1

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

PRECISIONSM RESOLUTION TARGETS

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.

Tracking Performance of Acoustic Echo Cancellers

submitted by

Timothy G. D. Webster, B.Eng (Elec)

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

Master of Engineering

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

Carleton University
Ottawa, Ontario, Canada K1S 5B6

September 12, 1995

© copyright
1995, Tim Webster

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

Quotient / Note de 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 0-612-08969-X

Name John Doe

Dissertation Abstracts International is arranged by broad, general subject categories. Please select the one subject which most nearly describes the content of your dissertation. Enter the corresponding four-digit code in the spaces provided.

U·M·I

SUBJECT CODE

THE HUMANITIES AND SOCIAL SCIENCES

Architecture	0729
Art History	0377
Cinema	0900
Dance	0378
Fine Arts	0357
Information Science	0723
Journalism	0391
Library Science	0399
Mass Communications	0708
Music	0413
Speech Communication	0459
Theater	0465

Psychology	0525
Reading	0535
Religious	0527
Sciences	0714
Secondary	0533
Social Sciences	0534
Sociology of	0340
Special	0529
Teacher Training	0530
Technology	0710
Tests and Measurements	0288
Vocational	0747

General	0513
Administration	0514
Adult and Continuing	0516
Agricultural	0517
Art	0273
Bilingual and Multicultural	0282
Business	0688
Community College	0275
Curriculum and Instruction	0727
Early Childhood	0518
Elementary	0524
Finance	0277
Guidance and Counseling	0519
Health	0680
Higher	0745
History of	0520
Home Economics	0278
Industrial	0521
Language and Literature	0279
Mathematics	0280
Music	0522
Philosophy of	0998
Physical	0523

Language	0679
General	0289
Ancient	0290
Linguistics	0291
Modern	

Literature	0401
General	0294
Classical	0295
Comparative	0297
Medieval	0298
Modern	0316
African	0591
American	0305
Asian	0352
Canadian (English)	0355
Canadian (French)	0593
English	0311
Germanic	0312
Latin American	0315
Middle Eastern	0313
Romance	0314
Slavic and East European	

Philosophy	0422
Religion	
General	0318
Biblical Studies	0321
Clergy	0319
History of	0320
Philosophy of	0322
Theology	0469

SOCIAL SCIENCES
American Studies 0323
Anthropology
Archaeology 0324
Cultural 0326
Physical 0327

Physical	0327
Business Administration	
General	0310
Accounting	0272
Banking	0770
Management	0454
Marketing	0338
Canadian Studies	0385

Economics	0501
General	0501
Agricultural	0503
Commerce-Business	0505
Finance	0508
History	0509
Labor	0510
Theory	0511

Folklore	0358
Geography	0366
Gerontology	0351
History	
General	0578

Ancient	0579
Medieval	0581
Modern	0582
Black	0328
African	0331
Asia, Australia and Oceania	0332
Canadian	0334
European	0335
Latin American	0336
Middle Eastern	0333
United States	0337
History of Science	0585
Index	0398

Political Science	0610
General	0615
International Law and Relations	0616
Public Administration	0617

Public Administration	0817
Recreation	0814
Social Work	0452
Sociology	

General	0626
Criminology and Penology	0627
Demography	0938
Ethnic and Racial Studies	0631
Individual and Family Studies	0628

Industrial and Labor Relations	0629
Public and Social Welfare	0630
Social Structure and Development	0700
Theory and Methods	0344

Theory and Methods	0344
Transportation	0709
Urban and Regional Planning	0999
Women's Studies	0453

Agriculture	
General	0473
Agronomy	0285
Animal Culture and Nutrition	0475
Animal Pathology	0476
Food Science and Technology	0359
Forestry and Wildlife	0478
Plant Culture	0479
Plant Pathology	0480
Plant Physiology	7817
Range Management	0777
Wood Technology	0746

Geodesy	0370
Geology	0372
Geophysics	0373
Hydrology	0388
Mineralogy	0411
Paleobotany	0345
Paleoecology	0426
Paleontology	0418
Paleozoology	0985
Palynology	0427
Physical Geography	0368
Physical Oceanography	0415

General	0306
Anatomy	0287
Biostatistics	0308
Botany	0309
Cell	0379
Ecology	0329
Entomology	0353
Genetics	0369
Limnology	0793
Microbiology	0410
Molecular	0307
Neuroscience	0317
Oceanography	0418
Physiology	0433
Radiation	0821
Veterinary Science	0778
Zoology	0472
Biophysics	
General	0788
Medical	0760

Environmental Sciences	0768
Health Sciences	0769

General	0564
Audiology	0300
Chemotherapy	0992
Dentistry	0567
Education	0350
Hospital Management	0765
Human Development	0758
Immunology	0987
Medicine and Surgery	0564
Mental Health	0347
Nursing	0567
Nutrition	0570
Obstetrics and Gynaecology	0380

Occupational Health and	035
Therapy	038
Ophthalmology	057
Pathology	041
Pharmacology	057
Pharmacy	038
Physical Therapy	057
Public Health	057
Radiology	057
Recreation	057

Speech Pathology	0460
Toxicology	0383
Home Economics	0386

Pure Sciences

Chemistry	
General	048:
Agricultural	074:
Analytical	048:
Biochemistry	048:
Inorganic	048:
Nuclear	073:
Organic	049:
Pharmaceutical	049:
Physical	049:
Polymer	049:
Radiation	075:
Mathematics	040:

Mathematics	0403
Physics	
General	0603
Acoustics	0980
Astronomy and Astrophysics	0600
Atmospheric Science	0600
Atomic	0740
Electronics and Electricity	0600

Electronics and Chemistry	060
Elementary Particles and	
High Energy	079
Fluid and Plasma	075
Molecular	060
Nuclear	061
Optics	075
Radiation	075
Solid State	061
Spectroscopy	046

Applied Sciences	
Applied Mechanics	034
Computer Science	098

Engineering	
General	0537
Aerospace	0538
Agricultural	0539
Automotive	0540
Biomedical	0541
Chemical	0542

Civil	0543
Electronics and Electrical	0544
Heat and Thermodynamics	0348
Hydraulic	0545
Industrial	0546
Marine	0547
Materials Science	0799
Mechanical	0548
Metallurgy	0743
Mining	0551
Nuclear	0552
Packaging	0549
Petroleum	0765
Sanitary and Municipal	0554
System Science	0790

Geotechnology	0428
Operations Research	0796
Plastics Technology	0795
Textile Technology	0994

General	062
Behavioral	038
Clinical	062
Developmental	062
Experimental	062
Industrial	062
Personality	062
Physiological	098
Psychobiology	034
Psychometrics	063
Special	045

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

Tracking Performance of Acoustic Echo Cancellers

submitted by

Timothy G. D. Webster, B.Eng (Elec)

in partial fulfilment of the requirements
for the degree of Master of Engineering

Thesis Supervisor

Chair, Department of Systems and Computer Engineering

Carleton University

September 12, 1995

ABSTRACT

This thesis examines the tracking performance of acoustic echo cancellers operating in handsfree teleconferencing terminals in which the loudspeaker and microphone are located in a common enclosure. Acoustic echo cancellers are used in such a system to achieve full-duplex, high-quality handsfree communication, thereby mitigating the effects of near-end electro-acoustic instability. The performance of an acoustic echo canceller is shown to be limited by a multitude of physical factors including background acoustic noise, nonlinearities, nonstationarities, and undermodelling effects. The ability of an acoustic echo canceller to track nonstationarities in a typical conference room is seen to be a major limitation of practical systems.

This thesis examines the tracking performance of five adaptive filtering algorithms for acoustic echo cancellation using experimental data; the fullband NLMS, fullband ESNLMS, accelerated SFTF, subband NLMS, and subband ESNLMS algorithms. The fullband NLMS algorithm is shown to be relatively ineffective in tracking room nonstationarities. The accelerated SFTF algorithm, however, is capable of dramatically improved convergence performance especially when excited with speech.

Unfortunately, improvements in tracking performance of the SFTF algorithm requires the use of more complex techniques like periodic re-initialization or alternate windows. With the fullband and subband ESNLMS algorithms, tracking improvements of 55% (+7.4dB) and 102% (+15.3dB) respectively are demonstrated when excited with noise, while improvements of 12% (+1.8dB) and 39% (+5.7dB) are demonstrated with speech excitation. The subband ESNLMS algorithm is less sensitive to the colour in the input than the fullband ESNLMS algorithm and thus exhibits better tracking and convergence performance with speech excitation.

“Any sufficiently advanced technology is
indistinguishable from magic.”

Arthur C. Clarke,
from The Lost Worlds of 2001

To my wife Kelli, my parents,
and my brother Scott.

ACKNOWLEDGEMENTS

It is my pleasure to acknowledge those who played an integral role in the development and completion of this thesis. I would especially like to thank my supervisor Dr. Rafik Goubran for his support, and constant motivation. I am grateful for the assistance of Neil Birkett and Rajeev Bector in helping to design, build, and improve the experimental setup. To Jeff Lariviere, Jim Ryan, and especially Trevor Moat for the invaluable conversations we shared on adaptive filter structures. Danny Lemay's help in keeping the PC network in the DSP lab operational, and in letting us borrow *the cart* at a moments notice were equally invaluable. To Jack Kelly for giving us unlimited access to the anechoic chamber, and to Bell Northern Research for the speech files used in chapter 6. Finally to NSERC, Carleton University, and Bell Northern Research for their financial support.

Most importantly, to my wife Kelli for her patience and understanding during this difficult process.

TABLE OF CONTENTS

ABSTRACT	iii
ACKNOWLEDGEMENTS	vi
TABLE OF CONTENTS	vii
LIST OF FIGURES	ix
LIST OF TABLES	xiv
LIST OF ACRONYMS	xv
LIST OF SYMBOLS	xviii
1.0 INTRODUCTION	1
2.0 REVIEW OF TECHNIQUES AND PRINCIPLES.....	4
2.1 Acoustic Echo Cancellation.....	4
2.2 Linear Adaptive Filter Theory	6
2.3 The NLMS Algorithm	9
2.4 The Exponentially-Weighted RLS Algorithm	11
2.5 Tracking Performance of the LMS and RLS algorithms	12
2.6 NLMS Algorithm Enhancements	15
2.6.1 The ESNLMS Algorithm	15
2.6.2 Time Weighted NLMS Algorithm.....	16
2.6.3 Variable Step Size LMS Algorithms	17
2.6.4 The Transform Domain LMS Algorithm.....	19
2.6.5 Subband Adaptive Filters	23
2.7 RLS Algorithm Enhancements	25
2.7.1 The SFTF Algorithm	25
2.7.2 The FNTF Algorithm	29
2.7.3 Accelerated Fast-RLS Algorithms	32
2.7.4 The Normalized Stochastic Newton Algorithm	33
2.8 Tracking Performance of Enhanced Algorithms	34
2.9 Infinite Impulse Response Adaptive Filters.....	37
2.10 Summary	43
3.0 PROBLEM DEFINITION.....	44

4.0	EXPERIMENTAL SETUP AND PHYSICAL LIMITATIONS	46
4.1	Experimental Setup	46
4.2	Maximum Achievable ERLE of the Experimental Setup	48
4.3	Steady-State Performance Limitations	50
4.3.1	Limitations due to Speakerphone Design	50
4.3.2	Limitations due to Room Characteristics and Background Noise	55
4.4	Nonstationary Performance	58
4.5	Summary of Observations	59
5.0	BANDLIMITED NOISE RESULTS	61
5.1	Experimental Conditions	61
5.2	The Fullband NLMS Algorithm	63
5.3	The Accelerated SFTF Algorithm	66
5.4	The Fullband ESNLMS Algorithm	71
5.5	The Subband NLMS Algorithm	75
5.6	The Subband ESNLMS Algorithm	80
5.7	Complexity of the Subband ESNLMS Algorithm	88
5.8	Discussion and Summary	89
6.0	SPEECH RESULTS	90
6.1	Experimental Conditions	90
6.2	The Fullband NLMS Algorithm	93
6.2.1	Performance with Adaptation Halted	93
6.2.2	Effect of Step Size	94
6.3	The Accelerated SFTF Algorithm	96
6.4	The Fullband ESNLMS Algorithm	99
6.5	The Subband NLMS Algorithm	102
6.6	The Subband ESNLMS Algorithm	105
6.7	Discussion and Summary	107
7.0	CONCLUSIONS	109
7.1	Summary of Contributions	111
7.2	Suggestions for Future Work	111
	REFERENCES	113
	Appendices	122

LIST OF FIGURES

FIGURE 2.1	Illustration of the combination of a hybrid echo canceller (HEC) and an acoustic echo canceller (AEC) in a handsfree phone.	5
FIGURE 2.2	The configuration of an adaptive filter (ADF) for system identification.	7
FIGURE 2.3	Transversal or finite impulse response (FIR) digital filter structure.....	7
FIGURE 2.4	Convergence behaviour of the NLMS algorithm with coloured input signals.	10
FIGURE 2.5	Illustration of the classic trade-off between noise immunity and tracking speed of the LMS algorithm (also true in other algorithms). The optimum step size is that which minimizes the MSE.....	13
FIGURE 2.6	General transform-domain LMS (TDLMS) adaptive structure.....	21
FIGURE 2.7	Subband acoustic echo canceller structure	24
FIGURE 2.8	Implementation of the $12M+2N$ FNTF algorithm.....	30
FIGURE 2.9	Basic Structure of an Adaptive IIR filter based on the Output Error Formulation.....	38
FIGURE 2.10	Basic structure of an adaptive IIR AEC based on the equation error method.	39
FIGURE 2.11	Representation of the Gamma Delay Operator	40
FIGURE 2.12	The generalized feed-forward filter	40
FIGURE 2.13	CAPZ Model proposed by Haneda, Makino, and Kaneda.	42
FIGURE 4.1	Block diagram of the experimental setup	47
FIGURE 4.2	Optimum steady-state ERLE performance of the experimental setup based on the converged ERLE of a 10 tap NLMS adaptive filter with a step size of 0.5 (Linear FIR curve), and an estimate of the SNR of the primary signal (Asymptote curve).	49
FIGURE 4.3	Comparison of the optimum steady-state ERLE of the Ariel, DAT and Ariel, and full setup based on a 10 tap NLMS adaptive filter with a step size of 0.5.	50

FIGURE 4.4	Breakdown of the LREM impulse response, and AEC configuration.....	51
FIGURE 4.5	Anechoic impulse response of the NT phone.	52
FIGURE 4.6	Anechoic impulse response of the Mitel Superset 430.....	52
FIGURE 4.7	Anechoic impulse response of the Mitel Superset 410.....	53
FIGURE 4.8	Anechoic impulse response of the Telex CP268A.....	53
FIGURE 4.9	Anechoic impulse response of the Panasonic KX-T2315	54
FIGURE 4.10	Converged ERLE for five speakerphones at various sound pressure levels measured at 1m from the phone in an anechoic chamber.	55
FIGURE 4.11	LREM impulse response measured with the NT phone in Minto 3033	56
FIGURE 4.12	LREM impulse response measured with the NT phone in Minto 2014	56
FIGURE 4.13	Converged ERLE of the NT phone in Minto 2014 and Minto 3033 vs. adaptive filter order, $\alpha=1.0$.	57
FIGURE 4.14	ERLE performance of NT phone in Minto 2014 during long-term nonstationarity caused by a person moving in spot 1m directly in front of the phone.....	59
FIGURE 5.1	Effect of α on the tracking and convergence performance of the NLMS algorithm with 1200 taps and excited with bandlimited white noise.	63
FIGURE 5.2	Short-term power spectrum of error signal associated with NLMS algorithm computed from 0.5s to 1.5s (initial convergence), and from 9s to 10s (steady-state).	64
FIGURE 5.3	Short-term power spectrum of error signal associated with NLMS algorithm computed during re-convergence following the nonstationarity over the windows 5.5-6.5s and 6-7s, and in final steady-state from 9-10s.....	65
FIGURE 5.4	Effect of the forgetting factor on the ERLE performance of the SFTF algorithm excited with bandlimited white noise, $N=1200$, $\zeta=200$, and $\lambda=1$, $\lambda=1-1/10$, and $\lambda_{acc}=1-1/N$ ($\lambda=1-1/3N$, $\rho=0.95$).	67
FIGURE 5.5	ERLE performance of the SFTF algorithm with $N=1200$, $\lambda_{acc}=1-1/N$ ($\lambda=1-1/3N$, $\rho=0.95$) , and $\zeta=200$ compared with the NLMS algorithm with $\alpha=1$.....	69

FIGURE 5.6	Illustration of the improvement in convergence performance following the nonstationarity of the SFTF algorithm by purging its memory after the nonstationarity at 6.3s, with $N=1200$, $\lambda_{acc}=1-1/N$ ($\lambda=1-1/3N$, $\rho=0.95$), and $\zeta=200$.	70
FIGURE 5.7	Parameterizing the step size vector used in the ESNLMS algorithm	71
FIGURE 5.8	ERLE performance of the fullband ESNLMS algorithm, best tracking performance (trial #8, table 5.4), compared with the NLMS algorithm with $\alpha=1$	73
FIGURE 5.9	ERLE Performance of the ESNLMS algorithm, best convergence performance (trial #11, table 5.4), compared with the NLMS algorithm with $\alpha=1$	74
FIGURE 5.10	Frequency response of 8-band AFB. 128pt Blackman BPF, 40% overlap.	76
FIGURE 5.11	Frequency response of 8-band SFB. 128pt Blackman BPF	76
FIGURE 5.12	Overall frequency response of combined AFB+SFB system excluding the ADFs. 0.015dB ripple in passband, and -1 dB gain out-of-band near dc and 4kHz.	77
FIGURE 5.13	Stationary ERLE performance of 4-band, 8-band, and fullband NLMS vs # taps/band with $\alpha=1.0$ in all cases, and the NT phone excited with bandlimited white noise in Minto 2014.	77
FIGURE 5.14	ERLE Performance of a 4-band NLMS algorithm with 1200 taps/band, and $\alpha=1.0$ compared with the NLMS algorithm with $\alpha=1$	79
FIGURE 5.15	ERLE Performance of an 8-band NLMS algorithm with 1200 taps/band, and $\alpha=1.0$ compared with the NLMS algorithm with $\alpha=1$	79
FIGURE 5.16	Block diagram of the subband ESNLMS algorithm	81
FIGURE 5.17	Technique to estimate the decay factors in each band of a subband ESNLMS algorithm.	82
FIGURE 5.18	Estimated 8-band decay characteristics of the LREM impulse response measured in Minto 2014 with the NT phone.	83

FIGURE 5.19	ERLE Performance of the 8-band ESNLMS algorithm (trial #1, table 5.6) compared with the fullband NLMS algorithm with $\alpha=1$; N=1200.....	84
FIGURE 5.20	ERLE Performance of the 8-band ESNLMS algorithm (trial #4, table 5.6) compared with the fullband NLMS algorithm with $\alpha=1$; N=1200.....	85
FIGURE 5.21	ERLE Performance of 8-band ESNLMS (trial #1, table 5.6) compared with 8-band ESNLMS (trial #4, table 5.6).....	86
FIGURE 5.22	ERLE Performance of 8-band ESNLMS (trial #1, table 5.6) compared with fullband ESNLMS (trial #8, table 5.4) ; N=1200.....	86
FIGURE 5.23	ERLE Performance of 8-band ESNLMS algorithm (trial #1, table 5.6) compared with 8-band NLMS algorithm with $\alpha=1$; N=1200.....	87
FIGURE 6.1	Time-series plot of the speech excitation signal.	91
FIGURE 6.2	Spectrogram plot of the speech excitation signal.	92
FIGURE 6.3	Time-series plot of the complete excitation signal.	92
FIGURE 6.4	ERLE performance of a 1200 tap NLMS algorithm with parameters $\mu=1.0$, $\epsilon=10^{-6}$, $\psi=1/1200$ (halting enabled), compared with the same algorithm $\psi=0$ (halting disabled).....	93
FIGURE 6.5	ERLE performance of the NLMS algorithm with adaptation halted for $\alpha=0.1$ and $\alpha=1.0$ with N=1200, $\epsilon=10^{-6}$, $\gamma=1$.	95
FIGURE 6.6	ERLE Performance of SFTF and NLMS algorithms excited with speech.....	96
FIGURE 6.7	Tracking performance of SFTF and NLMS algorithms excited with speech.....	97
FIGURE 6.8	Step size vector used in the fullband ESNLMS algorithm in this chapter with speech.	100
FIGURE 6.9	ERLE performance of the fullband ESNLMS algorithm (trial #7, table 6.5) compared with the fullband NLMS algorithm with $\alpha=1$, $\psi=1$, and N=1200.	100
FIGURE 6.10	Tracking performance of the fullband ESNLMS algorithm (trial #7, table 6.5) compared with the fullband NLMS algorithm with $\alpha=1$, $\psi=1$, N=1200 (closeup view).....	102

FIGURE 6.11 ERLE performance of the 4-band NLMS algorithm with speech, 1200taps/ band, $\alpha=1.0$, $\psi=1.0$ and compared with the fullband NLMS algorithm with similar parameters.....	103
FIGURE 6.12 Tracking Performance of 4-band NLMS algorithm with speech, compared with the fullband NLMS algorithm (close-up view)	104
FIGURE 6.13 ERLE performance of 8-band NLMS algorithm with speech, 1200taps/ band, $\alpha=1.0$, $\psi=1.0$, compared with the fullband NLMS algorithm with similar parameters.....	104
FIGURE 6.14 Tracking performance of the 8-band NLMS algorithm with speech, compared with the fullband NLMS algorithm (close-up view).....	105
FIGURE 6.15 ERLE performance of the 8-band ESNLMS algorithm, 1200taps/band, $\psi=1.0$ (table 6.7), compared with the fullband NLMS algorithm with $\alpha=1$, $\psi=1$	106
FIGURE 6.16 Tracking performance of the 8-band ESNLMS algorithm with speech, compared with the fullband NLMS algorithm (close-up view).....	107

LIST OF TABLES

TABLE 4.1	List of speakerphones tested	51
TABLE 5.1	List of adaptive filtering algorithms investigated	62
TABLE 5.2	Quantitative ERLE performance of the NLMS algorithm excited with bandlimited white noise.	63
TABLE 5.3	Quantitative comparison of the ERLE Performance of the accelerated SFTF algorithm with bandlimited white noise.	68
TABLE 5.4	Quantitative comparison of the ERLE Performance of the ESNLMS algorithm with bandlimited white noise.	72
TABLE 5.5	Quantitative comparison of the ERLE Performance of the SBNLMS algorithm with bandlimited white noise.	78
TABLE 5.6	Quantitative comparison of the ERLE Performance of the 8-band ESNLMS algorithm with white noise.	84
TABLE 6.1	Effect of halting the adaptation on the ERLE performance of the NLMS algorithm.	94
TABLE 6.2	Effect of step size on ERLE performance of the NLMS algorithm with speech.	95
TABLE 6.3	Parameter list for the NLMS and SFTF algorithms.	97
TABLE 6.4	Quantitative comparison of the ERLE performance of the SFTF and NLMS algorithms with speech excitation.	98
TABLE 6.5	Quantitative comparisons of the ERLE Performance of the ESNLMS Algorithm with speech excitation.	101
TABLE 6.6	Quantitative comparison of the ERLE performance of 4 and 8 band NLMS algorithms with speech excitation.	103
TABLE 6.7	Quantitative comparison of the ERLE performance of the 8 band ESNLMS algorithm with speech excitation.	106

LIST OF ACRONYMS

4B	4 Band
8B	8 Band
ADF	Adaptive Digital Filter
AEC	Acoustic Echo Cancellers
AFB	Analysis Filter Bank
ASFTF	Accelerated SFTF algorithm
BPF	Band Pass Filter
CAPZ	Common Acoustical Pole Zero Model
DAT	Digital Audio Tape
DCT	Discrete Cosine Transform
DFT	Discrete Fourier Transform
DSP	Digital Signal Processing
EC	Echo Cancellor
ERLE	Echo Return Loss Enhancement
ESLMS	Exponentially decaying Step size LMS algorithm algorithm
ESNLMS	Exponentially decaying Step size NLMS algorithm algorithm
FB	Full Band
FFT	Fast Fourier Transform
FIR	Finite Impulse Response
FNTF	Fast Newton Transversal Filter algorithm
FPAF	Fixed Pole Adaptive Filter
HEC	Hybrid or line Echo Cancellers
IC	Intergrated Circuit

IIR	Infinite Impulse Response
KLT	Karhunen Loeve Transform
LMS	Least Mean Squares algorithm
LREM	Loudspeaker Room Enclosure Microphone system
MA	Multiply Accumulate
MMSE	Minimum Mean Square Error
MSE	Mean Square Error
NLMS	Normalized Least-Mean-Squares algorithm
NSN	Normalized Stochastic Newton Algorithm
NT	Northern Telecom
PC	Personal Computer
PO2	Power of Two Transform
QMF	Quadrature Mirror Filterbanks
RLS	Recursive Least Squares algorithm
SB	SubBand
SBNLMS	Subband NLMS algorithm
SBESNLMS	Subband ESNLMS algorithm
SFB	Synthesis Filter Bank.
SFTF	Stabilized Fast Transversal Filter algorithm
SG-GALS	Stochastic Gradient with Gradient Adaptive and Limited Step algorithm.
SG-GAS	Stochastic Gradient with Gradient Adaptive Step algorithm.
SNR	Signal to Noise Ratio
SPL	Sound Pressure Level
TDLMS	Transform Domain LMS algorithm
TWLMS	Time weighted LMS adaptive algorithm

TWNLMS	Time weighted NLMS adaptive algorithm
WHT	Walsh Hadamard Transform

LIST OF SYMBOLS

Notation.

