ACOUSTIC TUBE SHAPE RECOVERY WITH SPECIFIC APPLICATION TO SPEECH ANALYSIS

GREGORY JOHN BIELBY, B.E. (HONS.), B.Sc.

Being a thesis submitted

for the

DEGREE OF DOCTOR OF PHILOSOPHY

in the

DEPARTMENT OF ELECTRICAL AND

ELECTRONIC ENGINEERING

THE UNIVERSITY OF ADELAIDE

AUGUST 1983

THIS THESIS IS DEDICATED TO MY PARENTS FOR THEIR UNDYING LOVE, ENCOURAGEMENT, PATIENCE AND SUPPORT.

TABLE OF CONTENTS

	page
ABSTRACT	vii
SUMMARY	viii
STATEMENT OF ORIGINALITY	x
ACKNOWLEDGEMENTS	xi
CHAPTER 1 INTRODUCTION	
1.1 BACKGROUND	1
1.2 THE PHYSICAL VOCAL TRACT AND SPEECH PRODUCTION	5
1.3 THE NEED FOR ACCURATE VOCAL TRACT SHAPE RECOVERY	8
1.4 THE NEED FOR FURTHER INVESTIGATION	12
1.5 THESIS ORGANIZATION	19
CHAPTER 2 ACOUSTIC TUBE MODEL AND LINEAR PREDICTION	
2.1 INTRODUCTION	24
2.2 THE ACOUSTIC TUBE MODEL	27
2.2.1 Defining the Acoustic Tubes	28
2.2.2 Waveform Description of a Single Acoustic Tube	30
2.2.3 Waveform Description at the Junction of Acoustic Tubes	35
2.2.4 Boundary Conditions	39
2.2.5 Complete Acoustic Tube Model	43
2.3 LINEAR PREDICTION	53
2.3.1 Basic Principles	55
2.3.2 Covariance Formulation	59
2.3.3 Autocorrelation Formulation	63
2.3.4 Lattice Formulation	68
2.3.4.1 Basic principles	69
2.3.4.2 Calculation of k_i	75
2.3.5 Excitation Requirements of Linear Prediction	80
2.4 RELATIONSHIP BETWEEN LINEAR PREDICTION AND THE ACOUSTIC TUBE MODEL	84
2.5 THE CONCEPT OF AREA DISTANCE	86
2.6 SUMMARY	90

CHAPTER 3 NON-WHITE EXCITATION OF ACOUSTIC TUBES

3.1 INTRODUCTION 92

3.2 EFFECTS OF NON-WHITE EXCITATION 95

3.2.1 The Linear Prediction Process 105

3.2.2 Understanding Some Effects of Non-White Excitation 114

3.3 SPECIAL ACOUSTIC TUBE SHAPES 124

3.4 RECOVERY OF SPECIAL ACOUSTIC TUBES 136

3.5 AREAS OF APPLICATION 148

3.6 CONCLUDING COMMENTS 159

CHAPTER 4 GLOTTAL PULSE EXCITATION OF ACOUSTIC TUBES

4.1 INTRODUCTION 161

4.2 EFFECTS OF GLOTTAL PULSE EXCITATION 167

4.3 EFFECTIVENESS OF CONVENTIONAL MEANS TO OVERCOME GLOTTAL PULSE EXCITATION 187

4.4 IMPROVED ADAPTIVE PRE-EMPHASIS 222

4.4.1 A New Adaptive Pre-emphasis Filter 223

4.4.2 Determination of Parameters 232

4.4.2.1 Parameter α 234

4.4.2.2 Parameter β 253

4.4.3 Investigation of Sampling Frequency Effects 260

4.4.3.1 Parameter α 263

4.4.3.2 Parameter β 271

4.4.4 The Two/Three Adaptive Pre-emphasis Filter 278

4.5 SUMMARY 283

CHAPTER 5 EVALUATION OF THE TWO/THREE ADAPTIVE PRE-EMPHASIS FILTER

5.1 INTRODUCTION 287

5.2 EVALUATION WITH GLOTTAL PULSES 290

5.2.1 Parameter α 292

5.2.2 Parameter β 301

5.2.3 Sampling Frequencies 307

5.2.3.1 Parameter α 308

5.2.3.2 Parameter β 315

CHAPTER 5 (cont.)

