1

**PM-1 3 1/2" x 4" PHOTOGRAPHIC MICROCOPY TARGET
NBS 1010a ANSI/ISO #2 EQUIVALENT**

PRECISIONSM RESOLUTION TARGETS

PIONEERS IN METHYLENE BLUE TESTING SINCE 1974

MicroD

1880 COUNTRY ROAD S. BURNINGWELL NEW MEXICO 88057 USA
TEL 512 428 7887 FAX 512 428 7887 TLA 5100000000

National Library
of Canada

Acquisitions and
Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale
du Canada

Direction des acquisitions et
des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file Votre référence

Our file Notre référence

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

ACOUSTIC ECHO CANCELLATION: PERFORMANCE AND STRUCTURES

submitted by

MICHAEL E. KNAPPE, B.Sc (Eng)

A thesis submitted to
the Faculty of Graduate Studies and Research
in partial fulfillment of
the requirements for the degree of

Master's of Engineering

**Ottawa-Carleton Institute for Electrical Engineering
Faculty of Engineering
Department of Systems and Computer Engineering**

**Carleton University
Ottawa, Ontario, Canada, K1S 5B6**

July 25, 1992

© copyright

1992 Michael E. Knappe

National Library
of Canada

Acquisitions and
Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale
du Canada

Direction des acquisitions et
des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Veuillez lire votre référence

Veuillez lire votre référence

The author has granted an irrevocable non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of his/her thesis by any means and in any form or format, making this thesis available to interested persons.

L'auteur a accordé une licence irrévocable et non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de sa thèse de quelque manière et sous quelque forme que ce soit pour mettre des exemplaires de cette thèse à la disposition des personnes intéressées.

The author retains ownership of the copyright in his/her thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without his/her permission.

L'auteur conserve la propriété du droit d'auteur qui protège sa thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

ISBN 0-315-79813-0

Canada

The undersigned hereby recommend to
the Faculty of Graduate Studies and Research
acceptance of the thesis

**ACOUSTIC ECHO CANCELLATION:
PERFORMANCE AND STRUCTURES**

submitted by

MICHAEL E. KNAPPE, B.Sc (Eng.)

in partial fulfillment of the requirements
for the degree of Master's of Engineering

Thesis Supervisor

Chair, Department of Systems and Computer Engineering

Carleton University

Sept. 25, 1992

ABSTRACT

Handsfree desktop terminals, with the loudspeaker and microphone closely coupled in the same enclosure, typically allow only half-duplex conversation in order to maintain stability. This usually results in less natural conversations as one party waits for the other to finish speaking. The use of acoustic echo cancellation between the coupled loudspeaker and microphone has previously been proposed as a means of maintaining stability while allowing full-duplex conversation.

This thesis looks at the physical environment in which an acoustic echo canceller would operate and the limitations this environment places on achievable cancellation. It is shown that for short acoustic echo cancellers, the impulse response decay rate of the terminal and room imposes a limit on achievable cancellation, while for longer cancellers, either room noise or system non-linearities will limit cancellation. Non-stationarities in the room are shown to significantly affect cancellation during the period of movement.

This thesis proposes structures to improve achievable cancellation for two of the most significant problems found: system non-linearities and room non-stationarities. A non-linear pre-processor structure is proposed that improves cancellation performance to the theoretical limit for a known non-linearity. An adaptive step-size algorithm based on LMS is shown to provide both good tracking during non-stationary periods and a good level of cancellation during stationary periods. This algorithm is shown to provide a 2.5 to 5 dB improvement in Echo Return Loss Enhancement (ERLE) over LMS, NLMS and SFTF under varying acoustic conditions with an insignificant increase in computational costs.

ACKNOWLEDGEMENTS

I would like to thank Dr. R. Goubran for his guidance and motivation during the course of my thesis. I would like to thank Dr. B. Shelton of the Audio Research Group at BNR for his interest in my thesis and for giving me unlimited access to the group's facilities. I would also like to thank Dr. A. Van Shyndel of the Acoustics Group at BNR for taking time to discuss various aspects of my thesis, particularly loudspeaker non-linearities (and Parseval's theorem!). Finally, I would like to thank Ray Perkins for help in designing and constructing the modified handsfree terminals and the adjustable gain box.

I would also like to thank BNR, NSERC, and Carleton for their financial support during this thesis.

TABLE OF CONTENTS

ABSTRACT.....	iii
ACKNOWLEDGEMENTS.....	iv
TABLE OF CONTENTS.....	v
LIST OF FIGURES	vii
LIST OF TABLES	xi
LIST OF APPENDICES	xii
LIST OF SYMBOLS	xiii
1.0 INTRODUCTION	1
2.0 REVIEW OF ACOUSTIC ECHO CANCELLATION TECHNIQUES.....	5
2.1 Brief Description of Problem.....	5
2.2 Adaptive Filtering Algorithms	7
2.2.1 Least-Mean Squares (LMS)	9
2.2.2 Normalized Least-Mean Squares	10
2.2.3 Recursive Least Squares (RLS).....	12
2.2.4 Fast Transversal Filters (FTF)	13
2.2.5 Acoustic Echo Canceller Structure.....	18
2.2.6 Acoustic Echo Cancellation Performance Issues using LMS	19
2.3 Terminal/Room Acoustic Environment	21
2.4 General Advances in Adaptive Filtering Algorithms.....	22
2.4.1 Fast Transversal Filters.....	22
2.4.2 Adaptive IIR Filter Algorithms	23
2.5 Algorithm Enhancements for Acoustic Echo Cancellation	26
2.5.1 Algorithms for Efficiently Covering Long Impulse Responses	26
2.5.2 Algorithms for Improved Convergence with Speech Input.....	28
2.5.3 Algorithms with Dynamic Adaptation Step-Size	34
2.6 Methods for Improving Loudspeaker Linearity.....	37
2.7 Acoustic Echo Cancellation Limitations	39
2.8 Literature Review Summary	40
3.0 EXPERIMENTAL SETUP FOR AEC MEASUREMENTS	42
3.1 Impulse Response Measurement Setup	43
3.2 Terminal/Room Impulse Responses	45
3.2.1 Choice of Adaptation Step-Size.....	52
4.0 ACOUSTIC ECHO CANCELLATION PERFORMANCE LIMITATIONS.....	56
5.0 ERLE LIMITATIONS DUE TO ROOM NOISE AND ADAPTIVE FILTER LENGTH.....	65
5.1 ERLE Limitations Due to Room Noise	65
5.2 ERLE Limitations Due to Adaptive Filter Length.....	67
5.3 ERLE vs. TIP/TP (Artificial Data).....	77

5.4	ERLE vs. TIP/TP (Real Data).....	78
6.0	ERLE LIMITATIONS DUE TO SYSTEM NON-LINEARITIES.....	82
6.1	ERLE vs. Loudspeaker Level Changes (Real Data).....	82
6.2	ERLE vs. Reference Level Changes for Artificially Clipped Data.....	83
6.3	Non-Linear Compensation.....	95
6.4	Significance of Loudspeaker Non-linearities.....	99
7.0	DYNAMIC ACOUSTIC ECHO CANCELLATION PERFORMANCE.....	101
7.1	Algorithm Performance (Stationary IR, Noise Input).....	102
7.1.1	Stationary, Artificial Data, Noise Input.....	102
7.1.2	Stationary, Real Data, Noise Input.....	113
7.1.3	Stationary, Real Data, Speech Input.....	126
7.2	Convergence Properties Using Pre-whitening Structures.....	131
7.3	Tracking Improvements for LMS.....	131
7.4	Tracking Improvements for Speech Input Conditions.....	140
7.5	Discussion of Convergence and Tracking Results.....	140
8.0	CONCLUSIONS.....	142
8.1	Discussion of Results.....	142
8.2	Future Research.....	144
8.2.1	Incorporation of Adaptive Non-Linear Filter in Pre-Processor Structure.....	144
8.2.2	Applicability of Cancellation Limits with Sub-band Structures.....	144
8.2.3	Double-Talk Detection.....	145
8.2.4	Variable Gain NLMS.....	145
8.3	Summary of Contributions.....	146
9.0	REFERENCES.....	147
10.0	APPENDICES.....	152

LIST OF FIGURES

FIGURE 1.	Switching (Half-Duplex) Handsfree System	6
FIGURE 2.	Acoustic Echo Cancellation in Handsfree Environment	7
FIGURE 3.	Generic FIR filter with computation of an estimation error	8
FIGURE 4.	Acoustic Echo Canceller Structure	18
FIGURE 5.	Acoustic Echo Canceller using Common Room Poles.....	25
FIGURE 6.	Modelling of Room Impulse Response As Short-Long Convolution.....	27
FIGURE 7.	Tap Selection/Replacement Mechanism	28
FIGURE 8.	Hansler's Decorrelation Structure.....	30
FIGURE 9.	Mboup, Bonnet and Bershad's Decorrelation Structure (Classical).....	31
FIGURE 10.	Mboup, Bonnet and Bershad's Decorrelation Structure (Simplified).....	32
FIGURE 11.	Sub-band Acoustic Echo Canceller Structure	34
FIGURE 12.	Non-Linear Preprocessor For Linearization of Loudspeaker	38
FIGURE 13.	Adaptive Non-Linear Processing Scheme	39
FIGURE 14.	Simplified Coupling Response.....	43
FIGURE 15.	Gain Box Schematic Diagram	45
FIGURE 16.	Terminal / Room Impulse Response Measurement Setup	46
FIGURE 17.	Wideband prototype set / DPN-A Conference room impulse response.....	50
FIGURE 18.	Wideband prototype set / anechoic chamber impulse response.....	50
FIGURE 19.	Low-cost Set / DPN-A Conference room impulse response.....	51
FIGURE 20.	Low-cost set / anechoic chamber impulse response	51
FIGURE 21.	ERLE vs. time for 250 tap LMS with $\mu=0.001$ (Low-cost set in DPN-A) 54	
FIGURE 22.	ERLE vs. time for 250 tap LMS with $\mu=0.004$ (Low-cost set in DPN-A) 54	
FIGURE 23.	ERLE vs. time for 250 tap LMS with $\mu=0.016$ (Low-cost set in DPN-A) 55	
FIGURE 24.	ERLE vs. time for 250 tap LMS with $\mu=0.064$ (Low-cost set in DPN-A) 55	
FIGURE 25.	Artificial Primary File Creation Process	56
FIGURE 26.	ERLE vs. Number of LMS Tap Weights for Artificial Data, $(S+N)/N = 40$ dB	59
FIGURE 27.	ERLE vs. Number of LMS Tap Weights for DPN-A Conf. Room	60
FIGURE 28.	ERLE vs. Number of LMS Tap Weights for Anechoic Chamber	61
FIGURE 29.	Generalized Obtainable ERLE Curve.....	63
FIGURE 30.	TIP/TP Measurement on Impulse Response.....	68
FIGURE 31.	TIP/TP (dB) for 1200 tap Artificial Impulse Response LWBNI200.....	70
FIGURE 32.	TIP/TP (dB) for Wideband (solid), Low-cost Set (dashed)	70
FIGURE 33.	Terminal coupling decay rate as a performance requirement	77
FIGURE 34.	Comparison of ERLE and TIP/TP Curves for Artificial Data.....	79

FIGURE 35.	Comparison of ERLE and TIP/TP Curves for Wideband, Anechoic	80
FIGURE 36.	Comparison of ERLE and TIP/TP Curves for Low-cost set, Anechoic	81
FIGURE 37.	Input-Output Relation for Soft-Clipper	85
FIGURE 38.	Artificial Primary File Creation Process for Clipped Data	85
FIGURE 39.	Original Fixed Impulse Response	86
FIGURE 40.	Avg Tap Weights for 50 tap LMS, $\mu=0.04$, root clipped, 0 dB ref attenuation.....	88
FIGURE 41.	Avg Tap Wgt Error for 50 tap LMS, $\mu=0.04$, root clipped, 0 dB ref attenuation.....	88
FIGURE 42.	Avg Tap Weights for 50 tap LMS, $\mu=0.04$, root clipped, 1 dB ref attenuation.....	89
FIGURE 43.	Avg Tap Wgt Error for 50 tap LMS, $\mu=0.04$, root clipped, 1 dB ref attenuation.....	89
FIGURE 44.	Avg Tap Weights for 50 tap LMS, $\mu=0.04$, root clipped, 2 dB ref attenuation.....	90
FIGURE 45.	Avg Tap Wgt Error for 50 tap LMS, $\mu=0.04$, root clipped, 2 dB ref attenuation.....	91
FIGURE 46.	Avg Tap Weights for 50 tap LMS, $\mu=0.04$, root clipped, 3 dB ref attenuation.....	91
FIGURE 47.	Avg Tap Wgt Error for 50 tap LMS, $\mu=0.04$, root clipped, 3 dB ref attenuation.....	92
FIGURE 48.	Avg Tap Weights for 50 tap LMS, $\mu=0.04$, root clipped, 4 dB ref attenuation.....	92
FIGURE 49.	Avg Tap Wgt Error for 50 tap LMS, $\mu=0.04$, root clipped, 4 dB ref attenuation.....	93
FIGURE 50.	Avg Tap Weights for 50 tap LMS, $\mu=0.04$, root clipped, 5 dB ref attenuation.....	93
FIGURE 51.	Avg Tap Wgt Error for 50 tap LMS, $\mu=0.04$, root clipped, 5 dB ref attenuation.....	94
FIGURE 52.	Use of Non-Linear Model in Acoustic Echo Canceller Structure.....	96
FIGURE 53.	Alternative Structure for use of Non-Linear Loudspeaker Model.....	99
FIGURE 54.	Averaged Tap Weights for 50 tap LMS, $\mu=0.04$, 10 second Convergence	105
FIGURE 55.	Averaged Tap Weight Error for 50 tap LMS, $\mu=0.04$, 10 second Convergence	105
FIGURE 56.	ERLE for 50 tap LMS, $\mu=0.04$	106
FIGURE 57.	Averaged Tap Weights for 50 tap NLMS, $\mu=1/\text{power}$, 10 second Convergence	106
FIGURE 58.	Averaged Tap Weight Error for 50 tap NLMS, $\mu=1/\text{power}$, 10 second Convergence	107

FIGURE 59.	ERLE for 50 tap NLMS, $\mu=1/\text{power}$	107
FIGURE 60.	Averaged Tap Weights for 50 tap RLS, $l=1$, 10 second Convergence	108
FIGURE 61.	Averaged Tap Weight Error for 50 tap RLS, $l=1$, 10 second Convergence	108
FIGURE 62.	ERLE for 50 tap RLS, $l=1$	109
FIGURE 63.	Averaged Tap Weights for 50 tap FTF, $l=1$, $\mu=1$, 10 second Convergence	109
FIGURE 64.	Averaged Tap Weight Error for 50 tap FTF, $l=1$, $\mu=1$, 10 second Convergence	110
FIGURE 65.	ERLE for 50 tap FTF, $l=1$, $\mu=1$	111
FIGURE 66.	Averaged Tap Weights for 50 tap SFTF, $l=1$, $\mu=1$, 10 second Convergence	111
FIGURE 67.	Averaged Tap Weight Error for 50 tap SFTF, $l=1$, $\mu=1$, 10 second Convergence	112
FIGURE 68.	ERLE for 50 tap SFTF, $l=1$, $\mu=1$	112
FIGURE 69.	Averaged Tap Weights for 50 tap LMS, $\mu=0.004$, 10 second Convergence	116
FIGURE 70.	Averaged Tap Weights for 50 tap LMS, $\mu=0.04$, 10 second Convergence	116
FIGURE 71.	Averaged Tap Weight Error (rel. to RLS) for 50 tap LMS, $\mu=0.004$	117
FIGURE 72.	Averaged Tap Weight Error (rel. to RLS) for 50 tap LMS, $\mu=0.04$	117
FIGURE 73.	ERLE for 50 tap LMS, $\mu=0.004$	118
FIGURE 74.	ERLE for 50 tap LMS, $\mu=0.04$	118
FIGURE 75.	Averaged Tap Weights for 50 tap NLMS, $\mu=1/\text{power}$, 10 second Convergence	119
FIGURE 76.	Averaged Tap Weight Error (rel. to RLS) for 50 tap NLMS, $\mu=1/\text{power}$	119
FIGURE 77.	ERLE for 50 tap NLMS, $\mu=1/\text{power}$	120
FIGURE 78.	Averaged Tap Weights for 50 tap RLS, $l=0.9995$, 10 second Convergence	120
FIGURE 79.	ERLE for 50 tap RLS, $l=0.9995$	121
FIGURE 80.	Averaged Tap Weights for 50 tap FTF, $l=0.9995$, $\mu=1$, 10 second Convergence	122
FIGURE 81.	Averaged Tap Weight Error (rel. to RLS) for 50 tap FTF, $l=0.9995$, $\mu=1$	122
FIGURE 82.	ERLE for 50 tap FTF, $l=0.9995$, $\mu=1$	123
FIGURE 83.	Averaged Tap Weights for 50 tap SFTF, $l=0.9995$, $\mu=1$, 10 second Convergence	123
FIGURE 84.	Averaged Tap Weight Error (rel. to RLS) for 50 tap SFTF, $l=0.9995$, $\mu=1$	124
FIGURE 85.	ERLE for 50 tap SFTF, $l=0.9995$, $\mu=1$	124

FIGURE 86.	Averaged Tap Weights for 50 tap SFTFAP, $l=0.9995$, $\mu=1$, 10 second Convergence	125
FIGURE 87.	Averaged Tap Weight Error (rel. to RLS) for 50 tap SFTFAP, $l=0.9995$, $\mu=1$	125
FIGURE 88.	ERLE for 50 tap SFTFAP, $l=0.9995$, $\mu=1$	126
FIGURE 89.	ERLE for 50 tap LMS ($\mu=0.04$), Male Speech input.....	128
FIGURE 90.	ERLE for 50 tap LMS ($\mu=0.004$), Male Speech input.....	128
FIGURE 91.	ERLE for 50 tap NLMS ($\mu=1/\text{power}$), Male Speech input.....	129
FIGURE 92.	ERLE for 50 tap NLMS ($\mu=0.5/\text{power}$), Male Speech input.....	129
FIGURE 93.	ERLE for 50 tap SFTF ($l=0.9995$, $\mu=1$), Male Speech input.....	130
FIGURE 94.	ERLE for 50 tap SFTF ($l=0.9995$, $\mu=10$), Male Speech input.....	130
FIGURE 95.	ERLE for 600 tap SFTF, $l=0.9995$, $\mu=1$, Non-stationary IR conditions..	135
FIGURE 96.	ERLE for 600 tap SFTFAP, $l=0.9995$, $\mu=1$, Non-stationary IR conditions ...	136
FIGURE 97.	ERLE for 600 tap LMS, $\mu=0.004$, Non-stationary IR conditions.....	136
FIGURE 98.	ERLE for 600 tap LMS, $\mu=0.01$, Non-stationary IR conditions.....	137
FIGURE 99.	ERLE for 600 tap LMS, $\mu=0.04$, Non-stationary IR conditions.....	137
FIGURE 100.	ERLE for 600 tap NLMS, $\mu=1/\text{power}$, Non-stationary IR conditions	138
FIGURE 101.	ERLE for 600 tap LMS-VSS, $a=0.97$, $g=0.48$, $\mu_{\max}=0.04$, $\mu_{\min}=0.004$	138
FIGURE 102.	Step-size vs. time for 600-tap NLMS under Non-stationary conditions .	139
FIGURE 103.	Step-size vs. time for 600 tap LMS-VSS under Non-stationary conditions ..	139

LIST OF TABLES

TABLE 1.	Files Used For Reference Impulse Response Measurements	47
TABLE 2.	Effects of Varying μ (250 tap LMS - Low-cost NB/ DPN-A conf. room data).....	53
TABLE 3.	Converged ERLE vs. Number of Taps, Artificial Data	57
TABLE 4.	Converged ERLE vs. Number of LMS Taps, Real Data.....	58
TABLE 5.	Comparison of Primary (S+N)/N Ratio to ERLE Asymptote.....	78
TABLE 6.	Final Converged ERLE vs. Loudspeaker Acoustic Output Levels.....	83
TABLE 7.	ERLE and Tap Error Vector Norm vs. Level Changes for Artificially Clipped Data	94
TABLE 8.	Converged ERLE Values for Different Clipping Conditions.....	97
TABLE 9.	Comparison of Tap Weight Error Vector Norm and Final ERLE, Artificial Data	103
TABLE 10.	Algorithm Convergence Speed, Artificial Data	103
TABLE 11.	Comparison of Tap Weight Error Vector Norm and Final ERLE, Real Data	114
TABLE 12.	Algorithm Convergence Speed, Real Data	114
TABLE 13.	Full List of .TIM Files From DAT Recordings.....	152
TABLE 14.	Hypersignal File Header Format.....	179

LIST OF APPENDICES

Appendix A - DAT Recordings	152
Appendix B - Fast Transversal Filter Algorithm Derivation	154
B.1 Adaptive Forward Linear Prediction	154
B.2 Adaptive Backward Linear Prediction	158
B.3 Extended Gain Vector Derivation	161
B.4 Conversion Factor	163
B.5 Final FTF Algorithm Derivation	170
Appendix C - Hypersignal File Format	179

LIST OF SYMBOLS

General Symbols and Abbreviations:

$\mathbf{u}_M(n)$: input vector with current input sample and previous $M-1$ input samples

$u(n)$: current input sample (scalar)

$d(n)$: current sample of desired response (scalar)

$e(n)$: current estimation error output (scalar)

$\hat{\mathbf{w}}_M(n)$: current estimated tap-weight vector of length M

μ : adaptation step-size (LMS-type algorithms)

μ : soft-constraint initialization weight (FTF and SFTF)

λ : forgetting factor (RLS, FTF and SFTF)

ERLE: Echo Return Loss Enhancement

Superscripted Symbols:

H : Hermitian transpose of vector or matrix

T : simple transpose of vector or matrix

Subscripted Symbols:

M : filter order (generally denotes length of vector as well)

1.0 INTRODUCTION

Handsfree telephony has become an important feature of desktop telephone terminals, allowing the user to converse without the need of a handset. This provides two primary benefits:

1. Allows the user while conversing to move about freely in his/her environment.
2. Allows a group of users to converse over one terminal

Handsfree terminals, however, require much larger electrical gains to maintain usable speech loudness levels to and from the public network. The additional gain is used to compensate for the loss associated with the increase in distance of the user from the speaker and microphone (as opposed to a handset user). This additional gain, coupled with the fact that most handsfree terminals place the microphone and loudspeaker together in the same acoustic enclosure, would generally cause electro-acoustic instability (feedback), and even if stability was maintained from outside the set, significant talker echo would be a problem. Talker echo, which is the return of a talker's delayed speech to himself, would become a problem for the far-end user connected to the handsfree user across the network, as the received signal at the handsfree would be directly coupled back to the far-end talker via the microphone, with little loss and possibly gain.

Typical conversations take place with only one person talking at a time, with only small intervals of overlapped speech (double-talk). Handsfree designers in the past have used this characteristic of conversations to solve the talker echo and stability problems by switching a significant amount of attenuation into the voice path (transmit or receive) that is currently not carrying speech. Although this solves the aforementioned problems, it cre-

ates new ones. First, the algorithm that controls the switching decisions will have difficulties with double-talk conditions, resulting in possible clipped or chopped conversation. This leads both the handsfree user and the far end talker to adopt a more cautious, restrained approach to conversation, generally known as talker etiquette, to ensure complete transmission of the conversation. This is somewhat akin to a voice actuated "push to talk" intercom system, and is known as half-duplex handsfree (only one voice path available at a time). The need for talker etiquette causes a certain amount of unnaturalness in half-duplex conversations that is not present in handset (full-duplex) conversations.

Echo cancellation techniques have been used for over a decade for hybrid echo mitigation on toll calls. It had been hoped that similar cancellation techniques could be employed in handsfree terminals to mathematically cancel the coupling between loudspeaker and microphone, avoiding the need for switched loss and allowing full-duplex conversations. Acoustic echo cancellation has turned out to be far more involved and complex a problem than electrical hybrid echo cancellation for several reasons:

1. Hybrid echo characteristics are typically stationary on a per-call basis, whereas the acoustic echo path is affected by any movement within its acoustic surroundings.
2. The length of cancellation required for hybrid cancellation is generally an order of magnitude less than required for acoustic echo cancellation.
3. The hybrid echo path is well modelled by a linear system (resistance, inductance and capacitance), while the acoustic echo path has at least one significant non-linear component (the loudspeaker).

This thesis takes a broad look at the complexities of acoustic echo cancellation. Performance issues such as the adaptive filter algorithm used, length of filter, impact of non-stationarities and impact of non-linearities are discussed. A digital recording system was set up to obtain high-quality two-channel recordings through actual handsfree terminal enclosures and transducers, from which various algorithms could be applied to this data in non-real time. This data was essential in examining real application issues generally not dealt with at a more theoretical level.

