segmentation of offline handwritten Bengali script. In: *Proceedings of 28th IEEE ACE, Science City, Kolkata*, pp. 171–174, December 2002
13. Arica, N., Yarman-Vural, F.T.: An overview of character recognition focused on off-line handwriting. *IEEE Trans. Syst., Man, Cybern. Part C: Appl. Rev.* **31**(2) (2001)
14. Sanmorino, A., Yazid, S.: A survey for handwritten signature verification. In: *Proceedings of IEEE Trans International Conference on Uncertainty Reasoning and knowledge Engineering*, pp. 54–57 (2012)
15. Ahmad, A.R., Khalid, M., Gaudin, C.V., Poisson, E.: Online handwriting recognition using support vector machine. In: *Proceedings of IEEE Region 10 Conference*, vol. 1, pp. 311–314 (2004)
16. Bahlmann, C., Haasdonk, B., Burkhardt, H.: On-line handwriting recognition with support vector machines—a kernel approach. In: *Proceedings of IWFHR*, vol. 02 (2002)

Chapter 16

Speaker Verification for Variable Duration Segments and the Effect of Session Variability

Rohan Kumar Das and S.R.M. Prasanna

Abstract With the current advancements achieved in the area of speaker verification, a significant performance is obtained under sufficient data conditions. Whereas when there comes a constraint in the amount of speech data, it reflects directly on the performance. This paper presents initial speaker verification studies under variable duration test segments over a standard canned database and, then, studies for variable duration test segments over a database collected from a practical speaker verification system. The latter case helps to explore session variability issues and its impact on speaker verification. This information is used for remodeling of the enrolled speaker models, which in turn improves the system performance significantly.

Keywords Speaker verification · Duration · Session variability

16.1 Introduction

Speaker recognition is the task of recognizing a person from his/her speech. Broadly, it can be classified as speaker verification (SV), which is the task of verifying a person's identity claim and speaker identification (SI), which is the task to identify which test speech belongs to which speaker. The basic difference between the two is, in verification there is always a claim associated with the test speech. Based on content of speech data, SV can be classified as text-dependent and text-independent SV [1]. In text-dependent SV the same text is used during training and testing sessions. Whereas in text-independent there is no such constraint on what is to be spoken by the speakers. This made to have just a few seconds of fixed phrase as data requirements in the former case, whereas the latter requires sufficient amount of data for better verification accuracy. But when there is some constraint kept on amount of speech data for text-independent system, the performance degrades drastically. This paper

R.K. Das (✉) · S.R.M. Prasanna
Indian Institute of Technology Guwahati, Guwahati
781039, India
e-mail: rohankd@iitg.ernet.in

S.R.M. Prasanna
e-mail: prasanna@iitg.ernet.in

focuses on text-independent SV studies for variable duration test segments and the significance of session variability for such a practical deployable system.

In recent times, the i-vector [2]-based SV system has outperformed other existing SV systems, which has made this as the state-of-the-art system for text-independent SV. Studies [3, 4] have shown that the performance for short utterances fails poorly as the test data is reduced below 10 s. Most studies on variable duration test segments have been reported mainly over some canned database. In case of some practical deployed system these studies are not much explored. Session variability has greater impact on speaker recognition, for which in i-vector-based SV system channel/session compensation is done on top of the i-vectors to get improved performance. However, for a practical deployable system this session variability can help in improved modeling, as the data from the speakers are collected over a period of time.

This paper concentrates to study the effect of amount of test data on system performance, when it is varied from sufficient data conditions to very small duration segments. The initial studies are done on NIST speaker recognition evaluation (SRE) 2003 database [5] to see its basic trend. Similar studies are then made on a database collected from a deployed SV system with attendance as an application. Session variability issues on the data obtained from practical deployable system are explored and then remodeling of the enrolled speakers on the practical system are made by pooling some of the data given by them during few testing. This gave a significant improvement on later testing cases while retrained models were used.

The rest of the paper is organized as follows: Sect. 16.2 describes the development of i-vector-based SV along with channel/session compensation techniques used for text-independent SV. Section 16.3 includes SV studies done for variable duration test segments over standard canned database and over a database collected from practical deployed system. In Sect. 16.4, the SV experiments to understand the session variability issues are explored and this information is used for improving the practical system performance. Finally, Sect. 16.5 presents a brief conclusion.

16.2 Development of i-vector-Based SV System

The i-vector-based SV system is a contribution of factor analysis, where each Gaussian mixture model (GMM) [6] supervector of a particular utterance is transformed into a low ranked compact representation called identity vector (i-vector) using the Total-variability matrix (T-matrix), which covers all the variabilities. In this type of SV system both the train and test data undergo similar kinds of processing as shown in Fig. 16.1. Every utterance is processed by considering them as a block of 20 with 10 ms as shift. 13-dimensional mel-frequency cepstral coefficient (MFCC) features including their first- and second-order derivatives are taken and making it as a 39-dimensional feature vector, the features are extracted for the entire train and test data. Energy-based voice activity detection (VAD) is done and only frames with sufficient energy are considered as frames of interest. These features are

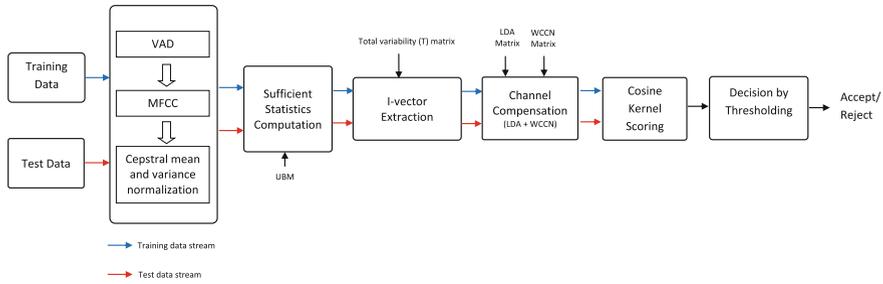


Fig. 16.1 Block diagram of i-vector-based speaker verification system [8]

then fit to zero mean unit variance. These normalized features are then used to extract the zeroth and first-order statistics, collectively called sufficient statistics using the universal background model (UBM). The UBM represents an average speaker information and is trained by development data of several hours which is not a part of train or test data. The T-matrix which covers all the variabilities is too learned with the development data. The i-vectors for each of the GMM supervectors are extracted by using their sufficient statistics and T-matrix. This low-dimensional i-vector represents each of the utterances in a robust manner, but channel/session compensation is needed on top of it to nullify the channel/session variabilities. In order to perform this linear discriminant analysis (LDA) [2] and within class covariance normalization (WCCN) [7] are performed, which further reduces the dimension of the i-vectors and make it a more compact robust representation. Finally, a cosine kernel classifier between the train and the test i-vectors is used to take a decision to accept or reject a claim based on some predetermined threshold. Given a model and a test i-vector $\hat{i}\mathbf{v}_{model}$ and $\hat{i}\mathbf{v}_{test}$, respectively, the verification for the claim is made by cosine kernel score computation between the two i-vectors as,

$$\frac{\langle \hat{i}\mathbf{v}_{model}, \hat{i}\mathbf{v}_{test} \rangle}{\|\hat{i}\mathbf{v}_{model}\| \|\hat{i}\mathbf{v}_{test}\|} \leq \gamma \text{ (Threshold)} \quad (16.1)$$

16.3 SV Studies: Variable Duration Test Segments

The i-vector-based SV system is developed using NIST SRE 2003 database as described in Sect. 16.2. Switchboard Corpus II cellular data is used as development data to train the UBM and T-matrix. A gender-independent UBM of 1024 Gaussian mixtures and T-matrix of size 400 columns are built. To compensate the channel/session variabilities the LDA matrix of 150 dimension is trained and full-dimensional WCCN matrix is taken. The system performance on the standard NIST SRE database is evaluated for variable duration of test segments considering from the full utterance upto duration of 2 s. Table 16.1 refers to the results of the system

Table 16.1 Results of i-vector-based SV system on NIST 2003 dataset for variable duration test segments

Test utterance duration (s)	Without compensation		With compensation	
	EER (%)	DCF	EER (%)	DCF
Full	4.74	0.0858	2.4	0.0474
20	5.51	0.1021	3.38	0.0606
15	6.41	0.1188	4.33	0.0813
10	8.85	0.1620	5.81	0.1090
5	13.91	0.2631	10.52	0.1977
3	19.82	0.3662	16.94	0.3100
2	25.38	0.4784	22.31	0.4128

performance in terms of equal error rate (EER) and decision cost function (DCF) for different durations of test segments. It is observed clearly that as the duration of the test segments is reduced, the performance falls drastically. Also, the performance improves when channel/session compensation is made on top of i-vectors. Studies with short utterances (<10 sec) are explored in the paper [4].

The results obtained for variable duration test segments infers that for an SV system to work properly sufficient amount of speech data are required, without which performance falls poorly. Figure 16.2 shows the score distributions for different

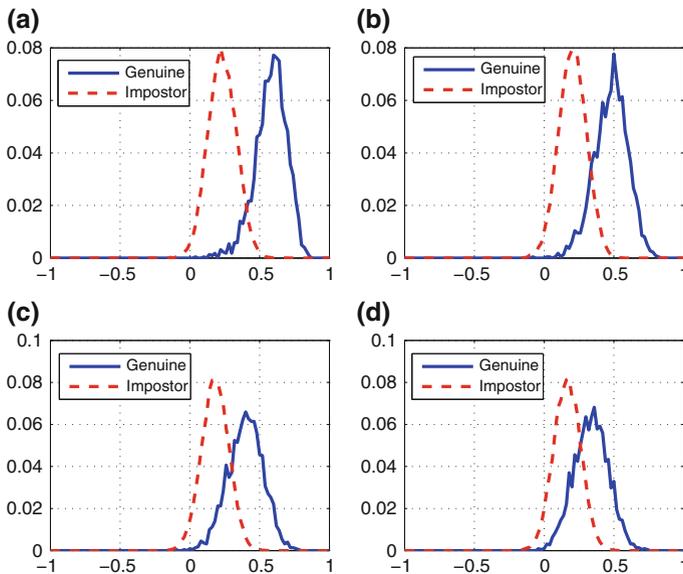
**Fig. 16.2** Score distribution for different durations of test data, **a** for 10 s test data, **b** for 5 s test data, **c** for 3 s test data, **d** for 2 s test data

Table 16.2 Performance of SV system for different durations of the test data for student attendance system

Duration (s)	EER with compensation (%)	EER with compensation and T-Norm (%)	Avg. no. of speech frames
Full	8.88	7.90	2800
40	9.8	8.56	1800
30	10.97	8.2	1200
20	12.18	9.2	700
10	13.08	11.60	350

durations of test segments, particularly for 10s or less. The distribution of genuine and impostor scores are much more separable for 10s test data case, whereas the test segment duration is reduced and the genuine and impostor scores tend to overlap more which leads to poor system performance. This also shows the way in which the threshold between the genuine and impostor score separation decreases, while reducing the amount of test data. This information is helpful for developing a practical SV-based deployable system.

A student attendance system [8] is developed over telephone network over which students regularly mark their attendance whenever they have classes. The initial studies mentioned above were useful for practical implementation of the system. The system performance is evaluated for the database collected through the attendance system for variable duration test segments. Table 16.2 shows the performance of the attendance system for different durations of test segments. As earlier results showed, better performance is achieved after channel/session compensation, therefore results for after compensation are only taken as interest. Cases with very short utterances (less than 10 s) are not considered while making the study over this database. This is because the initial studies made over NIST SRE 2003 data showed system performance falls badly, while duration of test segment goes below 10 s. The results show similar trend in downfall in performance as with the former study with NIST SRE 2003 dataset. Also when *test normalization* (T-norm) [9] is applied after channel/session compensation, shows visible improvement in performance. Table 16.2 also refers average number of speech frames for different duration test segments.

16.4 Session Variability: Effect and Remodeling

Session variability has lot of impacts on SV performance. This is much more important in case of a practical system, where there is regular testing from a group of people. The student attendance system is used regularly by students to mark their attendance. The test data collected over a period of semester is taken to study the effects of session variability issues. Figure 16.3 shows the score distribution obtained from the students over a period of 10 days. The figure clearly illustrates that the scores do not come

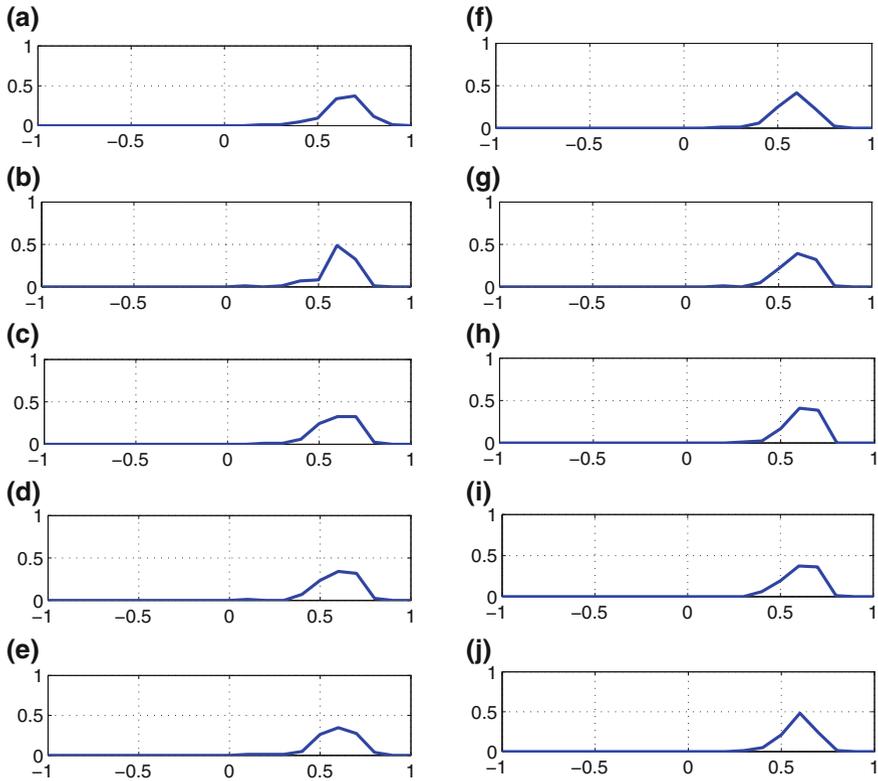


Fig. 16.3 Day-wise score distribution plots for 10 days. **a** Histogram plot for scores-day1. **b** Histogram plot for scores-day2. **c** Histogram plot for scores-day3. **d** Histogram plot for scores-day4. **e** Histogram plot for scores-day5. **f** Histogram plot for scores-day6. **g** Histogram plot for scores-day7. **h** Histogram plot for scores-day8. **i** Histogram plot for scores-day9. **j** Histogram plot for scores-day10

in a similar way for all days due to various conditions such as environmental noise present in the background, health issue of the speaker, and many such variabilities. This makes the system performance also to change everyday. A threshold is fixed by doing some offline studies to accept/reject a claim whenever a student mark his/her attendance. The success rate of the claims made to mark attendances for 10 days as mentioned in Table 16.3, which shows that the performance has variation for each of the days.

The student attendance data was collected for a period of a semester for a course. As the students used to test regularly a large database was available for each of the speakers. For better modeling the speaker with more amount of data as well as data collected in different sessions, all the speakers were retrained by adding test trial data over a month and performance analysis was made with three different days of testing data against both old model as well as the new retrained models of the speakers. Table 16.4 shows the performance comparison of three different testing

Table 16.3 Performance over a period of 10 days

Testing trial	Success rate (%)	Testing trial	Success rate (%)
Day 1	90.68	Day 6	82.59
Day 2	84.88	Day 7	89.53
Day 3	86.05	Day 8	90.70
Day 4	80.23	Day 9	82.59
Day 5	82.59	Day 10	84.88

Table 16.4 Attendance system performance comparison against retrained model

Testing trial	Success rate-old model (%)	Success rate-retrained model (%)
Test 1	75	88.33
Test 2	71	93.54
Test 3	80.4	93.49

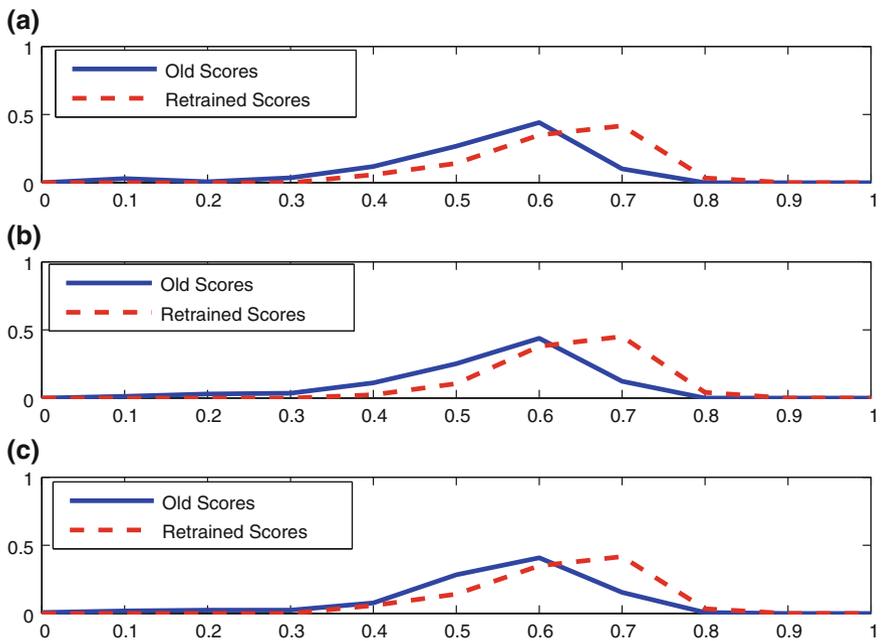


Fig. 16.4 Score analysis against retrained models. **a** Histogram plot for scores-old model versus retrained model-test 1. **b** Histogram plot for scores-old model versus retrained model-test 2. **c** Histogram plot for scores-old model versus retrained model-test 3

days for both old and retrained models. A significant improvement in performance is achieved for each of the three cases. The performance shoots up by at least 10% when the speaker models were retrained with some of the test trails collected over a period of time. Figure 16.4 shows how remodeling helps to boost the scores of the speakers with a significant margin, which in turn increases the success rate. This is due to the addition of some of the test data trails from different sessions that help in better modeling the speakers as it has speaker information for a period of time. This is helpful for a practical system deployment point of view, where multi-session data of the speakers can be used for better modeling and for improving the future system performance.

16.5 Conclusion

SV under variable durations, especially for low data case is a challenging task for the SV community. This paper presents a study of SV for variable duration test segments for a standard database and for a database collected from practically working SV system. It signifies the session variability importance in SV and how it can be used for a practical deployable system to achieve improved system performance. A significant improvement in performance is achieved when data collected over different sessions are used to retrain the speaker models over a period of time in a practical deployable system.

Acknowledgments This work is part of the ongoing project on the development of “Speech based multi-level person authentication system” funded by the e-security division of Department of Electronics and Information Technology (DeitY), Government of India.

References

1. Kinnunen, T., Li, H.: An overview of Text-Independent speaker recognition: from features to supervector. *Speech Commun.* **52**(1), 12–40 (2010)
2. Dehak, N., Kenny, P., Dehak, R., Dumouchel, P., Ouellet, P.: Front-end factor analysis for speaker verification. *IEEE Trans. Audio Speech Lang. Process.* **19**(4), 788–798 (2011)
3. Kanagasundaram, A., Vogt, R., Dean, D., Sridharan, S., Mason, M.: i-vector based speaker recognition on short utterances. In: *Interspeech* (2011)
4. Das, R.K., Abhiram, B., Prasanna, S.R.M., Ramakrishnan, A.G.: Combining source and system information for limited data speaker verification. In: *Interspeech* (2014)
5. The NIST Year 2003 Speaker Recognition Evaluation Plan. NIST. February 2003
6. Reynolds, D.A., Quatieri, T.F., Dunn, R.B.: Speaker verification using adapted Gaussian mixture models. *Digit. Signal Process.* **10**(1–3), 1941 (2000)
7. Hatch, A.O., Kajarekar, S., Stolcke, A.: Within-class covariance normalization for SVM-based speaker recognition. In: *Proceeding of ICSLP 2006*, pp. 1471–1474 (2006)
8. Dey, S., Barman, S., Bhukya, R.K., Das, R.K., Haris B.C., Prasanna, S.R.M., Sinha, R.: Speech biometric based attendance system. In: *National Conference on Communications* (2014)
9. Auckenthaler, R., Carey, M., Lloyd-Thomas, H.: Score normalization for text-independent speaker verification systems. *Digit. Signal Process.* **10**(13), 4254 (2000)

Chapter 17

Two-Dimensional Processing of Multichannel ECG Signals for Efficient Exploitation of Inter and Intra-Channel Correlation

Anurag Singh and S. Dandapat

Abstract Electrocardiogram signals acquired through different channels from the body surface are termed as Multichannel ECG (MECG) signals. They are obtained by projecting the same heart potential in different directions and hence share common information with each other. In this work a new two-dimensional (2-D) approach is proposed for MECG signal processing in order to exploit the correlated structure between the channels efficiently. Different channel data are arranged in a 2-D form giving them an image type arrangement and then 2-D discrete cosine transform (DCT) is applied in a blockwise manner over the whole data. The 2-D processing of MECG data ensures the efficient utilization of both inter-lead correlation (across the columns) and intra-lead correlation (across the rows). Since neighboring ECG samples across the channels are more correlated due to slowly varying nature of ECGs, blockwise processing of MECG data gives an effective way to exploit this. To quantify the performance of the proposed algorithm, it is evaluated on a compression platform. Each block after DCT transformation is undergone through a uniform scale zero-zone quantizer and entropy encoder to get the compressed bit streams. Performance metrics used are the compression ratio (CR), and widely used distortion measure, root mean square difference (PRD).

Keywords Multichannel ECG (MECG) · Discrete cosine transform (DCT) · Uniform scalar dead zone quantizer (USDZQ) · CR · PRD

17.1 Introduction

Electrocardiogram (ECG) signal furnishes detailed diagnostic information about the human heart. It is used by cardiologists as a noninvasive diagnostic tool for detecting various cardiac diseases. The ECG data acquisition is carried out through 12

A. Singh (✉) · S. Dandapat
Indian Institute of Technology Guwahati,
Guwahati 781039, India
e-mail: anurag.singh@iitg.ernet.in

S. Dandapat
e-mail: samaren@iitg.ernet.in

standard channels: Lead-I, II, III, aVR, aVL, aVF, V1, V2, V3, V4, V5, and V6, termed as multichannel ECG (MECG) and is the worldwide standardized. As the 12 channels/leads capture ECG signals from various regions of the heart, so different disease characteristics appear in either of these 12 leads and this makes MECG more clinically important than single lead ECG. Exploring higher lead systems, e.g., 64-leads or 128-leads [1] to extract more details about cardiac information is emerging as a potential research field. However, in real time, dealing with multiple channels by processing each channel individually is a computationally intensive task and also not an optimal way since MECGs are not independent and share common information. Hence, there is a need for efficient MECG processing techniques which can process MECGs jointly utilizing their common information effectively. Long-term ECG recording is very common while diagnosing the pathology of a patient. These long-term recordings and ever-increasing use of modern ECG monitoring devices lead to vast amount of raw digitized ECG data which becomes even higher in case of MECG recording. This poses a serious challenge in processing and storage of MECG signals, especially in resource constraint scenarios, e.g., for modern portable ECG monitoring devices. So effective processing of MECG signals becomes very important, which can compress the MECG data and give them a concise and compact representation. This can be done by exploiting the inter- and intra-channel correlation structure of MECGs efficiently. This helps in cutting down the redundant information present across the channels and compress the MECG signals for their practical and real-time processing. In this work, a two-dimensional (2-D) processing is proposed for 2-D arranged MECG data for better utilization of the inherent correlation structure. ECG signals can be compressed either in time domain taking directly the time-domain samples or in transform domain, where ECG signal is given some other time-frequency representation. Transform domain techniques proved more effective and efficient way of compressing ECG signal, since redundancy removal becomes easy in transform domains.