A boldface letter refers to a matrix or a vector, while a plain letter refers to a scalar.

All vectors are column vectors.

Subscripts on boldface letters indicate the length of the vector (which generally refers to the order of the filter).

Superscripts on boldface letters indicate transposition (T), inversion (-1), or the extraction of a particular element from a vector resulting in a scalar quantity, (e.g.. $M+1$).

General Symbols.

$x(n)$	current reference (input) sample at time n .
$d(n)$	current desired or primary sample at time n .
$y(n)$	current output of ADF at time n .
$e(n)$	current a priori estimation error at time n .
$v(n)$	additive uncorrelated noise introduced by the unknown system.
$H(n)$	Impulse response of unknown system or plant at time n .
$W(n)$	Impulse response of echo canceller at time n .
W_o	Optimum (Weiner) coefficient vector.
$X_N(n)$	Vector of current and past input samples of order N at time n . $\{x(n) \ x(n-1) \ \dots \ x(n-N+1)\}^T$
$R(n)$	Autocorrelation matrix of reference (input) vector, $X_N(n)$, at time n .

$\mathbf{p}(n)$	Cross-correlation vector of reference and desired signals at time n .
J_{ex}	Excess mean-squared error.
J_{min}	Minimum mean-squared error.
J_{noise}	Excess mean-squared error due to noise.
J_{lag}	Excess mean-squared error due to tracking lag.
α	Normalized step size constant used in the NLMS algorithm.
ϵ	Scalar constant used in the NLMS algorithm to bound the step size during periods of low excitation.
$\mathbf{C}(n)$	Kalman gain vector used in the SFTF algorithm.
$\tilde{\mathbf{C}}(n)$	Dual Kalman gain vector used in the SFTF algorithm.
λ	Forgetting factor in the (accelerated) SFTF algorithm.
λ_{acc}	Accelerated forgetting factor, determined by λ and ρ .
ζ	Soft-initialization constant in the (accelerated) SFTF algorithm.
ρ	Acceleration constant in the accelerated SFTF algorithm.
$\gamma_N(n)$	Likelihood variable of order N (SFTF).
γ	Decay factor in ESNLMS algorithm.
M	Starting tap # of exponential decay in ESNLMS algorithm.
α_0	Initial step size of the ESNLMS algorithm.
α_{min}	Minimum step size allowed in the ESNLMS algorithm.

- ψ** Threshold used in halting NLMS algorithm. Set as fraction of maximum power in tapped delay line which is calculated as $\mathbf{X}_N(n)^T \cdot \mathbf{X}_N(n)$.
- T** Orthonormal transformation matrix used in TDLMS algorithm.

Chapter 1

1.0 INTRODUCTION

Handsfree telephones are desirable in a variety of applications from teleconferencing systems, to mobile cellular phones, and multimedia terminals. High-quality full-duplex handsfree communication is difficult to achieve, however. In these systems the loudspeaker and microphone are typically located away from the talker thereby requiring large signal gains to maintain comfortable volume levels. In some handsfree systems like desktop terminals, the loudspeaker and microphone are located in a common enclosure causing high acoustic coupling between the components. These issues degrade communication quality due to the presence of acoustic echo in the signal received by the far-end talker. A second problem with this configuration is the potential for near-end instability due to the high level of electro-acoustic feedback, component leakage, and the large signal gains present.

A variety of solutions have been proposed, and implemented, to make handsfree telephony a feasible technology. Voice-controlled switching, and echo suppression reduce talker echo, and instability by implementing a half-duplex or pseudo full-duplex link by severely attenuating the silent path [1] [65] [96]. Unfortunately, the quality of the transmitted speech is generally compromised at the switching instant. True full-duplex handsfree telephony may be possible, however, with echo cancellation technology. Echo cancellers model the impulse response of the echo path and synthesize a replica of the actual echo signal for cancellation [56].

Echo cancellers come in two varieties. Line or hybrid echo cancellers cancel the echoes which leak through imperfect hybrid devices on the line. Acoustic echo cancellers (AEC), however, cancel the acoustic echo received by the microphone from the loudspeaker. AECs are more complicated than hybrid echo cancellers mainly because the duration of the acoustic echo is typically at least an order of magnitude longer than a hybrid/line echo.

AECs must also deal with greater levels of uncorrelated noise, nonstationarities, and nonlinearities.

This thesis investigates the physical limitations of acoustic echo cancellers operating in handsfree desktop terminals. The design of the phoneset, the acoustic conditions of the environment, the level of background acoustic noise in the room, and movements near the phone are shown to limit ERLE performance.

The remainder of the thesis investigates the nonstationary or tracking performance of acoustic echo cancellers. Results are based on experimental data captured in realistic operating conditions. The performance of the fullband NLMS algorithm is considered the baseline reference to which other algorithms are compared.

The fullband Exponentially-decaying Step size NLMS (ESNLMS) algorithm is examined to improve the tracking performance of AEC in noise and speech conditions based on work by Yuan [105] with noise signals. The accelerated Stabilized Fast Transversal Filter (SFTF) algorithm is then investigated based on the promising work of Petillon, Gilloire, and Theodoridis [106] in mobile AEC. Based on these results, a subband Exponentially-decaying Step size NLMS algorithm is proposed and shown to provide very good tracking performance.

This thesis is organized as follows:

Chapter 2 discusses the state-of-the-art in acoustic echo cancellers, including a discussion of adaptive filtering algorithms and structures.

Chapter 3 defines the focus of the thesis, and states the objectives.

Chapter 4 describes the experimental setup and conditions, and identifies the main physical limitations to achievable ERLE performance.

Chapter 5 investigates the tracking performance of several linear adaptive filtering algorithms in band-limited white noise conditions.

Chapter 6 repeats the above analysis with speech excitation.

Chapter 7 presents concluding remarks, a summary of contributions, and ideas for future work.

Chapter 2

2.0 REVIEW OF TECHNIQUES AND PRINCIPLES

The following topics are reviewed in this chapter:

- Acoustic echo cancellation.
- Adaptive filtering for system identification.
- Linear adaptive transversal filtering based on the LMS and RLS families of adaptive algorithms.
- Tracking properties of the LMS and RLS algorithms.
- General advances in FIR adaptive filtering.
- IIR adaptive filtering.

2.1 Acoustic Echo Cancellation

Two echo cancellers are required in a handsfree phone to prevent talker echo and near-end instability; a hybrid or line echo canceller (HEC) [11] [13] [56] [77] and an acoustic echo canceller (AEC) [2] [49] [52] [57] [65] as illustrated in figure 2.1. The AEC models the acoustic echo path between the loudspeaker and the microphone to reduce talker echo, while the HEC compensates for leakage through the hybrid and reflections on the line to prevent possible near-end instability in the feedback loop.

An AEC is more complicated to design than a HEC for the following reasons:

- The duration of an acoustic echo in a typical conference room is an order of magnitude longer than a hybrid or line echo. AEC must be capable of cancelling echoes of several hundred milliseconds in duration.

- While the characteristics of the hybrid echo path are stationary on a per call basis, the characteristics of the acoustic echo path changes with movements in the environment. AECs must be capable of tracking these changes.
- The hybrid echo path is generally well modelled by a linear system. The acoustic echo path, however, is complicated by the presence of nonlinearities introduced by the loudspeaker and the rattling of objects within and attached to the enclosure, as well as components due to air turbulence and whistling through openings in the enclosure in the vicinity of the microphone.
- AECs must also deal with potentially high levels of uncorrelated background acoustic noise picked up by the microphone. In HEC the noise level is much less severe.

FIGURE 2.1 Illustration of the combination of a hybrid echo canceller (HEC) and an acoustic echo canceller (AEC) in a handsfree phone.

Echo Return Loss Enhancement (ERLE) is a common method used to compare the performance of adaptive filtering algorithms. ERLE defines the difference in average power between the primary and error signals, as illustrated by equation (2.1). The expectation operator in equation (2.1) is typically replaced by a 256-point moving-average window as shown in equation (2.2). The latter equation is used in this thesis.

$$ERLE(dB) = 10 \cdot \log \left(\frac{E(d^2(n))}{E(e^2(n))} \right) \quad (2.1)$$

$$ERLE(dB) = 10 \cdot \log \left(\frac{\sum_{i=0}^{255} d^2(n-i)}{\sum_{i=0}^{255} e^2(n-i)} \right) \quad (2.2)$$

State-of-the-art AECs appear to be limited to a steady-state ERLE of 25dB, while greater than 40dB is desired in practice¹. HECs readily achieve a steady-state ERLE greater than 60dB, however [11] [13].

2.2 Linear Adaptive Filter Theory

Adaptive filter theory, and adaptive signal processing in general, has found widespread use in applications ranging from voiceband data modem to digital satellite transmission and speech compression [20] [26] [27] [56] [77] [78]. Most applications of adaptive filters can be grouped into two main categories; *system inversion* and *system identification*. The *system identification* configuration of an adaptive filter is discussed in the paragraphs below. The *system inversion* category, however, is beyond the scope of this thesis.

1. Based on conversations with Bell Northern Research.

FIGURE 2.2 The configuration of an adaptive filter (ADF) for system identification.

The *system identification* configuration of an adaptive filter is shown in figure 2.2. Understanding the operation and limitations of this very basic structure is important because it closely resembles the configuration of an acoustic echo canceller in a handsfree phone as shown in figure 2.1. The desired signal, $d(n)$, is equivalent to the signal received by the microphone which includes the acoustic echo signal, near-end speech, and background acoustic noise. The adaptive filter's internal parameters are selected to approximate the impulse response of the unknown system. This makes it possible to subtract an estimate of the actual acoustic echo signal from the desired signal to achieve cancellation. The adaptive filter is usually realized with a transversal or finite impulse response digital filter (FIR) for simplicity and stability reasons as shown in figure 2.3.

FIGURE 2.3 Transversal or finite impulse response (FIR) digital filter structure

The coefficients of the adaptive filter are generally selected to minimize the average power or mean-squared value of the a priori estimation error given by the *minimum mean-squared error criterion* (MMSE), $J_{\min}(\mathbf{W}_0) = E(e^2(n))$. The optimum coefficient vector, $\mathbf{W}_o$, can be found mathematically by solving the so-called *normal equation* given by $\mathbf{R} \cdot \mathbf{W}_o = \mathbf{p}$. The matrix $\mathbf{R}$ must be invertible and hence nonsingular for a unique $\mathbf{W}_o$ to exist [56].

A rigorous implementation of the normal equation is impractical because:

1. it requires a priori knowledge of the signal statistics, $\mathbf{R}$ and $\mathbf{p}$,
2. it involves matrix inversion which requires $O(N^3)$ multiplications with classical Gaussian elimination, and
3. in nonstationary environments $\mathbf{W}_o$ has to be computed repeatedly.

In real-time applications $\mathbf{W}_o$ must be computed on a sample-by-sample basis without a priori knowledge of the signal statistics, or requiring matrix inversion [77].

A relatively simple practical solution to this problem is the stochastic gradient approach, which forms the basis of the LMS algorithm and derivatives. Instead of inverting $\mathbf{R}$, an iterative, stochastic search methodology is employed in which the coefficient vector is updated in the direction of the steepest-descent (given by the negative gradient) of the error performance surface. The LMS algorithm has a computational complexity of $O(N)$ (which is low for adaptive algorithms), but generally suffers from poor dynamic performance especially when excited with highly coloured signals.

A second technique, which claims faster dynamic performance, is the RLS-family of algorithms. In this approach the normal equation is implemented directly in real-time by applying the so-called matrix inversion lemma to reduce computational complexity from $O(N^3)$ to $O(N^2)$. Fast techniques are available to reduce complexity to $O(N)$ without compromising dynamic performance.

2.3 The NLMS Algorithm

One of the simplest and therefore most widely used adaptive algorithm is the LMS algorithm introduced by Widrow and Hoff in 1960. The coefficients of the LMS algorithm are updated in the direction of steepest-descent of the error performance surface by an amount determined by the step size parameter. A large step size improves convergence and tracking performance but degrades steady-state performance, while a small step size improves steady-state performance at the expense of dynamic performance. In nonstationary environments, a large step size is needed to maximize convergence and tracking performance. This requires knowledge of the maximum expected power in the input signal to suitably bound the step size and prevent instability and divergence of the coefficients. In applications in which the dynamic range of the input signal is large, which is typical with speech, the maximum input power and thus the step size must be estimated conservatively, significantly degrading dynamic performance.

The NLMS algorithm, given by equations (2.3) - (2.5), solves this problem by normalizing the step size to the power in the input delay line at each iteration. It is deterministically stable at *every* sample instant provided the normalized step size, α , is less than two [103]. Convergence performance is maximized by setting α equal to one. The convergence behaviour of the NLMS algorithm depends quite strongly on the eigenvalue spread of the input autocorrelation matrix, R , as shown in figure 2.4. White noise excitation leads to a constant rate of convergence, while coloured excitation (eigenvalue spread > 1) leads to a faster rate of convergence initially but a much slower rate beyond a certain MSE threshold [103]. Whether or not the rapid reduction in MSE achieved before the threshold is useful or not depends on the application and the steady-state MSE required.

The NLMS Algorithm

Initialization. $W(n) = 0$ (2.3)

Algorithm. $e(n) = d(n) - W(n)^T \cdot X_N(n)$ (2.4)

$$W(n+1) = W(n) + \frac{\alpha \cdot e(n) \cdot X_N(n)}{\epsilon + \|X_N(n)\|^2} \quad (2.5)$$

where.

- α is the normalized step size constant.
- ϵ is a small positive constant used to place a lower bound on the input signal power.

Stability. $0 < \alpha < 2$.

FIGURE 2.4 Convergence behaviour of the NLMS algorithm with coloured input signals.

2.4 The Exponentially-Weighted RLS Algorithm

The Recursive Least-Squares (RLS) algorithm computes the optimum coefficient vector at each iteration by solving the *normal equation* directly. The RLS algorithm is often considered a high-performance adaptive filtering algorithm because it *generally* converges significantly faster than stochastic-gradient based approaches, especially with coloured input signals. To achieve good tracking performance, however, the standard RLS algorithm must be modified slightly to include an exponential or sliding window such that effects of past data are reduced over time. The *exponentially weighted* RLS algorithm is given by equations (2.6) - (2.10). [56] [77]

The Exponentially-Weighted RLS Algorithm

Initialization.

$$R^{-1}(0) = \delta^{-1} I, \quad \delta^{-1} = \text{large positive constant.}$$

$$W(n) = 0 \tag{2.6}$$

Algorithm.

$$e(n) = d(n) - W^T(n-1) X_N(n) \tag{2.7}$$

$$C(n) = \frac{\lambda^{-1} R^{-1}(n-1) X_N(n)}{1 + \lambda^{-1} X_N^T(n) R^{-1}(n-1) X_N(n)} \tag{2.8}$$

$$W(n) = W(n-1) + C(n) e(n) \tag{2.9}$$

$$R^{-1}(n) = \lambda^{-1} R^{-1}(n-1) - \lambda^{-1} C(n) X_N^T(n) R^{-1}(n-1) \tag{2.10}$$

where. • $C(n)$ is the $N \times 1$ gain vector, sometimes referred to as the Kalman gain vector.

- λ is the forgetting factor.

Stability.

$$0 < \lambda \leq 1.$$

2.5 Tracking Performance of the LMS and RLS algorithms

An AEC must *not* only converge quickly, it must be able to track nonstationary conditions as well. There is little dispute that adaptive algorithms designed to minimize the LS criteria (RLS family) converge more quickly than those which use the simple gradient-descent approach (LMS family), especially in conditions of low noise and high eigenvalue disparity. Whether the RLS algorithm has better tracking performance than the LMS algorithm remains a disputed issue. [25] [33] [74] [105] [56] [77]

A nonstationary condition exists when the signal statistics, $\mathbf{R}$ and $\mathbf{p}$, change with time. In AECs, and system identification in general, two types of nonstationary conditions can exist.

1. When the statistics of the unknown system changes with time, the cross-correlation vector, $\mathbf{p}$, is time-varying.
2. When the statistics of the reference signal changes with time, both the autocorrelation matrix, $\mathbf{R}$, and the cross-correlation vector, $\mathbf{p}$, are time-varying.

The first type is caused by variations in the acoustic environment caused predominantly by the movement of people in the room. Since $\mathbf{R}$ is unaffected, the eigenvalues and eigenvectors are constant and the shape of the *error performance surface* is fixed including $J_{\min}$. Only the location of the optimum is time-varying [12]. The second type is caused by the use of nonstationary signals like speech as the reference signal. This causes the shape of the *error performance surface* to change as well. In both cases the adaptive filter must not only seek out the optimum solution but must also track its changes.

The tracking performance of any adaptive algorithm can *not* be improved without degrading noise immunity (and hence steady-state performance) due to conflicting requirements on the width of the implicit data averaging window. The classic equation for misadjustment in a nonstationary environment is given by $J_{ex} = J_{noise} + J_{lag}$ [56]. Increasing the misadjustment due to noise, J_{noise} , decreases the misadjustment due to tracking lag, J_{lag} , and

vice versa, as shown in figure 2.5. Noise immunity and tracking performance are controlled by the step size and forgetting factor in the LMS and RLS algorithms respectively. Closed-form solutions to minimize MSE performance do not exist, however, making this type of optimization simply a trial-and-error exercise [12] [25] [77].

FIGURE 2.5 Illustration of the classic trade-off between noise immunity and tracking speed of the LMS algorithm (also true in other algorithms). The optimum step size is that which minimizes the MSE.

Eleftheriou and Falconer [25] derived theoretical results for the misadjustment of the LMS and RLS algorithms in simulated nonstationary conditions and proved the following general statements:

1. the tracking performance of the RLS algorithm is independent of the type of nonstationarity, the nature of the application, and the colour of the reference signal.
2. the tracking performance of the LMS algorithm is dependent on the colour of the reference signal.

This lead them to conclude the following points on the tracking performance of each algorithm:

1. the tracking performance of the RLS algorithm is *at least as good* as the LMS algorithm.
2. when the excitation signal is coloured, the RLS algorithm *should* track nonstationarities more effectively than the LMS algorithm, especially when the filter's order is large.
3. when the input is white, the RLS algorithm may *not* track any better than the LMS algorithm unless the optimum step size in the LMS algorithm violates the stability bound.

In chapter 15 of [77], Cioffi and Byun present a concise comparison of the tracking performance of the LMS and RLS algorithms. By expanding on Eleftheriou and Falconer's work [25] they show that the RLS algorithm is capable of tracking movements in a room more effectively than the LMS algorithm when the time-variation in $\mathbf{W}_0$ is the dominant effect provided the noise level is small. In situations in which the magnitude of the (uncorrelated) noise added by the room is larger than the magnitude of the time-variation in $\mathbf{W}_0$, the LMS and RLS algorithms are equivalent in tracking performance.

Knappe [74], and Yuan [105] compared the LMS and RLS algorithms in the application of acoustic echo cancellation in handsfree phones operating in anechoic and real conference rooms. Both found that the NLMS algorithm met or exceeded the tracking and steady-state performance of the SFTF algorithm (simply a fast RLS algorithm) for white noise excitation. Knappe also found the NLMS algorithm to out-perform the SFTF algorithm in real speech excitation conditions. The apparent contradiction with Eleftheriou and Falconer's [25] and Cioffi and Byun's [77] results might be explained in terms of the threshold MSE level illustrated in figure 2.4. The ERLE levels were all relatively small (high MSE) such that the NLMS algorithm was operating where convergence was relatively fast for coloured signals. At higher ERLE (smaller MSE) different results may have been obtained as illustrated in figure 2.4.

2.6 NLMS Algorithm Enhancements

2.6.1 The ESNLMS Algorithm

A variety of improvements to the NLMS have been proposed in the literature to deal with convergence and tracking problems in real applications. One of the most successful improvements is the *exponentially-decaying step size* NLMS (ESNLMS) algorithm, proposed by Makino et.al. [92], and further examined by Yuan [105]. The ESNLMS algorithm takes advantage of the decay characteristics of typical room impulse responses to improve convergence and tracking performance with white noise and speech inputs. The scalar step size, α , in equation (2.5) of the NLMS algorithm is replaced by a diagonal matrix, A , in which the diagonal elements are chosen to decay at an exponential rate given by the reverberation time of the room (or the decay of the impulse response). Effectively, the coefficients associated with large errors are adapted in larger steps, while those associated with smaller errors are adapted in smaller steps. The ESNLMS algorithm is defined by equations (2.11) to (2.15).

The ESNLMS Algorithm

Initialization. $W(n) = 0$ (2.11)

Algorithm. $e(n) = d(n) - W(n)^T \cdot X_N(n)$ (2.12)

$$W(n+1) = W(n) + \frac{A \cdot e(n) \cdot X_N(n)}{\epsilon + \|X_N(n)\|^2} \quad (2.13)$$

where: $A = \text{diag}(\alpha_0, \alpha_1, \dots, \alpha_{N-1})$ (2.14)

$$\alpha_i = \alpha_0 \cdot e^{\gamma \cdot i}, \quad i = 0, \dots, N-1 \quad (2.15)$$

- A defines the step size matrix which approximates the exponentially-decaying characteristics of the room's impulse response.
- α_0 is the initial step size, and γ is the decay factor.

Stability. [92]

For white noise: $0 < \bar{\alpha} < 2$. No bound on the individual steps, only the average.
 For speech: $0 < \alpha_i < 2$, $i=0, N-1$. Each step is bounded.

The ESNLMS improves convergence and tracking performance by a factor of three for white noise and a factor of two for speech over the NLMS algorithm [92]. Yuan [105] also investigated the use of a time-varying step size matrix, $A(n)$, but demonstrated only a minor improvement in performance.

