

**A Multiresolution Model of Auditory Excitation
Pattern and its Application to Objective
Evaluation of Perceived Speech Quality**

By

Abhijit Karmakar

Department of Electrical Engineering

*Submitted
in fulfillment of the requirements
for the degree of*

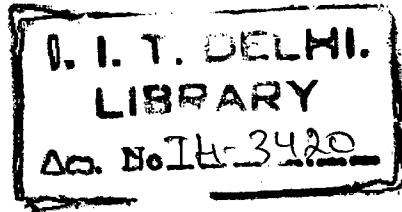
Doctor of Philosophy

to the



**Indian Institute of Technology Delhi, India
January 2007**

Sound recording @ Measurement of speech quality.

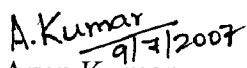


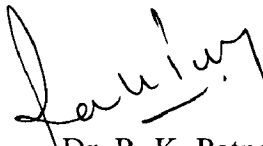
TH
681.841
KAR-M



Certificate

This is to certify that the thesis entitled “A Multiresolution Model of Auditory Excitation Pattern and its Application to Objective Evaluation of Perceived Speech Quality” being submitted by Mr. Abhijit Karmakar for the award of the degree of Doctor of Philosophy to the Department of Electrical Engineering, Indian Institute of Technology Delhi, is a record of bonafide work done by him under our joint supervision and guidance. In our opinion, the thesis has reached the standards fulfilling the requirements of the regulations relating to the degree. The matter embodied in this thesis has not been submitted to any other University or Institute for the award of any degree or diploma.


Dr. Arun Kumar
Associate Professor
Center for Applied Research in Electronics
Indian Institute of Technology Delhi
New Delhi – 110016
India


Dr. R. K. Patney
Professor
Department of Electrical Engineering
Indian Institute of Technology Delhi
New Delhi – 110016
India

Acknowledgement

I express my deep sense of gratitude to my supervisors, Dr. Arun Kumar and Prof. R. K. Patney for their valuable guidance and encouragement throughout the course of my work. I owe them a great deal for the precious time they shared with me and supporting me throughout the work. It has really been a rewarding experience, for the exposure I have gained from their breadth of knowledge and work ethics.

I would like to thank the student research committee (SRC) members, Dr. Brejesh Lall, Prof. Ranjan Bose, Prof. S. D. Joshi and Prof. M. Balakrishnan, for all the constructive ideas and suggestions throughout my research work. I would like to pay my special regards to Prof. G. S. Visweswaran and Prof. S. C. Dutta Roy for all the encouragement and support from time to time.

I would like to specially thank Mr. Arnab Das, Mr. S. S. Lokesh, Mr. M. Ravibabu, Mr. T. S. Rawat, Mr. B. Prabhu, Mr. Gaurav Saxena, Mr. Lalit Jiواني and Mr. Sanjay Nalwalwar for their lively company and the help they provided to me at various stages. I would also like to take this opportunity to thank all the speakers whose voices were recorded and all the listeners, who participated in the subjective listening test during my research work.

I would like to thank Dr. Chandra Shekhar, Director, CEERI, Pilani for his encouragement throughout the period of my PhD work. My special thanks are due to my colleagues, Mr. Raj Singh, Dr. S. C. Bose, Mr. Sudhir Kumar and Dr. A. S. Mandal, for the support they extended to me from time to time.

Finally, I would like to thank my parents for their love and inspiration. I express my special thanks to my wife for her encouragement, particularly during those not-so-great moments during the period of work.

Abhijit Karmakar

Abhijit Karmakar

Abstract

The work in this thesis proposes a wavelet packet (WP) based multiresolution auditory model (MRAM) for the peripheral auditory system and applies it to the problem of objective evaluation of wideband speech quality. MRAM produces the multiresolution auditory excitation pattern of an acoustic stimulus at different time-frequency resolutions that are matched to the perceptual frequency scale. The design of the WP-tree rests on an optimality criterion formulated to minimize a cost function based on the critical band (CB) structure of human auditory system. Also, the wavelet to be used for the WP-based decomposition is designed so that it transfers close to true energy of the input signal in the critical bandwidths, to the respective critical band channels.

The existing models of different auditory phenomena have been reformulated for the WP-based multiresolution framework. The instantaneous energy components available at the critical bands from the WP-based decomposition are obtained at different time resolutions dependent on the depth of the WP-tree. To take care of this, the OME weighting function is recast with a weighting function dependent on the duration of the CB energy components. Next, the OME-weighted energy components in the different CB channels are arranged in time-aligned blocks and spectral spreading is applied to them. For this purpose, the model of spectral spreading is reformulated and a scheme of multiresolution spectral spreading has been proposed. The model of temporal smearing is also applied at different rates determined by the time resolution of the critical band channels. Finally, to incorporate the effect of subjective loudness, a model of power-law of compression is applied.