Single channel ECG data compression techniques have been thoroughly researched so far but limited research is available for multichannel ECG data compression. In MECG signal compression techniques, the inherent correlation existing across the different channels is the main source of redundancies apart from inter-beat and intra-beat redundancies. Many authors have attempted to exploit it either in time domain [2–4] or in any other transform domain [5–7]. Cohen and Zigel [3] proposed a long-term prediction algorithm for MECG compression. Image-based compression algorithms have also been generalized for MECG signal compression. One such type of algorithm is based on vector quantization [4] which is used for both single lead ECG and MECG compression. Transform-based MECG compression algorithm is mainly based on KL transform, discrete cosine transform (DCT) [5], and wavelets [6, 8]. Recent advancements in the field of ECG compression include the techniques based on compressive sensing (CS) and sparse representation of ECG signals, either in fixed orthogonal basis (wavelet, DCT, etc.) or in some other nonorthogonal bases (learned bases/dictionaries) [9–12]. In ECG signals, a fair amount of sparseness exist in time-domain [9] which get enhanced many folds in some transform domains (e.g., wavelet domain). This sparseness is exploited for data reduction using CS-based

approaches. But these techniques are mainly dedicated for single channel ECGs and do not suggest for MECC processing.

In this paper, a 2-D processing of 2-D arranged MECC data is proposed to better exploit the inter as well as intra-lead correlation (exist across the columns and rows of 2-D arranged MECC data). For this an image-compression-based approach is followed which processes MECC data blockwise and decorrelates them using 2-D DCT. DCT is an orthogonal transformation which is well known for its energy compaction properties, helps here in better removing the redundancies across the blocks, and gives MECCs a concise representation.

17.2 Methodology

The proposed methodology can be described in mainly three stages: Preprocessing stage, orthogonal transformation and quantization, and entropy encoding stage. Figure 17.1 shows a flow of the proposed method. Each stage is discussed in detail in the following subsections.

17.2.1 Preprocessing

In the preprocessing stage, the baseline wander present in the ECG signal is removed using a low pass finite impulse response filter of cutoff frequency 0.5 Hz [13]. Further, ECG signals are amplitude normalized and mean removed. Here, all 12-leads are not processed and only 8 channels have been processed further after removing redundant

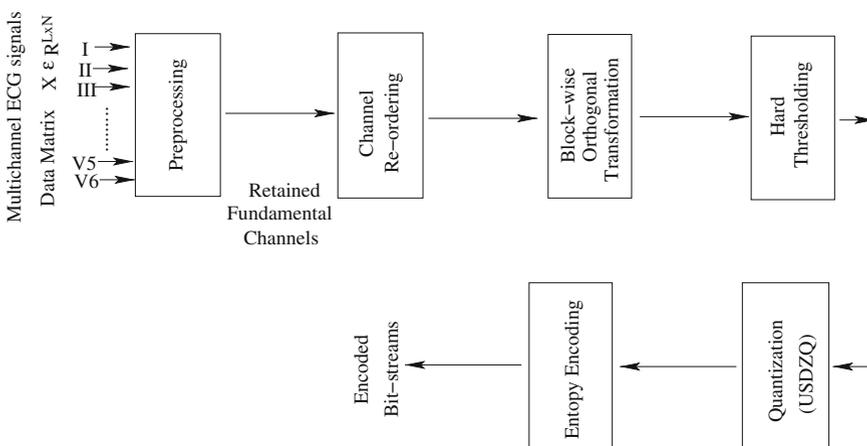


Fig. 17.1 Block-diagram of the proposed method

channels in the preprocessing stage since the rest of the 4 channels are supplementary, which can be derived from these 8 fundamental channels. As additional part of the preprocessing stage, channel reordering is done to keep the most correlated channels together to efficiently exploit their correlation. Since precordial leads capture the electrical heart vector amplitude from different narrow angles, it is more likely that they are highly correlated [5]. So we have arranged the sample points of 8 channel MEGC in matrix form keeping channels along the rows and number of samples along the columns in following order: V1, V2, V3, V4, V5, V6, lead I, lead II. So we having an $L \times N$ matrix to process with, here N and L are the number of sample points taken from each channel and number of channels (=8) respectively.

17.2.2 Orthogonal Transformation

The 2-D arranged MEGC data matrix is divided into blocks with M samples and 2-D DCT is applied on each $8 \times M$ size block in order to decorrelate the block samples. DCT of an N-dimensional 1-D signal is given as:

$$C(k) = \sqrt{\frac{\alpha_k}{N}} \sum_{n=0}^{N-1} x(n) \cos \frac{(2n+1)k\pi}{2N} \quad k = 0, 1, 2, \dots, N-1 \quad (17.1)$$

where $C(k)$ is the DCT coefficient, $\alpha_k = 1$ for $k = 1$ and $\alpha_k = 1/2$ elsewhere. The above defined 1-D DCT can be applied first on rows and then on columns of each block to get its fast 2-D implementation. The block-by-block processing of 2-D arranged MEGC samples seems intuitive as a block may have higher correlated samples compared to whole data matrix due to slowly varying nature of ECG in each lead. After transformation we get the same number of DCT-coefficients as the block size arranged in a zig-zag manner based on their clinical information content starting from low frequency toward high frequency. These DCT-coefficients are then scanned in zig-zag manner and one-dimensional vectors are formed corresponding to each block in which the coefficients are arranged in increasing order of frequency.

17.2.3 Block Thresholding

DCT squeezes the entire energy of a block into a few of the coefficients and produces many of the resulting transformed coefficients of very low amplitudes, which are insignificant for further processing. These coefficients can be truncated using any suitable thresholding technique. For threshold calculation for each block, the maximum coefficient from each block is calculated and a fraction of this is put as threshold. Coefficients above the threshold value are kept only, zeroing others as given below:

$$\alpha_i = \begin{cases} 0, & |\alpha_i| < p \times |\alpha_{\max}| \\ \alpha_i, & |\alpha_i| \leq p \times |\alpha_{\max}| \end{cases} \quad (17.2)$$

The fraction varies from 5 to 20 % of the peak value.

17.2.4 Quantization and Encoding

The thresholded coefficients are quantized by a uniform scalar dead zone quantizer (USDZQ) [14]. This is a special type of quantizer made from the modification of a mid-tread quantizer, in which a dead-zone is created across zero. The dead-zone is created across zero for further zeroing the high frequency noisy components. Coefficients outside the dead-zone are quantized uniformly without losing any significant clinical information. The introduction of the zero zone concept in the mid-tread quantizer is proved to be advantageous as it leads to noise elimination across zero and also improves signal quality and compression efficiency. The quantized coefficients are then entropy encoded by Huffman encoding strategy.

17.3 Results and Discussion

All experiments are conducted taking 4096 samples from each channel and thus making a data matrix of dimension 8×4000 . Datasets are taken from CSE-multilead library database [15]. The whole data matrix is processed blockwise. The idea of blockwise processing is motivated from JPEG image encoding technique, which helps here in exploiting both the inter-channel correlation (across the column) and intra-channel correlation (across the rows) between the neighboring ECG samples efficiently using 2-D DCT. The slowly varying nature of ECG signal in all leads contribute toward intra-channel correlation and simultaneous recording of the same heart activities through different channels contribute towards inter-channel correlation. Unlike conventional square block size used in image processing applications, here we chose a rectangular block of size $8 \times M$ to process the MECG data. Different values of M was chosen for experiments starting with $M = 8-64$ in power of 2 and it was found that block size of 8×32 gave overall optimum performance. Correlation structure between the ECG samples of different leads varies with different block sizes and that is the reason behind varying performances with different block sizes. To avoid any clinical loss while quantization, a specific quantizer is used here called USDZQ. It is well known for minimizing quantization error when its parameters are chosen judiciously. It has a specific dead zone created around zero which helps in eliminating noise and improving compression efficiency. All its specifications are chosen as suggested in [14]. Performance evaluation of the algorithm is done in terms of compression ratio (CR) and percentage root mean square difference (PRD) [16].

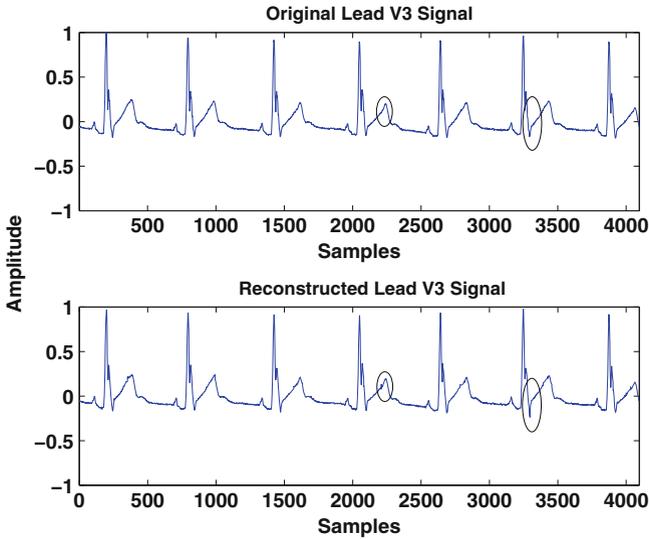


Fig. 17.2 Original ECG signal taken from Lead V3 and its corresponding reconstruction at CR = 6.93. Data-set M01_014 taken from CSE multilead library

Figure 17.2 depicts the plot of original signal from Lead V3 and its corresponding reconstruction at CR = 6.93 using a block size of 8×32 . All the clinical features of the ECG signals are preserved as verified by the reconstructed plot except a couple of points (black circled) where a slight alteration from the original signal is noticed. Distortion incurred during reconstruction is calculated in terms of widely used distortion measure PRD. Although it is not relevant from diagnostic point of view it helps in easy performance comparison with existing techniques. PRD values for different CR with 8×8 block size is depicted in Table 17.1 for data-set M01_024. It is noticeable that for pre-cordial leads (V1–V6), PRD value is lower compared

Table 17.1 PRD values for different leads for different CR values with 8×8 block size for data-set M01_024 taken from CSE Multilead library

Leads	CR = 5.07	CR = 5.07	CR = 6.04	CR = 6.58	CR = 7.08
V1	4.7136	4.8454	5.6239	6.5754	8.3005
V2	3.6364	3.7966	5.0843	5.9942	7.2245
V3	2.9736	3.4887	4.4517	5.0342	6.0347
V4	3.5264	3.9545	5.2246	6.3104	7.5899
V5	6.7100	7.8607	10.7068	13.1850	16.7586
V6	4.7215	5.3882	6.8894	8.6186	10.3809
Lead I	4.4473	4.8397	6.1996	7.2360	8.5536
Lead II	6.0259	6.5599	7.6425	9.1241	10.3926

to limb leads (Lead I and Lead II). This is because the former is generally less noisy compared to limb leads, which captures noise easily while acquisition. For noisy channels, the proposed algorithm reduces the noise level while reconstruction and so the PRD value increases as in the case of lead V5 (see Table 17.1). With increase in CR value the PRD value also increases with little distortion introduced in the reconstructed signal using fixed number of quantization bits (8 bits used here). Algorithm performance with different number of bits is shown in Fig. 17.3, which clearly shows performance improvement (low PRD) with increasing number of bits. So for any application, a tradeoff is required between maximum achievable CR, PRD, and quantization bits used.

A performance comparison of proposed technique is also done with recently reported MEEG compression techniques. Of these, one is proposed by Sharma et al. [13] which is based on multiscale principal component analysis (MSPCA) and the second is based on distributed compressive sensing (DCS) proposed by Polania et al. [11] for single channel ECG, but here we implemented it for MEEG for comparison purpose. Table 17.2 shows a lead-by-lead distortion comparison obtained through our technique and distortion reported in the existing techniques for dataset M01_014 at almost the same reported CR = 5.98. It is evident that the proposed method outperforms the reported techniques with low PRD values at the same CR. Even at higher CR = 6.93, we are getting better PRD values than the reported ones. Our algorithm also performs well for all leads in comparison to a recent reported image and volumetric coding-based MEEG compression algorithm [17] where the reported PRD (=9.97) is higher compared to our results at almost the same CR = 6.12. Although the CR achieved in our algorithm is not very high compared to state-of-art single channel CR (which might be possible due to lack of correlation in MEEG

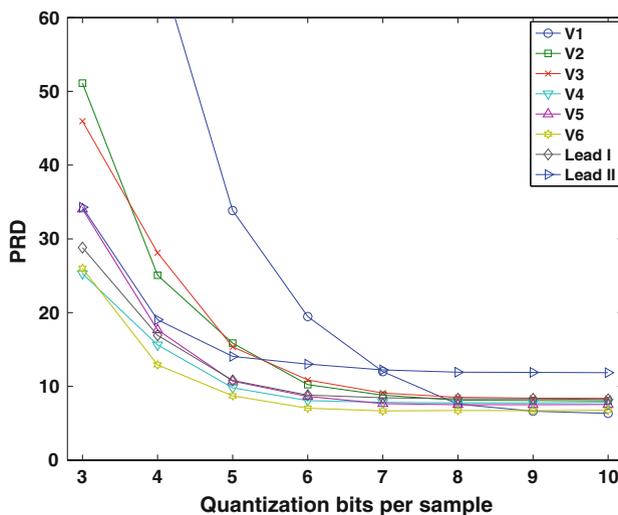


Fig. 17.3 Distortion variation in each fundamental lead with respect to quantization bits

Table 17.2 Performance comparison table

Reported techniques	Lead wise distortion (PRD)								CR
	Lead I	Lead II	V1	V2	V3	V4	V5	V6	
Sharma et al. [13]	9.82	11.74	14.08	10.33	7.63	6.82	18.76	20.74	5.98
Polania et al. [11]	16.92	11.21	7.78	8.20	8.41	7.36	23.78	23.74	5.97
Proposed	6.0	5.85	6.08	5.55	5.12	4.98	6.59	8.40	5.97
	7.58	8.12	8.51	7.75	7.52	6.73	8.32	11.92	6.93

compared to single lead ECG), yet the proposed method suggests a new way to jointly process and compress MECG data. The results can be improved further employing better block thresholding strategies. Discrete wavelet transform (DWT) may also be used for improved performance over DCT as it is known for its superior performance over DCT in case of JPEG2000 image encoding.

17.4 Conclusion

The work presented in this paper proposes a new way to exploit all the possible types of correlation structures present in multichannel ECG (MECG) signals. To achieve this, algorithm suggested a two-dimensional processing of MECG data arranged in a 2-D matrix so that the inter-channel correlation (exist across the columns) and intra-channel correlation (exist across the rows) could be exploited jointly using 2-D DCT. The significant DCT coefficients are quantized using a specific quantizer called uniform scalar zero zone quantizer which helps in minimizing the quantization error and suits for all type of ECG morphologies. The joint compression of MECG signals by the proposed algorithm may help in meeting high memory requirements for MECGs and facilitating their processing in resource constraint applications.

References

1. Haigh, A., Murray, A., Langley, P.: Separating the atrial and ventricular components in atrial fibrillation. are 64 leads better than 12? In: *Computers in Cardiology*, pp. 281–284 (2007)
2. Paggetti, C., Lusini, M., Varanini, M., Taddei, A., Marchesi, C.: A multichannel template based data compression algorithm. In: *Computers in Cardiology*, pp. 629–632 (1994)
3. Cohen, A., Zigel, Y.: Compression of multichannel ECG through multichannel long-term prediction. *IEEE Eng. Med. Biol. Mag.* **17**(1), 109–115 (1998)
4. Miaou, S.G., Yen, H.L.: Multichannel ECG compression using multichannel adaptive vector quantization. *IEEE Trans. Biomed. Eng.* **48**(10), 1203–1207 (2001), 109–115 (1998)
5. Cetin, A., Koymen, H., Aydin, M.: Multichannel ecg data compression by multirate signal processing and transform domain coding techniques. *IEEE Trans. Biomed. Eng.* **40**(5), 495–499 (1993)

6. Alesanco, A., Olmos, S., Istepanian, R., Garcia, J.: A novel real-time multilead ECG compression and de-noising method based on the wavelet transform. In: *Computers in Cardiology*, pp. 593–596 (2003)
7. Tan, Q., Fang, B., Wang, P.: Improved simultaneous matching pursuit for multilead ECG data compression. In: *2010 International Conference on Measuring Technology and Mechatronics Automation (ICMTMA)*, vol. 2, pp. 438–441 (2010)
8. Sharifahmadian, E.: Wavelet compression of multichannel ECG data by enhanced set partitioning in hierarchical trees algorithm. In: *28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society EMBS 06*, pp. 5238–5243 (2006)
9. Dixon, A., Allstot, E., Gangopadhyay, D., Allstot, D.: Compressed sensing system considerations for ECG and EMG wireless biosensors. *IEEE Trans. Biomed. Circuits Syst.* **6**(2), 156–166 (2012)
10. Mamaghanian, H., Khaled, N., Atienza, D., Vandergheynst, P.: Compressed sensing for real-time energy-efficient ECG compression on wireless body sensor nodes. *IEEE Trans. Biomed. Eng.* **58**(9), 2456–2466 (2011)
11. Polania, L., Carrillo, R., Blanco-Velasco, M., Barner, K.: Compressed sensing based method for ECG compression. In: *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 761–764 (2011)
12. Singh, A., Sharma, L., Dandapat, S.: Compressive sensing in eigenspace for multichannel electrocardiogram signals. In: *2013 International Conference on Advanced Electronic Systems (ICAES)*, pp. 166–170 September 2013
13. Sharma, L.N., Dandapat, S., Mahanta, A.: Multichannel ECG data compression based on multiscale principal component analysis. *IEEE Trans. Inf. Technol. Biomed.* **16**(4), 730–736 (2012)
14. Manikandan, M.S., Dandapat, S.: Wavelet threshold based ECG compression using USZZQ and huffman coding of DSM. *Biomed. Signal Process. Control* **1**(4), 261–270 (2006)
15. Willems, J.L.: CSE multilead Atlas, measurement results—data set 3. In: *Common Standards for Quantitative Electrocardiography*, Commission of the European Communities, Medical and Public Health Research. Ref. Nr. CSE. Leuven, 15th April 1988
16. Manikandan, M.S., Dandapat, S.: Wavelet energy based diagnostic distortion measure for ECG. *Biomed. Signal Process. Control* **2**(2), 80–96 (2007)
17. Srinivasan, K., Dauwels, J., Reddy, M.: Multichannel EEG compression: waveletbased image and volumetric coding approach. *IEEE J. Biomed. Health Inform.* **17**(1), 113–120 (2013)

Chapter 18

DCT-Based Linear Regression Approach for 12-Lead ECG Synthesis

Jiss J. Nallikuzhy and S. Dandapat

Abstract Synthesis of standard 12-lead electrocardiogram from reduced lead set without losing significant diagnostic information is a major challenge. In this work, we propose a patient specific method for synthesizing 12-lead electrocardiogram from reduced lead set by applying linear regression over the DCT domain. The proposed method is evaluated by standard distortion measures such as correlation coefficient, root mean square error, and wavelet energy-based diagnostic distortion. The results shows improvement from the existing systems without loss of significant diagnostic information.

Keywords Derived 12-lead ECG systems · Electrocardiogram (ECG) synthesis · Electrocardiography · Reduced lead sets

18.1 Introduction

Various cardiac arrhythmias can be detected by observing the commonly used noninvasive test known as electrocardiogram (ECG). A standard 12-lead ECG requires 10 electrodes to be positioned in various parts of the body for a continuous monitoring, which usually extends for several hours. This create inconvenience for the patient and the possibility of losing information due to detachment of electrodes being high. This generates a necessity for acquiring standard 12-lead ECG from less number of electrodes. Systems that synthesize standard 12-lead ECG from reduced number of leads are known as derived ECG systems.

J.J. Nallikuzhy (✉) · S. Dandapat
Indian Institute of Technology Guwahati, Guwahati 781039, India
e-mail: n.jiss@iitg.ernet.in

S. Dandapat
e-mail: samaren@iitg.ernet.in

Various systems with reduced leads are developed for generating standard 12-lead ECG. Dower, in 1968, introduced a derived 12-lead ECG system [3]. He used the orthogonal lead system developed by Frank for synthesizing the 12-lead ECG [4] using a linear transform. J.A Scherer and J.M. Nicklas developed a system for synthesizing the 12-lead ECG from a 3-lead semi-orthogonal subset using linear transformation arrays [10]. This work was based on their study where standard ECG are redundant and smaller subsets of electrodes can be used for constructing the 12-leads [10]. Reconstruction of 12-lead ECG from reduced lead sets was published by S.P. Nelwan et al. [9]. They used linear regression for obtaining the derived leads from a subset of the set of leads (I, II, V1-V6) with lead I, II and at least one precordial leads always selected [9]. The idea of deriving standard 12-lead ECG from four standard leads using information redundancy in 12-lead system was proposed by Wei [12] in 2002 based on Frank torso model. Sum-Che Man et al. published the reconstruction of standard 12-lead ECG from Mason-Likar electrode configuration in 2008 [7]. A novel neural-network model for deriving standard 12-lead ECG from three-lead ECG was proposed by Atoui et al. in 2010 [1]. 12-lead ECG reconstruction from sparse electrodes using Independent component Analysis was published by Tsouri and Ostertag in 2014 [11]. Most of these works consist of a linear transformation or more complex transforms such as the neural network. The results are evaluated by subjective measures, which may be less accurate in the case of ECG signal since the information content has diagnostic importance.

This study aims to extract the relationship between reduced leads and standard leads by regressing over the DCT coefficients between the two. The paper is organized as follows. Section 18.2 describes the methodology used for this study. Section 18.3 explains the observed results and Sect. 18.4 concludes the work.

18.2 Method

18.2.1 Preprocessing

ECG signal is preprocessed before conducting the experiment. Baseline wandering is removed by passing the signal through a low pass filter with a cut-off frequency of 0.7 Hz and then subtracting this signal from original signal. This is followed by mean removal and amplitude normalization. The signal is then divided into two parts for training and testing. Period normalization is applied to each ECG cycle after separating it out from the training and testing signal.