2.6.2 Time Weighted NLMS Algorithm

Kuo and Chen [91] appear to have taken the basic idea of the ESLMS algorithm and developed the Time Weighted LMS (TWLMS) algorithm. In TWLMS the N filter coefficients are first subdivided into M sections of length $N_1, N_2, N_3, \dots, N_M$. An updating order is then selected. During any given sampling instant only one section is updated. The selection of the updating order is dependent on the environment as is the decay factor, γ , in ESLMS, and determines how much time is devoted to the update of each section. Normalizing the step size matrix of TWLMS gives the TWNLMS. Results presented by the authors show that the TWLMS algorithm converges more quickly than the LMS algorithm (similar to ESLMS) at a fraction of the complexity.

2.6.3 Variable Step Size LMS Algorithms

The trade-off between steady-state and dynamic performance is classic to all adaptive algorithms. In time-invariant schemes, steady-state performance can *not* be improved without sacrificing dynamic performance, and vice versa. With the LMS algorithm, steady-state performance is improved by reducing the step size since this is equivalent to improving noise immunity through use of a longer data averaging window. Dynamic performance, however, is improved by increasing the step size.

One possible solution to this problem involves the use of a time-varying step size in which the step size is reduced during stationary periods to improve steady-state performance, and increased during nonstationary periods to improve dynamic performance [8]. The problem then becomes a matter of detecting the presence of the nonstationarity. A variety of time-varying algorithms have been developed in the literature for data echo cancellation which generally are not applicable to the AEC application. In these schemes the step size is set to an arbitrarily large value initially for a good initial rate of convergence, and gradually reduced as the adaptation proceeds to improve noise immunity for good steady-state performance [103]. Unfortunately these algorithms are generally not equipped with a mechanism to suddenly increase the step size during nonstationary periods, a particularly important property in acoustic echo cancellation systems.

Fortunately, a variety of algorithms employing time-varying step size sequences have been developed by researchers to accommodate nonstationary operating conditions [22] [73] [84] [100] [102]. In [22] the step size is increased or reduced by a factor, α , depending on the polarity of successive samples of the estimation error. [102] is similar to [22] except that the step size is updated through simpler addition and subtraction operations instead of multiplication and division. In [73] the step size update is based on the magnitude of the squared estimation error: $\mu(n+1) = \alpha\mu(n) + \gamma e^2(n)$. In all cases, upper and lower step size thresholds are used to maintain stability and reasonable tracking performance. The performance of each of these algorithms is highly sensitive to the selection of the

parameters which control the progression of the step size sequence, however. Furthermore, the optimal choice of these parameters is data dependent.

Unfortunately the performance of these VSS algorithms is affected by the presence of noise. One potential solution to this problem was proposed by Mayyas and Aboulnasr in [112]. Their algorithm is based on the simple fact that near the optimum not only is the error energy small, but the correlation between adjacent samples is also small. Such a correlation function is also less sensitive to noise than the energy in the error signal. The proposed algorithm, therefore, updates the step size in proportion to a time-averaged estimate of the autocorrelation between successive error samples. Improved performance is demonstrated with simulated results.

Mathews and Xie [100] proposed a time-varying step size algorithm in which the step size is updated in proportion to the negative gradient of the squared estimation error with respects to the previous step size. In effect the step size parameter is adapted in much the same way as the coefficients themselves. The algorithm is known as the *Stochastic Gradient Algorithm with Gradient Adaptive Step Size* (SG-GAS) and is described by equations (2.16)- (2.18) with $\rho > 0$.

$$e(n) = d(n) - W^T(n) \cdot X_N(n) \quad (2.16)$$

$$\mu(n) = \mu(n-1) + \rho \cdot e(n) \cdot e(n-1) \cdot X_N^T(n-1) \cdot X_N(n) \quad (2.17)$$

$$W(n+1) = W(n) + \mu(n) \cdot e(n) \cdot X_N(n) \quad (2.18)$$

The authors compared the performance of the SG-GAS algorithm with that of the NLMS algorithm and demonstrated improved steady-state performance without affecting dynamic performance in simulated conditions.

Recently, Sugiyama [84] demonstrated that the performance of the SG-GAS algorithm was affected by the level of uncorrelated noise in the environment. The use of soft-clipping

thresholds were proposed in the step size update equation to suppress the effects of noise and interference. The algorithm, called the *Stochastic Gradient Algorithm with Gradient Adaptive and Limited Step Size (SG-GALS)*, is derived by replacing (2.17) with (2.19)-(2.21).

$$\mu(n) = \mu(n-1) + \Delta\mu(n) \quad (2.19)$$

$$\Delta\mu(n) = \begin{cases} ThU(n) = \alpha \cdot \mu^2(n-1) & \text{for } \Delta\hat{\mu}(n) > ThU(n) \\ ThL(n) = -\beta \cdot \mu^2(n-1) & \text{for } \Delta\hat{\mu}(n) < ThL(n) \\ \Delta\hat{\mu}(n) & \text{otherwise} \end{cases} \quad (2.20)$$

$$\Delta\hat{\mu}(n) = \rho \cdot e(n) \cdot e(n-1) \cdot X_N^T(n-1) \cdot X_N(n) \quad (2.21)$$

The effect of (2.20) is to limit the change in $\mu(n)$ to a fraction of the square of $\mu(n-1)$ given by the constants α and β . This formulation was found to provide better performance than that obtained by simply imposing a limit that was linearly proportional to the step size.

2.6.4 The Transform Domain LMS Algorithm

The convergence rate of the NLMS algorithm is asymptotically bounded by the smallest eigenvalue of the autocorrelation matrix of the input signal [56]. This simply means that the convergence rate of the NLMS algorithm is, generally, quite slow when the input signal is coloured, and at its fastest when the input signal is white noise. One approach to improve the convergence performance of the NLMS algorithm with coloured signals is to excite the algorithm with a whitened version of the actual input signal. This is the basis of the *Transform Domain LMS (TDLMS)* algorithm described in this section [10] [23] [30] [47] [70] [97].

The TDLMS algorithm is described by equations (2.22) - (2.28) and illustrated in figure 2.6. The input vector $X_{N(n)}$ is first transformed into a new set of uncorrelated samples, $V(n)$, by a predefined transformation matrix, T (discussed in the next paragraph). The

power in each sample is then estimated using an exponential window and the ADF coefficients are adapted using different step sizes based on an estimate of the power that excites each coefficient. This is similar to a parallel bank of N one tap ADFs.

The TDLMS Algorithm

Initialization. $W(n) = 0$ (2.22)

Algorithm. $V(n) = T \cdot X_N(n)$ (2.23)

$$e(n) = d(n) - (W^T(n) \cdot V(n)) \quad (2.24)$$

$$W(n+1) = W(n) + \alpha \cdot e(n) \cdot \Gamma^{-1} \cdot V(n) \quad (2.25)$$

$$\Gamma^{-1} = \text{diag}(\sigma_{v_1}^2(n), \sigma_{v_2}^2(n), \dots, \sigma_{v_N}^2(n)) \quad (2.26)$$

$$\sigma_{v_i}^2(n) = \beta \cdot \sigma_{v_i}^2(n-1) + (1-\beta) \cdot v_i^2(n) \quad (2.27)$$

where: $\beta = 1 - \frac{1}{N}$ (2.28)

- T is any $N \times N$ unitary transformation matrix.
- $V(n)$ is an $N \times 1$ vector of transformed reference samples.

$$V(n) = \{v_1(n) \ v_2(n) \ \dots \ v_N(n)\}^T$$

- Γ^{-1} normalizes the step size to the power in each band.
- α is the adaptive step size.

Stability. $0 < \alpha < 2$

FIGURE 2.6 General transform-domain LMS (TDLMS) adaptive structure.

In general, the $N \times N$ transformation matrix, T , can be any orthogonal matrix. A matrix is orthogonal when $T^T T = M I$, where M is a scalar constant that typically equals one, and I is the $N \times N$ identity matrix. Some of the more commonly used transformations include the Discrete Fourier Transform (DFT), the Discrete Cosine Transform (DCT), the Walsh Hadamard Transform (WHT), and the Power-of-two transform (PO2) [4]. The optimal transform for the purpose of whitening a signal is the Karhunen-Loeve Transform (KLT). The KLT is generally considered too complex for real-time implementation, and is usually approximated by a sub-optimal transform [4].

Marshall, Jenkins, and Murphy [47] compared a number of real-valued suboptimal unitary transformations (DFT, DCT, WHT, DHT, and PO2) on their ability to improve the performance of LMS based adaptive filters. No single (suboptimal) transformation proved optimal with all types of input signals. The best transform for a certain application was found to depend almost entirely on the characteristics of the input signal. Even very simple transforms like the WHT in which all of its elements are either +1 or -1 out-performed more complex transforms like frequency domain approaches based on the DFT [5] [6] [21] [29] [32] [34] [69] [93] with certain input signal types.

Understanding the operation of the TDLMS algorithm is relatively straight-forward from an error surface description [47]. The optimal Karhunen-Loeve transform (KLT) rotates the axes of the error performance surface to achieve alignment with the principal axes of the hyper-ellipsoidal equal-error contours, corresponding to the eigenvectors of $\mathbf{R}$. This is equivalent to the well-known LMS/Newton (LMSN) algorithm [16] [97], and has the advantage that its convergence rate is independent of the degree of *colour* in the input signal. Sub-optimal transforms like the DFT or DCT still rotate the error surface but do not achieve perfect alignment with the principal axes.

Since unitary transforms simply rotate the error surface without affecting its shape (eigenvalues), their convergence rate is not improved unless power normalization is also employed. Power normalization refers to the process of optimizing the step size in each band to its current power level (accomplished by Γ^{-1} in (2.26)) such that coefficients associated with small eigenvalues are updated in large steps, while those associated with large eigenvalues are updated in smaller steps. This has the effect of reshaping the hyper-ellipsoidal error surface into a hyper-spherical error surface to match that of white noise. The realignment stage is necessary to uncouple the modes of the adaptive filter [47].

Although fast algorithms exist for computing most of the suboptimal transforms (DFT, and WHT), researchers have recently investigated the use of a technique known as *structured subband decomposition* to further reduce the complexity of TDLMS [98] [99] [101] (which requires an additional N^2 multiply-adds over the LMS algorithm). In *structured subband decomposition* the order of the transformation matrix, $\mathbf{T}$, is reduced to a size $M \times M$, where $M < N$, and adaptive sparse FIR filters are used in each band to make the total number of coefficient equal to N . This reduces complexity to an additional M^2 multiply-adds over the LMS algorithm, representing a tremendous computational savings for M much less than N . The performance of this technique in adaptive line enhancers, AEC, and adaptive channel equalization applications using the DFT was recently demonstrated in [98]. In [101] a similar comparison is made using the WHT. Improvements in the convergence performance of AEC in mobile applications was demonstrated in both cases. Surprisingly,

the difference in performance between a partial decomposition and a full decomposition was minor.

A common problem with all TDLMS algorithms based on time-invariant transforms is that their performance is dependent on the statistics of the input signal. This fundamental fact has prompted researchers to investigate time-varying transformations for use in the TDLMS algorithm [67] [97]. These techniques all approximate the optimal signal-dependent KLT transformation. In fact, the RLS algorithm can be viewed as one possible implementation of the KLT transform because it compensates for eigenvalue disparity in the input by computing $\mathbf{R}^{-1}$ at each iteration.

2.6.5 Subband Adaptive Filters

Adaptive filtering in subbands is a relatively recent development which closely parallels the TDLMS algorithm discussed in the previous section. The idea of subband adaptive filtering was popularized in a German paper published by Kellerman in 1985 [19], and later by Gilloire in 1987 [28], and again by Kellerman in 1988 [39]. In subband adaptive filtering the system's input and output signals are separated into independent frequency bands by an analysis filter bank (AFB), adapted individually, and then recombined into a single error signal with a synthesis filter bank (SFB) as shown in figure 2.7. Subband adaptive filters have been proposed to both reduce the complexity and improve the convergence rate of the LMS adaptive filter [94].

Etter, Carlson, and Margo [90] investigated critically sampled subband adaptive filters with perfect reconstruction filterbanks in terms of the model generated by the multirate system. They showed that subband adaptive filters with critical sampling can accurately identify an unknown system.

Gilloire and Vetterli [75] investigated the use of pseudo-QMF AFB/SFB with critical sampling using a truncated real room response with noise and speech signals but found that it converged more slowly than the full-band LMS algorithm. Conceivably, the inclusion of

adaptive cross-terms, which they proved were necessary to eliminate aliasing, and the large degree of spectral overlap in their AFB, degraded convergence performance.

FIGURE 2.7 Subband acoustic echo canceller structure

Yang, Cho, and Lee [89] examined the convergence behaviour of subband adaptive filters from a theoretical point-of-view based on error surface characteristics. They demonstrated that subbanding effectively whitens a signal by normalizing the step size to the power in different frequency bands. Improvements in convergence performance was demonstrated using simulated data. Unfortunately, they appear to have ignored the effects of aliasing distortion predominantly among adjacent subband in their analysis, and so failed to consider the use of adaptive cross-terms which Gilloire and Vetterli proved were necessary [75].

Ryan [71] and Wallace [59] investigated the performance of subband adaptive filters for noise cancellation and demonstrated improved rate of convergence with coloured signals over full-band algorithms. Ryan [71] also compared two techniques to reduce the complexity of subband filters. First, when modulated filterbanks were used such that a lowpass prototype lowpass filter was modulated to higher frequencies for full band coverage, the subband signals were decimated, reducing complexity by reducing the subband sampling rates. Decimation, however, lead to a reduction in convergence

performance. When the filterbanks were based on a parallel bank of bandpass filters, the subband signals could *not* be decimated because the subband signals were not bandlimited. Instead, sparse FIR filter were used in each subband to reduce complexity [15]. This caused a reduction in steady-state performance, however.

Petraglia and Mitra [88] investigated the oversampled case in which the subbands were not critically sampled. By oversampling the subband signals the authors were able to eliminate the adaptive cross-terms and improve convergence performance.

Unfortunately, the asymptotic convergence rate of oversampled subband adaptive filters is slow. As De Leon and Etter [116] demonstrated, this is caused by slow modes associated with the transition band region of the subband signals. By simply increasing the bandwidth of the analysis filterbank (AFB) asymptotic convergence rate is improved because the effects of these slow modes become attenuated by the SFB filter.

2.7 RLS Algorithm Enhancements

This section reviews methods proposed in the literature to reduce the complexity and improve the tracking performance of the exponentially-weighted RLS algorithm. Other RLS variants including QR decomposition, block RLS, and RLS lattice algorithms are discussed in [105].

2.7.1 The SFTF Algorithm

The complexity of the direct form exponentially-weighted RLS algorithm presented in section 2.4 can be reduced by taking advantage of redundancies in the standard RLS algorithm. One such algorithm is the Fast Transversal Filter (FTF) algorithm discussed in [14] [24]. Unfortunately, the standard FTF algorithm is plagued by numerical instability problems [35]. Stabilizing the FTF algorithm is an active area of research [14] [37] [58] [60] [68]. The technique used in this thesis is based on the method of error feedback

proposed in [37] and [60], in which complexity is increased from $7N$ to $8N$ operations per iteration. The SFTF algorithm typically yields a significant increase in convergence performance over the NLMS algorithm when coloured signals are involved.

The SFTF algorithm used in this thesis was proposed by Gilloire and Petillon in [51] and is described by equations (2.29) to (2.44):

The 8N-SFTF Algorithm

Initialization.

$$\alpha(0) = \zeta \cdot \lambda^N$$

$$\beta(0) = \zeta$$

ζ is the initial input variance and should be set large enough to prevent start-up divergence.

$$\gamma_N(0) = 1$$

$$X_N = A_N = B_N = \tilde{C}_N = H_N = 0$$

$$\tilde{C}_{N+1} = 0 \tag{2.29}$$

Phase 1: Prediction Part.

- Compute the forward prediction error.

$$\bar{e}(n) = x(n) - A_N(n-1)^T \cdot X_N(n-1) \tag{2.30}$$

- Compute the variance of the forward prediction error.

$$\alpha(n) = \lambda \cdot \alpha(n-1) + \gamma_N(n-1) \cdot \bar{e}(n)^2 \tag{2.31}$$

- Compute the likelihood variable of order $N+1$.

$$\gamma_{N+1}(n) = \frac{\lambda \cdot \alpha(n-1)}{\alpha(n)} \cdot \gamma_N(n-1) \quad (2.32)$$

- Compute the first auxiliary variable and the dual Kalman gain of order $N+1$.

$$\bar{S}_{N+1}(n) = \frac{\bar{e}(n)}{\lambda \cdot \alpha(n-1)} \cdot \begin{bmatrix} 1 \\ -A_N(n-1) \end{bmatrix} \quad (2.33)$$

$$\bar{C}_{N+1}(n) = \begin{bmatrix} 0 \\ \bar{C}_N(n-1) \end{bmatrix} - \bar{S}_{N+1}(n) \quad (2.34)$$

- Update the coefficients of the forward predictor.

$$A_N(n) = A_N(n-1) - \bar{e}(n) \cdot \gamma_N(n-1) \cdot \bar{C}_N(n-1) \quad (2.35)$$

- Compute the backward prediction error.

$$\bar{r}(n) = x(n-N) - B_N(n-1)^T \cdot X_N(n) \quad (2.36)$$

- Stabilize the backward prediction error.

$$\bar{r}^s(n) = 2 \cdot \bar{r}(n) + \lambda \cdot \beta(n-1) \cdot \bar{C}_{N+1}^{N+1}(n) \quad (2.37)$$

- Compute the likelihood variable of order N .

$$\gamma_N(n) = \frac{\gamma_{N+1}(n)}{1 + \gamma_{N+1}(n) \cdot \bar{r}^s(n) \cdot \bar{C}_{N+1}^{N+1}(n)} \quad (2.38)$$

- Compute the second auxiliary variable and the dual Kalman gain of order N .

$$\bar{U}_{N+1}(n) = -\bar{C}_{N+1}^{N+1}(n) \cdot \begin{bmatrix} -B_N(n-1) \\ 1 \end{bmatrix} \quad (2.39)$$

$$\begin{bmatrix} \tilde{\mathbf{C}}_N(n) \\ 0 \end{bmatrix} = \tilde{\mathbf{C}}_{N+1}(n) + \tilde{\mathbf{U}}_{M+1}(n) \quad (2.40)$$

- Update the coefficients of the backward predictor.

$$\mathbf{B}_N(n) = \mathbf{B}_N(n-1) - \tilde{r}^s(n) \cdot \gamma_N(n) \cdot \tilde{\mathbf{C}}_N(n) \quad (2.41)$$

- Compute the variance of the backward prediction error.

$$\beta(n) = \lambda \cdot \beta(n-1) + \gamma_N(n) \cdot \tilde{r}^s(n)^2 \quad (2.42)$$

Phase 2: Filtering Part.

$$e(n) = y(n) - \mathbf{H}_N(n-1)^T \cdot \mathbf{X}_N(n) \quad (2.43)$$

$$\mathbf{H}_N(n) = \mathbf{H}_N(n-1) - e(n) \cdot \gamma_N(n) \cdot \tilde{\mathbf{C}}_N(n) \quad (2.44)$$

Stability. $\left(1 - \frac{1}{2N}\right) < \lambda < 1$ (Note: the lower bound should be avoided) [106]

Conditional Re-initialization. The prediction part of the algorithm must be re-initialized following periods of poor excitation, which is typical with speech. This is accomplished by resetting *all* prediction variables whenever $\gamma_N(n)$ approaches zero without clearing $\mathbf{H}_N(n)$.

2.7.2 The FNTF Algorithm

The Fast Newton Transversal Filter proposed by Moustakidis and Theodoridis [62] is similar to the SFTF but takes advantage of the fact that in most cases the predictable part of the input signal can be extracted using predictors of much lower order than that of the filter itself ($M < N$). (Note that the forward and backward predictors of the SFTF algorithm are the same size as the filter). The dual Kalman gain vector, $\tilde{C}_N(n)$, and likelihood variable, $\gamma_N(n)$, are extrapolated from the corresponding reduced order variables, $\tilde{C}_M(n)$ and $\gamma_M(n)$ respectively, as described in detail in [62]. The FNTF algorithm allows complexity and performance to be traded by selecting the order of the predictors appropriately. At the extremes when $M=1$ and $M=N$, the FNTF algorithm is equivalent to the NLMS and SFTF algorithms respectively. Since speech can generally be modelled by a⁻ AR(16) to AR(20) process, significant computational savings in applications requiring long filters is possible.

While the extrapolation procedure can be based on both forward and backward predictors or just one of them, the use of both predictors appears to be more common [62] [64] [106]. This formulation results in a complexity of $12M+2N$, where M is the size of the predictors and N is the order of the overall filter [106]. This version of the FNTF algorithm is composed of four parts; delayed and non-delayed prediction, extrapolation, and filtering and adaptation as illustrated in figure 2.8. Phase 1 and 2 are identical to the prediction part of the SFTF algorithm except that the order of the predictors is reduced to M , and the delayed predictors are based on an appropriately delayed input vector. Phase 4 is also the same as the filtering part of the SFTF algorithm. The extrapolation phase is new and given by equations (2.45) - (2.49). The algorithm is initialized by setting all vectors to zero, all likelihood variables to 1, and the forward and backward prediction variances to suitable initial values.

FIGURE 2.8 Implementation of the 12M+2N FNTF algorithm.

The (12M+2N) FNTF Algorithm

Initialization.

$$A_M = B_M = \tilde{C}_M = X_M = 0$$

$$\tilde{C}_{M+1} = \tilde{S}_{M+1} = \tilde{U}_{M+1} = 0$$

$$C_N = H_N = 0$$

$$\alpha_{M,n}(0) = \alpha_{M,n_d}(0) = \lambda^N \cdot \zeta$$

$$\beta_{M,n}(0) = \lambda^{N-M} \cdot \zeta$$

$$\beta_{M,n_d}(0) = \zeta$$

$$\gamma_{M,n}(0) = \gamma_{M,n_d}(0) = \gamma_N(0) = 1 \quad (2.45)$$

Phase 1: Non-delayed Predictor.

Same as the prediction part of the SFTF algorithm (equations (2.30) - (2.42)) using the scalars $\alpha_{M,n}, \beta_{M,n}, \gamma_{M,n}$, with all parameters defined at time n .

Phase 2: Delayed Predictor.

Similar to phase 1 with scalars $\alpha_{M,n_d}, \beta_{M,n_d}, \gamma_{M,n_d}$ with all parameters defined at time $n_d=(n-N+M)$.

Phase 3: Extrapolation.

$$\tilde{C}_{N+1}(n) = \begin{bmatrix} 0 \\ \tilde{C}_N(n-1) \end{bmatrix} - \begin{bmatrix} \tilde{S}_{M+1}(n) \\ \mathbf{0}_{N-M} \end{bmatrix} \quad (2.46)$$

$$\begin{bmatrix} \tilde{C}_N(n) \\ 0 \end{bmatrix} = \tilde{C}_{N+1}(n) + \begin{bmatrix} \mathbf{0}_{N-M} \\ \tilde{U}_{M+1}(n_d) \end{bmatrix} \quad (2.47)$$

$$\delta_N(n) = \delta_N(n-1) + \tilde{U}_{M+1,n_d}^{M+1} \cdot \tilde{r}_{N,n_d}^s(n) - \tilde{S}_{M+1,n}^1 \cdot \tilde{e}_{N,n}(n) \quad (2.48)$$

$$\gamma_N(n) = \frac{1}{1 + \delta_N(n)} \quad (2.49)$$

Phase 4: Filtering and Adaptation.

Same as the SFTF algorithm; equations (2.43) and (2.44).

where:

- N is the order of the transversal filter (and estimated model of the unknown system)
- M is the estimated order of the predictors
- $\gamma_N(n)$, and $\gamma_{N+1}(n)$ are the likelihood variables of order N and $N+1$ respectively.

