INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

Orthogonal Adaptive Digital Filters
with Applications to Acoustic System Identification

prepared by

Trevor P.B.M. Moat, Dipl. Tech. (Mech., Elec.), B.Eng. (Elec.)

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

Master of Engineering

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

Carleton University
1125 Colonel By Drive
Ottawa, Ontario, Canada, K1S-5B6

© Copyright 1998
Trevor P.B.M. Moat

National Library
of Canada

Acquisitions and
Bibliographic Services

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale
du Canada

Acquisitions et
services bibliographiques

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

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

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

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

0-612-27022-X

Canada

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

**Orthogonal Adaptive Digital Filters
with Applications to Acoustic System Identification**

submitted by

Trevor P.B.M. Moat, Dipl. Tech. (Mech., Elec.), B.Eng. (Elec.)

in partial fulfilment of the requirements of the degree of

Master of Engineering

Research Supervisor

Chair, Department of Systems and Computer Engineering

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

1998 January 01

Abstract

The *Transform-Domain LMS Algorithm* (Narayan, 1983) is studied in the context of an acoustic system identification problem. The power estimator in this two-stage digital filter is shown to affect the achievable rates and depths of convergence significantly. Preferred values for the two tracking parameters, β and μ , are determined.

Dynamic Step-size Initialization is proposed to improve early convergence by accelerating the rate at which true power measurements replace (arbitrary) initial values. Later, linear estimators are shown to be sub-optimal, particularly where the spectral distribution of the reference changes rapidly. A simple non-linear *Peak Window Power Estimator* which eliminates these problems is described. It will be shown to improve the tracking rates and misadjustment simultaneously.

The benefits of these methods are demonstrated using FIR sequences representative of typical acoustic environments and using recordings from a commercial telephone set. The proposed structures surpass the existing algorithms consistently under all circumstances tested.

Acknowledgements

A single pole cannot escape the real axis without the help of at least one other.

I am deeply indebted to my research advisor Prof. Rafik Goubran who stepped beyond the call of duty several times on my behalf. I also wish to thank my research examiners, Profs. Ioannis Lambadaris, M.S. El Tanany, and W. Martin Snelgrove from Carleton, Prof. Tyseer Aboulnasr from l'Université d'Ottawa, and Dr. André Van Schyndel from Bell-Northern Research. I am doubly grateful to Drs. Aboulnasr, Snelgrove, and Van Schyndel, whose instruction and advice were instrumental to the formation of this work. The insights I gained from our lengthy chats and after-class discussions will not soon depart my memory.

Dr. T. Aaron Gulliver (Canterbury), Dr. Françoise Beaufays (Stanford), and Dr. James Bondar (Carleton) provided me with many useful suggestions and criticisms. My DSP lab mates Jeff Lariviere, Jim Ryan, and Dr. Neil Birkett provided valuable guidance in the early stages. Danny Lemay, our seemingly tireless LAN administrator, kept our systems operational, even from a distance, even through the holidays. Thanks to all of you.

Finally, I am grateful to the government and people of Canada, whose on-going support of our universities, colleges, and other training and social programmes have provided opportunities for us to learn and to grow; opportunities sadly unavailable to so many others. Financial assistance in support of this research was provided by Bell-Northern Research via the Telecommunications Research Institute of Ontario.

Cecile, Prairie, and Eliza provided all the inspiration, warmth, nourishment, and time I needed. *Ik houd verder van jullie dan welk konijntje ook kan bedenken.... en ook heel de weg weer terug.*

Charles P. Moat taught me to stand on my own two feet. I wish he was here to see how they run.

Precis

For all those who weather my imagination -

A tiny pole, real, sole, was placed at origin.
Fourth of three, mother Mary, sun-raised on salted spring,
scaled hidden harbour's cliffs across, above, within
the tunnels down below the ground, their treasures home to bring.

Grand pole well-travelled, graded, graveled, departed circle stable.
Guidance given, my debt unriden - form fact from this fine fable!
My complex focus, constrained locus, romantic eight-track trance,
then August Eve: blessings received! Two poles co-inside dance!

North, south-east, snow-thunder screech, midnight flatdeck flapping
on verdant Prairie, name her, dare we? Brandon: still no napping.
Skip-stone cast cracks harbour's glass - more circles to explore...
Whence Queen's meets King's new pole begins - two are we no more.

By rock cliff lea, Acacia: three, then one our range enhanced.
Released from grief, felled maple leaf, this day Remembrance!
Now deep in debt, we of spring forget, forfeit emancipation
'mongst medium/message, amused-to-death age, the unconscious civilization.

New poles refrain in each domain: 'transform, invert, emerge,
evolve, convolve, at last dissolve, towards greater depths converge'.
From sea to sea, from scene to seen, my prayer: it be no mir'cle -
stability, infinity, all poles in unit circle.

1996 December 13, Rockcliffe Park, Ontario
1997 April 12-13, May 18, Richmond Hill, Ontario
1997 August 22, Maracaibo, Saltspring Island, B.C.

Table of Contents

Abstract	ii
Acknowledgements	iii
Precis	iv
Table of Contents	v
List of Figures	xi
List of Tables	xiv
List of Acronyms	xv
Mathematical Symbols and Notation	xvii
1 The Acoustic Echo Cancellation Problem	1
1.1 Rudiments of Conference Telephony	1
1.2 Objectives of Acoustic Echo Cancellers	4
1.3 Obstacles to Complete Echo Suppression	5
1.3.1 System Excitation	5
1.3.2 Telephone Set Hardware	6
1.3.3 Operating Environment.....	8
1.4 Thesis Description	10
1.4.1 Objectives and Motivation	10
1.4.2 Contributions.....	11
1.4.3 Outline.....	12

2	Time-Domain Adaptive Filters	14
2.1	Structure	14
2.2	The Method of Steepest Descent	19
2.3	Least Mean Squares Algorithm	23
2.3.1	Derivation	24
2.3.2	Properties for Stationary Inputs	25
2.3.3	Considerations for Non-Stationary Inputs	25
2.4	Variable Step-Size LMS Algorithms	26
2.4.1	Power-Normalized LMS Algorithms	27
2.4.1.1	Moving-Average Normalized LMS Algorithm	28
2.4.1.2	Auto-Regressive Normalized LMS Algorithm	29
2.4.2	Proximity-Based LMS Algorithms	31
2.4.2.1	Diagonal Matrix Variable Step-Size Algorithm	31
2.4.2.2	Error Magnitude Variable Step-Size Algorithm	33
2.4.2.3	Gradient Adaptive Step-Size LMS Algorithm	35
2.4.3	Application-Specific LMS Algorithms	36
2.4.3.1	Reduced-Order Algorithms	36
2.4.3.2	The Exponentially-Weighted Step-Size LMS Algorithm	37
2.5	Sub-Band Architectures	38
2.5.1	Foundations	39
2.5.2	Perfect Reconstruction Systems	39
2.5.3	Polyphase Implementations	40
2.5.4	Sub-Band Adaptive Filters	41
2.5.5	Delayless Versions	43
2.6	Summary	45
3	Transform Domain Adaptive Filters	46
3.1	Overview	47
3.2	Block-Oriented Transform Domain Adaptive Filters	48
3.2.1	Chronological Development	49
3.3	Multi-Delay Implementations with Partial Overlap	52
3.3.1	Chronological Development	54
3.4	Sample-Oriented Transform Domain Adaptive Filters	55
3.4.1	Development from Time-Domain LMS	56
3.4.2	Chronological Development of Fixed-Transform Structures	60
3.4.3	Two-Layered Interpretation	65
3.5	Summary	67

4	Application to Acoustic Echo Cancellers	68
4.1	Some General Properties of the AEC Environment	69
4.1.1	Excitation	69
4.1.2	Acoustic Impulse Responses	71
4.1.3	Algorithm Behaviour	72
4.2	Specification of Available Parameters	78
4.2.1	Number of Taps - N	78
4.2.2	Global Step-Size - μ_g	79
4.2.3	Channel Power Estimation Parameter - β	79
4.2.3.1	Stationary Inputs	79
4.2.3.2	Step Non-Stationary Inputs	82
4.2.3.3	Towards a Rule for β Selection	84
4.3	Some Alternate Structures	86
4.3.1	Non-Rectangular Reference Windows	86
4.3.2	Re-Combination in the Transform Domain	87
4.4	Proposed Linear Pre-Processor Modifications	88
4.4.1	Dynamic Stepsize Initialization	88
4.4.2	Expansion of β	90
4.4.3	Adaptation of β	91
4.5	Non-Linear Power Estimation	93
4.5.1	Median and Rank-Order Filters	93
4.5.2	The Proposed Non-Linear Estimator	95
4.5.3	Mathematical Development	96
4.5.4	Comparison of Linear and Non-Linear Methods	98
4.5.5	Practical Implementation	100
4.6	Summary	102
5	Experimental Methods and Apparatus	104
5.1	Input Signals	104
5.2	Performance Metrics	106
5.2.1	Echo Return Loss Enhancement	106
5.2.2	TIP/TP Ratio	106
5.3	Sample Plants	109
5.4	Acoustic Recording Procedures and Apparatus	110
5.4.1	File Generation and Transfer from PC to DAT	110
5.4.2	Hardware and Recording Procedures	110
5.4.3	File Transfer from ADAT to PC	112

6	Experiments with Simulated Systems	114
6.1	Some Preliminaries.....	115
6.2	Comparison with Published Results.....	116
6.3	Algorithm Optimization.....	118
6.3.1	Selection of μ_g	118
6.3.1.1	Procedure.....	118
6.3.1.2	Observations.....	119
6.3.1.3	Analysis.....	119
6.3.2	Selection of β	120
6.3.2.1	Procedure.....	120
6.3.2.2	Observations.....	121
6.3.2.3	Analysis.....	123
6.4	Dynamic Step-Size Initialization.....	124
6.4.1	Procedure.....	124
6.4.2	Observations.....	124
6.4.3	Analysis.....	125
6.5	Performance Limitations Due to Undermodelling.....	126
6.5.1	Procedure.....	126
6.5.2	Observations.....	127
6.5.3	Analysis.....	128
6.6	Signal Identification Stage.....	128
6.6.1	Step Changes in Variance.....	128
6.6.1.1	Procedure.....	129
6.6.1.2	Observations.....	129
6.6.1.3	Analysis.....	130
6.6.2	Step Change in Distribution.....	131
6.6.2.1	Procedure.....	131
6.6.2.2	Observations.....	132
6.6.2.3	Analysis.....	133
6.6.3	Speech Sequences.....	134
6.6.3.1	Procedure.....	134
6.6.3.2	Observations.....	135
6.6.3.3	Analysis.....	135
6.7	Plant Identification Stage.....	137
6.7.1	Procedure.....	137
6.7.2	Observations.....	138
6.7.3	Analysis.....	139
6.8	Summary.....	140

7	Experiments with Acoustic Systems	142
7.1	Algorithm Parameter Specification	142
7.1.1	Activation Thresholds	142
7.1.2	Adaptive Filter Order	143
7.2	Experimental Notes - Board Room MC2014	146
7.3	Signal Identification Stage - MC2014	146
7.3.1	Procedures	146
7.3.2	Stationary Excitation	147
7.3.3	Step Change in Variance	149
7.3.4	Step Change in Distribution	150
7.3.5	Speech Bursts	151
7.3.6	Analysis	153
7.4	Plant Identification Stage - MC2014	154
7.4.1	Procedure	154
7.4.2	Observations	154
7.4.3	Analysis	157
7.5	Experimental Notes - Small Office ME4144	158
7.6	Plant Identification Stage - ME4144	158
7.6.1	Procedure	158
7.6.2	Observations	158
7.6.3	Analysis	159
7.7	Numerical Summary	162
7.8	Summary	164
8	Conclusions	166
8.1	Review of Thesis Objectives	166
8.2	How the Objectives Were Met	167
8.3	Summary of Experimental Results	169
8.3.1	Dynamic Step-Size Initialization	169
8.3.2	Preferred Values of Linear Parameters	169
8.3.3	Misadjustment	169
8.3.4	Signal Identification Capability	170
8.3.5	Plant Identification Capability	170
8.4	Suggestions for Further Research	171
8.4.1	Modified Structures	171
8.4.2	Linear Systems	171
8.4.3	Non-Linear Systems	172
8.5	Summary of Contributions	173

A1	A Review of Adaptive Filter Theory	174
A1.1	Elements of Digital Filters.....	174
A1.2	Adaptation of Digital Filters.....	176
A1.2.1	Objectives.....	176
A1.3	Some Pertinent Results from Matrix Theory.....	177
A1.3.1	Hyperelliptical Surfaces.....	178
A1.3.2	Translation to Standard Position	179
A1.3.3	Rotation into Standard Position	179
A1.4	Mathematical Framework in the Time Domain	180
A1.4.1	Input Crosscorrelation Vector	181
A1.4.2	Input Autocorrelation Matrix	181
A1.4.3	Mean Squared Error	183
A1.4.4	Performance Surface	184
A1.4.5	Modal Convergence	185
A1.5	Mathematical Framework in the Transform Domain	187
A1.5.1	Input Cross-Correlation Vector	187
A1.5.2	Input Autocorrelation Matrix	188
A1.5.3	Mean-Squared Error.....	189
A1.5.4	Performance Surface	190
A1.5.5	Modal Convergence	191
A1.6	Visualization of the Geometry.....	192
A1.6.1	Autocorrelation Matrices	192
A1.6.2	Performance Surfaces and Convergence Rates.....	197
A1.7	Some Notes Concerning Computational Expense.....	201
A2	Signals and Systems Used in Simulations	202
A2.1	Excitation.....	202
A2.2	FIR Transfer Functions.....	204
A2.3	Acoustic Transfer Functions.....	205
A3	List of References	207
A4	Bibliography	219

List of Figures

FIGURE 1.1:	A highly simplified conference telephone set.	2
FIGURE 1.2:	A conference telephone set with an acoustic echo canceller.	4
FIGURE 2.1:	The generalized parallel adaptive digital filter in the time domain.	15
FIGURE 2.2:	Power estimation using a first-order AR process.	30
FIGURE 2.3:	A sub-band adaptive filter with four channels.	42
FIGURE 2.4:	A four-band realization of Morgan's delayless sub-band filter.	44
FIGURE 3.1:	Dentino's frequency-domain LMS filter for $N=4$	49
FIGURE 3.2:	A first attempt in the implementation of the SOX-AF.	56
FIGURE 3.3:	SOX-AF with on-line channel power estimation.	58
FIGURE 3.4:	Channel power estimation using a first-order AR process.	58
FIGURE 3.5:	Two-Layer representation of the SOX-AF.	66
FIGURE 4.1:	Comparison of DCT output signals and output power time series.	74
FIGURE 4.2:	PSD plots of the time- and power series of figure 4.1.	75
FIGURE 4.3:	Frequency content of the transformed outputs and power sequences.	77
FIGURE 4.4:	Power estimation characteristics in one bin for sinusoidal excitation.	81
FIGURE 4.5:	Broadband tracking capability of AR-SOX-AF as a function of β	83
FIGURE 4.6:	Flowchart representing Dynamic Step-Size Initialization.	90
FIGURE 4.7:	Schematic representation of the proposed Peak Window Power estimator.	96
FIGURE 4.8:	Relationship between the actual and PWP-estimated sine power pulse.	97
FIGURE 4.9:	Comparison between three channel power estimators.	99
FIGURE 5.1:	An FIR impulse response showing TIP and TP measurements.	107
FIGURE 5.2:	Methods to transfer files between the remote apparatus and the PC.	111
FIGURE 5.3:	Apparatus used to make remote recordings	112
FIGURE 6.1:	Duplication of results of [Marshall89] for four reference inputs.	117
FIGURE 6.2:	Convergence rates as a function of μ_g in the MA-NLMS algorithm.	119
FIGURE 6.3:	Convergence rates of the AR-SOX-AF with broadband excitation for various values of β	121
FIGURE 6.4:	Convergence rates of the AR-SOX-AF with AR2(-1.5955, .95) excitation for various values of β	122
FIGURE 6.5:	Convergence rates of the AR-SOX-AF with AR12 excitation for various values of β	122

FIGURE 6.6:	Reduction in overshoot for broadband excitation using DSI.	125
FIGURE 6.7:	Cancellation limits for the three LMS algorithms with a measured FIR acoustic plant.	127
FIGURE 6.8:	Reference input tracking behaviour for the three algorithms when a step increase in variance is applied to a broadband reference input.	129
FIGURE 6.9:	Reference input tracking behaviour for three different algorithms when the excitation is a concatenation of AR2 processes.	132
FIGURE 6.10:	Signal Identification of SSGF2 track.	136
FIGURE 6.11:	Signal Identification of SSGM1 track.	136
FIGURE 6.12:	FIR Plant tracking capability with broadband inputs.	138
FIGURE 6.13:	FIR Plant tracking capability with AR12 excitation.	139
FIGURE 7.1:	Steady-state broadband ERLE for the time-domain MA-NLMS filter in MC2014.	144
FIGURE 7.2:	Steady-state broadband ERLE for the AR-SOX-AF in MC2014.	144
FIGURE 7.3:	Steady-state broadband ERLE for the PWP-SOX-AF in MC2014.	145
FIGURE 7.4:	AR2(-1.5955,.95) excitation applied to the three LMS filters.	148
FIGURE 7.5:	AR12 excitation applied to the three LMS filters.	148
FIGURE 7.6:	Tracking capability of three algorithms with broadband excitation whose variance is changed sharply in MC2014.	149
FIGURE 7.7:	Tracking capability of the three algorithms with the AR2 sequence whose PSD is changed sharply in MC2014.	150
FIGURE 7.8:	Comparison of the three algorithms when tracking the SSGF2 speech input file in MC2014.	152
FIGURE 7.9:	Comparison of the three algorithms when tracking the SSGM1 speech input file in MC2014.	152
FIGURE 7.10:	Comparison of plant tracking capabilities of the three algorithms with WSS broadband excitation in MC2014.	155
FIGURE 7.11:	Comparison of plant tracking capabilities of the three algorithms with WSS AR2(-1.5955,.95) excitation in MC2014.	155
FIGURE 7.12:	Comparison of plant tracking capabilities of the three algorithms using a WSS AR12 excitation signal in MC2014.	156
FIGURE 7.13:	Comparison of plant tracking capabilities of the three algorithms with SSGF2 speech excitation in MC2014.	156
FIGURE 7.14:	Comparison of plant tracking capabilities of the three algorithms with SSGM1 speech excitation in MC2014.	157
FIGURE 7.15:	Comparison of plant tracking capabilities of the three algorithms with broadband excitation in ME4144.	160
FIGURE 7.16:	Comparison of plant tracking capabilities of the three algorithms with AR2(-1.5955,.95) excitation in ME4144.	160

FIGURE 7.17: Comparison of plant tracking capabilities of the three algorithms with the SSGF2 speech input in ME4144.	161
FIGURE 7.18: Comparison of plant tracking capabilities of the three algorithms with the SSGM1 speech input in ME4144.	161
FIGURE A1.1: Visualization of autocorrelation matrices and their eigenvalues for the AR1(0.9) process.	193
FIGURE A1.2: Visualization of autocorrelation matrices and their eigenvalues for the AR2(-1.9114,.95) process.	194
FIGURE A1.3: Visualization of autocorrelation matrices and their eigenvalues for the AR2(-1.5955,.95) process.	195
FIGURE A1.4: Visualization of autocorrelation matrices and their eigenvalues for the AR12 process.	196
FIGURE A1.5: Comparative visualization of fast-mode convergence rates in time- and transform-domain algorithms.	199
FIGURE A1.6: Comparative visualization of slow-mode convergence rates in time- and transform-domain algorithms.	199
FIGURE A1.7: Comparative visualization of bi-modal convergence rates in time- and transform-domain algorithms.	200
FIGURE A2.1: Spectral characteristics of the four AR2 processes.	202
FIGURE A2.2: Spectral characteristics of the AR12 process.	203
FIGURE A2.3: Power spectral density plots of the synthetic speech sequences.	203
FIGURE A2.4: Time series and PSD plots of the FIR impulse response IRN(1200).	204
FIGURE A2.5: Time-delay measurement of the reflective feature distinguishing four FIR impulse responses recorded in MC2014.....	205
FIGURE A2.6: Layout of conference room MC2014 at Carleton University.	206
FIGURE A2.7: Layout of graduate office ME4144 at Carleton University.	206

List of Tables

TABLE 3.1	Algorithm for the standard SOX-AF described in the literature.	59
TABLE 3.2	Condition numbers for the SOX-AF with Markov-1 inputs.	63
TABLE 4.1	Tracing the influence of β on MSE performance in the AR-SOX-AF.	80
TABLE 5.1	Description of stochastic excitation sequences.	105
TABLE 5.2	Twelfth-order AR process parameters.	105
TABLE 5.3	Description of measured acoustic transfer functions for chapter 6.	109
TABLE 5.4	Details of hardware interfaces of file transfer apparatus.	112
TABLE 6.1	Concatenated sequence of AR2 processes.	132
TABLE 6.2	Sequence of plants to be identified.	138
TABLE 7.1	Measured steady-state ERLE for the MA-NLMS algorithm.	162
TABLE 7.2	Measured room-tracking ERLE for the MA-NLMS algorithm.	162
TABLE 7.3	Measured steady-state ERLE for the AR-SOX-AF.	163
TABLE 7.4	Measured room tracking ERLE for the AR-SOX-AF.	163
TABLE 7.5	Measured steady-state ERLE for the PWP-SOX-AF.	163
TABLE 7.6	Measured room-tracking ERLE for the PWP-SOX-AF.	163
TABLE 7.7	Summary of “averaged” results for the stationary plant.	164
TABLE 7.8	Summary of “averaged” results for the time-varying plant.	164
TABLE A1.1	Some characteristics of static FIR and IIR filters.	175
TABLE A1.2	Some characteristics of adaptive FIR and IIR filters.	175
TABLE A1.3	Peak eigenvalue ratios for various inputs with $N=64$	192
TABLE A4.1	Catalogue Numbers for cited texts.	219
TABLE A4.2	Titles and indices for serial publications.	220

List of Acronyms

ADC	Analog-to-Digital Converter
AEC	Acoustic Echo Cancellation, Acoustic Echo Canceller
AR	Auto-Regressive
ARMA	Auto-Regressive Moving-Average
BOX-AF	Block-Oriented Transform Domain Adaptive Filter
DAC	Digital-to-Analog Converter
DAT	Digital Audio Tape
DCT	Discrete Cosine Transform
DFT	Discrete Fourier Transform
DHT	Discrete Hartley Transform
DSP	Digital Signal Processing, Digital Signal Processor
DST	Discrete Sine Transform
ERLE	Echo Return-Loss Enhancement
FBAF	Frequency-Bin Adaptive Filter
FFT	Fast Fourier Transform
FIR	Finite Impulse Response
FSF	Frequency Sampling Filter
GMDF α	Generalized Multi-Delay Filter with parameter α

IIR	Infinite Impulse Response
KLT	Karhunen-Loeve Transform
LMS	Least-Mean Squares
LS	Least-Squares
MA	Moving-Average
MA-NLMS	Moving-Average Normalized Least-Mean Squares
MSE	Mean-Squared Error
NLMS	Normalized Least-Mean Squares
NPR	Near-Perfect Reconstruction
OLA	Overlap Add
OLS	Overlap Save
PR	Perfect Reconstruction
PSD	Power Spectral Density
PWP	Peak Window Power
QMF	Quadrature Mirror Filter
RLS	Recursive Least Squares
RMS	Root-Mean-Square
SNR	Signal-to-Noise Ratio
SPL	Sound Pressure Level
SOX-AF	Sample-Oriented (or Self-Orthogonalizing) Transform Domain Adaptive Filter
TIP/TP	Total Impulse Power to Tail Power Ratio
VSS	Variable Step-Size
XDAF	Transform Domain Adaptive Filter
WOLA	Weighted Overlap Add

Mathematical Symbols and Notation

General Matrix Conventions

Scalar variables are represented with unweighted symbols, vectors by lower-case bold symbols, and matrices by upper-case bold symbols. Vectors most often represent a set of observations at particular points in time. To avoid confusion, the time index will always be subscripted, whereas the index into the vector will follow in brackets. Either qualifier will be omitted if its use in the present context is redundant. By way of example:

$\mathbf{x}_k(n)$	refers to the n th element of vector $\mathbf{x}$ at time k ;
$\mathbf{u}_k$	refers to all elements of the vector $\mathbf{u}$ at time k ;
$\mathbf{u}(n)$	refers to the sequence of the n th elements of $\mathbf{u}$ over all k .

Operator Notation

$DCT[\mathbf{h}]$	Discrete Cosine Transform
$\det[\mathbf{H}]$	Determinant of a square matrix
$DFT[\mathbf{h}]$	Discrete Fourier Transform
$\text{diag}[\mathbf{H}]$	Elements extracted from the main diagonal of a square matrix
$E[\mathbf{x}]$	Expected value of a random variable
$FFT[\mathbf{h}]$	Fast Fourier Transform
$\max[\mathbf{H}]$	Vector of maxima in each column of a matrix
$T[\mathbf{h}]$	Non-specific discrete transform
$\text{tr}[\mathbf{H}]$	Trace of a square matrix

Arabic Symbols

A	General square matrix
B	General square matrix
b	General scalar variable
c	General scalar variable
d_k	Measured primary input at iteration k
D	Number of iterations over which DSI is active
e_k	Scalar error value at iteration k
f_k	Sub-window vector in PWP estimator
g_k	PWP vector
G_k	PWP matrix
h	General column vector
I	Identity matrix
i	General scalar index
j	Positive unit vector in the complex plane perpendicular to the real axis
k	Discrete time index
$L1$	Time window of each PWP matrix row entry
$L2$	Number of rows in the PWP matrix
M	Step-size matrix in the Diagonal Matrix Variable Step-Size Algorithm
M	Misadjustment
m	Sample rate conversion factor in sub-band algorithms
m_0	Step-size decrease period in Diagonal Matrix Variable Step-Size Algorithm
m_1	Step-size increase period in Diagonal Matrix Variable Step-Size Algorithm

n	Index into a vector of order N
N	Order of an adaptive filter, observation vector, or plant
p_k	Power estimate vector at iteration k
p	Data cross-correlation vector in the time domain
$\bar{p}$	Data cross-correlation vector in the transform domain
p_π	Total power estimate in one half-cycle of a sinusoidal signal
R_x	Autocorrelation function
R_{xx}	Input autocorrelation matrix in the time domain
R_{uu}	Input autocorrelation matrix in the transform domain
S	Eigenvector matrix
S_x	Power spectral density function
u_k	Input observation vector in the transform domain
u_k^2	Instantaneous power in the transformed observation vector
w_k	Adaptive filter coefficient vector in the time domain
$\mathbf{w}_k$	Adaptive filter coefficient vector in the transform domain
w_{opt}	Optimal adaptive filter solution in the time domain
$\mathbf{w}_{opt}$	Optimal adaptive filter solution in the transform domain
x_k	Reference input sample at iteration k
$\mathbf{x}_k$	Input observation vector in the time domain
y_k	Adaptive filter output sample at iteration k

Greek Symbols

α	Shift parameter in GMDF; Error Magnitude Step-size Algorithm parameter
β	Persistence factor in the first-order AR power estimator
γ	Convergence parameter in Error Magnitude Step-size Algorithm
Λ	Diagonal matrix containing the eigenvalues of R
λ_n	The n th eigenvalue of R
μ	General step-size in LMS-based algorithms
μ_g	Global step-size used in normalized LMS algorithms
μ_k	Overall step-size at iteration k
μ_p	Step-size derived from measured quantities at iteration k
μ'_k	Adjusted step-size in Error Magnitude Step-size Algorithm
τ	Delay variable in correlation and convolution integral equations
ξ	MSE performance surface
∇_k	Gradient of a quadratic MSE surface at iteration k

1

The Acoustic Echo Cancellation Problem

This research was motivated by the need to attenuate acoustic reverberation in typical multi-user hands-free communications systems. Significant advancements in *Acoustic Echo Cancellation* (AEC) have been made in recent years with the advent of end-to-end digital networks. Signal processing methods that were impossible even in state-of-the-art analog transmission systems are now commonplace with modest digital hardware. Even so, an optimal solution to the AEC problem remains elusive: the complicated natures of the unknown acoustic systems (including telephone hardware and the rooms in which they are installed) and the signals that excite them (the human voice, typically) present numerous mechanical, electronic, and acoustic problems.

The amount of research on-going in universities and corporate research centres world-wide is indicative of the value placed on a cost-effective means of echo attenuation. Indeed, the recent proliferation of wideband services to facilitate voice and video conferences, cellular and wireless systems, and the advent of ever-smaller Personal Communications Systems (PCS), all of which suffer to varying degrees from the undesirable coupling of acoustic signals from output to input transducers, suggest that research in AEC will be active for some time to come. The purpose of this opening chapter is to introduce the elements of the AEC problem, and to summarize the objectives, contributions, and outline of the thesis.

1.1 Rudiments of Conference Telephony

A *conference telephone set* allows one or more people in a room to converse hands-free (i.e. with no hand-held earpieces or sensors) with participants elsewhere as if all were in the same room. Essentially, it is a standard telephone equipped with a *power amplifier*, a *loudspeaker*, and a *microphone*. The first two allow signals received from the network to be heard in a given space,

while the third converts acoustic energy detected in that space into an electrical signal, which can be coded and transmitted back through the network. In practice, numerous other components and design changes are required to achieve acceptable fidelity. Modern conference telephone sets also incorporate numerous other features for user convenience which would not be required in regular hand-held telephones.

A minimal full-duplex conference telephone set is represented in figure 1.1. The *reference signal* is coupled from the four-wire network loop to the two-wire local loop by way of a *hybrid circuit*. After digital-to-analog conversion, the reference signal is amplified and fed to the loudspeaker so it can be heard by participants nearby. At the same time, acoustic energy is transduced by the microphone into an electrical signal, which is called the *primary signal* after low-pass filtering and analog-to-digital conversion. In this elementary case, the primary signal is coupled into the telephone network through the hybrid circuit.

FIGURE 1.1: A highly simplified conference telephone set. The terminology of the electronic signals is also shown.

The hybrid is not lossless in practice, so some energy from the primary signal will leak into the reference path, possibly causing some acoustic feedback or *howling*. An *adaptive network echo canceller* is installed to model the transfer function of the leakage path. Its structure is not unlike the adaptive filters which will be discussed in detail shortly, so it will not be described further here. Suffice it to say that after a brief convergence period, it will allow the leaked energy to be predicted and subtracted from the reference signal very effectively.

In the acoustic domain, the loudspeaker output is referred to as the *far-end signal*, while speech originating within the room is called the *near-end signal*. In most practical cases, an additional *noise signal* generated by other sources, such as fans, mechanical systems and general office activity, along with circuit noise arising from crosstalk and thermal properties of the electronic components are also present. It will be assumed hereafter that these sound sources occupy the same bandwidth in the same acoustic channel (i.e. the room) simultaneously.

Acoustic energy cannot be constrained to propagate along perfectly directional channels in free space, so a certain amount of the far-end signal will be radiated directly from the loudspeaker to the microphone. Further, the boundaries of the room (the ceiling, walls, and doors, for example) will reflect far-end signals, adding delay, distortion, attenuation, and scatter with each reflection. Left as-is, the aggregate acoustic energy in the room is integrated by the microphone and appears in the primary. Thus, the system of figure 1.1 will not work well in practice because it will transmit the desired near-end signal and the unwanted artifacts of the far-end and noise signals with equal amplification. This energy is synthesized as a smear of echos and noise whose varying delay and amplitude can be highly objectionable to participants across the network.

Some mechanism is required to remove artifacts of the far-end signal from the primary. This mechanism will be referred to hereafter as an *Acoustic Echo Canceller*, the design and operation of which is the subject of this thesis. Complete acoustic echo cancellation is achieved only when all energy due to the far-end signal is squelched from the primary while the near-end portion of the primary - speech that originated from the participants in that room - is passed unaltered. A refined conference telephone set with a parallel acoustic echo canceller is shown in figure 1.2.

FIGURE 1.2: A conference telephone set with an acoustic echo canceller. Several possible paths of acoustic coupling are shown.

1.2 Objectives of Acoustic Echo Cancellers

Acoustic Echo Cancellation is a *System Identification problem*. The acoustic echo cancellers of concern here attempt to model the acoustic transmission parameters of the telephone set and the conference room in which it operates using an *adaptive digital filter*. The filter is necessarily adaptive for two main reasons:

- the nature of the room(s) in which the set will be installed cannot be known in advance;
- the acoustic transfer function of the operating environment is subject to change continuously, such that a static digital filter will become sub-optimal.

The objectives of an acoustic echo canceller are to identify the unknown transfer function as quickly and as accurately as possible, and to adjust itself continually so as to track changes in the plant. It achieves these objectives by:

- generating a cancellation signal from the adaptive filter and the reference input;
- subtracting the cancellation signal from the primary input to generate an error signal;
- adjusting the adaptive filter so as to minimize the correlation between the reference and error signals.

These processes are not trivial, as the responses of the set and room can be very complicated and time-varying. Deep cancellation is often exacerbated by system non-linearities and interfering noise sources, such as vibration, near-end speech, and other audible disturbances. Further, the far-end signal is normally human speech, whose *Power Spectral Density* (PSD) and dynamic range may vary considerably, causing the filter to behave unevenly across the different spectral bands and at different times. More advanced algorithms monitor the time-frequency properties of the reference signal to accelerate tracking performance in under-excited modes.

1.3 Obstacles to Complete Echo Suppression

A number of conspirators make the AEC problem a particularly challenging one. The greater culprits are mechanical set dynamics, vibration, and the acoustic properties of the operating environment, none of which can be addressed effectively without adverse influence to cost. The major sources of difficulty are identified in this section, and their effects on performance are assessed.

In some applications of echo cancellers (those used for data streams in satellite communications, for example) it is possible to employ *training sequences* at regular intervals to assist the filter in tracking the evolving transmission properties of the channel. Such techniques are inappropriate in conference telephony, as they interrupt the free flow of conversation. In AEC, the filter design objectives must be met with no advance knowledge of the excitation signal.

1.3.1 System Excitation

In a hands-free telephone environment, the excitation of interest is far-end speech. The nature of human speech has been studied extensively, so a complete review will not be included here.

Interested readers are referred to any of several introductory tutorials and manuscripts on the subject [Crochiere75] [Schafer75] [Makhoul75] [Zelinski77] [Rabiner78]. Two properties are of seminal importance to adaptive digital filters:

- *Stationarity*: A speech signal acquired over a long time interval will tend to have no DC power, and its long-term PSD envelope will tend to decay (though not monotonically) with frequency. These same properties persist in subsequent intervals, suggesting speech signals are stationary when viewed over very long periods of time. However, effective AEC demands substantial filter convergence within a few seconds, and over these shorter intervals, speech must be regarded as being highly non-stationary [Schafer75] [Rabiner78]. However, when viewed through the lens of short-time analysis, a certain degree of stationarity returns to most speech signals [Schafer75], opening a window of opportunity for modelling these inputs over much shorter intervals (say, 10-30 msec). Accordingly, the filter parameters must enable the filter to tolerate rapid changes in the input signal statistics to achieve acceptable performance.
- *Spectral content*: Speech is known to be highly coloured, which is to say its power spectrum varies greatly from frequency to frequency. This is a particularly undesirable characteristic for signals incident to adaptive filters, because the adaptive algorithm has reduced means of achieving proper convergence in frequency bands which are relatively under-excited. This will be demonstrated for conventional filter structures in chapter 6. Adaptive filters receiving such signals may have to re-converge continually as changes in the far-end signal are detected.