18.2.2 Proposed Method

The proposed method consists of transforming the signal into DCT domain and then applying linear regression over these coefficients. The DCT is given by Eq. 18.1 as

$$X(k) = \left(\frac{2}{N}\right)^{\frac{1}{2}} \sum_{n=0}^{N-1} x(n) \cdot w(n) \cdot \cos\left[\frac{\pi k}{2N}(2n+1)\right], \quad k = 0, 1, \dots, N-1. \quad (18.1)$$

where $w(n) = 1/\sqrt{2}$ for $n = 0$ and is equal to 1, otherwise. The DCT transforms the time domain signal into a scale where the high energy, low frequency components of the signal are located at the beginning of the scale and low energy, high frequency components of the signal toward the end of the scale. This property can be utilized to capture a better linear relationship between predictor leads and response lead. DCT is calculated for each period normalized cycle of ECG signal. The resulting DCT coefficients is then arranged serially for a particular lead.

Linear regression is applied over the DCT coefficients according to Eq. 18.2 as

$$Vd_{V_i} = b_0 + b_1 \cdot Vd_{Lead I} + b_2 \cdot Vd_{Lead II} + b_3 \cdot Vd_{V_2}. \quad (18.2)$$

where b_0, b_1, b_2 and b_3 are the regression coefficients and Vd_{V_i} is the target lead which can be any one of the precordial leads $Vd_{V_1}, Vd_{V_3}, Vd_{V_4}, Vd_{V_5}$ and Vd_{V_6} . Regression coefficients are calculated by minimizing the squared error between predictor and response leads using gradient descent algorithm. The limb lead, *Lead III* and augmented limb leads *aVR*, *aVL* and *aVF* can be easily synthesized by linearly combining the leads *Lead I* and *Lead II* [6]. In this work, a subset of standard 12-lead, which contains *Lead I*, *Lead II* and V_2 are selected for predicting the response lead.

18.2.3 Training and Testing

In the training phase, DCT coefficients of *Lead I*, *Lead II*, and V_2 along with the response lead V_i is regressed as per Eq. 18.2 to find the regression coefficients. This regression coefficients is used to predict the response lead in the testing phase. The resulting ECG cycle is denormalized to obtain the actual period. The process is repeated for all cycles and is combined together to obtain the complete ECG signal.

18.2.4 Performance Evaluation

The performance of proposed model is evaluated by comparing the original leads with the response leads. Standard performance evaluation methods like correlation coefficient, root mean square error, and wavelet energy-based diagnostic distortion measure [8] are used to evaluate the performance of synthesized ECG signal. Equations 18.3 and 18.4 give correlation coefficient (r_x) and root mean square error ($RMSE$) respectively. Correlation coefficient and $RMSE$ calculate the similarity between leads original and response leads.

$$r_x = \frac{\sum_{k=1}^K [x(k) - \mu_o] \sum_{k=1}^K [\hat{x}(k) - \mu_r]}{\sqrt{\sum_{k=1}^K x(k)} \sqrt{\sum_{k=1}^K \hat{x}(k)}}. \quad (18.3)$$

$$RMSE = \sqrt{\frac{1}{K} \sum_{k=1}^K [x(k) - \hat{x}(k)]^2}. \quad (18.4)$$

Here, $x(k)$ represents the original signal, $\hat{x}(k)$ represents the reconstructed signal, k denotes the k th sample, and K denotes the total number of samples in the signal. Mean value of original signal is denoted as μ_o and mean value of synthesized signal as μ_r . In this experiment, μ_o and μ_r are zero since the mean was removed while preprocessing. Equation 18.5 gives the wavelet energy-based diagnostic distortion ($WEDD$). This distortion measure is used to evaluate the clinically important diagnostic information between original and reconstructed signal.

$$WEDD = \sum_{l=1}^{L+1} w_l WPRD_l. \quad (18.5)$$

where w_l is weight for l th subband given by Eq. 18.6 and $WPRD$ is the wavelet percentage root mean square difference given by Eq. 18.7.

$$w_l = \frac{\sum_{k=1}^{K_l} d_l^2(k)}{\sum_{m=1}^{L+1} \sum_{k=1}^{K_l} d_m^2(k)}, \quad l = 1, 2, \dots, (L + 1). \quad (18.6)$$

$$WPRD_l = \sqrt{\frac{\sum_{k=1}^{K_l} [d_l(k) - \hat{d}_l(k)]^2}{\sum_{k=1}^{K_l} [d_l(k)]^2}}. \quad (18.7)$$

Here, K_l is the number of wavelet coefficient in l th subband and $d_l(k)$ is the k th wavelet coefficient of signal in l th sub-band [8]. In this experiment, mean free ECG signal is decomposed to 7th level using Daubechies biorthogonal 4/4 wavelets filters.

18.3 Results and Discussions

The most commonly available PTB database is used for this study [2, 5]. The proposed method is verified for 70 healthy control data and 45 anterior myocardial infarction data. Each data of 10s duration is selected and is divided into two parts of 5s each for training and testing.

18.3.1 Performance Evaluation for Healthy Control Data

The model is tested for 70 healthy control data in the PTB database using performance evaluation criteria such as correlation coefficient (r_x), $RMSE$ and $WEDD$. Figure 18.1 shows the original and reconstructed precordial leads for healthy control data. It can be observed that the proposed model is able to reconstruct most of the leads in an efficient manner. Lead V_4, V_5 and V_6 shows a little amount of noise compared to the original leads. This is due to the fact that these leads are located away from the predictor lead V_2 and is commonly referred to as proximity effect.

Table 18.1 shows the comparison of proposed model with linear regression model and ICA-based model for various performance evaluation criteria. Here average values of evaluation criteria for 70 healthy control data are tabulated. It can be observed that correlation coefficient and $RMSE$ show much similarity in the case of linear regression and proposed model. Notable difference can be observed in the values of $WEDD$, which is an objective distortion measure. It can be inferred that the proposed model captures diagnostic information much more compared to the existing models. This is important since ECG signal contains very sensitive diagnostic information.

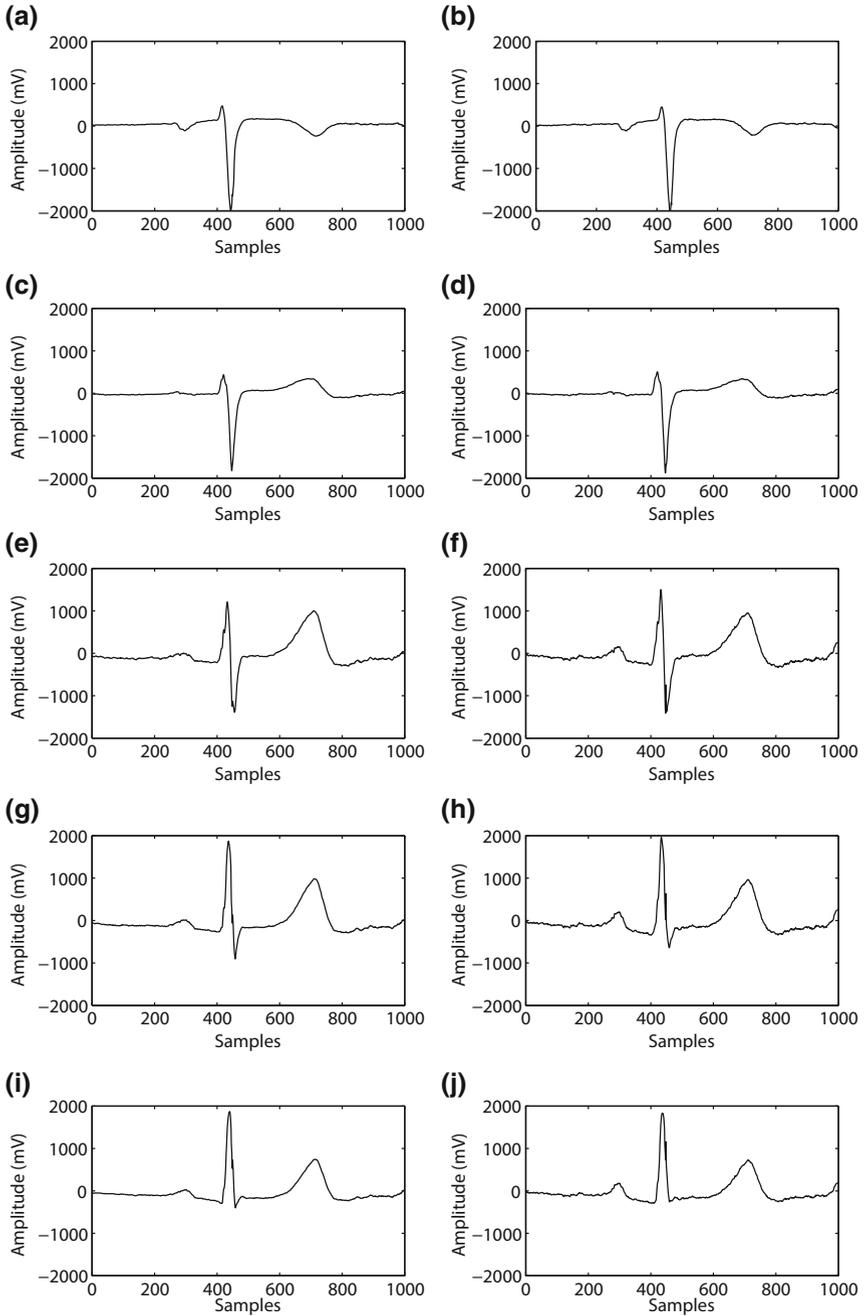


Fig. 18.1 Original and synthesized ECG leads V_1 , V_3 , V_4 , V_5 and V_6 for healthy control data. **a** Original lead V_1 . **b** Reconstructed lead V_1 . **c** Original lead V_3 . **d** Reconstructed lead V_3 . **e** Original lead V_4 . **f** Reconstructed lead V_4 . **g** Original lead V_5 . **h** Reconstructed lead V_5 . **i** Original lead V_6 . **j** Reconstructed lead V_6

Table 18.1 Averaged correlation coefficient (r_x), *RMSE* and *WEDD* between original and synthesized healthy control data for linear regression (LR), ICA method, and proposed model (PM)

Leads	Avg. correlation coefficient (r_x)			Avg. RMSE			Avg. WEDD (%)		
	LR	ICA	PM	LR	ICA	PM	LR	ICA	PM
V_1	0.97	0.95	0.97	0.032	0.036	0.033	10.15	23.76	9.31
V_3	0.98	0.96	0.98	0.032	0.036	0.033	15.39	19.27	13.98
V_4	0.96	0.93	0.96	0.049	0.047	0.051	32.58	26.9	31.67
V_5	0.97	0.95	0.96	0.04	0.038	0.042	29.96	24.99	29.48
V_6	0.98	0.96	0.97	0.033	0.032	0.034	22.53	21.28	22.31

18.3.2 Performance Evaluation for Anterior Myocardial Infarction Data

The proposed model is tested for performance in the case of pathological data. Here, 45 anterior myocardial infarction data are used for testing. Figure 18.2 shows the original and reconstructed leads for anterior myocardial infarction data. It can be observed that, the proposed model performs satisfactorily and is able to reconstruct most of the precordial leads. A similar observation as that of healthy control data can be noted here as well. The leads away from V_2 show a little noise in the signal, but still preserve the shape.

Table 18.2 shows the comparison of the model with linear regression model and ICA-based model for various performance evaluation criteria. The average performance evaluation value for 45 anterior myocardial infarction data is tabulated. In this case, it is observed that all the evaluation criteria show similar values for both regression and proposed model. The proposed model preserves as much diagnostic information as that of existing models.

In this experiment, the proposed model was evaluated only over two datasets, one normal dataset and one pathological dataset. The model may perform better in the case of other pathologies since DCT transforms the signal into a domain where a better linear relationship between predictor leads and response lead can be achieved. The experiment is conducted only for specific patients and hence the performance of the proposed model as a global model for synthesis is an open challenge.

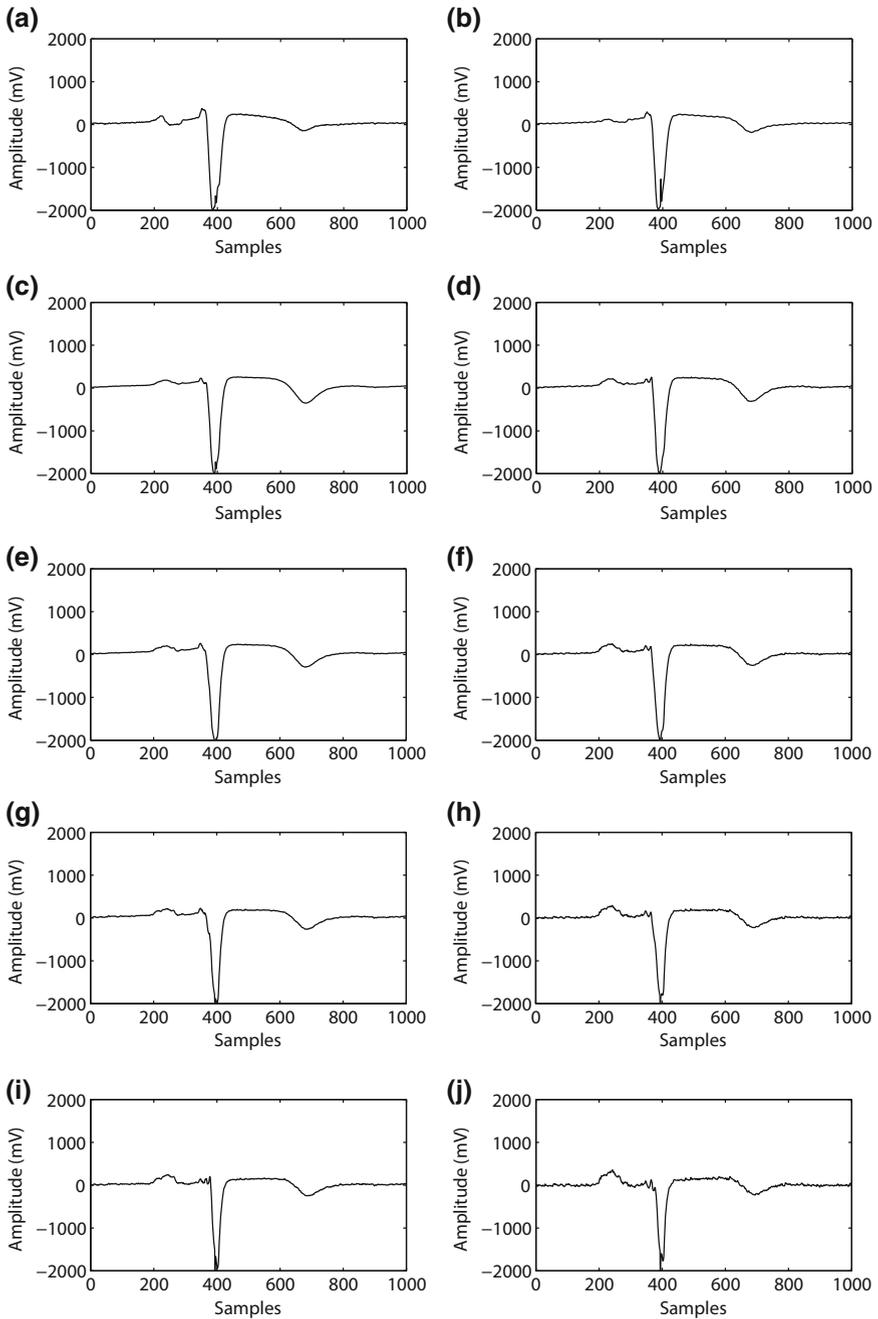


Fig. 18.2 Original and synthesized ECG leads V_1 , V_3 , V_4 , V_5 and V_6 for anterior myocardial infarction data. **a** Original lead V_1 . **b** Reconstructed lead V_1 . **c** Original lead V_3 . **d** Reconstructed lead V_3 . **e** Original lead V_4 . **f** Reconstructed lead V_4 . **g** Original lead V_5 . **h** Reconstructed lead V_5 . **i** Original lead V_6 . **j** Reconstructed lead V_6

Table 18.2 Averaged correlation coefficient (r_x), *RMSE* and *WEDD* between original and synthesized anterior myocardial infarction data for linear regression (LR), ICA method and proposed model (PM)

Leads	Avg. correlation coefficient (r_x)			Avg. RMSE			Avg. WEDD (%)		
	LR	ICA	PM	LR	ICA	PM	LR	ICA	PM
V_1	0.98	0.96	0.98	0.031	0.034	0.032	15.67	21.43	15.39
V_3	0.97	0.96	0.97	0.037	0.039	0.038	12.44	21.19	12.67
V_4	0.92	0.91	0.91	0.072	0.065	0.075	23.39	33.39	23.99
V_5	0.94	0.91	0.94	0.054	0.051	0.056	21.05	32.27	20.68
V_6	0.96	0.94	0.96	0.043	0.041	0.044	17.21	26.48	17.50

18.4 Conclusion

A patient-specific model for synthesizing standard 12-lead ECG from reduced lead set is proposed in this work. The proposed model utilizes the linear relationship between predictor leads and response leads in the DCT domain for synthesizing. Performance evaluation results show that the proposed model preserves the sensitive diagnostic information much more than the existing models.

References

1. Atoui, H., Fayn, J., Rubel, P.: A novel neural-network model for deriving standard 12-lead ECGs from serial three-lead ECGs: application to self-care. *IEEE Trans. Inf. Technol. Biomed.* **14**(3), 883–890 (2010)
2. Boussejot, R., Kreiseler, D., Schnabel, A.: Nutzung der ekg-signaldatenbank cardiodat der ptb ber das internet. *Biomedizinische Technik* **40**(1), 317–318 (1995)
3. Dower, G.E.: A lead synthesizer for the frank system to simulate the standard 12-lead electrocardiogram. *J. Electrocardiol.* **1**(1), 101–116 (1968)
4. Frank, E.: An accurate, clinically practical system for spatial vectorcardiography. *Circulation* **13**(5), 737–749 (1956)
5. Goldberger, A.L., Amaral, L.A.N., Glass, L., Hausdorff, J.M., Ivanov, P.C., Mark, R.G., Mietus, J.E., Moody, G.B., Peng, C.K., Stanley, H.E.: Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. *Circulation* **101**(23), e215–e220 (2000)
6. Malmivuo, J., Plonsey, R.: *Bioelectromagnetism: Principles and Applications of Bioelectric and Biomagnetic Fields*, 1st edn. Oxford University Press, New York (1995)
7. Man, S.C., Maan, A.C., Kim, E., Draisma, H.H., Schalij, M.J., van der Wall, E.E., Swenne, C.A.: Reconstruction of standard 12-lead electrocardiograms from 12-lead electrocardiograms recorded with the Mason-Likar electrode configuration. *J. Electrocardiol.* **41**(3), 211–219 (2008)
8. Manikandan, M.S., Dandapat, S.: Wavelet energy based diagnostic distortion measure for ECG. *Biomed. Signal Process. Control* **2**(2), 80–96 (2007)
9. Nelwan, S.P., Kors, J.A., Meij, S.H., van Bommel, J.H., Simoons, M.L.: Reconstruction of the 12-lead electrocardiogram from reduced lead sets. *J. Electrocardiol.* **37**(1), 11–18 (2004)

10. Scherer, J., Nicklas, J.: Synthesis of the 12 lead electrocardiogram from a 3 lead semi-orthogonal subset using patient-specific linear transformation arrays. In: Proceedings of Computers in Cardiology, September 1988, pp. 449–451 (1988)
11. Tsouri, G., Ostertag, M.: Patient-specific 12-lead ECG reconstruction from sparse electrodes using independent component analysis. *IEEE J. Biomed. Health Inform.* **18**(2), 476–482 (2014)
12. Wei, D.: Deriving the 12-lead electrocardiogram from four standard leads using information redundancy in the 12-lead system. *Int. J. Bioelectromagn.* **4**, 127–128 (2002)

Chapter 19

Design, Simulation, and Performance Evaluation of a High Temperature and Low Power Consumption Microheater Structure for MOS Gas Sensors

Kaushik Das and Priyanka Kakoty

Abstract The purpose of this work is to design, simulate, and to evaluate the performance of a low power microheater for a gas sensing system. A microheater is a microstructure incorporated in a MOS gas sensor in order to elevate the temperature of the sensor to an operating range for the reliable performance of a gas sensor. An approach is made in this work to find an optimized microheater structure by considering different membrane sizes and geometries and taking into account the temperature distribution and power consumption problems. The materials used for the analysis are Platinum and Polysilicon. After analyzing various microheater designs, a novel design is developed by optimization and varying geometry, layer dimension, and materials of the device. For the developed design, thermal profile and power consumption analysis are carried out. The entire work is carried out in COMSOL MULTIPHYSICS 4.2.

Keywords Microheater · Heater geometry · Low power · Uniform temperature

19.1 Introduction

Among all other existing types of gas sensors, MOS-based gas sensors have exhibited promising properties in the field of gas sensing due to their acceptable sensitivity, low cost, flexibility in production, and stability over wide working temperature range, large number of possible application field like environmental, medicinal, food industry, detection of industrial and toxic gases. One of the key components of a MOS gas sensor is the microheater. The temperature of the sensing layer is controlled by the microheater which is used as a hot plate. The semiconductor gas sensor utilizes the principle of chemiadsorption to detect changes in resistance as a function of varying concentrations of different gases. In order to detect these resistive changes, the heater

K. Das (✉) · P. Kakoty
Tezpur University, Tezpur 784028, India
e-mail: 54kaushik@gmail.com

P. Kakoty
e-mail: priyankak@tezu.ernet.in

temperature must be held constant and uniform over the heater area. Therefore, on the operating temperature homogeneity/uniformity of the microheater, different semiconductor gas sensor properties like sensitivity, selectivity, and response time are dependent [1, 2]. Additionally, the use of a microheater in a gas sensor, fabricated using compatible process with a good fabrication yield can improve the dynamic response in the sensing process and reduce the power consumption too [3].

A high temperature should be maintained on the entire gas sensing region in order to meet current and future electronic system requirements [4]. Generally, bulk Si substrate is used where the heater is micromachined. The high temperature is obtained by joules heating of the heater deposited on the membrane. Si has excellent mechanical properties. Another advantage of using Si substrate is that the electronic circuits and the sensor can be integrated on one single chip. Thus, for the appropriate performance of a semiconductor gas sensor, a microheater is in-built in it that plays a crucial role in maintaining and controlling the temperature of the sensing film.

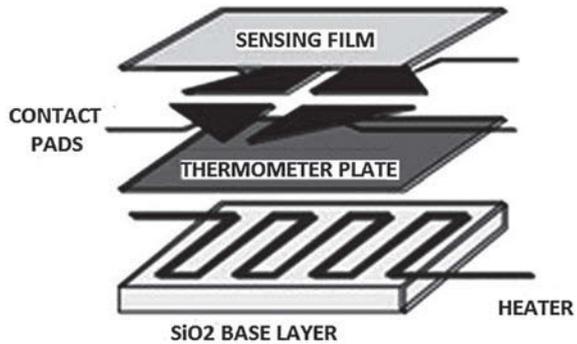
19.2 Selection of Heater Material

Choosing a perfect material for the microheater is a challenge for the better performance of the gas sensor. Most of the researches reported so far deal with the design of either platinum or polysilicon microheater, particularly applied for the higher temperature range (400–700 °C) for better high temperature stability and the process compatibility [5]. But the problem of high temperature sensing lies in the high power consumption and the reduced lifetime of the sensor. For polysilicon, the resistivity is very high, making it suitable for the heater element, but it cannot be deposited simply by electron beam deposition technique. Instead it has to be deposited using chemical vapor deposition (CVD) technique, which is much more costly method. In this work, we have come up with a design to reduce the power consumption of the platinum and polysilicon microheater based on geometrical variation which has resulted in optimum performance.