- $\tilde{\mathbf{C}}_N(n)$, and $\tilde{\mathbf{C}}_{N+1}(n)$ are the dual Kalman gain vectors of order N and $N+1$

Stability. $\left(1 - \frac{1}{2N}\right) < \lambda < 1$. Good numerical stability requires use of a tighter bound to prevent unrealistic a priori estimation errors in the backward prediction filters: $\left(1 - \frac{1}{3N}\right) < \lambda < 1$. [106]

In AECs for teleconferencing, the order of the predictors is complicated by the fact that the far-end talker's speech is convolved with the response of the far-end room [81]. Depending on the acoustic characteristics of that room, the required prediction order may be significant larger than that expected based on the speech alone, reducing any potential complexity reduction advantages.

Petillon, Gilloire, and Theodoridis [106] compared the performance of the FNTF and NLMS algorithms in terms of convergence rate, tracking performance, and steady-state error in the context of acoustic echo cancellation in the mobile domain. Using white noise, coloured noise, and real speech and an impulse response measured in a car, they demonstrated improved convergence and tracking performance with the SFTF and FNTF algorithms over the NLMS algorithm. Unfortunately, whether these results could be reproduced in teleconferencing systems installed in conference rooms is unclear.

2.7.3 Accelerated Fast-RLS Algorithms

The ability of the SFTF and FNTF algorithms to track nonstationarities is related to how effectively the algorithm is able to forget past data; a property determined by the forgetting factor. As the forgetting factor approaches zero, the memory of the algorithm approaches one sample. At the other extreme when the forgetting factor is equal to one, the algorithm has infinite memory and is ill-equipped to deal with sudden changes in the environment. A value closer to zero improves tracking performance. Unfortunately, the forgetting factor is bounded by $\left(1 - \frac{1}{2N}\right) < \lambda < 1$ to prevent instability in the prediction part of the SFTF

algorithm [42][74][105]. Even for moderate filtering tasks this bound severely limits the tracking performance of fast RLS algorithms.

Fortunately, a technique known as *acceleration* is available to improve the tracking performance of these algorithms. It works by increasing the Kalman gain in the joint-process (filtering) part without affecting the prediction part and thereby not affecting stability [50] [80]. Acceleration is achieved by replacing equation (2.44) with equation (2.50), where $\eta(t)$ is an arbitrarily defined gain vector.

$$H_N(t) = H_N(t-1) - \eta(t) \cdot e(t) \cdot \gamma_N(t) \cdot \tilde{C}_N(t) \quad (2.50)$$

Benallal and Gilloire [50] suggested that $\eta(t)$ be defined to exactly cancel the a posteriori filtering error, as given by equation (2.51), where ρ is a control parameter set close to but never greater than one.

$$\eta(t) = \frac{1}{1 - \rho \cdot \gamma_N(n)} \quad (2.51)$$

It can be shown that the time constants of the accelerated algorithm are dependent on an accelerated forgetting factor, λ_{acc} , and not λ [42] [51]. λ_{acc} is equivalent to λ when ρ equals zero (0), and approaches zero (0) as ρ approach one (1). The advantage of using acceleration in fast RLS algorithms is demonstrated in [42], [51] and [106].

2.7.4 The Normalized Stochastic Newton Algorithm

A second disadvantage of exponential-weighting in the RLS algorithm is a degradation of convergence performance following the nonstationarity especially in large order filtering applications where the forgetting factor must be set close to one. As the forgetting factor approaches one the Kalman gain vector approaches zero as the number of iterations are increased. This reduces the algorithm's ability to re-converge after a nonstationary period [54].

Davila [54] investigated this phenomena and proposed the Normalized Stochastic Newton (NSN) algorithm. The NSN algorithm is defined by replacing equation (2.44) of the SFTF algorithm by equation (2.52).

$$\mathbf{H}_N(n) = \mathbf{H}_N(n-1) - \frac{\mu \cdot e(n) \cdot \tilde{\mathbf{C}}_N(n)}{\mathbf{X}^T(n) \cdot \tilde{\mathbf{C}}_N(n)} \quad (2.52)$$

The normalizing factor in the denominator of equation (2.52) prevents the gain from going to zero as the number of iterations increase. The scalar constant, μ , is similar to the step size, α , in the NLMS algorithm. Although the NSN algorithm does not minimize the same LS criteria as the RLS algorithm it does converge to the optimum Weiner vector. In fact, the only difference between the NSN and SFTF algorithms is the size of the *step* taken towards the optimum vector at each iteration, since the search direction is the same in both algorithms.

2.8 Tracking Performance of Enhanced Algorithms

There appears to be an unfortunate absence in the literature on the tracking performance of the enhanced algorithms discussed in sections 2.6 and 2.7. The relative tracking performance of the LMS and RLS algorithms has been discussed in the literature and in section 2.5. Tracking performance was found to depend quite strongly on the degree of colour in the input reference signal, and the level of uncorrelated background noise in the primary signal. Other parameters like the number of taps in the adaptive filter and the value of the step size (or forgetting factor) also affects tracking performance [77]. Particularly, when the number of taps is large (as in teleconferencing applications), algorithms with dynamic properties that are independent of the colour in the reference signal (like those of the RLS family) maintain a higher level of ERLE during a nonstationary period (better tracking performance) than simpler algorithms which are sensitive to the colour in the reference signal (like those of the LMS family). The

tracking performance of the enhanced algorithms discussed in the sections above is described below.

ESNLMS.

The tracking performance of the ESNLMS algorithm is expected to exceed that of the NLMS algorithm because the adaptation is specialized to the requirements of a real conference room through the use of an exponentially-decaying step size matrix. The time constants of the algorithm are, however, sensitive to the eigenvalue disparity in the reference signal meaning that degraded performance can be expected in applications involving highly coloured signals.

TWNLMS.

The TWNLMS algorithm is expected to track at a level between that of the NLMS algorithm and the fullband ESNLMS algorithm. Its dynamic performance does, however, depend on the degree of colour in the reference signal.

VSSLMS.

Variable Step Size LMS (and NLMS) algorithms are used to trade-off steady-state performance with tracking performance by appropriately varying the step size parameter. The algorithm is used to achieve *good* tracking and steady-state performance concurrently when different scalar step sizes are usually required.

TDLMS.

The Transform Domain LMS algorithm is capable of improved tracking performance in applications involving highly coloured reference signals. Tracking performance is improved in these situations because of the orthogonalization achieved through use of the transformation matrix and power normalization process in the algorithm. No improvement over the ESNLMS algorithm is expected for white noise reference signals.

- Subband NLMS.** The subband NLMS algorithm is similar to the TDLMS algorithm in that improvements in tracking performance are expected only for highly coloured reference signals.
- SFTF.** The tracking performance of the SFTF algorithm is better than the NLMS algorithm when the reference signal is highly coloured, the uncorrelated noise level is small, and possibly the number of coefficients in the adaptive filter is relatively large. The time constants of the SFTF algorithm is insensitive to the colour in the reference signal.
- FNTE.** Similar to the SFTF algorithm with complexity and tracking performance between that of the NLMS and SFTF algorithms respectively.
- ASFTF.** The tracking performance of the ASFTF (and SFNTF) algorithms is expected to exceed that of the SFTF (and FNTE) algorithms because of the use of larger adaptive gain vectors than that which is possible with the SFTF (and FNTE) algorithm directly.
- NSN.** Theoretically, the tracking performance of the NSN algorithm *should* exceed that of the SFTF (and ASFTF) algorithm because the magnitude of the Kalman gain vector is normalized and as such does not decrease with increasing iterations. The NSN algorithm appears to be more sensitive to background noise than the SFTF algorithm and so does not provide any overall improvement in tracking performance.

2.9 Infinite Impulse Response Adaptive Filters

While the majority of adaptive filtering problems have traditionally been solved through use of *finite impulse response* (FIR) or *transversal* adaptive filters, a second class of filters, namely, *infinite impulse response* (IIR) adaptive filters would seem more appropriate for acoustic echo cancellation in teleconferencing applications. Modelling the *Loudspeaker-Room-Enclosure-Microphone* (LREM) impulse response with a simple FIR model leads to large order filters in rooms with a long reverberation time. More than 4000 coefficients are required to model an LREM system with a reverberation time of 500 ms. to an accuracy of -60dB (assuming 8kHz sampling). The inclusion of poles in the model, as in adaptive IIR filters, simplifies the task of *covering* long impulse responses since by definition the impulse response of IIR filters is infinite in duration. While use of adaptive FIR filters to model both the resonances (poles) and nulls (zeros) of the system's frequency response is simple and well understood, it leads to significant additional complexity over use of *the right* IIR model.

There are two fundamental formulations of adaptive IIR filters; *equation error*, and *output error* [48] [27]. In the *equation error* formulation the recursive coefficients corresponding to the poles of the filter are updated in an all-zero, non-recursive form and copied into a second filter which actually implements the all-pole section (figure 2.10). Since the all-pole coefficients are updated with the *noisy* desired signal instead of the reference which adapts the all-zero portion, the final converged tap-weight solution is *biased*. Techniques to reduce this bias are available, however [114]. The alternative is the *output error* formulation, shown in figure 2.9, in which the recursive coefficients are updated directly in pole-zero form. In this case the converged coefficients are unbiased and therefore represents an accurate model of the unknown system. It is possible, however, to converge to a local minimum which might be significantly higher than the global minimum.

FIGURE 2.9 Basic Structure of an Adaptive IIR filter based on the Output Error Formulation.

The vast majority of *pole-zero* echo cancellers are based on the *equation error* approach because of its trivial convergence and unimodal *error performance surface*. Long, Shwed, and Falconer [31] studied *equation-error pole-zero* adaptive filters for echo cancellation (figure 2.10). $A(z)$ and $B(z)$ are updated according to (2.53) and (2.54), where $X(n)$ and $D(n)$ represent the contents of the $A(z)$ and $B(z)$ filters, and μ is a scalar step size constant.

$$A(n+1) = A(n) + \frac{\mu \cdot e(n) \cdot X(n)}{\|X(n)\|^2} \quad (2.53)$$

$$B(n+1) = B(n) + \frac{\mu \cdot e(n) \cdot D(n)}{\|D(n)\|^2} \quad (2.54)$$

The authors compared the pole-zero ADF with a simple transversal ADF and concluded the following four points:

- The rate of convergence of the pole-zero equation error EC is not only dependent on the characteristics of the reference signal, but also on that of the echo path.
- Overspecifying the EC with an equal number of *extra* poles and zeros improves the convergence rate, steady-state ERLE, and the stability of the pole-zero ADF.

- Steady-state ERLE can be significantly higher for an over-specified pole-zero model than for a similar transversal model.
- In the context of modelling room acoustics, pole-zero structures perform consistently worse than transversal filters due to the large number of poles present in a typical conference room, and thus the large filter orders needed for overmodelling.

FIGURE 2.10 Basic structure of an adaptive IIR AEC based on the equation error method.

In modelling room acoustics, this particular pole-zero adaptive IIR filter appears to be inappropriate due to the large number of parameters required to achieve accurate modelling.

Gudvangen and Flockton [76] compared the adaptive LMS-FIR algorithm with adaptive IIR structures based on both the *equation error* and *output error* formulations. They showed that in the context of modelling a room it is not possible to achieve a given level of ERLE using significantly fewer parameters with an adaptive IIR filter than with an adaptive FIR filter. Quite surprisingly they found that as the reverberation time of the room was increased, the number of parameters in the IIR model rose at the same rate as the FIR model to achieve a given level of ERLE. While the *output error* IIR filter consistently out-

performed the FIR filter, the improvement was minor. The *equation error* IIR filter, however, performed consistently worse than the FIR filter because of the bias present in the steady-state tap-weight solution.

Principe et.al. [95] proposed the *generalized feed-forward* filter shown in figure 2.12 (and further developed in [86]). The structure reduces to the transversal filter when $G(z)=z^{-1}$. However, when $G(z)$ is some other suitably chosen tap-to-tap transfer function it can exhibit some of the advantages of both IIR and FIR filters, including infinite duration impulse response modelling (IIR), trivial stability (FIR), and simple adaptation algorithms (FIR). A particular instance of the *generalized feed-forward* filter investigated by the authors is the *gamma filter*. In this case $G(z) = \mu / (z - (1 - \mu))$ and is shown in figure 2.11.

FIGURE 2.11 Representation of the Gamma Delay Operator

FIGURE 2.12 The generalized feed-forward filter

The coefficients are adapted with the LMS algorithm. The step size is also adapted, however. This requires knowledge of the gradient, $\alpha_j(n)$, of the input, $x_j(n)$, with respects to μ as given in equation (2.55) with $\alpha_0(n) = 0$.

$$\alpha_j(n) = (1 - \mu) \cdot \alpha_j(n-1) + \mu \cdot \alpha_{j-1}(n-1) + x_{j-1}(n-1) - x_j(n-1) \quad , j = 1, \dots, K \quad (2.55)$$

The adaptation of μ proceeds as given in equation (2.56); $\mathbf{W}(n) = [w_0(n), w_1(n), \dots, w_K(n)]^T$, $\alpha(n) = [\alpha_0(n), \alpha_1(n), \dots, \alpha_K(n)]^T$, and η is a common step size parameter.

$$\mu(n+1) = \mu(n) + \eta \cdot e(n) \cdot \mathbf{W}(n) \cdot \alpha(n) \quad (2.56)$$

The parameter, μ , is called the *memory parameter* because it provides a mechanism for uncoupling the length of the filter's impulse response from its order. When $\mu=1$ the gamma filter reduces to Widrow's adaptive transversal filter. The gamma filter is particularly suited to modelling low-frequency signals since the group delay of $G(z)$, which defines the memory depth vs. frequency, is greater at low frequencies. In the context of general pole-zero filters discussed earlier, the gamma filter has a K^{th} order pole at $(1-\mu)$. However, instead of adapting K different poles, a single K^{th} order pole is adapted thus reducing computational complexity. Unfortunately, it may converge to a local minimum *not* the global minimum.

Haneda, Makino, and Kaneda [63] [107] proposed a *common-acoustical-pole and zero* model (CAPZ) consisting of a fixed AR filter and an adaptive MA filter in a series-parallel structure as shown in figure 2.13. The coefficients of the fixed AR filter are estimates of the *common acoustical poles* corresponding to the resonant properties of the room. These poles depend on the acoustical conditions of the room but not on the positions of the source and receiver and can be estimated from a large number of impulse responses measured in the room.

FIGURE 2.13 CAPZ Model proposed by Haneda, Makino, and Kaneda.

The CAPZ model demonstrated a 1.5 times improvement in convergence rate over an all-zero FIR model with the same number of total parameters. The improvement in the CAPZ model was due to a reduction in the number of variable parameters through use of a large number of fixed parameters. Unfortunately, the authors imply that the CAPZ model is only effective at low frequencies ($< 800\text{Hz}$) where the *common acoustical poles* can be estimated reasonably accurately. For signals occupying the complete telephony bandwidth ($300\text{Hz} - 3400\text{Hz}$) it is unclear whether the CAPZ model would continue to outperform a similar FIR model. Simulations presented by the authors were specific to frequencies below 800Hz . No comparison at higher frequencies is presented. Furthermore, use of the CAPZ model in an AEC is only applicable in systems fixed to a particular room due to the large number of impulse responses that have to be measured to estimate the fixed AR coefficients.

Williamson [85], and Zimmermann [87] studied *fixed pole adaptive filters* (FPAF) similar to the CAPZ model shown in Figure 2.13. Aside from a potential improvement in convergence rate, a reduction in misadjustment is also suggested because the number of adapted coefficients is reduced. As in the CAPZ model, the FPAF is comprised of an adaptive MA component and a fixed AR component. The fixed pole locations are chosen to approximate the exponential trends in a collection of sample impulse responses (based on a

modified Prony Method [3]). FPAF are shown to out-perform adaptive FIR filters given equal complexity.

In summary, a number of variations on the adaptive IIR filter have been investigated in recent years to deal with some of the problems inherent in transversal AEC. Three general approaches can be identified:

1. those which adapt both the poles and the zeros.
2. those which adapt both the poles and the zeros but restrict the model to a K^{th} order pole.
3. those which fix the location of the pole(s) and adapt the zeros.

The most promising approach appears to be the *fixed pole* structure, but further work is required to assess the implications of errors in estimating the pole locations.

2.10 Summary

The principles and techniques of linear adaptive filtering discussed in this chapter are particularly suited to the application of acoustic echo cancellation in teleconferencing applications. The basic NLMS and RLS algorithms based on the linear transversal FIR filter were presented first, followed by a comparison of the theoretical tracking performance of the two algorithms. General advances in each family were presented next including the ESNLMS, TWNLMS, time-varying NLMS, TDLMS, and subband techniques in the NLMS family, and the SFTF, FNTF, acceleration, and NSN algorithms in the RLS family. The chapter concluded with a review of infinite impulse response (IIR) algorithms and structures, and reviewed general tracking characteristics of these enhanced algorithms.

Although significant progress has been made in adaptive filtering, current AECs continue to operate below desired steady-state and dynamic performance levels.

Chapter 3

3.0 PROBLEM DEFINITION

The following general statements can be made based on the material provided in chapter 2.

1. An Acoustic Echo Canceller (AEC) is one solution to the problem of supporting full-duplex, high-quality handsfree communication.
2. The design of an AEC is complicated by a variety of factors which can be broadly classified into three categories; environmental/operational, algorithmic, and equipment related. The main limitations of AEC are due to:
 - uncorrelated noise - hardware, algorithmic, and room noise [72] [74]
 - undermodelling of the LREM impulse response [74]
 - enclosure vibration, rattling, and whistling effects
 - nonlinearities introduced by the loudspeaker [74] [111]
 - dynamic tracking of nonstationarities [12] [36] [105]
 - double-talk [9] [17]
3. Current AECs do *not* meet required levels of steady-state or dynamic performance and so must typically be coupled with other less desirable techniques like *center-clipping*, or *voice-controlled switching*.

The purpose of this thesis is to investigate the tracking performance of a variety of adaptive filtering algorithms for acoustic echo cancellation in typical conference rooms. Performance comparisons are made with band-limited white noise and speech excitation signals. The speech signal used in this thesis was provided by Bell Northern Research. Room non-stationarities were introduced by moving objects near the phoneset. Complexity, near-end speech, double-talk, and nonlinearities are considered out-of-scope.

Algorithms considered in this thesis include the following:

1. The exponentially-decaying step size NLMS algorithm provides improved convergence and tracking performance over the regular LMS family through an optimization of the algorithm to the characteristics of room's impulse response.
2. Subband adaptive filtering algorithms have been proposed to improve the convergence of the LMS algorithm with colored input signals.
3. Accelerated SFTF algorithm appears particularly suited to the tracking of room nonstationarities in the presence of speech excitation.

Algorithms not considered include the following:

1. A variety of methods have been proposed to improve the rate of convergence of the LMS algorithm with speech signals. The transform-domain LMS algorithm is one possibility. Unfortunately, the use of fixed transforms leads to results which are signal dependent.
2. Time-varying algorithms were previously studied by Knappe [74], and Yuan [105] and found to be only marginally effective at improving the dynamic performance of LMS type algorithms in nonstationary conditions.
3. While infinite impulse response or pole-zero adaptive filters appear particularly suited to acoustic echo cancellation due to the dominance of poles in the frequency response of typical conference rooms, considerable work remains in the development of techniques to estimate the pole locations accurately.

Chapter 4

4.0 EXPERIMENTAL SETUP AND PHYSICAL LIMITATIONS

This chapter investigates the physical limitations of acoustic echo cancellers from an experimental viewpoint. The reader is referred to Knappe [74] for a more theoretical and complete treatment of these issues. Our discussion is limited to the issues which are relevant to the focus of this thesis. The experimental setup is described first, followed by an investigation of the maximum achievable ERLE of the setup. Next, the steady-state limitations are quantified including the effects of phone set design, and the acoustic environment. The chapter concludes with an investigation of the limitations caused by room nonstationarities. The results presented in sections 4.1 and 4.2 were obtained in collaboration with Neil Birkett.

4.1 Experimental Setup

Since this thesis expands on that of Yuan [105] a similar experimental setup was used. Unfortunately, a variety of problems were uncovered with the old setup which were found to limited achievable ERLE to approximately 30dB. Testing conducted with the NT phone (results discussed in section 4.3.1) demonstrated a need for a higher quality experimental setup. Considerable work was involved in evaluating the limitations of the setup. Changes were made to the power amplifier, the primary and reference pre-DAT conditioning circuits, the telephony network (replaced with a telephony BPF), and the overall grounding strategy. The new experimental setup is displayed in figure 4.1.

A modified speakerphone¹ was positioned in either an anechoic chamber or a conference room and excited with either bandlimited white noise or speech depending on the

1. each phone was modified to allow access to the internal loudspeaker and microphone terminals.

conditions desired. Noise was generated using a General Radio Company 1390-B Random Noise Generator. Artificial speech recordings were provided by Bell Northern Research and reproduced using a Sony TC-D7 Digital Audio Tape (DAT) recorder. The source signal selected was then filtered to the telephony bandwidth (200Hz to 3400Hz) with a cascade of two National Semiconductor TP3040 switched capacitor filters to provide greater than 60dB of stopband attenuation. The filtered signal was amplified with a 75W dual-channel Samson Servo-150 studio-quality power amplifier and used to excite the loudspeaker in the modified phone.

A set of pre-conditioning circuits were used to attenuate the reference signal, and amplify the primary signal to equate their levels at the DAT's inputs. Both circuits were based on the Analog Devices AD524 instrumentation amplifier IC. The primary and reference signals were recorded to DAT using a Teac DA-P20 DAT recorder at a level of -9dB with noise, and a peak level of -6dB with speech. The data was later downloaded to a PC for off-line processing by resampling the data at 16kHz with an Ariel DSP-96 card. Schematics of the experimental setup can be found in [110].

FIGURE 4.1. Block diagram of the experimental setup

The datafiles were then decimated by a factor of two to a sampling frequency of 8kHz using a 300th order Blackman windowed FIR anti-aliasing filter with a cutoff frequency of 0.48π (in the Matlab environment). The dc-bias, introduced predominantly by offset errors on the Ariel board, were removed to prevent poor convergence rate. The resulting primary and reference data vectors were then issued to a variety of adaptive filtering programs written specifically for this thesis in Matlab and C.

4.2 Maximum Achievable ERLE of the Experimental Setup

Knowledge of the maximum achievable ERLE of the experimental setup is important because it provides knowledge about the *quality* of the setup. Uncorrelated noise or nonlinear components added to the primary or reference signals degrades achievable ERLE. This is understood by the simple fact that the unknown system modelled by an adaptive filter at the computer not only consists of the room environment but also the Ariel DSP circuitry, the DAT, and the conditioning circuitry. Note, any circuitry common to both paths has no effect on achievable ERLE, only signal quality.

The maximum achievable ERLE of the experimental setup was found by replacing the *Loudspeaker-Room-Enclosure-Microphone* (LREM) system (see figure 4.1) with a passive attenuator network. A delay was added to the primary (simulation stage) to avoid possible anti-causality issues caused by a difference in the delay between the reference and primary signal paths. The modified setup was excited with bandlimited white noise. The achievable ERLE was calculated using a 10 tap NLMS adaptive filter with a step size of 0.5. The resulting curve is shown in figure 4.2. Also plotted is an asymptotic curve which represents an estimate of the maximum achievable ERLE of the experimental setup based on an estimate of the primary signal's SNR. The maximum achievable ERLE of the measurement setup was found to be approximately 56dB. This limit is caused by imperfection in the experimental setup including noise and nonlinear components introduced by *any* of the electronic components.