1.3.2 Telephone Set Hardware

Modern conference telephone sets are designed to provide numerous features in addition to good acoustic performance. These features are most often achieved only with added material cost. As objects of mass production, cost is a very influential factor in mechanical and electronic design of these systems.

Many present-day conference sets employ *switch-loss insertion*, a self-tuning interlocked analog attenuation mechanism in the reference and primary signal paths, to mimic echo cancellation. Its intent is to maximize the gain in the reference path while maintaining the electro-acoustic loop

gain of the set below unity. When the reference signal is active, the primary signal is attenuated, and vice-versa. Switch-loss insertion is an inexpensive but awkward substitute for a discrete-time echo cancellation system.

It will be argued shortly that good acoustic design is a prerequisite for effective acoustic echo cancellation. Such a design would account for the following hurdles, listed here roughly in the order they are encountered by the reference signal:

- *DAC and analog reconstruction filter:* The reference signal must be converted to analog form prior to its presentation to the loudspeaker. The finite bandwidths of the reconstruction filter and the amplifier prevent ideal conversion, though these effects are normally negligible.
 - *Loudspeaker:* Loudspeakers introduce harmonic distortion and exhibit non-linear responses, regardless of how well they are constructed. Those installed in telephone sets are particularly suspect, since they must be inexpensive and small (typically less than eight cm diameter and rated between 0.5 and 3 Watts RMS [VanSchyn96]). Much research has been conducted into loudspeaker design, compensation [Gao91a] and modeling [Birkett97], and the topic remains open for further investigation. Loudspeaker effects will impair performance in real systems, however it is beyond the scope of this work to investigate them in detail.
 - *Mechanical coupling:* The loudspeaker is an electro-acoustic transducer which radiates energy through axial movement of its cone. This movement induces vibration in the loudspeaker frame, the mounting surface of the set, and ultimately in the microphone. Mechanical coupling is nearly instantaneous, highly non-linear, and very difficult to model. Noise due to mechanical coupling can be expected to worsen rapidly as the acoustic volume increases.
 - *Mechanical vibration:* Loose components in the telephone set may compound the mechanical coupling problem by adding further vibration and acoustic noise. The most likely sources include an on-hook handset, buttons (particularly those which are user-removable), warpage in the plastics or frame leading to stress concentration and uneven mounting on the table, and loose screws, components, or wires. The resulting noise is impulsive, and is aggravated by loud volume or excitation near resonant modes. Because of its broad spectral content and bursty nature, acoustic and mechanical vibrational noise is very difficult to remove.
-

-
- *Microphone properties:* The microphone is an acoustico-electric transducer which suffers from finite bandwidth, non-uniform spectral response, and a some degree of non-linearity. These effects are negligible when compared to the loudspeaker because the components are smaller and less susceptible to mechanical losses due to inertia, friction, and cone distortion.
 - *ADC and anti-alias filter:* The microphone signal must be passed through an analog low-pass filter before sampling to suppress aliases of in-band signals. Filtering, sampling, and coding processes will all add noise and delay to the primary signal, however these will be of minor consequence relative to the other offenders.
 - *Digital filter limitations:* Most acoustic transfer functions exhibit a complicated arrangement of many poles and zeros [Haneda94] [Gudvangen95]. Further, loudspeaker transfer functions contain (at least) one dominant pole, and they tend to impart significant non-linearities on the signals they transduce [Gao91a], particularly at higher amplitudes [Birkett97]. FIR filters cannot model systems with one or more poles perfectly, nor can they model non-linearities. Finally, there are numerous practical constraints on FIR filter length to consider, and these may result in the echo path being *undermodelled* [Knappe92], placing a theoretical limit on the depth of cancellation which can be achieved.
 - *Hybrid coupling:* Leakage of the primary signal into the reference path is certain to occur in the hybrid network, as indicated in figure 1.1. Fortunately, since the hybrid is a (nearly) linear electrical device, this leakage can be modelled and subtracted very effectively with the hybrid echo canceller.

1.3.3 Operating Environment

The environment in which the AEC system operates compromises performance expectations even further. Again in the order they are encountered by the far-end signal, these are:

- *Direct-path acoustic coupling:* A loudspeaker will radiate acoustic energy to the microphone through any direct acoustic path. There are an infinite number of such paths both internal and immediately external to the set. Directly-coupled responses appear in the lower-order portion of the acoustic impulse response, they are of relatively short duration, and are quantified most
-

readily through measurements in an anechoic chamber. These signals are often the largest contributors of echo signal power, so it is important for the filter to model them well. This is not a trivial task, given the loudspeaker effects noted above.

- *Indirect acoustic coupling:* In addition to the direct acoustic paths, an infinite number of indirect internal and external acoustic paths exist, incorporating walls, tables, people, and other reflective surfaces in the room and set. *Lower-order echoes* are those which arrive at the microphone quickly, having reflected off of only a few surfaces. They tend to be reasonably well-correlated with the far-end signal. *Higher-order echoes* are those which have been reflected by many surfaces. They tend to be less correlated with the reference signal due to acoustic scattering, absorption, and dispersion at reflective surfaces. These higher-order *tails* are very difficult to model, and they produce what is known as the *barrel effect*, similar to the echo one would expect to hear when speaking down a deep well shaft.
 - *Non-stationarities in acoustic paths:* Any change in the acoustic environment will affect the overall system transfer function. Perturbations as small as the movement of one's arm in a reverberant conference room can bear significant influence on the room transfer function [Goubran96]. A filter must track and adapt to these perturbations, and it will necessarily lag behind them. Rapid convergence rates are imperative to minimize the time over which these changes are improperly modelled.
 - *Transfer Function:* In the AEC problem, the transfer function of concern is that between two transducers - the loudspeaker and the microphone. The microphone sensor in particular is very small, and the field over which it integrates energy is minute. It is conceivable that if standing waves are produced from the loudspeaker, some may interfere destructively at or near the microphone, and in that case energy at those frequencies cannot be treated by the adaptive filter. Short of an array of microphones [Amand94] [Kellermann94] there is little that can be done to address this problem.
-

1.4 Thesis Description

Many variations of FIR adaptive filter structures have been proposed to maximize convergence rates and depths in hands-free telephones¹. Even so, the desired degree of echo suppression in conference telephone environments still cannot be achieved in practice [Gilloire94]. It appears that some physical barriers to the performance of time-domain algorithms are being broached, suggesting that a new filtering philosophy tailored specifically to AEC may be required.

1.4.1 Objectives and Motivation

The objectives of this thesis, then, are several:

- To review the development of adaptive FIR filter structures applicable to AEC from first principles to the state of the art;
- To investigate in detail a class of structures known as *Transform Domain Adaptive Filters* (XDAP's), with special consideration reserved for a particular sub-class of these structures which introduce no propagation delay in the primary path. This family of filtering algorithms were first presented in [Narayan81];
- To propose, justify, and test modifications to these structures which have not been described in the literature.

It is well known that an adaptive filter will converge to its optimal solution more rapidly with broadband, rather than narrowband, excitation. Accordingly, if tools are available which make narrow-band reference signals appear more broadband, then it seems reasonable to apply them to signals incident to the adaptive filter.

One set of such tools is the family of trigonometric expansions commonly referred to as *Fourier Transforms*. With the development of efficient algorithms for computing Fourier coefficients [Cooley65], it is now possible to calculate higher orders of the *Discrete Fourier Transform* (DFT) and its relatives in real-time - a key requirement for applications such as AEC. Adaptive

1. Several IIR methods have been proposed in the time domain [Mikhail86] [Mikhael89] [Mboup91] and more recently in other domains [Kim95], however IIR structures will not be considered in this work.

filters which employ this type of pre-processing are often said to operate (rather ambiguously) in the “Frequency Domain” (for the special case of the DFT) or in the “Transform Domain” (for other transformations, including non-trigonometric ones).

1.4.2 Contributions

Through the course of this research, a number of new XDAF structures and rules of thumb for controlling them were devised. Those showing the greatest promise were singled out for further investigation. The key contributions to the signal processing community unveiled here include:

- The successful application of Narayan’s structure to Acoustic Echo Cancellation;
 - The development of a procedure called *Dynamic Step-size Initialization* (DSI) which reduces the persistence of initial power estimates in the early stages of convergence so as to minimize MSE overshoot on start-up [Haykin96] (pages 469-477), [Lee86b] while having no effect on the steady-state performance;
 - The establishment of a universal rule for selecting the XDAF parameter β which determines the rate at which past measurements of instantaneous power are dismissed from the present power estimates. The preferred value of β for a given adaptive filter forges a near-optimal compromise between the tracking capability and the steady-state misadjustment of the filter, regardless of the filter order;
 - The introduction of a simple non-linear *Peak Window Power* (PWP) estimator which ensures filter stability at every iteration. Unlike the linear estimators described in the literature, the PWP estimator is delay-less when the incoming signal power is increasing, so it is able to reduce the channel step-sizes before any filter instability is introduced. The PWP estimator improves the plant tracking capability and the steady-state misadjustment of the transform domain filter simultaneously, and both of these metrics exceed the capabilities of similar structures based on linear estimators.
-

1.4.3 Outline

This chapter has served three functions:

- it introduced the nature and elements of the AEC problem;
- it outlined why transform domain algorithms warrant further study in this application;
- it summarized the key objectives and contributions of this research.

The remaining chapters are organized as follows:

An anthology of time-domain adaptive filtering methods will be presented in Chapter 2. Some familiarity with adaptive filter theory on the part of the reader will be assumed. A review of the fundamentals is provided in appendix A1 for those seeking a short introduction to the subject. The LMS algorithm has emerged as the workhorse driving most adaptive filter structures, so it is described in detail as its operation is central to all algorithms discussed in this work. Subband filters are also introduced in chapter 2, and it will be shown that these structures act as a bridge between the time-domain and transform-domain algorithms.

Transform-Domain methods are introduced in Chapter 3. Two sub-classes of transform-domain structures are identified: the first is a block-oriented family of algorithms which are known to be numerically efficient, but which introduce a flat delay; the second is a sample-oriented family that is numerically intensive, but delayless. Both exploit the advantages offered by transform-domain processing, but their operating principles are very different from one another. A detailed mathematical description of transform-domain processing is provided in appendix A1 also.

In chapter 4, the sample-oriented transform-domain algorithm will be explored in greater detail, with particular emphasis paid to the selection of its control parameters. Later, modifications to the algorithm will be proposed in an effort to tailor it to acoustic echo cancellation. None of these appear to have been presented explicitly in the literature. Those which seem to be most promising will be singled out for experimentation in subsequent chapters.

A summary of the processes and apparatus used to acquire and analyze the experimental data is provided in chapter 5. A more detailed description of the stochastic signals, speech tracks, and

room transfer functions and acoustic environments used to evaluate the adaptive filter structures is given in appendix A2.

Performance of the subject algorithms under tightly-controlled circumstances is examined in chapter 6. The purpose of these experiments is three-fold:

- to establish the performance limits and characteristics of the different stages of the filter;
- to determine the influence of the various control parameters on algorithm behaviour so they can be chosen to suit the problem at hand;
- to demonstrate the benefits of DSI and the PWP estimator in transform-domain filters.

These results will demonstrate the power of orthogonal pre-processors in the context of adaptive filtering with narrow-band and time-varying inputs.

In chapter 7, the ability of the algorithms to treat actual conference telephone signals will be demonstrated. The published sample-oriented transform domain algorithm will be shown to be superior to the time-domain methods in most cases. Further improvements due to the proposed modifications will also be demonstrated.

Conclusions and suggestions for further research will be provided in Chapter 8. References cited in the thesis will be listed appendix A3, and finally, a bibliography containing ISBN and library classification numbers will be provided in appendix A4 for the benefit of interested readers.

2

Time-Domain Adaptive Filters

The study begins with a review of conventional FIR adaptive filters in a system identification arrangement. Presently, the overview is restricted to those implemented in the *time domain*, in which case the data streams are presented to the filter in the sequences that they are received, with no additional pre-processing. In and after chapter 3, the benefits of pre-processing the incoming data prior to filtering will be investigated.

Familiarity with elementary vector calculus and linear algebra in the context of digital systems and adaptive filter theory is assumed hereafter. Readers seeking an in-depth introduction to these subjects are referred to any comprehensive text ([Widrow85] [Haykin96] [Treichler87], amongst others), and to appendix A1 where the more pertinent results are developed and summarized.

2.1 Structure

The ingredients of a system identification problem include an unknown plant, an adaptive filter, and some reference excitation sequence which is applied to both. The arrangement is shown in figure 2.1. To simplify the description of the adaptive filter, the following assumptions are made:

- the excitation is a *Wide-Sense Stationary* (WSS) stochastic process of great length;
- both the unknown plant and adaptive filter are FIR structures;
- the coefficients of the unknown plant are fixed;
- the coefficients of the adaptive filter are variable under some prescribed mechanism;
- the system is free of extraneous noise.

FIGURE 2.1: The generalized parallel adaptive digital filter in the time domain. The plant is shown as a linear FIR structure, though in practice it may feature any number of poles, non-linearities, and other noise sources.

The following quantities are defined and illustrated in figure 2.1: The *observation vector*

$$\mathbf{x}_k = [x_k \ x_{k-1} \ x_{k-2} \ \dots \ x_{k-N+2} \ x_{k-N+1}]^T \quad (2.1)$$

consists of the N most recent unweighted elements of the reference signal, and the *adaptive filter coefficient vector* in the time domain is constituted as

$$\mathbf{w}_k = [w_k(0) \ w_k(1) \ w_k(2) \ \dots \ w_k(N-2) \ w_k(N-1)]^T, \quad (2.2)$$

where k is the iteration number and N is the number of elements in (or the *order* of) the adaptive filter. The (scalar) output of the adaptive filter is computed as the inner product of $\mathbf{x}_k$ and $\mathbf{w}_k$, so

$$y_k = \sum_{n=0}^{N-1} w_k(n) x_{k-n}, \quad (2.3)$$

or, more compactly,

$$y_k = \mathbf{x}_k^T \mathbf{w}_k \quad (2.4)$$

The elements of $\mathbf{w}_k$ are adjusted at each iteration towards an optimal set of values $\mathbf{w}_{opt}$. The *instantaneous* error e_k is the difference between the outputs of the filter and the unknown plant,

$$e_k = d_k - y_k \quad (2.5)$$

where d_k is the current *primary input* sample. The adaptive filter performs a *linear convolution* between $\mathbf{x}_k$ and $\mathbf{w}_k$ (equation (2.4)) , and a *linear correlation* between y_k and d_k (equation (2.5)) over time¹. It should be clear that the error signal is a measure of the distance between vectors $\mathbf{w}_k$ and $\mathbf{w}_{opt}$, for as $\mathbf{w}_k$ approaches $\mathbf{w}_{opt}$, e_k moves towards zero, and artifacts of the reference signal are cancelled from the primary input. The objective, then, is to construct a robust algorithm which will adjust $\mathbf{w}_k$ and yield $\mathbf{w}_{opt}$ quickly for a wide variety of reference inputs. Note that if $\mathbf{w}_{k-1} = \mathbf{w}_k$ for all k , the filter is static or non-adaptive, rendering its subscript redundant.

To formalize the processes through which the adaptation will be undertaken, it is necessary to define several higher-order variables from which statistical knowledge of the inputs can be obtained. The theory of time-domain adaptive filters is introduced in greater detail in section A1.4 of Appendix A1, so only a cursory introduction will be provided here.

The *input autocorrelation matrix* is defined as

$$\mathbf{R}_{xx} \equiv E[\mathbf{x}_k \mathbf{x}_k^T] \quad (2.6)$$

and the *cross-correlation vector* between the reference and primary inputs is defined as

$$\underline{p} \equiv E[d_k \mathbf{x}_k] \quad (2.7)$$

1. The convolution in an adaptive filter is truly linear only at one instant of time, unless the coefficients are held constant over some interval, in which case it is linear over that interval.

Merging equations 2.4 and 2.5 and squaring the result produces the *instantaneous squared error*

$$e_k^2 = d_k^2 - 2d_k \mathbf{x}_k^T \mathbf{w}_k + \mathbf{w}_k^T \mathbf{x}_k \mathbf{x}_k^T \mathbf{w}_k \quad (2.8)$$

which is to be minimized in some sense by the adaptation algorithm. To keep the development valid for general stochastic inputs, the time-average of the squared error must be considered. Using the expectation operator and assuming that $\mathbf{w}_k$ has settled to become time-invariant, the *Mean Squared-Error* (MSE) is defined as

$$\xi = E[e_k^2] = E[d_k^2] - 2E[d_k \mathbf{x}_k^T] \mathbf{w} + \mathbf{w}^T E[\mathbf{x}_k \mathbf{x}_k^T] \mathbf{w} . \quad (2.9)$$

The gradient of ξ is, using equations 2.6 and 2.7 and assuming that d_k and $\mathbf{x}_k$ each have means equal to zero,

$$\nabla = \frac{\partial \xi}{\partial \mathbf{w}} = -2\mathbf{p}^T \mathbf{w} + \mathbf{w}^T \mathbf{R}_{xx} \mathbf{w} . \quad (2.10)$$

The minimum MSE is obtained where its gradient is zero, which yields

$$\mathbf{R}_{xx} \mathbf{w} = \mathbf{p} \quad (2.11)$$

and finally the solution

$$\mathbf{w}_{opt} = \mathbf{R}_{xx}^{-1} \mathbf{p} . \quad (2.12)$$

The MSE is quadratic in $\mathbf{w}$ and therefore its gradient is zero only at the unique point where the minimum MSE is achieved. This result can be viewed intuitively using figure 2.1 by expanding equation 2.11 and removing the expectation to produce

$$\mathbf{x}_k \mathbf{x}_k^T \mathbf{w}_{opt} = d_k \mathbf{x}_k \quad (2.13)$$

and reducing to

$$\mathbf{x}_k^T \mathbf{w}_{opt} = d_k , \quad (2.14)$$

which states that the adaptive filter output is equal to the primary input. If this is the case, then the error signal in equation 2.5 is zero and no further adjustment of the filter is required.

There are two dominant techniques for finding w_{opt} described in the literature:

Inversion Methods attempt to locate w_{opt} directly from the inverse of R_{xx} at each iteration by way of equation 2.12. These methods are *Least-Squares* (LS) algorithms, which are often used in curve-fitting exercises. LS-based algorithms provide rapid initial convergence for a wide variety of inputs, but they demand considerable processing power. Most practical implementations employ variants of the *Recursive Least-Squares* (RLS) algorithms which update the *inverse* of R_{xx} directly rather than through matrix inversion. Even so, these methods remain numerically intensive, prone to instability, and sensitive to noise and quantization errors. Whereas they have been proposed for use in certain AEC applications, they will not be investigated further here.

Gradient Search Methods, on the other hand, converge recursively towards w_{opt} at a rate derived from the magnitude of the error observed at each iteration. These algorithms are simpler than the inversion methods, but they can be much slower to converge, particularly when the PSD of the input is non-uniform. They have proven to be very popular in real-time system-identification problems such as AEC owing to their simplicity and stability. All of the algorithms which will be investigated here are adapted as a function of estimated gradients.

Recently, considerable attention has been directed towards *Sub-Band Adaptive Filter* (SBAF) structures. Here, filter banks split the incoming signals into complete sets of frequency bands, and the cancellation is achieved with a bank of parallel (gradient-based) adaptive filters, each attempting to cancel unwanted energy over its prescribed frequency range. These methods appear to require much more computational effort than conventional time-domain algorithms, however the burden can be reduced substantially by down-sampling the filtered signals.

On one hand, SBAF's operate in the time domain because each converged parallel adaptive filter models the impulse response of the plant for a particular portion of the frequency spectrum. On the other hand, the filter banks constitute a pre-processing stage, so the reference signal has been transformed to some extent prior to its presentation to the adaptive filter bank. SBAF algorithms are included here as a bridge to chapter 3, where a set of structures which operate in domains other than time will be discussed.

Finally, to cast the system identification problem into the context of AEC,

- the reference input x_k is the signal received from the network;
- The primary input d_k is the transduced near-end signal detected by the microphone;
- The error signal e_k is the system output that is coded and transmitted to the network;
- The unknown plant is the sum of all transmission paths from the adaptive filter input x_k to the microphone output, d_k . The plant includes all elements discussed in sections 1.3.2 and 1.3.3.

2.2 The Method of Steepest Descent

The MSE in equation 2.10 defines a *performance surface* that is quadratic in the N dimensions of w . An intuitive means of charting a trajectory to the bottom of a concave-up quadratic surface is to follow the path of steepest descent at every step. This is accomplished in an adaptive filter by adjusting the weight vector so as to push the present operating point in the direction of the steepest (negative) gradient of the surface. As the optimal solution is approached, the magnitude of the gradient becomes smaller. These thoughts are formulated into an algorithm as follows: Let

$$w_{k+1} = w_k + \mu \left[\frac{\partial \xi_k}{\partial w_k} \right] \quad (2.15)$$

where μ is a scalar *step-size*. The bracketed term represents the gradient from equation 2.10,

$$\frac{\partial \xi_k}{\partial w_k} = 2R_{xx}w_k - 2p \quad (2.16)$$

Thus,

$$w_{k+1} = w_k + \mu [p - R_{xx}w_k] \quad (2.17)$$

which is, (using equation 2.12),

$$w_{k+1} = w_k + \mu R_{xx}(w_{opt} - w_k) \quad (2.18)$$

where the constant 2 is merged into μ . The trouble is, the surface described in equation 2.18 is not generally in *standard position* [Anton] (page 423), so it is awkward to describe the trajectory of the operating point as the filter elements converge. Using the methods in section A1.3 of appendix A1 as a template, the performance surface can be relocated into standard position by re-writing equation 2.18 as

$$\mathbf{w}_{k+1} - \mathbf{w}_{opt} = (\mathbf{w}_k - \mathbf{w}_{opt}) - \mu \mathbf{R}_{xx} (\mathbf{w}_k - \mathbf{w}_{opt}) \quad (2.19)$$

and making the substitution

$$\underline{\mathbf{w}}_k = \mathbf{w}_k - \mathbf{w}_{opt} \quad (2.20)$$

Note that $\underline{\mathbf{w}}$ and $\mathbf{w}$ share the same coordinate space. Thus,

$$\underline{\mathbf{w}}_{k+1} = \underline{\mathbf{w}}_k - \mu \mathbf{R}_{xx} \underline{\mathbf{w}}_k \quad (2.21)$$

This completes the translation of the performance surface to a point at or above the origin of the $\underline{\mathbf{w}}$ coordinate system. The need remains to rotate the surface so that its principal axes are aligned with the axes of the coordinate system. $\mathbf{R}_{xx}$ can be diagonalized by writing

$$\mathbf{R}_{xx} = \mathbf{S} \mathbf{\Lambda} \mathbf{S}^T \quad (2.22)$$

where $\mathbf{\Lambda}$ is a diagonal matrix whose non-zero elements are the *eigenvalues* of $\mathbf{R}_{xx}$, and $\mathbf{S}$ is some unitary matrix whose columns are the *eigenvectors* of $\mathbf{R}_{xx}$. The unitary property ensures that

$$\mathbf{S}^{-1} = \mathbf{S}^T \quad (2.23)$$

Now, equation 2.21 can be re-written as

$$\underline{\mathbf{w}}_{k+1} = \underline{\mathbf{w}}_k - \mu \mathbf{S} \mathbf{\Lambda} \mathbf{S}^T \underline{\mathbf{w}}_k \quad (2.24)$$

Pre-multiplying each term by $\mathbf{S}^T$ produces

$$\mathbf{S}^T \underline{\mathbf{w}}_{k+1} = \mathbf{S}^T \underline{\mathbf{w}}_k - \mu \mathbf{S}^T \mathbf{S} \mathbf{\Lambda} \mathbf{S}^T \underline{\mathbf{w}}_k \quad (2.25)$$

which leads to, using equation 2.23,

$$S^T \underline{w}_{k+1} = S^T \underline{w}_k - \mu \Lambda S^T \underline{w}_k. \quad (2.26)$$

By making one final substitution

$$\bar{\underline{w}}_k = S^T \underline{w}_k \quad (2.27)$$

the desired coordinate system which renders the performance surface in standard position is obtained:

$$\bar{\underline{w}}_{k+1} = \bar{\underline{w}}_k - \mu \Lambda \bar{\underline{w}}_k, \quad (2.28)$$

or, in simpler form,

$$\bar{\underline{w}}_{k+1} = (I - \mu \Lambda) \bar{\underline{w}}_k \quad (2.29)$$

Equation 2.28 is in standard position because the minimum of the performance surface occurs at $\bar{\underline{w}}=0$, and the matrix Λ is diagonal. Once again, $\bar{\underline{w}}$ and $\underline{w}$ share the same coordinate space.

The next objective is to determine the range of μ which guarantees stability. Maintaining the assumption that the reference is WSS, the eigenvalues of R_{xx} are constant, so the bracketed term of equation 2.29 is constant for any given μ . Equation 2.29 can be re-written in recursive form as

$$\bar{\underline{w}}_{k+1} = (I - \mu \Lambda)^k \bar{\underline{w}}_0 \quad (2.30)$$

which renders it a geometric progression that is convergent if and only if its geometric ratio has magnitude less than unity, i.e.

$$|I - \mu \lambda_{max}| < 1 \quad (2.31)$$

which requires

$$-1 < (1 - \mu \lambda_{max}) < 1. \quad (2.32)$$

All the eigenvalues of R_{xx} are known to be positive, so solving both inequalities in equation 2.32 simultaneously produces the stability bounds for μ :

$$0 < \mu < \frac{2}{\lambda_{max}} . \quad (2.33)$$

Typically, μ is chosen well below $\mu_{max} = 2/\lambda_{max}$ so as to avoid conditions of marginal stability. Further, since it is difficult to compute the eigenvalues of (a high-order) R_{xx} , the maximum value of μ is often restricted to

$$0 < \mu < \frac{2}{tr[R_{xx}]} \quad (2.34)$$

which is certain to provide stability because

$$\lambda_{max} < tr[\Lambda] \quad (2.35)$$

at all times and

$$tr[\Lambda] = tr[R_{xx}] . \quad (2.36)$$

Now equation 2.30 can be re-written so as to obtain the convergence path in terms of w . This is accomplished through back-substitution (following [Widrow85] pp 56-60), with the final result being

$$w_k = w_{opt} + (I - \mu R_{xx})^k (w_0 - w_{opt}) . \quad (2.37)$$

Equation 2.37 is readily converted to an iterative form as

$$w_{k+1} = w_k + (I - \mu R_{xx})^k (w_0 - w_k) \quad (2.38)$$

which is the steepest descent algorithm.

Assuming μ is chosen to ensure stability, the rightmost term will decay geometrically with k and the unique minimal solution will be achieved when the remaining terms are equal. Note that if μ is set to zero, the adaptation is frozen and w_k remains constant. If it is set to $2/\lambda_{max}$, then the filter is only marginally stable, and cannot be expected to perform well. This algorithm is well-suited

to DSP devices as it requires repeated multiply-add operations to effect the convolution between two vector quantities.

Note that the trajectory is very dependent on w_0 . Thus, *steepest descent filters are slow to track changes in the excitation*. Older gradient estimates tend to restrain the filter taps when they should be changing course to suit the updated performance surface. The method of steepest descent as written has no means of dismissing outdated gradient estimates, and thus its use in applications where the reference signal is non-stationary is limited at best.

2.3 Least Mean Squares Algorithm

To make the steepest descent algorithm track non-stationary signals effectively, a mechanism to dismiss out-dated gradient estimates must be adopted. In applications such as AEC, new samples of the reference and primary signals are available at each iteration, so it seems reasonable to use these to estimate the gradient over shorter periods of time.

It should be clear that the portion of the primary which is to be cancelled is the exact linear convolution between the unknown plant and the reference signal. Thus one can assess the quality of the adaptive filter by placing it in parallel with the unknown plant as in figure 1.2, and measuring the correlation of its output with the primary signal. To be precise, *the error signal itself could be used as an instantaneous estimate of the gradient*. As it is updated each iteration, the error output becomes more dependent on x_k , d_k , and w_k and less dependent on w_0 as k increases.

The approach, as before, is to construct a cost function that is quadratic in w and then to find its minimum, but to use the error as a feedback signal by which the adaptive filter coefficients are adjusted. The error signal seems appropriate because the gradient of this quadratic function increases as w_k is moved radially outward from w_{opt} (i.e. as e increases), regardless of the shape or the dimension of the function. Further, the rate at which it increases is related to the slope of the surface in the direction of travel.

These are the principles underlying the *Least Mean Squares* (LMS) algorithm, described in 1960 by Bernard Widrow and generalized in [Widrow75] for the complex case. The LMS algorithm

has become the workhorse of many adaptive structures, and it has been investigated and refined by many researchers over the years. Modern LMS algorithms are especially popular owing to their simplicity, efficiency, and stability. However, under some (not uncommon) conditions, they can be very slow to converge [Morgan95a] [Slock93]. All the structures yet to be discussed are erected on the foundations of LMS, so a detailed review of the algorithm is in order.

2.3.1 Derivation

As before, the quadratic cost function is the MSE. Its derivative with respect to the evolving adaptive filter

$$\frac{\partial \xi_k}{\partial \mathbf{w}_k} = 2e_k \frac{\partial e_k}{\partial \mathbf{w}_k} \quad (2.39)$$

is submitted as an estimate of the instantaneous gradient. Using the definition for e_k in equation 2.5, the instantaneous gradient is

$$\frac{\partial \xi_k}{\partial \mathbf{w}_k} = 2e_k(-\mathbf{x}_k) = -2e_k\mathbf{x}_k \quad (2.40)$$

The iterative LMS algorithm falls together from equation 2.15 as

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \mu e_k \mathbf{x}_k \quad (2.41)$$

where once again the constant 2 is merged into the step-size.

The optimal solution is located where $\mathbf{w}_k = \mathbf{w}_{opt}$ which occurs where the derivative of ξ_k with respect to $\mathbf{w}_k$ is zero. Being of the steepest descent family, the LMS algorithm is guaranteed to be stable if the bounds of μ expressed in equation 2.33 are maintained at all times. This demand can be met easily if the reference input power is constant and known in advance, and if the PSD is uniformly distributed. However, in most practical cases, the reference power and PSD are time varying, in which case some careful consideration is warranted.

2.3.2 Properties for Stationary Inputs

If the reference input signal is known to be WSS for all time, an upper bound for μ valid for all time is prescribed by the largest eigenvalue of $\mathbf{R}_{xx}$. Then, if the PSD of the reference is flat, all the eigenvalues of $\mathbf{R}_{xx}$ are (nearly) equal, making the step-size equally valid in each eigenmode, and ensuring even convergence towards $\mathbf{w}_{opt}$. If the PSD is non-uniform however, the under-excited modes will be slower to converge than those with greater excitation. The *convergence time constant* of the n th eigenmode of the LMS filter is [Widrow85] (pages 107-109)

$$\tau_{MSE}(n) = \frac{1}{4\mu\lambda_n} \quad (2.42)$$

Thus, it is possible to choose μ to achieve a certain convergence rate for a particular eigenmode provided μ remains within its stability interval¹. The behaviour of the filter during tracking and in the steady state can vary considerably depending on where within these bounds μ is chosen. The steady-state error, commonly referred to as the *Misadjustment*, is given by [Widrow85] (pages 110-112)

$$M_T = \mu(\text{tr}[\mathbf{R}_{xx}]) \quad (2.43)$$

The trade-off between speed of convergence and misadjustment is now apparent: the former is inversely proportional while the latter is proportional to μ .

2.3.3 Considerations for Non-Stationary Inputs

The applicability of eigenanalysis is much less clear if the reference input is non-stationary, because $\mathbf{R}_{xx}$ can no longer be assumed to be time-invariant. Not only might the PSD of the input be varying, (in which case the behaviour in each eigenmode will be uneven in time), but the overall power in the reference might be changing also, in which case the stability of the entire filter may be in question.

1. In time-domain filters, individual step-sizes to suit each eigenmode cannot be specified unless it is known in advance that the eigenvectors of $\mathbf{R}_{xx}$ (the principal axes of the performance surface) are aligned with the coordinate axes of $\mathbf{w}$.

The presence of non-stationarities in the reference input presents two immediate problems:

- the tracking rates will be uneven in the various eigenmodes of the filter when the PSD of the reference signal is non-uniform;
- the tracking rates will be uneven from one time to the next, if the total power in the reference input is time-varying.

Since the derivation for the steepest-descent and LMS algorithms has been predicated on the provision of WSS inputs, it is inappropriate to make general statements about the behaviour of these algorithms when excited with signals that are not WSS. However, if the rate of change of the reference statistics is low relative to the sampling frequency, it may not be unreasonable to partition the reference input into finite time segments during which the statistics are assumed to be constant. Within each time segment, a short-term autocorrelation matrix could be defined, and its eigenanalysis could be used as the basis for ensuring convergence and specifying filter performance *within that time segment*. These are the principles supporting the development of a family of LMS algorithms whose step-sizes are time-varying. In the next section, it will be shown that variations in the unknown plant may also warrant the use of such algorithms even if the reference input is stationary.

2.4 Variable Step-Size LMS Algorithms

In many adaptive filtering problems such as AEC, fast convergence and low misadjustment are highly desirable. Static algorithms such as steepest descent and LMS, however, cannot achieve both simultaneously: a larger step-size will (to a point) accelerate convergence (equation 2.42), and a smaller step-size will always reduce misadjustment in the steady state (equation 2.43).

These equations suggest that convergence rates can be accelerated if the step-size is decreased when the excitation power is large, and increased when the excitation power is small. *Power Normalized LMS algorithms* which have this feature are particularly convenient because they adapt to the reference signal automatically, and eliminate any guesswork associated with the choice of μ .

A second means for varying the step-size might be based on the proximity of $\mathbf{w}_k$ to $\mathbf{w}_{opt}$. When the MSE is large, one would prefer a larger step-size so as to converge to the neighbourhood of $\mathbf{w}_{opt}$ quickly. Once near, a lower step-size is preferred so as not to push $\mathbf{w}_k$ too far beyond $\mathbf{w}_{opt}$ on each update. The step-size adjustment mechanism in these *Proximity-Based LMS Algorithms* works in a manner akin to driving at higher speeds on an inter-city freeway and lower speeds as an urban destination is approached. A review of the variable step-size time domain LMS filters follows shortly.