This thesis is organized as follows. The second chapter provides an introduction to some of the basic adaptive filtering algorithms employed in this thesis, as well as a literature review on previous work of significance in acoustic echo cancellation and related fields. The third chapter provides a discussion of the methods used in obtaining experimental data with further discussion on the acoustic echo cancellation problem. The fourth chapter presents a discussion of basic limitations on acoustic echo cancellation performance. The fifth chapter looks at the limitations on the obtainable Echo Return Loss Enhancement (ERLE - mathematically defined in Section 3.2.1) due to room noise and adaptive filter length. It provides derivations for the expected ERLE performance of a system along with simulations to support these derivations. The sixth chapter examines the effects of system non-linearities, and includes simulations on artificial data to clarify ERLE results from real data taken from non-linear loudspeaker conditions. A possible adaptive filtering structure is given that could be used to compensate for loudspeaker non-linearities. The seventh chapter looks at dynamic canceller performance and examines the effects of room non-stationarities on ERLE performance. An in depth look at the performance of various adaptive filtering algorithms is given, with possible alternative adaptive filtering structures to max-

imize convergence speed and ERLE performance under non-stationary room conditions.

The final chapter presents conclusions on the findings of this thesis, with discussion of areas requiring further study.

Appendices included in this thesis detail the recordings and files used for simulations, a derivation of the Fast Transversal Filter, as well as a description of the Hypersignal file header and format used for the data files.

2.0 REVIEW OF ACOUSTIC ECHO CANCELLATION TECHNIQUES

This section provides background and discusses previous work on acoustic echo cancellation (and related fields) including:

1. a brief description of the acoustic echo cancellation problem.
2. a concise summary of the principal algorithms used in this thesis for acoustic echo cancellation.
3. investigations of the acoustic environment in which the acoustic echo canceller is to perform.
4. particular adaptive filtering algorithms and structures applicable to acoustic echo cancellation.
5. methods for improving the linearity of loudspeakers.
6. environmental limitations to acoustic echo cancellation

2.1 Brief Description of Problem

Figure 1 shows a simplified block diagram of a handsfree system. Typically, the microphone and loudspeaker are placed within the same acoustic enclosure in order to market the handsfree as a single desktop unit. Closely coupling the microphone and loudspeaker does create several problems. First, the large fixed gains required to maintain usable receive and transmit levels at a distance from the set can cause electro-acoustic instability. Even if stability is maintained, the large amount of gain plus the close loudspeaker-microphone coupling provide a large echo path back to the talker conversing with the handsfree

terminal. This echo, along with 40 ms or more of network time delay can cause serious conversational impairment [4]. Traditionally, these problems have been overcome by assuming that two talkers will not converse at the same time and inserting loss in the unused voice path. This method still requires some sort of switching decision mechanism to find who is the more deserving talker, and requires a finite amount of switching time. This switching can cause some impairment of its own, most noticeably clipping and chopping of words or sentences. The fact that only one voice path is available at a time defines this type of system as *half-duplex*.

FIGURE 1. Switching (Half-Duplex) Handsfree System

This thesis looks at the problems and limitations in creating a *full-duplex* handsfree system, one in which both voice paths are open for conversation at all times (as in standard handset to handset conversations). It has been a goal of handsfree designers, even before digital signal processing hardware became economically viable, to employ an adaptive fil-

ter to mathematically remove the loudspeaker-coupled component from the microphone signal. Figure 2 illustrates the positioning of the adaptive filter for acoustic echo cancellation.

FIGURE 2. Acoustic Echo Cancellation in Handsfree Environment

2.2 Adaptive Filtering Algorithms

The adaptive filters discussed in this thesis have the following characteristics in common [49]:

1. the filter output provides an estimate of a desired response, with the estimation error 'as small as possible'.
2. the filter is linear, to make mathematical analysis easy (a linear filter is one which has an output computed solely from a linear combination of current and previous inputs and previous outputs).

- the filter operates in discrete time steps (iterations), for implementation using digital hardware.

A simple diagram of this filter structure (in this case an FIR with M weights), without any adaptation scheme, is given in Figure 3 (from [49]):

FIGURE 3. Generic FIR filter with computation of an estimation error

Figure 3 does not show how to obtain the optimum weights based on the given signals (the adaptation scheme). Typically, adaptive systems update the weights based on some statistical criterion of the estimation error. The most common criterion is the mean-squared estimation error.

The four primary algorithms employed in this thesis are listed below in concise form. The algorithms are listed in the general complex form (although all simulations performed for this thesis used only real data). Derivations for LMS, NLMS and RLS are relatively sim-

ple and have been well documented (see [49]). The derivation for the FTF algorithm, however, is considerably more complex. An expanded version of Haykin's FTF derivation [49] can be found in Appendix B.

2.2.1 Least-Mean Squares (LMS)

The Least-Mean Squares (LMS) algorithm has become a popular algorithm for adaptive filtering due to its simplicity, numerical robustness and relatively low computational requirements ($2M+1$ multiplications and $2M$ additions/subtractions, where M is the order of the filter [49]).

Symbol Key:

M - filter order

Inputs:

$\mathbf{u}(n)$ - M by 1 input vector

μ - adaptation step-size (must be between 0 and $2/(\text{total input power})$), where the total input power is equivalent to $M r(0)$, and where $r(0)$ is the variance of the input signal. This implies the μ must be chosen to satisfy this criterion for the largest expected input signal variance.

$d(n)$ - desired signal

Outputs:

$\hat{\mathbf{w}}(n)$ - M by 1 tap weight vector

$y(n)$ - filter output

$e(n)$ - estimation error (a priori)

Algorithm:

$$y(n) = \hat{\mathbf{w}}^H(n) \mathbf{u}(n) \quad (1)$$

$$e(n) = d(n) - y(n) \quad (2)$$

$$\hat{\mathbf{w}}(n+1) = \hat{\mathbf{w}}(n) + \mu \mathbf{u}(n) e^*(n) \quad (3)$$

Initialization:

$$\hat{\mathbf{w}}(n) = \mathbf{0} \quad (4)$$

2.2.2 Normalized Least-Mean Squares

NLMS is a variation of the LMS algorithm which changes the effective adaptation step-size depending on the input power. The computational requirements of this algorithm are identical to LMS except the need for a divide plus an additional multiply and two addition/subtractions every iteration.

Symbol Key:

M - filter order

Inputs:

$\mathbf{u}(n)$ - M by 1 input vector

$\tilde{\mu}$ - adaptation step-size (must be between 0 and 2)

$d(n)$ - desired signal

Outputs:

$\hat{\mathbf{w}}(n)$ - M by 1 tap weight vector

$y(n)$ - filter output

$e(n)$ - estimation error (a priori)

Algorithm:

$$y(n) = \hat{\mathbf{w}}^H(n) \mathbf{u}(n) \quad (5)$$

$$e(n) = d(n) - y(n) \quad (6)$$

$$\|\mathbf{u}(n)\|^2 = \sum_{k=0}^{M-1} |u(n-k)|^2 \quad (7)$$

$$\hat{\mathbf{w}}(n+1) = \hat{\mathbf{w}}(n) + \frac{\tilde{\mu}}{\|\mathbf{u}(n)\|^2} \mathbf{u}(n) e^*(n) \quad (8)$$

Initialization:

$$\hat{\mathbf{w}}(n) = \mathbf{0} \quad (9)$$

2.2.3 Recursive Least Squares (RLS)

The Recursive Least Squares (RLS) algorithm is a more elaborate algorithm that takes advantage of a knowledge of the input signal statistics that extends back to the initialization of the algorithm (LMS, on the other hand, relies on instantaneous estimates of the input signal statistics). This typically gives RLS an order of magnitude improvement in convergence speed (the time taken to reach the point of minimum error variance) over LMS [49]. The RLS algorithm, however, is a much more computationally demanding algorithm (an order $O(M^2)$ algorithm). It also suffers from numerical precision problems [49] (the standard algorithm is listed here, but an equivalent reformulation of the algorithm with improved numerical stability was implemented for simulations in this thesis - RLS version II from [49]).

Symbol Key:

M - filter order

Inputs:

$\mathbf{u}(n)$ - M by 1 input vector

λ - forgetting factor (set between 0 and 1)

$d(n)$ - desired signal

Intermediate States:

$\mathbf{P}(n)$ - M by M inverse correlation matrix

$\mathbf{k}(n)$ - M by 1 gain vector

Outputs:

$\hat{\mathbf{w}}(n)$ - M by 1 tap weight vector

$\alpha(n)$ - estimation error (a priori)

$$\mathbf{k}(n) = \frac{\lambda^{-1} \mathbf{P}(n-1) \mathbf{u}(n)}{1 + \lambda^{-1} \mathbf{u}^H(n) \mathbf{P}(n-1) \mathbf{u}(n)} \quad (10)$$

$$\alpha(n) = d(n) - \hat{\mathbf{w}}^H(n-1) \mathbf{u}(n) \quad (11)$$

$$\hat{\mathbf{w}}(n) = \hat{\mathbf{w}}(n-1) + \mathbf{k}(n) \alpha^*(n) \quad (12)$$

$$\mathbf{P}(n) = \lambda^{-1} \mathbf{P}(n-1) - \lambda^{-1} \mathbf{k}(n) \mathbf{u}^H(n) \mathbf{P}(n-1) \quad (13)$$

Initialization:

$$\mathbf{P}(0) = \delta^{-1} \mathbf{I}, \quad \delta = \text{small positive constant}$$

$$\hat{\mathbf{w}}(n) = \mathbf{0} \quad (14)$$

2.2.4 Fast Transversal Filters (FTF)

The Fast Transversal Filter algorithm is an exact implementation of the RLS algorithm that takes advantage of redundancies within the RLS algorithm to bring the computational

complexity from order $O(M^2)$ to $O(M)$ ($7M+12$ multiplications, 4 divisions and $6M+3$ additions/subtractions [49]). The FTF algorithm, however, is even more sensitive to numerical precision problems than the RLS, and is generally viewed as unstable. A method of stabilizing this algorithm was introduced by Slock and Kailath in [43], which increased the numerical complexity from $7M$ to $8M$ multiplications and $6M$ to $7M$ addition/subtractions. This algorithm, known as the Stabilized Fast Transversal Filter (SFTF), is very similar to the original FTF so is not listed here. A derivation and listing of the SFTF algorithm can be found in [43], along with a Fortran implementation of the algorithm.

Symbol Key:

M - filter order

Inputs:

$\mathbf{u}_M(n)$ - M by 1 input vector

$\mathbf{u}_{M+1}(n)$ - $M+1$ by 1 input vector

λ - forgetting factor (set between 0 and 1)

$d(n)$ - desired signal

Intermediate States:

$\mathbf{a}_M(n)$ - $M+1$ by 1 forward prediction error filter

$\eta_M(n)$ - forward a priori prediction error

$f_M(n)$ - forward a posteriori prediction error

$\mathcal{F}_M(n)$ - minimum value of the sum of weighted forward prediction-error squares

$c_M(n)$ - $M + 1$ by 1 backward prediction error filter

$\psi_M(n)$ - backward a priori prediction error

$b_M(n)$ - backward a posteriori prediction error

$\mathcal{B}_M(n)$ - minimum value of the sum of weighted backward prediction-error squares

$\tilde{\mathbf{k}}_M(n)$ - M by 1 normalized gain vector

$\tilde{\mathbf{k}}_{M+1}(n)$ - $M + 1$ by 1 normalized extended gain vector

$\tilde{k}_{M+1,M+1}(n)$ - last element of normalized extended gain vector

$\gamma_M(n)$ - conversion factor (order M)

$\gamma_{M+1}(n)$ - conversion factor (order $M + 1$)

Outputs:

$\hat{\mathbf{w}}_M(n)$ - M by 1 tap weight vector

$\alpha(n)$ - a priori estimation error

$e(n)$ - a posteriori estimation error

Rescue variable - rescue by restart (saving $\hat{\mathbf{w}}_M(n)$ as initial condition) if negative

Algorithm:

$$\eta_M(n) = \mathbf{a}_M^H(n-1) \mathbf{u}_{M+1}(n) \quad (15)$$

$$f_M(n) = \gamma_M(n-1) \eta_M(n) \quad (16)$$

$$\mathcal{F}_M(n) = \lambda \mathcal{F}_M(n-1) + \eta_M(n) f_M^*(n) \quad (17)$$

$$\gamma_{M+1}(n) = \lambda \frac{\mathcal{F}_M(n-1)}{\mathcal{F}_M(n)} \gamma_M(n-1) \quad (18)$$

$$\tilde{\mathbf{k}}_{M+1}(n) = \begin{bmatrix} 0 \\ \tilde{\mathbf{k}}_M(n-1) \end{bmatrix} + \lambda^{-1} \frac{\eta_M(n)}{\mathcal{F}_M(n-1)} \mathbf{a}_M(n-1) \quad (19)$$

$$\mathbf{a}_M(n) = \mathbf{a}_M(n-1) - f_M^*(n) \begin{bmatrix} 0 \\ \tilde{\mathbf{k}}_M(n-1) \end{bmatrix} \quad (20)$$

$$\psi_M(n) = \lambda \mathcal{B}_M(n-1) \tilde{k}_{M+1, M+1} \quad (21)$$

$$\gamma_M(n) = (1 - \psi_M^*(n) \gamma_{M+1}(n) \tilde{k}_{M+1, M+1}(n))^{-1} \gamma_{M+1}(n) \quad (22)$$

$$\text{Rescue variable} = (1 - \psi_M^*(n) \gamma_{M+1}(n) \tilde{k}_{M+1, M+1}(n)) \quad (23)$$

$$b_M(n) = \gamma_M(n) \psi_M(n) \quad (24)$$

$$\mathcal{B}_M(n) = \lambda \mathcal{B}_M(n-1) + \psi_M(n) b_M^*(n) \quad (25)$$

$$\begin{bmatrix} \tilde{\mathbf{k}}_M(n) \\ 0 \end{bmatrix} = \tilde{\mathbf{k}}_{M+1}(n) - \tilde{k}_{M+1,M+1}(n) \mathbf{c}_M(n-1) \quad (26)$$

$$\mathbf{c}_M(n) = \mathbf{c}_M(n-1) - b_M^*(n) \begin{bmatrix} \tilde{\mathbf{k}}_M(n) \\ 0 \end{bmatrix} \quad (27)$$

$$\alpha_M(n) = d(n) - \hat{\mathbf{w}}_M^H(n-1) \mathbf{u}_M(n) \quad (28)$$

$$e_M(n) = \gamma_M(n) \alpha_M(n) \quad (29)$$

$$\hat{\mathbf{w}}_M(n) = \hat{\mathbf{w}}_M(n-1) + \tilde{\mathbf{k}}_M(n) e_M^*(n) \quad (30)$$

Initialization:

$$\mathbf{a}_M(0) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\mathcal{F}_M(0) = \lambda^M \mu$$

$$\mathbf{c}_M(0) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$\mathcal{B}_M(0) = \mu$$

$$\mathbf{k}_M(0) = \mathbf{0}$$

$$\gamma_M(0) = 1$$

$$\hat{\mathbf{w}}_M(0) = \mathbf{0}, \quad \text{for initial startup}$$

$$\hat{\mathbf{w}}_M(0) = \hat{\mathbf{w}}_0, \quad \text{for restarts} \quad (31)$$

Little (if any) documentation on acceptable values for the soft-initialization constraint μ have been given in literature. Simulations performed during the course of this thesis have found that values between 0.5 and 10 appear reasonable for filter sizes between 50 and 1200 taps.

2.2.5 Acoustic Echo Canceller Structure

It is a simple matter to take the generic filter structure of Figure 3 and apply it as an acoustic echo canceller for handsfree applications. The receive (Rx) signal (before reaching the loudspeaker) is defined as the reference signal (also known as the input signal $u(n)$), and the microphone signal as the primary signal (the desired signal $d(n)$). The subsequent error signal $e(n)$ becomes the new microphone or transmit (Tx) signal. This is illustrated in Figure 4 below.

FIGURE 4. Acoustic Echo Canceller Structure

2.2.6 Acoustic Echo Cancellation Performance Issues using LMS

LMS is a widely used and well researched algorithm. Describing all the various LMS performance characteristics, however, is beyond the scope of this thesis, so the following discussion is limited to a few major performance factors [49]:

1. **Step-size selection** - step-size selection is an optimization of speed of convergence against the accuracy of the converged values. In general, the larger the step-size μ , the faster the convergence. There is an upper bound on the size of μ , which is dependent on the number of filter taps and the power and eigenvalue spread of the input data. Choosing μ larger than this bound will result in instability. Choosing μ to ensure stability in real applications where input signal characteristics are not fully known requires some guess as to the expected maximum values of the input signal. Typically, μ will be set smaller than this upper bound to ensure accuracy, because the smaller μ gets, the more accurate the converged tap weights become. The mean-squared error is asymptotically limited by the level of uncorrelated measurement noise in the primary signal. There is no theoretical reason why one cannot make μ as small as possible - however, in practical implementations the numerical accuracy of the number system used usually limits this direction. In practical systems, the optimum μ is the one that maintains a reasonably fast rate of convergence with a reasonable level of accuracy, for the application under consideration.

2. **Filter length** - the longer the filter, the slower the convergence. Given a highly non-stationary room impulse response, a shorter filter typically provides better performance in tracking the changing response than a longer one which covers the whole length of the response. The optimum filter length, then, is the shortest possible length that gives the required level of cancellation for the worst case echo path in the application.
3. **Input signal characteristics** - LMS takes longer to converge given a reference signal with a large eigenvalue spread (spread of the largest spectral peak to the lowest spectral valley) than for one with a smaller spread. White noise has the smallest possible eigenvalue spread and is therefore the best signal for training an LMS filter. A sinusoidal signal, on the other hand, will almost certainly lead to tap weight divergence due to a lack of persistency of excitation (insufficient exercise of all the modes in the response to be modelled).
4. **Measurement noise** - this term includes any noise or signal in the primary signal that is uncorrelated with the reference signal. Generally, the larger the measurement noise, the slower the convergence and the larger the converged mean-squared error. Theoretical work has dealt almost entirely with the effects of uncorrelated Gaussian noise.
5. **Non-linear components** - this includes any non-linear components in the primary signal that are *correlated* with the reference signal. The effects of non-linearities are not as well understood as Gaussian measurement noise, but should have similar effects of slowing convergence and increasing the converged mean-squared error.

These characteristics need to be considered in any handsfree design that employs an LMS echo canceller.

2.3 Terminal/Room Acoustic Environment

Early work on characterizing room impulse responses was performed by Neely and Allen [7]. They characterized the invertibility of a room response (whether or not the impulse response was minimum phase) in terms of wall reflectivities and separation of source and receiver. They concluded that for highly reflective rooms or wide separations of source and receiver that the room impulse response may not be minimum phase and hence non-invertible by standard echo cancellation techniques.

Mourjopoulos [11] extended the work of Neely and Allen by examining the change in the measured room impulse response caused by changes in the locations of the source and receiver. It was found that changes in the locations of the source and receiver had a significant effect on the measured room impulse response.

Muron and Sikorav [14] investigated the change in terminal/room impulse response due to human body movement. They found that only a subset of the total impulse response changed during movement and concluded that acoustic echo cancellation algorithms should concentrate on fast adaptation of the coefficients within this movement window.

Sikorav in [27] continued work on identification of non-stationarities in audio conference rooms. He confirmed again that non-stationarities typically affect only a few windows of coefficients, and discussed a possible method of detecting these windows of motion within an impulse response. He also described a simple method of improving LMS tracking by assigning a larger adaptation step-size μ for the window of coefficients in which movement was most likely to occur.

2.4 General Advances in Adaptive Filtering Algorithms

It is beyond the scope of this thesis to describe all previous work on adaptive filtering algorithms. This thesis reviews previous work on the Fast Transversal Filter (FTF), and the subsequent Stabilized Fast Transversal Filter (SFTF), which it had been hoped that the SFTF would prove to be a suitable algorithm for acoustic echo cancellation. The SFTF algorithm will be shown later in this thesis to be unsuitable for this application, however. This review also provides an introduction to previous work on adaptive IIR filters.

2.4.1 Fast Transversal Filters

The original Fast Transversal Filter was derived by Cioffi and Kailath [9], and was a further reformulation and distillation of pioneering works on fast-RLS and fast-Kalman filters by Ljung, Morf and Falconer [5]. Cioffi and Kailath derived an exact RLS implementation with $O(N)$ complexity needing only $7N$ multiplications per iteration (where N is the order of the filter), compared to the $O(N^2)$ complexity of RLS. This reduction in complexity was accomplished by removing redundancies in the RLS solution, specifically by replacing the updating of the N by N inverse correlation matrix with updates of forward and backward prediction vectors. This removal of redundancy from the RLS algorithm, however, had the undesired effect of making the FTF algorithm numerically unstable when implemented in finite precision math. The FTF algorithm also suffers from a somewhat less catastrophic problem in that the derived method of initialization from zero initial conditions as given in [9] may also result in numerical problems when implemented in finite precision math [9] [51]. Cioffi and Kailath also give a method of initialization from arbitrary initial conditions using a weighting method on the initial conditions. This, however, introduces a measure of suboptimality (deviation from the true RLS solu-

tion) to the FTF solution, depending on the choice of weighting factor and the forgetting factor λ .

A significant breakthrough in the numerical stabilization of the FTF algorithm was introduced by Slock and Kailath [43]. They found that the error propagation mechanism in the FTF algorithm could be stabilized using error feedback methods. Specifically, the backward a priori prediction error was calculated using two theoretically equivalent methods (one, a scalar multiplication, the other, a vector multiplication), with the difference between the two (due to finite precision effects) fed back (using negative feedback gains) to stabilize the algorithm.

2.4.2 Adaptive IIR Filter Algorithms

A full discussion of adaptive IIR filter algorithms is beyond the scope of this thesis, but it had been hoped that the area had matured to the point where an adaptive IIR filter could be successfully employed as an acoustic echo canceller for long impulse responses (requiring a much lower order filter for the same length of cancellation). A sampling of recent papers on this subject, however, showed stability and convergence rate issues to be a continuing concern.

Ardalan and Faber [25] proposed a transversal filter structure with both poles and zeros, with tap-weight updating performed using a variant of the fast Kalman algorithm. Simulation results showed successful modelling of the poles and zeros of a 10th-order bandpass filter. This paper did not, however, discuss finite precision implementation, a problem which has plagued most fast RLS-fast Kalman implementations.

Fan and Jenkins [29] investigated the use of an adaptive IIR filter for a hybrid echo canceller. They investigated the performance of the Adaptive Filtering Mode (AFM) algorithm, which uses LMS type updates for the coefficients of pole and zero equations. The algorithm is given as follows:

$$y(n) = \sum_{i=1}^{\hat{n}_a} \hat{a}_i(n) y(n-i) + \sum_{j=0}^{\hat{n}_b} \hat{b}_j(n) x(n-j) \quad (32)$$

$$e(n) = d(n) - y(n) \quad (33)$$

$$\hat{a}_i(n+1) = \hat{a}_i(n) + \mu e(n) y'(n-i), i = 1, 2, \dots, \hat{n}_a \quad (34)$$

$$\hat{b}_j(n+1) = \hat{b}_j(n) + \mu e(n) x'(n-j), j = 0, 1, \dots, \hat{n}_b \quad (35)$$

$$y'(n) = y(n) + \sum_{i=1}^{\hat{n}} \hat{a}_i(n) y'(n-i) \quad (36)$$

$$x'(n) = x(n) + \sum_{i=1}^{\hat{n}} \hat{a}_i(n) y'(n-i) \quad (37)$$

Simulations of this algorithm as an adaptive IIR hybrid echo canceller were performed in this paper, leading the authors to the following conclusions:

With respect to an adaptive FIR canceller, it is less advantageous to use an adaptive IIR canceller if

1. the orders of the echo path transfer function are not known exactly
2. there is considerable measurement noise
3. fast convergence is required

For acoustic echo cancellation, all three points are true, which likely makes the use of this particular adaptive IIR algorithm unsuitable.

Haneda, Makino and Kaneda [58] have proposed that the poles in a typical room response correspond to the resonance properties of the room and can be estimated off-line, then inserted in an acoustic echo canceller as fixed Auto-Regressive (AR) coefficients of a linear filter. The zeros of the room response (as specified in the Moving Average (MA) coefficients of the linear filter) would continue to be adaptively updated. A block diagram of their proposed structure is given in Figure 5.

FIGURE 5. Acoustic Echo Canceller using Common Room Poles

Although a 1.5 times increase in convergence speed is cited with the fixed pole structure over conventional adaptive FIR structures for the same converged ERLE, a significant number of different estimates of the acoustical poles (the authors suggest 30) must be taken from different points in the room to get accurate averaged estimates. This makes the solution dependent on that particular room (which may be reasonable for a teleconferencing system permanently stationed in a room). They do not, however, take into account the effects of differing numbers of people that may be present and their effect on the room's acoustical poles.