FIGURE 4.2 Optimum steady-state ERLE performance of the experimental setup based on the converged ERLE of a 10 tap NLMS adaptive filter with a step size of 0.5 (Linear FIR curve), and an estimate of the SNR of the primary signal (Asymptote curve).

Further experiments were conducted to evaluate the experimental setup more completely and provide an explanation for the limitation in performance. The results are displayed in figure 4.3. The line labelled *Ariel* corresponds to converged ERLE values for a variety of equivalent DAT record levels¹ obtained by injecting bandlimited white noise directly into the analog input ports of the Ariel board. An achievable ERLE in excess of 70dB was obtained for record levels greater than -13dB. The additional quantization noise added by first sampling the data to DAT degraded the achievable ERLE by approximately 8dB as shown in the curve labelled *DAT + Ariel*. Finally, the performance of the complete setup is illustrated by the line labelled *Full Setup*. This curve is similar to that shown in figure 4.2 except that the reference signal voltage is held constant at 1Vrms while the record level at the DAT is adjusted. These results suggest that the current limitation of 56dB is caused by noise added by the primary (microphone) amplifier. Improving the noise performance of

1. For the *Ariel* curve the input voltage is converted to an *equivalent* DAT record level with knowledge of the DAT's input/output transfer characteristics to support comparison between the curves.

this amplifier is unnecessary, however, because as demonstrated in the remainder of this chapter a variety of physical limitations act to limit ERLE to a level below 56dB.

FIGURE 4.3 Comparison of the optimum steady-state ERLE of the Ariel, DAT and Ariel, and full setup based on a 10 tap NLMS adaptive filter with a step size of 0.5.

4.3 Steady-State Performance Limitations

This section examines the steady-state performance of acoustic echo cancellers by examining the limitations to achievable ERLE caused by variations in the design of the speakerphone, room characteristics, and background acoustic noise.

4.3.1 Limitations due to Speakerphone Design

The five speakerphones listed in table 4.1 were tested in an anechoic chamber¹ to determine the magnitude of the limitation to ERLE caused by variations in their design. The purpose here is only to provide evidence of the potential spread in ERLE caused by variations in phoneset design, and *not* to isolate each of the multitude of factors which contribute to the

1. The anechoic chamber at Carleton University is located in the Psychology department in room A102A.

limitation (including the quality of the loudspeaker and microphone, and the design of the enclosure).

TABLE 4.1 List of speakerphones tested

Speakerphone Manufacturer/Model #
Northern Telecom / NT8B04AA
Mitel / Superset 430
Mitel / Superset 410
Telemax / CP268A
Panasonic / KX-T2315

Each phone was placed in an anechoic chamber and energized with bandlimited white noise to sound pressure levels ranging from 60 to 88dB measured at 1m. Primary and reference signals were recorded to DAT and then downloaded to an IBM PC for off-line processing. A 1200 tap NLMS adaptive filter with a step size of 0.5 was allowed to converge for 4 seconds. The steady-state coefficients were obtained by averaging the filter's instantaneous coefficient values over the final second of adaptation, to obtain an estimate of the impulse response of the unknown system. This system is comprised of the components shown in figure 4.4.

FIGURE 4.4 Breakdown of the LREM impulse response, and AEC configuration.

The anechoic impulse responses each speakerphone are plotted in figures 4.5 - 4.9.

FIGURE 4.5 Anechoic impulse response of the NT phone.

FIGURE 4.6 Anechoic impulse response of the Mitel Superset 430

FIGURE 4.7 Anechoic impulse response of the Mitel Superset 410

FIGURE 4.8 Anechoic impulse response of the Telemax CP268A

FIGURE 4.9 Anechoic impulse response of the Panasonic KX-T2315

Ideally, the anechoic impulse response of a phone should consist of a single impulse representing the direct path delay between the loudspeaker and microphone due to the absence of reverberation in an anechoic chamber. When the loudspeaker and microphones are located in a common enclosure, however, the anechoic response of a phone can be far from ideal due to the echo path within the enclosure. This is precisely the case with each of the phones tested. The response of the NT phone is closest to the ideal, with the response of both Mitel phones extend beyond 600 taps. While no general statement can be made to relate the length of the anechoic impulse response of a phone to its steady-state ERLE performance, striving for the ideal response should reap benefits in improved speech quality.

The steady-state ERLE performance of each phone was computed by averaging the short-term ERLE (computed with a 1200 tap NLMS filter and a step size of 0.5) over the last second of four seconds of adaptation. The result is plotted in figure 4.10. The ERLE performance is shown to vary over a range of approximately 16dB; 37dB with the NT phone, and 19dB with the Panasonic phone (measured at an SPL of 70dB at 1m). At high volumes the ERLE performance is seen to converge to a common level of approximately 23dB limited by strong nonlinear distortion in each phones introduced by low-quality

loudspeakers and rattling and vibration effects within the enclosure. At low volumes the ERLE performance is limited primarily by the SNR of the microphone amplifier.

FIGURE 4.10 Converged ERLE for five speakerphones at various sound pressure levels measured at 1m from the phone in an anechoic chamber.

4.3.2 Limitations due to Room Characteristics and Background Noise

Undermodelling of the LREM impulse response and the presence of background acoustic noise in real conference rooms also limits steady-state ERLE performance. In this section, the performance of the NT phone is evaluated in two typical conference rooms; Minto 3033 and Minto 2014¹. The layout and dimensions of each conference room are given in appendices A and B. The impulse response of each room captured with the NT phone excited with bandlimited noise at 70dB SPL are plotted in figures 4.11 and 4.12.

1. Minto 2014 and Minto 3033 are located in the Minto Building for Advanced Studies in Engineering at Carleton University.

FIGURE 4.11 LREM impulse response measured with the NT phone in Minto 3033

FIGURE 4.12 LREM impulse response measured with the NT phone in Minto 2014

The peaks in the impulse response below approximately 200 taps, which are absent from the anechoic plots, represent early reflections from the walls and surrounding surfaces. Energy beyond a few hundred taps correspond to reverberation which decays at a rate determined by the acoustic characteristics of the room. The reverberant energy decays more rapidly in Minto 2014 than in Minto 3033 because of differences in the absorption properties of each room.

Since the LREM impulse response of a real conference room extends several hundred milliseconds in duration due to resonances in the room, a large number of coefficients are required for complete coverage. As Knappe [74] demonstrated the steady-state ERLE varies with the number of taps in the adaptive filter such that only moderate filter lengths may be required for adequate performance. The converged ERLE vs. filter order for Minto 3033 and Minto 2014 using the NT phone are plotted in figure 4.13. In both cases the ERLE rises with the number of taps until a plateau is reached. The optimum number of taps (ignoring complexity) is the smallest number of taps which provides the greatest (or required) steady-state ERLE.

FIGURE 4.13 Converged ERLE of the NT phone in Minto 2014 and Minto 3033 vs. adaptive filter order, $\alpha=1.0$.

The optimum number of taps for adaptive filters operating in Minto 2014 and Minto 3033 is 1200 and 1600 respectively. More taps are required in Minto 3033 to maximize ERLE because of its longer reverberation time. The higher level of background acoustic noise in Minto 3033 caused a reduction in the steady-state ERLE over that of Minto 2014. Since an AEC always operates in an undermodelled situation, that is one in which the number of taps in the unknown system is greater than that of the ADF, this type of information is important because it enables the system designer to trade ERLE with complexity (# taps).

4.4 Nonstationary Performance

This section demonstrates the achievable ERLE performance of AEC based on the NLMS algorithm in nonstationary conditions. Good tracking performance is important to AEC due to the nonstationary behaviour of the room caused by the movement of people and other objects in the room. When the impulse response of a room changes, the adaptive filter coefficients lag the optimum leading to poor ERLE performance. A lag component is traditionally added to the overall misadjustment to account for this tracking lag.

To demonstrate the potential limitations, and further justify the usefulness of this research, a simple experiment was conducted in Minto 2014. The NT phone was placed on a table and excited with white noise to a moderate volume level (70dB SPL at 1m). At a predefined interval a person standing 1m directly in front of the phone moved his upper body arbitrarily. A 1200 tap NLMS adaptive filter was used with a step size of 1.0 (for maximum tracking performance) and the local ERLE performance was calculated. The result is plotted in figure 4.14. The ERLE performance was found to drop from a steady-state level of approximately 33dB to an average level of 22dB with a variation of +/- 5 dB during the nonstationarity.

FIGURE 4.14 ERLE performance of NT phone in Minto 2014 during long-term nonstationarity caused by a person moving in spot 1m directly in front of the phone.

4.5 Summary of Observations

The following points are demonstrated in this chapter:

- The setup is effective to an ERLE of approximately 56dB and appears to be limited by the SNR of the primary (and perhaps reference) amplifier (figures 4.2 and 4.3).
- At medium volume levels (70dB SPL at 1 m) the steady-state ERLE is limited by:
 1. the acoustic design of the enclosure including speaker/microphone separation and coupling, internal acoustic design, whistling through openings near the microphone, and rattling of keys or other loose objects on the phone (figure 4.10).
 2. background acoustic noise in the room (figure 4.13).
 3. undermodelling of the room transfer function (figure 4.13).

- Enclosure related effects appear to be the predominant limiting factor at mid-volumes. The variation in ERLE among the five speakerphones is approximately 16dB at 70dB SPL; 19dB with the Panasonic phone, and 37dB with the NT phone.
- nonlinearities introduced by the loudspeaker and possibly the enclosure act to limit ERLE performance at loud volumes.
- undermodelling of the acoustic transfer function typically required in practical implementations limits ERLE. The optimal number of taps is that point beyond which the ERLE does not improve (at least significantly). The optimal number of taps in Minto 2014 is 1200, while Minto 3033 benefits from up to 1600 taps because of its longer reverberation characteristic.
- nonstationarities caused by movements in the acoustic environment is shown to limit ERLE performance to a level determined by the ability of the algorithm to track the changes in the impulse response.

Chapter 5

5.0 BANDLIMITED NOISE RESULTS

In the previous chapter the achievable ERLE of an acoustic echo canceller (AEC) was shown to be limited by a variety of factors including the design of the enclosure and components, background acoustic noise, undermodelling, and nonstationarities. This chapter studies the dynamic performance of linear AECs operating in nonstationary environments excited with bandlimited white noise. All results are based on experimental data captured in realistic conditions in a typical conference room. Comparisons using simulated nonstationary conditions were avoided to maintain realism and uniqueness in the final results. The performance of the fullband NLMS algorithm is considered the baseline reference to which other algorithms are compared. A variety of fullband and subband LMS and RLS family of algorithms are investigated with the goal of improving the tracking performance of AECs.

5.1 Experimental Conditions

All results presented in this chapter are based on data collected in Minto 2014 with the NT phone positioned as illustrated in appendix B. Bandlimited white noise was injected into the loudspeaker of the modified phone to a sound pressure level of 70dB measured at 1m. The loudspeaker and microphone signals were recorded for 20 seconds to a Teac DAT recorder. At the 4 second mark, a stationary observer positioned near the phone moved his hand, with palm facing the phone, 6" directly above the phone in a direction parallel to the table on which the phone was placed. Four passes were made over the phone in a period of approximately 3 seconds. The observer was stationary during the remainder of the recording. A 10 second excerpt from the recorded data, starting at 0.5s, was downloaded from DAT to separate primary and reference PC datafiles for off-line processing. Complete details of the experimental setup, including operating procedures, is given in chapter 4.

Table 5.1 identifies the various adaptive filtering algorithms investigated in this chapter. Both fullband and subband LMS-type algorithms, and RLS-type algorithms were examined. Complexity issues were ignored in all algorithms to provide a best-case analysis of the potential benefits and performance of each algorithm. In all simulations the coefficient vector was initialized to the zero vector.

TABLE 5.1 List of adaptive filtering algorithms investigated

Full Algorithm Name	Acronym	Ref. Section
Normalized Least-Mean-Squares	NLMS	2.3
Accelerated exponentially-weighted Stabilized FTF	ASFTF	2.7.1, 2.7.3
Exponential-Decay Step size Normalized Least-Mean-Squares	ESNLMS	2.6.1
Subband Normalized Least-Mean Squares	SBNLMS	2.6.5
Subband Exponential-Step Normalized Least-Mean Squares	SBESNLMS	(proposed)

The algorithms are compared using both quantitative and qualitative measures based on ERLE performance computed using equation (2.2). Plots of short-term ERLE performance vs. time are compared to gain insight into the general convergence and tracking characteristics of each algorithm. Comparisons are also made based on average ERLE performance measured over stationary and nonstationary time windows. Average stationary ERLE is computed before the nonstationary from 2.5-3.5s, and following the nonstationarity from 8-10s. Nonstationary ERLE performance is measured from 3.7s to 6.7s.

5.2 The Fullband NLMS Algorithm

The first adaptive algorithm studied was that of the fullband NLMS algorithm. The algorithm was simulated with 1200 coefficients to maximize steady-state ERLE performance (see figure 4.13). Step sizes of $\alpha=1.0$ and $\alpha=0.1$ were used. Simulation results are summarized in table 5.2, and in figure 5.1.

TABLE 5.2 Quantitative ERLE performance of the NLMS algorithm excited with bandlimited white noise.

NLMS	Avr. Stat. ERLE (2.5-3.5s), (dB)	Avr. Stat. ERLE (8-10s), (dB)	Avr. Nonstat. ERLE (3.7-6.7s)
N=1200, $\alpha=1.0$	29.8	32.0	15.0
N=1200, $\alpha=0.1$	20.9	29.5	16.1

FIGURE 5.1 Effect of α on the tracking and convergence performance of the NLMS algorithm with 1200 taps and excited with bandlimited white noise.

Short-term ERLE curves shown in figure 5.1 illustrate the improvement in convergence that is possible with the NLMS algorithm by simply increasing the value of the step size. Maximum convergence performance is obtained by setting $\alpha = 1.0$ as shown by the solid curve [103]; 4s for initial convergence, 1.5s to converge following the nonstationarity. The difference in rates of convergence before and after the nonstationarity was due to the effects of slowly converging modes near the Nyquist frequency caused by the bandlimited nature of the input signal. This is proven in figures 5.2 and 5.3 in which the power spectral density of the error signal is computed using Welch's averaged periodogram method [41] while the algorithm is converging prior to and following the nonstationary period respectively. Beyond the 0.5s mark only those frequencies near the transition band at 3.5kHz remain to be identified (as indicated by the higher error power in this region). Convergence is slow because the eigenvalues (and hence the time constants) associated with the modes in this region are small. Convergence is faster after the nonstationarity because the modes changed by the nonstationary condition are excited equally by the reference signal. The transition band region is essentially unaffected by the nonstationarity.

FIGURE 5.2 Short-term power spectrum of error signal associated with NLMS algorithm computed from 0.5s to 1.5s (initial convergence), and from 9s to 10s.

The NLMS algorithm appears to be relatively ineffective at tracking nonstationarities introduced in typical conference rooms. The average ERLE performance dropped by approximately 50% during this period as shown in figure 5.1. With $\alpha=0.1$ the average nonstationary ERLE was found to be 1.1 dB greater than that obtained with $\alpha=1.0$. As explained in section 2.5 (and in figure 2.5), the misadjustment during a nonstationary period is given by the sum of the misadjustment due to noise, and due to tracking lag [25]. The step size parameter determines the relative magnitude of these conflicting factors in the overall misadjustment figure. A large step size reduces the misadjustment due to tracking lag, but increases that due to noise, and vice versa. While not shown in the plot, the curve for $\alpha=0.1$ will eventually converge to a higher steady-state ERLE than with $\alpha=1.0$. Steady-state was not achieved in either case because the algorithm was not allowed to adapt for a sufficient length of time.

FIGURE 5.3 Short-term power spectrum of error signal associated with NLMS algorithm computed during re-convergence following the nonstationarity over the windows 5.5-6.5s and 6-7s, and over the stationary window from 9-10s.

5.3 The Accelerated SFTF Algorithm

This section investigates the tracking performance of the accelerated SFTF algorithm, described in section 2.7.3, excited with bandlimited white noise. The motivation for pursuing this work originated in results presented by Knappe in [74], Yuan in [105], and Petillon, Gilloire, and Theodoridis in [106]. Both Knappe and Yuan found the NLMS algorithm to be superior to the SFTF algorithm in acoustic echo cancellers in teleconferencing applications. They suggested that the SFTF algorithm's poor performance was due to the limits on the forgetting factor, λ , for stability reasons. In large order filtering applications these limits, which are given by $1 - \frac{1}{2N} < \lambda \leq 1$, severely restrict λ , and hence tracking performance. For good numerical performance an even tighter bound is recommended $1 - \frac{1}{3N} < \lambda \leq 1$ [106]. A novel technique for overcoming these limits, known as acceleration, was proposed by Benallal and Gilloire in [42], and recently demonstrated in [106] for mobile AEC. The essence of the acceleration technique is to increase (or modify) the value of the Kalman gain vector in the filtering part of the algorithm without modifying the forgetting factor in the prediction part to improve the algorithm's rate of convergence and tracking performance. Unfortunately, in [106] improvements in performance were only demonstrated in mobile AECs in automobiles where the filter order was small ($N=256$). This section investigates the use of the acceleration technique in AEC in teleconferencing applications where much larger order filters are needed to cancel the acoustic echo.

The simulations presented in this section use both the accelerated and unaccelerated versions of the SFTF algorithm. The accelerated algorithm is used whenever λ_{acc} is specified, while the unaccelerated algorithm is used whenever λ is used. The unaccelerated algorithm is derived from the accelerated algorithm by setting the acceleration constant, ρ , equal to zero (0). In the accelerated algorithm, both the forgetting factor, λ , and the acceleration constant, ρ , must be specified. λ_{acc} is given by λ and ρ and determines the time constants of the algorithm. Effectively, whenever λ_{acc} is specified both λ and ρ must be given. When λ is specified, ρ is assumed to equal zero (0).

FIGURE 5.4 Effect of the forgetting factor on the ERLE performance of the SFTF algorithm excited with bandlimited white noise, $N=1200$, $\zeta=200$, and $\lambda=1$, $\lambda=1-1/10$, and $\lambda_{acc}=1-1/N$ ($\lambda=1-1/3N$, $\rho=0.95$).

The ERLE performance of the SFTF algorithm with $\lambda_{acc}=1-1/N$ ($\lambda=1-1/3N$, $\rho=0.95$), $\lambda=1-1/10N$, and $\lambda=1$ is compared in figure 5.4, and summarized in table 5.3. The order, N , equals 1200, while the soft-initialization constant, ζ , equals 200. The forgetting factor, λ , determines the width of the exponentially decaying window used in the algorithm to reduce the effect of past data on current performance. It is related to the step size, α , in the NLMS algorithm by the approximate relationship $\lambda=1-\alpha/N$, where N is the order of the filter [106]. As demonstrated in figure 5.4, increasing λ improves steady-state performance at the expense of convergence performance, and vice versa. Of particular interest is the apparent extreme sensitivity of the algorithm's convergence behaviour following the nonstationarity to the value of the forgetting factor. This latter phenomenon is described further in the next paragraph.

TABLE 5.3 Quantitative comparison of the ERLE Performance of the accelerated SFTF algorithm with bandlimited white noise.

Forgetting Factor λ	Avr. Stat. ERLE (2.5-3.5s), (dB)	Avr. Stat. ERLE (8-10s), (dB)	Avr. Nonstat. ERLE (3.7-6.7s), (dB)
1 - 1/N (accelerated)	33.1	33.0	15.0
1 - 1/10N	34.7	30.3	16.3
1	33.6	22.5	15.5
NLMS, $\alpha=1.0$	29.8	32.0	15.0

In general, the dynamic performance of an adaptive algorithm is affected primarily by the magnitude of the vector which updates the coefficients in the filtering phase (herein referred to as the *gain vector*) [54]. In the NLMS algorithm the gain vector is given by $\alpha \cdot X_N(n) / (\epsilon + \|X_N(n)\|^2)$ in equation (2.5), while in the SFTF algorithm it is given by $\gamma_N(n) \cdot \tilde{C}_N(n)$ in equation (2.44). Increasing the magnitude of the *gain vector* generally improves dynamic performance. In the NLMS algorithm, its magnitude is controlled by the step size parameter, α , while in the SFTF algorithm the gain vector is known as the Kalman gain vector and is controlled by the forgetting factor, λ . As λ approaches one, the amount of data used in estimating R^{-1} (and hence the Kalman gain vector) increases, causing its magnitude to decrease. This degrades the SFTF algorithm's tracking and convergence performance.

Figure 5.5 illustrates the relative performance of the SFTF and NLMS algorithms with similar levels of misadjustment; $\alpha = 1$ (NLMS), $\lambda_{\text{acc}} = (1-1/N)$ (SFTF). Both algorithms demonstrate almost identical tracking and convergence performance following the nonstationarity. The initial convergence of the SFTF algorithm, however, is much faster than the NLMS algorithm because of its insensitivity to the degree of colour in the input signal.

In designing an AEC, its initial convergence performance is less important than its tracking performance or its convergence performance following the nonstationarity. An AEC's coefficients are almost always initialized to a best guess of the current steady-state

LREM impulse response, and almost never initialized to zero. This is as simple as saving the coefficients between calls, and re-using them during the following call. Alternative techniques, like voiced-controlled switching, could also be used to mitigate the effects of slow initial convergence.

FIGURE 5.5 ERLE performance of the SFTF algorithm with $N=1200$, $\lambda_{acc}=1-1/N$ ($\lambda=1-1/3N$, $\rho=0.95$), and $\zeta=200$ compared with the NLMS algorithm with $\alpha=1$.

The rapid initial convergence performance of the SFTF algorithm could be used advantageously to improve convergence after the nonstationarity by initializing the algorithm following the nonstationarity. The problem becomes one of developing a suitable method of detecting the end of the nonstationarity. One possibility is to define an ERLE threshold (which could be software controlled) below which the algorithm would be initialized periodically. Although such system design issues are beyond the scope of this thesis, a brief look at the possible improvements in convergence performance using this technique is warranted. To verify the technique, the SFTF algorithm was re-simulated using the parameters of the previous section ($N=1200$, $\lambda=1-1/3N$, $\rho=0.95$, $\zeta=200$) and initialized at 6.3s. Figure 5.6 investigates the benefits of this technique.

FIGURE 5.6 Illustration of the improvement in convergence performance following the nonstationarity of the SFTF algorithm by purging its memory after the nonstationarity at 6.3s, with $N=1200$, $\lambda_{acc}=1-1/N$ ($\lambda=1-1/3N$, $\rho=0.95$), and $\zeta=200$.

As explained above, the convergence performance of the SFTF algorithm is affected by a number of factors, one of which is the magnitude of its Kalman gain vector (besides λ). The initial magnitude of this vector is determined by the magnitude of the soft-initialization constant, ζ , in an inverse relationship. ζ determines the initial magnitude of the estimation error energies in the algorithm, which affects the magnitude of the Kalman gain vector, and ultimately convergence performance. A relative small value for ζ was used in the simulations presented above to maximize convergence performance. By increasing the magnitude of the Kalman gain vector by re-initializing the algorithm, convergence performance is improved. Interestingly enough, large values for ζ are usually recommended in the literature to prevent initial divergence [106]. This, however, causes a noticeable reduction in convergence performance which may not be necessarily based on the simulations conducted during the course of this thesis.

5.4 The Fullband ESNLMS Algorithm

The *exponentially-decaying stepsize* NLMS algorithm (ESNLMS) was proposed by Makino, Kaneda, and Koizumi [92] and was investigated further by Yuan [105]. A set of 15 simulations were performed to evaluate its performance and sensitivity to initial conditions. Simulation results are summarized in table 5.4. The five parameters listed in columns two through six characterize the elements of the step size vector, α_i . These parameters are briefly described below, and are represented diagrammatically in figure 5.7. In this section, γ , $\alpha_{\min}$, $\bar{\alpha}$, and M are specified from which α_0 is calculated.