2.4.1 Power-Normalized LMS Algorithms

Earlier, the LMS algorithm was shown to provide benefits over steepest descent by using short-term gradient approximations. Here, the gradient approximation is extended to account for the overall steepness of the performance surface in addition to its evolving shape.

The slope of the performance surface in the direction of a given eigenvector of $\mathbf{R}_{xx}$ is directly proportional to the associated eigenvalue, which in turn is a measure of the signal power in that eigenmode¹. The average steepness of the performance surface can be *normalized* by dividing x_k by a short-term estimate of the reference power and using the result in the filter update equation.

The objective here is to specify a scalar *time-adaptive* step-size of the form

$$\mu_k = \mu_g \mu_p \quad (2.44a)$$

where

$$\mu_p = \frac{1}{c + p_k} \quad (2.44b)$$

is a scale factor which varies with the inverse of the estimated power p_k , the scalar $c \ll p_k$ is a small positive constant which ensures division by zero cannot occur, and μ_g is a user-specified time-invariant *global stepsize* selected from the interval $0 < \mu_g < 2$.

1. Details are provided in the adaptive filter texts and are reviewed in appendix A1.

The time-adaptive step-size can be expressed in one equation as

$$\mu_k = \frac{\mu_g}{c + p_k} . \quad (2.45)$$

The tap-weight update is then computed as

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \mu_k e_k \mathbf{x}_k . \quad (2.46)$$

There are two methods of estimating the short-term reference power commonly described in the literature. The first is a *Moving Average* (MA) technique, where the power estimate p_k is simply the power in the observation vector $\mathbf{x}_k$. The second is an *Auto-Regressive* (AR) method in which p_k is a weighted integration of past squares of reference samples x_k . Widrow-based algorithms employing the former will be referred to as MA-NLMS, and those based on the latter as AR-NLMS. The two will now be described formally.

2.4.1.1 Moving-Average Normalized LMS Algorithm

The instantaneous power in the observation vector is

$$p_k = \mathbf{x}_k^T \mathbf{x}_k , \quad (2.47a)$$

which produces

$$\mu_k = \frac{\mu_g}{c + \mathbf{x}_k^T \mathbf{x}_k} \quad (2.47b)$$

and thus the MA-NLMS algorithm may be expressed in one equation as

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \frac{\mu_g e_k \mathbf{x}_k}{c + \mathbf{x}_k^T \mathbf{x}_k} . \quad (2.47c)$$

The power estimate in the MA-NLMS algorithm is “exact” in the sense that it reflects the total power in the adaptive filter at each instant of time. However, it weights the N power estimates equally, which may not be optimal in cases where the excitation statistics are time-varying.

This method appears to have been introduced in [Nagumo67], with a refined geometrical description of its performance surface provided in [Kurosawa87]. The MA-NLMS algorithm has been found to be very robust, and was used in all time-domain NLMS implementations employed in this work.

2.4.1.2 Auto-Regressive Normalized LMS Algorithm

The AR-NLMS algorithm is a two-stage adaptive filter. The first stage tracks the power in the excitation according to

$$p_k = \beta p_{k-1} + (1-\beta) \|\mathbf{x}_k\|^2 \quad (2.48)$$

where the scalar *persistence factor* β^1 dismisses past power estimates in time. The weight update is then undertaken according to equation 2.46. The AR power estimator is transformed into the z -domain as follows, where, for convenience, $\chi_k = \|\mathbf{x}_k\|^2$:

$$p_k = \beta p_{k-1} + (1-\beta) \chi_k \quad (2.49)$$

$$P(z) [1 - \beta z^{-1}] = (1-\beta) \chi(z) \quad (2.50)$$

$$H(z) = \frac{P(z)}{\chi(z)} = \frac{1-\beta}{1-\beta z^{-1}} \quad (2.51)$$

The magnitude response of the filter can be determined by making the substitution

$$z = e^{j\omega} \quad (2.52)$$

to yield the frequency response

$$H(\omega) = \frac{1-\beta}{1-\beta e^{-j\omega}} \quad (2.53)$$

Invoking the Euler identity, equation 2.53 becomes

1. It is referred to in the literature as an *exponential weighting factor* [Haykin96], a *smoothing constant* [Lee86b][Sommen87], or a *forgetting factor* [Chao90]. The latter term will be avoided as it is a better name for the complement of β , i.e. $(1-\beta)$ would be a true forgetting factor.

$$H(\omega) = \frac{1-\beta}{1-\beta\cos(\omega) + j\beta\sin(\omega)} \quad (2.54)$$

which has a magnitude response of

$$|H(\omega)| = \frac{1-\beta}{\sqrt{1-2\beta\cos(\omega) + \beta^2}} \quad (2.55)$$

The scale factor $(1-\beta)$ ensures the integrator has unity gain at DC (0 dB at $\omega=0$). The phase response is, from equation 2.54,

$$\Theta(\omega) = -\text{atan}\left[\frac{\beta\sin(\omega)}{1-\beta\cos(\omega)}\right]. \quad (2.56)$$

Therefore, the AR power estimator is a low-pass filter whose single pole is located at $z = \beta$. The pole must be on the positive real axis within the unit circle in the z -plane to ensure stability. The closer the pole is to the origin, the faster the algorithm tracks changes in the excitation power, but the more noisy the estimates are likely to be. Smoother power estimates are obtained when the pole is closer to the unit circle, but these estimates will be slow to reflect changes in the input signal power. It is equally acceptable to view p_k as the time-integration of past power estimates which are dismissed geometrically at a rate determined by the scalar β . The AR power estimator is represented schematically in figure 2.2.

FIGURE 2.2: Power estimation using a first-order AR process.

The choice of a suitable β is not immediately obvious. Numerous references suggest it should be “close to, but less than unity” [Clark83] [Lee86b] [Beaufays95b] (amongst others) on the basis that smoother estimates are more desirable. It was argued in [Youn83] that β should be made

equal to $(1-\mu_g/p)$ after each adaptive filter update so as to make the persistence (i.e. the time constants) in the power tracking and filter update stages equal. A similar method was proposed in [Diniz92] and [DeCampos93], and then analyzed in [Diniz95] for an LMS-Newton algorithm. This method is appealing as it suggests the adaptation parameters of the input and output stages can work in concert with one another under the direction of the reference input power. Other similar estimators have been used in adaptive filter networks (see [Gitlin77] [Picci84] for the channel equalization problem and [Ogue83] for general applications), but they appear to have fallen into obscurity in the literature.

2.4.2 Proximity-Based LMS Algorithms

It is also possible to vary the step-size based on the nearness of $\mathbf{w}_k$ to $\mathbf{w}_{opt}$. Of course, $\mathbf{w}_{opt}$ is unknown, so some means for determining its approximate location relative to $\mathbf{w}_k$ is required. Several interesting approaches have been proposed, some whose step-size adaptation depends on the error signal itself [Kwong92] [Mathews93] [Douglas94a] [Douglas94b] (including the use of higher-order non-quadratic cost functions [Walach84] [Shah95] and time-averaging of the error signal [Aboulnasr97]) and some that are dependent on the instantaneous gradient [Harris86] [Evans93] [Farhang94b].

Space does not allow an exhaustive review, so a few representative algorithms will be discussed. It is disappointing that none of the proximity-based algorithms were compared experimentally to the MA-NLMS algorithm in these papers, as MA-NLMS had by the early 1980's become an accepted baseline algorithm. It would be interesting to investigate how well proximity-based algorithms work in conjunction with a power-normalization scheme such as MA-NLMS.

2.4.2.1 Diagonal Matrix Variable Step-Size Algorithm

An easy way to identify the proximity of $\mathbf{w}_k$ to $\mathbf{w}_{opt}$ is to examine the polarity (i.e. the sign) of the gradient at each iteration. The N -dimensional instantaneous gradient is the (unscaled) weight adjustment term in equation 2.41:

$$\nabla_k = e_k \mathbf{x}_k \quad . \quad (2.57)$$

Using back-substitution,

$$\nabla_k = (d_k - \mathbf{w}_k^T \mathbf{x}_k) \mathbf{x}_k \quad (2.58)$$

then, assuming that $\mathbf{w}_{opt}$ is of the same order as $\mathbf{w}_k$ and that no extraneous noise is present,

$$\nabla_k = (\mathbf{w}_{opt}^T \mathbf{x}_k - \mathbf{w}_k^T \mathbf{x}_k) \mathbf{x}_k \quad (2.59)$$

$$\nabla_k = [(\mathbf{w}_{opt}^T - \mathbf{w}_k^T) \mathbf{x}_k] \mathbf{x}_k \quad (2.60)$$

Equation 2.60 can be written as

$$\nabla_k = \mathbf{x}_k \mathbf{x}_k^T (\mathbf{w}_{opt} - \mathbf{w}_k) \quad (2.61)$$

and, if the reference input is WSS,

$$\nabla_k = \mathbf{R}_{xx} (\mathbf{w}_{opt} - \mathbf{w}_k) \quad (2.62)$$

Under these circumstances, *the gradient is proportional to the distance $\mathbf{w}_k - \mathbf{w}_{opt}$ in the N dimensions of $\mathbf{w}$* . Note that the gradient vector is computed automatically during the weight update of the LMS algorithm (equation 28 of [Widrow76]), and the larger the gradient, the greater the adjustment range of $\mathbf{w}_k$, as shown in equations 2.41 and 2.42.

Any uncorrelated noise in the primary signal forces finite misadjustment in the adaptive filter that is assumed to be constant for all k . When the gradient is very large, the noise amplitude is very small with respect to the adjustment range of $\mathbf{w}_k$, so the sign of the gradient is unlikely to change from one iteration to the next. As $\mathbf{w}_k$ nears $\mathbf{w}_{opt}$ however, the range of adjustment in $\mathbf{w}_k$ becomes smaller, and eventually the noise causes the gradient vector to rotate chaotically around $\mathbf{w}_{opt}$. Thus, the closer $\mathbf{w}_k(n)$ is to $\mathbf{w}_{opt}(n)$, the more likely it is that the sign of the gradient vector in the n th dimension will change due to the misadjustment noise.

The *Diagonal Matrix Variable Step-Size Algorithm* [Harris86] exploits this LMS filter property. In it, the weight update is obtained as

$$\mathbf{w}_{k+1} = \mathbf{w}_k + e_k \mathbf{M}_k \mathbf{x}_k \quad (2.63)$$

where $\mathbf{M}_k$ is a diagonal matrix whose non-zero elements are the (time-varying) step-sizes associated with each coefficient of $\mathbf{w}$. These step-sizes are bounded by the normal LMS requirement

$$0 < \mathbf{M}_k(n, n) < 2/\lambda_{\max}, \quad \forall n \in [0, 1, 2, \dots, N-1] \quad (2.64)$$

with $\lambda_{\max}$ being the largest eigenvalue of $\mathbf{R}_{xx}$. In practice, a minimum value for $\mathbf{M}_k(n, n)$ is often specified to prevent the filter from stalling and to achieve the desired level of misadjustment after convergence has been achieved.

The scalar stepsize $\mathbf{M}_k(n, n)$ is decreased by a factor of α if $\nabla_k(n)$ changes polarity over an interval of m_0 iterations, and increased by a factor of α if the polarity of $\nabla_k(n)$ is unchanged over an interval of m_1 iterations. It is recommended in [Harris86] that m_1 be chosen greater than or equal to m_0 , and that m_1 be finite. No recommendation for the scale factor α was made, though in the simulations provided in [Harris86], α was set to two, allowing multiplications and divisions by α to be effected through simple bit shifts. The simulations provided show that as the reference signal SNR increases, the convergence rate enhancement of the proposed algorithm over standard LMS decreases.

The diagonal matrix algorithm increases the computational burden over LMS by only about 15% [Harris86], and a modest increase in memory is also required to maintain the successive counts of sign change for each gradient. An arithmetic version, where the step-size adjustments were made by adding or subtracting a small value to the step-sizes rather than by scaling them, was discussed in [Evans93]. This method reduces the step-size quantization levels, allowing better tuning at the expense of increased complexity (the additions cannot be carried out by shifts). The arithmetic version is slightly faster in terms of initial convergence and tracking behaviour, while imposing no misadjustment penalty [Evans93].

2.4.2.2 Error Magnitude Variable Step-Size Algorithm

As discussed in sections 2.3, Widrow's LMS algorithm uses the instantaneous error itself as a measure of $\mathbf{w}_k - \mathbf{w}_{opt}$. This notion was extended in [Kwong92], where an adaptive scalar step-size was proposed to accelerate the static LMS update when the error is large, and to reduce it when

the error is small. The parameter updates in this *Error Magnitude Variable Step-size Algorithm* are carried out as follows:

The adjusted scalar step-size is computed as

$$\mu'_{k+1} = \alpha\mu_k + \gamma e_k^2, \quad 0 < \alpha < 1, \quad \gamma > 0 \quad (2.65)$$

where α and γ are arbitrary parameters chosen to govern the adaptation. Then, to ensure stability, the step-size is bounded by

$$\mu_{k+1} = \begin{bmatrix} \mu_{max} \\ \mu_{min} \\ \mu'_{k+1} \end{bmatrix} \text{ for } \begin{bmatrix} \mu'_{k+1} > \mu_{max} \\ \mu'_{k+1} < \mu_{min} \\ (otherwise) \end{bmatrix} \quad (2.66)$$

where

$$\mu_{max} \leq \frac{2}{3 \operatorname{tr}[R_{xx}]} \text{ and } 0 < \mu_{min} < \mu_{max} \quad (2.67)$$

The weight update is made using equation 2.42, however the adjusted step-size of equation 2.66 replaces the static μ . A distinct advantage of this algorithm is that (unlike the other embellished LMS algorithms) it requires almost no additional computing power nor memory than standard LMS. This is because the step-size remains scalar, which is not the case in the diagonal matrix algorithms. Its main disadvantage is that α and γ must be chosen, which implies that the filter may produce uneven results for successive trials, even when using stochastic inputs with similar statistics. Thus, the optimal pair (α, γ) for one trial might be quite different from that of another, and it will be difficult to determine these parameters in advance for a given application.

Simulations and analysis in [Kwong92] show that the proposed algorithm outperforms standard LMS in both stationary and non-stationary environments, and performs at least as well as one version of the diagonal matrix algorithm of [Harris86]. The numerous arbitrary parameters in these algorithms make it difficult (if not impossible) to compare their performance in general terms, however.

2.4.2.3 Gradient Adaptive Step-Size LMS Algorithm

An algorithm whose step-size was truly self-adjusting would be highly desirable for general applications. Such a method was originally proposed in [Shin85] and subsequently fine-tuned in [Mathews93], where it was labelled the *Stochastic Gradient Adaptive Filter with Gradient Adaptive Step-Size*. In it, the scalar step-size is adjusted recursively as a continuous function of the reference input and measured error outputs as follows:

$$\mu_k = \mu_{k-1} + \rho e_k e_{k-1} \mathbf{x}_k^T \mathbf{x}_{k-1} \quad (2.68)$$

Here, ρ is a small positive constant that determines the adjustment range of the step-size. The tap-weight update is conducted using equation 2.42 with the step-size of equation 2.68, and the (fixed) parameter ρ acts on the step-size rather like μ acts on the tap-weights in standard LMS. Specifically, a “step-size performance surface” is implied, where some μ_{opt} lies at the minimum, and ρ is a more primitive “step-size” which regulates the rate of travel of μ_k towards μ_{opt} using a gradient method not unlike that of conventional LMS. This analogy raises questions about the stability bounds of ρ . It is suggested in [Mathews93] that these limits are difficult to determine, so a simplified constraint on μ_k is recommended, such as that in equation 2.67.

This algorithm works well provided the reference is WSS. When the reference is not stationary, stability will be much more awkward to analyze and maintain using this kind of “saturation” approach. Simulations were provided in [Mathews93] demonstrating the algorithm’s excellent performance in stationary environments, and its tracking ability when abrupt non-stationarities are applied to plant. In these cases the time-varying step-size was shown to be very responsive, enabling re-convergence to be much faster than with standard LMS. One tremendous advantage to this algorithm is its computational simplicity: it requires only four extra multiplications and three extra additions per iteration, and an additional N memory locations. A matrix version of this algorithm was discussed in [Mathews90], however, its performance was not significantly better than that of the scalar version.

A robust algorithm incorporating the step-size update of [Kwong92] and error autocorrelation of [Mathews93] was described and analyzed in [Aboulnasr97]. This algorithm converges faster in stationary environments than the others almost regardless of the SNR in the primary signal.

2.4.3 Application-Specific LMS Algorithms

When certain properties of the unknown plant are known in advance, it is possible to jump-start the convergence somewhat by incorporating that knowledge into the algorithm. Computational effort is not wasted on previously known or less important coefficients or step-sizes, allowing faster identification of the more significant elements of the plant. Acoustic transfer functions are good candidates for such algorithms, as the impulse response is known to decay and changes in the plant tend to be localized in time.

2.4.3.1 Reduced-Order Algorithms

Lower-order systems are often easier to identify than higher order systems because the number of unknowns is reduced and the step-sizes can be made larger. Smaller autocorrelation matrices have smaller maximum eigenvalues than their higher-order counterparts for the same reference signal [Weslowski92]. As a result, the stability bounds on μ are broader when lower-order filters are implemented. This was made apparent in equation 2.35.

A *Variable-Length Adaptive Filter* was proposed in [Weslowski92] for adaptive line equalization of binary telephone channels. The length of the filter on start-up was made very small, and was increased by pre-determined amounts at specified intervals until the full desired filter length was achieved. The step-size was decreased as the filter order was increased so as to maintain stability at all times. This method provided faster initial convergence and equal misadjustment as the standard LMS algorithm in the simulations provided in [Weslowski92]. It is probably best suited to applications where most of the impulse response energy of the unknown plant is located in the lower-order terms, as is the case in AEC.

Algorithms tailored to systems with sparsely located taps were investigated in [Ikeda91] and [Margo93]. In the first case, the lower-order coefficients were assumed to be near-zero, implying the system to be identified exhibited a flat delay. A satellite transmission link is representative of such a system. This algorithm uses an innovative scheme to activate those taps which are most significant in the impulse response, and de-activate those of lesser importance automatically. The tap-significance detection mechanism runs concurrently, so the significant coefficients are promoted or removed from the active tap list continuously.

The method proposed in [Margo93] held similar objectives, however it was implemented as a bank of short-duration parallel adaptive filters, each delayed by different amounts to cover those areas of the impulse response known to contain significant energy. This algorithm is suitable for use in applications where the impulse response energy is known in advance to be time-localized. A secondary adaptation mechanism was discussed which allows the flat delay of each filter to be varied so as to maintain alignment between the “centres of gravity” of the adaptive filter and the time-block of energy it models in the unknown system, which may move from time to time in a non-stationary plant.

In certain applications, several unknown plants may require identification simultaneously by a single adaptive filter. An example might be a beamformer which is switched to detect several active sources at different periods of time. A *Time-Sequenced Adaptive Filter* [Ferrara81] was proposed to address such a problem. The structure maintains a bank of parallel adaptive filters, each of which models one unknown plant. Exactly one of the adaptive filters is active at any one time, and the activation is synchronized with the (already known) switching instants of the plant. This method negates the requirement for the adaptive filter to re-converge completely each time the plant is switched, since the structure retains knowledge of each of the unknown channels in its memory. Hence, the adaptation picks up where it was left off each time a filter is re-activated.

2.4.3.2 The Exponentially-Weighted Step-Size LMS Algorithm

A modified MA-NLMS algorithm tailored specifically for AEC problems was introduced [Makino90] and analyzed [Makino93] as the *Exponentially-Weighted Step-Size LMS Algorithm* (ES-NLMS). The algorithm was devised from experimental data which showed that the variance in tap-weight values decreases exponentially with filter order while the filter tracks a changing acoustic transfer function. The tap-weight updates were made as in the MA-NLMS algorithm, but with a diagonal step-size matrix reminiscent of that presented in [Harris86]. Simulations in [Makino90] compared the initial convergence of this algorithm to that of MA-NLMS in a two-channel sub-band structure; a study of its tracking behaviour was deferred until [Makino93]. Further refinements have been introduced as the *Exponential Step-Size Projection algorithm* and the more general *Exponential Step-Size Recursive Least Squares algorithm* [Makino95]. The motivation behind these orthogonalized structures is increased convergence and tracking rates.

Claims that the ES-NLMS algorithm requires no more computing power than MA-NLMS in [Makino93] are rather dubious, because it was implemented in that paper with multiple parallel DSP's, each programmed to run the standard MA-NLMS algorithm on a constant- μ portion of the full-length filter. A truly fair comparison would be made with each method implemented on a single DSP, in which case ES-NLMS has a vector step-size.

Experimental results provided in [Webster95] confirmed that full-band ES-NLMS filters provide significantly better tracking of time-varying acoustic transfer functions than similar MA-NLMS structures. The improvement is minor, however, when eight-channel sub-band configurations of these algorithms are compared.

2.5 Sub-Band Architectures

So far only wide-band adaptive filters have been discussed, in which case the filter inputs are assumed to occupy the entire telephony band. Since the PSD of the reference input is generally non-uniform, the convergence rates of the filter in any under-excited modes (hence the overall convergence of the filter) can be very slow [Slock93] [Morgan95a].

In this section, the notion of sub-banding will be introduced. Sub-band structures decompose the incoming signals into m parallel data streams, the bandwidth of each being confined to a certain portion of the reference spectrum. Once the data is partitioned, it is possible to assess the relative contribution of each channel to the overall signal, and to discard or de-emphasize those which matter least.

In the context of adaptive filters, sub-banding allows the convergence modes to be equalized to some degree by using m adaptive filters operating in parallel, each constrained to operate in their respective frequency bands. Each SBAF can be regulated using any of the power normalization methods¹ of section 2.4.1 so as to maximize the overall performance on a band-by-band basis.

1. It appears that no investigation into proximity-based normalization methods has been conducted in the context of sub-band adaptive filters, even though their use seems quite reasonable.

2.5.1 Foundations

Sub-band methods were employed in speech coding applications [Crochiere75], [Crochiere76] in the mid-1970's. In this case, transmission efficiency and control over quantization noise are important objectives. In these early implementations, an *articulation index* [Crochiere76] was proposed as a template for specifying the filter bank bandwidths. This index divides the voice band into twenty non-uniform contiguous bins, each of which is designed to contain the same amount of energy as perceived by the human auditory system. The quantization noise in a given channel of a sub-band coder is confined largely to that channel [Smith86], so pass-band signals tend to mask the quantization noise automatically [Chu85]. Concise mathematical descriptions for sub-band systems were provided in [Smith87] and [Vetterli87].

Since each output channel from the analysis filter banks occupies only a small portion of the total bandwidth, it is possible to *down-sample* (or decimate¹) each of these signals without the normal penalties imposed by Shannon's sampling theorem. Signals can be reconstructed through *up-sampling* and *interpolation* with suitable modulation processes. Principles of sample-rate conversion are outlined well in numerous papers and publications [Crochiere81], [Crochiere83], [Oppenheim89], [Vaidyan90], so they will not be reviewed here.

2.5.2 Perfect Reconstruction Systems

There has been much research conducted into the design of analysis/synthesis filter banks which possess properties of *Perfect Reconstruction* (PR) or *Near-Perfect Reconstruction* (NPR). A PR filter bank decomposes a wide-band signal into a set of down-sampled narrow-band signals which can be re-combined such that the aliasing distortion produced in each band is cancelled *exactly* by alias terms in other bands. PR filter banks are descendants of the *Quadrature Mirror Filters* (QMF) described in [Chu85] and [Vaidyan87a]. Unlike those described in [Crochiere76], these QMF and PR filter banks split the incoming signals into channels of equal bandwidth spaced evenly over the entire input spectrum.

1. The term *decimation* is very common in the literature, but it is used incorrectly, and hence will not be employed in this work. To decimate means to reduce *by* one-tenth. Its usage in common english would suggest it means to reduce *to* one tenth. Its meaning in sample-rate conversion is doubly questionable, as the number of samples dismissed need not be a factor of ten at all, and is normally greater than half.

The earliest PR filter bank design procedure was developed from a simple tree-structure using *conjugate quadrature filters* [Smith86]. The procedure was later generalized [Smith87]. An alternative frequency-domain technique based on *polyphase networks* was proposed almost coincidentally [Princen86]. Here, analysis filtering in the frequency domain is achieved with much less computational effort via the FFT and a suitable low-pass *prototype filter*. Interested readers are referred to any of the detailed papers describing the development and properties of PR systems [Vaidyan87b] [Vaidyan89] [Vaidyan90] [Vaidyan93] [Akansu92].

2.5.3 Polyphase Implementations

In the polyphase network, the columns of an FFT matrix modulate the (down-sampled) columns of the prototype filter into their appropriate spectral bands without the need (nor the propagation delay) of a lengthy time-domain analysis filter bank. The FFT is an efficient and delay-less filter bank in its own right. The leakage between channels can be reduced arbitrarily by increasing the order of the prototype filter. Though earlier implementations employed the FFT [Belanger76] [Rothweiler83], several *cosine-modulated filter banks* [Koilpillai91] [Ramstad91] [Lin95] with outputs that are purely real have become very popular, particularly with the advent of high-speed methods for designing cosine filter banks with the PR property [Rossi96].

The polyphase network also provides a very elegant means for effecting conversion between *Time-Division Multiplexed* (TDM) and *Frequency-Division Multiplexed* (FDM) signals known as *Transmultiplexers*. Interested readers are recommended to any of several introductory works on transmultiplexers [Narasimha79] [Freeny80] [Copeland82] as they fall outside the scope of this research.

The polyphase method, in summary, offers several important advantages:

- the analysis filters operate at a reduced sampling rate;
 - the FFT as a simultaneous filter bank and modulator is very efficient and delayless;
 - the rolloff characteristics of the prototype filter can be specified and modified easily;
 - the method allows perfect reconstruction
-

2.5.4 Sub-Band Adaptive Filters

PR systems promise substantial computational savings in applications such as speech coding. The benefits of PR systems in sub-band adaptive filtering applications, however, are less clear. The purpose of the filter, after all, is to effect cancellation of some portion of the primary input. Hence, by definition, perfect reconstruction is impossible, as the aliasing terms will be modified by the subtraction process. However, PR techniques are still appealing in echo cancellers as there are periods when the far-end signal is silent and the adaptation mechanism is suspended. It is highly desirable to pass the primary signal unaltered during these periods.

A typical sub-band filter is shown in figure 2.3. The reference and input signals are split into multiple (four, in this case) independent channels by the analysis filter banks. Cancellation is achieved in each with a conventional adaptive filter whose error signal is the residual energy in the corresponding channel. The sub-band errors are fed to the synthesis filter bank which strips out-of-band energy from each channel. Finally, the filtered outputs are re-combined through a simple summation to form a single wide-band output. It should be noted that the bandwidths of the analysis filters in the reference path are normally designed to be slightly wider than those in the primary path. This is to ensure that the eigenmodes at the band edges of each channel are not attenuated inadvertently by the filter bank itself [DeLeon95].

It is convenient to view the sub-band filter as a collection of N performance surfaces existing in parallel. The eigenvalues of each corresponding autocorrelation matrix are very widely spaced (deliberately so, as a result of the analysis filter bank) allowing the less-important eigenmodes to be ignored altogether¹ with only very minor consequences. The eigenvalue spread is reduced by down-sampling. Each filter is adapted independently of the others, so it is possible to specify the parameters of each to suit their inputs. Specifically, the power normalization methods discussed earlier can be employed in each filter so as to equalize spectral convergence across the entire input spectrum.

1. Within the given channel: these modes are modelled elsewhere in the parallel adaptive filter bank.

FIGURE 2.3: A sub-band adaptive filter with four channels. For simplicity, no sample rate conversion is shown in this example.

The benefits of sub-band filtering are not gained without cost and due consideration, however. For example:

- the bandwidth of each filter bank must be specified, and the optimal characteristics for each are not immediately obvious;
- the filter banks in the primary and error paths insert a finite propagation delay which may be intolerable for certain real-time applications like AEC;
- there is a significant computational cost associated with each filter bank that is related to the quality of the filters. This cost may be reduced substantially through sample rate conversions.

Each of these topics have been the concern of many researchers recently, as sub-band filtering techniques have proven to be quite fruitful in many speech processing applications.

Sub-band echo cancellers were first presented in the late 1980's [Gilloire87] [Gilloire88] and have continued to be of interest [Etter93] [Yasukawa93] [Schoenle93] [DeLeon95] [Jin96]. They have also been applied to general adaptive filtering problems [Kwan93] [Usevitch96], including those with polyphase implementations [Farhang96].

Detailed performance analyses of various sub-band structures have been undertaken [Gilloire92] [Petraglia93] with a simplified description of sub-band performance surface geometry provided in [Yang93]. Recent work in sub-band structures describes the use of time-varying analysis filter banks [Higa95] [Wada95] [Sodagar95] and filter banks with unequal bandwidths [Princen95]. The non-uniform methods are reminiscent of earlier sub-band structures [Crochiere76], and they provide a natural segue to the more recent wavelet transform methods [Vetterli92] [Ramchan96] [Guillemain96].

2.5.5 Delayless Versions

Some flat delay is unavoidable in all of the sub-band filters discussed so far because the primary signal is convolved with an analysis filter (the prototype filter in the case of polyphase networks) and the sub-band errors are convolved with the synthesis filters. Several means to reduce or eliminate this delay have been proposed, including the use of a parallel conventional MA-NLMS filter to model the lower-order impulse response [Chen88a] [Chen88b] [Dorbecker95] [Chen96] and the use of a routing switch in the primary path which bypasses the filter banks and disables cancellation during double-talk or when the reference signal power is very low [Amano91].

A pair of elegant and truly delayless sub-band filters were described in [Morgan95b], the first being a filtered- x structure intended for active noise control systems where any propagation delay is very detrimental, and the second suggested for use in acoustic echo control applications. The structure of the AEC version is shown in figure 2.4.

FIGURE 2.4: A four-band realization of Morgan's delayless sub-band filter.

In this structure, *the error signal is decomposed into sub-bands*, not the primary, so the correlations are made between the reference and the error signals directly. Hence this structure is a true closed-loop negative feedback system that will tune itself automatically to counteract any errors it introduces. Thus, there is no need for the bandwidths of the reference analysis filter bank to be extended artificially, because the inter-band noise will be emphasized in the sub-band errors and cancelled in the LMS filters. Round-off and scaling errors inherent in fixed-point processors should also be reduced by means of the same principle.

It is important to note that the error outputs are not propagated in the feedback path of this filter, as is the case in almost all other adaptive filters. Instead, the sub-band adaptive filter *coefficients* are transferred through the medium of the FFT to synthesize a single wide-band adaptive filter. The cancellation signal is generated from the linear (time-domain) convolution of the wide-band adaptive filter with the reference input, as is the case in conventional time-domain structures.

2.6 Summary

In this chapter, the core members of the time-domain adaptive filtering family were introduced. Widrow's LMS algorithm was derived, as it has emerged as the backbone of the more popular methods of steepest descent in applications requiring higher-order filters. It will be seen shortly that the popularity of the LMS algorithm extends readily into the transform domain.

The LMS algorithm is known to be impractical when the reference input PSD is non-uniform or time-varying. Two simple normalization schemes, MA-NLMS and AR-NLMS, which vary the step-size to account for changes in the short-term power of the reference signal, were outlined. Subsequently, several ad-hoc algorithms whose step-size (scalar or vector) varies according to the proximity of $\mathbf{w}_k$ to $\mathbf{w}_{opt}$ were also described.

It should be remembered that the tap-weights in time-domain LMS algorithms do not necessarily correspond to the eigenvectors of $\mathbf{R}_{xx}$. It would be incorrect, therefore, to conclude that the use of an N th order step-size vector in a time-domain filter automatically provides the processing benefits implied by N additional degrees of freedom, because the step-sizes are not independent of one another. This may explain why the tracking behaviour of the diagonal matrix algorithm was not found to be appreciably better than that of the scalar time-varying step-size algorithms.

The ES-NLMS algorithm was also discussed. This algorithm uses a step-size vector too, though in this case the step-sizes are tailored to suit the statistics of variations observed in the unknown plant, rather than to the relative error in a particular filter coefficient. The step-size vector can be changed arbitrarily to suit any application where the time-localized variance of the unknown plant is known reasonably well in advance.

Finally, sub-band structures in general voice processing applications including adaptive filtering were discussed, with particular emphasis allotted to a delayless closed-loop topology. As they are not strictly time-domain structures, sub-band filters set the stage for the orthogonal transform domain methods which will be the topic of the following and subsequent chapters.

3

Transform Domain Adaptive Filters

There are several time-domain tools which can be used to evaluate some superficial statistical properties of evolving signals from one time segment to the next, (peak amplitude and weighted energy measurements, zero-crossing frequency, short-term autocorrelation analysis, and so on), but these provide meagre grounds for the cultivation of concise adaptive filters. For example, these methods do not reveal the largest eigenvalue of R_{xx} , so one cannot use them alone to select a step-size that will guarantee the convergence of Widrow's LMS algorithm¹. A set of tools which decompose the reference signals into more manageable sub-signals will be introduced in this chapter. The benefits of their use in adaptive filters will be revealed thereafter.

A *Transform* is a mathematical tool which provides an alternate representation of a given input sequence. It reveals information about the signal which is not readily apparent in its natural state. For example, a prism is a physical *analysis transform* which splays white light into a continuous colour spectrum, revealing characteristics of the input which are not easily identified otherwise. Filtering can be achieved very easily in this *transform domain* by attenuating or phase-shifting energy over a certain range of frequencies. A second prism can be used to *synthesize* or *inverse-transform* the segregated band energies into a unified time-domain output signal. A piano is an example of a mechanical inverse-transform which synthesizes spatially-distributed mechanical energy into an acoustic time series.

Transforms can also be employed to simplify difficult or lengthy calculations. For example, the product of two numbers can be found using logarithms, where the logarithm and antilogarithm

1. The problem is readily overcome by using normalized LMS algorithms, some of which were described in chapter 2. However, the step-size in NLMS algorithms is varied as a function of reference signal power, and this too is statistical information unavailable to the conventional LMS algorithm.

operators act as forward and inverse transforms respectively, and multiplication in the “linear domain” is analogous to addition in the “log domain”; the Laplace Transform is an operator which converts a continuous-time integral or derivative problem into an algebraic one. In short, if one is prepared to make a computational investment by transforming a signal, then significant savings and filtering opportunities might be realized by exploiting the time-frequency properties revealed by the transform. Generally speaking, these investments are unprofitable for small N .

3.1 Overview

In this chapter, the structure and operation of a number of *Transform-Domain Adaptive Filters* (XDAPF) will be introduced. The term “Transform Domain” has been used quite liberally in the literature, even though the underlying structures may differ considerably from one other. To avoid confusion, all transform domain filters mentioned herein will be assigned to one of two classes:

- a *Block-Oriented Transform-Domain Adaptive Filter* (BOX-AF) buffers the incoming data streams and processes them at a reduced rate, thereby improving computational efficiency at the expense of a flat delay in throughput;
- a *Sample-Oriented Transform-Domain Adaptive Filter* (SOX-AF) operates on a sample-by-sample basis and is therefore free of flat delay, but it requires greater computational effort per output sample than its block-based cousin¹.