19.3 Microheater Placement

The functional layers in a MOS gas sensor with a single microhotplate element are shown in the Fig. 19.1. The most obvious features in a microhotplate are the suspended plate, supported beams, and the pit. The device shown also has four electrodes that are used for measuring the electrical characteristics of deposited sensing films. The layer structure within the suspended plate indicates other functionality within the basic microhotplate arrangement.

Fig. 19.1 Functional layers of a microhotplate (reproduced from [6])



19.4 Design and Electrothermal Analysis of Microheater

Microheaters are microstructures consisting generally a hot plate on a membrane micromachined from bulk silicon. The temperature rise is obtained by the Joules heating in a resistor deposited on a membrane a few micrometers thick. The objective of the microheater design is to achieve low power consumption and a uniform temperature distribution over the active area. This elevated uniform temperature of the sensing material is required for the smooth operation of the gas sensor as the chemical adsorption/desorption of the gas on the sensing surface can take place only at an elevated temperature, and the uniformity in temperature gives a better sensor response. The tool used for this purpose is COMSOL MULTIPHYSICS 4.2. Different types of heater elements, heater structures, and geometries are analyzed in order to obtain an optimal design.

The numerical approximation of the temperature distribution, total heat loss, etc., can be obtained by replacing the sensor by a set of points. This set of points is called the computation grid or mesh. The right choice of the computational grid is very important to obtain good approximation. By increasing the number node points can result in better accuracy but at the same time leads to an increase of computational time. Therefore, it is often better to use a nonuniform grid which models the parts of the sensor which are subjected to large changes with a narrow grid and regions with nearly no change in temperature with a rather rough grid. A fundamental premise of using the finite element procedure is that the body is subdivided up into small discrete regions known as finite elements. These elements defined by nodes and interpolation functions.

The commercial finite element model (FEM) programs COMSOL MULTIPHYSICS 4.2 have been employed for the coupled electrothermal simulations by means of three-dimensional (3D model) powerful finite element analysis program that can be used in many engineering analysis, including thermal, structural, and electrical analysis. COMSOL MULTIPHYSICS 4.2 3D Builder Module has been used for design of a microheater and COMSOL MULTIPHYSICS 4.2 Joules Heating Module is used for simulation of microheater. For temperature less than 400 °C

radiation losses are neglected [7]. Due to small size of the heated structure, the convection losses are negligible at medium-low temperature. Assuming the fluid motion to be negligible the air has been considered as a static fluid modeled as a solid on the top and bottom of the membrane. The temperature of the frame is assumed to remain at room temperature. Hence the boundary condition is 27°C.

19.5 Design Procedure

From the study it has been established that for a low-power semiconductor metal oxide gas sensor power consumption of the heater should be low as possible. There are several reports where researchers have designed microheaters with different shapes (meander shaped, fan shaped, spiral shaped, S-shaped, honeycomb shaped, etc.) [8] which yielded different results. Taking into consideration the response given by all the available designs, a novel design was developed and optimized to give a better response. In our design, we wanted to improve the temperature uniformity of the microheater. From the already known shapes, it was revealed that a square structure gives better temperature uniformity. And in order to improve the heating, the number of connectors between inner squares and outer square to increase the current density all over the structure was developed [9]. In our approach, we have tried to develop a square-shaped structured microheater based on the previous reports available. The operating voltage was selected from 50 mV for simulation. Platinum is chosen as the reference material. For the square gridded-type microheater, to start with, the power consumption is taken to be 10 mW which is given by

$$P = \frac{V^2}{R} \quad (19.1)$$

where V is the applied voltage and R is the resistance of the material. So to calculate the length of the design we have,

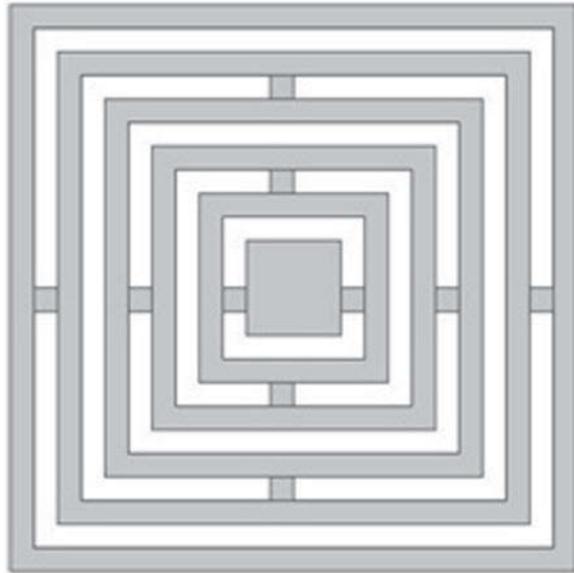
$$R = \rho \frac{L}{A} \quad (19.2)$$

where ρ is the resistivity of the material; L is the length of the heater and A is the active area of the microheater.

Thus taking all the above values in consideration the length is found out to be $\sim 117 \mu\text{m}$, which we take $120 \mu\text{m}$ to simplify our design.

The square gridded-type design is shown in the Fig. 19.2 the designed microheater has length of $120 \mu\text{m}$, width of the strip is $10 \mu\text{m}$, and thickness of $5 \mu\text{m}$.

Fig. 19.2 Square grilled type



19.6 Simulation of Platinum and Polysilicon Microheater

The Joule heating module is used to get the electrothermal simulation. The stationary solver was used to determine the steady state temperature distribution of microheater.

Fixed temperature and potentials are assumed at ends of the heater. Several material properties are required to solve the mathematical equations as mentioned above.

The material properties of platinum and polysilicon are shown in Table 19.1.

The potential applied across the heater ranges from 50 to 60 mV. The simulated result of 60 mV voltage is shown in Fig. 19.3 and the variation of temperature is shown in Fig. 19.4.

Table 19.1 Material properties of platinum and polysilicon

Material	Platinum	Polysilicon
Electrical resistivity in $\Omega\text{-m}$	10.6e-10	32.2e-8
Thermal expansion coeff α in 10 ⁻⁶ /K	8.8	2.8
Thermal conductivity at 300 K in W/M-K	70	29–34
Density in kg/m ³	2109	2330
Young's modulus GPa	168	169
Specific heat in J/kg-K	130	753

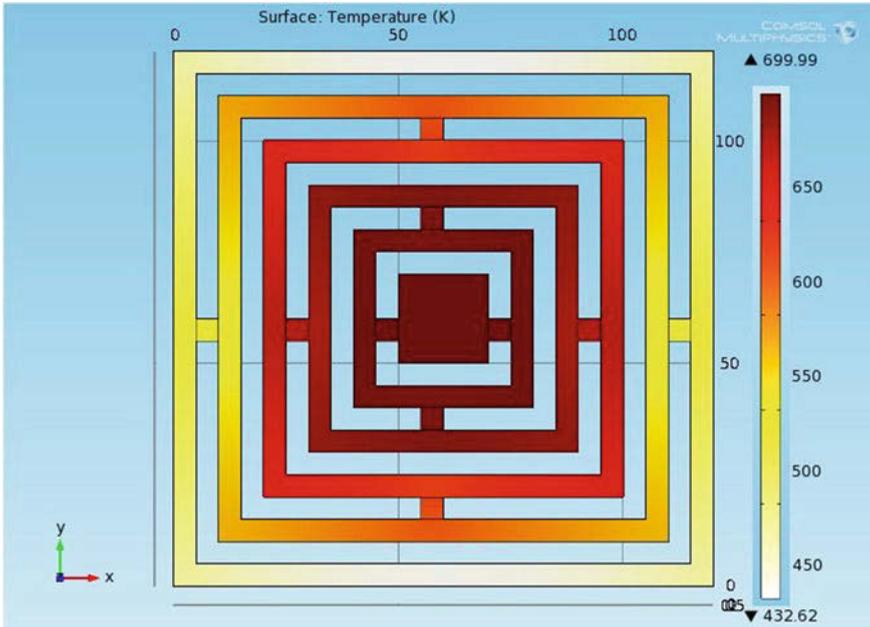


Fig. 19.3 Thermal profile of square gridded-type platinum microheater

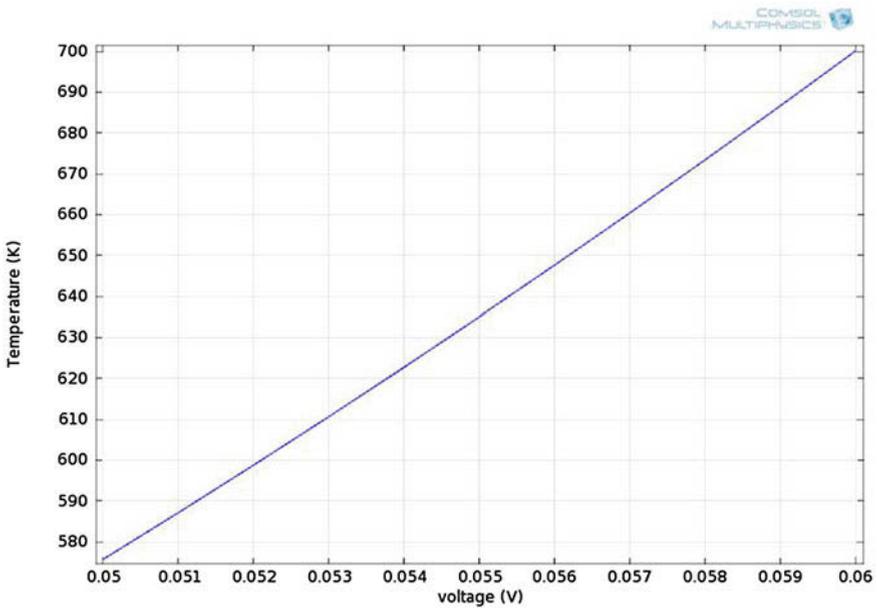


Fig. 19.4 Variation of temperature with voltage of square gridded-type platinum microheater

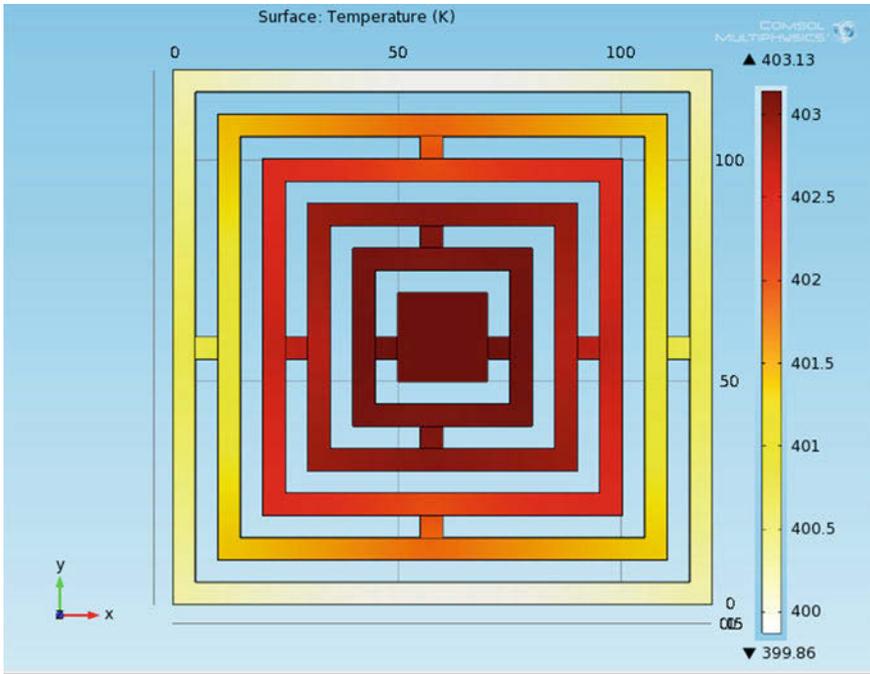


Fig. 19.5 Thermal profile of square gridded-type polysilicon microheater

The variation of temperature for the polysilicon microheater is shown below in Fig. 19.5.

There is a substantial difference in the average temperature recorded between the two microheaters with and without the power compensation. In order to objectively compare the average temperatures, only the heat over the area of the membrane was considered. By analyzing the simulation results and the thermal profile, we can compare the required temperature range, maximum and average sensor temperature, power consumption, required heater resistance for the given power consumption. For each design the difference between maximum and average temperature can easily be calculated. Comparing the simulation results shown in Figs. 19.4 and 19.6, it has been observed that the microheater with platinum as heating material yields higher temperature than that of polysilicon microheater for the same operating voltage.

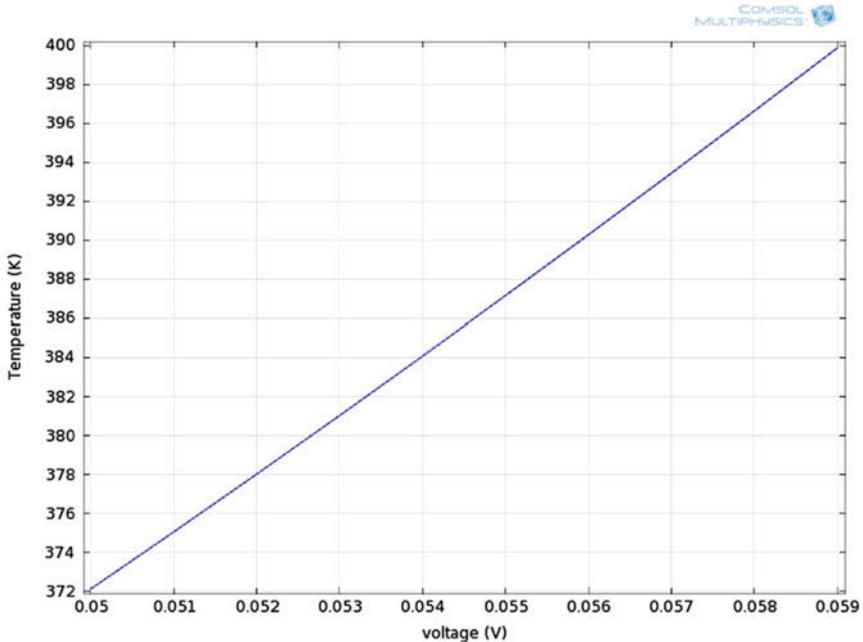


Fig. 19.6 Variation of temperature with voltage of square grided-type polysilicon microheater

19.7 Conclusion

Comprehensive thermal model of microheaters are designed and simulated using COMSOL 4.2. While examining the temperature output when the voltage applied is kept in the range of 50–60 mV the square grided platinum type design gives the optimum result. A uniform microheater temperature is a necessary requirement as it often enhances the operation of the sensor. The film overlying the microheater should be maintained at a uniform temperature for maximum sensitivity. A uniform temperature implies minimization of the heater “hot-spot” which is a crucial requirement for heater reliability. Thus the temperature uniformity depends on the membrane materials and also on their geometry, especially of the microheater. In this case, the inner part of the heater area experiences heat loss only to the ambient area and radiation, whereas the outer part additionally suffers from heat loss through the membrane. As our work is in basic level, it is our future concern to fabricate a microheater on the basis of these simulations.

References

1. Partridge, J.G., Field, M.R., Sadek, A.Z., Kalantar-Zadeh, K., Du Plessis, J., Taylor, M.B., Atanacio, A., Prince, K.E., McCulloch, D.G.: Fabrication, structural characterization and testing of a nanostructured Tin oxide gas sensor. *IEEE Sens. J.* **9**, 563–568 (2009)
2. Tao, C., Yin, C., He, M., Tu, S.: Thermal analysis and design of a micro-hotplate for si-substrated micro-structural gas sensor. In: *Proceedings of the 3rd IEEE International Conference on Nano/Micro Engineered and Molecular Systems*, pp. 284–287, Sanya, China, 6–9 January, 2008
3. Ruther, P., Ehmann, M., Lindemann, T., Paul, O.: Dependence of the temperature distribution in micro hotplates on heater geometry and heating mode. In: *The 12th International Conference on Solid State Sensors, Actuators and Microsystems*, pp. 73–76, Boston, 8–12 June, 2003
4. Chung, G.S., Jeong, J.M.: Fabrication of micro heaters on polycrystalline 3C-SiC suspended membranes for gas sensors and their characteristics. *Microelectron. Eng.* **87**, 2348–2352 (2010)
5. Mondal, S.S.: Development of MEMS based coplanar microheater structures. Master of Electronics and Tele-Communication Engineering, Thesis submitted by (2013)
6. Semancik, S., Cavicchi, R.E., Wheeler, M.C., Tiffany, J.E., Poirier, G.E., Walton, R.M., Suehle, J.S., Panchapakesan, B., Devoe, D.L.: Microhotplate platforms for chemical sensor research. vol. 77, pp. 579–591 (2001)
7. www.comsol.co.in, Introduction to COMSOL Multiphysics. Protected by U.S. Patents 7, 519, 518; 7, 596, 474; 7, 623, 991, pp. 1–162 (2013)
8. Velmathi, G., Ramshanker, N., Mohan, S.: 2D Simulations and Electrothermal Analysis of Microheater Designs Using COMSOL TM for Gas Sensor Applications. pp. 2–5 (2010)
9. Sujatha, L., Selvakumar, V.S., Aravind, S., Padamapriya, R., Preethi, B.: Design and analysis of microheaters using COMSOL multiphysics for MEMS based gas sensor. Excerpt from the *Proceedings of the 2012 COMSOL Conference in Bangalore*

Chapter 20

Experimental Analysis on the Performance of a New Preprocessing Method Used in Data Compression

P. Khanikar, M.P. Bhuyan, R.R. Baruah and H. Sarma

Abstract This paper presents a new text transformation method, which has a few similarities with the StarNT text transformation method. StarNT is a dictionary-based lossless text transform algorithm. Many different compression methods have been devised by researchers to find a suitable solution of data transmission that utilises the entire network bandwidth optimally and which also achieves a higher compression ratio. Most of the approaches that are being used, like the Prediction by Partial Matching (PPM), Burrows–Wheeler Transform have been unable to achieve the best possible output as provided by theoretical calculations and hence have left researchers to find more efficient techniques of text compression. Further in this paper we also provide the experimental results of the timing performance and space utilisation by compression of our algorithm by comparing with StarNT method.

Keywords Text transformation · StarNT · Dictionary based transform · Encoding · Compression

P. Khanikar (✉) · M.P. Bhuyan · H. Sarma
Assam Engineering College, Guwahati 781013, India
e-mail: pk_1992@yahoo.com

M.P. Bhuyan
e-mail: manash.cse@aec.ac.in

H. Sarma
e-mail: dugisharma2013@gmail.com

R.R. Baruah
Gauhati University, Guwahati 781014, India
e-mail: rashmibaruah05@gmail.com

20.1 Introduction

Transmission of data from one source to another over a network takes place through the utilisation of communication channels. The channel bandwidth is sometimes unable to accommodate and transmit data optimally because of the large amount of storage area required by the data which is primarily down to the presence of redundant data and its inefficient representation. Transmission of a considerable amount of redundant data is, therefore, one of the primary factors behind the slow conveyance of useful data over the Internet. File compression and decompression is one of the most effective ways to transmit data over the internet by removing redundant data and utilising some transformation algorithms to achieve lossless, reversible transmission of data [1–3]. Transformation algorithms are those algorithms which are applied to the source files to make them easier to be compressed by compression algorithms afterwards. Some of the transformation techniques are given below.

20.1.1 *Burrows–Wheeler Transform*

BWT [4] was introduced in 1994 by Michael Burrows and David Wheeler. According to the theory, the BWT is an algorithm that procures a block of data and restructures it using a sorting algorithm, then piped through a Move-To-Front (MTF) stage, then the Run Length encoder stage and finally an entropy encoder (Huffman coding or Arithmetic coding). Bhuyan et al. [5] had introduced security concept with Burrows–Wheeler Transform.

20.1.2 *LIPT Transform*

The basis of creation and utilisation of the LIPT [6] Transform is the length and frequency of the words occurring. LIPT uses a dictionary of the English language composed by 59,951 words, with a size of approximately 0.5 MB. The transformation dictionary has approximately 0.3 MB. The words that are not found in the dictionary are left unchanged in the coded text. To create the LIPT dictionary, the English dictionary needs to be sorted depending on the word length, and every block of a specific length to be sorted in descending order of the frequency of the words.

20.1.3 Star Transform

The Star Transform [6] is another lossless, reversible transformation algorithm which is implemented on a source file before actual compression process to make the compression process more easier. It works by making a very big dictionary of commonly used words expected in the input files. The dictionary should be set up beforehand and it must be introduced to the compressor and decompressor. Every word in the dictionary has a star encoded equivalent, and the '*' symbol is used to replace as many letters as possible.

20.1.4 StarNT Transform

The StarNT Transform [7] uses only letters {a...z, A...Z} as code words to gain a much better compression performance for the backend data compression algorithm. A dictionary is created, where every word in the dictionary is assigned a unique keyword. The first 26 words in the dictionary are assigned 'a', 'b', ..., 'z' as their code words. The next 26 words are assigned 'A', 'B', ..., 'Z'. After that the 53rd word is assigned 'aa', the 54th 'ab'. Following this order, 'ZZ' code word is assigned to the 2756th word in the Dictionary. The 2757th word is assigned 'aaa', the following 2758th word is assigned the 'aab' codeword, and so on.

20.2 Our Proposed Dictionary Based Transform Method

In this section we describe our text transformation method. First we present the algorithm, the dictionary design and the encoding procedure. Then we point out the differences and similarities between our method and StarNT method with an example. Further we present the experimental analysis of our method.

20.2.1 Algorithm

1. Create the dictionary, which consists of words and their respective code words.
2. Encode the input text file using dictionary.
3. If all words in the input text file are found in dictionary, go to step 5.
4. If new words are found in input text file, add them to a temporary file and encode the remaining words.
5. Encoding complete.

20.2.2 Dictionary

The initial part of the dictionary is made by using some of the most frequently used words in English language (upto 312 words). After the first 312 words, next words are arranged according to length in decreasing order. Words which are not found in the dictionary at the time of encoding are added to a temporary file, from which they are periodically updated in the dictionary. The code words are assigned in the following way:

1. Words having length ≤ 3 , the code word is '#' followed by the word itself.
For e.g. if the word is abc, code word is '#abc'.
2. Words having length > 3 , the code words are assigned by different combinations of letters of English language.
For e.g. code words will be like 'a', 'b', 'c',, 'aa', 'ab', 'ac', 'ad',, 'ba', 'bb',, etc.

20.2.3 Encoding

Any text file can be encoded using the dictionary. When the words in the input text file are found in the dictionary, these words are replaced by their respective codewords in the dictionary and the encoding process is completed. However, when some words are found in the text file, which are not present in the dictionary, these words are added to a temporary file.