2.5 Algorithm Enhancements for Acoustic Echo Cancellation

The majority of previous works on algorithm enhancements for acoustic echo cancellation capitalize on some particular a priori knowledge of the signal or environmental characteristics to improve canceller performance. The majority of these utilize some variant of the LMS algorithm.

2.5.1 Algorithms for Efficiently Covering Long Impulse Responses

Casar-Corredera and Alcazar-Fernandez in [17] proposed a method of reducing the computational requirements for long acoustic echo cancellers by assuming that a room impulse response could be modelled by a short impulse response convolved with a long, sparse impulse train. This model is illustrated in Figure 6 below:

FIGURE 6. Modelling of Room Impulse Response As Short-Long Convolution

This paper does not give a good justification why the early reflections (as denoted by $p(n)$), should continue to be redistributed throughout the remainder of the room impulse response. This is in fact contrary to most fundamental models of acoustics which treat early reflections and later reverberation as largely uncorrelated to each other [30].

A modified LMS algorithm which updates the coefficients of $p(n)$ and $u(n)$ is derived in this paper, but the fact that only artificially generated data was utilized to test the algorithm is another indication that their model may not be realistic.

Kawamura and Hatori in [18] proposed a tap selection scheme in which at a given instant in time, only a subset of possible taps were being used for adaptive filtering. The selection criteria is based on the assumption that the smaller the tap magnitude, the less that tap con-

tributes to the filter output. Tap selection is not performed on a per sample basis. After a certain pre-defined period of regular adaptive filter iterations, the tap with the smallest magnitude is replaced by the first tap in a first-in, first-out queue of previously unused or discarded tap positions. Figure 7 illustrates this concept.

FIGURE 7. Tap Selection/Replacement Mechanism

This algorithm does have certain advantages. It ensures that for a computationally limited system, any significant echoes within a possibly lengthy time frame are being covered. The system is also adaptive and will follow slow changes in the acoustic environment. The significant disadvantage to this system is the computational overhead required for the tap selection, which is dependent on the desired tap discard rate.

2.5.2 Algorithms for Improved Convergence with Speech Input

Several papers have been written on schemes for improving the convergence of LMS and Normalized LMS (NLMS) algorithms while using speech as the training signal. LMS and NLMS are known to perform poorly (converge slowly) under conditions of large eigen-

value spreads (i.e. large spreads between spectral peaks and valleys) [49]. The underlying assumption in these papers is that convergence speed can be increased by pre-whitening (reducing the eigenvalue spread of) the input signal to the adaptive filter.

In papers by Yasukawa, Shimada and Furukawa[21] and Yasukawa, Furukawa, and Ishiyama [32], a variant of LMS was implemented which employed input signal decorrelation. This algorithm (assuming real inputs) is summarized in the following steps:

1. Determine estimation error (same as standard LMS):

$$e(n) = d(n) - \hat{\mathbf{w}}^T(n) \mathbf{u}(n) \quad (38)$$

2. Determine first correlation coefficient:

$$\rho_1 = \frac{\mathbf{u}^T(n) \mathbf{u}(n-1)}{\mathbf{u}^T(n-1) \mathbf{u}(n-1)} \quad (39)$$

3. Decorrelate input vector:

$$\mathbf{Z}(n) = \mathbf{u}(n) - \rho_1 \mathbf{u}(n-1) \quad (40)$$

4. Update tap weight vector with decorrelated inputs:

$$\hat{\mathbf{w}}(n+1) = \hat{\mathbf{w}}(n) + \mu \frac{\mathbf{Z}(n) e(n)}{\mathbf{u}^T(n) \mathbf{Z}(n)} \quad (41)$$

Simulation results in these papers show improved convergence speed for this algorithm relative to NLMS for speech input signals. From a linear systems standpoint, however, there appears to be a problem in the fact that they are in effect applying a filter to the input (reference) signal only, and not to the desired (primary) signal. The decorrelation of the

input signal should also decorrelate the input and desired signals from each other, making convergence more difficult.

To this end, several papers have been published which incorporate the decorrelation in both the input and desired signal paths.

Hansler [40] proposed a structure that inserted block-updated decorrelation filters in the input and desired signal paths. This structure is illustrated below in Figure 8.

FIGURE 8. Hansler's Decorrelation Structure

Although this structure ensures that the reference and primary inputs to the NLMS algorithm are both filtered, it requires a complete re-filtering of the reference input vector each

time an update of the decorrelation filters is performed. This is due to the fact that the length of the reference input vector will typically be an order or two larger in magnitude than the length of the decorrelation filter (which would typically be 6-15 taps long).

There is one significant theoretical problem with this approach. Since the decorrelation filter affects the desired signal, any uncanceled portion of the desired signal (including room/measurement noise) may be amplified by the decorrelation filter, resulting in a lower ERLE than would be obtained without the filter.

Mboup, Bonnet and Bershad [59] proposed a structure that inserted low-order LMS pre-whitening filters in the reference input and error output paths. This structure is illustrated below in Figure 9.

FIGURE 9. Mboup, Bonnet and Bershad's Decorrelation Structure (Classical)

They also presented a simplified pre-whitening structure in which only the input signal was pre-whitened. This structure is illustrated in Figure 9.

FIGURE 10. Mboup, Bonnet and Bershad's Decorrelation Structure (Simplified)

In this structure, the pre-whitening filter is implemented as an adaptive LMS predictor. This paper demonstrates that the stability of the entire system is dependent on the mutual selection of the pre-whitener adaptation step-size and the acoustic echo canceller adaptation step-size. From their analysis, it appears that the pre-whitener step-size is required to stay in the range $[0, \frac{1}{P_w + MP_s}]$, where P_w is the input noise power, P_s is the input signal power, and M is the order of the predictor. An interesting point about this paper is that although adaptation is claimed to be faster with the pre-whitener than without, no simulations are given to back up this claim.

Although beyond the scope of this thesis, a discussion of current work in acoustic echo cancellation algorithms would not be complete without some mention of subband acoustic echo cancellation. The subbanded approach splits the input signals into several frequency bands, with separate adaptive filters performing cancellation for each band [36]. This approach allows a reduction in the sampling rate in each band, requiring fewer taps to cover the same echo length, and allows a reduction in the tap update frequency [36].

To keep the discussion reasonably brief, a few representative papers on the subband acoustic echo cancellation approach have been selected ([22], [28], [38], [39], and [47]), from which it appears the general subband structure can be traced back to a 1985 paper (in German) by Kellerman [12]. This general structure (taken from [22]) is displayed in Figure 11.

Gilloire [22] points out several major advantages of sub-band acoustic echo cancellers over their full-band counterparts:

1. **Computational Efficiency:** The sub-band signals in each band can be decimated by a factor up to the number of sub-bands, for which computations can be performed at the reduced rate. This allows a reduction in computational requirements close to the decimation factor (allowing for computational overhead required for the analysis/synthesis filterbanks). The length of the filter in each band can also be optimized to specific echo response length at that range of frequencies.
2. **Increased Adaptation Speed:** The adaptation gain (or step-size) may be adjusted and optimized according to the signal power present in each band.

FIGURE 11. Sub-band Acoustic Echo Canceller Structure

2.5.3 Algorithms with Dynamic Adaptation Step-Size

Several papers have been written which try to improve the convergence and/or tracking properties of LMS by varying the adaptation step size.

Chen and Vandewalle [26] proposed a variant of the NLMS algorithm in which the traditionally scalar step-size constant μ is replaced by a time varying μ -vector, in which more recent taps are given larger values of μ . The tap-update portion of the algorithm is given as follows:

$$P(n) = (1 - \alpha) P(n-1) + \alpha u^2(n) \quad (42)$$

$$q(n) = \frac{a}{P(n) + P_{\min}} \quad (43)$$

$$\mu_i(n) = (1 - \beta)^{i-1} q(n) \quad (44)$$

$$\mu_{\text{vec}}(n) = \begin{bmatrix} \mu_1(n) & & & 0 \\ & \mu_2(n) & & \\ & & \ddots & \\ 0 & & & \mu_M(n) \end{bmatrix} \quad (45)$$

$$\hat{\mathbf{w}}_M(n+1) = \hat{\mathbf{w}}_M(n) + 2\mu_{\text{vec}}(n) e(n) \mathbf{u}_M(n) \quad (46)$$

where

M is the filter order

$P(n)$ is the input power estimate

α is a small constant less than 1 (generally $1/M$)

$u(n)$ is the current reference input sample

$q(n)$ is the input power-dependant adaptation step-size (as used in NLMS)

a is the step-size constant

$P_{\min}$ is a small constant to maintain a numerical bound on $q(n)$

$\mu_i(n)$ are the individual elements of $\mu_{\text{vec}}(n)$

β is a forgetting factor between 0 and 1 ($\beta = 0$ results in NLMS operation)

Simulations in this paper were performed with the algorithm running in an Adaptive Line Enhancer (ALE) configuration, and showed improved output error performance over NLMS under non-stationary conditions. Although this paper was not directly aimed at acoustic echo cancellation applications, there are many similarities between it and a subsequent paper by Makino and Kaneda [37], which references the former paper.

Makino and Kaneda [37] appear to have taken the basic idea of a μ -vector NLMS algorithm and adapted it for use in an acoustic echo cancellation application. In this paper the authors propose that the elements of the μ -vector should be proportional to the mean-squared error of the corresponding tap-weight, in order to update tap-weights with large errors with large steps. They noted that in typical teleconferencing situations, the decay of the speaker to microphone impulse response was a decaying exponential. For this reason they initialized their algorithm to have an exponentially decaying μ -vector, after which the μ -vector should be updated with respect to the mean-squared error of the corresponding coefficient. For practical reasons, however, simulations were performed with a time-invariant exponentially decaying μ -vector. Further practical modifications were also tried, where the tap-weight vector was divided into sections, with each section given a fixed μ value.

Simulation results showed a tripling of convergence speed for white noise inputs and a doubling of convergence speed for speech inputs relative to conventional NLMS performance.

A recent paper by Kwong and Johnston [62] proposed a variable step LMS algorithm which varied the adaptation step-size based on the mean-squared estimation error. The

step-size itself was made adaptive, increasing for large mean-squared error and decreasing for small mean-squared error. This adaptation scheme is given as:

$$\mu'(n+1) = \alpha\mu(n) + \gamma e^2(n) \quad (47)$$

where

$$0 < \alpha < 1, \quad \gamma > 0 \quad (48)$$

and

$$\mu(n+1) = \begin{cases} \mu_{\max}, & \mu'(n+1) > \mu_{\max} \\ \mu_{\min}, & \mu'(n+1) < \mu_{\min} \\ \mu'(n+1), & \text{otherwise} \end{cases} \quad (49)$$

Some suggested values in the paper for α and γ are $\alpha = 0.97$ and $\gamma = 4.8 \times 10^{-4}$.

Simulation results in the paper show a considerably faster (2 to 3 times) convergence speed than standard LMS under stationary conditions, and demonstrates improved tracking performance under non-stationary conditions. This is all achieved with essentially the addition of an extra tap update.

2.6 Methods for Improving Loudspeaker Linearity

A major limitation of acoustic echo canceller performance discussed in this thesis is the impact of non-linearities (distortion) in the loudspeaker on the converged depth of cancellation.

Gao and Snelgrove in [48] addressed the general problem of low-frequency cone suspension non-linearities and proposed an adaptive non-linear Volterra filter that pre-compensates

sates the input signal to the loudspeaker so as to give overall linear performance from the system. This concept of pre-distortion is illustrated in Figure 12:

FIGURE 12. Non-Linear Preprocessor For Linearization of Loudspeaker

A complete block diagram of their proposed system is given in Figure 13. The two dashed lines indicate that either the output from the actual loudspeaker or the output from the non-linear adaptive filter block can be used to adapt the linear adaptive filter. The filter coefficients from the two adaptive filters are copied over to the filters in the pre-processor.

This linearization scheme is based on the assumption that the non-linearity is weak in comparison to the linear operation of the loudspeaker, and that the loudspeaker transfer function can be divided into a cascade of purely non-linear (Volterra) and purely linear filters. The linear filter can then be updated using simple LMS, while the Volterra filter is updated in a similar fashion except for the presence of additional multiplicative input terms in the tap update equation.

Simulation results showed a 27 dB reduction in the non-linear distortion of a simulated loudspeaker. The major drawback to this method is the large computational requirement. For the third order Volterra filter implemented for the simulations, 1117 filter taps were

required, with 1000 of them requiring 4 multiplications for each tap update. There was no discussion of model simplification and whether a smaller benefit could be achieved with far fewer filter coefficients.

FIGURE 13. Adaptive Non-Linear Processing Scheme

2.7 Acoustic Echo Cancellation Limitations

Little attention has been given in signal processing literature on the impact and limitations of physical and environmental conditions on acoustic echo cancellation performance. One

exception is the recent paper by van de Kerkhof and Kitzen on the tracking of a time-varying acoustic impulse response by an adaptive filter [61]. Numerous simulations were performed to gain an understanding of the effect a typical person walking through a room would have on the mean-squared error output of an LMS filter for different adaptation step-sizes. They concluded that the simple LMS algorithm is unable to track (maintain the same mean-squared error output for) the non-stationarities due to a person walking through a room.

2.8 Literature Review Summary

This review of previous work on acoustic echo cancellation and related problems show many issues remain to be solved. The issues of improving convergence speed with speech input and improving the tracking capabilities (to follow room non-stationarities) during periods of movement in the acoustical environment are two significant problems that affect cancellation performance which have been addressed. Less emphasis, however, has been placed on the impact of physical conditions which an acoustic echo canceller would be required to work under. Environmental conditions such as room noise, the acoustic impulse response length, and system non-linearities will be shown in this thesis to be basic limitations on canceller performance that have as much significance as the dynamic performance aspects of convergence and tracking.

This thesis explores how the various physical conditions relate to the overall performance of an acoustic echo canceller. Section 3.0 describes the collection of two-channel data from actual handsfree terminal and room environments and a description of the basic processing of the data. Section 4.0 examines five basic sets of data and the change in the con-

verged ERLE found as the number of adaptive filter taps is changed. Section 5.0 looks at the impact room noise and the decay rate of the terminal/room impulse response have on the achievable ERLE. Section 6.0 looks at the impact system non-linearities have on the achievable ERLE and presents an adaptive filtering structure that uses a non-linear pre-processor to improve the achievable ERLE for certain non-linear systems. Section 7.0 looks at dynamic performance issues such as convergence speed and tracking ability for several different adaptive filtering algorithms. Finally, Section 8.0 presents conclusions on the findings of this thesis.

3.0 EXPERIMENTAL SETUP FOR AEC MEASUREMENTS

The first major step in examining and understanding the performance of an acoustic echo canceller is to understand the length and shape of the impulse response between the coupled loudspeaker and microphone. In desktop terminal applications, this coupling will include three basic components:

1. the direct path (if any) between loudspeaker and microphone
2. early reflections and reverberation within the terminal enclosure (terminal impulse response)
3. early reflections and reverberation outside the terminal enclosure (room impulse response)

The direct path happens very early, and is limited by the speed of sound in the medium (air and plastic). The second and third components typically overlap to some degree, depending on the length of the terminal impulse response and the proximity of reflecting surfaces external to the set. [30]

Figure 14 shows a simplified coupling response for an arbitrary terminal/room combination.

FIGURE 14. Simplified Coupling

3.1 Impulse Response Measurement Setup

In order to estimate the coupling impulse response of handsfree sets in different room environments and under varying acoustic conditions (e.g. physical movement within the room), it was necessary to measure the coupling between the loudspeaker and microphone of typical sets. The measurement was made by applying a known signal to the loudspeaker and recording the signal coupled to the microphone. This was performed on two different terminal setups, each with different transducer and enclosure types, and in two different room environments.

The first terminal setup used a modified production set with low-cost narrowband transducers (intended for the 300-3400 Hz telephony bandwidth). The second terminal setup used a prototype set with improved wideband transducers (intended for a 100-7000 Hz bandwidth).

The first room used was a medium sized (3.5W x 5.5L x 2.5H) conference room (named DPN-A) with a large central table, several cushioned seats and hard walls. The second room used was a large anechoic chamber located in the administration building of the BNR Carling site. This chamber was essentially echo and noise free.

The combined terminal/room impulse response was measured with the following custom hardware developed for this purpose:

1. High SNR gain box - this box allowed arbitrary levels to the loudspeaker and from the microphone to be set independently. It also split the reference signal to allow simultaneous recording of the reference and primary signals to Digital Audio Tape (DAT). A schematic diagram of this gain box is shown in Figure 15.
2. Speech / Noise Digital Audio Tape (DAT) - a source tape was made using speech samples from the NTT "Speech Data Base for Telephony 1988" CD (Track 1) and white noise (bandlimited to 300-3300 Hz) from Denon's "Denon Audio Technical CD" (Track 79).

Digital audio tape was chosen as the recording format for its high quality recording and reproduction capabilities. Sound data is stored on the tape in 16-bit digital words at a sampling rate of 48 kHz. Although not available for this thesis, direct digital access to the original data for downloading to a PC file will eventually be possible. For this thesis, PC files were created by resampling at 8 kHz the analog output of the DAT recorder using a PC-based DSP board with 16 bit A/D and D/A capabilities.

The complete measurement setup is shown in Figure 16. One DAT recorder was employed as the source (one channel used) while a second DAT recorder recorded simultaneously

Gain Block Schematic Diagram for Set Echo Response Recordings

FIGURE 15. Gain Box Schematic Diagram

the source track from the first DAT machine on one channel and the returning echo on the other channel.

3.2 Terminal/Room Impulse Responses

After recording two-channel terminal/room response data to DAT, a representative ten second, two channel segment from each recorded terminal/room/source condition was sampled at 8 KHz and saved in 16 bit Hypersignal format (.TIM files - see Appendix C) as separate reference and primary signals. These files were then used as inputs to a C program implementing the LMS algorithm as an echo canceller in double floating point precision, from which an error signal could be computed and re-stored as a .TIM file. A full list of files generated from the DAT recordings is given in Appendix A.

FIGURE 16. Terminal / Room Impulse Response Measurement Setup

Four two-channel file sets were chosen as 'reference' recordings to compute reasonably accurate reference terminal/room impulse responses which acted as a starting point for further experimentation. These files had the following characteristics:

TABLE 1. Files Used For Reference Impulse Response Measurements

Filename	Terminal	Room
WBNR13.TIM WBNP13.TIM	Wideband Prototype	DPN-A Conference Room
NBNR16.TIM NBNP16.TIM	Low-cost Narrowband	DPN-A Conference Room
WBNR22.TIM WBNP22.TIM	Wideband Prototype	Admin building anechoic chamber
NBNR25.TIM NBNP25.TIM	Low-cost Narrowband	Admin building anechoic chamber

The DOS file names chosen for the .TIM files follow these conventions:

1. The first two letters signify the type of set used - 'WB' for a prototype set with improved wideband transducers, and 'NB' for a modified production set with low-cost narrowband transducers. This DOES NOT signify the bandwidth of the recording or source - all recordings were sampled at 8 kHz and all source material was bandlimited to 4 kHz.
2. The third letter signifies the source type - 'N' for 300-3300 Hz bandlimited Gaussian noise. 'S' was used for sample speech (Harvard sentences) bandlimited to 4 kHz (not used for these particular measurements).
3. The fourth letter signifies from which channel of the two channel segments the file was generated - 'R' for the Reference (in this case the signal before being played through the loudspeaker), and 'P' for Primary (in this case the signal recorded from the microphone output). This terminology is standard for most echo/noise cancelling literature.

4. The two digit number identifies the DAT track from which the ten second segment was sampled.

All files in Table 1 were recorded with no physical movement in the acoustic environment to get as close to a stationary room response as possible.

Representative impulse responses for each of the four terminal/room conditions in Table 1 were generated by averaging the tap weights obtained by running the reference and primary signals through an LMS filter. The LMS filter was chosen to have 1200 taps (150 ms) and an empirically determined step size constant of $\mu=0.001$ (slow convergence but small mean-squared error). The LMS filter was allowed to adapt over the full 8000 samples (10 seconds) of data, with the tap weights averaged over the last 5000 iterations well after the filter had converged. The averaging was performed to minimize the bias of any local statistical variations in the data. These tap weights provide an accurate estimate of the impulse response of the echo path of the terminal.

The averaged tap weights were plotted and are included here for each of the four terminal/room setups in Figures 17 to 20¹. By performing measurements in an anechoic chamber, with no significant echo contribution of its own, the coupling response of the terminal itself is measured (Figures 18 and 20). Comparing these to the combined terminal/room responses measured in the DPN-A conference room (Figures 17 and 19), one can see that the magnitude and power of the early terminal coupling response of the terminals is significantly greater than that of the room coupling, and also shows how the returning echoes

1. Note: although the length of the impulse responses is given in tap weight numbers, the actual length (given a sampling rate of 8 kHz) in ms at any point along the graph can be determined by using the relation $200 \text{ taps} = 25 \text{ ms}$. Thus, each impulse response shown here has been computed up to 150 ms.

from the room environment overlap the terminal coupling response (even the first 50 ms of the DPN-A and anechoic responses are not identical). It is interesting to note that the Wideband set impulse response measured in DPN-A also shows a remarkable similarity to a representative terminal/room impulse response measured by Hansler [40]. It is also interesting to note the differences between the Low-cost narrowband and Wideband prototype terminal responses. The Low-cost set has a much more tightly coupled impulse response (faster decay rate) than the Wideband prototype set, which intuitively should allow the use of less filter taps for the same amount of cancellation. The importance of this decay rate is investigated further in this thesis in Section 5.2.

FIGURE 17. Wideband prototype set / DPN-A Conference room impulse response

FIGURE 18. Wideband prototype set / anechoic chamber impulse response

FIGURE 19. Low-cost Set / DPN-A Conference room impulse response

FIGURE 20. Low-cost set / anechoic chamber impulse response

3.2.1 Choice of Adaptation Step-Size

To get some idea of different convergence speeds achievable with different choices for μ before running further simulations, a 250 tap LMS filter was applied to the Low-cost set / DPN-A conference room data (files NBNR16.TIM and NBNP16.TIM). Four different iterations were run, with μ values of 0.001, 0.004, 0.016 and 0.064. Convergence speed was measured by measuring the time it took for the filter to achieve a stable Echo Return Loss Enhancement (ERLE). The ERLE, a measure of power reduction between the primary signal (the desired signal) $d(n)$ and the error signal $e(n)$, is formally defined in equation (50).

$$\text{ERLE (dB)} \equiv 10 \log \frac{E[d^2(n)]}{E[e^2(n)]} \quad (50)$$

where $E[\]$ is the expectation operator [29].

For each iteration, the primary file and the generated error file were run through an 'ERLE filter', in which this usual long-term average definition of ERLE was substituted with the following moving average definition to obtain a smoothed measure of local ERLE performance:

$$\text{ERLE (dB)} = 10 \log \frac{\sum_{i=0}^{255} d^2(n-i)}{\sum_{i=0}^{255} e^2(n-i)} \quad (51)$$

This itself was plotted vs. time for the different μ cases. These results are plotted in Figures 21 to 24. These examples illustrate the general tradeoffs in choosing μ . Compared to

$\mu=0.004$, convergence speed for $\mu=0.001$ begins to decline rapidly with only a very slight improvement in the average ERLE (not visible or audible). This is due to the mean-squared error reaching its asymptotic limit based on measurement/room noise and other system non-linearities. Thus, reducing μ below 0.004 results in an overall degradation of performance. On the other hand, moving μ all the way to 0.064 results in a large improvement in convergence speed over $\mu=0.004$, but suffers a large drop in the converged mean-squared error. Thus, for any μ value between (approx.) 0.004 and 0.064 (above which stability becomes questionable), the design choice becomes a trade-off between convergence speed and converged mean-squared error, in which the application will typically make the decision. Table 2 gives approximate convergence time and ERLE numerical data for the varying μ cases.

TABLE 2. Effects of Varying μ (250 tap LMS - Low-cost NB/ DPN-A conf. room data)

μ	Approx. Convergence Time (s)	Converged ERLE (dB)
0.001	5.0	11.39
0.004	1.0	11.37
0.016	0.4	10.84
0.064	0.2	7.49

The choice of μ is also dependent on the power of the input signal. Speech has a much higher peak-to-RMS ratio than Gaussian noise, resulting in lower RMS levels with speech for a given maximum peak determined by maximum coding levels. For fixed μ , this may allow a larger value of μ to be used to optimize convergence speed/converged ERLE. The dynamic nature of speech, however, makes a variable step size scheme related to the input power (such as normalized LMS) more appropriate.