Henceforth, the term “Transform Domain” will be used only when reference is made to general adaptive filtering structures which employ a transform operator, whether they operate on a block- or sample-oriented basis. An understanding of discrete transform principles as applied to signal processing will be assumed hereafter. Readers seeking an introduction to the topic are referred to any of the standard texts on DSP [Proakis88] [Oppenheim89] and particularly to the excellent books by Brigham [Brigham88] and Rao and Yip [Rao90]. A short review of the key properties and implications of signal processing in the transform domain is provided in appendix A.1.

1. These structures are also referred to as *Self-Orthogonalizing Transform-Domain Adaptive Filters*, which can be tagged (conveniently) with the same acronym.

3.2 Block-Oriented Transform Domain Adaptive Filters

Block-processing algorithms were developed initially as a means to reduce complexity in higher-order adaptive filtering systems. There are two mechanisms through which these savings are realized. First, the transforms and adaptive filter updates are carried out only once per N iterations, where N is the number of elements in the adaptive filter. Second, the DFT of a purely real sequence can be calculated in $(N/2)\log_2(N)$ operations using the *Fast Fourier Transform* (FFT) [Cooley65]. The potential for large computational savings is apparent, especially for large N . These early incarnations provided no performance advantages over conventional algorithms, and were hampered by the inherent processing delay associated with the latency of the input buffers. In later implementations, the orthogonal properties of the DFT were employed to improve the convergence and tracking behaviour of the filter, and the recent multi-delay versions allow the user to control the amount of flat delay - an important consideration in real-time systems.

BOX-AF algorithms swap the time-domain convolution problem for one of frequency-domain multiplication via the convolution property of the DFT¹. Special consideration must be given to the buffering of input signals in these algorithms, however, as the multiplication of discrete-time spectra yields *circular convolutions* in time. There are two segmentation procedures described in the literature which can be used to reduce or eliminate the time-domain aliases introduced by unconstrained circular convolutions. These are called the *Overlap-Add* (OLA) and *Overlap-Save* (OLS) methods [Oppenheim89].

OLS is much more popular in BOX-AF structures because it is simpler and it often requires less memory [Clark83]. It also allows the (nearly) independent streams of the transform block output to be processed independently and in parallel [Shynk92]. The major difficulty with OLA is that the filter output computation is not linear in the strict sense, as the adaptive filter coefficients are updated between reference buffer updates [Clark83]. Strictly speaking, linearity holds only when the filter elements are time-invariant, which is not the case in general adaptive filtering situations.

1. Other transforms appear to be unsuitable for block processing because their convolution properties are considerably more complex. DFT-based filtering algorithms are often called "Block Frequency Domain Methods", where the *frequency domain* is a special complex space of N dimensions that is accessed strictly via the DFT/FFT.

3.2.1 Chronological Development

The first transform domain adaptive filters were designed to reduce the computational burden of systems modelling lengthy impulse responses [Dentino78]. A block diagram of Dentino's filter structure is shown in figure 3.1, where the transformations and all operations enclosed by them are undertaken once every N samples, with N being the length of the buffer and the order of the adaptive filter. A single tap weight is associated with each channel, and the tap-weight vector is adjusted using the *complex LMS algorithm* [Widrow75]. The error signals used to update the tap weights are channel-specific, and in this sense, Dentino's algorithm is a precursor to the sub-band structures discussed in section 2.5. The buffered error signals are transmitted sequentially to the output at the input sampling rate.

FIGURE 3.1: Dentino's frequency-domain LMS filter for $N=4$. The DFT's and the operations enclosed by the DFT blocks are undertaken at $1/N$ times the system sampling rate. The output exhibits artifacts of circular convolution.

Dentino's algorithm requires three length- N FFT's and one adaptive filter update per N samples. The MSE is dependent on the step-size and input signal-to-noise ratio [Bershad79]. There are no constraints specified, and thus one would expect to encounter *wrap-around noise* due to circular convolution when the output sequence is reconstructed.

Concurrent analysis of wrap-around effects was underway at Carleton University [Pelkowitz81]. The work was concerned only with frequency-domain equivalents of linear time-invariant filters in the time-domain, and thus the results are not applicable to adaptive filters generally, though perhaps to those which have converged to a steady state. The sectioning procedure used was OLS, though analytical extensions were made to cover OLA.

By this time, the first BOX-AF whose output was completely free of wraparound effects had been reported [Ferrara80]. Ferrara showed that the optimal (Weiner) solution could be obtained via fast convolution at the same rate as in time-domain LMS, but with much lower computational burden for $N > 64$. This was achieved with a constraint on the gradient [Shynk92] though OLS sectioning. The method employs 5 length- $2N$ FFT's and the tap-weight update is calculated in the frequency domain. This represents a major departure from Dentino's algorithm, where the adaptive filter was updated in the time domain [Clark81] and only three length- N FFT's were required. A sketch of Ferrara's structure is provided on page 21 of [Shynk92].

An unconstrained *Fast Convolution Algorithm* was reported in [Mansour82]. The claim is rather misleading, however, because the constraint was merely transferred to the reference input from the gradient [Clark83]. It retained the configuration proposed in [Dentino78] but extended the length of the three FFT's to $2N$. This was the first BOX-AF to exploit fully the orthogonality of the DFT by adjusting the step-size in each bin independently. A similar power normalization technique was developed very differently in [Ogue83], which described a power-normalized version of Dentino's algorithm.

Mansour's OLS-based algorithm achieves lower minimum MSE than was attainable in previous unconstrained implementations, and much faster convergence than LMS for highly coloured inputs [Mansour82]. It was shown analytically and experimentally that with wraparound effects removed, the structure converges to the DFT of the Weiner solution, and it thus invited a detailed analysis of the relationships between adaptive filtering algorithms in the time- and frequency domains. Such studies were undertaken in [Clark83], where several variants of frequency domain adaptive filters were analyzed. These structures were shown to produce identical adaptive filter solutions provided wraparound effects were nullified. The study confirmed earlier findings that the time-domain Block LMS (BLMS) algorithm could be implemented directly in the frequency

domain, and that it produced the same results as the existing BOX-AF algorithms [Clark81]. The analysis was extended in [Feuer85] [Lee89] to show how the filter order affects the convergence rate and misadjustment. Further, the minimum MSE was shown to be a function of the square of the delay in the plant [Reed81], such that Dentino's algorithm is rendered useless unless the data in the primary input buffer is nearly perfectly aligned with that in the reference signal buffer.

A link between Dentino's structure and the LS family of algorithms was described in [Hertz86]. The method was formulated from the unconstrained algorithm of [Mansour82], but employed a weight-update mechanism similar to that of [Ogue83]. The method is difficult to implement in practice because it requires vector divisions in the weight update equation. A simpler normalized block frequency domain algorithm was studied in [Bershad86] which used the MA rectangular windows of section 2.4.1.1 to estimate the power in each channel each iteration. An equivalent AR version of this algorithm was analyzed in [Florian88]. Hereafter the former will be referred to as the MA-BOX-AF and the latter as the AR-BOX-AF. Equal misadjustment and convergence rates can be achieved in the MA- and AR-BOX-AF's by setting $\beta = (N-1)/(N+1)$ [Florian88].

Frequency-weighting in the BOX-AF was proposed in [Lee85a]. This technique was motivated by advances in perceptual speech coding, where the frequency discrimination capability of the human auditory system and orthogonal properties of the transforms are key. The algorithm was designed to increase cancellation in the most audible frequency bands at the cost of increased noise in frequency bands which are of less importance. Lee's structure was erected on Mansour's "unconstrained" BOX-AF. It required two additional length- N FFT's to weight the error vector without introducing wrap-around effects. Multirate processing in the BOX-AF was proposed in [Lee86a] in an effort to reduce the computational load of the adaptive filter further. The effects of leaked energy between adjacent channels in the down-sampled signals was not discussed.

The earliest report investigating BOX-AF's applied to AEC appears to have been published in [Asharif86] and repeated in [Asharif87]. Here, a *Frequency Bin Adaptive Filter* (FBAF) was proposed to address the problem of undermodelling lengthy unknown plants, such as acoustic impulse responses. The FBAF partitions the reference input into blocks which are much shorter than the duration of the unknown plant. Undermodelling is addressed by feeding the outputs of each FFT channel into a conventional LMS adaptive filter.

The adaptive filter portion of the FBAF is manifested in matrix form, not unlike the sub-band structures discussed in section 2.5.2. However, due to the block processing in the FBAF, the delay associated with each element in this matrix tap-delay line equals the number of samples between successive computations of the FFT. It is recommended in [Asharif86] that a block overlap of at least 25% ($N/4$) produces the best tradeoff between performance and computational efficiency. When compared to the well-known BOX-AF algorithms discussed earlier, the FBAF requires less computational power and it imposes a much lower (and controllable) flat delay. Simulations provided in [Asharif86] are inconclusive, as the unknown plant was a synthetic FIR impulse response that is not truly representative of an acoustic transfer function. Furthermore, no comparison between the FBAF and a full-length BOX-AF was provided. It is also interesting to note that OLA rather than OLS sectioning was employed in Asharif's FBAF.

Application of the standard AR-BOX-AF to AEC appeared first in [Sommen87]. That structure included the use of a particular "cosine window" which was intended to exaggerate the influence of early reflections in the echo path. A derivation for a generalized window function to eliminate wraparound effects was provided, and a brief look at the influence of β on convergence rates and MSE was undertaken.

The review now turns to a set of transform domain algorithms which, though complicated, place control of the amount of flat delay, frequency resolution, and processing schedules in the hands of the designer.

3.3 Multi-Delay Implementations with Partial Overlap

All BOX-AF's which re-use exactly N samples of the data in the input buffers impose a flat delay of N samples in throughput. Delay as low as 10 msec has been identified as objectionable in AEC applications [VanSchyn96]. At a sampling rate of 8.000 kHz (typical in toll-quality speech coding), this translates to a maximum tolerable delay of 80 samples, well-short of the adaptive filter orders required to model known conference telephone set responses, let alone typical conference room responses. Until recently, flat delay has precluded the use of BOX-AF algorithms in AEC. In addition to the flat delay problem, conventional BOX-AF's tend to over-resolve the frequency

content of lengthy adaptive filters modeling very long impulse responses are being identified. The notion of over-resolution is best understood through a review of speech generation mechanisms (see, for example, [Makhoul75], [Schaefer75], and [Rabiner78] amongst others).

The human vocal system is often modelled as a time-varying ARMA filter whose excitation is switched between broadband noise (for *fricatives*) and an impulse train (for *voiced sounds*). This filter (comprised by the vocal tract, tongue, mouth, etc.) does not have a Q-factor large enough to produce orthogonal signals in very narrow frequency bins, certainly not relative to the Q's which can be obtained with (even modest) digital filters [VanSchyn96]. At a sampling frequency of 8 kHz, the effective power bandwidth for speech in telephony lies between about 100 Hz and 3.4 kHz, and a typical room impulse response of 500 msec can be represented with about 4000 taps. A BOX-AF of this order will be quite impractical for several reasons:

- it will impose an (intolerable) flat delay of 500 msec on the system throughput;
- its frequency response will extend beyond the telephony band (about 2 Hz to almost 4 kHz);
- a fixed-point transform will produce larger rounding errors for larger transforms;
- data streams are spaced by 2 Hz - much less than required in normal speech applications.

As a result, significant computing power is wasted. A better choice for the buffer size might be (for example) $N=64$, because:

- the flat delay will be limited to 64 samples (8.0 msec), which should not be objectionable;
- the response will not exceed appreciably the limits of human hearing and speech production;
- the output channels will be spaced by about 125 Hz, which is not unreasonable for speech;
- the transform will be numerically efficient as it is an integer power of two.

A block size of this order is impractical for conventional BOX-AF's, but it is not unreasonable in structures such as the FBAF [Asharif86].

Though it imposes a memory management penalty, there are several reasons why this "divide and conquer" approach to transform-domain adaptive filtering is particularly appealing in certain

time-critical applications such as AEC:

- it localizes impulse response energy in time and in frequency simultaneously;
- it reduces the size of the transforms which need to be calculated;
- computed transform blocks can be delayed and re-used;
- it reduces the flat delay inherent in BOX-AF structures;
- the temporal content of the adaptive filter remains in the control of the designer;
- the algorithm is “modular”, so it is well-suited to construction in hardware.

In this section, some similar two-dimensional structures are introduced which have been shown to provide very promising results in acoustic modelling problems.

3.3.1 Chronological Development

A *Time-Frequency Domain Adaptive Filter* was conceived independently of [Asharif86] and was presented in [Xu89]. In this paper, a *double-convolution model* - where the density of the time axis is decreased as the frequency resolution (i.e. the filter length) is increased - was used to describe the interplay between the block size, the amount of overlap, and the length of the filter matrix. The utility of the structure was demonstrated using an actual room impulse response, which was not the case in [Asharif86].

A third time-frequency structure called the *Multi-Delay Block Frequency Adaptive Filter* (MDF) was developed (again independently) in [Soo90]. The improved convergence of the MDF over the BOX-AF's of [Ferrara80] and [Mansour82] was credited to the increased rate at which the filter taps are updated, made possible ultimately by the reduced block size. This work proved to be a seminal contribution to the signal processing community, as it was the prototype for a very promising class of XDAF's known as the *Generalized Multi-Delay Adaptive Filter* (GMDF), which employed a *Weighted Overlap Add* (WOLA) reconstruction method instead of the usual OLS technique, also used in [Soo90]. The term “generalized” refers to the (variable) amount of overlap between adjacent transform blocks. The parameter α , which specifies the amount of overlap¹, allows trade-offs to be made between flat delay, computational efficiency, and memory.

The GMDF was first described in [Prado94], using both the FFT and the *Fast Hartley Transform* (FHT). A concise mathematical description of the GMDF was provided in [Moulines95]. Several hardware and software implementations designed specifically for AEC were suggested in [Martin94] and [Baganne95], and a detailed comparative analysis between numerous time- and frequency-domain algorithms in an AEC application was presented in [Naylor94]. Under almost all circumstances tested, the GMDF was found to outperform the other algorithms.

Recent experiments at Carleton University with the GMDF [Lariviere97] have shown that better cancellation can be achieved through judicious recombination of the channel outputs during the reconstruction phase. A low-order MA-NLMS filter was incorporated in an attempt to eliminate the propagation delay which would be imposed by the first block convolution, with mixed results. This technique was also discussed in [Gardner95].

3.4 Sample-Oriented Transform Domain Adaptive Filters

All of the transform domain filtering methods presented thus far insert some undesirable finite flat delay in the near-end path. Fortunately, the benefits of transform domain processing can be won without any flat day at all. Delayless transform-domain adaptive filtering can be achieved if one is prepared to transform subsequent blocks of data at *every* iteration. The operation of such an algorithm will be described shortly. They were developed initially to improve the tracking behaviour in cases where the PSD of the reference input signal is highly coloured and where rapid convergence is required. Their computational demands will always be greater than those of NLMS structures of the same order because the transform must be calculated every iteration.

1. The GMDF is also referred to in the literature as the $\text{GMDF}\alpha$, and also as the *Sliced Frequency Domain LMS* (SFLMS) algorithm [Amrane92].

3.4.1 Development from Time-Domain LMS

The time-domain LMS adaptive filter of figure 2.1 is duplicated in figure 3.2 with a transform operator inserted along the tap-delay line. In this case the reference input slides along the input to the transform rather than the adaptive filter. Once the transformation is complete, the outputs are fed *in parallel* to the LMS stage. Note that the time domain LMS filter can be realized exactly with this structure if the transform is set to the identity matrix¹.

FIGURE 3.2: A first attempt in the implementation of the SOX-AF. This structure offers no convergence rate improvement over conventional LMS algorithms since the step-size in the LMS stage is scalar.

The structure of figure 3.2 will converge if due care is exercised in choosing the step-size. As-is, however, it can offer no convergence acceleration over conventional LMS algorithms because the step-size remains scalar: the filter will converge at rates proportional to the eigenvalues of the transformed autocorrelation matrix R_{uu} , which are similar to those of R_{xx} . In geometrical terms, the transform will rotate and translate the performance surface, but leave its overall shape unchanged. The geometrical properties of the transform and their influence on the behaviour of the adaptive filter is discussed in detail in section A1.5 of Appendix A1.

1. The Identity matrix cannot orthogonalize signals because its off-diagonal elements all equal zero, so it cannot suppress the cross-terms as the conventional trigonometric transforms do.

The design and motivation for power-normalized algorithms in the time domain were discussed in section 2.4.1 and in appendix A.1. Each data stream fed to the LMS stage of the SOX-AF is (ideally) orthogonal to the others, so power normalization can be undertaken on a channel-by-channel basis rather than over the signal as a whole. Thus, convergence rates can be controlled independently in each eigenmode, allowing slower modes to be accelerated without jeopardizing stability in the faster ones. Any of the power normalization methods discussed in section 2.4.1 for wide-band time-domain LMS filters are equally valid for each narrow-band channel in the SOX-AF. Ultimately, the convergence rate in each eigenmode can be optimized with minimal regard to the others, such that all of them converge at a near-uniform rate determined only by μ_g .

Figure 3.2 is re-drawn in figure 3.3 to incorporate the channel power estimation block (marked “P”). This represents the *Transform Domain Adaptive Filter* first described in [Narayan81]. Time-normalization is achieved in each channel, and it is necessary (and fortunate) consequence that normalization of the power spectrum is achieved simultaneously. The filter operates as follows:

- sequential reference observation vectors x_k are transformed to yield u_k at each iteration;
- elements of u_k are fed in parallel to the adaptive filter;
- the power in each channel is estimated using one of the methods of section 2.4.1;
- channel step-sizes are computed as the inverses of the estimated channel powers scaled by the (user-defined) global step-size parameter, μ_g ;
- the filter output y_k , the error signal e_k and the adaptive filter coefficients are calculated as in the conventional LMS algorithm.

The AR-SOX-AF algorithm is summarized in table 3.1. This approach to power normalization appears to be the most common in the literature [Lee86b] [Marshall89] [Beaufays95a] though MA-SOX-AF’s have also been investigated [Narayan81] and analyzed [Florian88]. The AR step-size estimation procedure is illustrated in figure 3.4, which is a re-print of figure 2.2 with notation agreeable for the SOX-AF. Here, the operator $diag[A]$ extracts the elements on the main diagonal of the square matrix A and places them in sequence in a length- N column vector.

FIGURE 3.3: SOX-AF with on-line channel power estimation. Each channel is treated independently of the others, though they are fed by a common scalar error signal.

FIGURE 3.4: Channel power estimation using a first-order AR process.

TABLE 3.1 Algorithm for the standard SOX-AF described in the literature [Narayan83].

Operation	Governing Equations
Initialization	$\mathbf{p}_0 = \text{diag}[\delta \mathbf{I}]$, $\delta = \text{some small positive value}$
Transform Reference Vector	$\mathbf{u}_k = T[\mathbf{x}_k]$
Estimate Channel Powers	$\mathbf{p}_k = \beta \mathbf{p}_{k-1} + (1-\beta) \text{diag}[\mathbf{u}_k \mathbf{u}_k^T]$, $0 < \beta \leq 1$
Compute Step-size Vector	$\mu_k(n) = \mu_g \mathbf{p}_k^{-1}(n)$, $n = 0, 1, 2, \dots, N-1$
Generate AF Output	$y_k = \mathbf{u}_k^T \mathbf{w}_k$
Compute Error Signal	$e_k = d_k - y_k$
Update Filter Coefficients	$\mathbf{w}_{k+1}(n) = \mathbf{w}_k(n) + \mu_k(n) e_k \mathbf{u}_k(n)$, $n = 0, 1, 2, \dots, N-1$

The operational theory of the SOX-AF can now be contrasted with that of the time-domain LMS algorithm. The following points should be borne in mind throughout the remainder of the thesis as they are central to the operation of the SOX-AF.

Recall that in time-domain LMS, the input signal slides along the adaptive filter input such that the overall sequence is processed *in series*. Thus:

- each tap in the time-domain LMS filter is an estimate of the actual plant transfer function *in the corresponding time window of duration $1/f_s$ seconds* over the entire input spectrum;
- all taps in the time domain structures are exposed to the same signal energy over long periods of time, so one step-size is applicable to the filter as a whole¹;
- $\mathbf{R}_{xx}$ has the Toeplitz property.

1. Alternatively, one could view the step-size as a vector under the stipulation that the values of each of the N elements are equal to one another, and are bounded by $2/N\lambda_{\max}$ of $\mathbf{R}_{xx}$.

In the SOX-AF, the adaptive filter input is received *in parallel* from the transform stage, with the following implications:

- each tap weight in the SOX-AF is an estimate of the frequency response of the plant transfer function *in the corresponding band of width $f_g 2N$ Hz* (for the DCT) over the duration of the unknown plant;
- each tap in the SOX-AF is exposed to different (orthogonal) sequences, so it is appropriate to specify a length- N step-size vector whose elements are chosen to suit the spectral distribution of the incoming signals. Specifically, the step-size for the n th channel can be set (up to its maximum of $2/\lambda_n$) independently of the others.
- R_{uu} is *not* Toeplitz (unless the reference PSD happens to be flat). However, its off-diagonal elements are near to zero due to the elimination of cross-terms by the transform, regardless of the PSD of the reference input.

In short, one can view the time-domain filter as a single filter in N dimensions, and the SOX-AF as a bank of N independent filters operating in parallel, each of unit dimension. It should be made clear that *the SOX-AF is not a subband filter* because a wideband error is used in the tap adjustment mechanism, not a filtered error applicable only to one channel as was the case in the subband structures discussed in section 2.5.2.

3.4.2 Chronological Development of Fixed-Transform Structures

Narayan's pioneering work [Narayan81] employed only the DFT with MA power estimation. In a subsequent paper [Narayan83], other trigonometric transforms were investigated and shown to provide satisfactory results. The DCT was upheld as a particularly promising candidate, as it was already employed extensively in speech coders. A brief analysis showed that the DCT allowed substantially faster convergence rates than the DFT for various samples of speech. Whereas Narayan's algorithm appears to be very different from the well-established BOX-AF's, it was shown later to be a special case of the BOX-AF with a block length of one [Lee84]. It can also be considered to be a simplified *Frequency Sampling Filter* (FSF) which gained wide attention around the same time [Bitmead81] [Copeland82] [Shynk86].

A detailed convergence and misadjustment analysis of the SOX-AF was undertaken in [Lee86b] for the DFT, DCT, *Discrete Sine Transform* (DST), the *Symmetric Cosine Transform* (SCT) and a *Fast Karhunen-Loeve Transform* (FKLT). In all cases, the AR power estimator was used. It should be noted that the time-domain filters in [Lee86b] were AR-NLMS algorithms as opposed to the MA-NLMS methods used in this research. The effect of varying β was discussed briefly in [Lee86b], though no analysis to support its influence on the initial convergence rate was provided. The suggestion was made that a β between 0.75 and 0.9 would provide good results for a wide variety of adaptive filtering applications. Larger values of β were shown to increase the MSE overshoot in the early stages of adaptation when the initial power estimates were small.

The performance surface geometry of the SOX-AF was first illuminated in [Marshall89]. Some analysis and simple experiments proved that no optimal fixed transform exists for the SOX-AF: certain transforms perform better than others depending on the PSD of the excitation. It should be noted that all of the simulations provided in [Marshall89] appear to have been obtained from single trials applied to fixed plants of order 8.

A recursive DFT pre-processor was proposed in [Ogunfunmi90] to reduce the computational burden associated with higher-order adaptive filters. The recursive DFT is commonly called the *LMS Spectrum Analyzer*, and its properties are now well-known [Widrow87] [Beaufays95b] [Beaufays95c]. Other transforms can be computed using LMS recursion, (see for example [Liu88a] [Liu88b] [Liu93] [Ogunfunmi90] for the DHT, and [Beaufays95b] [Beaufays95c] [Chau94] [Chau95] for DCT implementations). Recursive transforms can reduce computational complexity considerably, however they add another adaptation stage to the SOX-AF. Time did not permit these structures to be examined here. Interested readers are referred to [Beaufays95b] for some practical comparisons of the various recursive DFT and DCT methods.

Two new performance indices for the SOX-AF were introduced in [Farhang91]. Before then, the *condition number* of R_{xx} (the ratio of its largest to smallest eigenvalues) was employed almost exclusively as the yardstick by which convergence rates between algorithms and reference signals could be compared. However, the transformation of the reference signal reveals other statistical information which inspired two quick methods for estimating the amount misadjustment and the rate of overall convergence in the SOX-AF.

From [Farhang91], *Performance Index 1* is

$$P(\mathbf{R}) = \frac{1}{N\lambda_{min}} \sum_n \lambda_n \quad (3.1)$$

and *Performance Index 2* is

$$\rho(\mathbf{R}) = \frac{\lambda_a}{\lambda_g} \quad (3.2)$$

where

$$\lambda_a = \frac{1}{N} \sum_n \lambda_n \quad (3.3)$$

is the arithmetic mean and

$$\lambda_g = \sqrt[N]{\prod_n \lambda_n} \quad (3.4)$$

is the geometric mean of the eigenvalues λ_n of the autocorrelation matrix¹. The larger these two indices are, the greater the variance of the eigenvalues, and hence the less even the distribution of the reference power spectrum. It is shown that

$$P(\mathbf{R}) = MT_{tr} \quad (3.5)$$

where M is the steady-state misadjustment of the algorithm, and

$$T_{tr} = 4\tau_{max} = \frac{1}{\mu\lambda_{min}} \quad (3.6)$$

is the transient time required for the slowest mode of the adaptive filter. Equation 3.6 indicates that the transient time and misadjustment are inversely proportional to one another, and that the ratio of the proportion is determined through equation 3.1. Equation 3.2 is more useful, however, because it can be expressed in closed form in terms of the elements of $\mathbf{R}$ (see section 3 of

1. The statement applies to $\mathbf{R}$ generally because the formulae are valid in different domains.

[Farhang91], which is not the case for $P(\mathbf{R})$. These metrics are appealing because they consider all the eigenvalues of $\mathbf{R}$ instead of just the largest and smallest ones, as was the case when the condition number was used.

Equation 3.2 for $\rho(\mathbf{R})$ was shown to be a measure the distance between any LMS-type algorithm and the optimal transform-domain algorithm obtained via the KLT [Farhang92]. In particular, the “Improvement Factor”

$$I_p = \ln[\rho \text{diag}[\mathbf{R}_{uu}]] \quad (3.7)$$

is a measure of the amount of acceleration in overall convergence due to the transformation and normalization of an LMS algorithm as opposed to standard LMS in the time domain. It is known that I_p can never be negative, so *the transformation/normalization operations will never degrade the overall convergence rate relative to LMS.*

Practical experiments provided in [Farhang92] showed that the performance indices can be used to predict the behaviour of different transforms when excited by reference signals with different spectral distributions. They suggest that the DCT is an excellent choice for a fixed-transform in cases where the reference signals are expected to be predominantly low-pass, such as speech.

The analytical performance limits of DFT- and DCT-based SOX-AF algorithms for low-pass Markov-1 signals were derived in [Beaufays95]. Table 3.2 summarizes the theoretical limits of the condition number of $\mathbf{R}$ for the Identity, DFT and DCT transforms, where $\rho \in [0, 1]$ is the autocorrelation coefficient of the Markov-1 reference input.

TABLE 3.2 Condition numbers for the SOX-AF with Markov-1 inputs.

Transform	Condition Number of $\mathbf{R}$
Identity	$(1 + \rho)^2 / (1 - \rho)^2$
DFT	$(1 + \rho) / (1 - \rho)$
DCT	$1 + \rho$

The condition numbers of the autocorrelation matrices for each of these algorithms increases as a function of ρ , however the condition numbers of the DFT and particularly the DCT grow much slower than that of the LMS algorithm (i.e. the identity transformation) as ρ approaches one. These results assume the *asymptotic equivalence* [Gray72] of R_{uu} and a diagonal matrix - perfect decorrelation in other words - an assumption which weakens as the order of the filter decreases from infinity. As ρ approaches unity, the excitation becomes more correlated, and slower overall convergence is expected, regardless of which transform is employed. A DC component (i.e. where $\rho=1$) in a time-domain filter input is known to be problematic [Shoval95], and it is proven here to be troublesome for the SOX-AF also.

Practical implementation issues of several recursive DFT and DCT algorithms were addressed in [Beaufays95b] [Beaufays95c], particularly the influence of round-off errors due to fixed-point processing. Open-loop *sliding-window* DFT/DCT algorithms were shown to diverge after some time because the round-off errors tend to accumulate. The LMS Spectrum Analyzer however, captures roundoff errors within its feedback loop, so they are minimized automatically once the adaptive transform coefficients have reached their steady states.

The plant tracking behaviour of the SOX-AF was first investigated in [Farhang96]. The authors conclude that orthogonalization and normalization do not yield any significant improvement in steady-state MSE over standard LMS. However,

- Neither the time- nor frequency-domain characteristics of the unknown plant at the time of start-up were provided;
- All the elements of the unknown plant vector were made to change with time according to a multivariate Markov-1 process, the details of which were not specified.

Whereas these assumptions about plant non-stationarities are not unusual in the literature, (see for example [Widrow76], [Farden81], [Widrow84], [Ewada85]) they are atypical of those one would expect to see in AEC. More detail about the nature of the plants anticipated in AEC is provided in chapter 4. Thus it would be imprudent to apply these conclusions blindly to AEC.

Early experiments conducted during this research yielded results that were quite contrary to those of [Farhang96]. For example, it will be shown in chapter 6 that a properly-tuned SOX-AF will provide substantially better plant tracking capability than the optimal MA-NLMS algorithm for stationary signals with non-uniform PSD, and for non-stationary stochastic signals generally. These results were consistent in both simulated and actual recorded environments. Accordingly, the remainder of this chapter and the entire following chapter are devoted to setting up the framework in which the tracking behaviour of the SOX-AF can be examined critically.

3.4.3 Two-Layered Interpretation

The SOX-AF features two layers of adaptation¹ - one to estimate the filter's channel step-sizes and a second to adjust the filter coefficients. These sub-systems have entirely different objectives, and it seems appropriate to decouple them to the greatest extent possible such that each may be examined independently and tuned for optimal performance.

To that end, the two-layer description originally presented in [Beaufays95b] is adopted hereafter. The stages are delineated in figure 3.5.

The responsibilities of the signal identification stage are to:

- extract a length- N sequence from the reference signal using a rectangular window;
- transform the windowed vector into another domain, the DCT in this case;
- estimate the short-term power in each channel of the transform output;
- determine an appropriate step-size for each channel based on the channel power estimate;
- feed the transformed reference signal coefficients to the Plant Identification stage;
- limit the channel step-sizes during times of particularly low channel energy.

1. A third would be constituted by a recursive transform such as the LMS Spectrum Analyzer, which was not considered in this study.

FIGURE 3.5: Two-Layer representation of the SOX-AF. The first stage tracks the input signal, the second tracks the unknown plant.

The responsibilities of the plant identification stage are to:

- implement an array of N single-tap LMS filters with a common wideband error signal;
- scale the step-sizes provided by the signal identification stage by a factor of μ_g ;
- suppress energy outside bands of interest (i.e. beyond voice band in telephony).

These ideas are formalized mathematically in Appendix A1.

3.5 Summary

Two types of LMS transform domain adaptive filters were introduced in chapter 3, and a detailed literature survey was provided for each. The first requires the reference signal to be sectioned into blocks, so the processing is undertaken on a block-by-block basis. These methods introduce a flat delay into the transmission path, and though they are computationally efficient, they are generally unsuitable for modeling lengthy impulse responses in real-time applications. Two hybrid time-transform domain algorithms - the FBAF and the GMDF α - have been developed to address this problem. The latter one has been found to be particularly effective.

The second family of transform domain filters - the subject of this research - processes the input sequence on a sample-by-sample basis. The method requires greater computational effort, but it is delayless, and no sectioning procedures are required to suppress circular convolution aliasing. Transform domain methods convert the 1-dimensional N th order adaptive filtering problem into an N -dimensional first-order problem. The advantage in doing this is that each channel can be treated independently of the others, allowing the convergence modes of the adaptive filter to be normalized.

The normalization is achieved by adjusting the step-size of each channel to suit the power in that channel. Thus, the SOX-AF has a signal identification stage, the purpose of which is to estimate the recent power in each channel of the transform, and to scale the step-sizes accordingly. It also has a plant identification stage, which uses the transformed data and the corresponding step-size estimates to update the filter coefficients.

The notion of a two-stage SOX-AF is central to the format of the remainder of the thesis. Efforts will be made in the following chapters to divorce the operations of the two stages to the greatest extent possible, as their purposes are entirely different. Once each stage has been made to work well on its own, the two will be re-married into a unified orthogonal adaptive filter.

4

Application to Acoustic Echo Cancellers

The proliferation of books and technical papers (even entire conferences) pertaining to speech analysis, indoor acoustics, and adaptive DSP attests to the complexity of these fields. The AEC problem is a particularly challenging one because it exists where these studies intersect.

In chapters two and three, the merits of time-domain and transform-domain adaptive filters were compared, and the virtues of the SOX-AF in particular were extolled. In this section, ways and means to marry the SOX-AF with the AEC problem are considered. Attempts will be made to identify those avenues most likely to lead to robust and practical algorithms, and then to wring them thoroughly in order to squeeze as much acoustic reverberation as possible from the primary signals detected in simulated and actual environments.

First, the nature of the AEC apparatus and its typical operating environment is reviewed. The excitation (stochastic processes generally, and speech in particular), the set hardware, and those acoustic properties likely to be encountered in typical conference rooms will be discussed. The section concludes with an examination of the mechanisms available to control the SOX-AF. The intent is to qualify the influence of these parameters on the rate and depth of cancellation, and to develop some general rules for their selection so as to determine the suitability of the published SOX-AF algorithm to AEC, and to establish practical baselines for purposes of comparison.

It will be shown that the published SOX-AF algorithms are inherently sub-optimal for AEC, regardless of what values are assigned to the control parameters. Some structural modifications will be proposed to address these limitations, none of which appear to have been presented in the literature previously. Those thought to be most promising were tested thoroughly in simulated and actual environments. The results are presented in chapters 6 and 7. The others are left for future researchers to explore.

Throughout this and subsequent chapters, the AR-SOX-AF and the MA-NLMS algorithm will act as baselines against which the new algorithms will be compared. These were chosen because they appear to be the most common in the literature and both are well understood. In most cases, analogies to other implementations can be inferred easily, and clarifications will be made where required.

4.1 Some General Properties of the AEC Environment

Previous work in AEC has shown that almost all structural modifications result in performance trade-offs. For example, lower-order filters may converge more rapidly than higher-order ones, but they cannot be expected to model longer impulse responses as well; orthogonal algorithms like the SOX-AF may converge faster than time-domain filters of the same order, but they require more computational effort; larger step-sizes yield faster convergence at the expense of increased misadjustment; the examples are plenty.