ESNLMS Parameter Summary.

1. γ is the exponential decay factor.
2. $\alpha_{\min}$ sets a lower limit on the step size.
3. α_0 specifies the initial step size.
4. $\bar{\alpha}$ is the mean step-size.
5. M defines the starting point of the exponential decay.

FIGURE 5.7 Parameterizing the step size vector used in the ESNLMS algorithm

TABLE 5.4 Quantitative comparison of the ERLE Performance of the ESNLMS algorithm with bandlimited white noise.

#	Parameters of Step Size Vector					Avr. Stat. ERLE (dB) (2.5-3.5s)	Avr. Stat. ERLE (dB) (8-10s)	Avr. Nonstat. ERLE (dB) (3.7-6.7s)
	γ	$\alpha_{\min}$	α_0	α	M			
1	0.005	0	6.0	1	1	30.0	32.2	19.5
2	0.005	0	4.1	1	100	30.3	32.1	18.5
3	0.005	0.1	4.0	1	100	32.0	32.4	18.5
4	0.01	0	11.9	1	1	22.3	23.6	19.3
5	0.01	0	6.1	1	100	22.7	24.0	18.4
6	0.01	0.1	5.8	1	100	31.1	32.2	19.9
7	0.02	0	23.4	1	1	18.9	18.4	17.4
8	0.02	0.1	21.6	1	1	29.9	31.5	22.4
9	0.02	0	8.1	1	100	18.9	19.5	17.0
10	0.02	0.1	7.5	1	100	31.8	30.1	20.7
11	0.02	0.1	6.9	1	100	32.5	32.4	20.3
12	0.02	0.1	11.2	1	50	30.0	31.7	21.7
13	0.02	0.2	10.3	1	50	32.4	32.3	21.3
14	0.02	0.1	5.2	0.5	50	31.8	33.0	20.7
15	0.02	0.1	1.1	0.1	50	23.2	28.0	17.4
NLMS, $\alpha = 1.0$						29.8	32.0	15.0
SFTF, $\lambda_{\text{acc}}=1-1/N$						37.1	33.0	15.0

Different sets of parameters were required to optimize tracking and convergence performance. The best-case tracking and convergence performance were found in trial #8, and #11 respectively. They are compared with the fullband NLMS algorithm in figures 5.8 and 5.9. Nonstationary performance improved by 7.4dB and 5.3dB in trials #8 and #11, while stationary performance from 8-10s remained relatively constant among all algorithms.

By carefully examining the simulation results in table 5.4 the following points can be deduced.

1. The minimum step size parameter, $\alpha_{\min}$, becomes important as the decay factor, γ , is increased because this equates to a large number of coefficients stuck at zero and so not adapted. In trials # 1, 4, and 7, γ was increased progressively from

0.005 (through 0.01) to 0.02. With $\alpha_{\min} = 0$, the stationary performance (8-10s) dropped progressively from 32.2dB ($\gamma=0.005$) to 18.4dB ($\gamma=0.02$). Nonstationary performance was approximately the same in trials # 1 and 4 at 19.5dB and 19.3dB respectively, but dropped to 17.4dB in trial # 7. In trial #6 with $\gamma=0.01$ and $\alpha_{\min} = 0.1$, the stationary and nonstationary performance increased to 32.2dB and 19.9dB respectively. Similarly, in trial #8 with $\gamma=0.02$ and $\alpha_{\min} = 0.1$ the stationary and nonstationary ERLE performance increased to 31.5dB and 22.4dB respectively.

FIGURE 5.8 ERLE performance of the fullband ESNLMS algorithm, best tracking performance (trial #8, table 5.4), compared with the NLMS algorithm with $\alpha=1$.

2. The starting point of the exponential decay, given by the parameter M , has a relatively small impact on stationary and nonstationary ERLE performance. In trials # 1 and 2 where $\gamma=0.005$ and $\alpha_{\min} = 0$, and $M= 1$ and 100 respectively, the stationary ERLE remained unchanged at approximately 32dB, while the nonstationary performance dropped from 19.5dB (trial #1) to 18.5dB (trial #2). Similar results were observed in trials # 4 and 5. Comparing trials # 8, 12, and 10

in which $\gamma=0.02$, $\alpha_{\min}=0.01$, and $M=1, 50$, and 100 respectively the stationary and nonstationary performance remained unchanged until $M=100$ at which point both measures became degraded.

3. Reducing the mean step size, $\bar{\alpha}$, improves stationary ERLE performance but degrades nonstationary ERLE performance. Comparing trials # 12 and 14 in which $\gamma=0.02$, $\alpha_{\min}=0.01$, $M=50$, and $\bar{\alpha}=1$, and 0.5 respectively the stationary ERLE increased from 31.7dB (trial #12) to 33dB (trial #14) respectively, while nonstationary performance dropped from 21.7dB (trial #12) to 20.7dB (trial #14)

FIGURE 5.9 ERLE Performance of the ESNLMS algorithm, best convergence performance (trial #11, table 5.4), compared with the NLMS algorithm with $\alpha=1$.

2 of/de 2

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

PRECISIONSM RESOLUTION TARGETS

5.5 The Subband NLMS Algorithm

The subbanding technique is capable of improving the convergence and tracking performance of adaptive filtering algorithms [59]. The subband NLMS algorithm is described in section 2.6.5.

The analysis and synthesis filterbanks used are based on 128th order Blackman-windowed bandpass FIR filters with cutoff frequencies optimized to minimize the overall magnitude distortion in the combined AFB/SFB system, excluding the intermediate adaptive filters. The transition band characteristics of the Blackman window were found best suited to minimizing the overall passband ripple of the system. The bandwidth of the AFB was increased by 40% relative to the SFB to minimize the effects of slow asymptotic convergence in the subbands [116] [121]. An automated version of the procedure proposed by Ryan [71] was implemented.

The magnitude response of the AFB and SFB filterbanks is plotted in figures 5.10 and 5.11 respectively. The combined frequency response (magnitude and phase) excluding the ADFs is plotted in figure 5.12. The overall combined magnitude response of the filterbank system is flat to within 0.015dB in the passband, and drops by 1dB near dc and the Nyquist frequency (out-of-band). The phase response is linear in the passband.

Simulations were performed to compare the performance of 4-band and 8-band subband NLMS algorithms with its fullband counterpart. To maintain reasonable simulation times, a maximum of 8 bands were implemented. The number of taps needed in each band of a subband scheme is not as straight-forward as in its fullband counterpart [71]. In figure 5.13, the stationary ERLE performance of each case (averaged from 8-10s) are compared for a range of values for the number of taps per band with all bands being equal. This plot helps to identify the optimum number of taps/band to maximize steady-state ERLE performance.

FIGURE 6.8 Step size vector used in the fullband ESNLMS algorithm in this chapter with speech.

FIGURE 6.9 ERLE performance of the fullband ESNLMS algorithm (trial #7, table 6.5) compared with the fullband NLMS algorithm with $\alpha=1$, $\psi=1$, and $N=1200$.

FIGURE 5.12 Overall frequency response of combined AFB+SFB system excluding the ADFs. 0.015dB ripple in passband, and -1 dB gain out-of-band near dc and 4kHz.

FIGURE 5.13 ERLE performance of 4-band, 8-band, and fullband NLMS vs # taps/band with $\alpha=1.0$ in all cases, and the NT phone excited with bandlimited white noise in Minto 2014.

The apparent improvement in steady-state performance with the subband NLMS algorithm over that of the fullband NLMS was also reported by Wallace [59] and Ryan [71]. Ryan proposed that the improvement was due to additional cancellation by the SFB. In our case, since the overall system is flat in the passband to within 0.015dB additional cancellation by the filterbanks is unlikely. Given a sufficient amount of data, both the subband and fullband NLMS algorithms should theoretically converge to the same level of steady-state ERLE (or MSE). The improvement in steady-state ERLE (better referred to as the final observed stationary ERLE) with the subband NLMS algorithm is therefore caused by the combination of an improvement in convergence speed and an insufficient amount data. Improved tracking performance also results in a higher level of ERLE during stationary periods since in real environments stationary conditions seldom exist.

In figures 5.14 and 5.15 the ERLE performance of 4-band, 8-band, and fullband NLMS algorithms are compared with the number of taps in each band fixed at 1200, and the step size set to 1.0. Averaged performance figures are listed in table 5.5 for stationary and nonstationary time-windows. Both subband cases improved stationary ERLE (8-10s window) by 2dB over the fullband NLMS algorithm. The tracking performance of the 4-band and 8-band algorithms were 2 dB and 5 dB better than their full-band counterpart. The subband NLMS algorithm tracks room nonstationarities more effectively than the fullband NLMS algorithm because of the inherent normalization per band.

TABLE 5.5 Quantitative comparison of the ERLE Performance of the SBNLMS algorithm with bandlimited white noise.

SBNLMS	Avg. Stat. ERLE (dB) (2.5-3.5s)	Avg. Stat. ERLE (dB) (8-10s)	Avg. Nonstat ERLE (dB) (3.7-6.7s)
4 bands, N=1200, $\alpha=1.0$	30.6	34.3	17.6
8 bands, N=1200, $\alpha=1.0$	33.1	34.0	20.4
NLMS, $\alpha=1.0$	29.8	32.0	15.0
SFTF, $\lambda_{acc}=1-1/N$	33.1	33.0	15.0
ESNLMS, trial #8	29.9	31.5	22.4

FIGURE 5.14 ERLE Performance of a 4-band NLMS algorithm with 1200 taps/band, and $\alpha=1.0$ compared with the NLMS algorithm with $\alpha=1$.

FIGURE 5.15 ERLE Performance of an 8-band NLMS algorithm with 1200 taps/band, and $\alpha=1.0$ compared with the NLMS algorithm with $\alpha=1$.

5.6 The Subband ESNLMS Algorithm

The improvement in tracking performance observed in previous sections with the fullband ESNLMS and subband NLMS algorithms over the fullband NLMS algorithm is improved further with the subband ESNLMS algorithm examined in this section. The subband ESNLMS algorithm is similar to the subband NLMS algorithm with the exception that the coefficients in each band are adapted using the ESNLMS algorithm instead of the NLMS algorithm. This section examines the performance of the subband ESNLMS algorithm by comparing it against that of the fullband NLMS, fullband ESNLMS, and subband NLMS algorithms. A simple technique is also described to simplify the task of selecting the decay factors in band.

The subband *exponentially-decaying step size* NLMS algorithm (ESNLMS) is obtained by merging the subband NLMS and fullband ESNLMS algorithms. A block diagram of the proposed structure is shown in figure 5.16. The reference signal, $x(n)$, and primary signal, $d(n)$, are decomposed into M subband signals, $x_k(n)$ and $d_k(n)$ respectively ($k=1..M$), by independent analysis filterbanks (AFB). The subband reference signals, $x_k(n)$, drive a parallel bank of M autonomous ESNLMS adaptive filters. The estimation error associated with each adaptive filter is obtained by subtracting $y_k(n)$ from its associated $d_k(n)$ in each band. The unified system error signal, $e(n)$, is found by a two stage process in which the subband error signals are first filtered by an appropriate synthesis filterbank (SFB) and then summed together. The overall system ERLE is computed using $e(n)$ and $x(n)$, while $e_k(n)$ and $x_k(n)$ can be used to find the ERLE in each subband, if desired. The reference signal, $x(n)$, must be delayed to account for the delay through the AFB and SFB filterbanks or the resulting system ERLE will be inaccurate.

FIGURE 5.16 Block diagram of the subband ESNLMS algorithm

Implementing the subband ESNLMS algorithm requires a number of parameters to be specified, including the number of bands, step size vector and number of taps in each band, and the characteristics of the AFB and SFB filterbanks. For consistency, the AFB and SFB filterbanks used in the simulations presented in this section are identical to those used in the subband NLMS simulations in section 5.5. The number of bands is fixed at eight (8) to maintain reasonable simulation times, and the number of taps/band is set to 1200 to maximize ERLE performance (see section 5.5). Unfortunately, the parameters of the step size vector in each band can not be specified as easily as in the fullband case simply because of the number of permutations, and hence simulations that would result. An automated or controlled method of obtaining parameters which are tailored to room characteristics is needed.

The reverberation time of a room is defined as the time required for sound to decay by -60dB. It is determined by a complicated set of variables related to the size and shape of the room, and the room's average absorption coefficient. In real rooms, the absorption coefficient, and hence its reverberation time, is frequency dependent [44]. This suggests a possible advantage in optimizing each band of the subband ESNLMS algorithm to the frequency dependent characteristics of a room. One method of estimating these decay

factors is illustrated in figure 5.17 and is based on decomposing the fullband LREM impulse response as proposed in [82].

The fullband LREM impulse response obtained by standard adaptive system identification (section 2.2) is first decomposed by an AFB into M bands. The AFB is chosen to correspond to that used in the subband ESNLMS algorithm. The resulting subband impulse responses are then analyzed to determine the effective decay factor in each band. This is achieved by calculating the slope of the exponentially-decaying envelope of each subband LREM impulse response on a logarithmic scale. The resulting slope determines the optimal decay factors, γ_i , for each band.

FIGURE 5.17 Technique to estimate the decay factors in each band of a subband ESNLMS algorithm.

For the Minto 2014 conference room and the NT phone the resulting optimal decay factor for each band is plotted in figure 5.18. Five simulations were conducted using these decay factors and a variety of other parameters to completely specify each step size vector. The performance data for each trial are summarized in table 5.6 including average stationary and nonstationary ERLE performance. The performance of the fullband NLMS, accelerated SFTF, fullband ESNLMS, and 8-band NLMS algorithms are included in the table for comparison purposes.

FIGURE 5.18 Estimated 8-band decay characteristics of the LREM impulse response measured in Minto 2014 with the NT phone.

The subband ESNLMS algorithm is seen to out-perform all other algorithms examined in this thesis in terms of tracking performance. The greatest improvement in average nonstationary ERLE performance was obtained in trial #1, while the fastest convergence performance was obtained in trial #4. These trials are compared graphically with the fullband NLMS algorithm in figures 5.19 and 5.20, and compared with each other in figure 5.21.

TABLE 5.6 Quantitative comparison of the ERLE Performance of the 8-band ESNLMS algorithm with white noise.

#	Parameters of Step Size Vector (same for each band)			Avg. Stat. ERLE (dB) (2.5-3.5s)	Avg. Stat. ERLE (dB) (8-10s)	Avg. Nonstat. ERLE (dB) (3.7-6.7s)
	α_{min}	$\bar{\alpha}$	M			
1	0.1	1.0	1	30.1	37.5	30.3
2	0.1	1.0	50	30.3	36.8	28.0
3	0.1	1.0	100	30.3	36.6	26.7
4	0.2	0.5	1	36.0	37.6	25.5
5	0.05	1.0	50	31.1	35.9	27.6
NLMS, $\alpha=1.0$				29.8	32.0	15.0
SPTF, $\lambda_{acc}=1-1/N$				33.1	33.0	15.0
ESNLMS, trial #8				29.9	31.5	22.4
8-band NLMS, $\alpha=1$				33.1	34.0	20.4

FIGURE 5.19 ERLE Performance of the 8-band ESNLMS algorithm (trial #1, table 5.6) compared with the fullband NLMS algorithm with $\alpha=1$; $N=1200$.

In terms of average nonstationary ERLE performance, the subband ESNLMS (trial #1, table 5.6) exhibited improvements over the fullband NLMS, accelerated SFTF, fullband ESNLMS (trial #8, table 5.4), and 8-band NLMS of 15.3dB, 15.3dB, 7.9dB, and 9.9dB respectively. The performance of the subband ESNLMS (trial #1, table 5.6) is compared graphically with the fullband ESNLMS (trial #8, table 5.4) and 8-band NLMS algorithms in figure 5.22 and 5.23 respectively.

FIGURE 5.20 ERLE Performance of the 8-band ESNLMS algorithm (trial #4, table 5.6) compared with the fullband NLMS algorithm with $\alpha=1$; $N=1200$.

FIGURE 5.21 ERLE Performance of 8-band ESNLMS (trial #1, table 5.6) compared with 8-band ESNLMS (trial #4, table 5.6).

FIGURE 5.22 ERLE Performance of 8-band ESNLMS (trial #1, table 5.6) compared with fullband ESNLMS (trial #8, table 5.4) ; $N=1200$.

FIGURE 5.23 ERLE Performance of 8-band ESNLMS algorithm (trial #1, table 5.6) compared with 8-band NLMS algorithm with $\alpha=1$; $N=1200$.

The use of different decay factors in each band of the 8-band ESNLMS algorithm improved tracking performance because it was better equipped to track the nonstationarity. The performance of the ESNLMS algorithm was shown by Yuan [105] to be relatively insensitive to the decay factor and initial step size parameters. These results are valid for the subband ESNLMS algorithm as well. Unfortunately, the additional freedom in being able to vary the decay parameter in each band leads to a more complicated sensitivity issue. The problem becomes one of understanding the sensitivity of the algorithm's tracking performance to the use of the *optimum* decay factors in each band verses the use of a constant decay factor in all bands. An additional simulation was performed using a constant decay factor of 0.02 in all bands instead of that shown in figure 5.18. The average nonstationary ERLE performance dropped from 30.3dB (optimum decay factors) to 28.5dB for a loss of -5.9%.

5.7 Complexity of the Subband ESNLMS Algorithm

The complexity of a subband adaptive filtering algorithm depends on the complexity of the analysis and synthesis filterbanks, the adaptive filters used in each band, and the sampling rate of the subband signals. Decimation is typically employed to reduce the complexity of a subband algorithm by reducing the sampling rate in the subbands [28] [61] [71] [75] [89] [94] [116]. In fact, subband algorithms have been used to *reduce* the complexity of the NLMS algorithm in processor limited applications [94].

The complexity of the 8-band ESNLMS adaptive algorithm implemented in section 5.6 is summarized below:

AFB/SFB Filtering Stage:

2 AFB filters of 128 taps each = 256 MA/sample

+ 1 SFB filter of 128 taps each = 128 MA/sample

Total = 384 MA/sample

Subband ADF Filtering Stage:

Eight 1200 tap subband FIR filters = 9600 MA/sample

Subband ADF Adaptation:

Adapting 8 subband ADFs = 8 * 3600 MA/sample

= 28800 MA/sample

Total number of MA/sample required:

38784 MA/sample

This represents an increase in computational complexity of a factor of eight over the fullband NLMS algorithm. To support a *best-case* analysis of each algorithm, complexity issues were ignored in this thesis; the subbands were not decimated, and the AFB/SFB filters were not optimized. A complete investigation of these implementation issues is suggested for future research.

5.8 Discussion and Summary

This chapter examined the tracking performance of a variety of linear adaptive filtering algorithms for acoustic echo cancellation in handsfree teleconferencing systems excited with bandlimited white noise. Results were based on data captured in a real conference room using a commercially available conference phone, with nonstationary conditions introduced by moving objects near the phoneset.

The tracking performance of the fullband NLMS and accelerated SFTF algorithms was shown to be relatively inefficient for the type of nonstationarities found in typical conference rooms. Improvements in the tracking performance of the accelerated SFTF algorithm is possible by using more complex techniques like periodic re-initialization or alternate windows. The accelerated SFTF algorithm definitely converged faster than the fullband NLMS algorithm, however.

Improved tracking performance was obtained by concentrating the adaptation to the first several hundred taps of the adaptive filter. This was accomplished with the ESNLMS algorithm. An improvement of 7 dB (best-case) was observed over the fullband NLMS algorithm. The ESNLMS algorithm was then extended to a subband approach, and additional improvements in tracking performance were obtained. An improvement of 15.3dB in tracking performance was demonstrated with an 8-band ESNLMS algorithm, representing an improvement of 102% over the fullband NLMS algorithm. Optimum performance, however, requires the use of step size parameters that are specific to a given conference room. Fortunately, the sensitive of the algorithm to these parameters appears to be relatively small. The main drawbacks of the subband ESNLMS algorithm is the inherent delay through the AFB/SFB filterbanks.

Chapter 6

6.0 SPEECH RESULTS

6.1 Experimental Conditions

The experimental conditions adopted in this chapter parallel those of the previous chapter with the sole exception that speech was used to excite the loudspeaker instead of bandlimited white noise. Unfortunately, severe distortion introduced by the enclosure [111] limited the sound pressure of the speech in the room to an average level of approximately 60-64 dBSPL measured at 1m from the phone. Volume levels were reduced to minimize the magnitude of the distortion. A complete investigation of its effects on ERLE performance including potential compensation techniques is beyond the scope of this thesis [111].

A second problem in working with speech files is the potential sensitivity of the final results to the actual characteristics of the speech signal used as the excitation signal. The speech signal used throughout this chapter was obtained from Bell Northern Research as a result of extensive analysis of typical conversations on telecommunication networks. A time-series plot of the 2.85s speech signal is shown in figure 6.1. In speech analysis, a spectrogram typically yields significantly more information than a simple time-series plot since a spectrogram illustrates how the spectrum or energy distribution of a signal changes with time. Spectrograms are an estimate of the short-term, time-localized frequency content of a signal. The standard method of computing a spectrogram involves computing a set of FFTs of non-overlapping segments of the signal each windowed by an appropriate windowing function and padded with zeros if necessary [41].

Using segments of 128 point each, a Hanning window, and 512 point FFTs, a spectrogram of the speech excitation signal was obtained. The result is plotted in figure 6.2. The black

areas correspond to peaks in the spectrogram, while the white areas represent valleys. It is, perhaps, important to note that multiple spectrograms are needed to accurately interpret the characteristics of real speech signals because of the conflicting requirements between time and frequency resolution in computing spectrograms. Good time resolution requires short windows, while good frequency resolution requires long windows. The parameters used to display the spectrogram in figure 6.2 were selected as a compromise to display the formant tracks clearly [43]. Of prime importance in adaptive filtering is simply that the spectrum is coloured and time-varying.

FIGURE 6.1. Time-series plot of the speech excitation signal.

A 20 second excitation signal was created by concatenating seven of the 2.85s speech signals described above. The resulting excitation signal is depicted in figure 6.3. Of particular importance is the bursty nature of the signal which is characteristic of typical conversations. In adaptive filtering, these quiescent periods represent periods of low excitation (low SNR) and require special precautions to prevent divergence of the adaptive weights of the AEC. Room nonstationarities were introduced during the fifth burst (11 to 14 sec. mark) using the procedure described in the previous chapter.

FIGURE 6.2 Spectrogram plot of the speech excitation signal.

FIGURE 6.3 Time-series plot of the complete excitation signal.

6.2 The Fullband NLMS Algorithm

6.2.1 Performance with Adaptation Halted

During periods of low excitation the step size may become quite large due to the division operation in equation (2.5) and cause the coefficients to diverge away from the optimum solution. The constant ε in the denominator of equation (2.5) can be used to limit the maximum value of the step size and therefore prevent divergence to a certain degree. While such a scheme might be useful in some extreme circumstances, typically it is more desirable to simply halt the adaptation during these periods since insufficient excitation energy is present to excite the unknown system. The algorithm was modified to halt adaptation by setting $\alpha=0$ whenever the power in the input tap-delay line fell below a pre-determined threshold (ψ). A suitable threshold for a 1200 tap NLMS algorithm with the given data was found to be $1/1200^{\text{th}}$ of the total power computed as $X_N(n)^T \cdot X_N(n)$. The short-term ERLE performance with $\psi=0$ (halting disabled) and $\psi=1/1200$ (halting enabled) is plotted in figure 6.4. ε is set to 10^{-6} in all simulations.

FIGURE 6.4 ERLE performance of a 1200 tap NLMS algorithm with parameters $\mu=1.0$, $\varepsilon=10^{-6}$, $\psi=1/1200$ (halting enabled), compared with the same algorithm $\psi=0$ (halting disabled).