FIGURE 21. ERLE vs. time for 250 tap LMS with $\mu=0.001$ (Low-cost set in DPN-A)

FIGURE 22. ERLE vs. time for 250 tap LMS with $\mu=0.004$ (Low-cost set in DPN-A)

FIGURE 23. ERLE vs. time for 250 tap LMS with $\mu=0.016$ (Low-cost set in DPN-A)

FIGURE 24. ERLE vs. time for 250 tap LMS with $\mu=0.064$ (Low-cost set in DPN-A)

4.0 ACOUSTIC ECHO CANCELLATION PERFORMANCE LIMITATIONS

As a starting point for examining the limitations on the obtainable ERLE, simulations were run on the four sets of stationary real 'reference' data described in Section 3.2 and on an artificial set of data. The term 'real' is used to describe the source of the data (from an actual physical system) and does not mean real (as opposed to complex) numbers. The artificial data consisted of a previously recorded reference file (WBNR13.TIM) and a primary file (AWBNP113.TIM) created from the convolution of the reference noise file with the 1200 tap impulse response file LWBN1200.FLP (displayed in Figure 17) plus Gaussian noise (to give a primary file signal+noise to noise ratio of 40 dB). This artificial primary file creation procedure is illustrated in Figure 38.

FIGURE 25. Artificial Primary File Creation Process

These simulations consisted of running the same data through LMS filters of varying length (25 to 1300 taps for the artificial data, 25 to 1200 taps for the real data) with the same adaptation step-size ($\mu=0.004$). In these simulations, the convergence speed was not important as long as the filter converged within the given data window (which it did for all filter lengths). As in the generation of the reference impulse responses in "Terminal/Room Impulse Responses" on page 45, the LMS filters were allowed to adapt over the full 10 seconds of data, generating an error signal $e(n)$. To obtain the final converged ERLE for each simulation the expectation operator $E[\]$ (see Equation (50)) was approximated by averaging over the last 5000 $d^2(n)$ and $e^2(n)$ samples (of 80000), where convergence was guaranteed. This converged ERLE data is summarized below in Table 3 for artificial data and Table 4 for real data and plotted in Figures 27 and 28 as obtainable ERLE curves (the ERLE achievable after convergence).

TABLE 3. Converged ERLE vs. Number of Taps, Artificial Data

# Taps	ERLE from WBNR13 & AWBNP113
25	1.59
50	3.97
75	5.48
100	7.25
150	10.22
200	12.19
250	13.00
300	14.59
350	16.09
400	17.18
500	18.72
600	21.16
700	22.64
800	24.76

TABLE 3. Converged ERLE vs. Number of Taps, Artificial Data

# Taps	ERLE from WBNR13 & AWBNP113
900	26.96
1000	28.33
1100	31.71
1200	38.59
1250	38.55
1300	38.50

TABLE 4. Converged ERLE vs. Number of LMS Taps, Real Data

# Taps	NB, DPN-A	NB, Anechoic	WB, DPN-A	WB, Anechoic
25	6.76	11.69	1.60	2.69
50	8.35	13.87	3.96	4.87
75	8.87	16.19	5.41	6.57
100	9.37	16.84	7.17	8.49
150	10.34	17.78	9.88	13.41
200	10.87	18.29	11.63	16.58
250	11.37	18.42	12.33	18.05
300	11.63	18.45	13.57	19.69
350	11.98	18.49	14.71	20.35
400	12.13	18.46	15.46	21.88
500	12.35	18.40	16.53	23.44
600	12.51	18.33	17.93	23.87
700	12.54	18.25	18.56	25.03
800	12.59	18.20	19.23	25.54
900	12.58	18.13	19.71	25.55
1000	12.52	18.04	19.85	25.79
1100	12.45	17.97	20.20	25.84
1200	12.38	17.89	20.38	25.85

ERLE Obtained for Stationary Artificial Data, Noise Input, LMS, $\mu=0.004$, 10s Adaptation

FIGURE 26. ERLE vs. Number of LMS Tap Weights for Artificial Data, $(S+N)/N = 40$ dB

ERLE Obtained for Noise Input, Stationary Room Response, LMS, $\mu=0.004$, 10s Adaptation

FIGURE 27. ERLE vs. Number of LMS Tap Weights for DPN-A Conf. Room

ERLE Obtained for Noise Input, Stationary Room Response, LMS, $\mu=0.004$, 10s Adaptation

FIGURE 28. ERLE vs. Number of LMS Tap Weights for Anechoic Chamber

Each of the ERLE vs. Number of Tap Weights curves in Figures 27 and 28 show two important characteristics. First, starting from a small number of taps, the obtainable ERLE increases as more filter taps are added. At a certain point, however, the obtainable ERLE curve approaches an asymptote, after which adding additional filter taps gives no improvement in the obtainable ERLE. This thesis proposes that the obtainable ERLE for a given filter length under stationary impulse response conditions is limited by one of three different factors:

1. Terminal/room impulse response decay rate: this factor is measured by the Total Impulse response Power to Tail Power ratio as defined in Section 5.2. This becomes the limiting factor of the obtainable ERLE for adaptive filter lengths significantly less than the terminal/room impulse response length.
2. Room/measurement noise: this factor becomes the asymptotic limit for the obtainable ERLE as the adaptive filter length approaches the terminal/room impulse response length if significantly greater than the effects of system non-linearities
3. System non-linearities: this factor becomes the asymptotic limit for the obtainable ERLE if significantly greater than the effects of room/measurement noise.

Figure 29 shows a simplified obtainable ERLE curve for stationary conditions and the hypothesized conditions which affect this ERLE. The limitations due to room/measurement noise and adaptive filter length, including a definition and derivation for the TIP/TP measure are discussed in Section 5.0. The limitations due to non-linear effects are discussed in Section 6.0.

FIGURE 29. Generalized Obtainable ERLE Curve

A fourth factor affecting the obtainable ERLE are impulse response non-stationarities resulting from such things as movement within a room. These are transient effects which affect the obtainable ERLE at a given instant in time - a measure of an adaptive filter's ability to track these non-stationarities becomes important. This topic is discussed in Section 7.0.

Finally, there are implementation specific factors that will affect the obtainable ERLE. Probably the most significant of these are numerical precision effects. A discussion of these effects is beyond the scope of this thesis, but these effects were minimized by using 16-bit precision (approx. 96 dB of dynamic range) for the input and output data, and double precision floating point for all algorithm implementations. The integer data was converted to floating point before use in any algorithm. A discussion of finite precision effects can be found in [49]. Some algorithms discussed later in this thesis did show sensitivity to finite precision computations within the algorithm, but this was a characteristic of the

algorithm that could not be avoided. These effects did not result in limitations on the accuracy of the outputs as such but generally resulted in complete unboundedness and failure of the algorithm.

Note: The sudden rise in ERLE for the artificial data before full modelling is achieved at 1200 taps (see Figure 26) appears suspect until a careful look at the definition of ERLE is given. The ERLE gives a measure of the ratio between the desired and error signals. In these simulations, the desired power (the power of the primary file) remains constant as the number of taps is changed. The change in the ERLE as the number of taps changes is thus due solely to the error power, which for a predominant part of the ERLE curve is due to uncanceled taps from the convolved impulse response. For the intermediate part of the ERLE curve (e.g. 400 - 800 taps), cancellation of 100 more taps does not have as significant a numeric effect on the error signal as cancelling 100 of the last 200 taps, which explains the sudden rise in the ERLE as full modelling is approached. The use of 50 and 100 tap steps also does not give a detailed picture of the behaviour of the ERLE as full modelling is achieved and over-modelling begins (> 1200 taps). The ERLE plot looks somewhat discontinuous plotted in these coarse tap steps, but with finer increases in the number of tap weights the curve should show a smoother transition to the room noise asymptote.

5.0 ERLE LIMITATIONS DUE TO ROOM NOISE AND ADAPTIVE FILTER LENGTH

5.1 ERLE Limitations Due to Room Noise

In acoustic echo cancellation systems, it can be shown that room/measurement noise (for a linear transfer function between the reference and primary signals) becomes the asymptote for the achievable converged ERLE. This is partly due to the definition of ERLE itself.

Since the error signal contains both uncanceled echoes and uncancellable measurement/room noise, as the amount of cancellation increases with the number of taps, the uncancellable noise becomes the predominant factor in the ERLE calculation. Thus, for a linear time invariant physical system, a converged adaptive FIR with N filter taps with finite tap weight misadjustment, and Gaussian room/measurement noise,

$$\begin{aligned}
 \lim_{N \rightarrow \infty} (\text{ERLE (dB)}) &= \lim_{N \rightarrow \infty} \left(10 \log \frac{E \{ d^2(n) \}}{E \{ e^2(n) \}} \right) \\
 &= \lim_{N \rightarrow \infty} \left(10 \log \frac{\sigma_d^2}{\sigma_e^2} \right) \\
 &= 10 \log \frac{\sigma_d^2}{\sigma_v^2} \\
 &\approx \frac{(\text{Signal} + \text{Noise})}{\text{Noise}} \text{Ratio (dB) of Primary (d(n))}
 \end{aligned} \tag{52}$$

This can be proven as follows:

The estimation error $e(n)$ can be defined as

$$e(n) = \sum_{i=0}^{\infty} w_{T/R}(i) u(n-i) - y(n) + v(n) \tag{53}$$

where

$w_{T/R}(i)$ is the terminal/room impulse response of infinite length

$u(n)$ is the reference input at time n

$y(n)$ is the adaptive filter output at time n

$v(n)$ is a sample at time n from a Gaussian random noise process,

and

$$d(n) = \sum_{i=0}^{\infty} w_{T/R}(i) u(n-i) + v(n) \quad (54)$$

Thus, for a converged system with $N \rightarrow \infty$,

$$y(n) \approx \sum_{i=0}^{\infty} w_{T/R}(i) u(n-i) \quad (55)$$

Substituting (55) into (53) gives

$$e(n) \approx v(n) \quad (56)$$

which explains the substitution in the final two steps of (52)¹.

Essentially, this derivation implies that under ideal conditions, the maximum ERLF

obtainable will be equivalent to the (signal+noise)-to-noise ratio of the returning echoed

1. The approximately equal to symbol ($\approx$) is necessary in the last two steps in equation 52 to account for finite adaptive filter weight misadjustments (i.e. for LMS, gradient noise), which should not be a significant factor after convergence.

signal. This is further demonstrated in the ERLE curve plotted for the artificially created data (see Figure 26, "ERLE vs. Number of LMS Tap Weights for Artificial Data, $(S+N)/N = 40$ dB," on page 59). Here, the amount of cancellation approaches but never exceeds the primary file's signal+noise to noise ratio of 40 dB.

For real data, however, this does not take into account any non-linearities in the terminal/room response, such as loudspeaker distortion, which may affect the convergence and misadjustment of the adaptive tap weights. This topic is explored further in Section 6.0.

It is important to remember that ERLE values are only measures of power reduction - that even perfect echo cancellation will result in a finite ERLE if there is any additional uncorrelated element (i.e. measurement/room noise).

5.2 ERLE Limitations Due to Adaptive Filter Length

The next step in examining the limitations on the achievable ERLE in an acoustic echo canceller is to determine the impact undermodelling (adaptive filter length < terminal/room impulse response) has on the ERLE.

If the impact of the room impulse response is disregarded for a moment, a ratio of the total anechoic terminal impulse response power to the power of a pre-defined length of the tail of that same impulse response can be defined. This gives a measure of the loudspeaker-microphone coupling decay rate for a given terminal. This will be called the Total Impulse response Power to Tail Power (TIP/TP) ratio (in dB), and is defined as follows:

$$\text{TIP/TP}(n, M) \text{ (dB)} = 10 \log \frac{\sum_{i=0}^{M-1} h^2(i)}{\sum_{i=n}^{M-1} h^2(i)} \quad (57)$$

where

h is an impulse response of length M , and n is the discrete point at which the "tail" is considered to start.

The TIP/TP measurement is illustrated in Figure 30.

FIGURE 30. TIP/TP Measurement on Impulse Response

This measurement is very similar to Early-to-Late Reflection power ratios ($[\text{Early reflection power} / \text{Late reflection power}]$ instead of $[(\text{Early reflection} + \text{Late reflection}) / \text{Late}$

reflection power]). The TIP/TP ratio, however, has a bit of an advantage for generating a curve of ratios for a given impulse response in that fewer computations are required.

Figure 31 shows the TIP/TP ratio vs. [total impulse length - tail length] for the 1200 tap finite impulse response used in the creation of the artificial data. Figure 32 shows the TIP/TP ratio vs. [total impulse length - tail length] for the real anechoic data (the actual impulse responses used are shown in Figures 18 and 20 taken from LMS simulations on real data). The real data curves assume that the anechoic impulse responses had zero amplitude past 1200 taps. In Figure 32, the dashed curve corresponds to the Low-cost narrowband set, and the solid curve corresponds to the wideband set. For all data sets, the tail power obviously reduces as its length decreases (increasing n), thus the TIP/TP ratio increases. The Low-cost set TIP/TP curve, however, rises (surprisingly) much more rapidly than the wideband set curve, indicating better coupling decay characteristics. The TIP/TP measure is only a measure of the coupling decay of a set. The TIP/TP measure is inherently normalized, and thus impulse responses of any magnitude can be compared.

FIGURE 31. TIP/TP (dB) for 1200 tap Artificial Impulse Response LWBNI200

FIGURE 32. TIP/TP (dB) for Wideband (solid), Low-cost Set (dashed)

The relevance of the TIP/TP ratio as an ERLE limit as the measurement/room noise power becomes insignificant can be derived in a similar fashion to the previous derivation for the converged ERLE limiting conditions.

It can be shown (for a white noise input) that if the power of the room/measurement noise approaches zero, the TIP/TP(n, M) is equivalent to the ERLE obtained from perfect cancellation of the first n taps of a finite M length impulse response. Starting with

$$\begin{aligned} \lim_{\sigma_v^2 \rightarrow 0} (\text{ERLE (dB)}) &= \lim_{\sigma_v^2 \rightarrow 0} \left(10 \log \frac{E[d^2(n)]}{E[e^2(n)]} \right) \\ &= \lim_{\sigma_v^2 \rightarrow 0} \left(10 \log \frac{\sigma_d^2}{\sigma_e^2} \right) \\ &\approx 10 \log \frac{\sigma_d^2}{\sigma_{tail}^2} \end{aligned} \quad (58)$$

where the estimation error $e(n)$ is defined as before:

$$e(n) = \sum_{i=0}^{\infty} w_{T/R}(i) u(n-i) - y(n) + v(n) \quad (59)$$

where

$w_{T/R}(i)$ is the terminal/room impulse response of infinite length

$u(n)$ is the reference input at time n

$y(n)$ is the adaptive filter output at time n

$v(n)$ is a sample at time n from a Gaussian random noise process

and

$$d(n) = \sum_{i=0}^{\infty} w_{T/R}(i) u(n-i) + v(n) \quad (60)$$

It can then be written, for a converged system with finite filter length N ,

$$y(n) \approx \sum_{i=0}^{N-1} w_{T/R}(i) u(n-i) \quad (61)$$

and substituting (55) into (53) with $v(n) = 0$ gives

$$e(n) \approx \sum_{i=N}^{\infty} w_{T/R}(i) u(n-i) \quad (62)$$

Next, this can be redefined as the returning signal due to the uncanceled echo,

$$tail(n) = \sum_{i=N}^{\infty} w_{T/R}(i) u(n-i) \quad (63)$$

with

$$\sigma_{tail}^2 = \sigma_e^2 \quad (64)$$

Then, to avoid confusion with the original terminal/room impulse response, the error can be redefined as

$$e(n) \approx \sum_{i=0}^{\infty} w_{Tail}(i) u(n-i) \quad (65)$$

where

$$w_{tail}(i) = 0 \text{ for } 0 \leq i \leq N-1;$$

$$w_{tail}(i) = w_{T/R}(i) \text{ for } N \leq i \leq \infty$$

At this point, the computation of the ERLE still requires an input $u(n)$ to a system. It is necessary to find under what conditions the ERLE can be calculated without use of the input.

Taking the discrete-time Fourier transforms of $e(n)$ and $d(n)$, the following is obtained:

$$E(e^{j\omega}) = W_{Tail}(e^{j\omega}) U(e^{j\omega}) \quad (66)$$

$$D(e^{j\omega}) = W_{T/R}(e^{j\omega}) U(e^{j\omega}) \quad (67)$$

Taking the energy of $E(e^{j\omega})$ gives

$$\text{Energy}[E(e^{j\omega})] = \frac{1}{2\pi} \int_{-\pi}^{\pi} |E(e^{j\omega})|^2 d\omega \quad (68)$$

Substituting (66) into (68),

$$\text{Energy}[E(e^{j\omega})] = \frac{1}{2\pi} \int_{-\pi}^{\pi} |W_{Tail}(e^{j\omega}) U(e^{j\omega})|^2 d\omega \quad (69)$$

Similarly,

$$\text{Energy}[D(e^{j\omega})] = \frac{1}{2\pi} \int_{-\pi}^{\pi} |W_{T/R}(e^{j\omega}) U(e^{j\omega})|^2 d\omega \quad (70)$$

For a random, wide-sense stationary input,

$$\text{Energy}[E(e^{j\omega})] = \frac{1}{2\pi} \int_{-\pi}^{\pi} W_{\text{Tail}}(e^{j\omega})^2 S_u(\omega) d\omega = \sigma_e^2 \quad (71)$$

and

$$\text{Energy}[D(e^{j\omega})] = \frac{1}{2\pi} \int_{-\pi}^{\pi} W_{\text{T/R}}(e^{j\omega})^2 S_u(\omega) d\omega = \sigma_d^2 \quad (72)$$

where $S_u(\omega)$ is the power spectral density of a wide sense stationary random process.

For any white noise input signal $u(n)$, $S_u(\omega)$ is a constant equal to σ_u^2 , and thus can be brought outside the integral:

$$\text{Energy}[E(e^{j\omega})] = \frac{\sigma_u^2}{2\pi} \int_{-\pi}^{\pi} W_{\text{Tail}}(e^{j\omega})^2 d\omega \quad (73)$$

$$\text{Energy}[D(e^{j\omega})] = \frac{\sigma_u^2}{2\pi} \int_{-\pi}^{\pi} W_{\text{T/R}}(e^{j\omega})^2 d\omega \quad (74)$$

Taking the ratio of the two energies (leaving the $\frac{1}{2\pi}$ factor) gives:

$$\text{D/E Energy Ratio} = \frac{\frac{1}{2\pi} \int_{-\pi}^{\pi} W_{\text{T/R}}(e^{j\omega})^2 d\omega}{\frac{1}{2\pi} \int_{-\pi}^{\pi} W_{\text{Tail}}(e^{j\omega})^2 d\omega} = \frac{\sigma_d^2}{\sigma_e^2} \quad (75)$$

Using Parseval's theorem,

$$\sum_{i=-\infty}^{\infty} w_{T/R}(i)^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} W_{T/R}(e^{j\omega})^2 d\omega \quad (76)$$

and

$$\sum_{i=-\infty}^{\infty} w_{Tail}(i)^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} W_{Tail}(e^{j\omega})^2 d\omega \quad (77)$$

For a finite impulse response $w_{T/R}(i)$ of length M , with $w_{Tail}(i)$ being a subset of this with length $M-n$, this can be rewritten as

$$\sum_{i=0}^{M-1} |w_{T/R}(i)|^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |W_{T/R}(e^{j\omega})|^2 d\omega \quad (78)$$

and

$$\sum_{i=n}^{M-1} |w_{T/R}(i)|^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |W_{Tail}(e^{j\omega})|^2 d\omega \quad (79)$$

Therefore, from (75),

$$\frac{\sum_{i=0}^{M-1} |w_{T/R}(i)|^2}{\sum_{i=n}^{M-1} |w_{T/R}(i)|^2} = \frac{\sigma_d^2}{\sigma_e^2} \quad (80)$$

Taking 10log of each side,

$$10 \log \frac{\sum_{i=0}^{M-1} w_{T/R}(i)^2}{\sum_{i=n}^{M-1} w_{T/R}(i)^2} = 10 \log \frac{\sigma_d^2}{\sigma_e^2} \quad (81)$$

giving (for a finite filter of length n and total impulse response length M)

$$\text{TIP/TP}(n,M) \text{ (dB)} = \lim_{\sigma_e^2 \rightarrow 0} (\text{ERLE (dB)}) \quad (82)$$

The TIP/TP measure now takes on increased significance as a method of predicting possible cancellation with a given impulse response and filter length. However, in using anechoic data, the TIP/TP measure remains a best-case analysis. Actual room conditions will degrade the obtainable ERLE below this limit. There is no reason, though, that the TIP/TP cannot be measured for a terminal/room combination (as long as the discrete impulse response obtained covers the predominant portion of the real impulse response).

Figure 33 illustrates how an ERLE performance criteria can be set up to ensure that a terminal's coupling decay rate (measured by the TIP/TP ratio) will allow adequate cancellation for a certain number of filter taps.

FIGURE 33. Terminal coupling decay rate as a performance requirement

5.3 ERLE vs. TIP/TP (Artificial Data)

The validity of the two derivations for limits on the obtainable ERLE were tested by plotting the ERLE obtained through LMS simulations on the artificial data (data recorded previously in Table 3) vs. filter length against the TIP/TP curve. This plot is displayed in Figure 34. There is very good agreement between the two curves up to the 1200 tap filter length (where the ERLE is limited to a value less than the $(S+N)/N$ ratio of the primary signal and the TIP/TP by definition goes to infinity), which demonstrates that:

1. The TIP/TP measure accurately predicts the amount of achievable cancellation for a given filter length when room/measurement noise has negligible impact (has less power than the power of the uncanceled tail). The prediction appears to hold even for band-limited (300-3400 Hz) white noise used for the simulations.

2. The signal+noise to noise ratio of the primary becomes the limiting factor for the obtainable ERLE as complete cancellation of the cancellable portion of the primary is approached.

5.4 ERLE vs. TIP/TP (Real Data)

Figures 35 and 36¹ show plots of the obtainable ERLE against TIP/TP for the Wideband and Low-cost sets in the anechoic chamber. Once again, there is very good agreement between the TIP/TP and obtainable ERLE curves for filter lengths much shorter than the length of the impulse response. As the length of the filter increases, the uncanceled tail takes on less significance as an asymptote is reached, beyond which point the ERLE does not improve. At first sight, this appears to be the room/measurement noise asymptote derived previously. However, measurements of the primary (signal+noise)/noise ratio show this not to be the limiting factor (see Table 5). This leaves the possibility of some other non-linearity becoming the significant limit or asymptote on the achievable ERLE.

TABLE 5. Comparison of Primary (S+N)/N Ratio to ERLE Asymptote

Terminal/Room Combination	Measured (S+N)/ N Ratio (dB)	ERLE Asymptote (approx. dB)
Wideband set, anechoic chamber	40.13	25.9
Low-cost set, anechoic chamber	23.79	18.5

1. For the purpose of graphing both the TIP/TP and ERLE curves using the same x-axis, the [total number of taps - tail length] TIP/TP x-axis measure is set equivalent to the # tap weights x-axis measure.

ERLE and TIP/TP for Artificial Stationary Data

FIGURE 34. Comparison of ERLE and TIP/TP Curves for Artificial Data

ERLE and TIP/TP for Wideband Set Stationary Anechoic Response

FIGURE 35. Comparison of ERLE and TIP/TP Curves for Wideband, Anechoic

ERLE and TIP/TP for Low-cost Set Stationary Anechoic Response

FIGURE 36. Comparison of ERLE and TIP/TP Curves for Low-cost set, Anechoic

6.0 ERLE LIMITATIONS DUE TO SYSTEM NON-LINEARITIES

Simulations in Section 5.0 pointed to the possibility that non-linear components of the primary signal other than room/measurement noise were limiting the achievable ERLE. This section proposes that the significant source of non-linearities in an acoustic cancellation system may come from the use of small, low-cost loudspeakers with large cone suspension non-linearities. Simulations on real and artificial data were performed to show the impact of changing the input level to the loudspeaker on the achievable ERLE, and possible structures to improve ERLE performance in real acoustic echo cancellation systems are given.

6.1 ERLE vs. Loudspeaker Level Changes (Real Data)

To further explore the possibility that non-linear characteristics of the loudspeaker were impacting the obtainable ERLE, an experiment was set-up where reference and primary data was recorded for four different loudspeaker output levels. In each case, the same reference file (white noise bandlimited to 300-3400 Hz) was played through the Low-cost set loudspeaker in an anechoic chamber, with the output from the set microphone recorded as the primary signal. The Low-cost set was used for these recordings as its loudspeaker was expected to have more pronounced non-linearities than the loudspeaker in the Wideband prototype set. The anechoic chamber was used to remove any possibilities that non-linearities in the room impulse response were affecting the primary data. For each recording, only the output level of the loudspeaker was changed. The recorded reference and primary data were then used as inputs to a 1200 tap LMS filter with $\mu=0.004$. The use of 1200 taps ensured the entire impulse response of the Low-cost set was being covered. The final con-

2

PM-1 3½"x4" PHOTOGRAPHIC MICROCOPY TARGET
NBS 1010a ANSI/ISO #2 EQUIVALENT

PRECISIONSM RESOLUTION TARGETS

PIONEERS IN METHYLENE BLUE TESTING SINCE 1974

MicroD

1980 COUNTY ROAD 2, HUNTSVILLE, AL 35891 USA
TEL: 256/485-7857 FAX: 256/485-7857 TUL: 814/885-1252

verged values of the ERLE for each simulation are recorded below along with the primary file's signal+noise to noise ratio. Along with the fact that the signal+noise to noise ratio was well above the obtainable ERLE in each case, the converged ERLE actually improved as the signal+noise to noise ratio of the primary degraded (as the volume of the source was decreased). A 15 dB decrease in loudspeaker levels resulted in an ERLE improvement of 9 dB. This change in ERLE performance with changes in level points again to rather severe non-linear effects in the primary file. The most likely source of these non-linearities in Low-cost set case is the loudspeaker, which is known to have poor low frequency distortion performance. Gao [48] and Olson [1] both point out that the most severe non-linear distortions occur at low frequencies due to non-linearities in the cone suspension, with the distortion increasing with input level. Olson [1] illustrates the phenomena as a soft-clipping effect, gradually limiting the acoustic output until a definite physical cone-excursion limit is reached.