Other researchers have achieved performance improvements by incorporating assumptions about the operating environment into their algorithms, thereby giving their systems a head-start in the system identification problem. A good example is the ES-NLMS filter of section 2.4.3.2. One would not expect it to perform so well in general system identification problems, because the assumptions which led to its improvements may be counter-productive in other applications.

4.1.1 Excitation

The reference signals of greatest interest in acoustic echo cancellation problems are derived from human speech. Over intervals of a few minutes, one expects to see the spectral distribution, peak and instantaneous power, signal-to-noise ratio, etc. vary widely in a given speech sequence. Further, these parameters can be expected to change considerably from language to language, individual to individual, and, in the context of AEC, from set to set. An adaptive filter designed to account for all of these unknowns under the blanket of a single set of static parameters (as in standard LMS) will be very unlikely to succeed. Even so, there are features common to almost all speech signals, particularly when they are received from telephone networks.

Some of these include:

- *Finite Bandwidth*: Few people generate meaningful vocal energy below 100 Hz [Schafer75]. Also, an anti-alias filter is required in discrete-time systems, rendering frequencies higher than 3400 Hz out-of-band for standard 64 kbps systems. It seems reasonable to focus as much effort as possible to cancelling spectral energy within the 100-3400 Hz channels. It might even be prudent to neglect the out-of-band channels altogether [Beaufays95b].
 - *Spectral Distribution*: The PSD of time-averaged speech is often assumed to be a band-pass signal with most energy concentrated in the lower quarter of the pass-band spectrum. The time-averaged PSD decays generally, though not monotonically, with frequency [Schafer75].
 - *Low Q*: Many models of the human vocal tract have been proposed [Schafer75] [Rabiner78] to represent typical voice spectra over long periods of time. AR models of fairly low order excited by white noise are not uncommon [Mansour82], [Narayan83]. As a continuous-time system, the vocal tract cannot be expected to provide particularly sharp filter characteristics. Modern AEC algorithms operate mostly in discrete-time, where very sharp filter responses can be implemented with relatively little effort. Thus, one need not over-design the frequency resolution of the adaptive filter, since the energy of voiced speech will not vary drastically between adjacent frequency channels of such narrow bandwidth.
 - *Time-varying Power*: Speech time series evolve as concatenated bursts of energy, so the power levels in each frequency channel will vary greatly, *even when the reference input is WSS*. This is because the observation windows used in AEC are relatively short, and from one time segment to the next the PSD cannot be expected to remain constant.
 - *Power Thresholds*: Provisions must be made to suspend the plant tracking mechanism when the reference power is low, otherwise the algorithm will tend to converge to noise. The adaptation must also be suspended during periods of double-talk. Accordingly, the thresholds which activate and de-activate the adaptation mechanism must be specified.
 - *Human Factors*: The human auditory system does not respond equally to all frequencies of stimulation. Human factors will not be considered in this research, though future researchers may want to investigate the utility of transform domain algorithms which, (unlike their time-domain counterparts) make the distinct frequency bands available for manipulation.
-

4.1.2 Acoustic Impulse Responses

The total acoustic impulse response of the unknown plant consists broadly of two overlapping portions: that which is due to the set alone, and that which is due to the room and its contents. In practice, it is impossible to segregate the two completely, because both exhibit (many) poles, implying that their impulse responses decay in time but never disappear entirely. Room acoustic impulse responses are typically wideband, with power envelopes that are often modelled by a sum of exponential functions which decay in time at various rates [Widrow85] (pages 105-109). Examples of some measured impulse responses and their spectra are provided in appendix A2.

Most of the set echo is represented in the lower-order terms of the impulse response because the set has usually a small enclosure with the microphone located close to the loudspeaker. Thus, most of the total echo response power is due to the set itself, rather than to the room in which it is installed. Tightly regulated manufacturing processes improve the likelihood that the response of one set will be substantially the same as those of all others of equal construction. Suggestions to set w_0 to the measured anechoic response [Goubran96] seem well-founded.

The room response, on the other hand, can be expected to appear in all parts of the total impulse response *except the lowest order terms* because some minimal time interval is required for the sound to propagate from the loudspeaker to the reflective features of the room and back to the microphone. The room response will be unique to the room in which the set is located, and thus it cannot be known in advance. The higher-order terms are also much more difficult to model properly for three important reasons:

- *their contribution to the overall echo power is small compared to that of the set* [Sommen87];
 - *they are more likely to change as the room characteristics change*. Seemingly minor changes in the room such as an opening door or a moving person can produce significant differences in the higher-order room transfer function [Goubran96].
 - *they tend to be less correlated with the reference signal than those of lower order*. Acoustic wavefronts cannot be constrained to propagate along highly directional paths. Rather they are disseminated in the open acoustic channel and on each encounter with a reflective surface, producing the phenomenon known as the *barrel effect*.
-

A key element of this research concerns the ability of an adaptive filter to track changes in the unknown plant. Previous researchers have proposed to distribute the total energy held in the step-size along the (time-domain) adaptive filter in proportion to the variance measured in the corresponding tap weights as the room transfer function changes [Makino93]. These findings appear to be much more useful to time domain rather than transform domain adaptive filters, because localization in time implies dissemination in frequency.

4.1.3 Algorithm Behaviour

Because the SOX-AF can be interpreted as a bank of N single-tap filters (see section 3.3.1), any of the power estimation methods discussed in section 2.4.1 for time-domain filters are applicable to each channel of the SOX-AF. The estimators most commonly used in the literature are MA [Narayan81] [Narayan83] [Bershad86] and AR [Mansour82] [Clark83] [Lee86a] [Sommen87] [Florian88] [Lee89] [Chao90] [Beaufays95a]. The latter has gained prevalence, particularly in recent research, so the AR estimator will be used in all linear SOX-AF algorithms discussed hereafter unless specified otherwise.

The SOX-AF was developed in chapter 3 as a two-stage tandem structure, the purpose of the first stage being to track the PSD of the reference excitation, and the purpose of the second being to update the filter coefficients using Widrow's LMS algorithm with N independent step-sizes. One would anticipate that the performance of the second stage is dependent to some extent on that of the first, and indeed this relationship will be demonstrated in chapter 6. The behaviour of the LMS filter is well-known already, so attention will be directed to the signal identification stage for the time being.

Once the filter order and orthogonal transform have been selected, only the step-size estimation parameter, β remains to be chosen. Before an optimal choice of β can be made, however, some understanding of the signals upon which it will act must be gained.

Using the model of the transform as a filter bank, the $u(n)$ outputs are narrow-band signals, which become more sinusoidal as N increases. After transformation by DCT (for example), one

would expect most of the energy in any $u(n)$ to be concentrated in a bin of bandwidth $f_s/2N$ Hz centred at $(n-1/2)f_s/2N$ Hz, $n=1, 2, 3, \dots N$.

The instantaneous power in the n th channel at iteration k is defined as

$$u_k^2(n) \equiv \|u_k(n)\|^2 \equiv \text{diag}[u_k u_k^T] \quad (4.1)$$

The squaring operation has two significant effects:

- it is a *rectifier*, so the positive *and* negative peaks in $u(n)$ correspond to positive peaks in $u^2(n)$. Thus one would expect the energy concentration in the PSD of $u^2(n)$ to be shifted to a digital frequency twice that observed in the PSD of $u(n)$.
- it *increases the variance*, so the range of $u^2(n)$ is much greater than that of $u(n)$ in general. This exemplifies the need for an effective means to estimate the average output power from each channel of the transform block.

To illustrate these points, some time-series examples of $u(n)$ and $u^2(n)$ are provided in figure 4.1. Here, WSS broadband excitation was used, the filter order was $N=256$ and the transform selected was the DCT. In each case, $u(n)$ is scaled to have a modest peak amplitude so as to keep $u^2(n)$ readily visible on the same plot.

It is apparent from figures 4.1 that the transform output signals $u(n)$ are narrow band, and that the centre frequency of $u(n)$ increases with n . This demonstrates the utility of the DCT as a filter bank. The power signals $u^2(n)$ are strictly positive, their maxima corresponding to peaks in $u(n)$, and their minima corresponding to the zero-crossings of $u(n)$ as expected. Note, however, that the highest frequencies in $u^2(n)$ occur at $n=N/2$, and as n increases above $N/2$ the apparent centre frequency of $u^2(n)$ *decreases* with n . This is how the frequency shifting property of the squaring operation manifests itself in the time domain.

The PSD plots corresponding to the time series in figure 4.1 are shown in figure 4.2. In each case, the power band moves linearly with n for $u(n)$. It moves linearly with $2n$ for $u^2(n)$ until $n=N/2$, where it is reflected back into the base frequency band.

FIGURE 4.1: Comparison of DCT output signals and output power time series. Raw channel outputs (dashed) and instantaneous channel output powers (solid) are shown for the DCT with $N=256$. From the top, $n=32, 96, 128, 192$, and 240 . In each case, 100 consecutive samples are shown.

FIGURE 4.2: PSD plots of the time- and power series of figure 4.1. In each case, the PSD of the raw signal (dashed) is shown with the PSD of the power sequence.

It is more instructive to demonstrate this property in the transform domain through example, as follows:

A reference signal of length $M=1000$ was fed to a DCT-based AR-SOX-AF with $N=128$. At each iteration $k > N$, the instantaneous transform outputs were written to the $(k-N)$ th rows of an $(M-N) \times N$ matrix. The instantaneous output powers were written sequentially to a similar matrix. The PSD of each column was computed once the matrices were full. The process was repeated for 100 independent trials, and the ensemble averages of the resulting matrices appear in figure 4.3. The PSD's of the outputs $u(n)$ are shown in figures 4.3 (a) and (c), and the PSD's of the power sequences $u^2(n)$ appear in figures 4.3 (b) and (d) for the broadband and AR2(-1.5955, .95) inputs respectively. When comparing the two, it is apparent that the peaks in the PSD's of each $u^2(n)$ channel are shifted in frequency (across the page) by a factor of two with respect to those of the corresponding PSD's of $u(n)$. The spectral energy of $u^2(n)$ is reflected back into the base-band for channels $n > N/2 + 1$ as expected.

Figures 4.3 are interpreted as follows: The digital frequency is read across the page, 0π (DC) at the left and 1π ($f_s/2$) at the right. The DCT channel number n reads in and out of the page, with the lower-frequency channels at the rear, and the higher frequency channels to the fore.

The plots show that the power estimation filter must be low-pass to accommodate the persistent DC component in the $u^2(n)$ sequences. Note that the graphs in figures 4.3 (a) and (b) resemble R_{uu} (see appendix A1) - in fact, they are short-term estimates of R_{uu} that have been compressed logarithmically. In due course, these properties will be used to customize algorithms for AEC.

FIGURE 4.3: Frequency content of the transformed outputs and power sequences. In each case, the DC to high-frequency channels are ordered from rear to front. The PSD for each channel n reads across the page. The output signal PSD's are shown in (a) and (b) for broadband and AR2(-1.5955, .95) reference inputs respectively. Corresponding PSD's for the channel power outputs are shown below in (c) and (d).

4.2 Specification of Available Parameters

The SOX-AF provides several “handles” through which its performance can be crafted. This implies the structure is highly flexible, but it also increases the potential for instability, as the parameter set must be chosen carefully so as to ensure that the two stages of the system work agreeably with one another. Here, attention is directed towards these parameters in an effort to alleviate pitfalls in the published SOX-AF algorithms before they are encountered in practice.

4.2.1 Number of Taps - N

In many practical systems, advance knowledge of the time-power distribution of the unknown impulse response is unavailable. Worse, impulse responses of a few thousand taps are not uncommon in acoustic modelling problems [Sommen87] [Knappe92] [Makino93] [Morgan95a]. In most cases, it is impractical to implement filters of this size, so lower-order filters are often used with the knowledge that there will be an increase in misadjustment due to the contribution of signal energy from the unmodelled taps.

This compromise is not normally problematic because most acoustic impulse response energy appears in the lower-order terms, as outlined in section 4.1.2. Further, there are complications which arise in the adaptive filter itself when implementing higher-order filters. These are:

- The eigenvalue disparity in R_{xx} will probably increase, which in turn increases the likelihood that some will take on more extreme values. Thus, some (admittedly more narrow-band) eigenmodes may converge slower than would be the case had they been incorporated with their neighbours under the blanket of a single eigenmode in a lower-order filter.
 - The error in the gradient estimate will increase, adding uncorrelated noise to the output. This is important for decaying impulse responses such as those of AEC because the gradient noise is distributed evenly over the impulse response whereas the echo is bound to fade with time. Thus, the ratio of echo power to gradient noise will increase with filter order.
 - The overall step-size must be reduced to account for the increased power in the observation vector. This reduction occurs naturally in power-normalized LMS algorithms.
-

In summary, one should have sound reasons for employing higher-order filters when modelling acoustic systems for telephony, regardless of the processing domain chosen. The improvements one would anticipate with filter orders in the thousands are marginal at best in many cases. An effort should be made, however, to encompass the the set's anechoic impulse response within the adaptive filter wherever possible.

4.2.2 Global Step-Size - μ_g

In the conventional LMS algorithm, the choice of μ is bounded by the largest eigenvalue in R_{xx} . Typically, a designer would choose μ smaller than the maximum to reduce misadjustment and the likelihood of instability. Setting $\mu_g=1$ yields the fastest convergence rate for the MA-NLMS algorithm (see section 6.3.1) even though that structure is stable for $0<\mu_g<2$.

The step-size-adjusted inputs to the LMS stage of the SOX-AF are normalized over frequency (and, by extension, over time), assuming the signal identification stage has performed its duty properly. The wideband error e_k is fed to each coefficient, so μ_g must be reduced to $1/N$ to apportion the error associated with each (normalized) filter channel. Thus, $1/N$ seems to be a reasonable choice for μ_g in the SOX-AF. This choice has been found to be acceptable in most cases, however, it has been found to be too high for certain combinations of reference input PSD and β .

4.2.3 Channel Power Estimation Parameter - β

In the early stages of this research, the choice of β was found to bear significant influence on the convergence rate, misadjustment, and tracking capabilities of the AR-SOX-AF. Even so, the specification of an optimal β has received very little attention in the literature. The purpose of this section is to examine the influence of β in the AR-SOX-AF algorithm in the context of AEC, and to present some guidelines for selecting it in general applications.

4.2.3.1 Stationary Inputs

An explicit analysis of the dependence of w_k on β is not trivial, nor is the analytical framework for the choice of β simple (see for example [Sommen87], [Florian89]). This point is illustrated

in table 4.1 where the ancestry of the output sequence e_k from β is unveiled for the n th channel of the AR-SOX-AF.

TABLE 4.1 Tracing the influence of β on MSE performance in the AR-SOX-AF.

Parameter	Governing Equation	Independent Variable
e_{k+1}	$e_{k+1} = d_k - \mathbf{u}_k^T \mathbf{w}_k$	$\mathbf{w}_k$
$\mathbf{w}_k(n)$	$\mathbf{w}_k(n+1) = \mathbf{w}_k(n) + 2\mu_g \mu_k(n) e_k \mathbf{u}_k(n)$	$\mu_k(n)$
$\mu_k(n)$	$\mu_k(n) = p_k^{-1}(n)$	$p_k(n)$
$p_k(n)$	$p_k(n) = \beta p_k(n-1) + (1-\beta) \mathbf{u}_k^2(n)$	β
β	Selection criteria?	Possibly?

The tap weights of an LMS filter in the time domain are calculated recursively [Duhamel93], and the scalar step-size parameter μ specifies the rate at which each tap weight can change. In the SOX-AF, μ_k is an N -dimensional vector whose elements are adjusted at rates which vary from one channel to the next, depending on the channel power distribution. Furthermore, the step-sizes in the AR-SOX-AF are computed as the reciprocals of the channel power estimates $p_k(n)$, which in turn are computed recursively with parameter β in the signal identification stage. Thus, in order to determine the influence of β on overall performance of the AR-SOX-AF, one must consider the treatment of a non-linear inversion sandwiched between two linear recursions.

A small β will dismiss past power estimates quickly and yield step-sizes which are more representative of the instantaneous channel powers. This should accelerate the tracking of changes in the input spectrum. However, a small β will increase the variance in the step-size estimates, which may cause undue variance in $\mathbf{w}_k$. A larger β will reduce the step-size variances because the power will be averaged over longer periods of time. The penalty paid for smoother step-size estimates is sluggishness in response to changes in the input signal power distribution. This too is problematic, as the channel power estimates can vary considerably over time, even in the steady state for a WSS input process. This is easily demonstrated:

An AR-SOX-AF was constructed with $N=128$. A reference sequence was created with 1024 samples of a pure sinusoid with digital frequency of about 0.1π , pre-pended with 128 zeros. All AR-SOX-AF vectors were initialized to values very close to zero. Several trials were run with all parameters unchanged except β . The transform output channel with the greatest power ($n=14$ of 128) was examined, and the results are displayed in figure 4.4.

FIGURE 4.4: Power estimation characteristics in one bin for sinusoidal excitation. In (a), the transient part of the instantaneous power of the channel output is shown. Power estimates are computed using $\beta=1-b/N$ with $b=2.0$ (fastest), 1.0, 0.5, and 0.25 (slowest) are shown in (b), $b=4$ is shown in (c) and $b=8$ is shown in (d).

In figure 4.4 (a), the transient part of $u(14)$ is shown. Note the output is erratic for $128 < k < 256$ while the initial zeros are flushed out of the observation window and are replaced by those of the sinusoid. Once the sinusoid fills the entire window, $u(14)$ is locked into an oscillatory steady state. In figure 4.4 (b), (c), and (d), the running power estimates for this input with $b=0.25, 0.5, 1.0, 2.0, 4.0$, and 8.0 are shown, where β is computed as

$$\beta = 1 - \frac{b}{N}, \quad 0 < b < N. \quad (4.2)$$

The following observations can be made, and these results are true in general:

- Lower values of b produce higher values of β that grow asymptotically towards one as β approaches zero;
- Higher values of b produce lower values of β that accelerate the transient response and increase the variance of the step-size estimates in the steady state;
- If $b=0$ then $\beta=1$, giving the AR power estimator infinite memory so that it is never updated;
- If $b=N$ then $\beta=0$, giving the power estimator zero memory, in which case $p(n)$ equals $u^2(n)$;
- The mean value of the power estimate is the same in the steady state, regardless of the value of β [Florian88] provided $0 < \beta < 1$.

4.2.3.2 Step Non-Stationary Inputs

The ability of the AR-SOX-AF to track abrupt increases in the reference signal power will now be discussed. Recall that the derivation of the LMS algorithm (and hence its descendants, such as the SOX-AF) is predicated entirely on the assumption that the reference signal is WSS, in which case the upper bound on μ is prescribed by the maximum eigenvalue of R . However, R is defined as an expected value *over all time*, so in non-stationary environments it is unreliable for estimating μ_k . These facts spawned the normalization methods described in section 2.4.1, where μ_k was made proportional to some estimate of the inverse-square of the *recent* average power.

Regardless of which time-averaging method is chosen, however, a *linear step-size estimate lags behind the instantaneous power*. The amount of that lag is dependent on N (for MA versions) or β (for AR versions). The filter may diverge if any of the $p(n)$ lag the $u(n)$ after a sharp increase in reference signal power has been applied, because the $\mu(n)$ will be too large for the incoming signal. Stability will be regained when all the $p(n)$ increase to levels sufficient to bring all $\mu(n)$ within the bounds of equation 2.34. This is demonstrated in figure 4.5.

FIGURE 4.5: Broadband tracking capability of AR-SOX-AF as a function of β . An 8-fold step-increase in reference amplitude was applied at $k=5000$. The power evolution for each channel is shown in (a); the average power in the transform is shown in (b) for $b=4$ (leftmost), 3, 2, and 1 (rightmost) with $\beta=1-b/N$; In (c), the instantaneous power in the primary is compared to the average input power with $b=4$, showing the lag of the power estimate (and hence the step-size) behind the primary for the fastest case; In (d), the filter divergence is shown for $b=1$ (deepest), 2, 3, and 4 (shallowest).

This problem is aggravated when most of the impulse response power is concentrated in the lower order plant, (which is exactly the case in AEC), because such a distribution accentuates the differences between d_k and y_k during the transient phase, inflating the error signal e_k , thereby pushing w_k further from w_{opt} at a time when the step-sizes are already too large.

4.2.3.3 Towards a Rule for β Selection

Several suggestions to select β have been made by previous researchers:

- it should be chosen close to, but less than unity [Lee86b];
- it should be made as large as possible [Sommen87];
- it should be a function of the (time-varying) step-size [Youn83] [Cowan85] (p. 174).

In most cases, these conclusions are applicable only for certain classes of inputs. For example, if the reference input is truly WSS, β may be set artificially close to unity once the μ_k have converged to their steady-state values because the PSD of such inputs will remain nearly constant for all time. Of course, if the reference input statistics change, the step-size estimates will be very slow to adjust, rendering this choice of β sub-optimal, or even detrimental. The application of interest here demands an ability to track highly non-stationary inputs such as speech, so a rule for prescribing a more generalized β is sought.

It is clear from table 4.1 that β determines the *rate* at which the step-sizes are updated, not the step-size values directly. Further, the AR-SOX-AF as discussed so far employs a scalar β that is applicable to all of the channelized power sequences emanating from the transform. It is also clear from the discussion in this chapter that:

- *the inputs to the power estimator are strictly-positive pulsed-DC sequences;*
 - *the pulse frequency is a function of the transform selected and the channel index n ;*
 - *the pulse power envelope in each channel varies slowly with respect to the pulse frequency.*
-

From these facts flow the following guidelines:

- *A low-pass characteristic is required for the power estimator.*

Every channel must pass DC with unity gain, otherwise the power estimates will be biased. Further, the power *envelope* of the input is of interest to a linear estimator, not so much the instantaneous pulse power, so β (the pole of the estimation filter) must reside on the *positive* real axis in the z -domain.

- *The frequency response of the power estimator should encompass the passband of the lowest channel's output.*

The transform splits the reference signal into N channels of equal bandwidth. The longest wavelength which can be represented entirely within the observation vector has a length of N samples for the DFT¹. The estimator is already known to be a low-pass filter, so the location of its pole should be chosen so as to encompass this frequency in its pass-band.

- *The higher frequency bins are of less significance than the lower ones.*

There are two reasons why this is so: first, the energy in the higher frequencies is attenuated somewhat due to the roll-off of the anti-aliasing filter; second, fixed transforms (particularly the DCT) have greater difficulty orthogonalizing the higher frequency channels.

It was felt that the choice of β should be directed by the *lower* frequency content of the reference signal. From previous work, it is known that the estimator integrates the energy in each channel. An integrator is a low-pass filter, so the concern is in specifying the upper bound of the (low-pass) frequency response of this filter.

Accordingly, a fixed β of

$$\beta = 1 - \frac{1}{N} \quad (4.3)$$

will be used in all AR-SOX-AF structures hereafter.

1. Since the basis functions of the DCT are spaced at intervals of $1/2N$, a DCT-based SOX-AF has twice the frequency resolution of the DFT. Accordingly, the DCT allocates a channel which can represent a frequency whose period is $2N$ samples.

This choice for β is appealing for several reasons:

- it is “adaptive” in the sense that it varies automatically with the (user-specified) filter order;
- it is very near to $N-1/N+1$ for all reasonable values of N , which is the value of β that yields power estimates with the same step-size variance as the N th order MA-SOX-AF algorithm [Florian88].
- it represents the reduction in tap-weight variance expected due to the power normalization in the BOX-AF [Bershad86];
- it is reminiscent of the *forgetting factor* employed in most RLS algorithms.

In the remainder of this chapter, some suggestions will be made to improve the performance of the SOX-AF, and a description of those changes which warranted implementation will be given in detail.

4.3 Some Alternate Structures

This section provides a summary of some structural modifications to the existing SOX-AF that were conceived during the course of this research. The changes proposed in this section affect both stages in the filter.

4.3.1 Non-Rectangular Reference Windows

So far, the observation vector of the reference signal has been weighted uniformly before being fed to the transform. Whereas rectangular windows are employed in almost all of the transform-domain filter literature, it is not inconceivable that other window functions may provide better results in certain applications. A non-uniform window would allow certain time-periods of the input to be emphasized and others to be de-emphasized in the transformed representation.

Many researchers have reported that most of the impulse response power in AEC lies in the lower order plant [Sommen87] [Makino90] [Knappe92] [Vanderkooy94] [Bradley96]. Further, the variance in the elements of w_k due to room non-stationarities has been identified as having a

decaying exponential envelope with the filter order [Makino93]. These observations led to the ES-NLMS algorithm given in section 2.4.3.2. The step-sizes in all the transform-domain filters investigated here are derived from the statistics of the reference input alone, and are independent of the unknown plant. It would seem that altering the step-size adjustment mechanism based on statistical variances of the plant may be counterproductive in a transform-domain filter because localization in the time domain implies dissemination in the transform domain.

A Cosine window $g(n)$ was proposed in [Sommen87] for use with the BOX-AF when the impulse response of the unknown plant is known to decay in time. It was shown to reduce computational complexity (it allows the unconstrained BOX-AF to be realized using three N -point FFT's rather than five) with virtually no performance penalty.

4.3.2 Re-Combination in the Transform Domain

It will be shown in section 6.5 that a longer time window will allow the unknown plant to be modelled more accurately. However, as argued in section 3.2, increasing the frequency resolution beyond that warranted by the reference input is also counterproductive. It seems that the two outcomes are diametrically opposed to one another, as the number of frequency channels in the transform domain filter is the same as the order of the observation window. Here, a method to gain the benefits of both options is suggested. The proposed method works as follows:

Let N be chosen so as to model the desired portion of the unknown plant, even though this will over-resolve the transform spectrum. Compute the transform of x_k as before, but decrease the frequency-domain resolution (the order of the adaptive filter) by re-combining adjacent outputs. Estimate the step-sizes in each *grouped* channel rather than on each individual channel from the transform block, reducing memory and computational requirements simultaneously. Clearly, an immediate problem is how to define the re-combination boundaries. Here are some possibilities:

- *Uniform*: Make all the re-combined output channels of the transform have equal bandwidth. This is the simplest method, and it is the only one mentioned here from which a semblance of the Weiner solution can be obtained.
-

-
- *Constant Q* : Combine adjacent transform outputs into non-overlapping groups whose ratio of bandwidth to centre frequency is nearly constant. This method is not dissimilar from the wavelet-based methods discussed elsewhere in the literature [Vetterli92] [Vetterli95] [Giullemain96] [Ramachan96] .
 - *Constant Power*: Use the PSD of the reference signal to define filter channels with nearly equal power. This will make the step-size vector in the adaptive filter more-or-less uniform on the average.
 - *Perception-based*: Combine the transform outputs so as to maintain the resolution of those frequencies which are of greatest significance to the human auditory system [Lee84], while de-emphasizing those that are of little significance.

Neither of these structural alterations are investigated further in this research. Preliminary trials yielded filters that in most cases converged, but whose performance was clearly substandard to that of the known AR-SOX-AF algorithm. It was decided to concentrate on modifications to the signal identification stage alone, as much research into the significance of this stage still needed to be undertaken.

4.4 Proposed Linear Pre-Processor Modifications

Whereas previous researchers have alluded to the significance of β in the signal identification stage, an exhaustive literature search has uncovered little research into any means to optimize the selection of β for generalized N . When this work was started, this section was intended to introduce the core contributions to the field. It summarizes several original ideas to improve the performance of the signal identification stage through the manipulation of β . Those methods which held the most promise are developed intuitively and mathematically and implemented in chapters 6 and 7.

4.4.1 Dynamic Stepsize Initialization

Since μ_k is calculated as the inverse of p_k , some non-zero p_0 must be assigned on start-up of the AR-SOX-AF algorithm. The choice is significant because values of β close to unity (which is

not uncommon in the literature) will cause p_0 to persist for a long time. A large p_0 will prescribe a small μ_0 which will limit the convergence rate unnecessarily if the reference input power is low. On the other hand, a small p_0 will set μ_0 high, and may cause the filter to be unstable until step-size estimates based on measured reference samples are available. Higher β values tend to be desirable in the SOX-AF because they reduce the variance of the channel step-sizes in time¹.

Here, a *Dynamic Step-size Initialization* (DSI) process is introduced. It dismisses initialization artifacts quickly without compromising stepsize estimation in the steady state. In practice, the method has been found to provide a secondary benefit by addressing the overshoot-on-start-up problem described in [Lee86b], [Florian88], and apparent (though not identified explicitly) in [Haykin96] (pages 471-477).

Let p_0 be set to its maximum value so as to guarantee stability on start-up. Let β be some small positive value to allow the stepsize vector to track the reference signal very quickly. For each iteration $k < N$, one zero is shifted out of the observation vector and a valid reference sample is shifted in to it. Hence, the transform output becomes more representative of the reference signal as the initial conditions are flushed. Thus it seems appropriate to increase β from some low β_0 to its steady-state value of $1 - 1/N$ at $k = N$. DSI is conducted as shown in the flowchart of figure 4.6.

For $k > N$, the influence of p_0 has effectively disappeared, so the algorithm operates as described previously. DSI prescribes a “gearshifting” [Snelgrove97] of β , and it ensures the weighting of the step-size estimates in time is related to the knowledge of the reference signal available. It also ensures a smooth and stable transition from initialized estimated step-size values. The DSI method can be tailored to suit the MA-SOX-AF by setting the length of the power normalization window equal to the lesser of k and N for all k .

DSI was found to be very effective at accelerating initial convergence rates and eliminating overshoot. The experimental results supporting this claim are provided in section 6.4. Unless otherwise noted, DSI will be undertaken in all SOX-AF simulations presented hereafter.

1. This problem was termed “step-size noise” in [Lee86b], but no detailed analysis nor justification of the phenomenon was provided. Small β makes the reference appear progressively more non-stationary, as it reduces the period over which R_{uu} is estimated.

FIGURE 4.6: Flowchart representing Dynamic Step-Size Initialization.

4.4.2 Expansion of β

The major advantages offered by the SOX-AF over its time domain counterparts are afforded by virtue of its orthogonal pre-processor, whose mathematical properties have promised dramatic improvements in convergence rates and tracking capability for narrow-band signals. Even so, it appears that full advantage of the orthogonalization stage has not been realized in the published algorithms.

If an algorithm encourages different step-sizes for different channels, it seems quite reasonable that a different step-size estimator should be used in each channel also. In other words, β could

be expanded into an N -dimensional vector so as to prescribe a different time constant $\beta(n)$ for each channel. Under this scheme, the step-sizes in each bin are estimated over the same number of pulses of the channel power signals rather than over the same period of time, and hence the method allows the modes of the excitation PSD to be tracked in an equitable manner, regardless of which channel is considered. Though no extra computational effort would be required, some additional memory space would be needed to store the vector (as opposed to the scalar) β .

Numerous experiments were undertaken in the earlier stages of this research to determine what advantage, if any, could be leveraged with a fully orthogonal step-size estimator, as above. The investigation was suspended, however, because the results were mixed and they appeared to be signal-dependent. A deeper investigation into the spectral character of each channel revealed that this method may be counterproductive in general. The reasons why became apparent with the help of figure 4.3 (d), and similar plots for other non-uniform distributions.

It is clear that the power spectrum of each (power) channel $u^2(n)$ is not confined to a particular frequency band. This is due in part to leakage in the transform, but moreso to the frequency shifting of the power spectral energy in each channel brought about by the squaring operation. The frequency shifting causes energy to be reflected back as aliases at the plot boundaries. The aliases do not coincide in frequency with the fundamentals which caused them: the energy often re-combines in the baseband in curious and fascinating ways, as exemplified in figure 4.3 (d). Thus, it would appear unwise to assume that the power spectrum of $u^2(n)$ is narrow-band.

4.4.3 Adaptation of β

One might also question why β should be static if the process it is used to estimate is (over short intervals at least) not WSS. It was suggested in [Youn83] that it is desirable to maintain the same time constants in the signal- and plant-identification stages of the (time-domain) adaptive filter. A doubly-adaptive time-domain filter was introduced in [Youn83] for system identification. The notion of double-adaptation refers to the adjustment of μ_k with a β -controlled reference signal power estimate as in equation 2.48, followed by the adjustment of β based on the updated μ_k :

$$\beta = 1 - \mu_k . \quad (4.4)$$

This method will increase β when the reference input power is high, and therefore reduce the variance in the step-size when the greatest amount of cancellation is required. At the same time, it will reduce β when the reference input power is low, which will allow the estimation stage to be more responsive to sudden increases in power. As discussed in section 4.2.3, this is a highly desirable feature for a two-stage algorithm such as the AR-SOX-AF. This method was proposed for use in the AR-SOX-AF previously [Lee86b] (page 507), however, it was suggested only as a means to adjust the (scalar) global step-size and the (scalar) β .

In an effort to exploit the orthogonality of the AR-SOX-AF fully, a new filter structure was built which implemented double-adaptation *in each channel*. The step-sizes were estimated for each channel as before, but $\beta(n)$ (the β associated with the n th channel) was set to $1 - \mu_k(n)$ after each filter update. The filter could be stabilized for various speech inputs with careful selection of the initial conditions and switching thresholds. However, the filter could not match the AR-SOX-AF in convergence rate nor misadjustment when the preferred values of μ_g and β were selected. This is not to say that the method does not work - rather, it merely invites further investigation into the idea of a transform-domain adaptive filter whose power estimation is truly orthogonal.

By the time Youn's power estimation method had been implemented with reasonable success, it had become apparent that linear estimation methods cannot provide all of the features required by an ideal power estimator in the transform-domain adaptive filter. Linearity in this application seems to prescribe that any gains made in one respect due to the adjustment of the estimation parameter are offset by losses in other respects. Further, the linear SOX-AF estimators were found to be rather delicate when treated with any form of time adaptation. Thus, it seemed wise to seek a more suitable power estimation scheme outside the confines of linear systems.

4.5 Non-Linear Power Estimation

So far only linear step-size estimation methods have been considered in this work, and this is apparently the case in the transform-domain filter literature also. However, there are at least two reasons why a linear estimator may be inadequate:

- *A linear filter has a symmetrical response, meaning it learns and forgets at the same rate.*

A static first-order AR process such as that used to estimate the power in the AR-SOX-AF operates without knowledge of the power in $u^2(n)$, and it cannot determine whether the trend of $u^2(n)$ is static, increasing, or decreasing: it always lags its input, and it behaves in the same manner regardless of what its input is. It was shown in section 4.2.3 that the estimator should be very responsive when the channel power is increasing. When the channel power is static or decreasing, there is no such requirement.

- *A linear filter averages over time, so it will tend to under-emphasize peak measurements.*

Stability bounds for the LMS algorithm were derived in section 2.2. Choosing $\mu < 2/\lambda_{max}$ will ensure stability *on the average*, but will not guarantee stability *continually*. To guarantee stability at all instants of time, μ must be constrained to the bounds of equation 2.34 *at each iteration*. Accordingly, it would appear that the step-sizes should be derived from the peak power, rather than from the average power. Thus there is no means to ensure the $\mu_k(n)$ are within their stability bounds at every instant using the AR power estimator, regardless of what β is (are) chosen.