20.2.4 Comparison of StarNT with Our Proposed Transformation Method

20.2.4.1 Similarities

1. In both the text transform methods [8, 9], most frequently used words are listed at the beginning of the dictionary (upto 312 words).
2. In both the methods, different combinations of English letters [8, 9] are used as codewords.

20.2.4.2 Differences

1. In both the methods after the first 312 words in the dictionary, the remaining words are stored in according to their lengths. Words with longer lengths are stored after words with shorter lengths in StarNT [8, 9], while in our method the scenario is opposite.

2. In our method words having length ≤ 3 , the code word is '#' followed by the word itself and words having length > 3 , the code words are assigned by different combinations of letters of English language. But in StarNT [8, 9], different combinations of letters of English language are only used as code words.
3. In the encoded output text of our method, the '*' symbol is used to differentiate between the different code words. However, in StarNT [8, 9] method, the '*' symbol is prefixed before a word in the encoded output file, when that word is not found in the dictionary.

20.2.5 Example

Let us consider the sentence 'My name is Rahul'. We consider the scenario where the word 'name' in the given sentence is not present in the dictionary. Let the code words of 'My', 'is', 'Rahul' be 'a', 'b', 'c', respectively, in StarNT and '#My', '#is', 'a', respectively, in our method. Now the encoded outputs will be:

a *name b c (in StarNT)

#My*name*#is*a (in our method)

Now we consider the scenario where all the words are present in the dictionary. Let the code words of 'My', 'name', 'is', 'Rahul' be 'a', 'b', 'c', 'd', respectively, in StarNT and '#My', 'a', '#is', 'b', respectively, in our method. Now the encoded outputs will be:

a b c d (in StarNT)

#My*a*#is*b (in our method)

20.2.6 Performance Evaluation

20.2.6.1 Analysis of Our Algorithm

Here we have performed an experiment on the dictionary creation time for our method, encoding time for different sizes of files and the comparison of the size of encoded files in our method and the starNT method. We have considered files starting from size 100 kb upto 1024 kb for the experiments as shown in Tables (20.1, 20.2 and 20.3).

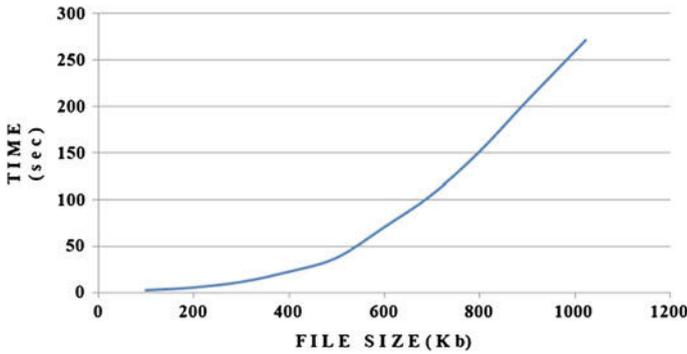


Fig. 20.1 Time analysis graph for dictionary creation for files of different size

From Fig. 20.1 we can claim that the dictionary creation time for our method is bounded by $O(n^2)$. From Fig. 20.2 it is seen that the encoding time of the input files in StarNT is less than our proposed method. From Fig. 20.3 it is clear that before the

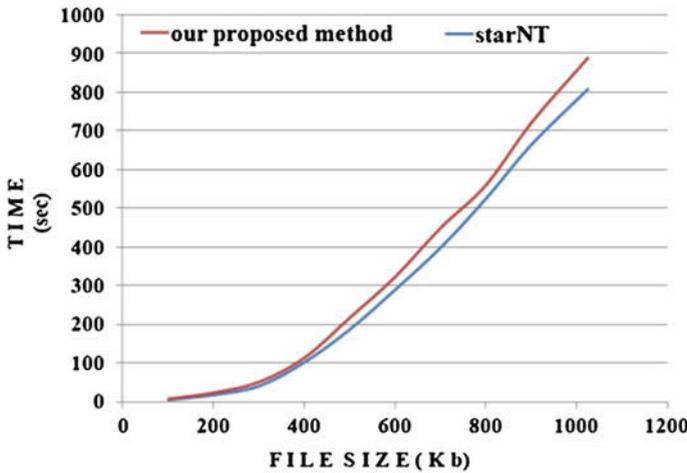


Fig. 20.2 Time analysis graphs for encoding of files of different size for StarNT and our proposed method

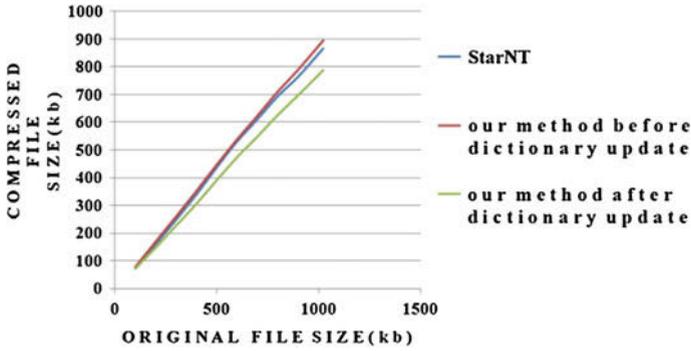


Fig. 20.3 Original file size versus compressed file size graphs for files of different size for StarNT and our proposed method

Table 20.1 Dictionary creation time for files of different sizes

Size of file to be encoded (Kb)	Dictionary time (s)
100	2
200	5
300	11
400	22
500	37
600	70
700	105
800	151
900	206
1024	272

dictionary update the compression of the input text file is slightly better in StarNT than our method. However, after the dictionary update it is seen that the compression is considerably better in our method than StarNT.

Table 20.2 Encoding time for files of different sizes

File size (Kb)	Encoding time for our method (s)	Encoding time for StarNT (s)
100	6	4
200	22	17
300	50	40
400	113	102
500	217	187
600	323	290
700	450	399
800	561	526
900	723	666
1024	890	809

Table 20.3 Encoded file size for files of different sizes for StarNT and for our proposed method before and after dictionary update

File size (Kb)	Compressed file size before dictionary update for our method (Kb)	Compressed file size after dictionary update for our method (Kb)	Compressed file size for StarNT (Kb)
100	78	72	72
200	169	147	159
300	259	226	247
400	351	305	338
500	446	390	436
600	537	473	529
700	622	547	609
800	709	625	692
900	788	696	762
1024	894	787	864

20.3 Conclusion

From the above experimental analysis it can be concluded that though the time required to encode files in our method before and after dictionary update is more, but it improves the performance of compression compared to StarNT. So the use of these type of preprocessing methods will help in improving the compression ratio of the other existing algorithms if we feed the files after applying reversible preprocessing method in those algorithms for text compression.

References

1. Jeyanthi, P., Anuratha, V.: Analysis of lossless reversible transformation. *J. Glob. Res. Comput. Sci.* **3**(8), (2012)
2. Awan, F., Zhang, N., Motgi, N., Iqbal, R., Mukherjee, A.: LIPT: a reversible lossless text transform to improve compression performance. In: *Proceedings of the IEEE Data Compression Conference 2001*, Utah, S., Storer, J., Cohn, M. (eds.) 481, (2001)
3. Radu RADESCU (2011) LIPT-Derived transform methods used in lossless compression of text files. *U.P.B. Sci. Bull., Seri. C*, **7**(32). ISSN 1454–234x (2011)
4. Burrows, M., Wheeler, D.J.: A block-sorting lossless data compression algorithm. *Digit. Syst. Res. Cent. Res. Rep.* **124**, (1994)
5. Bhuyan, M.P., Deka, V., Bordoloi, S.: Burrows Wheeler based data compression and secure transmission, In: *Proceedings of 1st National Conference on Research and Higher Education in Information Technology (RHEIT 2013)*, 4–5th February 2013, pp 35–40
6. Radu RADESCU.: Transform methods used in lossless compression of text files. *Romanian. J. Sci. Technol.* **12**(1). 101–115 (2009)
7. Rexline, S.J.: Robert L dictionary based preprocessing methods in text compression - A Survey. *Int J Wisdom Based Comput* **1**(2), 13–18 (2011)
8. Baruah, R.R., Deka, V., Bhuyan, M.P.: Enhancing dictionary based preprocessing for better text compression. *Int. J. Comput. Trends. Technol. (IJCTT)* **8**(3), 2231–2803 (2014)
9. Sun, W., Mukherjee, A., Zhang, N.: A dictionary-based multi-corpora text compression system. In: *Proceedings of Data Compression Conference, Snowbird, Utah, March 2003*

Chapter 21

Modeling a Nano Cylindrical MOSFET Considering Parabolic Potential Well Approximation

Jyotiskha Deka and Santanu Sharma

Abstract In this paper an analytical surface potential-based model considering the quantum mechanical effect at the semiconductor-oxide interface of a nanoscale cylindrical MOSFET is developed. The model considers the decoupling of the Poisson's and Schrodinger's equations via parabolic potential well approximation instead of fully self-consistent approach. Using the developed model, the effect of variation on surface potential, threshold voltage, drain current, with the extension into the saturation regime alongwith the variation of substrate doping, silicon pillar diameter, drain to source voltage, and gate to source voltage are observed. While obtaining the results, a large discrepancy in the device characteristics from the classical analysis is seen and this proves the need for quantum analysis to be done for highly doped substrates.

Keywords Cylindrical MOSFET · Schrodinger equation · Quantum mechanical effect (QME) · Threshold voltage · Drain current

21.1 Introduction

With the development of MOS devices into deep submicron regime, the scaling theory requires higher substrate doping concentration and thinner gate oxide thickness [1]. This leads to a high transverse electric field at the oxide/semiconductor interface. Such a strong field at the interface gives rise to splitting of the conduction band into discrete subbands and quantum mechanical effects (QMEs) becomes significant [2, 3]. Quantum mechanical effects manifest themselves in nanoscale MOSFETs in two ways. First, the threshold voltage shifts to a higher value due to higher electron

J. Deka (✉)
Assam Down Town University, Guwahati 781026, India
e-mail: jyotiskha145@gmail.com

S. Sharma
Tezpur University, Tezpur 784028, India
e-mail: sss@tezu.ernet.in

ground-state energy. Second, the electron concentration peaks away from the surface in contrast to the classical solution in which the carrier density peaks at the surface [4].

21.2 Threshold Voltage Model

To estimate the quantum mechanical V_T -shift with a 2-D confinement in a cylindrical structure, it is required to solve the Schrodinger equation and Poisson's equation self-consistently. Although this method gives accurate result, it is quite cumbersome [5]. A simple approximation for this problem is triangular potential well approximation. But this approximation does not give the exact shape of band bending. Here a parabolic potential well approximation in the cylindrical geometry has been assumed. As per this assumption, the potential is given as follows (Figs. 21.1 and 21.2):

$$V(r) = F \frac{t_{si}}{2} \left[1 - \left\{ \frac{r}{\frac{t_{si}}{2}} \right\}^2 \right], r \leq t_{si}/2 \tag{21.1}$$

$$V(r) = \infty, r > t_{si}/2 \tag{21.2}$$

Fig. 21.1 Cylindrical gate all around MOSFET

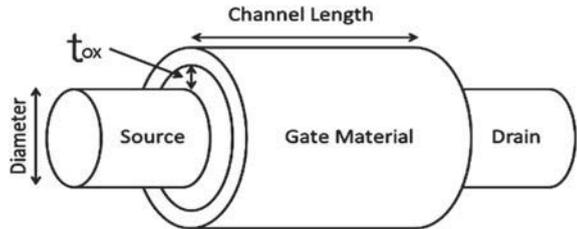
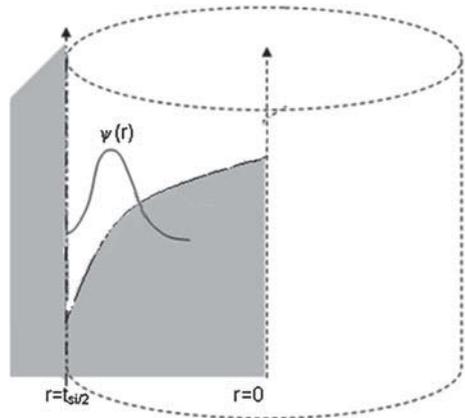


Fig. 21.2 The parabolic potential well in a cylindrical structure



The Schrödinger equation in cylindrical coordinate with a parabolic potential well can be presented as follows [5]:

$$-\frac{\hbar^2}{2m^*} \nabla^2 \Psi + F \frac{t_{si}}{2} \left[1 - \left\{ \frac{r}{\frac{t_{si}}{2}} \right\}^2 \right] \Psi = E \Psi \quad (21.3)$$

$$\begin{aligned} -\frac{\hbar^2}{2m^*} \frac{1}{r} \frac{\delta}{\delta r} \left(r \frac{\delta \Psi}{\delta r} \right) - \frac{\hbar^2}{2m^*} \frac{1}{r^2} \left(r \frac{\delta^2 \Psi}{\delta \Phi^2} \right) - \frac{\hbar^2}{2m^*} \left(r \frac{\delta^2 \Psi}{\delta z^2} \right) \\ + F \frac{t_{si}}{2} \left[1 - \left\{ \frac{r}{\frac{t_{si}}{2}} \right\}^2 \right] \Psi = E \Psi \end{aligned} \quad (21.4)$$

The ground-state wave function obtained is independent of z and Θ . So the above equation can be reduced to

$$-\frac{\hbar^2}{2m^*} \frac{1}{r} \frac{\delta}{\delta r} \left(r \frac{\delta \Psi}{\delta r} \right) + F \frac{t_{si}}{2} \left[1 - \left\{ \frac{r}{\frac{t_{si}}{2}} \right\}^2 \right] \Psi = E \Psi \quad (21.5)$$

which can finally be reduced to the following form:

$$\frac{\delta^2 \Psi}{\delta r^2} + \frac{1}{r} \frac{\delta \Psi}{\delta r} + \frac{2m^*}{\hbar^2} \left[E - F \left(\frac{t_{si}}{2} \right) \left\{ 1 - \frac{r^2}{\left(\frac{t_{si}}{2} \right)^2} \right\} \right] \Psi = 0 \quad (21.6)$$

$$\frac{\delta^2 \Psi}{\delta r^2} + \frac{1}{r} \frac{\delta \Psi}{\delta r} + \frac{2m^*}{\hbar^2} \left[E - F \left(\frac{t_{si}}{2} \right) + \left\{ F \frac{r^2}{\left(\frac{t_{si}}{2} \right)^2} \right\} \right] \Psi = 0 \quad (21.7)$$

The second term in the above equation can be neglected near the semiconductor insulator interface. Consequently, the Schrödinger equation becomes

$$\frac{\delta^2 \Psi}{\delta r^2} + \frac{2m^*}{\hbar^2} \left[E - F \left(\frac{t_{si}}{2} \right) + \left\{ F \frac{r^2}{\left(\frac{t_{si}}{2} \right)^2} \right\} \right] \Psi = 0 \quad (21.8)$$

Now let us introduce the variable ξ , where

$$\xi = \left\{ \frac{r^2}{\frac{t_{si}}{2}} + \frac{E - F \left(\frac{t_{si}}{2} \right)}{F} \right\} \left(\frac{2m^* F}{\hbar^2} \right)^{\frac{1}{5}} \quad (21.9)$$

And the Schrödinger equation reduces to

$$\frac{\delta^2\Psi}{\delta\xi^2} + \xi\Psi = 0 \tag{21.10}$$

The solution of the equation is

$$\Psi(\xi) = A\phi(-\xi) \tag{21.11}$$

$$\phi(\xi) = \frac{1}{\sqrt{\pi}} \int_0^\infty \cos\left(\frac{1}{3}u^3 + u\xi\right) du \tag{21.12}$$

$\phi(\xi)$ is known as Airy function and A is known as normalization constant [6].

$$A = \frac{(2m)^{\frac{1}{5}}}{\pi^{\frac{1}{2}} F^{\frac{1}{6}} \hbar^{\frac{2}{5}}} \tag{21.13}$$

$$F = qE_s \tag{21.14}$$

E_s is the effective electric field at the semiconductor surface region and can be expressed as [7]

$$E_s = -\frac{0.5Q_{inv} + Q_{dep}}{\epsilon_{si}} \tag{21.15}$$

For a MOSFET device the inversion charge density is given by [8, 9]

$$Q_{inv} = C_{ox}(V_{GS} - V_{th}) \tag{21.16}$$

Consequently, for a cylindrical device the same can be expressed as

$$Q_{inv} = C_{ox,cyl}(V_{GS} - V_{th,cyl}) \tag{21.17}$$

And the depletion charge is given by

$$Q_{dep} = -E_{si}E_s \tag{21.18}$$

where $E_s = -\frac{\delta\Psi}{\delta\rho} \Big|_{\rho=\frac{r_{si}}{2}}$ is the electric field at the surface when the width of depletion layer is maximum [9]. The asymptotic behaviors of Airy functions are as follows [7]:

$$\phi(\xi) = \frac{A}{2|\xi|^{\frac{1}{4}}} \exp\left(\frac{2}{3}|\xi|^{\frac{3}{2}}\right) \quad \text{for } \xi < 0 \tag{21.19}$$

$$\phi(\xi) = \frac{A}{2|\xi|^{\frac{1}{4}}} \sin\left(\frac{2}{3}|\xi|^{\frac{3}{2}} + \frac{1}{4}\pi\right) \quad \text{for } \xi > 0 \tag{21.20}$$

At $r = t_{si}/2$, the value of

$$\xi = \left\{ \frac{t_{si}}{2} + \frac{E - F \frac{t_{si}}{2}}{F} \right\} \left(\frac{2m^*F}{\hbar^2} \right)^{\frac{1}{5}} \quad (21.21)$$

$$\xi = \left\{ \frac{t_{si}F + 2E - 2F \frac{t_{si}}{2}}{2F} \right\} \left(\frac{2m^*F}{\hbar^2} \right)^{\frac{1}{5}} \quad (21.22)$$

$$\xi = \frac{E}{F} \left(\frac{2m^*F}{\hbar^2} \right)^{\frac{1}{5}} > 0 \quad (21.23)$$

Therefore, the second form has to be used for this. At $r = t_{si}/2$, $\Psi(\xi) = 0$ and hence the second asymptotic function becomes as shown below

$$\begin{aligned} 0 &= \frac{A}{2 \left[\frac{E_n}{F} \left(\frac{2m^*F}{\hbar^2} \right)^{\frac{1}{3}} \right]^{\frac{1}{4}}} \sin \left[\frac{2}{3} \left\{ \frac{E_n}{F} \left(\frac{2m^*F}{\hbar^2} \right)^{\frac{1}{3}} \right\}^{\frac{3}{2}} + \frac{\pi}{4} \right] \\ \Rightarrow \sin n\pi &= \sin \left[\frac{2}{3} \left\{ \frac{E_n}{F} \left(\frac{2m^*F}{\hbar^2} \right)^{\frac{1}{3}} \right\}^{\frac{3}{2}} + \frac{\pi}{4} \right] \\ \Rightarrow n\pi - \frac{\pi}{4} &= \frac{2}{3} \left\{ \frac{E_n}{F} \left(\frac{2m^*F}{\hbar^2} \right)^{\frac{1}{3}} \right\}^{\frac{3}{2}} \\ \Rightarrow \left[\left(n - \frac{1}{4} \right) \pi \right]^{\frac{2}{3}} &= \frac{2}{3} \left\{ \frac{E_n}{F} \left(\frac{2m^*F}{\hbar^2} \right)^{\frac{1}{3}} \right\} \\ \Rightarrow E_n &= \left(\frac{3}{2} \pi F \right)^{\frac{2}{3}} \left(\frac{\hbar^2}{2m^*} \right)^{\frac{1}{3}} \left(n - \frac{1}{4} \right)^{\frac{2}{3}} \quad \text{for } n = 1, 2, 3 \quad (21.24) \end{aligned}$$

For a moderately doped semiconductor, and at a low gate voltage there is low electron density, hence it can be considered that only the lowest subband ($n = 1$) is occupied.

$$E_1 = \left(\frac{\hbar^2}{2m^*} \right)^{\frac{1}{3}} \left(\frac{3}{2} \pi F \right)^{\frac{2}{3}} \left(\frac{3}{4} \right)^{\frac{2}{3}} \quad (21.25)$$

Now, using the expression of E_1 , the effective increase in threshold voltage can be expressed as follows:

$$\Delta V_{thqm} = \frac{E_1}{q} \quad (21.26)$$

Cylindrical oxide capacitance per unit area is [5]

$$C_{ox-cyl} = \frac{2\epsilon_{ox}}{t_{si} \ln \left(1 + 2 \frac{t_{ox}}{t_{si}} \right)} \quad (21.27)$$

Hence the new threshold voltage is given as

$$\begin{aligned} V_{thqm} &= V_{th} + \Delta V_{thqm} \\ &= V_{th} + \frac{E_1}{q} \end{aligned} \quad (21.28)$$

where

$$V_{th} = V_{FB} + 2|\phi_F| + (|Q_B|)/C_{(ox-cyl)} \quad (21.29)$$

V_{FB} = Flatband voltage

ϕ_F = Bulk Fermi potential

$$\phi_F = (kT/q) \ln(N_a/n_i)$$

k = Boltzmann's constant

$$= 1.38 \times 10^{-23} \text{ J/K}$$

q = electronic charge

$$= 1.6 \times 10^{-19} \text{ C}$$

T = temperature (K)

kT/q = thermal voltage (volts)

N_a = Acceptor doping density in the substrate (cm^3)

n_i = intrinsic carrier concentration (cm^3).

21.3 Drain Current Model

When quantum mechanical effects (QME) is taken into account, the drain current equation for a cylindrical gate all around MOSFET becomes

For active region

$$I_D = \mu C_{ox,cyl} \frac{W_{eff}}{L} \left[(V_{GS} - V_{thqm}) V_{DS} - \frac{V_{DS}^2}{2} \right] \quad (21.30)$$

For saturation region

$$I_D = \mu C_{ox,cyl} \frac{W_{eff}}{2L} (V_{GS} - V_{thqm})^2 \quad (21.31)$$

21.4 Results and Discussions

Figure 21.3 shows the variation of threshold voltage shift due to quantum mechanical effects (QME) with diameter of the silicon pillar of the cylindrical MOSFET for aluminum gate and n^+ polysilicon gate. With decrease in diameter of the silicon pillar, the quantum confinement becomes tighter leading to larger shift of threshold voltage at smaller size.

Figure 21.4 shows the variation of threshold voltage shift due to quantum mechanical effects (QME) with doping concentration of the cylindrical MOSFET for aluminum gate and n^+ polysilicon gate. With increase in doping concentration, the threshold voltage shift also increases.

Figures 21.5, 21.6, and 21.7 show the dependence of threshold voltage (with and without including QME) on doping concentration for aluminum gate for different values of silicon film thickness. When QME is taken into account, the conduction band edge goes up. As a result of this, the gate voltage required is little more to bring the conduction band below the Fermi level by the same amount and thus increasing the threshold voltage. For higher value of silicon film thickness, threshold voltage increases sharply.