5.3 EVALUATION WITH SYNTHETIC SPEECH DATA 322
5.3.1 Parameter α 324
5.3.2 Parameter β 343
5.3.3 Sampling Frequencies 353
5.3.3.1 Parameter α 354
5.3.3.2 Parameter β 374
5.4 EVALUATION WITH REAL SPEECH 391
5.5 SUMMARY 403

CHAPTER 6 ACOUSTIC TUBES WITH A LOSSY TERMINATION

6.1 INTRODUCTION 410
6.2 TERMINATION MODEL DEFINITION 414
6.3 THE LOSSY TERMINATION AND CONVENTIONAL ANALYSIS PROCEDURES 436
6.4 REQUIREMENTS FOR IMPROVED ANALYSIS OF A LOSSY TERMINATION 451
6.5 LOSSY TERMINATION ANALYSIS VIA AUTOCORRELATIONS 456
6.5.1 Derivation of an Analysis Process 457
6.5.2 Calculation of Reflection Coefficients 465
6.5.2.1 Evaluations with correct lossy termination 476
6.5.2.2 Evaluations with incorrect lossy termination 489
6.5.3 Parcor Lossy Termination Analysis of Real Speech 499
6.6 LOSSY TERMINATION ANALYSIS VIA A TRANSFER FUNCTION OF THE ACOUSTIC TUBE MODEL 509
6.7 SUMMARY 521

CHAPTER 7 AN ANALYSIS PROCESS FOR SPEECH

7.1 INTRODUCTION 526
7.2 COMBINATION OF PROCEDURES FOR GLOTTAL PULSE EXCITATION AND A LOSSY TERMINATION 527
7.3 EVALUATION WITH SYNTHETIC SPEECH 531
7.3.1 Evaluations with Correct Lossy Termination 534
7.3.2 Evaluations with Incorrect Lossy Termination 548
7.4 EVALUATION WITH REAL SPEECH 561
7.5 SUMMARY 579

CHAPTER 8 CONCLUSIONS 584

APPENDIX A IMPULSE RESPONSE OF TWO ACOUSTIC TUBES 595

APPENDIX B GENERATION OF SYNTHETIC SPEECH DATA WITH NON-WHITE EXCITATION 599

APPENDIX C ACOUSTIC TUBE SHAPES AND REFLECTION COEFFICIENTS FOR FIVE VOWELS 602

APPENDIX D LOOKUP TABLES FOR TWO/THREE ADAPTIVE PRE-EMPHASIS FILTER 605

APPENDIX E GENERATION OF SYNTHETIC SPEECH DATA WITH A LOSSY TERMINATION 614

APPENDIX F EQUIVALENCE BETWEEN BACKWARD AND MINIMUM FORM OF μ_i 617

APPENDIX G NON-LINEAR SIMULTANEOUS EQUATIONS FOR TRANSFER FUNCTION ANALYSIS 622

APPENDIX H PROCEDURE AND CONDITIONS UNDER WHICH REAL SPEECH WAVEFORMS ARE DIGITALLY RECORDED 625

APPENDIX I GENERATION OF SYNTHETIC SPEECH DATA WITH A NON-WHITE EXCITATION AND A LOSSY TERMINATION 627

REFERENCES 630

ABSTRACT

A lossless acoustic tube model and linear predictive analyses are reviewed, which represent the conventional model and analysis procedure for vocal tract shape recovery from the speech waveform. The effects of a non-white excitation on acoustic tube shapes recovered by linear prediction are investigated, and result in the definition of special acoustic tube shapes which can be recovered for non-white excitations of certain durations.

Conventional pre-emphasis techniques for removing glottal pulse excitation effects from the acoustic tube shape recovered by linear prediction are evaluated and, in general, are shown to perform poorly. A new adaptive pre-emphasis filter is defined, and shown to produce improved acoustic tube/vocal tract shape recovery in comparison with conventional pre-emphases for both synthetic and real speech waveforms of five vowel sounds.

A lossy termination of the acoustic tube model is investigated, and two new analysis procedures are presented, one based on autocorrelation functions, and another based on a transfer function of the acoustic tube model. The lossy termination analysis based on autocorrelations is shown to be applicable to real time applications of speech analysis, and both analyses produce an improvement in acoustic tube/vocal tract shape recovery in comparison with conventional analyses, when a lossy termination is present.

The new adaptive pre-emphasis filter and the lossy termination analysis based on autocorrelations are combined into a single speech analysis procedure. This new speech analysis procedure is evaluated with synthetic and real speech for five vowel sounds, and is shown to produce improved acoustic tube/vocal tract shape recovery in comparison with existing techniques.

SUMMARY

Acoustic tube models are widely used as models of the human vocal tract, and permit the acoustic waveforms in the vocal tract to be specified from the geometry and properties of the vocal tract. However, the inverse problem of determining the geometry or properties of a set of acoustic tubes, which model the vocal tract, is complex, and not completely resolved today. This thesis investigates the inverse problem to improve the accuracy of recovering acoustic tube geometries from the radiated acoustic waveform.

A lossless acoustic tube model and linear predictive analyses are reviewed, and the relationship between the results of a linear predictive analysis and the lossless acoustic tube model is defined. The conditions that the acoustic tube model must satisfy for a predictive analysis of its output or radiated acoustic waveform to recover its shape are presented. Of these conditions, many are not satisfied in the human vocal tract or during the production of speech; therefore, a linear prediction of a speech waveform does not identify an acoustic tube model the shape of which is the same as the vocal tract. An area distance measure is defined to provide a quantitative measure of the similarity between an acoustic tube shape recovered by an analysis of the output waveform from a set of acoustic tubes and the shape of those acoustic tubes.