Forthcoming results (section 6.6) show the inability of the linear AR estimator to track abrupt changes in PSD. Notably, sudden changes from low-pass to higher-pass distributions can cause serious *divergence* of the filter until the lagging low power estimates are sufficiently flushed.

4.5.1 Median and Rank-Order Filters

Several non-linear filters have been described in the literature. An important class of these are *Rank-Order Filters*, whose outputs, as the name implies, is concerned with the sequence of the values in the input, and not just the values themselves. The *Median Filter* is a particularly simple rank order filter whose output is the median value in a window of fixed length which slides

across the input signal. Median filters have been found to be very effective in certain cases where linear filters fail miserably.

The first concise description of median filters was given in [Gallagher81]. Several broad and introductory papers discuss convergence properties [Fitch85] [Wendt86a], design [Wendt86b] [Arce87] and the application of median filters to 2-dimensional filtering problems [Wendt86a] [Wendt86b]. Median filters are very easy to implement in hardware as no numerical processing is required: the outputs are simply selected from the input. Several hardware implementations of median and other rank-order filters have been provided in the literature [Arce87] [Murthy92].

The domain of non-linear filters (and median filters in particular) carries its own set of terms and definitions. The power estimator which is about to be presented is a non-linear structure which has much in common with median filters, so a review of some salient definitions and concepts is warranted.

- A *Constant Neighbourhood* is a sequence of $M+1$ consecutive samples of equal value, where M is the order of the median filter;
- An *Edge* is a region between two constant neighbourhoods whose values are either increasing or decreasing monotonically, and include no constant neighbourhoods;
- An *Impulse* is a region of up to M samples whose values differ from those of the identically-valued constant neighbourhoods which surround it;
- A *Root Signal* is any sequence which is invariant to the application of a certain median filter.

A signal which is comprised of only constant neighbourhoods and edges is a root¹. Further, any signal which is not a root will converge to a root after a finite number of passes through the same median filter. Median filters remove impulses but preserve edges. Thus it is incorrect to view the median filter as being high-pass or low-pass. The “pass-band” of a median filter [Wendt86b] and a particular sequence consists of a root signal and, to a lesser extent, the intermediate signals generated through successive applications of the filter.

1. Provided end effects are taken into account. Precise details are provided in [Fitch85].

4.5.2 The Proposed Non-Linear Estimator

A new non-linear method of channel step-size estimation is proposed in this section. The method is very simple, and it will guarantee that the $\mu_k(n)$ are held within their stability bounds for all n and for all k , regardless of the reference input statistics. It will be called the *Peak Window Power* (PWP) estimator. In it the p_k (and hence μ_k) are derived strictly from the maxima (rather than over a time-weighted average) in a sliding rectangular window of the $u^2(n)$. The PWP estimator appears to be a certain rank-order filter, though it was conceived and tested in this work with no prior investigation into rank-order nor median filters.

The simplest form of the PWP estimator works as follows¹ for the SOX-AF: the instantaneous power vector $u^2(n)$ is calculated and written over the oldest data row of an $N \times N$ circular buffer. The PWP estimate for the n th channel, $p_k(n)$, is simply half of the largest value in the n th column of the circular buffer at time k . The channel step-size μ_k is the reciprocal of $p_k(n)$ as before. The method offers several immediate benefits:

- The step-sizes are guaranteed to be within the stability limits prescribed by LMS for all k ;
- The step-size estimate exhibits a lag of zero when the channel power is increasing;
- It is fully automatic - no arbitrary parameters such as β need to be specified;
- No initial power/step-size vector estimates are required;
- No divergence on start-up is observed;
- The step-sizes tend to be locally constant, reducing “step-size noise” [Lee86b];
- The step-sizes tend to be more conservative than with the AR estimator;
- It is applicable to other transform domain filters such as the BOX-AF and the GMDFA;
- A modified PWP algorithm (to be described shortly) can be implemented in hardware or firmware very efficiently.

1. This is a simple description of the method which illustrates its intent. As written, this method is very cumbersome for large N . A simplified version will be introduced shortly which requires much less computation and slightly more memory than the existing SOX-AF algorithms of the same order.

A schematic representation of the PWP estimator for the n th channel is provided in figure 4.7. This diagram can be substituted directly into the “P” block of the SOX-AF in figure 3.3.

FIGURE 4.7: Schematic representation of the proposed Peak Window Power estimator. The diagram shows how $\mu_k(n)$ is obtained from $u_k(n)$.

4.5.3 Mathematical Development

Assume N is very large so that the channel bandwidth is very narrow and the channel outputs are nearly sinusoidal. Consider the output of the n th channel with $n \ll N/2$ such that $u^2(n)$ is highly over-sampled and its envelope approaches that of a continuous function. For simplicity, the index n is omitted and the function is assumed to be continuous. Then:

$$u_k = A \sin(2\pi nt/N) \quad (4.5)$$

where A is the (arbitrary) peak amplitude of the pulse being considered. Thus, the instantaneous power is

$$u_k^2 = A^2 [\sin(2\pi nt/N)]^2 \quad (4.6)$$

The power in one pulse of $u^2(n)$ is sought, so

$$P_\pi \approx A^2 \int_t^{t+\pi} [\sin(2\pi nr/N)]^2 dr \quad (4.7)$$

which can be written as

$$P_\pi \approx \frac{A^2}{2} \int_t^{t+\pi} [1 - \cos(4\pi nr/N)] dr \quad (4.8)$$

where r is a dummy variable for the integration. The second term under the integral is zero because the integration is performed over an effective interval of 2π . Thus,

$$p_\pi \approx \frac{A^2}{2} r \Big|_t^{t+\pi} \quad (4.9)$$

which yields

$$p_\pi \approx \frac{\pi A^2}{2} \quad (4.10)$$

Thus the power in one half-period of $u^2(n)$ can be re-distributed into a rectangular area of length π and amplitude $A^2/2$, as shown in figure 4.8. The rectangular area represents a constant power whose reciprocal will ensure the stability of the LMS algorithm throughout the pulse duration. More importantly, *the estimated power $p_k(n)$ should never be less than half of the instantaneous peak power $u^2_k(n)$* . If it is, the step-sizes produced will be too large and the stability of the LMS stage will not be guaranteed.

FIGURE 4.8: Relationship between the actual and PWP-estimated sine power pulse. The solid curve is a (heavily over-sampled) transform output power sequence. The area under the dotted line is equivalent to the area under the solid curve. The dashed line is the PWP estimate which will ensure the stability of the LMS stage for this pulse.

4.5.4 Comparison of Linear and Non-Linear Methods

From the arguments and observations of section 4.2.3, it seems clear that a “good” estimator will be very responsive when the signal power is increasing so as to ensure stability over very short periods of time. When the power is high and relatively constant, however, the estimator output will have little variance so as to minimize noise in the weight-vector update.

Linear filters do not discriminate: their characteristics are independent of the power sequences to which they are applied. All of the SOX-AF’s proposed in the literature use static linear power estimators, so they cannot be expected to guarantee fast and stable convergence while delivering low misadjustment. The doubly-adaptive scheme described in section 4.4.3 was proposed as a means to achieve both ends simultaneously, but (in this early incarnation, at least) this method has not been proven to be robust.

The PWP estimator *does* achieve both ends. Since its output is based solely on the measured peak power, it is delayless when the power envelope is increasing. It allows the LMS stage to converge using the largest step-sizes possible which ensure stability in each channel at every iteration. Its tracking ability may be reduced somewhat when the channel power is decreasing, however, because the PWP estimate lags significantly on the trailing side. This lag may impose some limitation on the maximum convergence rate, but once convergence is achieved, the lower effective step-sizes will reduce the total misadjustment.

It is instructive to compare the PWP estimator to the linear estimators through an example. A random reference input was generated and fed to a 256-point DCT. The instantaneous power sequences from each channel were computed by squaring the DCT outputs at every sampling interval. The first 2,000 power estimates for the $n=7$ th channel are plotted in figure 4.9. Superimposed on the plot are:

- an MA power estimate, made using a length- N rectangular window;
 - an AR estimate where $\beta=1-1/N$, with DSI enabled for the first N samples;
 - the PWP estimate.
-

The MA and AR estimates are quite similar, especially as time increases. This was expected as the choices of MA and AR parameters specified yield estimates with almost identical mean and variance [Florian88]. The linear estimates differ somewhat at first because the AR estimator must flush its initial conditions in the early stages. Note that both of these estimators lag the instantaneous output power.

The PWP estimator, however, rises instantly in response to power outputs that are greater than twice the current PWP estimate. It remains constant until a higher peak power is detected, or until the current estimate is “forgotten” - i.e. over-written in the circular buffer after N sampling periods. Further, the PWP estimator usually yields larger power estimates than those produced by linear estimators, which means its step-sizes are almost always smaller. During periods where the instantaneous power is relatively stable, the three techniques produce similar estimates.

FIGURE 4.9: Comparison between three channel power estimators. A broadband sequence was fed to a 256-point DCT. The measured output power sequence (dotted) from the $n=7$ th channel is shown along with the estimates provided by the AR (dashed), MA (dash-dot), and PWP (solid) methods.

4.5.5 Practical Implementation

The N -channel PWP estimator described in section 4.5.2 requires N maxima to be extracted from the columns of an $N \times N$ matrix at each iteration. Clearly, this method is unnecessarily inefficient. Here is a general description of the PWP estimator used to obtain the results to be presented in chapters 6 and 7. This version is suitable for implementation in MATLAB, which is capable of vector and matrix processing. Practical considerations for real-time DSP implementations will be considered shortly.

The scalars $L1$ and $L2=N/L1$ are defined, where $L1$, $L2$, and N are chosen such that each is a whole number. The operation $\max[A]$ returns a $1 \times N$ vector containing the largest elements in each column of matrix A . It may take more than one argument, provided each has the same number of columns, so $\max[A;B]$ returns the largest values from the N columns of A or B .

Three data structures are defined, and the elements of each are initialized to zero:

- the $1 \times N$ Sub-Window vector f_k
- the $1 \times N$ PWP vector g_k ;
- the $L2 \times N$ PWP matrix G_k ;

The following operations are performed every iteration:

- The instantaneous power vector u^2 is calculated as before:
- $f_k = \max[f_k; (u^2_k)^T]$;
- $p_k = \max[f_k; g_k]$;

Every $L1$ iterations, three operations are carried out:

- f_k is written to the oldest data row in G_k ;
 - $g_k = \max[G_k]$;
 - the elements of f_k are all reset to zero.
-

The following MATLAB code segment will serve to clarify the procedure:

```
%*** PWP Initialization appended to other SOX-AF definitions

L1=16;
L2=N/L1;
fk=zeros(1,N);          % maximum power in last L1 raw channel power o/p's
gk=zeros(1,N);          % maximum power in last N sample intervals
Gk=zeros(L2,N);         % where last L2 sub-window PWP estimates are kept
. . .
. . .
. . .
%*** SOX-AF algorithm starts here and enters into an infinite loop
% Uk2=conj(U).*U; % new instantaneous channel powers activates PWP subroutine
. . .
. . .
. . .
%*** PWP subroutine - executed whenever new Uk2 vector is available

fk=max(fk,Uk2');          % highest Uk2 in most recent L1 iterations
if rem(count-L1-1,L1)==0  % updates Gk every L1 iterations
    Gk(2:L2,:)=Gk(1:L2-1,:); % overwrite oldest fk from Gk
    Gk(1,:)=fk;             % with current fk
    gk=max(Gk);             % find new peak channel powers in Gk
    fk=zeros(1,N);         % reset fk for next L1 time intervals
end
pk=max(fk, gk)';          % PWP is max pwr detected in last N iterations
mk=0.5./pk;              % subroutine returns stepsize column vector

% end of PWP subroutine
```

Note that $L1$ determines the “forgetting rate” of the estimator, as old values of f_k are dismissed from G_k once per $L1$ iterations. In all of the experiments conducted with this accelerated PWP algorithm, the values $L1=16$ and $L2=N/L1$ were chosen, with N an integer power of two greater than 128. The best choices for $L1$ and $L2$ remain to be determined. A good starting point might be to choose $L1$ and $L2$ as integers in the neighbourhood of $\sqrt{N}$, while keeping the product $L1*L2$ equal to or (if need be) slightly greater than N .

Once u_k^2 is available, the accelerated PWP method requires $2N$ compares each iteration¹ and $N*L2$ compares every $L1$ th iteration; no extra multiplications or additions are required. The AR estimator, on the other hand, requires $2N$ multiplications and N additions every iteration. Again, this implementation is intended for use only in higher-level languages such as MATLAB.

1. Most DSP's incorporate a “free” right shift in the multiply instruction for the divide-by-two.

The requirements of embedded stand-alone DSP systems are entirely different. When fitting the PWP estimator into such systems, care should be taken to:

- *minimize arithmetic operations.* The $N p_k$ calculations are not required at every iteration in the PWP estimator as they are in the linear estimators. Since $\|u_k(n)\|^2 = (\text{abs}[u_k(n)])^2$, it is more efficient to compare the transform output *magnitudes* (which are already available) and avoid the channel power calculation unless the step-size is to be updated.
- *minimize memory accesses.* Though the PWP algorithm demands relatively few arithmetic operations, specific implementations will likely require substantial memory to maintain and update channel power estimates and step-sizes. One should be prepared to pay close attention to proper memory management when crafting low-level PWP estimators.

It is significant that few PWP estimates require multiplications or additions. As a result, it might be possible to perform many of the operations either in external hardware or in parallel with other operations, further reducing the computational burden on the DSP. In any case, a true PWP estimator will track the $u^2(n)$ with zero lag as the channel power increases, but it will force the step-sizes to remain artificially low when the channel power is decreasing steadily. In practice, this compromise is likely to be inconsequential because the μ_k are certain to remain within their stability bounds during these periods.

4.6 Summary

Three important problems were tackled in this chapter. First, the nature of the AEC environment was reviewed. Statistical properties of anticipated reference signals and transfer functions were discussed, and it was pointed out that very little control over them is available in most cases. This brought about the second problem, which was to determine how appropriate the known transform domain algorithms are for use in AEC and how best to apply them. The third problem was identified and targeted only after it became apparent through experimentation that even the “optimized” linear SOX-AF’s impose penalties on convergence and misadjustment in a dynamic application such as AEC.

Several deviations to the overall filter architecture were proposed in section 4.3, none of which provided the promise of better test results to warrant further investigation here. The focus was narrowed to consider only the well-known two-stage AR-SOX-AF. It was felt that little would be gained in a study of the plant identification stage as the operation and capabilities of the LMS filter are very well-known in the literature already. Comparatively little research effort has been directed towards the signal identification stage, (apart from the wealth of information available concerning the selection of orthogonal transforms and how best to compute them), so the early focus of this research concerned the optimal choice of the AR persistence factor β .

Justification was provided to fix β to $1-1/N$. The method of Dynamic Step-Size Initialization (DSI) was presented which reduces the dependence of the step-size on arbitrary initial values in the early stages of convergence of the AR-SOX-AF. Efforts to expand β into a vector quantity and to make this vector adaptable to the reference excitation yielded some interesting results, but by this time the limitations of linear estimators in this application had been recognized.

A new and very simple non-linear power estimator was proposed - one which overcomes the lag inherent in linear estimators during the critical periods when the channel power is increasing. The method is called the Peak Window Power (PWP) estimator, and it was shown to guarantee stability in the LMS stage of the filter at all times, regardless of the properties of the reference input. This method shows great promise for use in real-time applications such as AEC because the estimator relies heavily on comparisons and shifts - operations which are realized readily in hardware - rather than on arithmetic operations. The linear estimators require accumulators to carry out multiplications and additions, and thus they increase the load on the DSP.

In the next chapter, the experimental techniques are described. In Chapter 6, the key algorithms (MA-NLMS, the AR-SOX-AF, and the PWP-SOX-AF) will be tested in a variety of simulated environments, each designed to illuminate a particular facet of the filter's performance. Then in chapter 7, these experiments will be repeated for a number of recordings made in legitimate acoustic environments.

5

Experimental Methods and Apparatus

This chapter summarizes the experimental context surrounding the data presented in chapters 6 and 7. A more detailed description of the input sequences and transfer functions mentioned here is provided in appendix A2.

5.1 Input Signals

Low-pass and band-pass reference signals were generated using several MATLAB *mfiles* written by the author. These programs generate and feed gaussian sequences to various AR processes. Once generated, signals were offset to zero mean so as not to confuse the adaptation algorithm [Shoval95] and then scaled to have unit variance to maintain a decent and consistent signal-to-noise ratio. Signals intended for use in chapter 7 were re-scaled to prevent saturation distortion during fixed-point file transfers. No further filtering of these sequences was carried out before they were applied as reference inputs to the simulated systems in chapter 6. All of the reference signals referred to in chapter 7, however, were up-sampled to 48 kHz and filtered in the digital and analog domains as described in section 5.4.2 to standardize the trials as much as possible.

All sequences will be referred to hereafter by the names of the *mfiles* used to generate them, as shown in table 5.1. The coefficients for the AR12 process were drawn from [Mansour82] and are intended to represent the long-term average distribution of speech. The AR2 parameter values were used in [Haykin96] (section 8.4) to illustrate the significance of the reference input PSD and the initial filter vector on the filter convergence rate. Pole locations and PSD plots of the signals are provided in appendix A2.

TABLE 5.1 Description of stochastic excitation sequences.

Name	Description
RANDN	Broadband noise with Gaussian distribution
AR1(r)	First-order AR filter of parameter r
AR2($r1, r2$)	Second -order AR filter of parameters $r1$ and $r2$
AR12	Twelfth-order AR filter with parameters listed in table 5.2

TABLE 5.2 Twelfth-order AR process parameters.

Coefficient	Value	Coefficient	Value
0	1.000000	1	-1.460340
2	1.266380	3	-0.850841
4	0.629542	5	-0.497503
6	0.273701	7	-0.168227
8	0.257914	9	-0.238396
10	0.508109	11	-0.379440
12	0.204267	13+	0.000000

Two synthesized speech tracks were also employed in these experiments. These sequences were provided by Bell-Northern Research Ltd. and are intended to sound like plausible speech bursts whose time-frequency characteristics are representative of "average" male and female speech. Each burst has a duration of approximately two seconds, and test sequences consist of numerous identical bursts concatenated together and separated by short intervals of silence. The phrase might be represented phonetically as:

pār è bō de la mō ' sàwm nō leeyong dū dàw

A male track (SSGM1) and a female track (SSGF2) were selected. The phrase and its inflection are very similar in both cases. It contains a variety of *voiced*, *fricative* (*unvoiced*), and *plosive* segments [Crochiere78], and the pitch is fairly consistent throughout the sequence.

5.2 Performance Metrics

Several unbiased metrics are employed in chapters 6 and 7 which enable algorithm performance to be compared fairly. As they are very common in the acoustic echo cancellation literature but relatively uncommon elsewhere, a detailed description of each is provided in this section.

5.2.1 Echo Return Loss Enhancement

The *Echo Return Loss Enhancement* (ERLE) is an important figure of merit for acoustic echo cancellers. It provides a measure of the cancellation improvement achieved by a given filter over that measured in a system with no cancellation at all. A “raw” definition of ERLE will be used throughout this work so as to provide a measure of performance as independent as possible of external influences. The instantaneous ERLE is defined as

$$ERLE(dB) = 10 \log \left[\frac{e_k^2}{c + d_k^2} \right] \quad (5.1)$$

and it is unitless. The constant c is chosen to be much smaller than the expected primary signal power. As the ERLE fluctuates tremendously with k , all ERLE plots shown hereafter will be smoothed by a low-pass FIR filter of order N and a very low cutoff frequency. The ERLE is used as a measure of cancellation performance in both simulated and actual trials.

5.2.2 TIP/TP Ratio

An order- N adaptive filter cannot model an unknown plant of order $N' > N$ perfectly. To wit, an unknown plant with one or more non-zero poles (as is the case in AEC) will exhibit an infinite impulse response, rendering complete echo cancellation by an FIR filter impossible. It has been shown [Knappe92] that, where an adaptive FIR filter is of lower order than a static FIR plant, the limit of cancellation is dependent on the energy distribution in the plant impulse response. The quantities which prescribe this limit are the powers in the entire impulse response (the *Total Impulse Power* (TIP)) and in the unmodelled portion (the *Tail Power*, (TP)) of the impulse response. These quantities are illustrated in figure 5.1 for a typical acoustic transfer function.

FIGURE 5.1: An FIR impulse response showing TIP and TP measurements.

The TIP/TP ratio, which specifies the maximum attenuation possible for a given filter order in an FIR system identification problem, is defined as:

$$\zeta_{TIPTP} = 10 \log \left[\frac{\sum_{i=0}^{N'-1} h_i^2}{\sum_{i=N}^{N'-1} h_i^2} \right] \quad (5.2)$$

where $\mathbf{h}$ is a static FIR transfer function vector of order N' . Some examples of TIP/TP plots for FIR approximations to measured room transfer functions are shown in appendix A2. It is important to remember that *the TIP/TP ratio is valid for broadband excitation only* where there is no flat delay present in the adaptive filter. When the TIP/TP ratio is plotted as a function of filter length for a given plant, the plot is certain to increase (improve) monotonically as the filter order N increases. A similar metric called the *Early-to-Late Sound Energy Ratio* was derived in [Marshall96]. It was intended for use in larger auditoriums with excitation such as music.

The TIP/TP ratio provides the *maximum* cancellation limit only. In practical applications, many factors conspire to limit the cancellation depth at a level below that predicted by the TIP/TP ratio. Even so, the TIP/TP ratio provides a convenient guideline for specifying the filter order. Roughly in order of most- to least influential, these factors include [Birkett97]:

- distortion due to non-linear elements in the transfer function (the loudspeaker, primarily);
- noise due to vibrations of the set and its internal components;
- ambient room noise and other uncorrelated noise sources;
- measurement noise and truncation errors in fixed-point DSP's.

Experimental results in [Birkett97] show that the limits prescribed by vibrations and loudspeaker distortion are functions of the amplitude of the reference input. Vibrational noise is impulsive and quite uncorrelated with the reference input, and it tends to increase with the excitation power. Loudspeaker distortion, on the other hand, is manifested by non-linear components in the transfer function. This distortion increases when the excitation power is very low (because the two-point suspension of the speaker cone imparts a small but non-zero discrepancy between two restoring forces [VanSchyn96]), or very high (due to distortion of the cone as it nears its limits of travel). Adaptive FIR filters are most effective (almost by definition) when used to model linear system responses that are in some way correlated with the reference input. They can experience great difficulty when presented with non-linear distortions and uncorrelated noise.

It is interesting to note that the least significant contributors to practical cancellation limits are due to the room transfer function, whereas the most significant contributors are solely dependent on the the reference input amplitude and the acoustic design of the telephone set, suggesting that *even the best adaptive filtering algorithms will be foiled if put to work in poorly-designed sets.*

5.3 Sample Plants

In Chapter 6, a collection of FIR filters are employed to act as unknown plants. Time-series and PSD plots of these FIR transfer functions are provided in appendix A2. All were derived from recordings made in room MC2014 at Carleton University using the following procedure:

- Record and up-load approximately 30 seconds of unfiltered broadband reference inputs and the corresponding primary signals at 48 kHz using the apparatus of figure 5.3;
- Transfer the files to the PC for off-line processing as shown in figure 5.2 (b);
- Remove energy above 3.8 kHz and notch out 60 Hz using brick-wall FIR filters implemented in MATLAB and down-sample the output sequences to 8.0 kHz;
- Feed the reference and primary signals into a 1200-tap floating-point MA-NLMS adaptive filter with $\mu = 0.5$ and allow the filter to converge to the greatest depth possible;
- Save the final adaptive filter coefficients as the approximated FIR impulse response.

Four acoustic transfer functions were obtained, all from room MC2014. They are intended to be identical except for the intentional presence and location of a reflective surface in front of the set. The size of the surface was approximately 1.0 x 1.5 metres, and it was kept perpendicular to the table top on which it and the telephone set were placed. Details of the simulated transfer functions are provided in table 5.3.

TABLE 5.3 Description of measured acoustic transfer functions for chapter 6.

Name	Description
IRN	No reflective surface present
IR1	Reflective surface mounted 1.0 metres on-axis from set
IR2	Reflective surface mounted 2.0 metres on-axis from set
IR3	Reflective surface mounted 3.0 metres on-axis from set

5.4 Acoustic Recording Procedures and Apparatus

This section is applicable to chapter 7 only, as it provides details of the hardware and file transfer procedures used to acquire reference and primary signals from the chosen acoustic venues.

5.4.1 File Generation and Transfer from PC to DAT

Reference signals used in this research were pre-processed so as to minimize the influence of all parameters other than those pertaining directly to the adaptive filter. The pre-processing was undertaken using full floating-point precision with MATLAB functions written by the author. The sequences were generated in or imported into the MATLAB environment and conditioned in the following order before being transferred to DAT:

- up-sample from 8.0 kHz to 48.0 kHz;
- brick-wall low-pass filter to 3.80 kHz;
- offset to zero the mean;
- scale to make peak amplitude 32,767 to avoid saturation and to maximize SNR;
- write header information and convert the sequence to 16-bit fixed-point PCM format.

Details of the file transfer protocols and interfaces are summarized in table 5.4, which is to be read in conjunction with figure 5.2.

5.4.2 Hardware and Recording Procedures

During each trial recorded for chapter 7, the DAT output (i.e. the reference signal) was filtered with a switched-capacitor elliptical filter (described in [Birkett97]) to simulate the anti-aliasing filter in the microphone input of a typical telephone set. The filtered output was coupled into the loudspeaker by way of a precision power amplifier, bypassing the (low-cost) power amplifier in the set. The gain was set to provide about 70 dB SPL for broadband noise at 0.5 metres on-axis from the loudspeaker. This volume level is just below the level at which vibrational noise limits the achievable cancellation for this set [Birkett97] (figure 4.14).

The hardware arrangement is sketched in figure 5.3. The ambient noise level 0.5 metres above loudspeaker was about 56 dB SPL in MC2014 and 62 dB ME4144. The handset was left off-hook and moved well away from the set for all recordings. The reference and primary signals were fed to the ADAT inputs through low-pass variable-gain analog buffers.

FIGURE 5.2: Methods to transfer files between the remote apparatus and the PC. (a): Apparatus to transfer reference signals from PC to remote DAT. The DAT output is then applied to the input of the hardware apparatus shown in figure 5.3; (b): Apparatus to transfer in-situ reference and primary recordings from the ADAT to the PC. Details of the hardware and software interfaces marked A through G are provided in table 5.4.

TABLE 5.4 Details of hardware interfaces of file transfer apparatus.

I/F	Description
A	Stereo ISA-bus 16-bit parallel DMA transfer.
B	Stereo 48 kHz S/PDIF transfer via shielded RCA-coax-RCA cable.
C	Proprietary Alesis optical fibre interface, 1 of maximum 8 channels used at 48 kHz.
D	Analog transfer between 1/4 inch mono jack and one channel of 1.8 mm stereo plug.
E	Proprietary Alesis optical fibre interface, 2 of maximum 8 channels used at 48 kHz.
F	Stereo 48 kHz S/PDIF transfer via shielded RCA-coax-RCA cable.
G	Stereo ISA-bus 16-bit parallel DMA transfer.

FIGURE 5.3: Apparatus used to make remote recordings. All the electronics in the set except for the speaker and microphone were bypassed. The set was otherwise unmodified, with its handset left off-hook during all recordings. (Figure adapted from [Birkett97], and used by permission).

5.4.3 File Transfer from ADAT to PC

The Alesis ADAT is a studio-quality eight-channel S-VHS digital recorder with a fixed sampling rate of 48.0 kHz. The ADAT employs eight discrete A/D converters so its analog inputs are not multiplexed. This feature ensures perfect alignment between the reference and primary signals. The ADAT records information in 16-bit PCM format. Its outputs were converted to S/PDIF by the Alesis AI-1 format translator, and transferred to the PC hard-disk through the Digital-Only

Card-D sound card from Digital Audio Labs under control of the Software Audio Workshop (SAW) package designed for this sound card.

Note that all transfers from ADAT to PC are conducted in discrete-time, and are *lossless*. Hence, great care was taken to condition the recorded signals so as to ensure that the results presented were not tainted by spurious noise nor artifacts of the experimental environment. The following sequence of signal conditioning operations was applied to all of the measured recordings once they had been translated into MATLAB floating point format:

- low-pass filtered at 48 kHz to 3.80 kHz using a 1024-point low-pass hanning FIR filter;
 - notch-filtered to cut the 58-62 Hz band using a 256-point hanning FIR filter;
 - down-sampled to 8.000 kHz;
 - shifted to zero the mean and scaled to achieve the desired variance or peak amplitude;
 - stored as reference and primary .MAT files ready for off-line adaptive processing.
-

6

Experiments with Simulated Systems

Chapter 6 is presented in two parts. The first concerns only the linear filters (MA-NLMS and AR-SOX-AF), and its objectives are as follows:

- to validate the SOX-AF against those reported previously in the literature (section 6.2);
- to demonstrate the influence of μ_g on the MA-NLMS filter (section 6.3.1);
- to show the influence of β and μ_g on the convergence of the AR-SOX-AF (section 6.3.2);
- to confirm the “optimal” β and μ_g proposed in chapter 4 from these experiments.

The “optimized” versions of the SOX-AF and MA-NLMS filters will act as reference algorithms in all subsequent simulations in chapters 6 and 7.

In the second part of chapter 6, the proposed DSI method for the AR-SOX-AF is assessed, and the PWP-SOX-AF filter is compared to the “optimized” linear structures. The experiments are arranged as follows:

- DSI is evaluated for adjustment windows of various lengths (section 6.4);
- misadjustment is measured as a function of filter order for the three algorithms (section 6.5);
- the plant is held constant and the variance of the reference input is changed (section 6.6.1);
- the plant is held constant and the PSD of the reference input is changed (section 6.6.2);
- the plant is held constant and the reference input is non-stationary speech (section 6.6.3);
- A step-change to the plant is applied for various reference inputs (section 6.7).

These experiments will assist in anticipating and analyzing the experimental results in chapter 7, where very little control over the operating environment can be secured.

6.1 Some Preliminaries

Most simulations in this chapter are conducted with WSS reference signals, and all model plants employed are time-invariant FIR filters that are subsets of a measured 1200-tap acoustic transfer function. Accordingly, conclusions about the algorithms' behaviours in practical environments cannot be drawn from these results alone. Care was taken to ensure that the comparisons made between algorithms were fair. Unless stated otherwise, the following statements apply to all results presented in chapter 6:

- results for stochastic input signals represent ensemble averages over 32 independent trials;
- results for speech input signals were acquired from single trials;
- all stochastic inputs were scaled to have unit variance and zero mean;
- all speech inputs were scaled to have equal peak amplitude and zero mean;
- the plants to be identified were the N lowest-order elements of several 1200-tap FIR transfer function measurements made in MC2014 under various static conditions. Details of these transfer functions and how they were obtained are provided in appendix A2;
- each element of p_0 was set to a small positive constant and β was set to $1-1/N$ in the steady state for the AR-SOX-AF;
- For the PWP estimator, $L1$ was set to 16, and $L2$ to $N/16$ for N a composite of sixteen.

The eigenanalysis methods described in section A1.7 of Appendix A1 were used to calculate the fair re-start values for the adaptive filters where appropriate. The computational requirements of doing so were orders of magnitude greater than for the examples shown in the appendix, as the filters used here were of much higher order.

6.2 Comparison with Published Results

An attempt was made to duplicate the experimental results in [Marshall89] to verify the proper operation of the transform domain algorithms discussed here. The same algorithm parameters were used (unless they were unspecified) to facilitate good correlation between the two works.

Four reference input distributions were generated by passing gaussian noise through a particular 32-tap FIR filter. The filter coefficients and plots of the spectral distributions they yield are given in [Marshall89]. Each plot in figure 6.1 is associated with one of these distributions, and shows the convergence path for three transforms - the identity (i.e. the MA-NLMS algorithm), the DFT, and the DCT. In all cases, the order of the model plants and filters were fixed at $N=8$, and a small measurement noise -100 dB relative to unity was added to the primary input. Each curve is an ensemble average of 32 independent trials. It appears that only single trials were used in [Marshall89].

The results are presented in figures 6.1 (a), (b), (c), and (d), and these plots correspond to figures 6, 7, 8, and 10 in [Marshall89] respectively. It is important to note that these simulations cannot be used to support general conclusions about the convergence rates of the various transforms and distributions, as the experiments were not arranged so as to allow fair comparisons to be made. For example:

- different step-sizes were prescribed for each algorithm;
- corresponding filter eigenmodes were not identified and traversed explicitly in each domain;
- the convergence plots shown represent a single trial only instead of an ensemble average.

These results serve only to demonstrate that there is no optimal fixed transform - some are better decorrelators than others for certain reference input distributions. The results obtained here are in good general agreement with those presented in [Marshall89].

FIGURE 6.1: Duplication of results of [Marshall89] for four reference inputs. In each of these plots, the MA-NLMS algorithm (dotted), the DFT-based AR-SOX-AF (dashed) and the DCT-based AR-SOX-AF (solid) are shown over ensemble averages of 32 trials. The plots suggest that there is no optimal fixed transform.

6.3 Algorithm Optimization

Before meaningful comparisons are made, the behaviour of the reference algorithms must be understood and the parameters to achieve their best performance must be determined. The aim of this section is to justify the choice of β and μ_g through demonstration. For the time being, the *rate* of convergence rather than the *depth* of convergence will be the metric of choice. Less emphasis is placed on the depth of convergence because it is unlikely that the misadjustment of any of these algorithms will be below the noise floor in a typical AEC environment. The major factors responsible for this noise floor were described in section 4.1.

The standard algorithms for purposes of comparison are the MA-NLMS algorithm described in section 2.4.1.1 and the AR-SOX-AF [Narayan83] with the step-size estimates made as discussed in section 4.2.3. Optimal values of μ_g must be determined for each, and an optimal choice for β is required for the latter. Once the parameter selection for algorithms is stabilized, the effects of new modifications can be quantified in isolation.

6.3.1 Selection of μ_g

The governing parameter in all LMS-based algorithms (and hence the plant identification stage in the SOX-AF) is μ_g , the significance of which has been studied extensively (comprehensive summaries are provided in [Widrow85] and [Haykin96]).

6.3.1.1 Procedure

An MA-NLMS filter was constructed with $N=256$ to identify the fixed 256-tap FIR acoustic transfer function $\text{IRN}(256)$. No measurement process noise was added. The global step-size μ_g was set to 0.2, and the average ERLE was computed over 32 independent trials with unfiltered broadband reference sequences. The experiment was repeated for μ_g set to all multiples of 0.2 less than two, and the results are displayed in figure 6.2. These experiments were repeated for the SOX-AF, with β fixed at $1-1/N$.

FIGURE 6.2: Convergence rates as a function of μ_g in the MA-NLMS algorithm. In each case, the solid line shows $\mu_g=1$. The dotted lines show convergence rates for μ_g set to 0.8, 0.6, 0.4, 0.2 (a) and 1.2, 1.4, 1.6, 1.8 (b) from fastest to slowest.