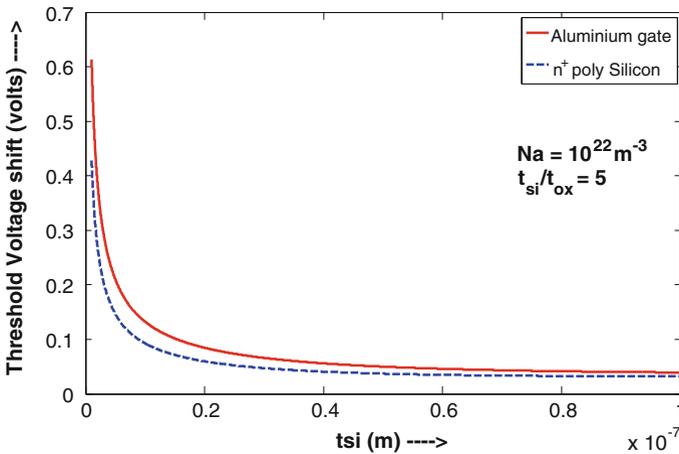


Fig. 21.3 Threshold voltage shift versus silicon pillar diameter for aluminum gate and n^+ polysilicon gate

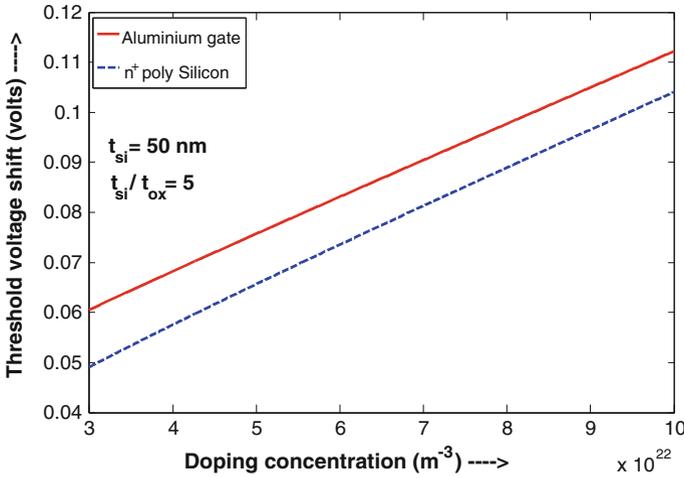


Fig. 21.4 Threshold voltage shift versus doping concentration for aluminum gate and n^+ polysilicon gate

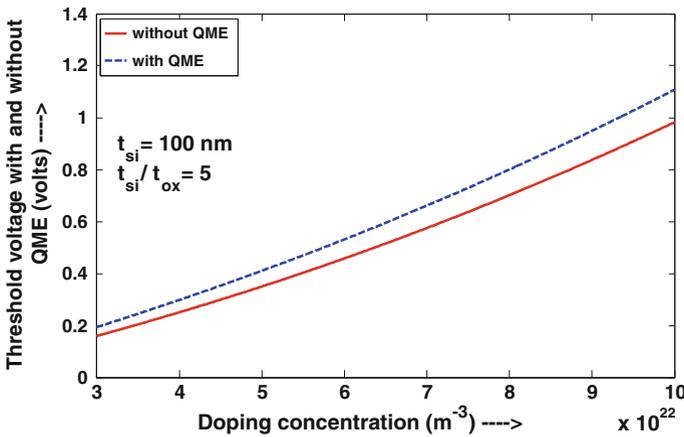


Fig. 21.5 Threshold voltage (with and without including QME) versus doping concentration for aluminum gate for silicon pillar thickness (t_{si}) of 100 nm

Figure 21.8 shows the dependence of threshold voltage including QME on silicon pillar diameter for aluminum gate and n^+ polysilicon gate. It is observed that when the silicon pillar diameter increases the threshold voltage decreases steeply and then attains a constant value.

Figure 21.9 shows the variation of drain current considering QME with channel length for various values of silicon film thickness in case of aluminum gate. Drain current increases as the channel length is decreased.

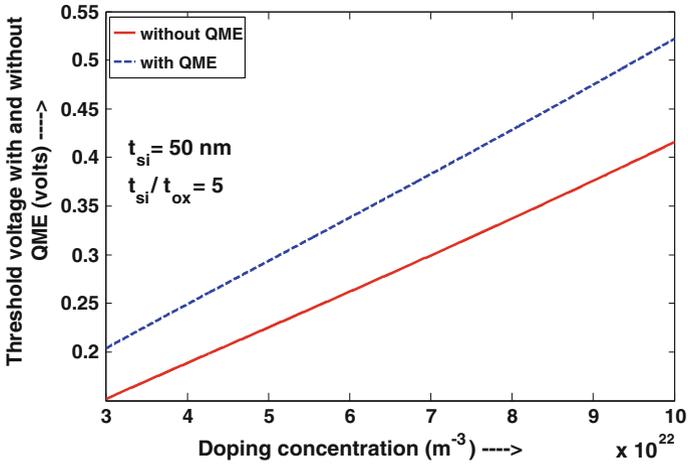


Fig. 21.6 Threshold voltage (with and without including QME) versus doping concentration for aluminium gate for silicon pillar thickness (t_{si}) of 50 nm

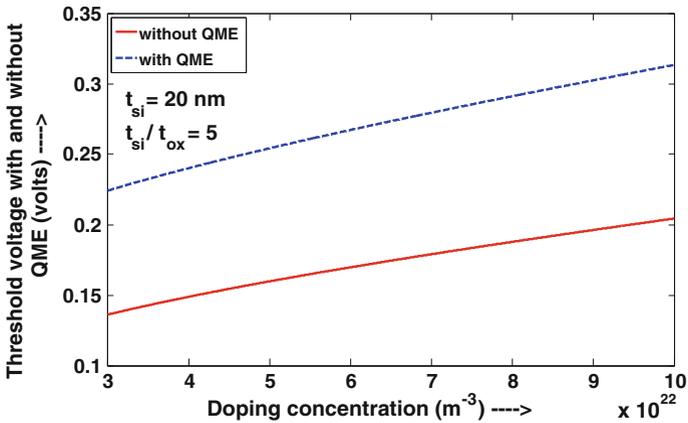


Fig. 21.7 Threshold voltage (with and without including QME) versus doping concentration for aluminum gate for silicon pillar thickness (t_{si}) of 20 nm

Figure 21.10 shows the variation of drain current considering QME with silicon film thickness for various values of channel length in case of aluminum gate. Drain current increases as the silicon film thickness is increased. Figure 21.11 shows the threshold voltage (considering QME) dependence on channel length due to short channel effect. As the channel length decreases, the fraction of charge in the channel region controlled by the gate decreases. This results in a reduction in threshold voltage value.

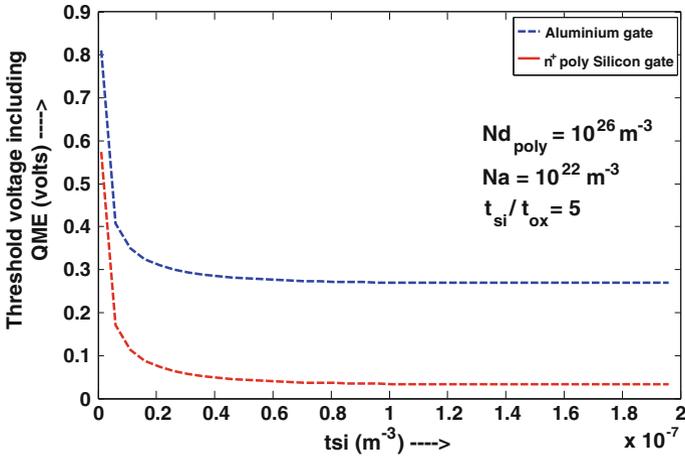


Fig. 21.8 Threshold voltage including QME versus silicon pillar diameter for aluminum gate and n⁺ polysilicon gate

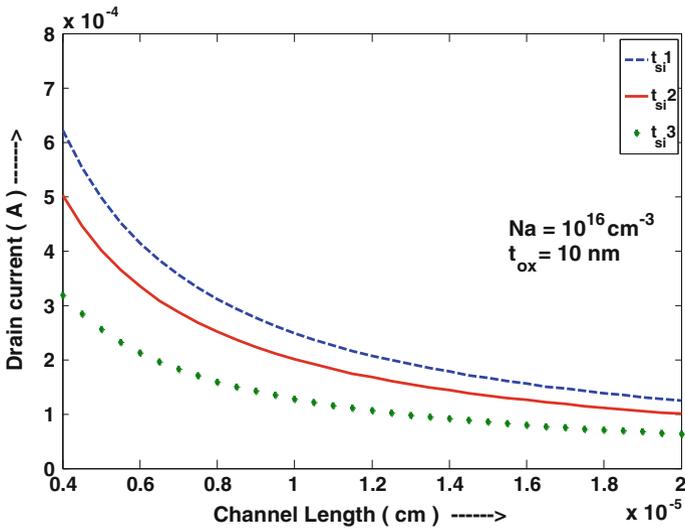


Fig. 21.9 Drain current considering QME versus channel length for various values of silicon film thickness (tsi1 = 100 nm, tsi2 = 60 nm and tsi3 = 40 nm) for aluminum gate

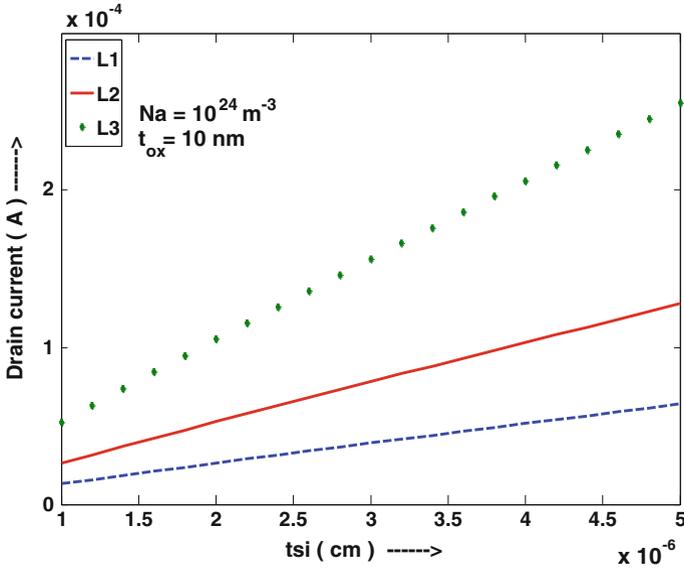


Fig. 21.10 Drain current considering QME versus silicon film thickness for aluminum gate at different channel lengths (L1 = 200 nm, L2 = 100 nm and L3 = 50 nm)

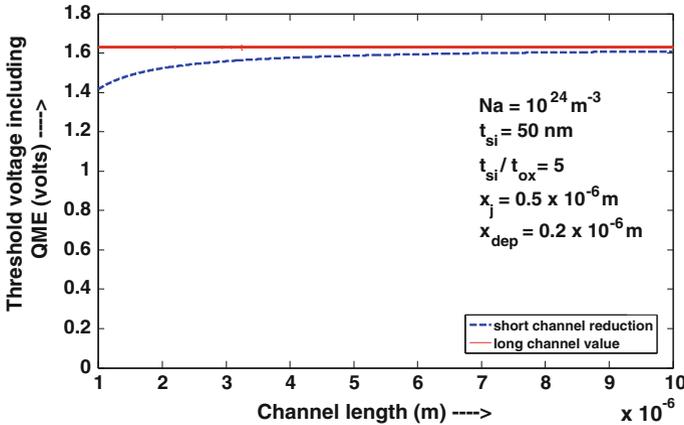


Fig. 21.11 Threshold voltage (including QME) dependence on channel length due to short channel effect (charge sharing)

21.5 Conclusion

In this paper, a quantum threshold voltage model for a nano cylindrical MOSFET is proposed by solving the Poisson’s and Schrodinger’s equations considering moderate doping concentration. The model considers the decoupling of Poisson’s and

Schrodinger's equations via parabolic potential well approximation. The threshold voltage analysis is based on a definition applicable for the quantum approach. By comparing the results with those available in the literature, the parabolic potential well approximation is found to be accurate and efficient for modeling the I-V characteristics of the nano cylindrical MOSFET. A large discrepancy in the device characteristics is seen from the classical analysis while obtaining the results and this proves the need for quantum analysis to be done. Finally, to conclude, this model provides an analytical and useful way for threshold voltage evaluations in nano cylindrical MOSFETs.

References

1. Baccarani, G., Wordeman, M.R., Dennard, R.H.: Generalized scaling theory and its application to a micrometer MOSFET design. *IEEE Trans. Electron Devices* **31**(4), 452–462 (1984)
2. Stern, F.: Self-consistent results for N-type Si silicon inversion layers. *Phys. Rev. B Condens. Matter* **5**(12), 4891–4899 (1972)
3. Ando, T., Fowler, A.B., Stern, F.: Properties of 2-dimensional electron system. *Rev. Mod. Phys.* **54**(2), 437–672 (1982)
4. Song, J., Yu, B., Yuan, Y., Taur, Y.: A review on compact modeling of multiple-gate MOSFETs. *IEEE Trans. Circuits Syst. I Regul. Pap.* **56**(8), 1858–1869 (2009)
5. Sharma, S.: Modeling and simulation of nanobioelectronic device: the cylindrical ion sensitive field effect transistor. Ph.D. thesis. Tezpur University (2009)
6. Singh, J.: *Quantum Mechanics Fundamentals and Applications to Technology*. Wiley, New York (1997)
7. Janik, T., Majkusiak, B.: Analysis of the MOS transistor based on the self-consistent solution to the Schrodinger and Poisson equations and on the local mobility model. *IEEE Trans. Electron Devices* **45**(6), 1263–1271 (1998)
8. Jang, S.L., Liu, S.S.: An analytical surrounding gate MOSFET model. *Solid State Electron.* **42**(5), 721–726 (1998)
9. Streetman, B.G., Banerjee, S.: *Solid State Electronic Devices*, 5th edn. Prentice Hall of India Private Limited, Englewood (2001)

Chapter 22

SVD Analysis on Reduced 3-Lead ECG Data

Sibasankar Padhy and S. Dandapat

Abstract This paper presents synthesis of Electrocardiogram (ECG) leads from reduced set of leads. The Singular Value Decomposition (SVD) is used to train subject-specific all desired leads for minimum of three beat periods. Then, in the testing phase, only 3-leads are used to reconstruct all other leads. The singular value matrix of the reduced 3-lead data is transformed to a higher dimension using a transform matrix. For evaluation purpose, the proposed method is applied to a publicly available database. It contains number of 12-lead ECG recordings with different cardiac patients data. After synthesis of ECG data, the performance of the method is measured using percent correlation present between the original and synthesized data.

Keywords Electrocardiogram (ECG) · Singular Value Decomposition (SVD) · Reduced 3-lead ecg · Percent correlation coefficient · PRD · WEDD

22.1 Introduction

Electrocardiogram (ECG) is the standard graphical tool used to diagnose the patients suffering from Cardiovascular Disease (CVD), a major life threatening disease all over the world [1]. The significant clinical information of an ECG signal, present in the P-wave, QRS-complex, ST-segment, and T-wave [2], are used for diagnosis and monitoring of these patients. It is difficult to analyze all types of cardiac problems using single-lead ECG [3]. A single-lead ECG provides selective information of heart condition [4] in a particular view. Hence, physicians use the standard 12-lead ECG system for critical diagnostic decision. During monitoring of 12-lead ECG signal, ten (nine for ECG acquisition and one as the ground) electrodes are placed over the surface of body. Also number of wires and careful placement of electrodes are needed

S. Padhy (✉) · S. Dandapat
Indian Institute of Technology Guwahati, Guwahati 781039, India
e-mail: sibasankar@iitg.ernet.in

S. Dandapat
e-mail: samaren@iitg.ernet.in

for proper acquisition of ECG data. Patients find difficult in long-term monitoring of the 12-lead ECG data using traditional recording systems.

Researchers have proposed different algorithms to derive 12-lead ECG data from reduced number of leads [5–9]. These works focus to develop a reduced-lead ECG system that can be able to analyze the diagnostic information more close to that of 12-lead ECG system. The algorithms used in these methods use linear transformation techniques to reconstruct other precordial leads from three or four independent leads. Ostertag and Tsouri [8] have proposed patient-specific independent coefficients for other lead reconstruction using Independent Component Analysis (ICA).

The 12-lead ECG data have high redundant information with interlead correlation. The significant redundancy in precordial leads may be due to their close proximal. In this work, we have used singular value decomposition (SVD) to reconstruct other leads from reduced lead set. SVD exploits the redundancy among these leads. Analysis on the synthesis of other precordial leads suggest a competitive result with the existing work [8].

The rest of the paper is organized as follows. Section 22.2 describes the proposed method which discusses the ECG processing and SVD-based training and testing. Results and discussions of the proposed method are discussed in Sect. 22.3. Conclusion with future direction of this method is presented in Sect. 22.4.

22.2 Proposed Method

In this section, the proposed method for synthesizing other leads using three leads (I, II, V2) is discussed. First, preprocessing with QRS peak detection of each lead is performed. Then, Singular Value Decomposition (SVD) is performed during training stage. Then, only three ECG leads (I, II, V2) are processed during testing phase to reconstruct other precordial leads. Details about each step are presented as follows.

22.2.1 ECG Preprocessing

Each ECG leads are preprocessed for removal of baseline wander problem. A Butterworth digital high-pass filter with cutoff frequency $f_c = 0.5$ Hz is used to remove the baseline wander. Amplitude normalization and mean removal are performed to have all the coefficients within the range ± 1 [10]. It helps in faster processing and better reconstruction as the error in smaller coefficients during reconstruction is less. Then, R-peak of a single lead is detected. Each beat period has length of consecutive R-peaks, and is spanned 100 samples (100 ms) left from the current R-peak to 100 samples (100 ms) left from the next R-peak.

22.2.2 SVD-based Training and Testing

After detecting the R-peaks, the data matrix is decomposed using the SVD. The SVD is used to identify and order the data dimensions along which the data points have most variation. After identifying the dimension where it has most variation, it is possible to determine the best approximation of the original data with fewer dimensions. The SVD [11, 12] of a matrix $X_{m \times n}$ is given as $X = U \Sigma V^T$ where $U \in \mathfrak{R}^{m \times m}$, $V \in \mathfrak{R}^{n \times n}$, and $\Sigma_{m \times n} = [\text{diag}\{\sigma_1, \dots, \sigma_r\}; 0]$, $r = \min(m, n)$, and $\sigma_1, \sigma_2, \dots, \sigma_r$ are the singular values. The reduced SVD of the data matrix can be given as $X = \hat{U}_{m \times n} \hat{\Sigma}_{n \times n} V_{n \times n}^T$ where n and m represent the number of leads and corresponding samples of each lead, respectively [11]. The \hat{U} and V represent the orthonormal basis for the column and row space, respectively. In other words, the \hat{U} matrix contains the orthonormal vectors corresponding to the data matrix [13].

Initially, desired number of ECG leads are acquired during training phase. Training is done for minimum of three-beat period data for more accurate reconstruction. The SVD of the training ECG leads is given as

$$X_{m \times n'} = \hat{U}_{m \times n'} \hat{\Sigma}'_{n' \times n'} V_{n' \times n'}^T \quad (22.1)$$

where n' is the number of desired leads (includes reduced 3-leads and other precordial leads) for reconstruction and m has a minimum of three-beat period information. As discussed earlier, the \hat{U} matrix has the information about the ECG leads.

During testing phase, the trained \hat{U} and V matrices are used for reconstruction. However, $\hat{\Sigma}$ matrix is used to transform the reduced 3-lead to a higher dimension using a transform matrix. This $\hat{\Sigma}$ matrix with newly generated $\hat{\Sigma}'$ matrix helps in reconstructing other beat periods. The desired ECG leads can be reconstructed as

$$\tilde{X}_{m \times n'} = \hat{U}_{m \times n'} \hat{\Sigma}'_{n' \times n'} V_{n' \times n'}^T \quad (22.2)$$

The transformation of $\hat{\Sigma}'$ matrix to $\hat{\Sigma}$ is performed by multiplying a transformation matrix.

The transformation of $\hat{\Sigma}'$ matrix to $\hat{\Sigma}$ is performed by multiplying a transformation matrix ($\hat{\Sigma}' \xrightarrow{T} \hat{\Sigma}$) which is given as

$$T_{n' \times 3} = \Sigma_{m \times n'}^{-1} \Sigma'_{m \times 3} \quad (22.3)$$

The dimension of two singular value matrices are enhanced by appending $m - n$ or $m - n'$ rows of zeros.

After reconstruction of all desired leads, the similarity between the original and reconstructed signals is measured using percent correlation coefficient [14]. It is given as

$$\rho = \frac{\sum_{i=1}^m x_i \tilde{x}_i}{\sqrt{\sum_{i=1}^m x_i^2} \sqrt{\sum_{i=1}^m \tilde{x}_i^2}} \times 100 \% \tag{22.4}$$

where x and \tilde{x} represent the original and reconstructed each ECG signal, respectively. Also, two distortion measures namely percentage rms difference (PRD) [15] and wavelet energy-based diagnostic distortion (WEDD) [16] are evaluated for accessing the performance of the synthesized ECG signals.

The PRD is a widely used nondiagnostic distortion measure for its simplicity and is defined as

$$PRD = \sqrt{\frac{\sum_{i=1}^m (x(i) - \tilde{x}(i))^2}{\sum_{i=1}^m x^2(i)}} \times 100 \% \tag{22.5}$$

The WEDD is defined as

$$WEDD = \sum_{j=1}^{L+1} w'_j WPRD_j \tag{22.6}$$

where w'_j and $WPRD_j$ are the energy-based weight factor and the PRD of the j th subband, respectively, and are defined as

$$w'_j = \frac{\sum_{k=1}^{N_j} w_{j,k}^2}{\sum_{j=1}^{M+1} \sum_{k=1}^{N_j} w_{j,k}^2} \tag{22.7}$$

$$WPRD_j = \sqrt{\frac{\sum_{k=1}^{N_j} (w_{j,k} - \tilde{w}_{j,k})^2}{\sum_{k=1}^{N_j} w_{j,k}^2}} \tag{22.8}$$

where $w_{j,k}$ and $\tilde{w}_{j,k}$ are the original and wavelet decomposed coefficients, respectively, and N_j is the number of coefficients in the j th subband.

During evaluation of WEDD, each MEG data is wavelet decomposed at $L = \lceil \log_2(F_s) - 2.96 \rceil$ level [17] where F_s is the sampling frequency.

22.3 Results and Discussions

For evaluation purpose of the proposed method, we have used the 12-lead ECG data from the publicly available database (Physikalisch-Technische Bundesanstalt (PTB) Diagnostic ECG database [18]) in the physionet. This database contains ECG data from 290 subjects (each subject having one to five records) with a total of 549 subjects. Each record includes 15 simultaneously measured signals: the conventional 12 leads (I, II, III, aVR, aVL, aVF, V1-V6) and 3 Frank leads (vx, vy, vz). This database has various pathological data like myocardial infarction, heart failure, bundle branch block, myocardial hypertrophy, etc. Details about the database and extra pathological conditions are available in the database itself [18].