TABLE 6. Final Converged ERLE vs. Loudspeaker Acoustic Output Levels

Loudspeaker Acoustic Output @ 1 m (dBSPL)	Primary (S+N)/N Ratio (dB)	Final Converged ERLE (dB)
75	51.14	14.94
70	47.96	15.50
65	41.35	20.50
61	36.55	23.50

6.2 ERLE vs. Reference Level Changes for Artificially Clipped Data

To further explore the impact of non-linearities on the achievable ERLE, artificial data files were created with conditions thought to be similar to the real recordings generated in the Section 6.1. For each set of artificial data, the same reference file was used (WBN-

R13.TIM). The artificial primary was varied for each case by changing the level of the reference level before being soft-clipped (or limited). After clipping, the convolution of the impulse response and the level of the Gaussian noise remained constant for each primary file. The clipping shape and levels were chosen somewhat arbitrarily to obtain a soft clipping shape with clipping beginning at -6 dB relative to full-scale coding with a full-scale attenuation of -3 dB. For a 16-bit two's complement coding scheme, the clipping was performed as follows:

$$y(n) = x(n) \quad (83)$$

for $|x(n)| \leq 16384$,

and

$$y(n) = \text{sgn}(x(n)) \cdot (16384 + (|x(n)| - 16384)^{0.9092}) \quad (84)$$

for $|x(n)| > 16384$.

A plot of the soft-clipping input-output relation of the soft-clipper is given in Figure 37.

Six different primary files were created with attenuation ranging from 0 to 5 dB (in 1 dB increments). The artificial primary file creation process is illustrated in Figure 38.

FIGURE 37. Input-Output Relation for Soft-Clipper

FIGURE 38. Artificial Primary File Creation Process for Clipped Data

FIGURE 39. Original Fixed Impulse Response

The 50 tap fixed impulse response used in the convolution with the clipped reference signal is displayed in Figure 39. The level of Gaussian noise was chosen to give a 40 dB signal + noise to noise ratio of the primary for 0 dB of reference attenuation, with this noise power kept constant for the other attenuation settings (gradually giving worse primary signal+noise to noise ratios).

This primary file creation process is meant to roughly model an acoustic echo cancellation system with a non-linear loudspeaker where the non-linearity occurs before linear convolution and noise addition. The attenuation of the reference level before clipping is meant to simulate changes in a loudspeaker driver volume control. The artificial clipping process has been simplified by applying it broadband as opposed to being a function of the input

frequency, and is probably considerably harsher clipping than would actually be experienced in real systems.

These six 80000 sample data sets (1 reference file, 6 different primary files) were run through a 50 tap LMS filter with $\mu=0.04$. Plots of the average tap weights (averaged over the last 5000 samples) and average tap weight error (relative to the original impulse response) are displayed in Figures 40 to 42 for the six different reference attenuation levels. The weight and weight error plots have been normalized to the zero dB reference attenuation case for ease of comparison. Final converged ERLE values (averaged over the last 5000 iterations), the norm of the averaged tap weight error vector (relative to the original impulse response) and the primary signal + noise to noise ratio are given in Table 7 for the same six cases. These results show a similar trend to the results obtained for the real data. Here, the ERLE becomes limited by the primary $(S+N)/N$ ratio only after the reference signal has been attenuated by 4-5 dB (i.e. it is no longer significantly) affected by clipping. For attenuation less than 4 dB, clipping becomes an increasingly severe limit on the converged ERLE as the input level to the clipper is increased. The weight error plots show an interesting phenomena in that for the more distorted cases (0, 1 and 2 dB of reference attenuation), the error of each tap weight is directly correlated with that tap weight's value. The tap weight error vector becomes uncorrelated from the tap weight values as the clipping becomes insignificant.

FIGURE 40. Avg Tap Weights for 50 tap LMS, $\mu=0.04$, root clipped, 0 dB ref attenuation

FIGURE 41. Avg Tap Wgt Error for 50 tap LMS, $\mu=0.04$, root clipped, 0 dB ref attenuation

FIGURE 42. Avg Tap Weights for 50 tap LMS, $\mu=0.04$, root clipped, 1 dB ref attenuation

FIGURE 43. Avg Tap Wgt Error for 50 tap LMS, $\mu=0.04$, root clipped, 1 dB ref attenuation

FIGURE 44. Avg Tap Weights for 50 tap LMS, $\mu=0.04$, root clipped, 2 dB ref attenuation

FIGURE 45. Avg Tap Wgt Error for 50 tap LMS, $\mu=0.04$, root clipped, 2 dB ref attenuation

FIGURE 46. Avg Tap Weights for 50 tap LMS, $\mu=0.04$, root clipped, 3 dB ref attenuation

FIGURE 47. Avg Tap Wgt Error for 50 tap LMS, $\mu=0.04$, root clipped, 3 dB ref attenuation

FIGURE 48. Avg Tap Weights for 50 tap LMS, $\mu=0.04$, root clipped, 4 dB ref attenuation

FIGURE 49. Avg Tap Wgt Error for 50 tap LMS, $\mu=0.04$, root clipped, 4 dB ref attenuation

FIGURE 50. Avg Tap Weights for 50 tap LMS, $\mu=0.04$, root clipped, 5 dB ref attenuation

FIGURE S1. Avg Tap Wgt Error for 50 tap LMS, $\mu=0.04$, root clipped, 5 dB ref attenuation

TABLE 7. ERLE and Tap Error Vector Norm vs. Level Changes for Artificially Clipped Data

Reference Attenuation (dB)	Primary (S+N)/N Ratio (dB)	Final Averaged ERLE (dB)	Norm of Final Averaged Tap Weight Error Vector
0	40.03	26.72	0.0139
1	39.10	29.76	0.0071
2	38.14	33.03	0.0034
3	37.16	35.41	0.0018
4	36.17	35.83	0.0014
5	35.17	35.07	0.0016

6.3 Non-Linear Compensation

Typical desk-top handsfree terminals employ low-cost loudspeakers for cost-competitive reasons. Such loudspeakers may exhibit compromised non-linear performance, generating frequency and amplitude dependent distortions. Adaptive algorithms that attempt to solve the linear Wiener-Hopf equations, such as LMS and RLS, are inherently ill-suited to cope with non-linear discrepancies between the reference input signal and the desired signal. If an assumption is made that the non-linearities occur before convolution with an unknown, linear impulse response, a model of these non-linearities can be applied to the reference input data before the adaptive algorithm. Figure 52 illustrates an acoustic echo canceller utilizing a non-linear model. The idea of using non-linear pre-compensation to improve the performance of an essentially linear system is not new. Both Aboukarr [42] and Gao and Snelgrove [48] have proposed the idea of a non-linear pre-processor, though not in the context of improving the ERLE of an echo canceller.

The non-linear model of the loudspeaker used in this setup could be made either a fixed input-output relation (difficult to make frequency dependent), or an adaptive system. Gao and Snelgrove [48] has proposed a combination of third order Volterra filters for adaptive modelling of loudspeaker non-linearities. It may be possible to use the error signal from the actual acoustic echo canceller after some initial convergence period to then adapt the Volterra filter to obtain a better converged ERLE.

FIGURE 52. Use of Non-Linear Model in Acoustic Echo Canceller Structure

As an exercise to show the validity of this structure, an LMS filter (50 tap, $\mu=0.04$) with a non-linear preprocessor was used as an echo canceller for the reference file WBN-R13.TIM and the artificially clipped primary file AWBNP525.TIM (the 0 dB attenuation file generated for Section 6.2). Several simulations were run, in which the clipping threshold characteristics of the non-linear pre-processor was changed and the final converged ERLE recorded. For each simulation, a start and end point (thresholds) for the clipping was chosen. The characteristics of the clipper were then chosen as follows. given thresh-

olds T_1 and T_2 . Threshold T_1 is the point at which the non-linearity begins, whereas threshold T_2 is the maximum output value possible with the clipper.

1. Generate required root-power for clipper (where 32768 is the absolute value of full-scale coding for 16-bit two's complement samples):

$$P = \frac{\log (T_2 - T_1)}{\log (32768 - T_1)} \quad (85)$$

2. The clipping scheme is then implemented as in Section 6.2:

$$y(n) = x(n) \quad (86)$$

for $|x(n)| \leq T_1$,

and

$$y(n) = \operatorname{sgn}(x(n)) (T_1 + (|x(n)| - T_1)^P) \quad (87)$$

for $|x(n)| > T_1$.

Table 8 shows the final converged ERLE values for different combinations of thresholds (shown as the number of dB down from full scale coding).

TABLE 8. Converged ERLE Values for Different Clipping Conditions

Threshold 1 (dB relative to full-scale coding)	Threshold 2 (dB relative to full-scale coding)	Final Converged ERLE (dB)
No clipping	No clipping	26.72
-6	-3	39.94
-6	-1.5	32.61
-6	-4.5	33.27

TABLE 8. Converged ERLE Values for Different Clipping Conditions

Threshold 1 (dB relative to full-scale coding)	Threshold 2 (dB relative to full-scale coding)	Final Converged ERLE (dB)
-7.5	-3	31.76
-4.5	-3	32.28
-4.5	-1.5	29.65
-7.5	-1.5	33.15
-7.5	-4.5	26.65

This data demonstrates that for a perfect match of the non-linear pre-processor to the non-linearity in the primary signal (in this case the -6 and -3 dB thresholds using the root-clipper), ERLE values equivalent to the theoretical maximum (40 dB) are achievable. This data also shows that some benefit from the pre-processor is still possible even for mismatches in the clipping thresholds, with the improvement in ERLE decreasing as the pre-processor model moves away from the characteristics of the actual non-linearity (leading to eventual degradation of the ERLE relative to cancellation without the pre-processor).

These simulations do not show the effect of changing the base model of the non-linear pre-processor. Simulations performed with the root-clipper on real data (not shown here) showed no improvement in ERLE with the pre-processor, showing that some knowledge of the base characteristics of the non-linearity to be compensated for is essential. Thus, an adaptive approach to 'learn' the characteristics of the non-linearity appears promising.

An alternative to the structure of Figure 52 would place an inverse model of the loudspeaker transfer function before the loudspeaker but not in the path of the reference input to the adaptive filter. This structure is displayed in Figure 53. Although this structure would have the benefit of improving loudspeaker performance for the listener, at the same

time any transient artifacts generated while adapting to the non-linearity would be heard by the user.

FIGURE 53. Alternative Structure for use of Non-Linear Loudspeaker Model

6.4 Significance of Loudspeaker Non-linearities

A significant amount of previous research on acoustic echo cancellation has sought ways in which to improve the converged ERLE of a system by finding ways in which to cancel a larger portion of the echo impulse response (e.g. [17], [18], [22], [38], [39], [40], [41], and [55]). From the results in this thesis, it can be seen that depending on the level of non-linearities present in an acoustic echo cancellation system, attempting to cancel a larger

portion of the terminal/room impulse response may actually show no improvement in the converged ERLE. This then becomes a system design exercise - what number of taps and what level of loudspeaker performance is needed to make optimum use of computing resources and to provide a desired level of cancellation.

7.0 DYNAMIC ACOUSTIC ECHO CANCELLATION PERFORMANCE

Previous sections of this thesis have focused on steady-state acoustic echo cancellation performance limitations (after convergence of the adaptive filter). Although the LMS algorithm was used for all previous simulations, it was used mainly as a tool to highlight performance limitations caused by the physical environment in which the adaptive filter operated. Its own operation and characteristics as such were not under direct examination (except for Section 3.2.1 on convergence speed which was primarily meant to determine a useful range of LMS adaptation step-sizes).

This section, however, begins to look at algorithm specific performance as it relates to acoustic echo cancellation. The two primary performance benchmarks this thesis examines are convergence speed (the time required to reach a steady-state mean-squared error variance from algorithm initialization) and tracking ability (the ability to maintain a constant mean-squared error variance while changes are occurring in the impulse response to be modelled). Although these two criteria appear to be very similar, it is not necessarily the case that good performance for one implies good performance for the other [49]. For acoustic echo cancellation performances, the dynamic nature of conversation and body movement (non-stationarities) may make tracking performance a more important consideration than convergence speed (as long as the algorithm converges to some required ERLE early enough in the conversation or during some pre-conversation training period). This section will begin with a look at convergence speed considerations, but will eventually move to a discussion of tracking ability. This section will also look at converged tap-weight accuracy - what final converged ERLE is achievable by each algorithm.

7.1 Algorithm Performance (Stationary IR, Noise Input)

This section looks at fundamental convergence speed properties and the final depth of cancellation (ERLE) of several algorithms: LMS, NLMS, RLS, FTF and SFTF. These algorithms are discussed briefly in Section 2.2 (with the FTF algorithm derived in Appendix B). Of particular interest is a comparison between the LMS (low complexity) and RLS (higher complexity) families of algorithms. Although the noise input and stationary impulse response conditions used here are not truly indicative of real acoustic cancellation environments, it is a useful exercise to compare these algorithms under these idealized conditions.

7.1.1 Stationary, Artificial Data, Noise Input

As a starting point for examining the performance of the LMS and RLS algorithm families, stationary data was created artificially. This process required a previously recorded band-limited (300-3400 Hz) reference channel noise file, which was then convolved with a fixed finite length impulse response (shown in Figure 39), to which a stationary Gaussian noise process was added to create an artificial primary file. This process is the same as that shown in Figure 38. The additive noise power was chosen to give a primary file (Signal+Noise)/Noise ratio of 40 dB. Several algorithms were then compared using this reference channel file and artificially generated primary channel file. Adaptation (using 50 tap filters in each case) was allowed to continue for the duration of the input files (10 seconds). For each algorithm, the final averaged tap weights, final averaged tap weight error, the norm of the final tap weight error vector, short-term averaged ERLE curve and final ERLE value are presented. The 10 second period for which adaptation continued is what is referred to in the Figure titles as '10 second Convergence'. The norm of the final tap

weight error vector and ERLE values are given in Table 9. Tap weights were averaged over the last 5000 samples (0.625 seconds) of the input files before plotting. The tap weight error vector was found by subtracting the final averaged tap weight vector of the adaptive filter from the original tap weight vector. The norm of the final tap weight error vector was used as a simple measure of 'closeness' to the original impulse response. The final ERLE result was obtained by averaging the ERLE over the last 5000 samples. The short-term averaged ERLE curve, however, was obtained by using a moving average window of width 256 over the length of the data.

TABLE 9. Comparison of Tap Weight Error Vector Norm and Final ERLE, Artificial Data

Algorithm	Norm of Final Averaged Tap Weight Error Vector	Final Averaged ERLE (dB)
LMS ($\mu=0.04$)	8.8416×10^{-4}	39.93
NLMS ($\mu=1/\text{power}$)	0.0020	37.13
RLS ($\lambda=1$)	8.2879×10^{-4}	40.20
FTF ($\lambda=1, \mu=1$)	8.3931×10^{-4}	40.20
SFTF ($\lambda=1, \mu=1$)	8.3931×10^{-4}	40.20

TABLE 10. Algorithm Convergence Speed, Artificial Data

Algorithm	Approx. Convergence Time (seconds)
LMS ($\mu=0.04$)	0.45
NLMS ($\mu=1/\text{power}$)	0.09
RLS ($\lambda=1$)	0.04
FTF ($\lambda=1, \mu=1$)	0.4
SFTF ($\lambda=1, \mu=1$)	0.4

Table 9 shows the various RLS implementations (RLS, FTF and SFTF) to be virtually identical (which they should be) in terms of final achieved ERLE. This table also shows the RLS algorithms to be superior in terms of tap weight accuracy relative to LMS and NLMS under these idealized conditions, but in the case of LMS this difference is insignificant. Based on previous discussions (see Section 5.1) the final ERLE should have been very close to 40 dB given the primary signal + noise to noise ratio was 40.03 dB (averaged over the entire primary file).

Table 10 shows the approximate convergence speeds for the different algorithms using the stationary, artificial data. The choice of μ for LMS was arbitrary and given the constant power of the input signal, could have been set to give equivalent convergence speed performance to NLMS. RLS demonstrates the best convergence speed, which theoretically should have been equalled by the FTF and SFTF algorithms. The need for a soft-constraint initialization method which weighted the biased zero initial conditions and the fact that the effect of these conditions did not exponentially decay (infinite memory with $\lambda=1$) gave these two algorithms very slow convergence speeds. It is interesting to note that with $\lambda=1$, the FTF algorithm was less susceptible to finite precision effects (as noted in [9]) and did not fail during the course of the simulation data.

FIGURE 54. Averaged Tap Weights for 50 tap LMS, $\mu=0.04$, 10 second Convergence

FIGURE 55. Averaged Tap Weight Error for 50 tap LMS, $\mu=0.04$, 10 second Convergence

FIGURE 56. ERLE for 50 tap LMS, $\mu=0.04$

FIGURE 57. Averaged Tap Weights for 50 tap NLMS, $\mu=1/\text{power}$, 10 second Convergence

FIGURE 58. Averaged Tap Weight Error for 50 tap NLMS, $\mu=1/\text{power}$, 10 second Convergence

FIGURE 59. ERLE for 50 tap NLMS, $\mu=1/\text{power}$

FIGURE 60. Averaged Tap Weights for 50 tap RLS, $\lambda=1$, 10 second Convergence

FIGURE 61. Averaged Tap Weight Error for 50 tap RLS, $\lambda=1$, 10 second Convergence

FIGURE 62. ERLE for 50 tap RLS, $\lambda=1$

FIGURE 63. Averaged Tap Weights for 50 tap FTF, $\lambda=1$, $\mu=1$, 10 second Convergnece

FIGURE 64. Averaged Tap Weight Error for 50 tap FTF, $\lambda=1$, $\mu=1$, 10 second Convergnece

FIGURE 65. ERLE for 50 tap FTF, $\lambda=1, \mu=1$

FIGURE 66. Averaged Tap Weights for 50 tap SFTF, $\lambda=1, \mu=1$, 10 second Convergece

FIGURE 67. Averaged Tap Weight Error for 50 tap SFTF, $\lambda=1$, $\mu=1$, 10 second Convergece

FIGURE 68. ERLE for 50 tap SFTF, $\lambda=1$, $\mu=1$

7.1.2 Stationary, Real Data, Noise Input

The next set of simulations looked at the accuracy and convergence speed of several algorithms under actual recorded data conditions (noise input). The computational storage (and time) requirements restricted the RLS algorithm to 50 taps (as in the artificial data case). For real data, this prompted the use of data with the shortest impulse response so as to give the least amount of undermodelling. The stationary anechoic chamber data for the Low-cost set (NBNR25.TIM and NBNP25.TIM) was chosen, which had an impulse response length of about 300 taps (see Figure 20). Simulations were run for LMS (two different step sizes), NLMS, RLS, FTF, SFTF and SFTF with a priori estimation error output (instead of a posteriori). The two versions of SFTF were run to ensure the original SFTF was operating correctly (SFTF tended to drive its a posteriori error low if inappropriate values for μ or λ were chosen, which gave the appearance of better performance than was actually occurring). As in Section 7.1.1, adaptation was allowed to continue for the duration of the input files (10 seconds). Since any real conditions could not be guaranteed to be completely stationary, a forgetting factor less than 1 was chosen for the RLS type algorithms ($\lambda=0.9995$). For each algorithm, the final averaged tap weights, final averaged tap weight vector, the norm of the final tap weight error vector, short term averaged ERLE curve and final ERLE value were computed. The norm of the final tap weight error vector and ERLE values are presented in Table 11. Tap weights were averaged over the last 5000 samples (0.625 seconds) of the input files before plotting. For the real data (as opposed to the artificial data) there was no reference tap weight vector to which the averaged tap weight vectors could be compared. They were instead compared to the RLS averaged tap weight vector (mainly to see how close the FTF and SFTF solutions came to the RLS solution) by subtracting the final averaged tap weight vector of the adaptive filter in question

from the RLS averaged tap weight vector. The norm of this error vector was then used as a measure of 'closeness' to the RLS solution. The final ERLE result was obtained by averaging the ERLE over the last 5000 samples. The short-term averaged ERLE curve, however, was obtained by using a moving-average window of width 256 over the length of the data.

TABLE 11. Comparison of Tap Weight Error Vector Norm and Final ERLE, Real Data

Algorithm	Norm of Final Averaged Tap Weight Error Vector (relative to RLS)	Final Averaged ERLE (dB)
LMS ($\mu=0.004$)	0.0700	13.87
LMS ($\mu=0.04$)	0.0236	13.71
NLMS	0.0526	11.13
RLS	Reference	13.88
FTF	0.2522	numerically unstable
SFTF	4.6108×10^{-6}	14.09
SFTF (a priori error output)	4.6108×10^{-6}	13.88

TABLE 12. Algorithm Convergence Speed, Real Data

Algorithm	Approx. Convergence Time (seconds)
LMS ($\mu=0.004$)	2
LMS ($\mu=0.04$)	0.3
NLMS ($\mu=1/\text{power}$)	0.15
RLS ($\lambda=0.9995$)	0.15
FTF ($\lambda=0.9995, \mu=1$)	0.15
SFTF ($\lambda=0.9995, \mu=1$)	0.15
SFTF (a priori error output) ($\lambda=0.9995, \mu=1$)	0.15

Table 11 shows the SFTF and RLS solutions to be virtually identical. The final averaged ERLE for the a posteriori error output SFTF was slightly better than the RLS or a priori estimation error output SFTF. The FTF algorithm proved to be numerically unstable under these simulation conditions. Once again, the RLS solutions (RLS or SFTF), were just marginally better in terms of final converged ERLE than the best LMS solution ($\mu=0.004$).

Table 12 shows the approximate convergence speeds for the algorithms tested. For these simulation conditions, there appears to be no advantage in convergence speed in going to the computationally more complex RLS type algorithms from the LMS type algorithms except for the slight improvement in final converged ERLE.

It is interesting to note here that given a forgetting factor less than 1, the weighting of initial conditions in the soft-constraint initialization method for the SFTF algorithm did not affect the convergence speed as it did for the artificial data case where infinite memory was used ($\lambda=1$).

FIGURE 69. Averaged Tap Weights for 50 tap LMS, $\mu=0.004$, 10 second Convergence

FIGURE 70. Averaged Tap Weights for 50 tap LMS, $\mu=0.04$, 10 second Convergence

FIGURE 71. Averaged Tap Weight Error (rel. to RLS) for 50 tap LMS, $\mu=0.004$

FIGURE 72. Averaged Tap Weight Error (rel. to RLS) for 50 tap LMS, $\mu=0.04$

FIGURE 73. ERLE for 50 tap LMS, $\mu=0.004$

FIGURE 74. ERLE for 50 tap LMS, $\mu=0.04$

FIGURE 75. Averaged Tap Weights for 50 tap NLMS, $\mu=1/\text{power}$, 10 second Convergence

FIGURE 76. Averaged Tap Weight Error (rel. to RLS) for 50 tap NLMS, $\mu=1/\text{power}$

FIGURE 77. ERLE for 50 tap NLMS, $\mu=1/\text{power}$

FIGURE 78. Averaged Tap Weights for 50 tap RLS, $\lambda=0.9995$, 10 second Convergence

FIGURE 79. ERLE for 50 tap RLS, $\lambda=0.9995$

FIGURE 80. Averaged Tap Weights for 50 tap FTF, $\lambda=0.9995$, $\mu=1$, 10 second Convergence

FIGURE 81. Averaged Tap Weight Error (rel. to RLS) for 50 tap FTF, $\lambda=0.9995$, $\mu=1$

FIGURE 82. ERLE for 50 tap FTF, $\lambda=0.9995, \mu=1$

FIGURE 83. Averaged Tap Weights for 50 tap SFTF, $\lambda=0.9995, \mu=1$, 10 second Convergence

FIGURE 84. Averaged Tap Weight Error (rel. to RLS) for 50 tap SFTF, $\lambda=0.9995$, $\mu=1$

FIGURE 85. ERLE for 50 tap SFTF, $\lambda=0.9995$, $\mu=1$

FIGURE 86. Averaged Tap Weights for 50 tap SFTFAP, $\lambda=0.9995$, $\mu=1$, 10 second Convergence

FIGURE 87. Averaged Tap Weight Error (rel. to RLS) for 50 tap SFTFAP, $\lambda=0.9995$, $\mu=1$

FIGURE 88. ERLE for 50 tap SFTFAP, $\lambda=0.9995$, $\mu=1$

7.1.3 Stationary, Real Data, Speech Input

This set of simulations looked at the convergence performance of various algorithms (reduced now to LMS (two different step sizes), NLMS (two different step-size constants), and SFTF (two different initialization weighting constants) for stationary conditions with speech input. Real data from the same recording conditions as Section 7.1.2 (anechoic chamber, Low-cost set) was used except for the use of a male speech signal as input (files NBSR23.TIM and NBSP23.TIM). The short-term averaged ERLE curve (for a moving window of 256 samples) is displayed for the different algorithms in Figures 89 to 94.