6.3.1.2 Observations

The results show that the fastest convergence is achieved for broadband excitation when μ_g is half of its maximum value. Further, the convergence rates for MA-NLMS are nearly symmetric about $\mu_g=1$. It was found that convergence in the AR-SOX-AF is similarly dependent on μ_g for any (reasonable) fixed β .

Convergence rates for the AR-SOX-AF were very similar to those of MA-NLMS with β set to $1-1/N$. However, it was found that when β was set much above or below this value, the filter *diverged* for some trials, even for $\mu_g < 2$. The range of β values for which the filter behaved well tended to broaden as the PSD of the excitation became more uniform. In all cases, stability could be re-gained for a given input by reducing μ_g or moving β closer to $1-1/N$. A definite interplay between μ_g and β was observed in these experiments.

6.3.1.3 Analysis

In the MA-NLMS filter, μ_g determines the range of the tap-weight update vector. As an LMS algorithm, the gradient estimate is made each iteration over a short period of time (in this case, the most recent N samples), so it likely to vary somewhat from one iteration to the next because

the stationarity of the reference over such short intervals cannot be assured. This phenomenon has been referred to elsewhere as *gradient noise*.

Smaller values of μ_g will reduce the gradient noise, because less distance is travelled at each iteration. Though the trajectory will tend to be fairly smooth over a number of iterations, the overall convergence rate will be reduced because the step-size limits the maximum distance which can be traversed. Larger values of μ_g will increase the total distance travelled, but the trajectory will be more erratic because the gradient noise is proportional to the step-size. It appears that the best trade-off between the two is achieved with $\mu_g = 1$. These findings are agreeable with those of previous researchers [Sommen87] [Makino93]. Therefore, μ_g was set to one in all MA-NLMS filters implemented herein.

The AR-SOX-AF also employs an LMS filter, so its plant identification stage has a very similar dependence on μ_g . However, the channel step-sizes vary each iteration at a rate determined by β and around a value determined by the recent time-averaged power in each channel. Evidently, the choice of β in these algorithms is significant, and its influence will be investigated now:

6.3.2 Selection of β

The governing parameter of the signal identification stage in the AR-SOX-AF is β . It will be shown here to have a direct effect on the overall filter performance. This was anticipated in table 4.1 where the lineage of w_k was found to originate with β . It has been observed already that the AR-SOX-AF may become unstable if β is chosen outside a fairly narrow range.

6.3.2.1 Procedure

All plants and adaptive filters employed in this section were of order $N=128$. In the AR-SOX-AF trials, μ_g was held constant at $0.5/N$ to reduce the potential for instability; μ_g for the MA-NLMS filter was fixed at unity at all times. Uncorrelated measurement noise 80 dB below the reference signal was added to the primary.

Once all stages of the structures had reached their steady states, the distance vector $w_k - w_{opt}$ was set to provide 0 dB ERLE in the direction of the slowest convergence mode. The convergence

modes were identified using the eigenanalysis techniques described in appendix A 1. In addition to providing comparable starting points for each algorithm, this method also ensured that valid power estimates were loaded in the SOX-AF when the re-initialization occurred. Thus, the only free variable in these simulations was β , which was chosen to be $1-b/N$ where b was set to .125, .25, .5, 1, 2, 3, 4, 5, 6, and 8. The excitation sequences employed were broadband (figure 6.3), AR2(-1.5955,.95) (figure 6.4), and AR12 (figure 6.5).

FIGURE 6.3: Convergence rates of the AR-SOX-AF with broadband excitation for various values of β . Here, β is computed as $\beta=1-b/N$. In (a), the curves from the fastest to slowest represent $b=2, 1, 3, 4, 5, 6, 8$. In (b), the curves from fastest to slowest represent $b=1, 0.5, 0.25, 0.125$, and for the MA-NLMS filter implemented in the time domain.

6.3.2.2 Observations

When the excitation is broadband, β has a minor but definite influence on the convergence rate and misadjustment of the AR-SOX-AF. The most rapid convergence was obtained when β was set close to one, as expected. When β was set smaller than about $1-4/N$ (figure 6.3 (a)), the rate of convergence fell off rapidly, and the maximum depth of convergence was noticeably reduced. When β was set larger than $1-1/N$ (figure 6.3 (b)), the convergence rate decreased somewhat, but not as dramatically as before. There was little change in the misadjustment when β was made larger. The MA-NLMS algorithm was slower to converge than these, but not significantly so.

FIGURE 6.4: Convergence rates of the AR-SOX-AF with AR2(-1.5955, 95) excitation for various values of β . Here, β is computed as $\beta=1-b/N$. In (a), the curves from fastest to slowest represent $b=1, 2, 3, 4, 5, 6, 8$. In (b), the curves for $b=0.5, 0.25, 0.125$ are clustered at the far left, the curve for $b=1$ is slightly to the right of the cluster; the slowest curve is for the MA-NLMS algorithm.

FIGURE 6.5: Convergence rates of the AR-SOX-AF with AR12 excitation for various values of β . Here, β is computed as $\beta=1-b/N$. In (a), the curves from the left represent $b=1, 2, 3, 4$. In (b), the curves that represent $b=1, 0.5$, and 0.25 are clustered at the far left, the curve for $b=0.125$ is slightly to the right of the cluster, and the slowest curve is for MA-NLMS.

Similar observations were made for the AR2 and AR12 processes, however, the range of the convergence rates was more extreme as β was varied, and in both cases the MA-NLMS filter was found to be significantly slower than the AR-SOX-AF, even over a wide range of β values. Again, the fastest convergence was obtained for $\beta=1-1/N$ for these inputs, and larger values of β appeared not to affect the misadjustment significantly. Certain trials in the AR12 experiments diverged when β was chosen less than $1-5/N$ or greater than $1-.5/N$ for $\mu_g=1$.

6.3.2.3 Analysis

Smaller values of β increase the variance in the estimated step-sizes, and this in turn increases the misadjustment. If the algorithm noise is greater than the added measurement noise, it will be the limiting statistic in the steady-state performance of the filter. If the measurement noise was removed, the gradient noise would eventually limit the depth of cancellation regardless of the choice of β .

These results indicate that a preferred β exists for these WSS inputs. When similar experiments were conducted with $N=64$, 256, and 512, the results obtained agreed closely with those shown here. In almost all trials, convergence rates reduced as β was moved further away from $1-1/N$.

To summarize, β greater than $1-1/N$ is generally undesirable because it will make the linear channel power estimators sluggish, holding the step-sizes artificially high when the channel power increases, and possibly making the LMS stage temporarily unstable. As demonstrated in section 4.5.3, the channel power can fluctuate considerably even when the reference input is WSS. Values of β less than $1-1/N$ are undesirable because they will force the linear channel power estimators to track the instantaneous channel power rather than the power envelope. Since the instantaneous power is known to be pulsed DC, certain step-size estimates will be very large as the instantaneous power nears zero, yielding step-sizes with large variance even though the average power may be quite constant, therefore increasing the algorithm noise unnecessarily.

6.4 Dynamic Step-Size Initialization

In sections 4.2.3 and 6.3.2, justification was provided for specifying β to be $1-1/N$, which is very near unity for large N . However, it has been observed in many trials that this value of β can cause serious divergence or overshoot in the early stages and that the convergence behaviour in this transient phase is highly dependent on the choice of p_0 when β is close to unity.

DSI was introduced in section 4.4.1 as a simple variable- β technique which allows the arbitrary values of p_0 to be dismissed more rapidly than would be the case with a static β . Faster initial convergence is anticipated with DSI because p_k for $k \leq N$ becomes more representative of the actual spectral characteristics of the input.

The purpose of DSI is to accelerate the rate at which p_0 is dismissed from p_k for small k so that μ_k becomes less a function of p_0 and progressively more a function of p_k in the early stages of convergence. The purpose of these experiments is to demonstrate the effectiveness of DSI in reducing the initial divergence, particularly when β is very close to unity, which is most often the case in practice.

6.4.1 Procedure

In these experiments, the plant used was IRN(256), and the adaptive filter order was $N=256$. The excitation was unfiltered broadband noise normalized to have unit variance and zero mean. On start-up, β_k was set to $1/N$, and it was increased linearly to $\beta_k = 1-1/N$ over D iterations, after which it was held constant. The adjustment interval D was set in each ensemble to be $D = 0$ (i.e. DSI disabled), 8, 32, 64, 128, 256, 512 to 1024. In all cases, $\mu_g = 0.5/N$ and all elements of p_0 were set to unity. Uncorrelated measurement noise 80 dB below the reference signal was added to the primary. The results presented are ensemble averages over 32 independent trials. The experiments were repeated for several AR2 and AR12 inputs.

6.4.2 Observations

The results of the broadband-input DSI experiment are provided in figure 6.6. The least amount of divergence was achieved when β was set to increase for $D = N/8 = 64$ iterations, followed by

$D = N/4$, $N/2$, and N iterations. The overshoot was not serious in any of these cases, rather, the onset of convergence was delayed slightly. Divergence is much more apparent when D was set greater than $2N$, and it is particularly bad when D is less than $N/8$.

These trials were repeated for a multitude of different reference input distributions. The best choice for D varied with each distribution, but the range was always between $N/8$ and $2N$. The amount of overshoot varied tremendously in some sample paths, and was heavily dependent on p_0 . These results were not shown as they were similar in form to those presented in figure 6.6.

FIGURE 6.6: Reduction in overshoot for broadband excitation using DSI. In (a), the fastest convergence is achieved with $K=N/8$ iterations, followed by $K=N/4$, $K=8$, and $K=0$. In (b), the fastest convergence is obtained with $K=N/2$ samples, followed by $K=N$, $2N$, and $4N$.

6.4.3 Analysis

Smaller values of β yield power estimates that are more representative of the instantaneous power in each channel. Thus, the signal identification capability is accelerated, but a penalty in net convergence rate and misadjustment is paid because the range of the step-sizes is increased, sometimes beyond the filter's stability limits. DSI attempts to merge the best of both options by reducing the adoption rate of the channel power estimates (i.e. β) in proportion to the amount of knowledge gained by the filter of the actual power distribution.

The choice of a suitable D appears to be highly dependent on the spectral distribution of the input, so it is difficult to draw general conclusions about the optimal choices of D when using DSI. In this work, $D = N$ was selected as a first approximation. Though in some cases it delayed the onset of convergence (as in figure 6.6 (b)), it did not exhibit any serious divergence in any of the simulations. It was found to provide the best trade-off between initial overshoot and sluggishness for the trials conducted here.

6.5 Performance Limitations Due to Undermodelling

The maximum steady-state cancellation capabilities of the “optimized” algorithms are compared in this section, for cases where the adaptive filter is of lower order than the plant to be identified. The metric employed in this section is the TIP/TP ratio [Knappe92] of the unknown plant.

6.5.1 Procedure

The unknown plant in this section was the FIR sequence IRN(1200). Details of its properties and how it was obtained are provided in appendix A2. The reference input was unfiltered broadband noise. The convergence parameters μ_g and β were set to their preferred values per sections 4.2 and 6.3.

The theoretical limit of cancellation for a given filter order is computed using the TIP/TP ratio of equation 5.2. For the experiments, the filter order N was set to 64 and to integer multiples of 128 to a maximum of 1152. A single trial was run for each algorithm, and the average converged ERLE for each filter was calculated well after the steady-state cancellation depth was reached. The results were plotted along with the theoretical limit curve.

FIGURE 6.7: Cancellation limits for the three LMS algorithms with a measured FIR acoustic plant. The solid curve is the TIP/TP ratio for each N to the full length of the plant. The measured steady-state limits for various N are provided for the MA-NLMS algorithm (circle), the AR-SOX-AF (plus), and the PWP-SOX-AF (cross).

6.5.2 Observations

The results shown in figure 6.7 represent the average ERLE over the interval of $k=152,000$ to $160,000$ (19 to 20 seconds at 8 kHz) by which time the filters appeared to have converged to their limits of cancellation for at least 5,000 iterations.

The AR-SOX-AF performance is almost identical to that of MA-NLMS for $N < 1000$, which is in agreement with the analysis developed in [Florian88]. It fails to keep up with MA-NLMS for $N > 1000$, and this has been observed in practice for many different reference inputs. The PWP-SOX-AF provides about 3 dB deeper cancellation than the other algorithms, and it outperforms the MA-NLMS filter for N large.

6.5.3 Analysis

The better steady-state performance of the PWP-SOX-AF is quite easy to explain in light of the development of the algorithm provided in section 4.4.1. It was shown in figure 4.5.3 that the PWP estimate is consistently greater than those obtained using the preferred values of the MA or AR methods. The channel step-sizes are bound to be lower in the PWP algorithm as they are the reciprocals of the channel power estimates. It is well-known that lower step-sizes yield deeper convergence in the absence of interfering noise sources.

It should be remembered that *the TIP/TP ratio is a valid metric only if the PSD of the excitation is uniform*. The observed depth of convergence may well exceed the theoretical limit prescribed by the TIP/TP ratio if the excitation is narrow band, because the out-of-band components in the plant become less significant. It will be seen in chapter 7 that wider band signals are the most difficult reference inputs to track and to cancel effectively, and that greater ERLE is achieved for inputs with progressively narrower bandwidth in practice. This is not to say that the optimal adaptive filter solutions have been reached in such cases.

6.6 Signal Identification Stage

In this section the performance of the PWP-SOX-AF is compared to the AR-SOX-AF and the MA-NLMS filter in circumstances where prescribed changes are applied to the reference input while the unknown plant is held constant. In all cases, the adaptive filter order is the same as that of the unknown plant, and an independent and uncorrelated noise sequence 80 dB down from the reference is added to the primary input so as to make the convergence depth finite.

6.6.1 Step Changes in Variance

The experiments in this section were conducted to determine how readily the power estimation mechanism in the SOX-AF tracks abrupt increases and decreases in the reference input PSD when all other parameters are held constant. The ability to track large variances in dynamic range is essential in speech applications due to the bursty nature of speech signals.

6.6.1.1 Procedure

The plant used was IRN(256), the adaptive filter order was $N=256$, μ_g was fixed at 1 (for the MA-NLMS algorithm) and $1/N$ (for the SOX-AF's), and β was set to $1-1/N$. The excitation was unfiltered broadband noise with unit variance and zero mean. Once the filter had converged to a steady-state, the variance of the reference signal was doubled suddenly in three short and spaced intervals. The results shown in figure 6.8 are ensemble averages of 32 independent trials that have been smoothed with a hanning window of 257 taps.

6.6.1.2 Observations

The results of the experiment are shown in figure 6.8. The MA-NLMS algorithm appears to track the step-change well, with only a slight lag in convergence. The AR-SOX-AF achieves the same convergence depth as the MA-NLMS algorithm, however it lags behind noticeably. At first, it appears to track the leading edge of the burst perfectly, but very soon the ERLE reduces before it catches up to the MA-NLMS filter.

FIGURE 6.8: Reference input tracking behaviour for the three algorithms when a step increase in variance is applied to a broadband reference input. In (a), the MA-NLMS algorithm (dotted) is shown with the AR-SOX-AF; in (b), MA-NLMS (dotted) is compared to the PWP-SOX-AF (solid).

The PWP-SOX-AF achieves cancellation levels that are consistently 2 dB deeper than those of the other two algorithms. Further, it does not exhibit the delay of the AR-SOX-AF; rather, it mimics the performance of the MA-NLMS algorithm in terms of its responsiveness to a sudden

increase in reference power. Very similar results were obtained when the experiments were repeated with AR2 and AR12 distributions, so only the broadband results are shown here.

6.6.1.3 Analysis

When the amplitude (variance) of a signal is increased and all other parameters are held the same, the eigenvalues of R increase, but the orientation of the eigenvectors remains unchanged. Hence, the performance surface remains in the same position and it retains its proportion, but its slope is scaled uniformly in all directions. Ideally, the filter coefficients would remain the same, but the step-sizes which control their movement would be scaled back instantly in proportion to the increase in their associated eigenvalues.

The MA-NLMS filter is able to track step changes in PSD very well in this experiment because there is a direct mapping between the elements of the adaptive filter and the plant. Thus, when a sharp increase in power is detected, it will propagate through the two filters in a very similar manner. Looking back to equation 2.5, y_k will remain very close to d_k so e_k will remain small, and the adaptive filter correction term (equation 2.41) in the LMS filter update equation is small as a result. Once the limit of convergence is achieved it will be maintained almost regardless of what changes are applied to the excitation in a time-domain LMS filter.

In the transform domain, however, a period of re-adjustment is required when the step-change is detected. This is because each tap in the SOX-AF corresponds to a different spectral mode of the plant, each of which is dependent in turn on the entire plant vector. Since the primary signal is the linear convolution of the reference input with the plant, the modes of the settled primary are not available immediately. Thus, a lack of correspondence between the primary signal d_k and the adaptive filter output y_k can be expected during the transient phase.

Furthermore, the channel power estimates tend to move away from their previous states as the leading edge of the step-change propagates through the observation window. The p_k values are transient during this time, and (if the reference power has increased suddenly) the sluggishness of the estimator may force the step-sizes to remain too large and to cause temporary instability in the adaptive filter.

The AR-SOX-AF responds well until the lack of correspondence between d_k and y_k is detected. When it is, the filter becomes temporarily unstable because the step-sizes are too slow to change from their present (higher) values to their required (lower) values. The AR-SOX-AF exhibits the worst response in this experiment because its linear estimator always lags behind the measured instantaneous power values.

The PWP estimator, on the other hand, reduces the step-size estimates with zero lag. Hence, the the PWP-SOX-AF is able to track the step-change in reference power without exhibiting the temporary instability of the AR-SOX-AF. The PWP-SOX-AF achieves consistently deeper depths of convergence because its channel power estimates tend to be larger (as shown in figure 4.9) and therefore its step-size estimates tend to be smaller. Smaller step-sizes reduce the steady-state misadjustment in the absence of other more significant noise sources or non-linearities.

6.6.2 Step Change in Distribution

This experiment was designed to show how well the various adaptive filters track reference inputs whose total power is constant, but whose PSD is changing. This corresponds to a change in the eigenvectors of R , which in turn indicates that the shape of the performance surface is changing. Accordingly, the power estimates in the SOX-AF should adjust themselves to reflect the new eigenvalues as quickly as possible. The location of the minimum remains the same because the plant is constant, so the filter coefficients should also remain unchanged. In the full-band MA-NLMS filter, the total power estimate should remain unchanged because the trace of R (the total power in the reference signal) is constant in this case.

6.6.2.1 Procedure

The reference signal was constructed by concatenating sequences of several AR2 processes together. Each sequence was scaled to have unit variance and zero mean. The reference signal, summarized in table 6.1, was chosen to cycle through every possible transition between three AR2 processes. A random measurement noise 80 dB below the reference signal was added to the primary to limit the convergence depth. The plots show ensemble averages of 32 independent trials, smoothed with a hanning window of 257 taps.

TABLE 6.1 Concatenated sequence of AR2 processes.

Iteration Interval	Reference Signal
1-8,999	AR2(-.195,.95)
9,000-16,999	AR2(-.975,.95)
17,000-24,999	AR2(-1.5955,.95)
25,000-32,999	AR2(-.195,.95)
33,000-40,999	AR2(-1.5955,.95)
41,000-48,999	AR2(-.975,.95)
49,000-56,999	AR2(-.195,.95)

FIGURE 6.9: Reference input tracking behaviour for three different algorithms when the excitation is a concatenation of AR2 processes. The solid curves show the responses of (a) the AR-SOX-AF and (b) the PWP-SOX-AF. The dotted curves are the responses for the MA-NLMS filter. The AR2 sequences used in as reference are listed in table 6.1.

6.6.2.2 Observations

The plots show that the MA-NLMS filter tracks the changes almost perfectly. The AR-SOX-AF, however, can track changes from higher bandpass to lower bandpass well, but shows substantial divergence when tracking changes from lower bandpass to higher bandpass. This is indicated by

the downward spikes in ERLE at $k=25000$, 41,000 and 49,000. The echo signal was *amplified* by 10 dB on average at $k=25000$ in this experiment. The PWP-SOX-AF diverges coincidentally with, but not nearly as severe as the AR-SOX-AF.

6.6.2.3 Analysis

The power estimates in the AR-SOX-AF are known to lag the instantaneous power signals from the transform. When the input PSD is changed to lower bandpass after a long period of middle band-pass excitation, neither the lower nor higher channel step-sizes are particularly large. Thus, when the PSD peak in the reference excitation is shifted suddenly, the changes in the channel step-sizes are not extreme, and the transient misadjustment of the filter is minimal.

On the other hand, after a long period of low band-pass excitation the step-sizes for the higher bins are very large because the power is concentrated at the lower extreme of the spectrum. When higher frequency excitation appears suddenly, the step-sizes are too large for the channel outputs, and the filter diverges temporarily. The amount of divergence is dependent on the lag in the power estimator and the relative change in the channel power. The adaptive filter taps will gradually re-converge as the step-sizes move to their desired values.

The PWP-SOX-AF was not expected to show divergence, since the estimator guarantees zero lag when the power is increasing. The minor divergence shown may come about while the step change is propagating through the observation window. During this phase, the reference and primary signals are in flux, and the power spectra of d_k and y_k may not be agreeable with one another. Accordingly, the error signal may not be representative of the “true” difference between the plant and w_k .

To treat divergence in AR-SOX-AF, one could either shut off the offending taps with a low-power channelized switch, or add a small white noise signal to reference input. Both alternatives were suggested in [Beaufays95b], the latter reported to produce slightly better results. However, it is difficult to develop design rules to specify the noise level or the shut-off threshold which should be applied. In this work, a shut-off mechanism was employed in each channel, because the addition of noise in the reference signal is highly undesirable in acoustic applications.

6.6.3 Speech Sequences

The previous two experiments demonstrated the behaviour of the three adaptive filters when they were confronted with abrupt and solitary changes in the otherwise stationary reference input. While these experiments are useful to gain insight, they are of limited utility in practice.

Two speech sequences - one male and one female - are used here for the first time in this research. These signals are continuously non-stationary, so the power spectrum and the total power in the observation vector is evolving constantly. The first purpose of this experiment is to determine whether the power estimators in both the time- and transform-domain algorithms can track these changes effectively.

This experiment carries a secondary purpose as well. The adaptive filters here are of lower order than the FIR plant, so there will be a finite limit to the convergence depth. The filter order was chosen here to be $N=512$ based on practical results obtained in chapter 7, where it was found that filters of this size yielded the best compromise between misadjustment and tracking ability. Thus, the results of this section act as a benchmark of ideal performance for several experiments in chapter 7. The performance in those experiments will undoubtedly be impaired by factors in the acoustic environment which are not apparent here.

6.6.3.1 Procedure

The unknown plant was IRN(1200) as described in Appendix A2. The filter order in all trials was $N=512$, and all filter parameters were set to their preferred values. The reference signals were scaled to have a peak magnitude of 32,767.

A shutoff mechanism was employed so as to prevent the filter from attempting to converge to noise when the reference input power was very small. The threshold in this case was chosen by trial and error to be the smallest value possible, since the only noise in the primary was due to the filter's inability to model the unknown plant completely. The minimum threshold for the MA-NLMS was found to be very small, and none was required for the PWP-SOX-AF at all, as it is a zero-lag estimator. The AR-SOX-AF, however, required a significant shut-off threshold to maintain stability.

6.6.3.2 Observations

The results are shown in figures 6.9 and 6.10 for the female and male sequences respectively. The MA-NLMS filter and PWP-SOX-AF achieve similar convergence depths which average about 40 dB when the reference signal is active. Note that the ERLE remains quite high between speech bursts for the MA-NLMS and PWP-SOX-AF curves in this experiment. This is because those filters remain active at all times, regardless of the power in the reference input.

The same is not true for the AR-SOX-AF. Through trial and error, the minimum power threshold which allowed the filter to remain stable was about 35. Thus, the AR-SOX-AF cannot continue to cancel reference signals when the power estimates fall below this threshold. This explains why the cancellation is lower than the other algorithms during periods of relative quiet. During the active periods, the convergence of the AR-SOX-AF is shallower than the other algorithms.

6.6.3.3 Analysis

The equal steady-state performance of the MA-NLMS filter and the PWP-SOX-AF suggests that deeper convergence is not possible for this combination of excitation, plant, and filter order. In previous experiments, the PWP-SOX-AF achieved deeper steady-state cancellation because its step-sizes were consistently lower than those of the other filters. Here, it would appear that the noise due to undermodelling has swamped out the misadjustment due to the smaller step-sizes.

The failure of the AR-SOX-AF to achieve the same depth of cancellation is indicative of a structural limitation, due in part to the need for the shut-off mechanism. The performance surface in these cases is evolving continuously, and it has been seen already that the AR estimator always lags the true power, often to the detriment of the overall filter performance. When a performance surface is evolving rapidly and continuously as is the case here, it appears that the ability of the signal identification stage in the AR-SOX-AF compromises the overall performance of the filter.

General conclusions should not be drawn from this experiment as the signals chosen are single trials. Nonetheless, speech signals provide good insight. Results from similarly sized structures with actual acoustic signals will be studied in chapter 7, so these plots will act as benchmarks for the algorithms in noise-free environments.

FIGURE 6.10: Signal Identification of SSGF2 track. No measurement noise added in primary. MA-NLMS (dotted) and PWP-SOX-AF (solid) required no thresholds. The AR1-SOX-AF (dashed) required a threshold of 35.

FIGURE 6.11: Signal Identification of SSGM1 track. No measurement noise added in primary. MA-NLMS (dotted) and PWP-SOX-AF (solid) required no thresholds. The AR1-SOX-AF (dashed) required a threshold of 35.

6.7 Plant Identification Stage

LMS filters in the time domain are known to be relatively slow to identify unknown plants when the input excitation is non-uniform [Morgan95a]. In this section, the SOX-AF will be shown to enable much faster plant identification than the time-domain NLMS algorithm of equal order for WSS narrow-band inputs.

Unlike the previous experiments, the reference excitation will remain constant here, and a step change to the unknown plant will be applied. In all cases, the unknown plant will be an FIR plant that represents a measured impulse response from the same equipment in the same room. The only difference between the plants being used is the location of a large reflective surface. Plots of the room transfer functions associated with these files can be seen in the appendix A2.

6.7.1 Procedure

A stationary reference signal of 20,000 samples was applied to an FIR plant and adaptive filter, each with 256 taps. Several reference signal distributions were used, with an ensemble of 32 independent trials run for each experiment. The adaptive filter was initialized to its optimal value for each trial because the initial convergence of the filter is not being examined here. The adaptive filtering loop was activated with a measurement noise of -80 dB (relative to unity) added to the primary signal.

Once the steady-state was achieved, the plant transfer function was changed suddenly, forcing the adaptive filter to converge towards a new solution. The plant was changed after each interval of 4,000 samples in the sequence specified in table 6.2. For each ensemble, the MA-NLMS filter ERLE is plotted against those of the AR-SOX-AF and the PWP-SOX-AF. The plots have been smoothed with a 257-tap hanning window.

TABLE 6.2 Sequence of plants to be identified.

Iteration Interval	Plant
1-1,999	IRN(256)
2,000-5,999	IR3(256)
6,000-9,999	IR2(256)
10,000-13,999	IR1(256)
14,000-18,000	IRN(256)

6.7.2 Observations

The results for the broadband input are shown in figure 6.12. There appears to be little difference between the convergence rates of the AR-SOX-AF and the MA-NLMS algorithm in this case. The PWP-SOX-AF, however, converges at a slightly slower rate.

When the experiment was repeated with narrow band excitation (several AR2 distributions and AR12), the MA-NLMS filter was very slow to re-converge. The transform domain filters, on the other hand, were able to re-converge almost at the rate they did for broadband excitation. Only the results for the AR12 reference trials are shown in figure 6.13, as those recorded for the AR2 distributions were quite similar.

FIGURE 6.12: FIR Plant tracking capability with broadband inputs. The AR-SOX-AF (solid) is shown in (a) and the PWP-SOX-AF (solid) is shown in (b). In both plots, the MA-NLMS algorithm (dotted) is shown.

FIGURE 6.13: FIR Plant tracking capability with AR12 excitation. The AR-SOX-AF (solid) is shown in (a) and the PWP-SOX-AF (solid) is shown in (b). In both plots, the MA-NLMS algorithm (dotted) is shown.

6.7.3 Analysis

When the reference input is broadband, the eigenvalues of R are approximately equal, so the MA-NLMS algorithm is able to converge evenly regardless of the trajectory. The AR-SOX-AF was not expected to achieve faster convergence, as shown in figure 6.12. The faster convergence in the SOX-AF may arise from the continual “tuning” of the step-sizes in each channel in response to the short-term variances in the reference PSD. The MA-NLMS algorithm employs a scalar step-size, and thus cannot track the power in individual eigenmodes. The PWP-SOX-AF converges at a slower rate because on average its step-sizes are smaller, as shown in figure 4.9.

The SOX-AF’s converge much faster than the MA-NLMS algorithm when the reference input is not of uniform distribution. These experiments reveal the key advantage offered by orthogonal algorithms over their time-domain counterparts: the transform decouples the eigenmodes of the performance surface, and the step-sizes adjust the filter gain in each channel to suit the reference distribution. The net effect of the two operations is the “pre-whitening” of the reference input before it is fed to the LMS stage.

In summary, the convergence rate of the MA-NLMS filter to a new optimal solution is governed always by the largest eigenvalue of R_{xx} , whether or not the associated eigenmode is traversed during the course of convergence. In a transform domain filter, however, the convergence rate is governed more by the step-sizes in the eigenmodes which are being traversed. Because the step-sizes are scaled to suit the power in each eigenmode, the overall convergence rate of the filter is often independent of the reference input distribution, provided that the reference is WSS. The plant-tracking capabilities of the self-orthogonalizing filter are highly desirable in AEC, as room transfer functions can be expected to change rapidly and continually in most applications.

6.8 Summary

The first tangible results of the research have been presented in this chapter. In the first part, a collection of experiments were crafted to determine the preferred values of the convergence parameters for the previously known MA-NLMS and AR-SOX-AF algorithms. It was found that $\mu_g=1$ (for the MA-NLMS filter), $\mu_g=1/N$ and $\beta=1-1/N$ (for the AR-SOX-AF) yielded the best results, as had been suggested in section 4.2.

In the second part of chapter 6, the usefulness of the proposed Dynamic Step-Size Initialization procedure was demonstrated, though no claims as to its general performance were made. Later, the behaviours of the (optimized) MA-NLMS and AR-SOX-AF algorithms were compared against the proposed PWP-SOX-AF. In each experiment, the control parameters were tightly regulated so as to examine the ability of each algorithm to track a particular stimulus introduced in the environment.

First, the steady-state limits of cancellation were compared against the theoretical limit, given by the TIP/TP ratio described in section 5.2.2. It was found that the MA-NLMS and AR-SOX-AF algorithms achieve similar steady-state cancellation almost regardless of filter order. The latter was found to be inferior only in the highest-order systems. The PWP-SOX-AF achieved depths of cancellation that were consistently 2 dB deeper than the other two structures, most likely due to its smaller effective step-size.

The ability of the algorithms to track changes in the variance and distribution of the reference input was examined in section 6.6. It was found that the lag of the linear AR estimator impaired the ability of the AR-SOX-AF to track abrupt power increases in some portion of the reference spectrum. The non-linear PWP estimator was able to reduce or eliminate these effects, as it is delayless when the reference power increases. The PWP-SOX-AF and MA-NLMS algorithms achieved similar depths of cancellation when speech bursts were used as the reference inputs. The AR-SOX-AF was clearly inferior by comparison.

When the input is stationary, however, the AR-SOX-AF achieves the fastest convergence of the three. This was the topic of discussion in section 6.7. The PWP-SOX-AF was considerably faster than the MA-NLMS algorithm, but because its average step-sizes are smaller than those of the AR-SOX-AF, it is slower to re-converge when the plant is changed.

7

Experiments with Acoustic Systems

In this chapter are presented the penultimate results of the research. All reference and primary signals mentioned here were acquired in real-time from a commercial telephone set situated in the Minto Centre conference room MC2014 or in the graduate student office ME4144, Carleton University, using the methods and apparatus described in chapter 5. The results are summarized numerically in tables at the end of the chapter. These tables facilitate comparison between the ERLE achieved by the different algorithms when they are confronted with stationary and non-stationary excitation in both static and time-varying environments.

7.1 Algorithm Parameter Specification

Based on the results presented in section 6.3, μ_g was set to one in the MA-NLMS algorithm and to $1/N$ in the AR-SOX-AF. The persistence factor β was set to $1-1/N$ in all implementations of the AR-SOX-AF. For the PWP-SOX-AF, $L1$ was set to 16 and $L2$ to $N/16$, as discussed at the end of section 4.5.5.

7.1.1 Activation Thresholds

In cases where the reference excitation is non-stationary, it is important to disable the adaptation when the reference power drops below some prescribed threshold so as to prevent the filter from converging to noise in the primary signal. This is particularly important in AEC, because the excitation is known to be very bursty in time, and to have an uneven PSD. The switch in single-channel systems (such as the NLMS algorithms) is activated by the total input power. In the SOX-AF, however, a separate switch is used in each channel to ensure that only those channels with sufficient excitation are activated.

The maximum theoretical cancellation depth anticipated in this application is on the order of 40 dB. Hence, there is no advantage to be gained by setting the switching thresholds more than 40 dB below the peak amplitude in the primary. The recorded reference and primary speech signals were both scaled to have peak magnitudes of 32,767, which yielded *root-mean-square* (RMS) amplitudes between 6,000 and 8,000 over a typical speech burst. The activation thresholds were chosen to be slightly less than 40 dB below the RMS value, which corresponded to a numerical threshold of about 50.

Recall that in the (noiseless) simulations of section 6.6.3, the minimum threshold value required to maintain stability in the AR-SOX-AF was found to be about 35, so the choice of 50 for use in non-ideal circumstances is not unreasonable in this sense either. The switching threshold will limit the depth of convergence to about 40 dB, however this is nearly 10 dB beyond the best AEC results reported in the literature [Gilloire94] [Naylor94] [Gilloire95].

7.1.2 Adaptive Filter Order

Here, unlike chapter 6, the unknown plant has an infinite impulse response. FIR adaptive filter algorithms will be used, and therefore some noise due to undermodelling will be unavoidable. The purpose of this section is to demonstrate the maximum depth of cancellation as a function of the filter order for each algorithm, and to determine which N provides the best trade-off between the rate and depth of convergence in cases where the input is stationary.

Though it is known that the noise due to undermodelling will limit cancellation, other factors in the environment have been found to limit the maximum cancellation depth to levels below those which would be prescribed by undermodelling noise [Birkett97]. These sources were discussed in section 5.2.2.

The reference and primary signals employed in this section were derived from broadband noise inputs recorded in MC2014. Three algorithms were compared: MA-NLMS, the AR-SOX-AF, and the PWP-SOX-AF. In each case, filter orders of 128, 256, 512, and 1024 were employed. The results for each algorithm are presented in figures 7.1, 7.2, and 7.3 respectively.