After preprocessing (baseline wander removal and amplitude normalization) of all ECG leads, R-peak of a single lead is detected. In this work, a minimum of three ECG beat periods (approximately 2400 samples) are trained and other beats are used for testing purpose. The orthonormal matrices generated during training phase are directly used for reconstruction of other leads. The singular value matrix of testing phase is transformed with the help of a transformation matrix.

Figure 22.1 shows the actual and reconstructed ECG signals of two bipolar (Lead-I and Lead-II) and six precordial (V1–V6) leads. These signals show the 3-beat period after training phase (after 4th beat period). It can be marked that the leads V1 and V6 have more deviation from the actual one. V2–V5 have more correlation with the actual signals. This is because of selecting Lead-II and V2 as training leads. Other precordial leads, V3–V5, are in the direction of Lead-II and V2, but electrical forces of V1 and V6 are not in accordance with Lead-II and V2. In most of the datasets, it is observed that there is significant deviation in the R-peak and the S-wave onset of Lead-I, V1, and V6.

Table 22.1 shows the percent correlation coefficient (ρ) with PRD and WEDD as a factor to measure the performance of the reconstructed signal compared to the original signal. These values are in well accordance with the Fig. 22.1 as discussed in earlier paragraph. High ρ values of Leads II, V2–V5 suggest better reconstruction of these leads. For validation of the proposed method, ECG beats are synthesized after 15th beat period. The percent correlation coefficients for this case suggest that the proposed method can able to synthesize more close to the original ECG leads. Although lead-II has good percent correlation, the high PRD is due to presence of more noise in the original signal. It is observed that the synthesized leads I, V1, and V6 deviate more significantly from the original signal. This is significant mismatch in R-wave and S-wave of these leads. The high ρ , PRD, and WEDD values of these leads also suggest the poor synthesis of these leads. However, the proposed

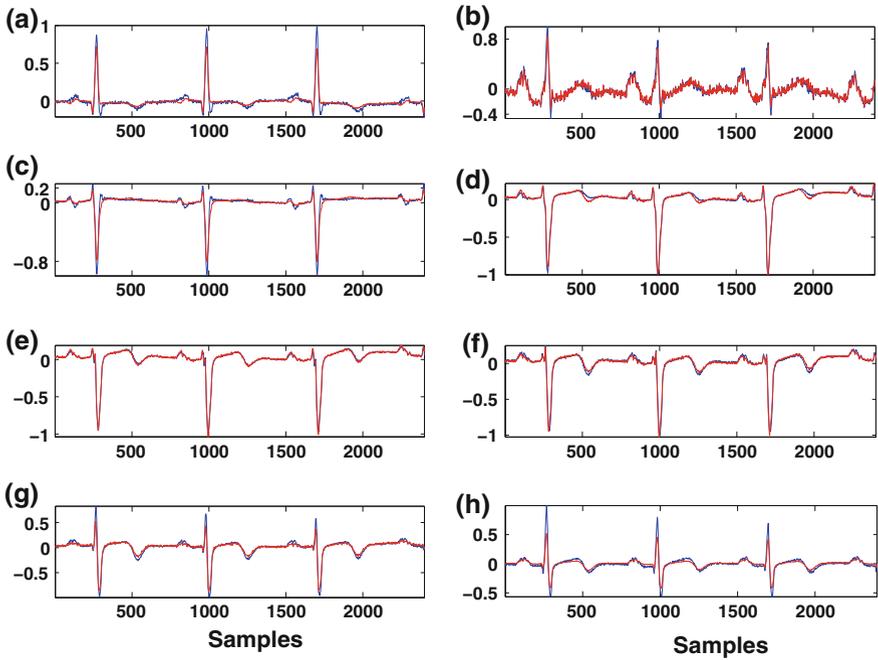


Fig. 22.1 Comparison plot of actual (blue line) and reconstructed (red line) leads after 4th-beat period of 2400 samples. (a) Lead-I, (b) Lead-II, (c) V1, (d) V2, (e) V3, (f) V4, (g) V5, (h) V6, Dataset: *s0046lrem*, PTB Diagnostic Database, Description: Male, Age 75, anterior myocardial infarction

Table 22.1 Percent correlation coefficient

Leads	I	II	V1	V2	V3	V4	V5	V6
ρ	91.5	96.7	92.0	97.9	97.5	97.3	95.8	90.0
	91.9	96.1	91.6	97.5	96.8	97.2	95.1	87.8
PRD	7.85	6.96	8.17	4.52	2.93	3.55	7.41	8.92
	8.59	8.78	9.84	6.38	3.48	4.87	8.32	10.28
WEDD	4.25	3.48	4.02	3.32	2.08	2.31	3.37	4.38
	4.26	3.57	4.17	3.41	2.19	2.87	3.51	4.97

1st row shows the ρ , PRD, and WEDD values after 4th beat period and 2nd row shows the ρ , PRD, and WEDD values after 15th beat period

algorithm is unable to synthesize ECG signals for the subjects suffering from cardiac dysrhythmia. It is due to irregularity of ECG periods.

This work is performed with an offline database (PTB). This method is applied for different cardiac disease data. It is observed that this method fails where the ECG information varies rapidly. It can also be used for online purpose as it has very less computational time. One of the advantage of this method is that the ECG data has to be trained for few seconds, and then the data can be transmitted or reconstructed

with few singular value coefficients. This may help saving energy of the transmitter, and can enhance the battery life of these devices. However, its potential for different pathological data has to be tested. Also, reconstruction performance for long-term data has to be examined.

22.4 Conclusion

In this paper, we presented a SVD-based technique to synthesize the missing ECG leads from a set of 3-lead reduced ECG data. This reduced-lead system will help patient comfort. This preliminary work has no so much improvement compared to the existing methods. However, the performance of this work can be enhanced for future purpose. Hence, it is a great challenge to develop an algorithm that can adapt to the transform coefficients over period of time.

References

1. WHO http://www.who.int/topics/cardiovascular_diseases/en
2. Zigel, Y., Cohen, A., Katz, A.: ECG signal compression using analysis by synthesis coding. *IEEE Trans. Biomed. Eng.* **47**(10), 1308–1316 (2000)
3. Surawicz, B., Knilans, T.K.: *Chous Electrocardiography in Clinical Practice*, 6th edn. Elsevier, Philadelphia (2008)
4. Tsouri, G., Ostertag, M.: Patient-specific 12-Lead ECG reconstruction from sparse electrodes using independent component analysis. *IEEE J. Biomed. Health Inform.* **18**(2), 476–482 (2014)
5. Guillem, M., Sahakian, A., Swiryn, S.: Derivation of orthogonal leads from the 12-lead ECG. accuracy of a single transform for the derivation of atrial and ventricular waves. *Comput. Cardiol.* 2006 249–252 (2006)
6. Mason, R.E., Likar, I.: A new system of multiple-lead exercise electrocardiography. *Am. Heart J.* **71**(2), 196–205 (1966). <http://www.sciencedirect.com/science/article/pii/0002870366901827>
7. Nelwan, S.P., Kors, J.A., Meij, S.H., van Bommel, J.H., Simoons, M.L.: Reconstruction of the 12-lead electrocardiogram from reduced lead sets. *J. Electrocardiol.* **37**(1), pp. 11–18 (2004). <http://www.sciencedirect.com/science/article/pii/S0022073603001729>
8. Ostertag, M., Tsouri, G.: Reconstructing ECG precordial leads from a reduced lead set using independent component analysis. In: *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*. pp. 4414–4417 August 2011
9. Trobec, R., Tomasic, I.: Synthesis of the 12-lead electrocardiogram from differential leads. *IEEE Trans. Inf. Technol. Biomed.* **15**(4), 615–621 (2011)
10. Rajoub, B.A.: An efficient coding algorithm for the compression of ECG signals using the wavelet transform. *IEEE Trans. Biomed. Eng.* **49**(4), 355–362 (2002)
11. Padhy, S., Dandapat, S.: A new multilead ECG data compression method using higher-order singular value decomposition. In: *Twentieth National Conference on Communications (NCC 2014)*. 14 February 2014
12. Wei, J., Chang, C., Chou, N., Jan, G.: ECG data compression using truncated singular value decomposition. *IEEE Trans. Inf. Technol. Biomed.* **5**(4), 290–299 (2001)
13. Clifford, G.D., Azuaje, F., McSharry, P.E.: *Advanced Methods and Tools for ECG Data Analysis. Engineering in Medicine and Biology*. Artech House, Boston (2006)

14. Horek, B.M., Warren, J.W., Wang, J.J.: On designing and testing transformations for derivation of standard 12-lead/18-lead electrocardiograms and vector cardiograms from reduced sets of predictor leads. *J. Electrocardiol.* **41**(3), 220–229 (2008). <http://www.sciencedirect.com/science/article/pii/S0022073608000587>
15. Zigel, Y., Cohen, A., Katz, A.: The weighted diagnostic distortion (WDD) measure for ECG signal compression. *IEEE Trans. Biomed. Eng.* **47**(11), 1422–1430 (2000)
16. Manikandan, M.S., Dandapat, S.: Wavelet energy based diagnostic distortion measure for ECG. *Biomed. Signal Process. Control, Elsevier* **2**, 80–96 (2007)
17. Al-Fahoum, A.S.: Quality assessment of ECG compression techniques using a wavelet-based diagnostic measure. *IEEE Trans. Inf. Technol. Biomed.* **10**(1), 182–191 (2006)
18. Goldberger, A.L., Amaral, L.A.N., Glass, L., Hausdorff, J.M., Ivanov, P.C., Mark, R.G., Mietus, J.E., Moody, G.B., Peng, C.K., Stanley, H.E.: Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. *Circulation* **101**(23): e215–e220. [Circulation Electronic Pages; <http://circ.ahajournals.org/cgi/content/full/101/23/e215>] (June 13, 2000)

Chapter 23

Large Library-Based Regression Test Cases Minimization for Object-Oriented Programs

Swapan Kumar Mondal and Hitesh Tahbildar

Abstract Large library-based regression test cases minimization technique for object-oriented programs has been depicted in this paper. These works have been carried out in three steps. In the first step, the original program is instrumented and executed with test cases. Library is made on the basis of these test cases, coverage of codes, and then the program is modified. In the second step, the modified program is analyzed by latent semantic analysis. It is making the latent semantic matches automatically between users given values and linear combination of its small text objects or variables or database of the software. Therefore, modified code is recorded by latent semantic analysis. Data flow sensitivity and context sensitivity are used for statically and dynamically analyzing the affected and unaffected objects along with the recorded modified codes. After precision data flow analysis, test cases are generated from affected objects with same test cases coverage and affected objects with new test cases coverage. Therefore, redundant test cases are reduced by new optimal page replacement algorithm and updated the library along with code coverage records. In the third step, the test cases of former and modified program are collected and sent to the test cases repository. Now the new optimal page replacement algorithm is implemented on the test cases repository and reduced the regression test suites. An Illustrative example has been presented to establish the effectiveness of our methodology.

Keywords Context sensitivity · Data flow sensitivity · Latent semantic analysis (LSA) · Library · New optimal page replacement algorithm

S.K. Mondal (✉)
University of Science and Technology, Meghalaya 793101, India
e-mail: swapan3121968@gmail.com

H. Tahbildar
Assam Engineering Institute, Guwahati 781003, India
e-mail: tahbil@rediffmail.com

23.1 Introduction

Regression testing is a challenging task before software tester and developer because large numbers of redundant test cases are generated for each changing of software and making complex dependency among programming elements. Sometime original codes are deleted during several times using the library by third parties [1]. Therefore, regression test suites minimization technique is a challenging task. Most of the researchers have been discussed in different literatures [2–6] about procedural program but very few research papers discussed about the object-oriented programs [7, 8]. In this paper, an approach has been made for the regression test cases minimization in object-oriented programs based on large library. Here, the test cases are collected by instrumented code and the program is executed by existing library. The existing library generates the objective summary function based on structural analysis of code coverage of the software. Then the test cases are stored in the repository of diamond structure of UML activity diagram. Now, the program is modified. The modification may be carried out through codes addition, deletion, and changing or new class addition or deletion. The changing elements are analyzed through latent semantic analysis. The latent semantic analysis is making the semantic relationship automatically with the change components of modified software along with their affected statements. The analysis has generated the test cases after analyzing the affected and unaffected objects with the help of data flow sensitivity and context sensitivity. The redundant test cases along with new test cases are collected from affected and unaffected portion of software. The redundant test cases are then removed by the new optimal page replacement algorithm and then updated the library of modified program along with code coverage and library summary informations. Now, the test cases from updated library are sent to the repository of test cases. Therefore, repository contains a huge number of test cases along with redundant test cases (e.g., test cases from former library and modified library). The new optimal page replacement algorithm is implemented on the repository of test cases and reduced test suites. The existing test cases are incremented along with the new test cases of modified library. In this way all the reduced regression test suites are selected which are coverage of the modified program and do not introduce new bugs to the former program.

This paper is organized as follows: In Sect. 23.2, we discussed about the review concepts of different components that provide in detail the need to understand our proposed technique. Our proposed approach has been discussed in Sect. 23.3. An illustrative example has been given in Sect. 23.4. Related work has been explained in Sect. 23.4.1. In Sect. 23.4.2, we explained about the analysis and discussion about our proposed approach and finally conclude the paper in Sect. 23.5.

23.2 Background Concept

23.2.1 *Latent Semantic Analysis*

Semantic analysis is checking the actual meaning of the statements from parser tree. Peoples have used tremendous diversity of words to represent the text objects. For larger text objects, lexical matching between users requirements and database of program is an imprecise and incomplete work of traditional semantic analysis. For overcoming those problems, latent semantic analysis (LSA) theory and methods are used for extracting and representing the contextual-usage meaning of words by statistical computations applied to a large corpus of text [9]. In LSA, the large text objects are decomposed with higher order of indexing structure along with a very high-dimensional semantic space and form the linear combination with the help of singular value decomposition methods (SVD) [10]. The SVD automatically form the lexical matching between user requirements and database of program with orthogonal and diagonal factor of matrix. Therefore, LSA is more helpful for recognizing the changing elements along with relevant statements in programs with iterative programming environment [10].

23.2.2 *Library*

Object-oriented program library is the static representation of the program [11] and also dynamic analysis is carried out. Library-based analysis of program can generate the maximum precision results [11] through precision coverage of object-oriented program paradigm. The library contains with summary of objects about the program under test. Data flow sensitivity and context sensitivity are the main constituents of library and has been used for component-level analysis of affected and unaffected portion of software. Cost of component level analysis is substantially lower than the cost of corresponding whole program analysis of the program. The generated pre-existing intra-procedural and inter-procedural control flow graph can also be reused for software testing, maintenance, verification, and optimization. The flow sensitivity represents the flow of informations through intra-procedurally valid paths and context sensitivity represents through inter-procedurally valid paths. This technique is followed by Rountev1 et al. [12]. The data flow sensitivity processed the informations trough fixed point iteration and meet-over-all-path solution (MOP). During the analysis of affected and unaffected portions of software three basic criteria are needed to fulfill, e.g., transfer or monotonic function [13], confluence operators (e.g., \sqcup and \sqcap) which are used for processing the informations between upper bound and lower bound values and direction of propagation. The monotonic data flow functions work on the principal of lattice. Lattice analyses the whole program based on the tuple (G, L, F, M) where

1. $G = (N, E)$, $N = \text{node}$, $E = \text{edge}$ generates ICFG.
2. $L =$ meet semi lattice, with partial order of set (POSET), meet operation (problem is must), and upper-bounded value.
3. $F \setminus f \mid f : L \rightarrow L$ is monotonic function space that close under functional composition and functional meet.
4. $M : E \rightarrow F$ Monotonic function is functionally dependent on edges.
5. L is the solution at the start node of main. But the context sensitive analysis considers the calling path for forwarding message to other nodes and jumped back to the called sites [14, 15] nodes. Therefore, information is processed from one method to other method by method traversing edge along with monotonic function is called data flow sensitivity, otherwise is called flow insensitive. The path coverage is computed by the mathematical equation: $MVPn = qVP(n)f q()$ where $VP(n)$ is the set of all valid paths. q leading the path ($q(\text{edge}) = e1, e2, e3, ek$) from starting node of main $f q()$ to destination node $p(n)$. Implementation of monotonic function in our approach has already been explained by Sharir and Pnueli [16] in his literature.

23.2.3 Computation of Library Summary Objects

For computation of library summary function, we are considering a simple object-oriented program below (Fig. 23.1).

According to lattice theory

$$f0(x) = (x - 0)U\{(length, 5), (width, 2), (radius, 1)\}$$

$$= \{(length, 5)(width, 2), (radius, 1)\}$$

$[x =$ computed value of area in initial method $f o(x)]$

$$f1(x) = f1(f o(x)) = f1U\{(length, 5)(width, 2), (radius, 1)\}$$

$$= (x - lengthwidth)U\{(length, 5), (width, 2), (radius, 1)\}$$

$$= \{(area, 10), (length, 5), (width, 2), (radius, 1)\}$$

```

L1 Class rect {length=5, width=2; radius=1
L2 Public: void arearect () {
L3 area =length * width;}
L4};
L5 Class squ: public rect {
L6 public: void areasqu() {
L7 Areasqu= width*width;}
L8};
L9 Class circle: public rect {public: void areacir() {
L10 areacir= Π* radius* radius;}
L11};

```

Fig. 23.1 Object-oriented program sample

Here, confluence operator is used for meet operation where problem is must and U operator is used for joining operation where problem may be present. Therefore, library summary information for the object arearect is

$$L1 = (area, 10)U(length, 5), (width, 2)$$

L3

In a similar way the summary information of objects areasqu and areacir are

$$L5 = (areasqu, 4)U(width, 2), (width, 2)$$

L7

$$L9 = (areacir, 3.14)U(radius, 1)respectively.$$

L10

23.2.4 UML Activity Diagram

UML activity diagram is a flow chart of activity. It represents the flow of activity, sequence of activity, parallel flow, branch flow, and concurrent flow of activity of the system. This diagram also captures the dynamic behavior of the system. Our proposed model structure has been represented with UML activity diagram. Rounded rectangles represent action, diamond represents decision, bar represents split and joint of concurrent activities, black circle represents the start and an encircled black circle represents the final states.

23.2.5 Regression Testing Cycle

H. Do et al. have proposed a maintenance model in software development life cycle. According to their proposed model resolution is done after each change requirements. During resolution, failure track reports and change requirements are accumulated. After completion of resolution the regression testing is done. This cycle would be continuing until more regression test failure is occurred. This is called regression testing cycle.

23.2.6 New Optimal Page Replacement Algorithm

Operating system uses optimal page replacement algorithm for virtual memory management. Different algorithm is used for managing the page table with valid reference bit strings. Traditional page replacement algorithm decides which page frame to be swapped out (write to disk) when a page of memory needs to be allocated for updating. The occupied page table with swapped in is used in main memory for CPU processing. Therefore, the page replacement algorithm has the optimal capacity to select the reference strings (e.g., test cases) which are not be used for longest period of

time in future [17]. Though, future detection is impossible. Therefore, this algorithm cannot be implemented practically, though it is a very powerful page replacement algorithm. But the anomalies of this algorithm are illustrated by FIFO (first-in-first-out). LRU (Least recently used) is another page replacement algorithm which is used for virtual memory management. But traditional optimal page replacement algorithm could not be eliminated in the redundant test cases from test suites but it generates the lowest page fault rate than FIFO and LRU. For regression test suites minimization, lowest page fault means lowest number of reduced test suites is desirable. Therefore, optimal page replacement algorithm generates the lowest page fault rate by eliminating redundant test cases. We used traditional optimal page replacement algorithm in combination with insertion sort algorithm. In insertion sort algorithm, each test case is remarkable by integer number and rearranged the test cases along with redundant test cases according to increasing cost. We named it new optimal page replacement algorithm. Here the reference bit string is represented as test cases. Page frame number represents the test suites. Divide the sorted set of test cases as per page frame number for each page. Therefore, we can recognize the redundant test cases among the large number of test suites and is eliminated in a shortest time period by this modified new optimal page replacement algorithm.

23.2.7 Effectiveness of Regression Test Suite

Shin et al. [6] have defined the concept of regression test cases effectiveness in his literature. A set of test requirements r_1, r_2, \dots, r_n for the test suite T . These test requirements must be satisfied to provide the adequate testing of program. Subset of T is T_1, T_2, \dots, T_n . Any one of the test cases t_j that belongs to T_i can be used to achieve the requirements of r_i associated with each r_i of modified program and former program where s is the set of specification.

23.3 Our Proposed Approach

Our proposed approach is minimized the regression test cases based on precision coverage of structural analysis of the program. We now describe the different activities of regression test suites minimization technique in Fig. 23.2 that are carried out during first regression testing cycle.

23.4 An Illustrative Example

We explain our proposed precision coverage-based regression test cases minimization using the example of inherited object-oriented programs.

The existing library contains the test cases and its executed codes. Test cases here are considered for normal execution and caught exception. Now, cout code is

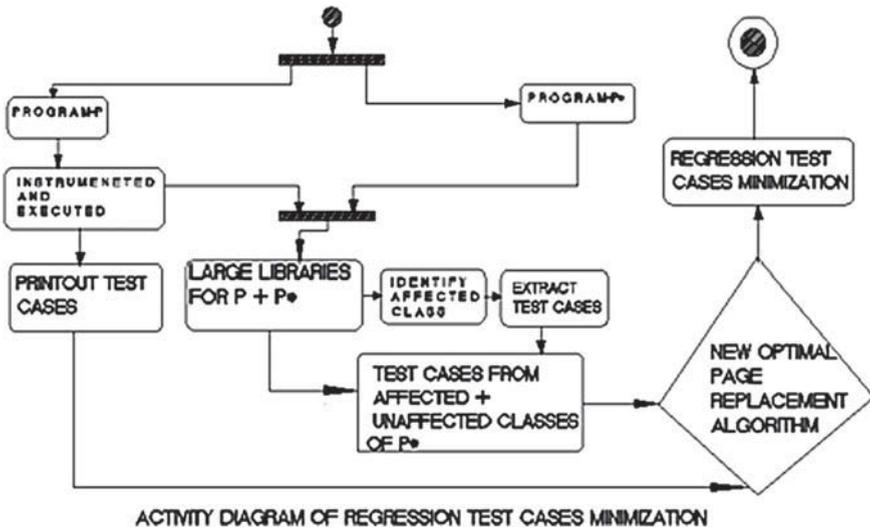


Fig. 23.2 Activity diagram representation of large library-based regression test cases minimization

introduced to the program and taking out print from the program. In this way, all the test cases are taking out print from the program. During execution of the program, cout code is stripped out. Therefore, after instrumented and executed the program, all test cases are collected and send to test cases repository. Then program P is modified and named the modified program is P^* (Table 23.1, Figs. 23.3 and 23.4).