The excitation pattern so obtained is used to define an objective measure of auditory distortion of a distorted speech signal compared to the undistorted one. The performance of this objective measure is evaluated with a database of various kinds of NOISEX-92 degraded wideband speech signals for predicting the subjective mean opinion score (MOS) and is compared with the STFT-based ITU-T PESQ P.862.2 algorithm. The proposed measure is found to achieve improved correlation between subjective MOS and objective MOS as PESQ P.862.2, with a trend suggesting better correlation for the non-stationary degradations compared to the stationary ones.

In this thesis, we also propose a design method for optimal two-channel quadrature mirror filter and cyclic orthogonal filterbank using convex optimization technique. This simple method of optimization minimizes the stopband energy of the filter impulse response and poses the problem as a semidefinite programming problem.

Contents

List of Figures	(vii)
List of Tables	(xi)
List of Symbols	(xiii)
List of Abbreviations	(xv)
1 Introduction	1
1.1 Background and Motivation	1
1.2 Some Important Auditory Phenomena and their Psychoacoustic Models	4
1.2.1 Absolute Hearing Threshold	5
1.2.2 Simultaneous Masking and Critical Band Rate Scale	6
1.2.3 Temporal Masking	10
1.2.4 Loudness Perception	11
1.2.5 Subjective Loudness	12
1.3 Objective Estimation of Subjective Speech Quality and STFT-based Auditory Models	13
1.4 Conceptual Framework of the Proposed Multiresolution Auditory Model	15
1.5 Objective and Scope of the Work	18
1.6 Outline of the Thesis	21
2 Design of an Optimal Wavelet Packet Tree based on Auditory Perception	25
2.1 Introduction	25
2.2 Discrete Wavelet Transform	26
2.2.1 Decomposition of Signal Space and DWT Coefficients	26
2.2.2 Implementation of DWT	28

2.3	Wavelet Packet Transform	31
2.3.1	Library of Wavelet Packet Basis Functions	32
2.3.2	Ordering of Frequency Bands	34
2.4	Design of Perceptually Motivated Optimal Wavelet Packet Tree	35
2.4.1	Perceptual Frequency Scale	35
2.4.2	Criterion of Optimality for the WP-tree based on Bark scale	36
2.4.3	Optimal WP-tree and Frequency Band Ordered WP basis for 16 kHz	
	Sampling Frequency	41
2.5	Results	45
2.6	Conclusions	47
3	Design of an Optimal Wavelet based on Auditory Perception	49
3.1	Introduction	49
3.2	Design of Perceptually Motivated Wavelet	50
3.3	Results	62
3.4	Conclusions	66
4	Multiresolution Auditory Model	67
4.1	Introduction	67
4.2	Time-Frequency Decomposition and the Auditory Band Indexed WPT Coefficients	71
4.3	Energy Computation and Duration Dependent OME Weighting	73
4.4	Multiresolution Spectral Spreading	76
4.5	Temporal Smearing and Subjective Loudness	82
4.6	Results	85
4.7	Conclusions	88
5	Wideband Speech Quality Evaluation Using MRAM	89
5.1	Introduction	89
5.2	Objective Speech Quality Measures	90
5.2.1	Conventional Objective Measures	90
5.2.2	Subjective versus Objective Measure	92

5.2.3	Perceptually Motivated Speech Quality Measures	93
5.3	Speech Quality Evaluation in the Framework of MRAM	97
5.3.1	Model of Perceptual Transformation: MRAM	98
5.3.2	Measure of Auditory Distortion	99
5.3.3	Distortion-to-MOS Mapping	100
5.4	Preparation of Speech Database and Subjective Listening Test	100
5.4.1	Preparation of Clean Speech Database	101
5.4.2	Preparation of Degraded Speech Database	101
5.4.3	Subjective Listening Test and Performance Evaluation Strategy	102
5.4.4	Measures of Performance	103
5.5	Results	106
5.6	Conclusions	122
6	Design of Optimal Two-Channel FIR Orthogonal Filterbank	123
6.1	Introduction	123
6.2	Semidefinite Programming	125
6.3	Properties of QMF Filterbank and the Criterion of Optimality	127
6.4	Design of Energy Compaction Two-Channel FIR QMF Filterbank	128
6.4.1	Objective Function in Terms of Autocorrelation Sequence	128
6.4.2	PSD Characterization of Autocorrelation Sequence	129
6.4.3	QMF Constraints in SDP Framework	130
6.4.4	Filter Design Problem in SDP Framework	131
6.4.5	A Design Example	131
6.5	Design of Energy Compaction Two-Channel Orthogonal Cyclic Filterbank	133
6.5.1	Properties of Orthogonal CFB and the Criterion of Optimality	133
6.5.2	Objective Function and Orthogonal CFB Constraints in SDP Framework	135
6.5.3	SDP Problem Formulation and its Solution	137
6.5.4	A Design Example	138
6.6	Conclusions	140

7	Conclusions	141
7.1	Key Contributions	141
7.2	Limitation of the Work	142
7.3	Scope for Future Work	143
A	Appendix	145
A.1	Speech Database	145
A.2	ACR-LQ Test	147
B	Appendix	149
B.1	Delay Compensation in MRAM	149
B.2	Wideband Speech Quality Measurement Using Delay-Compensated MRAM	151
	References	155
	List of Publications	165