These curves show the NLMS algorithm to give the best performance under these condi-

tions, although this algorithm should have been coupled with a double-talk detector to stop adaptation during no-speech periods between words (where the ERLE jumped to artificially high values). The surprise in this set of simulations is the relatively poor performance of the SFTF algorithm, which should have performed better than NLMS for speech inputs (SFTF should not have been as sensitive to the eigenvalue spread of speech). The SFTF algorithm may have been more sensitive to non-linearities in the recordings, and may have needed a smaller forgetting factor. The SFTF algorithm forgetting factor, however, is bounded for reasons of stability between 1 and $1-(1/2N)$, where N is the order of the filter [43]. For a 50 tap filter, this works out to a bound between 0.99 and 1, although from experience in simulations performed for this thesis the forgetting factor should not be set too close to the lower bound (the a posteriori output error becomes unrealistically low).

The SFTF simulations also showed the initialization weight to have a fairly significant impact on convergence speed, where the larger the weight used, the longer it took for the SFTF algorithm to ‘purge’ the memory of zero initial conditions and the corresponding sub-optimal tap-weights.

FIGURE 89. ERLE for 50 tap LMS ($\mu=0.04$), Male Speech input

FIGURE 90. ERLE for 50 tap LMS ($\mu=0.004$), Male Speech input

FIGURE 91. ERLE for 50 tap NLMS ($\mu=1/\text{power}$), Male Speech input

FIGURE 92. ERLE for 50 tap NLMS ($\mu=0.5/\text{power}$), Male Speech input

FIGURE 93. ERLE for 50 tap SFTF ($\lambda=0.9995, \mu=1$), Male Speech input

FIGURE 94. ERLE for 50 tap SFTF ($\lambda=0.9995, \mu=10$), Male Speech input

7.2 Convergence Properties Using Pre-whitening Structures

Several pre-whitening structures were implemented during the course of this research, including structures by Yasukawa, Shimada, and Furukawa [21], Hansler [40], or Mboop, Bonnet and Bershad [59], with the hope that the convergence speed of the NLMS algorithm could be improved for speech inputs. However, for simulations performed with a male speech input, none of these structures appeared to give any convergence improvement over NLMS, and for the brevity of this thesis, these simulation results have been omitted.

7.3 Tracking Improvements for LMS

It has been shown (van de Kerkhof and Kitzen [61]) that the LMS algorithm is unable to successfully track (maintain a constant ERLE for) typical changes in a room impulse response due to body movement (i.e. walking through the room). In such instances, the ERLE typically degrades until the movement stops and the LMS adaptive filter can once again converge to a stable impulse response.

For best performance under non-stationary conditions, it would be desirable to increase the adaptation step-size during these conditions to speed convergence towards the desired impulse response and improve tracking. Sikorav [27] proposed an alteration to the standard LMS algorithm in which three different adaptation step sizes were used for three different portions of the impulse response. Sikorav proposed that most movement of coefficients in the room impulse response due to body movement would occur in a particular time window or portion of the impulse response to be cancelled. For this window, a larger adaptation step-size would be used to improve tracking. This model, however, does

not take into account the possibility of movement outside this impulse response window nor does it address the poorer performance obtained with a large step size if the impulse response within that window remains stationary.

An alternative way of addressing the tracking problem is to assume that if the ERLE has become degraded (the mean-squared error has increased), some room non-stationarity is occurring and a larger step-size is needed to improve tracking. It has been known for some time that the adaptation step-size should itself be made adaptive (Smalley [3]). Kwong and Johnston [62] have developed an algorithm that increases the LMS adaptation step-size during periods of large mean-squared error and decreases it during periods of small mean-squared error. This adaptive adaptation step-size scheme is very simple and requires the equivalent of an additional tap update within the context of a much larger adaptive filter. The algorithm was discussed in Section 2.5 but is rewritten here for convenience:

$$\mu'(n+1) = \alpha\mu(n) + \gamma e^2(n) \quad (88)$$

where

$$0 < \alpha < 1, \quad \gamma > 0 \quad (89)$$

and

$$\mu(n+1) = \begin{cases} \mu_{\max}, & \mu'(n+1) > \mu_{\max} \\ \mu_{\min}, & \mu'(n+1) < \mu_{\min} \\ \mu'(n+1), & \text{otherwise} \end{cases} \quad (90)$$

To compare the tracking ability of this algorithm (abbreviated as LMS-VSS for LMS with Variable Step-Size) with LMS, NLMS and SFTF, acoustic echo cancellation simulations

using files where a significant acoustic non-stationarity occurred were used. These files, WBNR14.TIM and WBNP14.TIM, used the Wideband prototype set in the DPN-A conference room with a bandlimited noise input. During the recording, a hand was waved directly in front of the set every 5 seconds. For these files, the non-stationarity occurred twice, between 1.5 and 3.5 seconds into the file, and between 7 and 9 seconds into the file.

The LMS-VSS algorithm was given the following characteristics:

$$\alpha = 0.97$$

$$\gamma = 0.48$$

$$\mu_{\min} = 0.004$$

$$\mu_{\max} = 0.04$$

Note that the value for γ is two-orders of magnitude larger than suggested in [62]. It was necessary to increase this value to obtain reasonable adaptation of the step-size value for the given data.

Short-term averaged ERLE curves (256 point moving average window) for the different algorithms are displayed in Figures 95 to 101. Plots of adaptation step-size versus time are given for the NLMS and LMS-VSS simulations in Figures 102 and 103.

Comparing the tracking ability during the 1.5-3 second and 7-9 second periods with the other algorithms, the LMS-VSS algorithm gives better tracking performance (approx. 5 dB) than the SFTF algorithms or the fixed step-size LMS algorithms (for $\mu=0.004$ and $\mu=0.01$), while it gives equivalent tracking performance to the NLMS algorithm and LMS

with a large step-size ($\mu=0.04$). ERLE performance of LMS-VSS during stationary periods (between 3.5 and 7 seconds) exceeded that of NLMS or the fixed LMS (with $\mu=0.04$) by 2.5 dB, while achieving an equivalent ERLE to the fixed LMS with $\mu=0.004$. The limits on μ for LMS-VSS were purposely set at 0.004 and 0.04 to ensure that better performance was not being achieved simply because the step-size was allowed to range further than the μ 's set for the fixed step-size LMS cases. Therefore, for this white noise input case, LMS-VSS outperformed NLMS or any fixed LMS with μ inside the range covered by LMS-VSS. It is interesting to note that even with the best tracking shown here, the dynamic ERLE performance still shows a 5-10 dB drop under these non-stationary conditions. This correlates well with findings by van de Kerkhof and Kitzen [61].

A note on SFTF: Good tracking performance with RLS type algorithms can only be achieved by setting the forgetting factor below 1. The degree with which the forgetting factor in the SFTF algorithm can set below 1 without causing numerical problems is dependant on the filter size. With a filter size of 600 and $\lambda=0.9995$ (not far below 1!), there is a large discrepancy between the steady-state ERLE (between 3.5 to 7 seconds) of the a posteriori estimation error output SFTF algorithm (SFTF) and the a priori estimation error output SFTF algorithm (SFTFAP). From experience in other SFTF simulations, it is likely that this discrepancy would diminish if λ were moved closer to 1, and a more reasonable a posteriori estimation error would be generated. Increasing λ , however, would further degrade already poor tracking ability. To this end, the SFTF algorithm appears unsuitable for large order filtering applications where good tracking ability is required. Eleftheriou and Falconer [19] also concluded that the LMS algorithm could always be set to have equivalent or superior tracking performance to an RLS algorithm.

FIGURE 95. ERLE for 600 tap SFTF, $\lambda=0.9995$, $\mu=1$, Non-stationary IR conditions

FIGURE 96. ERLE for 600 tap SFTFAP, $\lambda=0.9995$, $\mu=1$, Non-stationary IR conditions

FIGURE 97. ERLE for 600 tap LMS, $\mu=0.004$, Non-stationary IR conditions

FIGURE 98. ERLE for 600 tap LMS, $\mu=0.01$, Non-stationary IR conditions

FIGURE 99. ERLE for 600 tap LMS, $\mu=0.04$, Non-stationary IR conditions

FIGURE 100. ERLE for 600 tap NLMS, $\mu=1/\text{power}$, Non-stationary IR conditions

FIGURE 101. ERLE for 600 tap LMS-VSS, $\alpha=0.97, \gamma=0.48, \mu_{\max}=0.04, \mu_{\min}=0.004$

FIGURE 102. Step-size vs. time for 600-tap NLMS under Non-stationary conditions

FIGURE 103. Step-size vs. time for 600 tap LMS-VSS under Non-stationary conditions

7.4 Tracking Improvements for Speech Input Conditions

Although the case of tracking room non-stationarities occurring during speech input was not fully examined, preliminary simulations showed that it was not possible to give the LMS-VSS algorithm a sufficient range of μ values to match the convergence speed and tracking ability of NLMS with speech inputs without causing instability. Thus, for speech inputs, it is likely that LMS-VSS will continue to outperform any fixed- μ LMS algorithms having a μ within the range given to the LMS-VSS algorithm, but will probably not match NLMS performance. For systems that cannot afford the additional computational burden of the NLMS algorithm, this may still provide a significant improvement in performance over fixed- μ LMS.

It may also be possible to adapt the LMS-VSS algorithm for use with NLMS, where the step-size constant $\bar{\mu}$ (see Equation (8) in Section 2.2) would be adapted to give NLMS better steady-state ERLE performance. This is discussed in the conclusions as a possible area of further research.

7.5 Discussion of Convergence and Tracking Results

From the simulations performed in this section, it has been shown that for large filter orders and non-stationary environments, LMS type algorithms will give better overall performance than RLS type algorithms. Certainly, for the large increase in computational costs for even fast-RLS implementations such as the SFTF, there can be no current justification for their use over LMS type algorithms in acoustic echo cancellation applications. NLMS was shown to provide best convergence for speech inputs, at the expense of increased computational requirements. For tracking ability under white noise input condi-

tions, a variant of the LMS algorithm with an adaptive variable step-size was shown to give better overall convergence speed, tracking ability and steady-state ERLE performance than fixed LMS or NLMS. Although the case of room non-stationarities occurring during speech input was not fully examined, some preliminary simulations showed that the variable LMS algorithm would still outperform fixed- μ LMS algorithms but not the NLMS algorithm.

8.0 CONCLUSIONS

8.1 Discussion of Results

The results presented in this thesis show the impact several environmental conditions have on the achievable cancellation (ERLE) in an acoustic echo cancellation system. The results also show the impact a non-stationarity in the room environment has on canceller performance. An acoustic echo canceller structure to improve canceller performance for non-linear loudspeakers has been proposed. The use of an LMS with adaptive step-size algorithm has been proposed to improve tracking performance during non-stationary periods.

Room noise and system non-linearities have been shown to impose an asymptote on the achievable ERLE as the number of taps in the adaptive filter is increased. With room noise as the asymptote, the maximum achievable ERLE is equivalent to the signal+noise to noise ratio of the primary signal (by the definition of the ERLE). For low-cost handsfree systems, it has been shown that system non-linearities (such as loudspeaker distortion) may have a much larger impact on the achievable ERLE than typical room noise. For short adaptive filter lengths, however, the uncanceled echoes remaining in the error signal may take precedence as the limit on achievable ERLE. For white noise inputs, this ERLE limit was shown to be measured directly by the Total Impulse response Power to Tail Power (TIP/TP) ratio.

To mitigate the effects of non-linearities in the loudspeaker transfer function, a structure has been proposed which would pre-distort the reference input to the adaptive filter based on a model of the loudspeaker transfer function, or alternatively linearize the loudspeaker

by inserting an inverse model of the loudspeaker's non-linear response before the loudspeaker. The former structure was used to show that given an exact model of the non-linearity, the theoretical limit for achievable ERLE could be attained using this structure.

To improve tracking under non-stationary room impulse response conditions, application of an adaptive step-size LMS algorithm was proposed. This algorithm, which increased the adaptation step-size when the squared output error term from the canceller was large and reduced the step-size when the squared error term was small, was shown to offer better cancellation performance overall than either a fixed step-size LMS or NLMS. Relative to fixed step-size LMS with the step size optimized for good cancellation under stationary conditions (small μ), the adaptive step-size LMS showed an approximate 5 dB improvement in tracking of severe non-stationarities (hand waving directly in front of the hands-free terminal) and showed equivalent ERLE performance during stationary periods. Relative to fixed step-size LMS with the step size optimized for tracking (larger μ) or NLMS, the adaptive step-size LMS showed equivalent tracking capability but approximately 2.5 dB improvement in ERLE performance during stationary periods.

The use of the Stabilized Fast Transversal Filter (SFTF), a computationally reduced implementation of the Recursive Least Squares (RLS) algorithm, was examined for acoustic echo canceller applications. It was found that the effects of the soft-constraint initialization method and restrictions on the allowable forgetting factor gave the algorithm poorer convergence speed and tracking abilities than simpler algorithms such as NLMS.

8.2 Future Research

Results from this thesis have pointed to several areas where benefit could be obtained from further research:

8.2.1 Incorporation of Adaptive Non-Linear Filter in Pre-Processor Structure

Further research could be done on incorporating Gao and Snelgrove's [48] adaptive Volterra filter for loudspeaker linearization into the acoustic echo cancellation structures proposed in this thesis. These structures would apply knowledge of the loudspeaker's non-linear characteristics gathered by the adaptive non-linear filter to provide better ERLE. The major obstacle in this work would be finding a suitable output signal from the loudspeaker for adapting the non-linear filter. Ideally, a microphone placed in front of the loudspeaker in anechoic conditions would be used to provide the adaptive non-linear filter with the distorted loudspeaker output. This would not be feasible with a desktop terminal, although it may be possible to provide the adaptive non-linear filter with the error signal from the linear acoustic echo canceller after some initial period of its convergence, after which the adaptive non-linear filter would attempt to improve the ERLE.

8.2.2 Applicability of Cancellation Limits with Sub-band Structures

An exploration of the cancellation limits due to number of adaptive filter taps could be examined for sub-band acoustic echo cancellers. This would involve a comparison of achievable ERLE for a full-band canceller with a given number of taps to the achievable ERLE obtained for a sub-band canceller with the same number of taps spread across several frequency bands.

8.2.3 Double-Talk Detection

An important facet of an acoustic echo cancellation system that was not discussed in this thesis is the double-talk detector. This detector is required to halt adaptation during periods of significant input (speech or noise) from the microphone uncorrelated to the reference input that would cause divergence of the tap weights. Typically, double-talk conditions are identified by a sudden drop in the ERLE (or sudden rise in the mean-squared error output). Currently, little emphasis has been placed on this aspect of acoustic echo cancellation in literature, but it remains a crucial part of any real time system. It would be useful to determine the fastest and most reliable methods of detecting double-talk conditions (when adaptation should be halted), and finding ways of distinguishing these conditions from non-stationary conditions (when adaptation should continue).

8.2.4 Variable Gain NLMS

The NLMS algorithm has been shown to provide excellent tracking ability and convergence speed (especially for speech inputs). Its converged mean-squared error performance for stationary environments is generally not as good as an optimized fixed- μ LMS algorithm, however. It may be possible to incorporate an algorithm similar to the variable step-size LMS [62] in the context of the NLMS algorithm, where under conditions of low mean-squared error output the magnitude of change in the already variable step-size of the NLMS algorithm would be attenuated by an adaptive gain factor. Similarly, under high mean-squared error conditions (such as during a room non-stationarity), this gain could be adaptively increased.

8.3 Summary of Contributions

- 1. Derivation of basic acoustic echo cancellation limits based on terminal/room impulse responses, room noise and system non-linearities.**
- 2. Non-linear pre-processing architecture in reference path for compensating non-linear effects of loudspeakers.**
- 3. Application of LMS with adaptive step-size algorithm for improved tracking in acoustic environments.**
- 4. High quality two-channel recordings of terminal/room impulse responses.**
- 5. Application of SFTF for acoustic echo cancellation.**

9.0 REFERENCES

- [1] Olson, H. F., *Elements of Acoustical Engineering*, 2nd Ed., D. Van Nostrand Co., Toronto, 1947.
- [2] Beranek, L. L., *Acoustics*, McGraw-Hill, Toronto, 1954.
- [3] Smalley, A. R., "Time Varying Applications of Practical Adaptive Systems Using Gradient Algorithms", PhD Thesis, Carleton University, Sept., 1973.
- [4] Cavanaugh, J. R., R. W. Hatch and J. L. Sullivan, "Models for the Subjective Effects of Loss, Noise, and Talker Echo on Telephone Connections", *Bell Systems Technical Journal*, Nov., 1976.
- [5] L. Ljung, M. Morf and D. Falconer, "Fast calculation of gain matrices for recursive estimation schemes", *Int. J. Control.*, vol. 27, pp. 1-19, Jan. 1978.
- [6] Falconer, D. and L. Ljung, "Application of fast-Kalman estimation to adaptive equalization", *IEEE Trans. Commun.*, vol. COMM-26, pp. 1439-1446, Oct. 1978.
- [7] Neely, S. T. and J. B. Allen, "Invertibility of a Room Impulse Response", *Journal of the Acoustical Society of America*, 66(1), pp. 165-169, July 1979.
- [8] Carayannis, G., D. Manolakis, and N. Kalouptsidis, "A fast sequential algorithm for least squares filtering and prediction", *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-31, pp. , Dec. 1983.
- [9] Cioffi, J. M. and T. Kailath, "Fast, Recursive-Least-Squares Transversal Filters for Adaptive Filtering", *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-32, pp. 304-338, April 1984.
- [10] Papoulis, A., *Probability, Random Variables, and Stochastic Processes*, 2nd ed., McGraw-Hill, Toronto, 1984.
- [11] Mourjopoulos, J., "On the Variation and Invertibility of Room Impulse Response Functions", *Journal of Sound and Vibration*, vol. 102, no. 2, pp. 217-228, 1985.
- [12] Kellerman, W., "Kompensation Akustischer Echos in Frequenzteilbandern", *Frequenz*, vol. 39, no. 7/8, 1985, pp. 209-215.
- [13] Widrow, B. and S. D. Stearns, *Adaptive Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, 1985.
- [14] Muron., O. and J. Sikorav, "Modelling of Reverberators and Audioconference Rooms", *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Tokyo, Japan, April, 1986, pp. 921-924.
- [15] Yasukawa, H., M. Nishino, K. Ishimaru and H. Murakami, "Echo Return Loss Required for Acoustic Echo Canceller Based on Subjective Assessment", *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Tokyo, Japan, April, 1986, pp. 1309-1312.

- [16] Madiseti, V. K., D. G. Messerschmitt and N. Nordstrom, "Dynamically-Reduced Complexity Implementation of Echo Cancellers", *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Tokyo, Japan, April, 1986, pp. 1313-1316.
- [17] Casar-Corredera, J. R. and J. Alcazar-Fernandez, "An Acoustic Echo Canceller for Teleconference Systems", *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Tokyo, Japan, April, 1986, pp. 1317-1320.
- [18] Kawamura, S. and M. Hatori, "A Tap Selection Algorithm for Adaptive Filters", *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Tokyo, Japan, April, 1986, pp. 2979-2982.
- [19] Eleftheriou, E. and D. D. Falconer, "Tracking Properties and Steady State Performance of RLS Adaptive Filter Algorithms", *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 1097-1110, Oct. 1986.
- [20] Eriksson, L. J., M. C. Allie and R. A. Greiner, "The Selection and Application of an IIR Adaptive Filter for Use in Active Sound Attenuation", *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, pp. 433-437, April 1987.
- [21] Yasukawa, H., S. Shimada and I. Furukawa, "Acoustic Echo Canceller with High Speech Quality", *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Dallas, Texas, April, 1987, pp. 2125-2128.
- [22] Gilloire, A., "Experiments with Sub-band Acoustic Echo Cancellers for Teleconferencing", *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Dallas, Texas, April, 1987, pp. 2141-2144.
- [23] Miyoshi, M. and Y. Kaneda, "Inverse Filtering of Room Acoustics", *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-36, pp. 145-152, Feb. 1988.
- [24] Toplis, B. and S. Pasupathy, "Tracking Improvements in Fast RLS Algorithms Using a Variable Forgetting Factor", *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-36, pp. 206-227, Feb. 1988.
- [25] Ardalan, S. H. and L. J. Faber, "A Fast ARMA Transversal RLS Filter Algorithm", *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-36, pp. 349-358, March 1988.
- [26] Chen, J. and J. Vandewalle, "A μ -Vector LMS Adaptive System for Enhancing Non-stationary Narrow-band Signal", *Proc. IEEE Int. Symp. Circuits and Systems*, Espoo, Finland, June, 1988, pp. 771-774.
- [27] Sikorav, J., "Experiments to Identify and Track Non Stationarities In Audio Conference Rooms", *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, New York, NY, April, 1988, pp. 2566-2569.
- [28] Kellerman, W., "Analysis and Design of Multirate Systems For Cancellation of Acoustical Echoes", *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, New York, NY, April, 1988, pp. 2570-2573.

- [29] Fan, H. and W. K. Jenkins, "An Investigation of an Adaptive IIR Echo Canceller: Advantages and Problems", *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-36, pp. 1819-1833, Dec. 1988.
- [30] Benson, K. B. (editor), *Audio Engineering Handbook*, McGraw-Hill, Toronto, 1988.
- [31] Rossi, M., *Acoustics and Electroacoustics*, Artech House, Norwood, MA, 1988.
- [32] Yasukawa, H., I. Furukawa and Y. Ishiyama, "Acoustic Echo Control for High Quality Audio Teleconferencing", *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Glasgow, Scotland, May, 1989, pp. 2041-2044.
- [33] Gierlich, H.W., "A New Concept for a Hands-Free Telephone", *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, NY, October, 1989.
- [34] Leon-Garcia, A., *Probability and Random Processes for Electrical Engineering*, Addison-Wesley, Don Mills, 1989.
- [35] Oppenheim, A.V. and R.W. Schaffer, *Discrete-Time Signal Processing*, Prentice-Hall, NJ, 1989.
- [36] Murano, K., S. Unagami, and F. Amano, "Echo Cancellation and Applications", *IEEE Communications Magazine*, Jan., 1990.
- [37] Makino, S. and Y. Kaneda, "Acoustic Echo Canceller Algorithm Based on the Variation Characteristics of a Room Impulse Response", *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Albuquerque, New Mexico, April, 1990, pp. 1133-1136.
- [38] Gay, S. L. and R. J. Mammone, "Fast Converging Subband Acoustic Echo Cancellation Using RAP on the WE DSP16A", *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Albuquerque, New Mexico, April, 1990, pp. 1141-1144.
- [39] Hatty, Birger, "Recursive Least Squares Algorithms Using Multirate Systems for Cancellation of Acoustical Echoes", *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Albuquerque, New Mexico, April, 1990, pp. 1145-1148.
- [40] Hansler, E., "Adaptive Echo Cancellation Applied to the Hands-Free Telephone Problem", *Proc. IEEE Int. Symp. Circuits and Systems*, New Orleans, Louisiana, May, 1990, pp. 279-282.
- [41] Fleischmann, G., R. Frenzel, B. Hatty, and E. Bjarnason, "Multi Signal Processor Systems for Acoustical Echo Control", *Proc. IEEE Int. Symp. Circuits and Systems*, New Orleans, Louisiana, May, 1990, pp. 1430-1433.
- [42] Aboukarr, B., "Adaptive Predistortion Linearization of Nonlinear Power Amplifiers", M. Eng. Thesis, Carleton University, Sept., 1990.
- [43] Slock, D. T. M. and T. Kailath, "Numerically Stable Fast Transversal Filters for Recursive Least Squares Adaptive Filtering", *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-39, pp. 92-113, Jan. 1991.