Table 23.1 Library of program P

Test cases	Length	Width	Code execution	Class coverage	Methods coverage	Result
T1	30	20	Length, width, area r	Rectangle	Arearect()	NE
T2	30	Nil	“	“	“	CE
T3	Nil	20	“	“	“	CE
T4	Nil	Nil	“	“	“	CE
T5		20	Width, area s	Square	Areasquare()	NE
T6		Nil	“	“	“	CE

```

class rectangle{private: int length, width, area;// Program P.
public: void arearect ({
arear=length*width;
} }
class square: public
rectangle{
public; int width;
public void
areasquare (){
areas=width* width;}
}
main() {
square s;
s. arearect (30, 20);
s. areasquare (20);
}
    
```

Fig. 23.3 Sample object-oriented program

```

The modified program structure of P* is assumed below: // Program P*
L1: Class rectangle {private: int length, width, arear;
L2: Public: void rectangle ( ) {
L3: arear=length*width;
L4:}}
L5: Class square: public rectangle {int length, width;
L6: Public: void areasquare ( ) {
L7: areas=width*width; }
L8: Public: void diagonalsquare ( ) {
L9: diagonal=v (width + length);}
L10: }
main({
square s;
    
```

Small Text	L1	L3	L5	L7	L9
length	1	1	1	0	1
width	1	1	1	1	1
arear	1	1	0	0	0
areas	0	0	1	1	0

Singular value decomposition methods

```

s. rectangle (30,20);
s. areasquare (20);
s. diagonalsquare (4,9); }
    Modified Program
    
```

Latent semantic analysis showing change components automatically

Small Text	L1	L3	L5	L7	L9
length	1	1	1	0	1
width	1	1	1	1	1
arear	1	1	0	0	0
areas	0	0	1	1	0
diagonal	0	0	1	0	1

Codes / Methods	Statements
diagonal	L5, L9
diagonalsquare ()	L8, L9

L= Related statements for computation.
Recorded modified code (manually)

Fig. 23.4 Modified program

23.4.1 Analyzed Modified Program

See Fig. 23.5, Tables 23.2 and 23.3

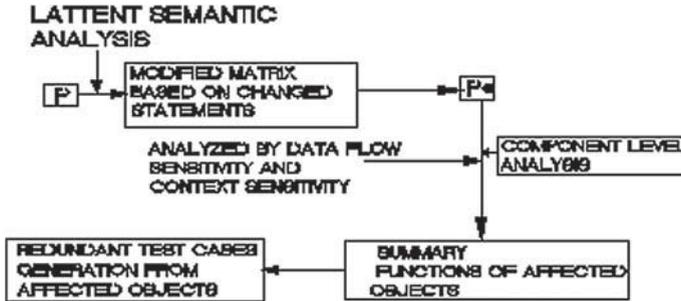


Fig. 23.5 Redundant test cases generation from affected objects

Table 23.2 Recorded modified code

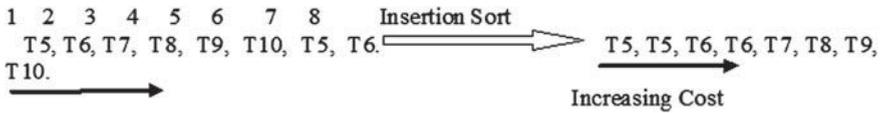


Table 23.3 Redundant test cases from affected objects (T5 and T6 are redundant test cases)

Test Cases	length	width	Code coverage	Method coverage	Class coverage	result
T5		20	width, areas	area square ()	square	NE area=400
T6		20	"	"	"	CE area=410
T7	4	9	length, width, diagonal	Diagonal square ()	"	NE
T8	4	Nil	"	"	"	CE
T9	Nil	9	"	"	"	CE
T10	Nil	Nil	"	"	"	CE

Failure!

23.4.2 Reduced Redundant Test Cases from Affected Objects

It is seen that program P^* in Table 23.3. is contains with T5 and T6 test cases which are coming from former test cases of program P and showing error result. The error result is passed through data flow sensitivity and context sensitivity. Therefore, affected objects are generated the test cases of: T5, T6, T7, T8, T9, T10, T5, T6. Now insertion sort algorithm is used to rearrange the test cases in the following manner. Now, we are assuming page frame is 3. Each test case represented as reference bit string for page frame. We discussed details in Sect. 23.4.1 how optimal page replacement algorithm is used for reducing the redundant test cases. Therefore, the test cases are divided as per page frame number below (Fig. 23.6).

Now, the affected objects contain with the test cases of $T5, T6, T7, T8, T9, T10$. The program P^* contains with inherited classes. The base class is rectangle and derived or inherited class is square. As per testing of inheritance classes, we need to test the base classes also along with derived classes. Therefore, the modified library of program P^* is given in Table 23.4.

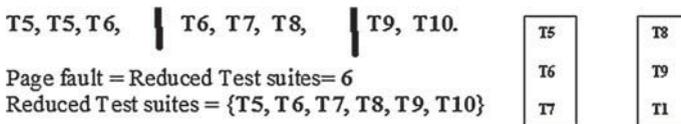


Fig. 23.6 Page frame contains with test cases

Table 23.4 Test cases of modified program P^*

Test cases	Code coverage	Class coverage	Method coverage	Result
T1	area, length=30, width=20,	rectangle	rectangle ()	NE
T2	“	“	“	CE
T3	“	“	“	CE
T4	“	“	“	CE
T5	areas, width=20	square	areasquare ()	NE(areas=400)
T6	“	“	“	CE (areas=400)
T7	length, width, diagonal	“	diagonalsquare ()	NE
T8	“	“	“	CE
T9	“	“	“	CE
T10	“	“	“	CE

Passed

23.4.3 *Reduced Redundant Test Cases from Test Cases Repository*

Algorithm 1: Pseudocode to minimize regression test cases through new optimal page replacement algorithm (Figs. 23.7 and 23.8).

23.5 Related Work

Some research results have been published on regression test cases minimization technique in [18–22]. They proposed firewall-based technique [23], design model-based technique [18, 23, 27], program model-based technique [24]. Very few researchers have proposed regression test cases minimization technique in object-oriented programs.

Eric Wong et al. [27]

This paper, different techniques of regression test cases minimization have been proposed which have reduced the cost of testing, e.g., modification-based test selection, coverage-based test case selection, and risk analysis-based test cases selection. The cost is reduced by this better proposed technique. Faults are analyzed through execution slice. It is different from static, dynamic, backward and forward slicing

```

Input: {T1,T2,T3,T4,T5,T6,T5,T6,T7,T8,T9,T10}
       $\xrightarrow{\text{Increasing Cost}}$ 
Output: {T1, T2, T3, T4, T5, T6, T7, T8, T9, T10}

1. Test suite reduction
2. Rearrange test cases according to incremental cost through
   insertion sort
3. Test cases ( $T_i$ ) = reference bit string
4. Assume each page contains =3 test cases (page frame)
5. Page fault number = reduced test suites
6. while targets to cover != ? and each page with frame no. <=3
7. min. page faults (MPF) =Targets to cover, Page faults=0
8. while not covered (MPF) and page faults < max Page faults
9. Select test cases whose cost is minimum and send to page as
   page frame
10. Update page with increment per unit cost of test cases
11. If test cases having same cost discard its
12. Replaced the test cases whose cost is comparatively less
13. If covered (MPF) break
14. Page faults = Page faults + 1
15. end while
16. end while

```

Fig. 23.7 Pseudocode of new optimal page replacement algorithm



Fig. 23.8 Witness of regression test selection

techniques. In execution slice, test cases along with coverage statements are represented as block. The dependent affected blocks are used to find out the causes of errors. Execution slice not only finds out the error but also helps for decision-making with respect to the corresponding blocks. Greedy heuristic search algorithm is used here for minimization of test cases within exponential in time because regression test cases minimization is a NP complete problem [3, 20]. Yoo [6] and Chen [25] proposed the regression test minimization technique. The minimization technique is represented as a minimal hitting set problem. The minimal hitting set problem is a NP complete problem. Because minimized test suites are used for dual coverage. It is used for testing of original program and modified program. Different heuristics have been utilized like GE and GRE heuristics and compared the results with HGS (Harrold–Gupta–Sofa). In GE heuristic first essential test cases are selected from test suites. Additional greedy algorithm is used to select the remaining test cases. In GRE heuristic, the redundant test cases are first removed from the test suites. Then heuristic is applied on the reduced test suites. **Critical Evaluation** Greedy heuristic search-based regression test cases minimization is a very low cost technique. It is easy to use and cheap procedure. It works on the principal of observed usability problem with reference to established usability principle [16]. The reduced test suites are also more effective and efficient coverage of former and modified program. But utilization of proposed technique in object-oriented programs has not been described. But heuristic search-based technique does not give the accurate solution. Therefore, approximate solution is obtained by this proposed algorithm [30]. Moreover, greedy heuristic search loses the structure and overloop some states for computation of better states. Therefore, it is risk to find out the test cases and does not allow a way to assess the quality of redesign. Wong also proposed in greedy algorithm to select each test case first whose cost is zero. For the remaining test cases selection, increment maximal coverage with per unit cost is selected. But test cases whose cost is zero maybe never possible because minimum cost is generated for each test case due to hardware and software requirements during testing.

23.6 Analysis and Discussion

23.6.1 Cycle Time is Reduced

LSA analyzing details about the changed statements and changed impact on other component of software by singular value decomposition method. This analysis also reduced time for editing and execution [9]. As a result regression testing cycle time is reduced.

23.6.2 Simple Model Structure

The minor changed components affect on traversing paths. Those traversing paths are recognized by meet-over-all-valid-paths-selection equation. Therefore, affected and unaffected class can be analyzed by LSA along with library summary informations. No UML state diagram [19] is recurred to find out the affected statements. As a result the model structure becoming simple and can be reused by developer.

23.6.3 Safe Testing Technique

The data flow sensitivity and context sensitivity return the informations to the calling nodes without losing their precision. It is safe, valid, and relevant information flow technique [11].

23.6.4 Efficiency Development of Model Driven Development

UML activity diagram can represent two or more threads of concurrent execution. When two or more class acts as input to the computation, then those set of classes are called thread. As a result flow chart of activity can be extending. This diagram also supports the all features of flow graph. As a result control flow-based technique model can be developed and maybe useful for path coverage [26].

Page replacement algorithm offers the near-optimal performance. This optimality has been increased with insertion sort algorithm. Therefore, theoretically the developed optimal page replacement algorithm can be used in huge number of redundant test cases for identifying and removing the redundant test cases permanently. We have implemented the structure of optimal page replacement algorithm but practically each test case is considered here as the incrementing cost. The lowest cost of test cases in page frame is always replaced because those test cases should never be used and have arranged in increasing order of cost through insertion sort (Fig. 23.9).

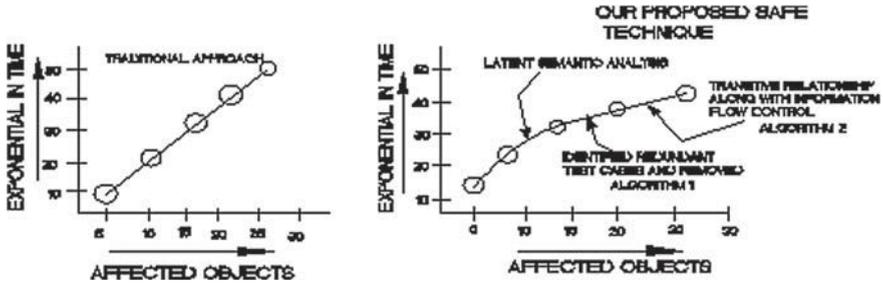


Fig. 23.9 Pseudocode of new optimal page replacement algorithm

23.7 Conclusion

Large library-based regression test cases minimization technique in object-oriented programs have been depicted in this paper. The library summary function represents the test cases along with code coverage and executed result of existing program. The modified program is analyzed by latent semantic analysis and precise data analytical technique. The new test cases and redundant test cases are recognized from modified library. The redundant test cases are removed by the new optimal page replacement algorithm. Finally, the existing test cases of former program are incrementally updated by new test cases from existing library and modified library. This technique is more effective to revealing faults from modified program and more cost effective.

References

1. Sen, T., Mall, R.: State-model-based regression test reduction for component-based software. *Int. Sch. Res. Netw. ISRN Softw. Eng.* 9 pp. (2012) Article ID 561502
2. Suri, B., Mangal, I., Srivastava, V.: Regression test suite reduction using an hybrid technique based on BCO and genetic algorithm. pp. 165–172. *IJCSI, ISSN (PRINT):2231–5292*
3. Sharir, M., Pnueli, A.: Two approaches to inter procedural data flow analysis. In: *Program Flow Analysis: Theory and Applications*. pp. 189–234. (1981)
4. Frechette, N., Badri, L., Badri, M.: Regression test reduction for object-oriented software: a control call graph based technique and associated tool. *ISRN Softw. Eng.*, Article ID 420394, 10 pp. (2013)
5. Rao, S., Medeiros, H. Kak, A.: Comparing incremental latent semantic analysis algorithms for efficient retrieval from software libraries for bug localization. Published. In: *2nd International Workshop on Software Mining, Palo Alto, California, November 11 2013*
6. Supriya, Chinky Aneja Mtech research scholar.: Test case minimization for object oriented technique on the basis of object oriented coupling. *J. Res. Electr. Electron. Eng. (ISTP-JREEE)*, ISSN: 2321–2667, 2(4), pp. 11–17 (2013)
7. Mansour, N., El-Fakih, K.: Simulated annealing and genetic algorithms for optimal regression testing. *J. Softw. Maint. Res. Pract.* 11(1), 1934 (1999)
8. Biswas, Swarnendu, Mall, Rajib, Satpathy, Manoranjan, Sukumaran, Srihari: Regression test selection techniques: a survey. *Informatica* 35, 289–321 (2011)

9. Yoo, S., Harman, M.: Regression Testing Minimization, Selection and prioritization: A Survey. Wiley Inter Science (www.interscience.wiley.com). 2007, DOI: 10.1002/000, Reliab. pp. 17
10. Rothermel G.G., Harrold, M.: Selecting regression tests for object-oriented software. In: International Conference on Software Maintenance, p. 1425, March 1994
11. <http://www.cl.cam.ac.uk/teaching/1011/L111/ip-dfa.pdf>
12. Rountev1, A., Kagan1, S., Marlowe2, T.: Inter Procedural Dataflow Analysis in the Presence of Large Libraries. 1 Ohio State University, Columbus, OH, USA, 2 Seton Hall University, South Orange, NJ, USA, pp. 2–16 (2006)
13. www.google.co.in/#q=monotone+data+flow+problem+ppt
14. http://en.wikipedia.org/wiki/Data_flow_analysis
15. www.cs.umd.edu/class/spring2011/cmsc430/lectures/lec09.pdf
16. <http://www.scribd.com/doc/13301539/Advantages-and-Disadvantages-of-Performing-Heuristic-Evaluations>
17. <http://knewpedia.com/the-optimal-page-replacement-algorithm>
18. Ali, A., Nadeem, A., Iqbal, Z., Usman, M.: Regression testing based on UML design models. In: Proceedings of the 13th Pacific Rim International Symposium on Dependable Computing, pp. 85–88 (2007)
19. Panigrahi, C.R., Mall, R.: A hybrid regression test selection technique for object-oriented programs. Proc. Int. J. Softw. Eng. Appl. **6**(4) (2012)
20. Leung, H., White, L.: Insights into regression testing. In: Proceedings of the Conference on Software Maintenance, pp. 60–69 (1989)
21. Lin, J., Huang, C., Lin, C.: Test suite reduction analysis with enhanced tie-breaking techniques. In: 4th IEEE International Conference on Management of Innovation and Technology. ICMIT 2008, pp. 1228–1233, September 2008
22. Chen, T.Y., Lau, M.F.: Dividing strategies for the optimization of a test suite. Inf. Process. Lett. **60**(3), 135–141 (1996)
23. Kung, D., Gao, J., Hsia, P., Toyoshima, Y., Chen, C.: Firewall regression testing and software maintenance of object—oriented systems. J. Object-Oriented Program. (1997)
24. Orso, A., Shi, N., Harrold, M.: Scaling regression testing to large software systems. In Proceedings of the 12th ACM SIGSOFT Twelfth International Symposium on Foundations of Software Engineering, pp. 241–251, November 2004
25. Ye Wu, Mei-Hwa Chen and Howard M. Kao.: Regression Testing on Object-Oriented Programs. wuye, mhc, kao) @cs.albany.edu (<http://staff.unak.is/andy/MScTestingMaintenance0607/Lectures/RegressionTesting.pdf>)
26. www.phindia.com/rajibmall/chapters/LECTURE16.ppt
27. W. Eric Wong.: Reduce Cost of Regression Testing <http://www.utdallas.edu/ewong>. (2012)
28. Harrold, M.J., Gupta, R., Sofia, M.L.: A methodology for controlling the size of a test suite. ACM Trans. Softw. Eng. Methodol. **2**(3), 270–285 (1993)
29. http://en.wikipedia.org/wiki/Singular_value_decomposition

Author Index

A

Abdulkareem, V., 81

B

Baishya, Rubi, 171
Barbhuiya, Ferdous A., 3
Baruah, R.R., 231
Baruah, Sunandan, 41
Basishtha, Subhash, 183
Bhuyan, M.P., 231
Biswas, Santosh, 3
Bora, P.K., 81

C

Choudhury, Hussain Ahmed, 149

D

Dandapat, S., 17, 73, 161, 201, 211, 253
Das, Kaushik, 221
Das, Rohan Kumar, 193
Datta, Sumit, 115
Deka, Bhabesh, 115
Deka, Jyotisikha, 241
Dey, Samrat, 103

H

Halder, Anindya, 103

K

Kakoty, Priyanka, 221
Kalita, Mayuri, 61
Karthik, K., 81

Khanikar, P., 231
Kumar, Ansuman, 103
Kumar Mondal, Swapan, 261

N

Nallikuzhy, Jiss J., 211
Nandi, Sukumar, 3
Nath, Keshab, 183

P

Padhy, Sibasankar, 253
Patanaiik, Chinmaya K., 3
Prasanna, S.R.M., 127, 139, 193
Priya, Bhanu, 161

S

Saikia, Monjul, 149
Saikia, Torali, 93
Sarma, H., 231
Sarma, Biswajit Dev, 127, 139
Sarma, Kandarpa Kumar, 61, 93
Sarma, Meghamallika, 127
Sarma, Mousmita, 139
Sarma, Utpal, 171
Sharma, L.N., 17
Sharma, Santanu, 241
Singh, Anurag, 201
Sri Ranjani, S., 81

T

Tiru, Banty, 171
Tripathy, R.K., 17, 73

Subject Index

Symbols

12-lead electrocardiogram, 212

A

ALFKNN, 111

ANFIS, 94

Antivirus, 3

ARPS, 149

Artificial Neural Network, 62

Assamese, 183

Atrio-ventricular (AV), 18

Average pitch value (APV), 133

AWEDD, 75, 79

B

Back-propagation, 62

Bagle.A, 3

Bandgap, 41

Beat synchronous chroma, 82

Botnets, 3

Burrows-Wheeler Transform, 232

C

Cardiac arrhythmias, 211

Cardiovascular, 18

Clustering, 162

Code obfuscation, 4

Codewords, 233

Competitive learning, 62

Compressed sensing (CS), 115

Compression ratio (CR), 201

COMSOL MULTIPHYSICS 4.2, 223

Context sensitivity, 263

Contraction, 18

Cosine, 162

Cosine kernel classifier, 195

Couplers, 179

CR, 62

CS-MRI, 118

CTSLs, 154

Cylindrical MOSFET, 241

D

DCT, 202

Decomposition, 65

Defuzzification, 95

Delta chroma, 82

Diagnostic, 18

Diphthongs, 142

Doping, 246

Drain current, 246

DSSC, 43

DWT, 62

Dynamic time warping (DTW), 82

E

Electrocardiogram (ECG), 18, 201, 253

Electromagnetic wave, 175

Energy, 41

Euclidean, 22

Euclidean distance, 105

Executable, 4

F

Feature, 187

Feedforward, 62

Filtering, 18

Finite state automata, 4

Frequency, 62
Fuzzy Inference System, 94

G

Gaussian mixture model (GMM), 194
Genotype, 104
Glottal, 130
GRE heuristics, 272

H

Haar DWT, 65
Hanning window, 166
Heart, 18
Hidden Markov Models (HMMs), 140
HIS bundle, 18
HMM tool kit (HTK), 143
Hydrothermal, 42
Hypercalcemia, 19
Hyperkalemia, 19
Hypokalemia, 19

I

Inner product space, 162
Intrusion Detection System (IDS), 3

K

Key, 82
K-space, 116

L

Large library based regression, 262
Lattice analysis, 263
LCR circuit, 177
Learning Vector Quantization (LVQ), 62

M

Magnetic resonance (MR), 115
Magnetic Resonance Imaging (MRI), 115
Mahalanobis distance, 105
Malwares, 3
Matlab, 83
Mean square error (MSE), 22
MECG, 73
Medulloblastomas, 110
Microheater, 222
MoCA, 171
Monophone, 143
Motion estimation, 149

MPRD, 75
MSPCA, 75
Multichannel ECG (MECG), 202
Multi-Layer Perceptron (MLP), 62
Multilead ECG, 73

N

Nanomaterials, 41
Nanotechnology, 42
Neuro-Fuzzy Systems (NFS), 94
Neurons, 62
NIST SRE 2003, 197

O

On-line, 185

P

Pacemaker, 18
Pathological, 17
Perceptual, 81
pH, 43
Phenotypic, 104
Phonetics, 127
Photovoltaic, 42
Physionet, 257
Piezogenerators, 41
Pitch, 128
Pitch marking, 127
Platinum, 222
Point spread function (PSF), 116
Poissons equation, 242
Poly-Silicon, 222
Power Line Communication (PLC), 171
PRD, 17
Primal-Dual Interior-Point Method (PDIPM), 120
Principal Component Analysis (PCA), 62
Probabilistic induction, 104
Prosodic, 128
Prosody, 127
Protocol, 4
Proximity effect, 215
PSNR, 62, 150
Purkinje fibers, 18
P-wave, 18, 253

Q

Quantummechanical effects (QMEs), 241

R

Recognizer, 168
Reconstruction, 65
Relaxation, 18
Rhythm, 127
RMSE, 18
Rootkits, 3
RTSLS, 154

S

Scattering matrices, 172
Schrödinger equation, 242
SCPRD, 74, 79
Self-Organizing Maps (SOM), 62
Semantic analysis, 263
Sensors, 42
Septum, 18
SHAZAM, 82
Silicon, 42
Sino-atrial (SA), 18
Sinusoidal model, 162
Solar cells, 42
Speaker verification, 193
Spywares, 3
StarNT, 233
StraceNt, 6
SVD, 254
Syllables, 139
Synthesis, 18

T

Takagi-Sugeno FIS, 94
Tempo, 82, 127
Thin films, 42

Thresholding, 188

Timbre, 82

T-matrix, 195

Tones and break, 128

Truncated Newton Interior-Point Method
(TNIPM), 120

T-wave, 19

U

UML activity diagram, 265

Universal background model (UBM), 195

Unsupervised, 62

USDZQ, 205

V

Vapor-deposition, 42

Variable density, 116

Ventricles, 18

Virtual memory, 265

Viruses, 3

Voice activity detection (VAD), 194

W

Wavelets, 117

WEDD, 17

WEWPRD, 27

Worms, 3

Z

Zero frequency filtering (ZFF), 130

ZMRTSLS, 154

ZnO, 41