Table 6.1 compares the effects of halting the adaptation in quantitative terms. The *stationary* and *nonstationary* ERLE values were obtained by averaging the short-term ERLE performance over the seventh and fifth bursts respectively. By halting the adaptive process where appropriate, a steady-state ERLE of 20.6dB (average) was obtained, compared to only 12dB when no halting was used. The divergence in the coefficients between speech bursts when the adaptation was *not* halted forced the algorithm to re-converge during each burst leading to inferior ERLE performance. The difference in performance during the nonstationary burst is a bit misleading, however, because of the difference in the state of convergence between the two algorithms.

TABLE 6.1 Effect of halting the adaptation on the ERLE performance of the NLMS algorithm.

Conditions of the NLMS algorithm	Stationary ERLE (Speech burst # 7)	Nonstationary ERLE (Speech burst #5)
Update halted during periods of low excitation.	20.6 dB	14.6dB
No halting performed.	12.0 dB	9.8 dB

6.2.2 Effect of Step Size

To illustrate the effect of changing the step size parameter, α , in the NLMS algorithm on steady-state and dynamic ERLE performance, two sufficiently extreme values were selected for comparison. Figure 6.5 compares the ERLE performance of a 1200 tap NLMS algorithm with step sizes of $\alpha=0.1$ and $\alpha=1.0$. The adaptive weights were frozen between speech bursts to prevent divergence as described in the previous section.

In table 6.2 the ERLE performance with $\alpha=1.0$ and $\alpha=0.1$ is quantified for the same stationary and nonstationary bursts averaged in the previous section. Of particular importance here is the extremely slow *best-case* convergence of the NLMS algorithm with speech excitation (figure 6.5). With $\alpha=1.0$ the algorithm converged in approximately 10s to a steady-state ERLE of 20.6dB, and achieved an average nonstationary ERLE of 14.6dB. With $\alpha=0.1$, a steady-state ERLE of only 11.6dB was achieved after 20s of convergence.

FIGURE 6.5 ERLE performance of the NLMS algorithm with adaptation halted for $\alpha=0.1$ and $\alpha=1.0$ with $N=1200$, $\varepsilon=10^{-6}$, $\psi=1$.

TABLE 6.2 Effect of step size on ERLE performance of the NLMS algorithm with speech.

Normalized Step Size Constant	Stationary ERLE (Speech burst #7)	Nonstationary ERLE (Speech burst #5)	Rate of Convergence (seconds)
$\alpha=0.1$	11.6 dB	10.0dB	>20s
$\alpha=1.0$	20.6 dB	14.6dB	approx. 10s

6.3 The Accelerated SFTF Algorithm

In this section the steady-state and dynamic performance of the accelerated SFTF algorithm (with exponential weighting) is examined. Its dynamic performance is expected to exceed that of the NLMS algorithm with speech signals because the rate of convergence of the SFTF algorithm is known to be insensitive to the degree of colour in the signal.

The performance of the SFTF algorithm in terms of short-term ERLE is plotted in figure 6.6 and 6.7. The performance of the NLMS algorithm with a similar level of misadjustment is superimposed on the same graph for comparison purposes. A summary of the relevant simulation parameters for the NLMS and SFTF algorithms used in generating this plot are given in table 6.3. Setting the soft initialization constant, ζ , in the SFTF algorithm to a value of 50 was found sufficient to ensure numerical stability of the forward and backward prediction filters.

FIGURE 6.6 ERLE Performance of SFTF and NLMS algorithms excited with speech.

TABLE 6.3 Parameter list for the NLMS and SFTF algorithms.

NLMS Algorithm		SFTF Algorithm	
Number of Taps	1200	Number of Taps	1200
Step Size	$\alpha = 1.0$	Forgetting Factor	$\lambda = 1 - \frac{1}{3N}$
Halting Threshold	$\psi = 1.0$	Acceleration Factor	$\gamma = 0.95$
Stepsize Bounding Parameter	$\epsilon = 10^{-6}$	Soft initialization constant	$\zeta = 50$

FIGURE 6.7 Tracking performance of SFTF and NLMS algorithms excited with speech.

The convergence performance of the SFTF algorithm exceeds that of the NLMS algorithm with highly coloured signals. The SFTF algorithm converged in less than 1s, while close to 10s was required for the NLMS algorithm to converge.

Average stationary and nonstationary ERLE results are summarized in table 6.4. Additional simulations were also performed using smaller values for λ to more fully characterize the performance of the SFTF algorithm. Similar comparisons with the NLMS algorithm were not possible due to its relatively slow *best-case* convergence.

For the 20s time-window used in the simulations, the SFTF algorithm demonstrated superior steady-state and convergence performance over the NLMS algorithm. The best-case steady-state ERLE of the SFTF algorithm was 21.5dB, approximately 1dB greater than that of the NLMS algorithm. In terms of tracking performance the SFTF algorithm was more than 2dB better than the NLMS algorithm.

TABLE 6.4 Quantitative comparison of the ERLE performance of the SFTF and NLMS algorithms with speech excitation.

SFTF Algorithm			NLMS Algorithm		
Forgetting Factor, λ	Stationary ERLE (dB)	Nonstationary ERLE (dB)	Step Size Constant, μ	Stationary ERLE (dB)	Nonstationary ERLE (dB)
$1 - \frac{1}{N}$	19.5	15.2	1.0	20.6	14.6
$1 - \frac{1}{10N}$	21.4	16.6	0.1	11.63	10.0
$1 - \frac{1}{50N}$	21.4	17.0			
$1 - \frac{1}{100N}$	21.5	16.7			
$1 - \frac{1}{200N}$	21.5	16.6			
1	21.5	16.4			

In general, the SFTF algorithm suffers from a similar divergence of the coefficients during periods of low excitation as that of the NLMS algorithm. Extended periods of low excitation cause a *locking* of the prediction variables of the SFTF algorithm [106]. To prevent *locking* all prediction variables, including the dual Kalman gain vector, should be re-initialized whenever the likelihood variable falls below a preset threshold. A threshold of 0.1 was found sufficient to prevent locking of the algorithm with the data used in this chapter.

6.4 The Fullband ESNLMS Algorithm

In this section the performance of the fullband ESNLMS algorithm with speech is examined. Since adjacent samples of speech are highly correlated, restrictions must be placed on each element of the step size vector (as opposed to just the average with white noise) to prevent divergence. Limiting each element to a maximum value of 2.0 was found sufficient in this work, and also in [92], to prevent divergence.

The algorithm was examined through a set of 15 simulations with results summarized in table 6.5. Due to the restrictions placed on the step size vector, obtaining a steady-state and/or tracking performance advantage over the NLMS is more difficult. Instead of fixing $\alpha_{\min}$, and varying $\alpha_{\max}$ to obtain a desired $\bar{\alpha}$ (as in the previous chapter), $\alpha_{\max}$ must be fixed at 2.0 while $\alpha_{\min}$ is varied. When $\bar{\alpha}$ is reasonably large, however, $\alpha_{\min}$ settles close to one diminishing any potential performance advantage (see trials #1-3). As $\bar{\alpha}$ is reduced, $\alpha_{\min}$ is also reduced and tracking performance and steady-state performance are improved. The best performance was obtained in trial #7 with $\gamma=0.005$, $\bar{\alpha} = 0.335$, $\alpha_{\min}=0.005$, $\alpha_{\max}=2.0$, and $M=1$. Steady-state performance improved by 0.4dB to 21dB, and nonstationary performance improved by 1.8dB to 16.4dB over the *best-case* fullband NLMS algorithm. The step size vector and short-term ERLE performance of trial #7 is plotted in figures 6.8 and 6.9 respectively. A close-up of the nonstationary ERLE performance during the fifth burst is reproduced in figure 6.10 for clarity. A rather serious short-coming of the fullband ESNLMS algorithm observed in figure 6.9 is that its convergence performance is similar to the fullband NLMS algorithm; requiring 10s to converge. Alternative techniques are needed to improve convergence performance with coloured signals.

FIGURE 6.8 Step size vector used in the fullband ESNLMS algorithm in this chapter with speech.

FIGURE 6.9 ERLE performance of the fullband ESNLMS algorithm (trial #7, table 6.5) compared with the fullband NLMS algorithm with $\alpha=1$, $\psi=1$, and $N=1200$.

TABLE 6.5 Quantitative comparisons of the ERLE Performance of the ESNLMS Algorithm with speech excitation.

#	Parameters of Step Size Vector				Mean Steady-State ERLE (dB) (7 th burst)	Mean Nonstationary ERLE (dB) (5 th burst)
	γ	$\bar{\alpha}$	$\alpha_{\min}$	M		
1	0.005	1.0	0.95	1	20.4	14.6
2	0.01	1.0	0.98	1	20.5	14.7
3	0.02	1.0	0.99	1	20.6	14.7
4	0.005	0.5	0.32	1	21.1	16.1
5	0.01	0.5	0.43	1	21.0	16.0
6	0.02	0.5	0.47	1	20.8	15.9
7	0.005	0.335	0.005	1	21.0	16.4
8	0.01	0.2	0.06	1	19.9	16.1
9	0.02	0.2	0.14	1	19.6	15.9
10	0.005	1.0	0.89	50	20.2	14.5
11	0.01	1.0	0.93	50	20.3	14.5
12	0.02	1.0	0.95	50	20.3	14.5
13	0.005	0.5	0.21	50	21.0	16.0
14	0.01	0.5	0.35	50	20.9	15.9
15	0.02	0.5	0.40	50	20.9	15.9
NLMS, $\alpha=1$, $\psi=1$					20.6	14.6
SFTF, $\lambda=1-1/N$					19.5	15.2

FIGURE 6.10 Tracking performance of the fullband ESNLMS algorithm (trial #7, table 6.5) compared with the fullband NLMS algorithm with $\alpha=1$, $\psi=1$, $N=1200$ (closeup view).

6.5 The Subband NLMS Algorithm

In the previous chapter the subband NLMS algorithm was shown to have a steady-state and tracking advantage over its fullband counterpart with bandlimited white noise. Similar advantages are expected with speech. The subband algorithm (including parameters) used in this section is identical to that used in the previous chapter, and are not re-iterated here. Figures 6.11 and 6.13 plot the ERLE performance of 4-band and 8-band NLMS algorithms respectively. A close-up plot of the tracking performance of each case is given in figures 6.12 and 6.14. Quantitative results are summarized in table 6.6

As expected, the subband approach improves convergence, tracking, and steady-state performance over its fullband counterpart. In fact, the 4-band (and 8-band) NLMS algorithm converged almost as quickly as the SFTF algorithm, in approximately 1 second. The 4-band NLMS algorithm gave a steady-state ERLE performance of 22.3dB,

and a nonstationary tracking performance of 18dB. The 8-band NLMS algorithm, however, produced 22.8dB of steady-state ERLE, and 18.8dB of nonstationary ERLE. This equates to an improvement of 2dB in steady-state ERLE, and 4dB in nonstationary ERLE performance due to improved tracking. The subband NLMS algorithm proved superior to the fullband ESNLMS algorithm as well; 2dB steady-state advantage, 2.4dB nonstationary advantage.

TABLE 6.6 Quantitative comparison of the ERLE performance of 4 and 8 band NLMS algorithms with speech excitation.

SBNLMS	Avr. Steady-State ERLE (dB) (7th burst)	Avr. Nonstationary ERLE (dB) (5th burst)
4 bands, 1200taps, $\alpha=1.0$	22.3	18.0
8 bands, 1200taps, $\alpha=1.0$	22.8	18.8
NLMS, $\alpha=1.0$, $\psi=1.0$	20.6	14.6
SFTF, $\lambda=1-1/N$	19.5	15.2
ESNLMS, trial #7	21.0	16.4

FIGURE 6.11 ERLE performance of the 4-band NLMS algorithm with speech, 1200taps/band, $\alpha=1.0$, $\psi=1.0$ and compared with the fullband NLMS algorithm with similar parameters.

FIGURE 6.12 Tracking Performance of 4-band NLMS algorithm with speech, compared with the fullband NLMS algorithm (close-up view)

FIGURE 6.13 ERLE performance of 8-band NLMS algorithm with speech, 1200taps/band, $\alpha=1.0$, $\psi=1.0$, compared with the fullband NLMS algorithm with similar parameters.

FIGURE 6.14 Tracking performance of the 8-band NLMS algorithm with speech, compared with the fullband NLMS algorithm (close-up view).

6.6 The Subband ESNLMS Algorithm

The final algorithm investigated in this chapter is that of the subband ESNLMS algorithm. Only the 8-band algorithm was simulated, due to its similarity with the 4-band structure. 1200 taps were used in each band, the halting threshold was set $\psi=1.0$, and the step size vector was defined according to the parameters obtained in the previous chapter with noise excitation. The performance of the 8-band ESNLMS is compared with that of the fullband NLMS algorithm in figures 6.15 and 6.16. Quantitative results are summarized in table 6.7.

The 8-band ESNLMS algorithm achieved a nonstationary ERLE of 20.3 dB, and a steady-state ERLE of 23.8dB. This represents an improvement in nonstationary ERLE of 5.8dB over the fullband NLMS and SFTF algorithms, 4dB over the fullband ESNLMS algorithm, and 1.6dB over the 8-band NLMS algorithm. The algorithm also converged more quickly than the fullband NLMS, and ESNLMS algorithms.

TABLE 6.7 Quantitative comparison of the ERLE performance of the 8 band ESNLMS algorithm with speech excitation.

Algorithms	Avr. Steady-State ERLE (dB) (7 th burst)	Avr. Nonstationary ERLE (dB) (5 th burst)
parameters identical to trial #1 of 8-band ESNLMS in table 5.6, chapter 5.	23.8	20.4
NLMS, $\alpha=1.0$, $\psi=1.0$	20.6	14.6
SFTF, $\lambda=1-1/N$	19.5	15.2
ESNLMS, trial #7	21.0	16.4
8-band NLMS, $\alpha=1.0$	22.8	18.8

FIGURE 6.15 ERLE performance of the 8-band ESNLMS algorithm, 1200taps/band, $\psi=1.0$ (table 6.7), compared with the fullband NLMS algorithm with $\alpha=1$, $\psi=1$.

FIGURE 6.16 Tracking performance of the 8-band ESNLMS algorithm with speech compared with the fullband NLMS algorithm (close-up view).

6.7 Discussion and Summary

In this chapter the steady-state, convergence, and tracking performance of a variety of fullband and subband adaptive filtering algorithms excited with speech signals in acoustic echo cancellers were investigated. The fullband NLMS and ESNLMS algorithms proved to be relatively ineffective at tracking nonstationarities with speech signals, and demonstrated poor convergence performance caused primarily by the coloured nature of the speech signal. The use of the accelerated SFTF algorithm improved convergence performance, due to its insensitivity to eigenvalue disparity, but demonstrated only a minor improvement in tracking performance over the fullband NLMS algorithm. The exponential characteristics of the window, and the high filter order required to achieve sufficient steady-state ERLE degraded tracking performance. The use of more complex techniques like periodic re-initialization or alternate windows might support additional improvements in tracking performance.

With bandlimited white noise, the ESNLMS algorithm was found to improve tracking performance quite dramatically. With speech, however, the advantage was much less dramatic because of the additional restrictions on the elements of the step size vector (with highly correlated signals) to prevent divergence. The convergence of the ESNLMS algorithm proved to be similar to that of the NLMS algorithm due to its sensitivity to the degree of colour in the excitation signal.

The use of subband adaptive filtering techniques provided an improvement in convergence, and tracking performance over the fullband algorithms. Convergence performance was improved because the step size vector was normalized to the power in each band which reduced its sensitivity to eigenvalue disparity. An improvement in tracking performance was obtained through the use of exponentially-decaying step size vectors in each band.

Chapter 7

7.0 CONCLUSIONS

This thesis has examined the tracking performance of acoustic echo cancellers operating in handsfree teleconferencing systems using adaptive digital signal processing techniques. A high-quality experimental setup was developed, tested, and used to capture real data in anechoic and typical conference rooms for use in computer simulations. These simulations were conducted using both bandlimited white noise and simulated speech excitation signals. Nonstationary conditions were introduced by moving objects near the speakerphone.

The achievable ERLE of the experimental setup was investigated to verify the accuracy of final results, and to examine the main physical limitations of AECs. The setup itself was found to be fundamentally limited to a maximum ERLE of 56dB when the unknown system was replaced with an attenuator box (no acoustics). The achievable ERLE dropped to 37dB once the effects of the speakerphone's enclosure were taken into account, and further to 30dB to 33dB (different rooms) when room reverberation and background acoustic noise effects were incorporated. Truncation, finite precision, and nonlinear effects were also found to degrade ERLE performance at soft and loud volume levels.

The tracking performance of the fullband NLMS algorithm was investigated to form the baseline reference to which all other algorithms would be compared. The fullband NLMS algorithm demonstrated poor convergence and tracking performance with bandlimited white noise and simulated speech. When excited with bandlimited white noise, convergence required approximately 4s, and ERLE performance dropped by 55% during the nonstationarity. When excited with simulated speech, 10s was required to converge to steady-state, and a drop of 29% was observed in ERLE performance during the nonstationary event.

The accelerated SFTF algorithm was then examined to improve tracking performance over the fullband NLMS algorithm based on promising results presented in the literature. The main advantage of the SFTF algorithm is that its performance is insensitive to the colour of the input signal. Convergence was improved by a factor of 4 for noise and a factor of 10 for simulated speech. A 4% improvement in tracking performance was also observed with simulated speech, while no improvement was observed with noise excitation. Improvements in the tracking performance of the accelerated SFTF algorithm should be possible by using more complex techniques like periodic re-initialization, or alternate data windowing functions.

The fullband ESNLMS algorithm was investigated next, and found to improve tracking performance dramatically over the fullband NLMS algorithm when excited with noise. Improvements of 50% for noise and 12% for speech were demonstrated. Unfortunately, its performance was found to be as sensitive to the degree of colour in the input excitation signal as that of the fullband NLMS algorithm. No improvement in convergence performance was observed with the fullband ESNLMS excited with simulated speech over the fullband NLMS algorithm.

Next, the subband NLMS algorithm was investigated. Improved tracking, and convergence performance were observed with noise and speech inputs. Improvements of 16% in tracking performance was observed for noise excitation, while improvements of 29% for tracking performance was obtained with speech inputs.

The final algorithm examined was the subband ESNLMS algorithm. This algorithm has the ability to model frequency dependent decay characteristics which occur in typical conference rooms due to variations in the absorption characteristics of the room with frequency. Its performance is relatively insensitive to the colour in the excitation signal as well. This results in improved convergence, and tracking performance over the fullband NLMS algorithm. Improvements of 102% in tracking performance with noise, and 39% with speech were obtained when the parameters of the algorithm were matched to

the characteristics of the room. Fortunately, its performance appears to be reasonably insensitive to its parameters.

7.1 Summary of Contributions

The following list summarizes the main contributions of this thesis.

- Participation in the design and optimization of a high-quality real-time data acquisition system for the evaluation of AECs.
- Identification of the practical limitations of AEC.
- Evaluation of the tracking performance of fullband and subband adaptive filtering algorithms for AEC with band-limited white noise and speech signals.
- Application of the accelerated SFTF algorithm with re-initialization to AEC.
- Application of the fullband and subband ESNLMS algorithms to AEC.

7.2 Suggestions for Future Work

As a result of the work done in this thesis, the following areas are suggested for further research.

Real-Time Implementation.

Real-time implementation of the structures proposed in this thesis is an obvious extension to this work. Of particular interest would be the selection of the most appropriate architecture, its implementation, and limitations caused by such an implementation. Techniques to reduce the complexity of the subband ESNLMS algorithm proposed in this thesis would be required. The effects of near-end speech and double-talk could be examined.

Time-varying Algorithms.

A detailed investigation of time-varying algorithms including the SG-GALS, and TW-NLMS algorithms might lead to improved performance. Based on the work done in this thesis, incorporating a time-varying approach into the subband ESNLMS algorithm appears to be particularly promising.

Detailed Investigation of Periodically Initializing the SFTF Algorithm.

Based on the preliminary evidence provided in this thesis, a detailed investigation of the effect of re-initialization on tracking performance, and convergence rate of the accelerated SFTF algorithm appears worthwhile. The major challenge would be the investigation of techniques to support re-initialization at the most appropriate moments to maximize performance advantages.

Reduction of Enclosure Related Distortions.

A major limitation of current speakerphone design is the nonlinear distortion caused by rattling and whistling within the enclosure. A detailed investigation of the nature and characteristics of these types of disturbances might lead to a method of reducing their effects. Such a structure would achieve improved steady-state ERLE at mid to loud volumes, as well as improving perceived speech quality.

REFERENCES

- [1] O.M. Mracek Mitchell, and David A. Berkley, "A Full-Duplex Echo Suppressor Using Center-Clipping", Bell Sys. Tech. Journal, Vol. 50, No. 5, May/June 1971, pp. 1619-1631.
- [2] David A. Berkley, and Olga M. Mracek Mitchell, "Seeking the Ideal in 'Hands-Free' Telephony", Bell Laboratories Record, Nov. 1974, pp. 318-325.
- [3] F.B. Hilderbrand, Introduction to Numerical Analysis, Dover Publications Inc, New York, 1974.
- [4] N Ahmed and K.R. Rao, Orthogonal Transforms for Digital Signal Processing, Springer-Verlag, 1975.
- [5] Mauro Dentino, John McCool, and Bernard Widrow, "Adaptive Filtering in the Frequency Domain", Proceedings of the I.E.E.E., Vol. 66, No. 12, Dec 1978, pp. 1658-1659.
- [6] Francis A. Reed, and Paul L. Feintuch, "A Comparison of LMS Adaptive Cancellers Implemented in the Frequency Domain and the Time Domain.", I.E.E.E. Transactions on Circuits and Systems, Vol CAS-28, No. 6, June 1981, pp. 610-615.
- [7] Gregory A. Clark, Sanjit K. Mitra, and Sydney R. Parker, "Block Implementation of Adaptive Digital Filters", I.E.E.E. Transactions on ASSP, Vol. 29, No. 3, June 1981, pp. 744-752.
- [8] Kenneth R. Perry, "A Distributed- μ Implementation of the LMS Algorithm", I.E.E.E. Transactions on ASSP, Vol. 29, No. 3, June 1981, pp. 753-762.
- [9] Tatsuki Hayashi, et.al., "Echo Canceller with Effective Double Talk Control", I.E.E.E. Globecom 1983, Vol. 3, pp. 1389-1393.
- [10] S. Shankar Narayan, Allen M. Peterson, and Madihally J. Narasimha, "Transform Domain LMS Algorithm", I.E.E.E. Transactions on Acoustic, Speech, and Signal Processing, Vol. ASSP-31, No. 3, June 1983, pp. 609-615.
- [11] David G. Messerschmitt, "Echo Cancellation in Speech and Data Transmission", I.E.E.E. Jour. Sel. areas of Commun., Vol. 2, No. 2, Mar. 1984, pp. 18-28.
- [12] Bernard Widrow, and Wugene Walsh, "On the Statistical Efficiency of the LMS Algorithm with Nonstationary Inputs", I.E.E.E. Transactions on Information Theory, Vol IT-30, No. 2, March 1984, pp 211-221.
- [13] C.W.K. Gritton, D.W. Lin, "Echo Cancellation Algorithms", I.E.E.E. ASSP magazine, Apr. 1984, pp. 30-37.