- [44] Abel, J. S. and J. O. Smith, III, "Restoring A Clipped Signal", *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Toronto, May, 1991, pp. 1749-1752.
- [45] Kim, K. and E. J. Powers, "Analysis of Initialization and Numerical Instability of Fast RLS Algorithms", *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Toronto, May, 1991, pp. 1861-1864.
- [46] Baik, H. K. and V. J. Matthews, "Adaptive Algorithms for Identifying Recursive Nonlinear Systems", *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Toronto, May, 1991, pp. 2077-2080.
- [47] Amano, F., and H. Perez, "A New Subband Echo Canceller Structure", *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Toronto, May, 1991, pp. 3585-3588.
- [48] Gao, F. and W. M. Snelgrove, "Adaptive Linearization of a Loudspeaker", *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Toronto, May, 1991, pp. 3589-3592.
- [49] Haykin, S., *Adaptive Filter Theory*, 2nd ed., Prentice-Hall, Toronto, 1991.
- [50] Chao, J. and S. Tsuji, "A Stable and Distortion-Free IIR Echo and Howling Canceller", *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-39, pp. 1812-1821, Aug. 1991.
- [51] Moustakides, G. V. and S. Theodoridis, "Fast Newton Transversal Filters - A New Class of Adaptive Estimation Algorithms", *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-39, pp. 2184-2193, Oct. 1991.
- [52] Champagne, B., A. Lobo and P. Kabal, "Efficient Methods for Simulating a Moving Talker in a Rectangular Room", *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, NY, October, 1991.
- [53] Chu, P., "Audio Compression and Echo Cancellation for Low Bit Rate Video Teleconferencing", *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, NY, October, 1991.
- [54] Sondhi, M. M. and D. R. Morgan, "Acoustic Echo Cancellation for Stereophonic Teleconferencing", *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, NY, October, 1991.
- [55] Frenzel, R. and M. E. Hennecke, "A Robust Echo Compensator: Implementation and Realtime Measurements", *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, NY, October, 1991.
- [56] *Hypersignal Users Manual (revision 2.0)*, Hyperception Inc., 1991.
- [57] Cho, Y. S. and E. J. Powers, "Two-Tone vs. Random Process Inputs For Nonlinear Distortion Estimation", *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, San Francisco, March, 1992, vol. II, pp. 209-212.

- [58] Haneda, Y., S. Makino and Y. Kaneda, "Modelling of a Room Transfer Function Using Common Acoustical Poles", *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, San Francisco, March, 1992. vol. II, pp. 213-216.
- [59] Mboup, M., M. Bonnet and N. Bershad, "Coupled Adaptive Prediction and System Identification: A Statistical Model and Transient Analysis", *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, San Francisco, March, 1992, vol. IV, pp. 1-5.
- [60] Cho, Y. S., S. B. Kim, E. L. Hixson and E. J. Powers, "A Digital Technique to Estimate Second-Order Distortion Using Higher Order Coherence Spectra", *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-40, pp. 1029-1040, May 1992.
- [61] Van de Kerkhof, L. M. and W. J. W. Kitzen, "Tracking of a Time-Varying Acoustic Impulse Response by an Adaptive Filter", *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-40, pp. 1285-1294, June 1992.
- [62] Kwong, R. H., and E. W. Johnston, "A Variable Step Size LMS Algorithm", *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-40, pp. 1633-1642, July 1992.
- [63] Marshall, D. F. and W. K. Jenkins, "A Fast Quasi-Newton Adaptive Filtering Algorithm", *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-40, pp. 1652-1662, July. 1992.

10.0 APPENDICES

Appendix A - DAT Recordings

TABLE 13. Full List of .TIM Files From DAT Recordings

Filename	Terminal	Room	Source	Conditions
WBSR12.TIM WBSP12.TIM	WB	DPN-A Conference Room	Male speech (300-3400 Hz)	Stationary
WBNR13.TIM WBNP13.TIM	WB	DPN-A Conference Room	White noise (300-3400 Hz)	Stationary
WBNR14.TIM WBNP14.TIM	WB	DPN-A Conference Room	White noise (300-3400 Hz)	Non-stationary (L-R hand movement in front of set every 5 seconds)
NBSR15.TIM NBSP15.TIM	NB	DPN-A Conference Room	Male speech (300-3400 Hz)	Stationary
NBNR16.TIM NBNP16.TIM	NB	DPN-A Conference Room	White noise (300-3400 Hz)	Stationary
NBNR17.TIM NBNP17.TIM	NB	DPN-A Conference Room	White noise (300-3400 Hz)	Non-stationary (L-R hand movement in front of set every 5 seconds)
WBSR21.TIM WBSP21.TIM	WB	Anechoic chamber	Male speech (300-3400 Hz)	Stationary
WBNR22.TIM WBNP22.TIM	WB	Anechoic chamber	White noise (300-3400 Hz)	Stationary
NBSR23.TIM NBSP23.TIM	NB	Anechoic chamber	Male speech (300-3400 Hz)	Stationary
NBNR25.TIM NBNP25.TIM	NB	Anechoic chamber	White noise (300-3400 Hz)	Stationary
NBNR33.TIM NBNP33.TIM	NB	DPN-A Conference Room	White noise (300-3400 Hz)	Stationary
NBNR34.TIM NBNP34.TIM	NB	DPN-A Conference Room	White noise (300-3400 Hz)	Non-stationary (L/R Body movement: 1 metre from set every 5 seconds)
NBNR38.TIM NBNP38.TIM	NB	Anechoic chamber	White noise (300-3400 Hz)	75 dBSPL Loudspeaker output at 1 m
NBNR39.TIM NBNP39.TIM	NB	Anechoic chamber	White noise (300-3400 Hz)	70 dBSPL Loudspeaker output at 1 m

TABLE 13. Full List of .TIM Files From DAT Recordings

Filename	Terminal	Room	Source	Conditions
NBNR40.TIM NBNP40.TIM	NB	Anechoic chamber	White noise (300-3400 Hz)	65 dBSPL Loudspeaker output at 1 m
NBNR41.TIM NBNP41.TIM	NB	Anechoic chamber	White noise (300-3400 Hz)	61 dBSPL Loudspeaker output at 1 m
NBNR43.TIM NBNP43.TIM	NB	DPN-A Conference Room	White noise (300-3400 Hz)	75 dBSPL Loudspeaker output at 1 m, stationary conditions
NBNR44.TIM NBNP44.TIM	NB	DPN-A Conference Room	White noise (300-3400 Hz)	75 dBSPL Loudspeaker output at 1 m, alternating near-end speech with body movement in front of set
NBSR45.TIM NBSP45.TIM	NB	DPN-A Conference Room	Male speech (300-3400 Hz)	68 dBSPL Loudspeaker output at 1 m, stationary conditions
NBSR46.TIM NBSP46.TIM	NB	DPN-A Conference Room	Male speech (300-3400 Hz)	68 dBSPL Loudspeaker output at 1 m, alternating near-end speech with body movement in front of set

Appendix B - Fast Transversal Filter Algorithm Derivation

The derivation for the Fast Transversal Filter is based on the removal of redundancy from the RLS algorithm by employing 4 separate transversal filters of order N to obtain the RLS solution. These filters are the forward linear prediction filter, the backward linear prediction filter, the extended gain vector and the joint process extension filter. The original FTF derivation by Cioffi and Kailath [9] used a geometrical approach with slightly different nomenclature. The derivation included here follows the algebraic FTF derivation of Haykin [49], with additional intermediate steps included for clarity.

B.1 Adaptive Forward Linear Prediction

Starting with a forward linear predictor of order M , optimized in a least squares sense with weight vector $\hat{\mathbf{w}}(n)$, we can define the a posteriori forward prediction error as

$$f_M(i) = u(i) - \hat{\mathbf{w}}^H(n) \mathbf{u}_M(i-1), \quad 1 \leq i \leq n \quad (91)$$

where

$$\mathbf{u}_M^T(i-1) = [u(i-1), u(i-2), \dots, u(i-M)] \quad (92)$$

We can then rewrite $f_M(i)$ by defining a forward prediction-error filter $\mathbf{a}_M(n)$, where

$$\mathbf{a}_M(n) = \begin{bmatrix} 1 \\ -\hat{\mathbf{w}}(n) \end{bmatrix} \quad (93)$$

Redefining the input vector as being of order $M+1$

$$\mathbf{u}_{M+1}(i) = \begin{bmatrix} u(i) \\ \mathbf{u}_M(i-1) \end{bmatrix} \quad (94)$$

which allows us to write

$$f_M(i) = \mathbf{a}_M^H(n) \mathbf{u}_{M+1}(i), \quad 1 \leq i \leq n \quad (95)$$

The forward prediction error filter $\mathbf{a}_M(n)$ is also used in the augmented normal equations (see [49]) as defined for forward linear predictors:

$$\Phi_{M+1}(n) \mathbf{a}_M(n) = \begin{bmatrix} \mathcal{F}_M(n) \\ \mathbf{0} \end{bmatrix} \quad (96)$$

This compactly defines the least squares solution, where $\mathcal{F}_M(n)$ is the minimum value of the sum of weighted forward a posteriori prediction error squares:

$$\mathcal{F}_M(n) = \sum_{i=1}^n \lambda^{n-i} |f_M(i)|^2 \quad (97)$$

The $M+1$ by $M+1$ matrix $\Phi_{M+1}(n)$ is defined as the correlation matrix of the input vector $\mathbf{u}_{M+1}(i)$:

$$\Phi_{M+1}(n) = \sum_{i=1}^n \lambda^{n-i} \mathbf{u}_{M+1}(i) \mathbf{u}_{M+1}^H(i) \quad (98)$$

This matrix can be partitioned as follows:

$$\Phi_{M+1}(n) = \begin{bmatrix} \mathbf{U}_1(n) & \boldsymbol{\theta}_1^H(n) \\ \boldsymbol{\theta}_1(n) & \Phi_M(n-1) \end{bmatrix} \quad (99)$$

where

$$U_1(n) = \sum_{i=1}^n \lambda^{n-i} u(i)^2 \quad (100)$$

$$\theta_1(n) = \sum_{i=1}^n \lambda^{n-i} \mathbf{u}_M(i-1) u^*(i) \quad (101)$$

and

$$\begin{aligned} \Phi_M(n-1) &= \sum_{i=1}^n \lambda^{n-i} \mathbf{u}_M(i-1) \mathbf{u}_M^H(i-1) \\ &= \sum_{i=1}^{n-1} \lambda^{(n-1)-i} \mathbf{u}_M(i) \mathbf{u}_M^H(i) \end{aligned} \quad (102)$$

The RLS recursive equation for updating the tap weight vector,

$$\hat{\mathbf{w}}(n) = \hat{\mathbf{w}}(n-1) + \mathbf{k}(n-1) \alpha_M^*(n) \quad (103)$$

where

$$\alpha_M(n) = d(n) - \hat{\mathbf{w}}^H(n-1) \mathbf{u}(n) \quad (104)$$

can be rewritten in terms of a forward prediction:

$$\hat{\mathbf{w}}(n) = \hat{\mathbf{w}}(n-1) + \mathbf{k}_M(n-1) \eta_M^*(n) \quad (105)$$

where the forward a priori prediction error is

$$\eta_M(n) = u(n) - \hat{\mathbf{w}}^H(n-1) \mathbf{u}_M(n-1) \quad (106)$$

and the gain vector $\mathbf{k}_M(n-1)$ is defined as

$$\mathbf{k}_M(n-1) = \Phi_M^{-1}(n-1) \mathbf{u}_M(n-1) \quad (107)$$

Equation [105] can be substituted back into [93] to obtain the time update equation for the forward prediction-error filter $\mathbf{a}_M(n)$:

$$\begin{aligned} \mathbf{a}_M(n) &= \begin{bmatrix} 1 \\ -\hat{\mathbf{w}}(n) \end{bmatrix} \\ &= \begin{bmatrix} 1 \\ -\hat{\mathbf{w}}(n-1) \end{bmatrix} - \begin{bmatrix} 0 \\ \mathbf{k}_M(n-1) \end{bmatrix} \eta_M^*(n) \\ &= \mathbf{a}_M(n-1) - \begin{bmatrix} 0 \\ \mathbf{k}_M(n-1) \end{bmatrix} \eta_M^*(n) \end{aligned} \quad (108)$$

The forward a priori prediction error $\eta_M(n)$ can also be rewritten as follows:

$$\begin{aligned} \eta_M(n) &= u(n) - \hat{\mathbf{w}}^H(n-1) \mathbf{u}_M(n-1) \\ &= [1, -\hat{\mathbf{w}}^H(n-1)] \begin{bmatrix} u(n) \\ \mathbf{u}_M(n-1) \end{bmatrix} \\ &= \mathbf{a}_M^H(n-1) \mathbf{u}_{M+1}(n) \end{aligned} \quad (109)$$

From the RLS recursion for the minimum value of the sum of weighted error squares,

$$\mathcal{E}_{\min}(n) = \lambda \mathcal{E}_{\min}(n-1) + \alpha^*(n) e(n) \quad (110)$$

we can rewrite this for the forward prediction case,

$$\mathcal{F}_M(n) = \lambda \mathcal{F}_M(n-1) + \eta_M(n) f_M^*(n) \quad (111)$$

where

$$\eta_M(n)f_M^*(n) = \eta_M^*(n)f_M(n) \quad (112)$$

For the forward linear prediction section of the FTF algorithm, Equations (109) and (111) are needed for the final list of algorithm steps.

B.2 Adaptive Backward Linear Prediction

A derivation similar to that for forward linear prediction can be made for backward linear prediction. Starting with a backward linear predictor of order M , optimized in a least squares sense with weight vector $\mathbf{g}(n)$, we can define the a posteriori backward prediction error as

$$b_M(i) = u(i-M) - \mathbf{g}^H(n) \mathbf{u}_M(i) \quad (113)$$

where

$$\mathbf{u}_M^T(i) = [u(i), u(i-1), \dots, u(i-M+1)] \quad (114)$$

We can then rewrite $b_M(i)$ by defining a prediction-error filter $\mathbf{c}_M(n)$, where

$$\mathbf{c}_M(n) = \begin{bmatrix} -\mathbf{g}(n) \\ 1 \end{bmatrix} \quad (115)$$

Redefining the input vector as being of order $M+1$,

$$\mathbf{u}_{M+1}(i) = \begin{bmatrix} \mathbf{u}_M(i) \\ u(i-M) \end{bmatrix} \quad (116)$$

this gives

$$b_M(i) = \mathbf{c}_M^H(n) \mathbf{u}_{M+1}(i), \quad 1 \leq i \leq n \quad (117)$$

Similar to forward prediction, $\mathbf{c}_M(n)$ is also used in the augmented normal equations, here defined for backward linear predictors:

$$\Phi_{M+1}(n) \mathbf{c}_M(n) = \begin{bmatrix} \mathbf{0} \\ \mathcal{B}_M(n) \end{bmatrix} \quad (118)$$

This also compactly defines the least squares solution, where $\mathcal{B}_M(n)$ is the minimum value of the sum of weighted backward a posteriori prediction error squares:

$$\mathcal{B}_M(n) = \sum_{i=1}^n \lambda^{n-i} |b_M(i)|^2 \quad (119)$$

Partitioning $\Phi_{M+1}(n)$ (see Equation (98)) as follows,

$$\Phi_{M+1}(n) = \begin{bmatrix} \Phi_M(n) & \theta_2(n) \\ \theta_2^H(n) & U_2(n) \end{bmatrix} \quad (120)$$

this gives

$$\begin{aligned} U_2(n) &= \sum_{i=1}^n \lambda^{n-i} |u(i-M)|^2 \\ &= \sum_{i=1}^n \lambda^{(n-M)-i} |u(i)|^2 \end{aligned} \quad (121)$$

$$\theta_2(n) = \sum_{i=1}^n \lambda^{n-i} \mathbf{u}_M(i) u^*(i-M) \quad (122)$$

and

$$\Phi_M(n) = \sum_{i=1}^n \lambda^{n-i} \mathbf{u}_M(i) \mathbf{u}_M^H(i) \quad (123)$$

Similar to Equation (103) for forward prediction, we can generate a recursive update equation for the backward prediction tap-weight vector:

$$\mathbf{g}(n) = \mathbf{g}(n-1) + \mathbf{k}_M(n) \psi_M^*(n) \quad (124)$$

where the backward a priori prediction error $\psi_M(n)$ is defined as

$$\psi_M(n) = u(n-M) - \mathbf{g}^H(n-1) \mathbf{u}_M(n) \quad (125)$$

and the gain vector $\mathbf{k}_M(n)$ is defined as

$$\mathbf{k}_M(n) = \Phi_M^{-1}(n) \mathbf{u}_M(n) \quad (126)$$

Substituting Equation (124) back into Equation (115),

$$\begin{aligned} \mathbf{c}_M(n) &= \begin{bmatrix} -\mathbf{g}(n) \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} -\mathbf{g}(n-1) \\ 1 \end{bmatrix} - \begin{bmatrix} \mathbf{k}_M(n) \\ 0 \end{bmatrix} \psi_M^*(n) \\ &= \mathbf{c}_M(n-1) - \begin{bmatrix} \mathbf{k}_M(n) \\ 0 \end{bmatrix} \psi_M^*(n) \end{aligned} \quad (127)$$

The backward a priori prediction error $\psi_M(n)$ can also be rewritten as follows:

$$\begin{aligned}
\psi_M(n) &= u(n-M) - \mathbf{g}^H(n-1) \mathbf{u}_M(n) \\
&= \begin{bmatrix} -\mathbf{g}^H(n-1) & 1 \end{bmatrix} \begin{bmatrix} \mathbf{u}_M(n) \\ u(n-M) \end{bmatrix} \\
&= \mathbf{c}_M^H(n-1) \mathbf{u}_{M+1}(n)
\end{aligned} \tag{128}$$

Similar to the recursion for the minimum value of the sum of forward a posteriori weighted error squares (Equation (111)), we can write a recursion for the backward prediction case,

$$\mathcal{B}_M(n) = \lambda \mathcal{B}_M(n-1) + \psi_M(n) b_M^*(n) \tag{129}$$

where

$$\psi_M(n) b_M^*(n) = \psi_M^*(n) b_M(n) \tag{130}$$

For the backward linear prediction section of the FTF algorithm, Equation (129) is needed for the final list of algorithm steps.

B.3 Extended Gain Vector Derivation

Similar to Equation (126), a gain vector of extended order $M+1$ can be defined:

$$\mathbf{k}_{M+1}(n) = \Phi_{M+1}^{-1}(n) \mathbf{u}_{M+1}(n) \tag{131}$$

The inverse of the correlation matrix may be expressed as follows (see [49]),

$$\Phi_{M+1}^{-1}(n) = \begin{bmatrix} 0 & \mathbf{0}_M^T \\ \mathbf{0}_M & \Phi_M^{-1}(n-1) \end{bmatrix} + \frac{1}{\mathcal{F}_M(n)} \mathbf{a}_M(n) \mathbf{a}_M^H(n) \tag{132}$$

Postmultiplying Equation (132) by $\mathbf{u}_{M+1}(n)$, using the partitioned form from Equation (94) for the first term on the right hand side, and using the definition of the forward a posteriori error from Equation (95), we get

$$\begin{aligned}
 \Phi_{M+1}^{-1}(n) \mathbf{u}_{M+1}(n) &= \begin{bmatrix} 0 & \mathbf{0}_M^T \\ \mathbf{0}_M & \Phi_M^{-1}(n-1) \end{bmatrix} \mathbf{u}_{M+1}(n) + \frac{\mathbf{a}_M(n) \mathbf{a}_M^H(n) \mathbf{u}_{M+1}(n)}{\mathcal{F}_M(n)} \\
 \therefore \mathbf{k}_{M+1}(n) &= \begin{bmatrix} 0 & \mathbf{0}_M^T \\ \mathbf{0}_M & \Phi_M^{-1}(n-1) \end{bmatrix} \mathbf{u}_{M+1}(n) + \frac{1}{\mathcal{F}_M(n)} \mathbf{a}_M(n) \mathbf{a}_M^H(n) \mathbf{u}_{M+1}(n) \\
 &= \begin{bmatrix} 0 & \mathbf{0}_M^T \\ \mathbf{0}_M & \Phi_M^{-1}(n-1) \end{bmatrix} \begin{bmatrix} u(i) \\ \mathbf{u}_M(i-1) \end{bmatrix} + \frac{1}{\mathcal{F}_M(n)} \mathbf{a}_M(n) \mathbf{a}_M^H(n) \mathbf{u}_{M+1}(n) \\
 &= \begin{bmatrix} 0 \\ \Phi_M^{-1}(n-1) \mathbf{u}_M(i-1) \end{bmatrix} + \frac{1}{\mathcal{F}_M(n)} \mathbf{a}_M(n) f_M(n) \\
 &= \begin{bmatrix} 0 \\ \mathbf{k}_M(n-1) \end{bmatrix} + \frac{f_M(n)}{\mathcal{F}_M(n)} \mathbf{a}_M(n)
 \end{aligned} \tag{133}$$

Similarly, $\Phi_{M+1}^{-1}(n)$ can be expressed in terms of backward prediction (see [49]),

$$\Phi_{M+1}^{-1}(n) = \begin{bmatrix} \Phi_M^{-1}(n) & \mathbf{0}_M \\ \mathbf{0}_M^T & 0 \end{bmatrix} + \frac{1}{\mathcal{B}_M(n)} \mathbf{c}_M(n) \mathbf{c}_M^H(n) \tag{134}$$

Postmultiplying Equation (134) by $\mathbf{u}_{M+1}(n)$, using the partitioned form from Equation (116) for the first term on the right hand side, and using the definition of the backward a posteriori error from Equation (117), we get

$$\Phi_{M+1}^{-1}(n) \mathbf{u}_{M+1}(n) = \begin{bmatrix} \Phi_M^{-1}(n) & \mathbf{0}_M \\ \mathbf{0}_M^T & 0 \end{bmatrix} \mathbf{u}_{M+1}(n) + \frac{\mathbf{c}_M(n) \mathbf{c}_M^H(n) \mathbf{u}_{M+1}(n)}{B_M(n)}$$

$$\therefore \mathbf{k}_{M+1}(n) = \begin{bmatrix} \Phi_M^{-1}(n) & \mathbf{0}_M \\ \mathbf{0}_M^T & 0 \end{bmatrix} \mathbf{u}_{M+1}(n) + \frac{1}{B_M(n)} \mathbf{c}_M(n) \mathbf{c}_M^H(n) \mathbf{u}_{M+1}(n)$$

$$= \begin{bmatrix} \Phi_M^{-1}(n) & \mathbf{0}_M \\ \mathbf{0}_M^T & 0 \end{bmatrix} \begin{bmatrix} \mathbf{u}_M(i) \\ u(i-M) \end{bmatrix} + \frac{1}{B_M(n)} \mathbf{c}_M(n) \mathbf{c}_M^H(n) \mathbf{u}_{M+1}(n)$$

$$= \begin{bmatrix} \Phi_M^{-1}(n) \mathbf{u}_M(i) \\ 0 \end{bmatrix} + \frac{1}{B_M(n)} \mathbf{c}_M(n) b_M(n)$$

$$= \begin{bmatrix} \mathbf{k}_M(n) \\ 0 \end{bmatrix} + \frac{b_M(n)}{B_M(n)} \mathbf{c}_M(n)$$

(135)

For the extended gain vector section of the FTF algorithm, no equations are used as is from the previous derivations for the final algorithm. The extended gain vector derivations will be reconfigured in the section "Final FTF Algorithm Derivation" on page 170.

