

LB/DON/69/02

6



University of Moratuwa

DSP BASED SPEECH TRAINING OF HEARING IMPAIRED CHILDREN



University of Moratuwa, Sri Lanka
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Submitted to the department of
Electronic & Telecommunication Engineering
in partial fulfillment of the requirements
for the Degree of Master of Engineering

සමස්ත කාර්යය
සාරාංශය විසින් විද්‍යාලය, ශ්‍රී ලංකාව
සමරවනු ලබයි.

D.G. WEERASINGHE

April 2002

621.38 "02

621.391

TH

074679



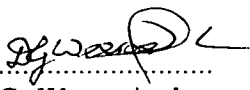
University of Moratuwa

74679

74679

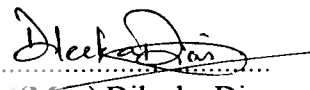
Declaration

The work presented in this dissertation has not been submitted for the fulfillment of any other degree.


.....
D.G. Weerasinghe
(Candidate)



University of Moratuwa, Sri Lanka
Electronic Theses & Dissertations
www.lib.mrt.ac.lk


.....
Dr. (Mrs.) Dileeka Dias
(Supervisor)

ABSTRACT

A study based on several digital signal processing (DSP) techniques to be used in the development of a computer-based speech trainer for hearing impaired children is presented.

Children with congenital hearing impairments have difficulties in speaking, and even in making the basic sounds associated with speech. Speech therapists use specialized training methods to train such children. The dearth of qualified speech therapists and other facilities hinder the speech development of many children in need of such training in most of the third world countries. The speech trainer described in this dissertation was developed as an alleviation to the above problem.

The training tool developed, will aid a child with initial guidance from an adult, to master the pronunciation of initial sounds taught in a speech therapy programme, in a game-like environment, with only a PC having multimedia facilities.



Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Three DSP techniques were studied for application to the trainer. The objective was to identify whether an utterance by a trainee was acceptable or not compared to an utterance by a normal person. The three techniques were based on spectral analysis, formant analysis and neural networks. The results with the spectral technique were found to be superior and were selected for use in the development of the training tool.

In its current status, the training tool can guide children in pronouncing the five vowel sounds, the first step in a speech therapy course.

List of Figures

1.1	Block diagram of the speech trainer	3
3.1	Spectrograms for normal and hearing impaired speakers for /eI/ sound	11
3.2	Flow chart for /a:/ sound	25
3.3	Flow chart for /æ/ sound	26
3.4	Flow chart for /u:/ sound	27
3.5	Flow chart for /o/ sound	28
3.6	Flow chart for /eI/ sound	29
3.7	Plot of target values and a correct attempt of a speaker	30
3.8	Plot of target values and an incorrect attempt of a speaker	30
4.1	Vocal tract for /a:/ sound	35
4.2	Vocal tract for /æ/ sound	35
4.3	Vocal tract for /u:/ sound	36
4.4	Vocal tract for /o/ sound	36
4.5	Vocal tract for /eI/ sound	37
4.6	Location of formants for /a:/ sound	38
4.7	Location of formants for /æ/ sound	38
4.8	Location of formants for /u:/ sound	39
4.9	Location of formants for /o/ sound	39
4.10	Location of formants for /eI/ sound	40
4.11	Area for the location of /a:/ sound	43
4.12	Area for the location of /æ/ sound	44
4.13	Area for the location of /u:/ sound	44
4.14	Area for the location of /o/ sound	45
4.15	Area for the location of /eI/ sound	45
5.1	Structure of a 3 layer feed forward neural network	48
5.2	Inputs, outputs and weights of the network	49
5.3	Unipolar sigmoidal function	51
5.4	Bipolar sigmoidal function	51
5.5	Flow chart for back-propagation learning algorithm	53
7.1	Display sequence with all color balloons for correct pronunciation	62
7.2	Display with colorless balloons for incorrect pronunciation	63
7.3	Flow chart for visual interface operation	64
7.4	Project workspace of a modal dialog box	65
7.5	Self-learn speech trainer	66
7.6	Active dialog box	67