- [14] John M. Cioffi, and Thomas Kailath, "Fast, Recursive-Least-Squares Transversal Filters for Adaptive Filtering", I.E.E.E. Transactions on ASSP, Vol. 32, No. 2, April 1984, pp. 304-337.
- [15] Yrjo Neuvo et.al., "Interpolated Finite Impulse Response Filters", I.E.E.E. Transactions on Acoustics, Speech, and Signal Processing, Vol ASSP-32, No. 3, June 1984, pp. 563-570.
- [16] Bernard Widrow, and Samuel D. Stearns, Adaptive Signal Processing, Prentice-Hall, 1985
- [17] S. Minami and T. Kawasaki, "A Double-Talk Detection Method for an Echo Canceller", International Conference on Communications, Vol. 3, 1985, pp 1492-1495.
- [18] John M. Cioffi, and Thomas Kailath, "Windowed Fast Transversal Filters Adaptive Algorithms with Normalization", I.E.E.E. Transactions on Acoustic, Speech, and Signal Processing, Vol. ASSP-33, No. 3, June 1985, pp. 607-625.
- [19] Kellermann, W., "Kompensation akustischer Echos in Frequenz-teilbandern", Frequenz, Vol. 39, No. 7/8, pp. 209-215, July/August 1985.
- [20] S. Thomas Alexander, Adaptive Signal Processing: Theory and Applications, Springer-Verlag, 1986.
- [21] David Hertz, David Mansour, and Isaac Engel, "On Least Square Frequency-Domain Adaptive Filters", I.E.E.E. Transactions on Circuits and Systems, Vol CAS-33, No. 3, March 1986, pp. 335-337.
- [22] Richard W. Harris, Douglas M. Chabries, and F. Avery Bishop, "A Variable Step (VS) Adaptive Filter Algorithm", I.E.E.E. Transactions on ASSP, Vol ASSP-34, No. 2, April 1986, pp. 309-316.
- [23] Jae Chon Lee and Chong Kwan Un, "Performance of Transform-Domain LMS Adaptive Digital Filters", I.E.E.E. Transactions on Acoustics, Speech, and Signal Processing, Vol ASSP-34, No. 3, June 1986, pp. 499-510.
- [24] S.T. Alexander, "Fast Adaptive Filter: A Geometrical Approach", I.E.E.E. ASSP magazine, Oct. 1986, pp. 18-28.
- [25] Evangelos Eleftheriou, and David D. Falconer, "Tracking Properties and Steady-State Performance of RLS Adaptive Filter Algorithms", I.E.E.E. Transactions on ASSP, Vol 34, No. 5, Oct 1986, pp. 1097-1109.
- [26] Maurice G. Bellanger, Adaptive Digital Filters and Signal Analysis, Marcel Dekker, Inc., 1987
- [27] John R. Treichler, C. Richard Johnson Jr, and Michael G. Larimore, Theory and Design of Adaptive Filters, John Wiley & Sons, 1987.
- [28] Gilloire, A., "Experiments with Sub-band Acoustic Echo Cancellers for Teleconferencing", Proc. ICASSP 1987, pp. 2141-2144.
- [29] Mohammed Reza Asharif et al., "Acoustic Echo Cancellation Based on Frequency Bin Adaptive Filtering (FBAF)", Proceedings Globecom, 1987, pp. 1940-1944.

- [30] W.K.Jenkins and D.F.Marshall, "On the Analysis and Design of Unitary Transformations for Adaptive FIR Digital filters", 29th Midwest Symposium on Circuits and Systems, 1987, pp. 534-537.
- [31] Guozhu Long, Dennis Shwed, and David D. Falconer, "Study of a Pole-Zero Adaptive Echo Canceller", I.E.E.E. Transactions on Circuits and Systems, Vol. CAS-34, No. 7, July 1987, pp. 765-769.
- [32] Piet C. W. Sommen et. al., "Convergence Analysis of a Frequency-Domain Adaptive Filter with Exponential Power Averaging and Generalized Window Function", I.E.E.E. Transactions on Circuits and Systems, Vol. CAS-34, No. 7, July 1987, pp. 788-798.
- [33] Eweda Eweda, and Odile Macchi, "Convergence of the RLS and LMS Adaptive Filters", I.E.E.E. Transactions on Circuits and Systems, Vol. CAS-34, No. 7, July 1987, pp. 799-803.
- [34] Bernard Widrow, et.al., "Fundamental Relations Between the LMS Algorithm and the DFT", I.E.E.E. Transactions on Circuits and Systems, Vol. CAS-34, No. 7, July 1987, pp. 814-820.
- [35] John M. Cioffi, "Limited-Precision Effects in Adaptive Filtering", I.E.E.E. Transactions on Circuits and Systems, Vol. CAS-34, No. 7, July 1987, pp. 821-833.
- [36] Jacques Sikorav, "Experiments to Identify and Track nonstationarities in Audio Conference Rooms", Proc. ICASSP 1988, Vol. 5, 1988, pp. 2566-2569.
- [37] A. Benallal and A. Gil'oire, "A new method to stabilize fast RLS algorithms on a first order model of the propagation of numerical errors.", Proc. ICASSP 1988, New York, pp.1365-1368.
- [38] Jiande Chen, Hugo Bes, Joos Vandewalle, and Paul Janssens, "A New Structure for Sub-band Acoustic Echo Canceler", Proc. ICASSP 1988, Vol. 5, 1988, pp. 2574-2577.
- [39] Walter Kellerman, "Analysis and Design of Multirate Systems for Cancellation of Acoustical Echoes", Proc. ICASSP 1988, Vol. 5, 1988, pp. 2570-2573.
- [40] Hong Fan, W.K.Jenkins, "An Investigation of an Adaptive IIR Echo Canceller: Advantages and Problems", I.E.E.E. Trans on ASSP, Vol. 36, No. 12, Dec. 1988, pp. 1819-1833.
- [41] Oppenheim and Shafer, Discrete-Time Signal Processing, Prentice-Hall, 1989
- [42] Ahmed Benallal and Andre Gilloire, "Improvement of the Tracking Capability of the Numerically Stable Fast RLS Algorithms for Adaptive Filtering", Proc. ICASSP 1989, pp. 1031-1034.
- [43] Sensimetrics, SpeechStation: User's Guide, Version 3.0, Semsimetrics Corp., Cambridge, Massachusetts, 1989-1993.
- [44] Donald G. Fink, and Donald Christiansen (editors) , Electronics Engineers' Handbook, 3rd ed., McGraw-Hill, 1989.

- [45] Wasfy B. Mikhael and Frank H. Wu, "A Fast Block FIR Adaptive Digital Filtering Algorithm with Individual Adaptation Parameters", I.E.E.E. Transactions on Circuits and Systems, Vol. 36, No. 1, Jan 1989, pp. 1-10.
- [46] Dirk T.M. Slock, and Thomas Kailath, "Fast Transversal Filters with Data Sequence Weighting", I.E.E.E. Transactions on Acoustics, Speech, and Signal Processing, Vol. 37, No. 3, March 1989, pp. 346-359.
- [47] Daniel F. Marshall, W. Kenneth Jenkins, and J.J. Murphy, "The Use of Orthogonal Transforms for Improving Performance of Adaptive Filters", I.E.E.E. Transactions on Circuits and Systems, Vol. 36, No. 4, April 1989, pp. 474-483.
- [48] John J. Shynk, "Adaptive IIR Filtering", I.E.E.E. Signal Processing Magazine, April 1989, pp. 4-21.
- [49] Hiroshi Yasukawa, Isao Furukawa, and Yasuzou Ishiyama, "Acoustic Echo Control for High Quality Audio Teleconferencing", Proc. ICASSP 1989, Glasgow Scotland, May. 1989, pp. 2041-2044.
- [50] A. Benallal, and A. Gilloire, "Improvement of the convergence speed and tracking capability of the numerically stable FLS algorithms for adaptive filtering", Proc. ICASSP 1989, Glasgow Scotland, May 1989, pp. 1031-1034.
- [51] Andre Gilloire, and Thierry Petillon, "A Comparison of NLMS and Fast RLS Algorithms for the Identification of Time-Varying Systems with Noisy Outputs", Signal Processing V: Theories and Applications, L. Torres et.al. editors, Elsevier Science Publishers B.V., 1990, pp. 417-420.
- [52] Kazuo Murano, Shigeyuki Unagami, and Fumio Amano, "Echo Cancellation Applications", I.E.E.E. Communications magazine, Jan. 1990, pp. 49- 55.
- [53] P.P. Vaidyanathan, "Multirate Digital Filters, Filter Banks, Polyphase Networks, and Applications: A Tutorial", Proceedings of the I.E.E.E., Vol. 78, No. 1, Jan 1990, pp. 56-93.
- [54] Carlos E. Davila, "A Stochastic Newton Algorithm with Data-Adaptive Step Size", I.E.E.E. Transactions on ASSP, Vol ASSP-38, No.10, Oct 1990, pp. 1796-1798.
- [55] Athanasios Papoulis, Probability, Random Variables, and Stochastic Processes, 3rd ed., McGraw Hill, 1991
- [56] Simon Haykin, Adaptive Filter Theory, 2nd ed., Prentice-Hall, 1991
- [57] C. Weiss, F. Berlin, D. Geller, "Advanced voice communication with acoustic echo cancelling and enhanced sound quality", ICC 91. International Conference on Communications, Vol. 1, 1991, pp. 478-83.
- [58] Ki-Ho Kim, and Edgard J. Powers, "Analysis of Initialization and Numerical Instabilities of Fast RLS Algorithms.", Proc. ICASSP 1991, pp. 1861-1863.
- [59] Bruce Wallace, Parallel Adaptive Noise Cancellation Filter Structures, M.Eng Thesis, Carleton Univ., Jan. 1991

- [60] D.T.M. Slock and T. Kailath, "Numerically Stable Fast RLS Transversal Filters for Recursive Least-Squares Adaptive Filtering", I.E.E.E. Transactions on Acoustics, Speech, and Signal Processing, vol. ASSP-39, Jan. 1991, pp. 92-113.
- [61] Fumio Amano, and Hector Perez, "A New Subband Echo Canceller Structure", Proc. ICASSP 1991, Toronto, May 1991, pp. 3585-3588.
- [62] George V. Moustakides, and Sergios Theodoridis, "Fast Newton Transversal Filters - A New Class of Adaptive Estimation Algorithms", I.E.E.E. Transactions on Signal Processing, Vol. 39, No. 10, Oct 1991, pp. 2184-2193.
- [63] Yoichi Haneda, Shoji Makino, and Yutaka Kaneda, "Modeling of a Room Transfer Function Using Common Acoustical Poles", Proc. ICASSP 1992, pp. II-213 to II-216.
- [64] Thierry Petillon, Andre Gilloire, and Sergios Theodoridis, "Complexity Reduction in Fast RLS Transversal Adaptive Filters", Proc. ICASSP 1992, pp. IV-37 to IV-40.
- [65] Eberhard Hansler, "The Hands-Free Telephone Problem", International Symposium on Circuits and Systems, 1992, Vol 4, pp. 1914-1917.
- [66] R. Wehrmann, "Concepts of Improving Hands-Free Speech Communication", I.E.E.E. International Symposium on Circuits and Systems, 1992, Vol 4, pp. 1918-1921.
- [67] Rudi Frenzel and Marcus E. Hennecke, "Using Prewhitening and Stepsize Control to Improve the Performance of the LMS Algorithm for Acoustic Echo Compensation", International Symposium on Circuits and Systems, 1992, Vol. 4, pp. 1930-1932.
- [68] H. Schutze, "Stabilized Fast Adaptation Algorithms for Acoustic Echo Control", International Symposium on Circuits and Systems, 1992, Vol. 4, pp. 1926-1929.
- [69] John J. Shynk, "Frequency-Domain and Multirate Adaptive Filtering", I.E.E.E. Signal Processing Magazine, Jan. 1992, pp 14-37.
- [70] S. Gazor, and B. Farhang-Boroujeny, "Quantization Effects in Transform-Domain Normalized LMS Algorithm", I.E.E.E. Transactions on Circuits and Systems - II, Vol. 39, No. 1, Jan. 1992, pp. 1-7.
- [71] James G. Ryan, Subband Adaptive Filters, M.Eng Thesis, Carleton Univ., March 1992
- [72] H. Yasukawa, "An acoustic echo canceller with sub-band noise cancelling", Electronics Letters, Vol. 28, No. 15, 16 July 1992, pp. 1403-4.
- [73] R.H. Kwong, and E.W. Johnson, "A Variable Step Size LMS Algorithm", I.E.E.E. Transactions on Signal Processing, Vol. 40, No. 7, pp. 1633-1642, July 1992.
- [74] Michael Knappe, Acoustic Echo Cancellation: Performance and Structures, M.Eng. Thesis, Carleton Univ., Sept 1992

- [75] A. Gilloire, M. Vetterli, "Adaptive filtering in subbands with critical sampling: analysis, experiments, and applications to acoustic echo cancellation", I.E.E.E. Transactions on Signal Processing, Vol. 40, No. 8, Aug 1992, pp. 1862-75.
- [76] S. Gudvangen, S.J. Flockton, "Comparison of pole-zero and all-zero modelling of acoustic transfer functions", Electronics Letters, Vol. 28, No. 21, 8 Oct 1992, pp. 1976-8.
- [77] Mitra and Kaiser, Handbook for Digital Signal Processing, Wiley, 1993
- [78] Emmanuel C. Ifeachor and Barrie W. Jervis, Digital Signal Processing: A Practical Guide, Addison Wesley, 1993.
- [79] Vaidyanathan, Multirate Systems and Filterbanks, Prentice Hall, 1993
- [80] Nicholas Kalouptsidis and Sergios Theodoridis, Adaptive System Identification and Signal Processing Algorithms, Prentice-Hall, 1993.
- [81] S. Gudvangen, and S.J. Flockton, "Intermediate Complexity Adaptive Filters in Acoustic Echo Cancellation", IEE Colloquium on 'New Directions in Adaptive Signal Processing'-Digest 39, Vol. 39, 1993, pp. 5/1 - 5/4.
- [82] M.Schoenle, N.Fliege, U. Zoelzer, "Parametric Approximation of Room Impulse Responses by Multirate Systems", Proc. ICASSP 1993, Vol. 1, 1993, pp. I-152 to I-156.
- [83] Luke S.L. Hsieh, and Sally L. Wood, "Performance Analysis of Time Domain Block LMS Algorithms", Proc. ICASSP 1993, Vol. 3, 1993, pp. 535-538.
- [84] Akihiko Sugiyama. "An Interference-Robust Stochastic Gradient Algorithm with a Gradient-Adaptive Step-Size", Proc. ICASSP 1993, Vol. 3, 1993, pp. 539-542.
- [85] Geoffrey A. Williamson, "Globally Convergent Adaptive Filters with Infinite Impulse Responses", Proc. ICASSP 1993, Vol. 3, 1993, pp. 543-546.
- [86] Tomas Oliveira e Silva, "Generalized Feed-Forward Filters: Some Theoretical Results", Proc. ICASSP 1993, Vol. 3, 1993, pp. 109-112.
- [87] Sergio Zimmermann, and Geoffrey A. Williamson, "Performance Properties of Fixed Pole Adaptive Filters", International Symposium on Circuits and Systems, 1993, Vol. 1, pp. 56-59.
- [88] Mariane R. Petraglia and Sanjit K. Mitra, "Performance Analysis of Adaptive Filter Structures Based on Subband Decomposition", International Symposium on Circuits and Systems, 1993, Vol. 1, pp. 60-63.
- [89] Yoon Gi. Yang, Nam Ik Cho, and Sang Uk Lee, "On the Performance Analysis of the Subband Adaptive Digital Filter", International Symposium on Circuits and Systems, 1993, Vol. 1, pp. 64-67.
- [90] D.M. Etter, N.C. Carlson, and V.A. Margo, "Adaptive System Modeling with Critically Decimated Multirate Filters", International Symposium on Circuits and Systems, 1993, Vol. 1, pp. 68-71.

- [91] Sen M. Kuo, Jier Chen, "Multiple-Microphone Acoustic Echo Cancellation System with the Partial Adaptive Process", *Digital Signal Processing*, Vol. 3, No. 1, Jan. 93, pp. 54-63.
- [92] Shoji Makino, Yutaka Kaneda, and Nobuo Koizumi, "Exponentially Weighted Step Size NLMS Adaptive Filter Based on the Statistics of a Room Impulse Response", *I.E.E.E. Trans on Speech and Audio Processing*, Vol. 1, No. 1, Jan. 1993, pp. 101-108.
- [93] Arie Feuer, Roberto Cristi, "On the Steady State Performance of Frequency Domain LMS Algorithms", *I.E.E.E. Transactions on Signal Processing*, Vol. 41, No. 1, Jan 1993, pp. 419-423.
- [94] Hiroshi Yasukawa, Shoji Shimada, "An Acoustic Echo Canceller Using Subband Sampling and Decorrelation Methods", *I.E.E.E. Trans. on Signal Processing*, Vol. 41, No. 2, Feb. 1993, pp. 926-930.
- [95] Jose C. Principe, Bert de Vries, and Pedro G. de Oliveira, "The Gamma Filter - A New Class of Adaptive IIR Filters with Restricted Feedback", *I.E.E.E. Transactions on Signal Processing*, Vol. 41, No. 2, Feb. 1993, pp. 649-656.
- [96] Georg Niedrist, "Echo Suppression for Loudspeaker-Microphone System Measurements", *J. Audio Eng. Soc.*, Vol. 41, No. 3, March 1993, pp. 143-153.
- [97] B. Farhang-Boroujeny, "Application of Orthonormal Transforms to Implementation of Quasi-LMS/Newton Algorithm", *I.E.E.E. Transactions on Signal Processing*, Vol. 41, No. 3, March 1993, pp. 1400-1405.
- [98] Mariane R. Petraglia, and Sanjit K. Mitra, "Adaptive FIR Filter Structure Based on the Generalized Subband Decomposition of FIR Filters", *I.E.E.E. Transactions on Circuits and Systems- II*, Vol. 40, No 6., June 1993, pp. 354-362.
- [99] Sanjit K. Mitra, and Abhijit Mahalanobis, "A Generalized Structural Subband Decomposition of FIR Filters and Its Application in Efficient FIR Filter Design and Implementation", *I.E.E.E. Transactions on Circuits and Systems - II*, Vol 40., No. 6, June 1993, pp. 363-374.
- [100] V. John Mathews, and Zhenhua Xie, "A Stochastic Gradient Adaptive Filter with Gradient Adaptive Step Size", *I.E.E.E. Transactions on Signal Processing*, Vol. 41, No. 6, June 1993, pp. 2075-2087.
- [101] A. Mahalanobis, et.al., "Adaptive FIR Filters Based on Structural Subband Decomposition for System Identification Problems", *I.E.E.E. Transactions on Circuits and Systems- II*, Vol. 40, No. 6, June 1993, pp. 375-381.
- [102] Joseph B. Evans, Ping Xue, and Bede Liu, "Analysis and Implementation of Variable Step Size Adaptive Algorithms", *I.E.E.E. Trans. on Signal Processing*, Vol. 1, No. 8, Aug. 1993, pp. 2517-2533.
- [103] Dirk T.M. Slock, "On the Convergence Behaviour of the LMS and Normalized LMS Algorithms", *I.E.E.E. Trans. on Signal Processing*, Vol. 41, No. 9, Sept 1993, pp. 2811-2825.

- [104] Michael E. Knappe, Rafik Goubran, "Steady-State Performance Limitations of Full-Band Acoustic Echo Cancellers", Proc. ICASSP 1994, Adelaide, South Australia, Vol. 2, pp. 73-76.
- [105] Haiyan Yuan, Dynamic Behaviour of Acoustic Echo Cancellation, M.Eng Thesis, Carleton University, 1994.
- [106] Thierry Petillon, Andre Gilloire, Sirgios Theodoridis, "The Fast Newton Transversal Filter: An Efficient Scheme for Acoustic Echo Cancellation in Mobile Radio", I.E.E.E. Transactions on Signal Processing, Vol. 42, No. 3, March 1994, pp. 509-517.
- [107] Yoichi Haneda, Shoji Makino, and Yutaka Kaneda, "Common Acoustical Pole and Zero Modeling of Room Transfer Functions", I.E.E.E. Trans. on Speech and Audio Processing, Vol. 2, No. 2, April 1994, pp. 320-328.
- [108] J.A. Bucklew and W.A. Sethares, "The Covering Problem and μ -Dependent Adaptive Algorithms", I.E.E.E. Transactions on Signal Processing, Vol. 42, No. 10, Oct. 1994, pp. 2616-2627.
- [109] C. Herley, and M. Vetterli, "Orthogonal Time-Varying Filter Banks and Wavelet Packets", I.E.E.E. Transactions on Signal Processing, Vol. 42, No. 10, Oct. 1994, pp. 2650-2663.
- [110] A.N. Birkett, Adaptive Linear and Nonlinear Structures for Acoustic and Audio Environments, Ph.D. Thesis Proposal, Carleton University, Jan. 1995.
- [111] A.N. Birkett, Rafik Goubran, "Acoustic Echo Cancellation using NLMS-Neural Network Structures", Proc. ICASSP 1995, Detroit Michigan, Vol. 5, pp. 3035-3038.
- [112] K. Mayyas, and T. Aboulnasr, "A Robust Variable Step Size LMS-type Algorithm: Analysis and Simulations", Proc. ICASSP 1995, Detroit Michigan, Vol. 2, pp. 1408-1411.
- [113] V.L. Stonick, "Time-varying Performance Surfaces for Adaptive IIR Filters: Geometric Properties and Implications for Filter Stability", I.E.E.E. Transactions on Signal Processing, Vol. 43, No. 1, Jan. 1995, pp. 29-42.
- [114] K.C. Ho, and Y.T. Chan, "Bias Removal in Equation-Error Adaptive IIR Filters", I.E.E.E. Transactions on Signal Processing, Vol. 43, No. 1, Jan. 1995, pp. 51-62.
- [115] C.W. Therrien and C.H. Velasco, "An Iterative Prony Method for ARMA Signal Modeling", I.E.E.E. Transactions on Signal Processing, Vol. 43, No. 1, Jan. 1995, pp. 358-361.
- [116] Phillip L. De Leon II, and Delores M. Etter, "Experimental Results with Increased Bandwidth Analysis Filters in Oversampled Subband Acoustic Echo Cancelers", I.E.E.E. Signal Processing Letters, Vol. 2, No. 1, Jan. 1995, pp. 1-3.
- [117] F. Beaufays, "Transform-Domain Adaptive Filters: An Analytical Approach", I.E.E.E. Transactions on Signal Processing, Vol. 43, No. 2, Feb. 1995, pp. 422-431.

- [118] M. Montazeri and P. Duhamel, "A Set of Algorithms Linking NLMS and Block RLS Algorithms", I.E.E.E. Transactions on Signal Processing, Vol. 43, No. 2, Feb. 1995, pp. 444-453.
- [119] C. Richard Johnson, "Yet Still More on the Interaction of Adaptive Filtering, Identification, and Control", I.E.E.E. Signal Processing Magazine, Vol. 12, No. 2, March 1995, pp 22-37.
- [120] P.S.R.Diniz, M.L.R.de Campos, and A. Antoniou, "Analysis of LMS-Newton Adaptive Filtering Algorithms with Variable Convergence Factor", I.E.E.E. Transactions on Signal Processing, Vol. 43, No. 3, March 1995, pp. 617-627.
- [121] D.R. Morgan, "Slow Asymptotic Convergence of LMS based AEC", I.E.E.E. Transactions on Speech and Audio Processing, Vol. 3, No. 2, March 1995, pp. 126-136.

APPENDIX A

Minto 3033 Conference Room

Characteristics:

- rectangular : 12'W x 18'L x 10'H
- reflective walls and floor.
- two tables, 6 padded chairs arranged around one table.
- ventilation fan on

APPENDIX B

Minto 2014 Conference Room

Characteristics:

- rectangular conference room: 19'W x 38' L x 10'H
- furnished. padded chairs, paintings, one small sound absorbing tile.
- floor is carpetted.
- ventilation fan on.

END

0 5-0 6-9/6

FIN