B.4 Conversion Factor

Define the conversion factor $\gamma_M(n)$ as the least squares estimation error for the filter $\mathbf{k}(n)$ to the first coordinate (pinning) vector $[1, \mathbf{0}_{n-1}]$, giving

$$\begin{aligned} \gamma_M(n) &= 1 - \mathbf{k}_M^H(n) \mathbf{u}_M(n) \\ &= 1 - \mathbf{u}_M^H(n) \Phi_M^{-1} \mathbf{u}_M(n) \end{aligned}$$

(136)

Substituting the recursive equation for the inverse of the correlation matrix from the use of the matrix inversion lemma in the RLS derivation (see Haykin [49]),

$$\Phi_M^{-1}(n-1) = \lambda^{-1} \Phi_M^{-1}(n-1) - \frac{\lambda^{-2} \Phi_M^{-1}(n-1) \mathbf{u}(n) \mathbf{u}^H(n) \Phi_M^{-1}(n-1)}{1 + \lambda^{-1} \mathbf{u}^H(n) \Phi_M^{-1}(n-1) \mathbf{u}(n)} \quad (137)$$

this gives

$$\begin{aligned} \gamma_M(n) &= 1 - \mathbf{u}_M^H(n) \lambda^{-1} \Phi_M^{-1}(n-1) \mathbf{u}_M(n) \\ &\quad + \mathbf{u}_M^H(n) \frac{\lambda^{-2} \Phi_M^{-1}(n-1) \mathbf{u}(n) \mathbf{u}^H(n) \Phi_M^{-1}(n-1)}{1 + \lambda^{-1} \mathbf{u}^H(n) \Phi_M^{-1}(n-1) \mathbf{u}(n)} \mathbf{u}_M(n) \end{aligned} \quad (138)$$

which simplifies to

$$\gamma_M(n) = \frac{1}{1 + \lambda^{-1} \mathbf{u}^H(n) \Phi_M^{-1}(n-1) \mathbf{u}(n)} \quad (139)$$

This conversion factor can be interpreted in a number of ways:

1. As the ratio between the estimation a priori error and a posteriori error:

$$\begin{aligned}
\frac{e_M(n)}{\alpha_M(n)} &= \frac{d(n) - \mathbf{w}_M^H(n) \mathbf{u}_M(n)}{d(n) - \mathbf{w}_M^H(n-1) \mathbf{u}_M(n)} \\
&= \frac{d(n) - [\hat{\mathbf{w}}_M^H(n-1) + \mathbf{k}_M^H(n) \alpha_M(n)] \mathbf{u}_M(n)}{d(n) - \hat{\mathbf{w}}_M^H(n-1) \mathbf{u}_M(n)} \\
&= \frac{d(n) - \hat{\mathbf{w}}_M^H(n-1) \mathbf{u}_M(n) - \mathbf{k}_M^H(n) \alpha_M(n) \mathbf{u}_M(n)}{d(n) - \hat{\mathbf{w}}_M^H(n-1) \mathbf{u}_M(n)} \\
&= \frac{\alpha_M(n) - \mathbf{k}_M^H(n) \alpha_M(n) \mathbf{u}_M(n)}{\alpha_M(n)} \\
&= 1 - \mathbf{k}_M^H(n) \mathbf{u}_M(n) \\
&= \gamma_M(n)
\end{aligned} \tag{140}$$

2. As a ratio between the a priori and the a posteriori forward linear prediction error:

Postmultiplying the Hermitian transposed sides of (108) by $\mathbf{u}_{M+1}(n)$,

$$\begin{aligned}
\mathbf{a}_M^H(n) \mathbf{u}_{M+1}(n) &= \mathbf{a}_M^H(n-1) \mathbf{u}_{M+1}(n) - \eta_M(n) \begin{bmatrix} 0 \\ \mathbf{k}_M(n-1) \end{bmatrix}^H \mathbf{u}_{M+1}(n) \\
&= \mathbf{a}_M^H(n-1) \mathbf{u}_{M+1}(n) - \eta_M(n) \begin{bmatrix} 0 \\ \mathbf{k}_M(n-1) \end{bmatrix}^H \begin{bmatrix} u(n) \\ \mathbf{u}_M(n-1) \end{bmatrix}
\end{aligned} \tag{141}$$

Substituting for the definitions of the a posteriori and a priori forward prediction error, using (95) and (109) respectively,

$$\begin{aligned}
 f_M(n) &= \eta_M(n) - \eta_M(n) \begin{bmatrix} 0 \\ \mathbf{k}_M(n-1) \end{bmatrix}^H \begin{bmatrix} u(n) \\ \mathbf{u}_M(n-1) \end{bmatrix} \\
 &= \eta_M(n) [1 - \mathbf{k}_M^H(n-1) \mathbf{u}_M(n-1)] \\
 &= \eta_M(n) \gamma_M(n-1)
 \end{aligned} \tag{142}$$

Giving

$$\frac{f_M(n)}{\eta_M(n)} = \gamma_M(n-1) \tag{143}$$

3. As a ratio between the a priori and the a posteriori backward linear prediction error:

Postmultiplying the Hermitian transposed sides of (127) by $\mathbf{u}_{M+1}(n)$,

$$\begin{aligned}
 \mathbf{c}_M^H(n) \mathbf{u}_{M+1}(n) &= \mathbf{c}_M^H(n-1) \mathbf{u}_{M+1}(n) - \psi_M(n) \begin{bmatrix} \mathbf{k}_M(n) \\ 0 \end{bmatrix}^H \begin{bmatrix} u(n) \\ \mathbf{u}_M(n-1) \end{bmatrix} \\
 &= \mathbf{c}_M^H(n-1) \mathbf{u}_{M+1}(n) - \psi_M(n) \begin{bmatrix} \mathbf{k}_M(n) \\ 0 \end{bmatrix}^H \begin{bmatrix} \mathbf{u}_M(n) \\ u(n-M) \end{bmatrix}
 \end{aligned} \tag{144}$$

Substituting for the definitions of the a posteriori and a priori backward prediction error, using (117) and (128) respectively,

$$\begin{aligned}
 b_M(n) &= \psi_M(n) - \psi_M(n) \begin{bmatrix} \mathbf{k}_M(n) \\ 0 \end{bmatrix}^H \begin{bmatrix} \mathbf{u}_M(n) \\ u(n-M) \end{bmatrix} \\
 &= \psi_M(n) [1 - \mathbf{k}_M^H(n) \mathbf{u}_M(n)] \\
 &= \psi_M(n) \gamma_M(n)
 \end{aligned} \tag{145}$$

Giving

$$\frac{b_M(n)}{\psi_M(n)} = \gamma_M(n) \quad (146)$$

Several order and time update recursions can be derived for the conversion factor:

1. Time and order update using the minimum sum of weighted forward prediction-error squares:

Postmultiplying the Hermitian transposed sides of () by $\mathbf{u}_{M+1}(n)$,

$$\begin{aligned} \mathbf{k}_{M+1}^H(n) \mathbf{u}_{M+1}(n) &= \mathbf{k}_M^H(n-1) \mathbf{u}_{M+1}(n) + \frac{f_M^*(n)}{\mathcal{F}_M(n)} \mathbf{a}_M^H(n) \mathbf{u}_{M+1}(n) \\ &= \mathbf{k}_M^H(n-1) \mathbf{u}_M(n-1) + \frac{f_M^*(n)}{\mathcal{F}_M(n)} f_M(n) \\ &= \mathbf{k}_M^H(n-1) \mathbf{u}_M(n-1) + \frac{f_M(n)^2}{\mathcal{F}_M(n)} \end{aligned} \quad (147)$$

Substituting for the definitions of $\gamma_{M+1}(n)$ and $\gamma_M(n-1)$,

$$\gamma_{M+1}(n) = \gamma_M(n-1) - \frac{f_M(n)^2}{\mathcal{F}_M(n)} \quad (148)$$

2. Order update using the minimum sum of weighted backward prediction-error squares:

Postmultiplying the Hermitian transposed sides of (135) by $\mathbf{u}_{M+1}(n)$,

$$\begin{aligned}
\mathbf{k}_{M+1}^H(n) \mathbf{u}_{M+1}(n) &= \mathbf{k}_M^H(n) \mathbf{u}_{M+1}(n) + \frac{b_M^*(n)}{B_M(n)} c_M^H(n) \mathbf{u}_{M+1}(n) \\
&= \mathbf{k}_M^H(n) \mathbf{u}_M(n) + \frac{b_M^*(n)}{B_M(n)} b_M(n) \\
&= \mathbf{k}_M^H(n) \mathbf{u}_M(n) + \frac{b_M(n)^2}{B_M(n)}
\end{aligned} \tag{149}$$

Substituting for the definitions of $\gamma_{M+1}(n)$ and $\gamma_M(n)$,

$$\gamma_{M+1}(n) = \gamma_M(n) - \frac{b_M(n)^2}{B_M(n)} \tag{150}$$

3. Using Equations (111) and (143) to remove $f_M(n)^2$ from Equation (148),

$$\begin{aligned}
\gamma_{M+1}(n) &= \gamma_M(n-1) - \frac{f_M(n)^2}{F_M(n)} \\
&= \frac{f_M(n)}{\eta_M(n)} - \frac{f_M(n)f_M^*(n)}{\mathcal{F}_M(n)} \\
&= \frac{f_M(n)\mathcal{F}_M(n) - f_M(n)f_M^*(n)\eta_M(n)}{\eta_M(n)\mathcal{F}_M(n)} \\
&= \frac{f_M(n)\mathcal{F}_M(n) - f_M(n)[\mathcal{F}_M(n) - \lambda\mathcal{F}_M(n-1)]}{\eta_M(n)\mathcal{F}_M(n)} \\
&= \frac{\lambda f_M(n)\mathcal{F}_M(n-1)}{\eta_M(n)\mathcal{F}_M(n)} \\
&= \frac{\lambda\mathcal{F}_M(n-1)}{\mathcal{F}_M(n)}\gamma_M(n-1)
\end{aligned} \tag{151}$$

4. Using Equations (129) and (146) to remove $b_M(n)^2$ from Equation (150),

$$\begin{aligned}
\gamma_{M+1}(n) &= \gamma_M(n) - \frac{b_M(n)^2}{\mathcal{B}_M(n)} \\
&= \frac{b_M(n)}{\psi_M(n)} - \frac{b_M(n)b_M^*(n)}{\mathcal{B}_M(n)} \\
&= \frac{b_M(n)\mathcal{B}_M(n) - b_M(n)b_M^*(n)\psi_M(n)}{\psi_M(n)\mathcal{B}_M(n)} \\
&= \frac{b_M(n)\mathcal{B}_M(n) - b_M(n)[\mathcal{B}_M(n) - \lambda\mathcal{B}_M(n-1)]}{\psi_M(n)\mathcal{B}_M(n)} \\
&= \frac{\lambda b_M(n)\mathcal{B}_M(n-1)}{\psi_M(n)\mathcal{B}_M(n)} \\
&= \frac{\lambda\mathcal{B}_M(n-1)}{\mathcal{B}_M(n)} \gamma_M(n)
\end{aligned} \tag{152}$$

For the conversion factor section of the FTF algorithm, Equations (140) (solving for $e(n)$), (142), (145) and (151) are needed for the final list of algorithm steps.

B.5 Final FTF Algorithm Derivation

With the background derivations complete, the derivation for the FTF algorithm proceeds as follows:

First, the a priori forward prediction error is computed (from Equation (109)):

$$\eta_M(n) = \mathbf{a}_M^H(n-1) \mathbf{u}_{M+1}(n) \tag{153}$$

The a posteriori forward prediction error is obtained by multiplying $\eta_M(n)$ by the delayed conversion factor $\gamma_M(n-1)$,

$$f_M(n) = \gamma_M(n-1) \eta_M(n) \quad (154)$$

The recursion derived in Equation (111) is used to time update the minimum value of the sum of weighted forward prediction-error squares:

$$\mathcal{F}_M(n) = \lambda \mathcal{F}_M(n-1) + \eta_M(n) f_M^*(n) \quad (155)$$

The conversion factor is then time and order updated by using the recursion derived in Equation (151):

$$\gamma_{M+1}(n) = \frac{\lambda \mathcal{F}_M(n-1)}{\mathcal{F}_M(n)} \gamma_M(n-1) \quad (156)$$

The updated extended gain vector is then computed by using Equation (133), substituting Equation (108) for the forward prediction-error filter $\mathbf{a}_M(n)$,

$$\begin{aligned} \mathbf{k}_{M+1}(n) &= \begin{bmatrix} 0 \\ \mathbf{k}_M(n-1) \end{bmatrix} + \frac{f_M(n)}{\mathcal{F}_M(n)} \mathbf{a}_M(n) \\ &= \begin{bmatrix} 0 \\ \mathbf{k}_M(n-1) \end{bmatrix} + \frac{f_M(n)}{\mathcal{F}_M(n)} \left(\mathbf{a}_M(n-1) - \begin{bmatrix} 0 \\ \mathbf{k}_M(n-1) \end{bmatrix} \eta_M^*(n) \right) \\ &= \left(1 - \frac{\eta_M^*(n) f_M(n)}{\mathcal{F}_M(n)} \right) \begin{bmatrix} 0 \\ \mathbf{k}_M(n-1) \end{bmatrix} + \frac{f_M(n)}{\mathcal{F}_M(n)} \mathbf{a}_M(n-1) \end{aligned} \quad (157)$$

By using Equations (111) and (151), the following simplification can be made:

$$\begin{aligned}
1 - \frac{\eta_M^*(n)f_M(n)}{\mathcal{F}_M(n)} &= \frac{\mathcal{F}_M(n) - \eta_M^*(n)f_M(n)}{\mathcal{F}_M(n)} \\
&= \lambda^{-1} \frac{\mathcal{F}_M(n-1)}{\mathcal{F}_M(n)} \\
&= \frac{\gamma_{M+1}(n)}{\gamma_M(n-1)}
\end{aligned} \tag{158}$$

Using Equations (143) and (151), the following can be written:

$$\begin{aligned}
\frac{f_M(n)}{\mathcal{F}_M(n)} &= \frac{\eta_M(n)\gamma_M(n-1)}{\mathcal{F}_M(n)} \\
&= \lambda^{-1} \frac{\eta_M(n)\gamma_{M+1}(n)}{\mathcal{F}_M(n-1)}
\end{aligned} \tag{159}$$

Substituting Equations (158) and (159) into (157),

$$\mathbf{k}_{M+1}(n) = \begin{bmatrix} \gamma_{M+1}(n) & 0 \\ \gamma_M(n-1) & \mathbf{k}_M(n-1) \end{bmatrix} + \lambda^{-1} \frac{\eta_M(n)\gamma_{M+1}(n)}{\mathcal{F}_M(n-1)} \cdot \mathbf{a}_M(n-1) \tag{160}$$

We can then define a normalized gain vector as

$$\tilde{\mathbf{k}}_M(n) = \frac{\mathbf{k}_M(n)}{\gamma_M(n)} \tag{161}$$

The extended gain vector update equation can then be rewritten as

$$\begin{aligned}
 \tilde{\mathbf{k}}_{M+1}(n) &= \frac{\mathbf{k}_M(n-1)}{\gamma_M(n-1)} + \lambda^{-1} \frac{\eta_M(n)}{f_M(n-1)} \mathbf{a}_M(n-1) \\
 &= \tilde{\mathbf{k}}_M(n-1) + \lambda^{-1} \frac{\eta_M(n)}{f_M(n-1)} \mathbf{a}_M(n-1)
 \end{aligned} \tag{162}$$

The normalization of Equation (161), as well as a substitution from Equation (143), can be used to rewrite the update recursion for the forward prediction-error filter from Equation (108):

$$\begin{aligned}
 \mathbf{a}_M(n) &= \mathbf{a}_M(n-1) - \frac{\mathbf{k}_M(n-1)}{\gamma_M(n-1)} \eta_M^*(n) \\
 &= \mathbf{a}_M(n-1) - \tilde{\mathbf{k}}_M(n-1) \eta_M^*(n) \gamma_M(n-1) \\
 &= \mathbf{a}_M(n-1) - \tilde{\mathbf{k}}_M(n-1) f_M^*(n)
 \end{aligned} \tag{163}$$

Next, it is necessary to derive a relation for the current normalized gain vector of order M ($\mathbf{k}_M(n)$), and to perform a time update on the backward prediction-error filter. Starting by substituting Equation (127) into Equation (135) and collecting common terms,

$$\begin{aligned}
\mathbf{k}_{M+1}(n) &= \begin{matrix} \mathbf{k}_M(n) \\ 0 \end{matrix} + \frac{b_M(n)}{\mathcal{B}_M(n)} \mathbf{c}_M(n) \\
&= \begin{matrix} \mathbf{k}_M(n) \\ 0 \end{matrix} + \frac{b_M(n)}{\mathcal{B}_M(n)} \left(\mathbf{c}_M(n-1) - \begin{matrix} \mathbf{k}_M(n) \\ 0 \end{matrix} \psi_M^*(n) \right) \\
&= \left(1 - \frac{\psi_M^*(n) b_M(n)}{\mathcal{B}_M(n)} \right) \begin{matrix} \mathbf{k}_M(n) \\ 0 \end{matrix} + \frac{b_M(n)}{\mathcal{B}_M(n)} \mathbf{c}_M(n-1)
\end{aligned} \tag{164}$$

Denoting the last element of $\mathbf{k}_{M+1}(n)$ as $k_{M+1,M+1}(n)$, and given that the last element of $\mathbf{c}_M(n-1)$ is unity, this gives

$$k_{M+1,M+1}(n) = \frac{b_M(n)}{\mathcal{B}_M(n)} \tag{165}$$

Normalizing with respect to $\gamma_{M+1}(n)$, and substituting in Equations (146) and (152),

$$\begin{aligned}
\tilde{k}_{M+1,M+1}(n) &= \frac{k_{M+1,M+1}(n)}{\gamma_{M+1}(n)} \\
&= \frac{b_M(n)}{\gamma_{M+1}(n) \mathcal{B}_M(n)} \\
&= \frac{\gamma_M(n) \psi_M(n)}{\gamma_{M+1}(n) \mathcal{B}_M(n)} \\
&= \frac{\psi_M(n)}{\lambda \mathcal{B}_M(n-1)}
\end{aligned} \tag{166}$$

Solving for $\psi_M(n)$,

$$\psi_M(n) = \lambda \mathcal{B}_M(n-1) \tilde{k}_{M+1, M+1}(n) \quad (167)$$

Rearranging Equation (152) to solve for $\gamma_M(n)$,

$$\gamma_M(n) = \frac{\mathcal{B}_M(n)}{\lambda \mathcal{B}_M(n-1)} \gamma_{M+1}(n) \quad (168)$$

Assuming $\mathcal{B}_M(n) \neq 0$, we can take Equation (129) and the second line from Equation (166) and perform the following manipulation:

$$\begin{aligned} \lambda \mathcal{B}_M(n-1) &= \mathcal{B}_M(n) - \psi_M(n) h_M^*(n) \\ \therefore \frac{\lambda \mathcal{B}_M(n-1)}{\mathcal{B}_M(n)} &= 1 - \frac{\psi_M(n) h_M^*(n)}{\mathcal{B}_M(n)} \\ &= 1 - \frac{\psi_M(n) (\gamma_{M+1}(n) \mathcal{B}_M(n) \tilde{k}_{M+1, M+1}^*(n))}{\mathcal{B}_M(n)} \\ &= 1 - \psi_M(n) \gamma_{M+1}(n) \tilde{k}_{M+1, M+1}^*(n) \end{aligned} \quad (169)$$

Substituting Equation (169) into (168), we get

$$\begin{aligned} \gamma_M(n) &= \frac{\mathcal{B}_M(n)}{\lambda \mathcal{B}_M(n-1)} \gamma_{M+1}(n) \\ &= (1 - \psi_M(n) \gamma_{M+1}(n) \tilde{k}_{M+1, M+1}^*(n))^{-1} \gamma_{M+1}(n) \end{aligned} \quad (170)$$

Noting that

$$\psi_M^*(n) \tilde{k}_{M+1, M+1}(n) = \psi_M(n) \tilde{k}_{M+1, M+1}^*(n) \quad (171)$$

Equation (170) can be rewritten as

$$\gamma_M(n) = (1 - \psi_M^*(n) \gamma_{M+1}(n) \tilde{k}_{M+1, M+1}(n))^{-1} \gamma_{M+1}(n) \quad (172)$$

Equation (145) is then used to compute the backward a posteriori prediction error:

$$b_M(n) = \gamma_M(n) \psi_M(n) \quad (173)$$

The recursion from Equation (129) can then be used to update the minimum value of the sum of weighted backward a posteriori prediction-error squares:

$$\mathcal{B}_M(n) = \lambda \mathcal{B}_M(n-1) + \psi_M(n) b_M^*(n) \quad (174)$$

The computation of the vector $\tilde{\mathbf{k}}_M(n)$ can now be completed, which is accomplished by substituting the first line of Equation (170), the second line of (166), and Equations (152) and (161) into Equation (164), rearranging to solve for $\tilde{\mathbf{k}}_M(n)$:

$$\begin{aligned}
\mathbf{k}_{M+1}(n) &= \begin{pmatrix} \lambda \mathcal{B}_M(n-1) \\ \mathcal{B}_M(n) \end{pmatrix} \mathbf{k}_M(n) + \begin{pmatrix} b_M(n) \\ \mathcal{B}_M(n) \end{pmatrix} \mathbf{c}_M(n-1) \\
&= \begin{pmatrix} \lambda \mathcal{B}_M(n-1) \\ \mathcal{B}_M(n) \end{pmatrix} \gamma_M(n) \frac{\mathbf{k}_M(n)}{\gamma_M(n)} + \begin{pmatrix} b_M(n) \\ \mathcal{B}_M(n) \end{pmatrix} \mathbf{c}_M(n-1) \\
&= \gamma_{M+1}(n) \begin{pmatrix} \tilde{\mathbf{k}}_M(n) \\ 0 \end{pmatrix} + \begin{pmatrix} b_M(n) \\ \mathcal{B}_M(n) \end{pmatrix} \mathbf{c}_M(n-1)
\end{aligned}$$

$$\begin{aligned}
\therefore \begin{pmatrix} \tilde{\mathbf{k}}_{M+1}(n) \\ 0 \end{pmatrix} &= \begin{pmatrix} \tilde{\mathbf{k}}_M(n) \\ 0 \end{pmatrix} + \gamma_{M+1}(n) \begin{pmatrix} b_M(n) \\ \mathcal{B}_M(n) \end{pmatrix} \mathbf{c}_M(n-1) \\
&= \begin{pmatrix} \tilde{\mathbf{k}}_M(n) \\ 0 \end{pmatrix} + \tilde{k}_{M+1, M+1}(n) \mathbf{c}_M(n-1)
\end{aligned}$$

$$\therefore \begin{pmatrix} \tilde{\mathbf{k}}_M(n) \\ 0 \end{pmatrix} = \begin{pmatrix} \tilde{\mathbf{k}}_{M+1}(n) - \tilde{k}_{M+1, M+1}(n) \mathbf{c}_M(n-1) \\ 0 \end{pmatrix}$$

(175)

This now allows the update of the tap-weight vector for the backward prediction-error filter using Equation (127), with substitutions from Equations (146) and (161):

$$\begin{aligned}
\mathbf{c}_M(n) &= \mathbf{c}_M(n-1) - \psi_M^*(n) \frac{\mathbf{k}_M(n)}{0} \\
&= \mathbf{c}_M(n-1) - \frac{h_M^*(n)}{\gamma_M(n)} \frac{\mathbf{k}_M(n)}{0} \\
&= \mathbf{c}_M(n-1) - h_M^*(n) \frac{\bar{\mathbf{k}}_M(n)}{0}
\end{aligned} \tag{176}$$

The only remaining steps in the FTF algorithm are those required for the joint process extension filter. The a priori estimation error can be computed using the error computation performed in the RLS algorithm (Equation (11)):

$$\alpha_M(n) = d(n) - \mathbf{w}_M^H(n-1) \mathbf{u}_M(n) \tag{177}$$

The a posteriori estimation error, not directly available in the RLS algorithm, can be computed using Equation (140):

$$e_M(n) = \gamma_M(n) \alpha_M(n) \tag{178}$$

Finally, the joint process extension tap weights can be updated using the same recursion used for RLS (Equation (12)), with substitutions from Equations (161) and (168):

$$\begin{aligned}
\hat{\mathbf{w}}_M(n) &= \hat{\mathbf{w}}_M(n-1) + \mathbf{k}_M(n) \alpha_M^*(n) \\
&= \hat{\mathbf{w}}_M(n-1) + \frac{\mathbf{k}_M(n)}{\gamma_M(n)} (\gamma_M(n) \alpha_M^*(n)) \\
&= \hat{\mathbf{w}}_M(n-1) + \bar{\mathbf{k}}_M(n) e_M^*(n)
\end{aligned} \tag{179}$$

Appendix C - Hypersignal File Format

All audio files saved with a .TIM DOS extension followed a standard binary file format used in the commercial software package Hypersignal Workstation by Hyperception. This file format consists of a fixed length 20 byte (10 16-bit integer words) header followed by a variable length stream of 16-bit two's complement binary data. The contents of this header are illustrated below:

TABLE 14. Hypersignal File Header Format

Header Word Number	Integer Parameter	Description of Parameter
0	AMPLITUDE	Maximum Absolute Value of Waveform Data (0 - 32767)
1	FRAMESIZE	Analysis Framesize
2	SAMPLING FREQUENCY, REM	Used to calculate sampling frequency, where $F_s = \text{SAMPLING FREQUENCY, DIV} \times 32767 + \text{SAMPLING FREQUENCY, REM}$
3	FFT ORDER	FFT Order not used (set to zero) for time data
4	NUMBER OF DATA VALUES, REM	Used to calculate total number of data values, where $N = \text{NUMBER OF DATA VALUES, DIV} \times 16384 + \text{NUMBER OF DATA VALUES, REM}$
5	FRAME OVERLAP	Used for FFT Generation, set to zero for time data
6	DATA TYPE	Generally set to zero for time data (except output of IIR or FIR filters which is set to two)
7	NUMBER OF CHANNELS	Set to zero for single channel data
8	SAMPLING FREQUENCY, DIV	Used to calculate actual sampling frequency; see parameter 2
9	NUMBER OF DATA VALUES, DIV	Used to calculate actual number of samples, see parameter 4

3 of/de 3

PM-1 3 1/2"x4" PHOTOGRAPHIC MICROCOPY TARGET
NBS 1010a ANSI/ISO #2 EQUIVALENT

PRECISIONSM RESOLUTION TARGETS

PIONEERS IN METHYLENE BLUE TESTING SINCE 1974

MICRO

1000 COUNTY ROAD A, E., HUNTSVILLE, AL 35891, USA
TEL: 256/445-7857 FAX: 256/445-7857 TLA 048884848

END

2|4 0|5 93

FIN