For linear prediction to recover the shape of a set of acoustic tubes, those acoustic tubes must be excited by a white excitation. The effects of a non-white excitation of acoustic tubes on the acoustic tube shape recovered by linear prediction are investigated, and show that poor acoustic tube shape recovery occurs. A class of acoustic tubes is determined which can be recovered after a linear predictive analysis for non-white excitations of restricted duration. The necessary post analysis to recover the acoustic tube shape is presented, and the areas of application are discussed.

A new adaptive pre-emphasis filter is presented, the form of which is defined from measurements of the required pre-emphasis to whiten a range of glottal pulse waveforms, which are representative of the glottal pulse waveforms used to excite the vocal tract during the production of voiced speech. Evaluation of the new pre-emphasis filter with real and synthetic speech waveforms shows that improved acoustic tube shape recovery is achieved in comparison with the acoustic tube shapes recovered by previously used pre-emphasis techniques. The evaluations consider a wide range of sampling frequencies, and show a consistent improvement in acoustic tube shape recovery.

Conventional acoustic tube shape recovery by linear prediction requires a lossless termination of the acoustic tubes, which is not the case for the vocal tract. The effects on the acoustic tube shape recovered by linear prediction for a lossy termination of acoustic tubes are presented, and are shown to be significant. A

general model of the loss at the termination of acoustic tubes due to radiation is reviewed, and simplified by the known conditions which exist at the lips, i.e. the termination of the vocal tract.

A number of autocorrelation analysis procedures for a lossy termination of acoustic tubes are defined, and evaluated with synthetic and real speech waveforms. An analysis procedure is also presented which is derived from a transfer function of the acoustic tube model and uses constraints, based on physical restrictions in the vocal tract, to overcome an ambiguity problem. The autocorrelation analyses are shown to be applicable to a wider range of situations than the transfer function analysis, especially for real time applications. These new analysis methods are shown to produce improved acoustic tube/vocal tract shape recovery in comparison with conventional analyses, when a lossy termination is present.

The new adaptive pre-emphasis filter and a lossy termination analysis are combined into a single speech analysis procedure. This new speech analysis procedure is evaluated with real and synthetic speech waveforms of five vowel sounds, and is shown to produce improved vocal tract/acoustic tube shape recovery in comparison with existing analysis techniques.

STATEMENT OF ORIGINALITY

This thesis contains no material which has been accepted for the award of any other degree or diploma in any University, and to the best of the author's knowledge and belief contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

G.J. BIELBY

ACKNOWLEDGEMENTS

Over the five years it has taken to perform the research presented in this thesis, I have been helped by many different people at an academic, technical and personal level, and I am in debt to them all.

I am greatly indebted to Professor R.E. BOGNER for his guidance and academic help in my first years of research, and his then allowing me to develop my research skills to a point of self-sufficiency. I have indeed been fortunate to have been supervised by a person of the highest calibre, academically and personally, who is able to open up new paths, from generalizing aquired scientific facts to disclosing truths which others have not perceived. I would like to thank Dr. B.R. DAVIS for his help and guidance during the year that Professor R.E. BOGNER was on study leave.

There have been many fellow post graduates who have helped me over the years, but none more than Mr. D.S. FENSOM. He has always been interested and willing to help with difficulties and, where possible, find a solution or a line of attach which has been fruitful. His honest opinions of many of my ideas and methods are also greatly appreciated, and I am especially indebted to DAVID FENSOM for supplying the real speech data. I acknowledge Dr. N.I. SMITH, Dr. T.K. KO, Dr. K. ESHRAGHIAN, Mr. P. BUDIMIR, Mr. M. LIEBELT, Mr. N. WOTTON, Mr. M. RAYMOND, Dr. R.J. CLARKE, Dr. A. ROY, Mr. R.C. BRYANT, Mr. J. NOONAN, Mr. B. COWLEY, Mr. N. MARTIN, and Mr. A. DAY for their moral support and their making life interesting and bearable as a post graduate student.

Special thanks must go to Mr. R.C. NASH, who managed to keep a troublesome computing system running long enough for me to finish my research work. Thanks are also directed to the staff and ancillary staff of the Department of Electrical and Electronic Engineering for all the help they have given.

I am also much indebted to ETTA CLARK for typing the thesis manuscript, and for using her expert typing skills to make this task relatively simple for me. The help of ELIZABETH HENDERSON with the production of some of the figures and tables is also acknowledged.

Finally, but most importantly, the research work presented in this thesis would never have been started or completed without the love, encouragement, support and patience of my parents, brothers and their families. Although they did not understand what I was doing and why, they have provided the greatest contribution to this thesis.