List of Tables

2.1	Different types of symbols used for vowels	7
3.1	Total power in each frequency band and the percentage	13
3.2	Percentages of correct decisions	14
3.3	Approach to the best possible algorithm for /a:/ sound	15
3.4	Approach to the best possible algorithm for /æ/ sound	16
3.5	Approach to the best possible algorithm for /u:/ sound	17
3.6	Approach to the best possible algorithm for /o/ sound	18
3.7	Approach to the best possible algorithm for /eI/ sound	19
3.8	Extracted characteristics for /a:/ sound	20
3.9	Extracted characteristics for /æ/ sound	21
3.10	Extracted characteristics for /u:/ sound	22
3.11	Extracted characteristics for /o/ sound	23
3.12	Extracted characteristics for /eI/ sound	24
3.13	Percentages of correct decisions (Improved algorithms)	31
3.14	Results for new samples	32
4.1	Comparison of formant frequencies	40
4.2	Possible area for /a:/ sound with percentages of correct decisions	41
4.3	Possible area for /æ/ sound with percentages of correct decisions	41
4.4	Possible area for /u:/ sound with percentages of correct decisions	42
4.5	Possible area for /o/ sound with percentages of correct decisions	42
4.6	Possible area for /eI/ sound with percentages of correct decisions	43
4.7	Summary of results for percentages in formant method	46
5.1	Summary of test results for neural method	55
5.2-5.5	Test results for hearing impaired samples	57
5.6	Test results for combination of best results for each vowel	58
5.7	Summary of results for neural method	59
6.1	Comparison of accuracies of test results	60

CONTENTS

Abstract	I
List of figures	II
List of tables	III
Chapter 1	
Introduction	1
1.1 Research background	1
1.2 Overview of the work	2
1.2.1 Speech training	2
1.2.2 Methods used for speech signal processing	2
1.2.3 Basic operation	3
Chapter 2	
Speech Processing	5
2.1 Speech	5
2.1.1 Organs of speech	5
2.1.2 Speech production	5
2.1.3 Hearing and perception	5
2.1.4 Features of speech	6
2.1.5 Speech as symbols	6
2.2 Speech processing techniques	7
Chapter 3	
Spectral Analysis	10
3.1 Spectrogram analysis of speech signals	10
3.2 Spectrographic speech processing	10
3.3 Evaluation of speech characteristics extraction and improvements	11
3.4 Best possible algorithms and flow charts	19
3.5 Coding into Matlab and testing	32
3.6 Real time speech recording	32
Chapter 4	
Formant Estimation	34
4.1 Formant frequencies	34
4.2 Formant estimation	34
4.3 Vowel recognition using formants	37
4.4 Average formant values	40
4.5 Specific regions of vowels	40

Chapter 5	
Neural Network Analysis	47
5.1 Neural network approach for vowel recognition	47
5.2 Selection of a suitable neural network	47
5.3 Designing a multi-layered neural network for vowel recognition	47
5.4 Selection of sigmoidal as activation function	50
5.5 Training procedure of the network	53
5.6 Testing the neural network	54
Chapter 6	
Analysis of Results	60
6.1 Comparison of methods used	60
6.2 Comparison of results obtained	60
6.3 Method selected for the speech trainer	61
6.4 Possible improvements to formant estimation method	61
6.5 Possible improvements to neural network method	61
Chapter 7	
Visual Interface	62
7.1 Training methodology and visual indication of results	62
7.2 Visual interface design	63
7.3 Conversion of Matlab into Visual C++	64
7.4 Designing dialog boxes in Visual C++	65
7.5 Connecting files to the dialog box	65
7.6 Operation of the speech trainer	66
7.7 Viewing a video clip	67
7.8 Training a vowel sound	67
Chapter 8	
Conclusion	68
8.1 Problems encountered	68
8.2 Further improvements and future work	68
8.3 Summary	69
References	70
Appendix(A)	
(i) Matlab code to find power and percentages of power in each frequency band for normal speakers	72
(ii) Matlab code to find percentages of power in each frequency band for hearing impaired speakers	73
(iii) Matlab code for comparison of template values and speaker utterances according to selected algorithms and flow charts	73
(iv) Matlab code for real time speech recording and comparison	75

Appendix(B)		
Power variation according to the number of frequency bands		77
Appendix(C)		
Percentages of power for normal and hearing impaired speaker samples		82
Appendix(D)		
Test results of the algorithm and results according to a normal listener		111
Appendix(E)		
Graphical representation of target values and the speaker performance		121
Appendix(F)		
Percentages of power for new samples		131
Appendix(G)		
Matlab source code for formant analysis		134
Appendix(H)		
(i) Initial weights applied for the neural network		136
(ii) Weight values after training the network		137
Appendix(I)		
Matlab source code for training and testing neural network		140
Appendix(J)		
Matlab source code for visual indication of results		142
Appendix(K)		
Visual C++ source code for the speech trainer		146



Acknowledgement

The author is indebted to

Dr. Dileeka Dias

for all the valuable guidance, advice, encouragement, inspiration and most of all for proposing this title for the dissertation.

Prof. I. J. Dayawansa

Dr. Nishantha Nanayakkara

for the valuable comments and suggestions provided at the progress review sessions.

Prof. N. Ratnayaka

Director – Post graduate research studies.

Asian Development Bank
for financial support.



University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Dr. Gihan Dias and the members of the Research Group

Mr. Ruwan Gajaweera

Koojana, Janaka, Namunu and Sankassa

Mr. Jayantha Perera and Mr. Philip Terrence

D.D. Sumanapala and Thushara

My parents, wife, daughter and son.