FIGURE 7.1: Steady-state broadband ERLE for the time-domain MA-NLMS filter in MC2014. Faster initial convergence is obtained at the cost of greater misadjustment for $N=128$ (dotted), 256 (dashdotted), 512 (dashed), and 1024 (solid).

FIGURE 7.2: Steady-state broadband ERLE for the AR-SOX-AF in MC2014. Note the convergence rates are less dependent on N than in the MA-NLMS filter. The SOX-AF, however, seems to encounter difficulty with higher-order systems. In the plot, $N=128$ (dotted), 256 (dashdotted), 512 (dashed), and 1024 (solid).

FIGURE 7.3: Steady-state broadband ERLE for the PWP-SOX-AF in MC2014. The convergence rates are almost independent of filter order for $N=128$ (dotted), 256 (dash-dotted), 512 (dashed), and 1024 (solid).

As expected, higher-order filters generally achieve deeper cancellation, regardless of which algorithm is employed. An anomaly appears in the AR-SOX-AF, where the steady state depth of cancellation for $N=1024$ is noticeably lower than that for $N=512$. The orthogonal filters achieve slightly better cancellation depth in the steady state than MA-NLMS filters of the same order in this experiment. It is significant that the PWP-SOX-AF achieves better steady-state cancellation than the AR-SOX-AF in every trial shown.

The rate of convergence in the MA-NLMS algorithm decreases as the filter order is increased. Although convergence rates in the AR-SOX-AF are noticeably faster than those of MA-NLMS¹, they remain dependent on N to some extent. In the PWP-SOX-AF, however, the convergence rate is almost independent of the filter order, and these rates are faster generally than those of even the AR-SOX-AF.

1. It is not really proper to compare convergence rates here, as no attempt was made to ensure the same convergence trajectories were traversed in each algorithm.

These results are a little surprising in that the MA-NLMS algorithm should perform as well as any orthogonal algorithm when the reference input spectrum is flat. However, in these trials, the reference signal had been passed through an elliptical band-pass filter, so its distribution was not truly uniform. The filter order N was chosen to be 512 in the remaining experiments presented in this chapter based on these results. This value of N appeared to provide the best combination of initial convergence rate and steady-state cancellation in each of the algorithms under test.

7.2 Experimental Notes - Board Room MC2014

MC2014 at Carleton University is a fairly large, ordinary conference room. A brief description of it is provided in section A2.3 of appendix A2. The set was placed at one end of the table so as to radiate sound directly to the longitudinal majority of the room.

The results are presented in a similar format to those of chapter 6. Throughout section 7.3, great care was taken to keep the room transfer function fixed so as to allow the signal identification portion of the adaptive filter to be studied in the presence of different reference inputs. In section 7.4, severe variations lasting about three seconds were made near to the set after two seconds of stillness. The variations were effected by moving a hard-bound notebook in an erratic manner about 30 cm above the set on-axis with its loudspeaker. These results will demonstrate the behaviour of the plant identification stage of the different filter structures.

7.3 Signal Identification Stage - MC2014

7.3.1 Procedures

The stationary stochastic sequences used in this section were generated using MATLAB. They and the synthetic speech recordings were filtered in MATLAB and transferred to a portable DAT recorder using the procedures described in section 5.4.1. The apparatus was connected as shown in figure 5.3. All accessible interference sources (air conditioning equipment, fluorescent lights,

and the like) were suppressed to the greatest extent possible. The ambient noise measured in the neighbourhood of the set was 56 dB SPL.

The test signals were applied in turn to the experimental apparatus. Movements in the room were not allowed while these recordings were being made. Real-time recordings of the reference and primary signals were made using the ADAT recorder, and the files were transferred to a PC using the methods of section 5.4.3. All of the filter structures were implemented in MATLAB.

Three algorithms were tested on the reference and primary data recorded for each study. Re-use of the same files in each section allows direct comparisons to be made between their outputs. The parameters of each filter were set to their preferred values as discussed in sections 4.2 and 6.3. In all cases, the filter orders were fixed at $N=128, 256, 512$, and 1024 .

In section 7.3.2, several stationary input sequences are studied, each having a different spectral distribution. The ability to track abrupt changes in reference power is examined in section 7.3.3. As in section 6.6.1, broadband excitation was used for this experiment, and after four seconds its amplitude was alternately doubled and halved in subsequent 1-second intervals. The same AR2 sequence specified in table 6.1 was used in section 7.3.4 to determine how well the algorithms track step-changes in PSD in true acoustic situations. Finally, the results for the continuously non-stationary synthetic speech sequences are presented in section 7.3.5.

7.3.2 Stationary Excitation

Results for broadband stationary inputs were presented in section 7.1, so they are not repeated. Figure 7.4 (a) shows the convergence plots for an AR2(-1.5955,.95) input and the MA-NLMS algorithm. The PWP-SOX-AF is compared to MA-NLMS in figure 7.4 (b). The PWP-SOX-AF offers slightly better steady-state performance than the AR-SOX-AF, and both the algorithms achieve more than 3 dB deeper cancellation than MA-NLMS for this excitation.

Results from similar experiments with the AR12 process are shown in figures 7.5 (a) and 7.5 (b). Here, the AR-SOX-AF is unable to match the cancellation afforded by MA-NLMS, as was the case for the broadband input shown in figure 7.2. The PWP-SOX-AF appears to offer the same steady state performance as MA-NLMS, but its initial convergence rate is clearly superior.

FIGURE 7.4: AR2(-1.5955,.95) excitation applied to the three LMS filters. In (a) the MA-NLMS algorithm (dotted) and the AR-SOX-AF (solid), and in (b) the MA-NLMS algorithm (dotted) and the PWP-SOX-AF (solid) in MC2014 are shown.

FIGURE 7.5: AR12 excitation applied to the three LMS filters. In (a) the MA-NLMS algorithm (dotted) and the AR-SOX-AF (solid), and in (b) MA-NLMS (dotted) and the PWP-SOX-AF (solid) in MC2014 are shown.

7.3.3 Step Change in Variance

The purpose of this experiment is to determine how well the algorithms track abrupt changes in reference input power in a practical situation. The results are shown for the three algorithms in figures 7.6 (a) and 7.6 (b). Here, no divergence was seen when the reference input power was suddenly increased, which was not the case in section 6.6.1. On the contrary, slight divergence appears for all the algorithms at the *trailing* edge of the burst, when the reference power drops abruptly. The divergence is about 3 dB for the SOX-AF's and about 4 dB for the MA-NLMS algorithm. Though it is quite apparent visually, this trailing divergence may be quite inaudible because:

- it is of very short duration, which suggests the filter coefficients do not deviate too far from their steady-state values;
- it occurs when the signal power is rapidly decreasing, so the error signal amplitude during this period is likely not increasing, making it difficult for the ear to perceive.

Both of the SOX-AF's achieve deeper cancellation than the MA-NLMS algorithm, whether the filter is in a transient state or in the steady state.

FIGURE 7.6: Tracking capability of three algorithms with broadband excitation whose variance is changed sharply in MC2014. In (a), the MA-NLMS algorithm (dotted) and AR-SOX-AF (solid) are compared. In (b) the MA-NLMS (dotted) and PWP-SOX-AF (solid) are shown. In these trials, $N=256$; the response was slightly slower for $N=512$.

7.3.4 Step Change in Distribution

It was shown in section 6.6.2 that sharp discontinuities in the reference PSD can cause serious divergence in orthogonalized adaptive filters. Here, the experiment is repeated with the same excitation in an acoustic context.

Figure 7.7 (a) shows the AR-SOX-AF suffers temporary divergence relative to the MA-NLMS algorithm, but not nearly to the same extent as in the (noise-free) simulation. The divergence is most apparent after the sixth and eighth seconds, where the PSD transition is from lower- to higher band-pass. Note that the AR-SOX-AF performs better when the reference distribution is lower pass (between the fifth and sixth seconds and the seventh and eighth seconds) than higher pass (between the sixth and seventh seconds and the ninth and tenth seconds).

The PWP-SOX-AF offers the better of both algorithms, providing cancellation similar to the AR-SOX-AF during lower pass periods yet matching the MA-NLMS algorithm during higher pass periods. Its only downfall is that it too exhibits divergence, but it is not significantly worse than that in MA-NLMS. The orthogonal filters are compared explicitly in figure 7.7 (b).

FIGURE 7.7: Tracking capability of the three algorithms with the AR2 sequence whose PSD is changed sharply in MC2014. In (a), the MA-NLMS algorithm (dotted) and the AR-SOX-AF (solid) are compared. In (b) the AR-SOX-AF (dotted) and PWP-SOX-AF (solid) are compared.

7.3.5 Speech Bursts

The study now turns from short-term WSS inputs to speech, whose statistical nature changes continuously. The results for simulated female and male speech are presented in figures 7.8 and 7.9 respectively. In each figure, the first few bursts are shown along with a close-up of a single burst once convergence was complete. No further convergence was achieved after 25 seconds. There is little difference between the AR-SOX-AF and the MA-NLMS algorithms in these trials, however the PWP-SOX-AF delivers consistently better performance than both.

FIGURE 7.8: Comparison of the three algorithms when tracking the SSGF2 speech input file in MC2014. In (a), the first few bursts are shown. In (b) the region of a single burst after steady-state is achieved is expanded. No further convergence was achieved even after 25 seconds. In all cases, the traces shown are for the MA-NLMS algorithm (dotted), the AR-SOX-AF (dashed), and the PWP-SOX-AF (solid).

FIGURE 7.9: Comparison of the three algorithms when tracking the SSGM1 speech input file in MC2014. In (a), the first few bursts are shown. In (b) the region of a single burst after steady-state is achieved is expanded. In all cases, traces shown are for the MA-NLMS algorithm (dotted), the AR-SOX-AF (dashed), and the PWP-SOX-AF (solid).

7.3.6 Analysis

The PWP-SOX-AF has demonstrated better depth of cancellation than both the conventional AR-SOX-AF and the MA-NLMS algorithm in virtually all cases where the acoustic transfer function was held constant. This result was not completely unexpected, as similar observations were made when these same algorithms were compared using simulated FIR transfer functions.

The results here are less dramatic than in chapter 6, and this is due to certain elements of the acoustic environment. For example, divergence of the AR-SOX-AF was not observed when the reference power was increased sharply in section 7.3.3. This may be because the loudspeaker has a low-pass response characteristic, and is unable to propel such a sharp transient in the acoustic domain, allowing time for the signal identification stage to reduce the channel step-sizes before substantial divergence occurs.

Similarly, the divergence due to the sharp changes in reference spectral power distribution were of less significance than might be suggested by the results presented in section 6.6.2. As before, it may be the case that the speaker cushions the abrupt changes in channel power, leaving time for the channel step-sizes to be reduced before the acoustic power increase is detected by the microphone.

Finally, it is useful to revisit section 6.6.3, where the performance of these three algorithms was compared when the speech tracks were used to excite a static FIR transfer function. In that case, the filter order was 512, the plant order was 1200, and no extraneous noise nor non-linearities in the plant were present. The cancellation depth achieved in those experiments hovered around 40 dB for both the female and male inputs. Here, cancellation in the order of 30 dB is achieved, so some feeling for the performance penalty imposed by the acoustic telephony environment in this room can be gained. The maximum depth of cancellation in a practical system is limited by any of the factors listed in section 1.3, as those are physical properties of the telephone hardware and the environment in which the observations are made, rather than of the algorithms themselves.

7.4 Plant Identification Stage - MC2014

The purpose of the experiments outlined in this section is to compare the behaviour of the three algorithms as they track variations in the room transfer function, both when the reference input is stationary, and then again when it is non-stationary.

7.4.1 Procedure

The apparatus and methods used in this section were identical to those of section 7.3, however, a severe variation of about three seconds' duration was introduced after a short period of stability in the room transfer function. The room transfer function was left unaltered initially to allow time for the filters and signal identification stages to reach their steady-states. The variations were effected by moving a hard-bound notebook continuously in an erratic manner about 30 cm above the set on-axis with its loudspeaker for a period of about three seconds.

7.4.2 Observations

The plots shown in this section are centred deliberately around the period of disturbance in the room transfer function. Figures 7.10, 7.11, and 7.12 show the filter performance results for the broadband, AR2(-1.5955, .95) and AR12 excitation sequences respectively.

The MA-NLMS algorithm performs consistently better during the disturbance for broadband inputs, though its improvement over the PWP-SOX-AF is almost negligible. The MA-NLMS algorithm is unable to keep up with the SOX-AF's when the reference input power spectrum is non-uniform. It is also quite clear that the PWP-SOX-AF outperforms the AR-SOX-AF in these cases, particularly for the AR12 sequence.

When the speech inputs were applied, the PWP-SOX-AF again was superior to the others, and the MA-NLMS algorithm consistently underperformed. Those results are shown in figures 7.13 and 7.14 for the female and male tracks respectively.

FIGURE 7.10: Comparison of plant tracking capabilities of the three algorithms with WSS broadband excitation in MC2014. In (a) the MA-NLMS algorithm (dotted) and AR-SOX-AF (solid) are shown; in (b) the AR-SOX-AF (dotted) is and the PWP-SOX-AF (solid) are shown.

FIGURE 7.11: Comparison of plant tracking capabilities of the three algorithms with WSS AR2(-1.5955, 95) excitation in MC2014. In (a) the MA-NLMS algorithm (dotted) and the AR-SOX-AF (solid) are shown; in (b) the AR-SOX-AF (dotted) is compared to the PWP-SOX-AF (solid).

FIGURE 7.12: Comparison of plant tracking capabilities of the three algorithms using a WSS AR12 excitation signal in MC2014. In (a) the MA-NLMS algorithm (dotted) and AR-SOX-AF (solid) are shown; in (b) the AR-SOX-AF (dotted) and the PWP-SOX-AF (solid) are shown.

FIGURE 7.13: Comparison of plant tracking capabilities of the three algorithms with SSGF2 speech excitation in MC2014. In (a) the full series is shown. In (b) the region of a single burst during the room variation is expanded. In all cases, traces shown are for the MA-NLMS algorithm (dotted), the AR-SOX-AF (dashed), and the PWP-SOX-AF (solid).

FIGURE 7.14: Comparison of plant tracking capabilities of the three algorithms with SSGM1 speech excitation in MC2014. In (a) the full series is shown. In (b) the region of a single burst during the room variation is expanded. In all cases, traces shown are for the MA-NLMS algorithm (dotted), the AR-SOX-AF (dashed), and the PWP-SOX-AF (solid).

7.4.3 Analysis

The room variation that was applied was very close to the set, so one would anticipate that the entire room transfer function (including the elements of lowest order) would be in a state of flux while the variation was applied. In cases where the input is WSS and broadband, one would not expect the orthogonal structures to offer significant advantages over the MA-NLMS algorithm, because all convergence modes are traversed at nearly equal rates prior to the transformation. The orthogonal filters offer much better performance, however, when the excitation is narrow-band, as shown in figures 7.11 and 7.12. In those cases, the reference was WSS, allowing the signal identification stage to remain in a steady-state while the room variation was unfolding.

The most complicated case - and that of greatest interest in practice - occurs when a speech input is applied to a time-varying plant, as shown in figures 7.13 and 7.14. In this situation, the signal identification stages in the orthogonal algorithms are tracking changes in the reference input while the plant identification stage tracks the changing room transfer function. Even though this second level of adaptation is adopted, the performance of the orthogonal structures - particularly that of the PWP-SOX-AF - is better than that of the MA-NLMS algorithm at all times.

7.5 Experimental Notes - Small Office ME4144

A set of recordings similar to those made in MC2014 were made in a small graduate student office, room ME4144 in the MacKenzie Engineering Building at Carleton University. A diagram and description of this room is provided in Appendix A2.

Only a small subset of those results are presented here in the interest of brevity. The results obtained when the room transfer function was constant were almost identical to those obtained in MC2014, so they have been omitted entirely. The disturbance applied in ME4144, however, was much less severe than in MC2014. The recording period commenced with no one in the room and with the door closed. After approximately two seconds of stillness, the door was opened fully, a person entered the room, closed the door, and walked quietly towards the set. This event sequence is a typical room variation which can be expected to occur in practice.

7.6 Plant Identification Stage - ME4144

7.6.1 Procedure

The same excitation sequences used in MC2014 were also employed in ME4144. All recording equipment was kept outside the room. The ambient noise level measured in the vicinity of the set was 62 dB SPL in this room. Files were transferred to the PC as described in section 5.4.3, and off-line processing was carried out with MATLAB. As before, the PWP-SOX-AF was compared to the AR-SOX-AF and to the MA-NLMS algorithm.

7.6.2 Observations

Results are presented for WSS broadband and AR2(-1.5955, .95) inputs in figures 7.15 and 7.16, and for the synthetic female and male speech segments in figures 7.17 and 7.18 respectively.

In the case of the stationary inputs, the point at which the door is opened is readily identified at the left hand side of the plots, where there is a sudden decrease and recovery in ERLE. In the broadband case, the MA-NLMS algorithm exhibits similar capability to the AR-SOX-AF, but

they are both consistently outdone by the PWP-SOX-AF. This result is a little surprising given that the MA-NLMS algorithm was shown in figure 7.10 to provide the best results under similar circumstances in MC2014. The PWP-SOX-AF was also the best performer in the case of the AR2 excitation. Here, the limitations of the MA-NLMS algorithm are clearly evident.

The plots for the female and male speech tracks relate the ERLE of the three algorithms while the subject is walking across the room. Once again, the orthogonal algorithms outperform the MA-NLMS algorithm, and the best performance is achieved with the PWP-SOX-AF.

7.6.3 Analysis

The room variation tracked in this section was much less severe than that of section 7.4, so the cancellation depth remains relatively deep throughout the room variation. The variation occurred at a much larger distance from the set than in section 7.4 also, which suggests that the lower-order elements of the filter were not required to deviate too far from their previous steady-state values.

The orthogonal algorithms achieve significantly deeper cancellation when the input sequences are WSS because their signal identification stages are not required to track the reference input. The eigenmodes of the adaptive filter remain well-equalized, allowing the filter to track the plant evolution very effectively.

When speech inputs are used, the advantages of the orthogonal filters are less apparent, as both stages of the filter are evolving continuously. Even so, the PWP-SOX-AF consistently delivers depths of cancellation which are not attainable by the other algorithms. This is most likely due to the ability of the PWP estimation scheme to reduce the step-size instantly as the channel power increases.

FIGURE 7.15: Comparison of plant tracking capabilities of the three algorithms with broadband excitation in ME4144. In (a) the period during and after the door is opened is shown. In (b) the period during which the door is opened and closed is expanded. In all cases, the traces shown are for the MA-NLMS algorithm (dotted), the AR-SOX-AF (dashed), and the PWP-SOX-AF (solid).

FIGURE 7.16: Comparison of plant tracking capabilities of the three algorithms with AR2(-1.5955, 95) excitation in ME4144. In (a) the period during and after the door is opened is shown. In (b) the period over which the door is opened and closed is expanded. In all cases, the traces shown are for the MA-NLMS algorithm (dotted), the AR-SOX-AF (dashed), and the PWP-SOX-AF (solid).

FIGURE 7.17: Comparison of plant tracking capabilities of the three algorithms with the SSGF2 speech input in ME4144. In (a) the full sequence after the door is opened is shown. In (b) the period during which the door is opened and closed is expanded. In all cases, the traces shown are for the MA-NLMS algorithm (dotted), the AR-SOX-AF (dashed), and the PWP-SOX-AF (solid).

FIGURE 7.18: Comparison of plant tracking capabilities of the three algorithms with the SSGM1 speech input in ME4144. In (a) the full sequence after the door is opened is shown. In (b) the period during which the door is opened and closed is expanded. In all cases, the traces shown are for the MA-NLMS algorithm (dotted), the AR-SOX-AF (dashed), and the PWP-SOX-AF (solid).

7.7 Numerical Summary

This section provides a numerical review of the results presented in sections 7.3 through 7.6. For each algorithm, the mean ERLE achieved in a matching interval of the experiments is provided for each filter order. These allow the influence of filter order on the steady-state cancellation depth to be evaluated for each algorithm. They also facilitate comparisons between the three algorithms for a given filter order.

An “average” is computed for each N which provides some measure of the performance one might expect to achieve for arbitrary inputs. The term “average” is used very loosely here: it is calculated as the mean of the entries above it in each column, however, it combines only a very small sample of possible distributions, some of which are WSS, some of which are not. These “averages” are tabulated at the end of the section for quick reference.

TABLE 7.1 Measured steady-state ERLE for the MA-NLMS algorithm.

Figure	Input	Interval	$N=128$	$N=256$	$N=512$	$N=1024$
7.1	Broadband	9-10 sec	18.54	23.60	28.30	30.05
7.4	AR2	9-10 sec	27.33	27.50	26.74	25.46
7.5	AR12	9-10 sec	25.82	27.99	28.10	25.88
7.8	SSGF2	9-11 sec	25.60	25.88	25.80	24.89
7.9	SSGM1	9-11 sec	25.23	26.87	27.83	27.00
-	“Average”	-	24.50	26.37	27.35	26.66

TABLE 7.2 Measured room-tracking ERLE for the MA-NLMS algorithm.

Figure	Input	Interval	$N=128$	$N=256$	$N=512$	$N=1024$
7.10	Broadband	3-7 sec	16.85	13.28	9.28	5.81
7.11	AR2	3-7 sec	19.10	18.87	18.33	17.82
7.12	AR12	3-7 sec	12.48	11.26	10.64	10.08
7.13	SSGF2	9-11 sec	17.75	16.95	16.39	15.92
7.14	SSGM1	9-11 sec	13.88	12.48	11.19	9.51
-	“Average”	-	16.01	14.57	13.17	11.83

TABLE 7.3 Measured steady-state ERLE for the AR-SOX-AF.

Figure	Input	Interval	N=128	N=256	N=512	N=1024
7.2	Broadband	9-10 sec	19.93	24.73	28.50	26.77
7.4	AR2	9-10 sec	28.21	27.23	27.35	25.70
7.5	AR12	9-10 sec	26.58	27.26	23.10	22.34
7.8	SSGF2	9-11 sec	28.14	27.81	25.88	22.54
7.9	SSGM1	9-11 sec	25.48	27.65	27.40	24.67
-	"Average"	-	25.67	26.94	26.45	24.40

TABLE 7.4 Measured room tracking ERLE for the AR-SOX-AF.

Figure	Input	Interval	N=128	N=256	N=512	N=1024
7.10	Broadband	3-7 sec	15.06	11.44	8.09	5.59
7.11	AR2	3-7 sec	21.09	20.31	19.76	18.83
7.12	AR12	3-7 sec	20.57	17.74	14.36	11.99
7.13	SSGF2	9-11 sec	20.85	19.69	18.69	17.18
7.14	SSGM1	9-11 sec	16.66	15.56	13.33	11.48
-	"Average"	-	18.85	16.95	14.85	13.01

TABLE 7.5 Measured steady-state ERLE for the PWP-SOX-AF.

Figure	Input	Interval	N=128	N=256	N=512	N=1024
7.3	Broadband	9-10 sec	19.82	25.11	29.79	30.44
7.4	AR2	9-10 sec	28.42	29.15	28.86	27.70
7.5	AR12	9-10 sec	26.71	28.99	28.39	23.63
7.8	SSGF2	9-11 sec	28.46	28.36	28.40	26.34
7.9	SSGM1	9-11 sec	27.35	28.80	29.48	28.32
-	"Average"	-	26.15	28.08	28.99	27.29

TABLE 7.6 Measured room-tracking ERLE for the PWP-SOX-AF.

Figure	Input	Interval	N=128	N=256	N=512	N=1024
7.10	Broadband	3-7 sec	14.60	10.95	7.71	5.31
7.11	AR2	3-7 sec	23.10	21.57	20.44	19.48
7.12	AR12	3-7 sec	20.98	18.79	15.98	13.77
7.13	SSGF2	9-11 sec	22.24	20.76	20.02	18.69
7.14	SSGM1	9-11 sec	19.28	16.42	14.46	12.67
-	"Average"	-	20.04	17.70	15.72	13.98

TABLE 7.7 Summary of “averaged” results for the stationary plant.

Algorithm	$N=128$	$N=256$	$N=512$	$N=1024$
MA-NLMS	24.50	26.37	27.35	26.66
AR-SOX-AF	25.67	26.94	26.45	24.40
PWP-SOX-AF	26.15	28.08	28.99	27.29

TABLE 7.8 Summary of “averaged” results for the time-varying plant.

Algorithm	$N=128$	$N=256$	$N=512$	$N=1024$
MA-NLMS	16.01	14.57	13.17	11.83
AR-SOX-AF	18.85	16.95	14.85	13.01
PWP-SOX-AF	20.04	17.70	15.72	13.98

The significance of these results should not be overlooked: The PWP-SOX-AF achieves better cancellation than the best-case MA-NLMS and AR-SOX-AF algorithms, whether the plant is static or time-varying, when the excitation is stationary. Its advantage is compromised somewhat when the input is non-stationary. This is probably because the non-stationarities applied in MC2014 were very serious and perhaps non-linear, limiting the depth of convergence which could be expected without modelling the non-linearity explicitly. On balance, the PWP-SOX-AF seems to offer consistent improvement over the known structures.

7.8 Summary

The most significant results of the thesis were presented in this chapter. These results represent filter performance in true acoustic environments, where many complicated factors ensure that complete cancellation cannot be achieved. Three algorithms were compared: the MA-NLMS algorithm with $\mu_g=1$, the AR-SOX-AF with $\mu_g=1/N$ and $\beta=1-1/N$, and the PWP-SOX-AF with $\mu_g=1/N$ and $L/l=16$.

In section 7.1, the steady-state performance of the three algorithms was compared for four filter orders. In each case, the PWP-SOX-AF achieved depths of cancellation that were unreachable

by the other algorithms. The experiments were duplicated in section 7.3 for two narrow-band excitation sequences, and again the PWP-SOX-AF was found to provide the best performance.

The ability of the orthogonal algorithms to track changes in the excitation were also examined in section 7.3. The lack of divergence when abrupt changes to the reference signal statistics were applied were attributed the mechanical inertia of the loudspeaker and other factors implicit in the acoustic domain. The PWP-SOX-AF provided superior performance to the other two algorithms in the transient *and* steady-state phases, suggesting that the method embodies the better features of each. It was also found to achieve cancellation levels that were consistently deeper than the others when tracking speech inputs.

The plant identification capabilities of the algorithms were tested in sections 7.4 for a severe plant variation in a large conference room, and for a relatively minor variation in a small office. Though no improvement over the MA-NLMS algorithm was achieved when the input was WSS and broadband, the PWP-SOX-AF outperformed the MA-NLMS algorithm by up to 5 dB and the AR-SOX-AF by up to 3 dB when the input was WSS and narrow-band. Similar benefits were observed when the speech tracks were examined.

These experiments were duplicated in section 7.6 for the small office. Even though the room variation was less severe, the PWP-SOX-AF provided improvements in both steady state and transient performance comparable to those observed in the large conference room. A numerical summary of the average cancellation depths of each algorithm was given in section 7.7, which allowed quick comparisons to be made of the algorithms under various operating conditions. In all cases, the PWP-SOX-AF was found to provide better cancellation than the AR-SOX-AF and the MA-NLMS algorithm, whether or not the excitation was WSS, whether or not the room transfer function was static, all regardless of the filter order.

8

Conclusions

8.1 Review of Thesis Objectives

This manuscript was prepared with several objectives in mind. The first, and the simplest, was to outline the intricate nature of the acoustic echo cancellation problem, and to provide the reader with an appreciation for its nuance and immensity. This was achieved largely in the opening chapter, with supporting arguments given in the first half of chapter 4.

The second objective was to summarize the algorithms which have been proposed and employed by previous researchers to treat system identification problems such as AEC. The fundamentals of LMS filters were reviewed in the first halves of chapter 2 and Appendix A 1. The latter half of chapter 2 was intended to outline several well-reasoned and ad-hoc methods for improving the performance of the original time-domain LMS algorithm.

Transform-domain LMS filters were introduced in chapter three as one well-founded method holding great promise for these improvements. Block-oriented algorithms were discussed first, as these were known in the literature long before the first sample-oriented filter algorithms were published. However, all subsequent attention in this research was directed to the sample-oriented method as the flat delay imposed by block sectioning is untenable in AEC. A concise description of the AR-SOX-AF was provided in the latter halves of chapter 3 and in appendix A 1.

The third objective of the manuscript was to examine the performance of the known SOX-AF algorithms with a critical eye, and to determine how well a SOX-AF might fare in the context of a complicated real-time acoustic modelling problem. Early experiments revealed that the known AR- and MA-SOX-AF's suffer stability problems when the spectral character of the reference

input varies rapidly - clearly, a severe limitation for AEC. It was also found that these two-stage linear filters are very sensitive to (arbitrary) initial conditions. Very quickly, the third objective ripened into one of creation: to propose and to evaluate solutions to these problems and to verify their usefulness through experiment.

8.2 How the Objectives Were Met

The sensitivity of the SOX-AF to initial conditions was fairly easy to address because a direct cause-and-effect relationship exists between the initialized power values, the choice of β , and the delay in the onset of convergence. The method of Dynamic Step-size Initialization formed a natural solution to the problem, though a rigorous analysis of it appears to be quite difficult. However, this solution is isolated from the greater problem of filter divergence because it makes itself irrelevant shortly after start-up.

A quick examination of the SOX-AF literature revealed that considerable efforts had been made to find the “best” transform for the structure. Several researchers showed through analysis and experiment that no optimal fixed transform exists for inputs with general spectral distributions, and that the transform by itself is unable to improve convergence rates or misadjustment. As the behaviour of the LMS stage was well-understood and known to be stable under all conditions in those early experiments, it seemed likely that neither the transform or the LMS stage were responsible for observed stability problems.

A more critical examination of the literature made it apparent that very little was known about the importance of the power normalization procedure. This lack of enlightenment provided the motivation for the detailed inquiry into the role of β provided in chapter 4. The key contributions of this work came about directly from that inquiry. As all of the known SOX-AF's derive their step-sizes from time-averaged power estimates, it was reasonable to conclude that the power estimator had a direct influence on the stability of the overall structure.

At first, the focus of this research was directed to the search for an optimal choice of β . It was soon realized that no optimal value for it exists, as it plays signal tracking responsiveness against

steady-state misadjustment. However, its preferred value was found to be approximately $1-1/N$. Even when this value of β was employed, however, the structure still exhibited divergence when abrupt changes in the amplitude or in the distribution were applied to the reference input. The estimator still seemed at fault, because the problem only occurred when under-excited modes were presented with sharp increases in power, even if those increases were relatively small.

It was concluded that all linear power estimators - be they MA or AR - necessarily lag the true power envelope of their inputs. After a sharp increase in the true power, the estimated power returns step-sizes which are momentarily outside of their stability bounds, and while they are the filter is unstable. At this point, it became clear that a non-linear method was required. The Peak-Window-Power estimator was crafted to address exactly this problem, and it was introduced at the end of chapter 4.

The PWP estimator has fortified the stability of the SOX-AF, and thereby treated one of the key outstanding problems in the SOX-AF literature. When it was first conceived, it was not expected to provide such reliable improvements in tracking capability and steady-state performance. That it does both simultaneously was completely unanticipated: such win-win situations are rare in adaptive filtering problems, where most every gain is mirrored by consequence. It is fortuitous that PWP estimates are obtained using only bit comparisons and shifts, as it facilitates a parallel hardware implementation that liberates the DSP to perform more complicated mathematical functions.

The remainder of the thesis was concerned with the verification of the proposed algorithms. A concise description of the methods, apparatus, and metrics used to evaluate them in simulated and genuine acoustic environments was provided in chapter 5. Experiments were orchestrated in chapter 6 to validate the conclusions reached in chapter 4 concerning the choice of linear power estimation parameters, and claims that the PWP estimator addresses those limitations. Backed with those favourable results, the same algorithms were applied to legitimate acoustic system identification problems in chapter 7.

8.3 Summary of Experimental Results

8.3.1 Dynamic Step-Size Initialization

The DSI procedure was shown to limit the amount of overshoot in the MSE shortly after start-up when its parameters were properly selected. A number of simulations were run in an effort to optimize the DSI parameters (the choice of β at $k = 0$ and the number of iterations D over which β is varied), but these optimal values remained elusive. It appears that the best choice of D is very dependent on the PSD of the reference input and the proximity of the initial power estimate vector to it. This suggests that no globally-optimal value for D exists. The value of β at $k = 0$ appears to be much less important, though DSI appears to be more reliable when it is made very small. An unanticipated benefit of the PWP estimator is that it eliminates the need to dismiss previous power estimates completely, and therefore it renders DSI redundant.

8.3.2 Preferred Values of Linear Parameters

The best choice for μ_g was found to be one for the MA-NLMS filter. For the AR-SOX-AF, $\mu_g = 1/N$ and $\beta = 1 - 1/N$ were found to provide the best convergence rates once initialization effects had been discharged. For two reasons, convergence and tracking abilities were deemed more important than misadjustment in this application: first, the excitation signals and the plant evolve continually in AEC, so steady-state operation is rare; second, factors other than filter order limit the depth of convergence achievable in practical AEC problems below that which would be imposed by misadjustment in the algorithm.

8.3.3 Misadjustment

Even so, it was important to compare the misadjustment of the preferred linear algorithms and the proposed PWP-SOX-AF to ensure no anomalies were introduced by either. This assessment could only be made under the tightly-regulated noise-free conditions upheld in chapter 6. The linear algorithms achieve the same cancellation depth when the FIR plant approximates a typical acoustic transfer function, unless the filter order is very close to the full length of the plant, in which case the MA-NLMS algorithm excels. The PWP-SOX-AF consistently achieves 2-3 dB

deeper cancellation than either linear algorithm regardless of filter order. It appears that this improvement is due to the smaller average step-size produced by the PWP estimator.

8.3.4 Signal Identification Capability

Orthogonal algorithms offer the promise of uniform convergence across the eigenmodes of the filter, but this promise carries with it a certain risk: if the power in each eigenmode is not tracked properly, the filter may become unstable. Indeed, the AR-SOX-AF was seen to suffer instability when sharp increases in reference power were detected, even when the preferred parameters were selected. This was demonstrated most clearly in section 6.6.2.

The PWP estimator is delayless, so it never exhibits instability. However, its benefits may be not be so apparent in AEC, because elements in the acoustic transfer function (most notably the loudspeaker) also feature a limited frequency response, reducing the discrepancy between the estimated and measured primary signals during sharp transitions in the reference input power. Experiments on actual data presented in sections 7.3.2 and 7.3.3 showed that the PWP-SOX-AF appears to provide the better behaviour of the MA-NLMS algorithm and the AR-SOX-AF when tracking abrupt changes in the input. On the average, the advantage is on the order of 1.6 dB over the MA-NLMS algorithm and 2.5 dB over the AR-SOX-AF for $N=512$ once the filter has reached its steady state.

8.3.5 Plant Identification Capability

All of the algorithms tested use similar LMS plant-tracking mechanisms. Thus, differences in the plant tracking capability of the three algorithms with stationary excitation can be attributed, at least to some extent, to the signal identification stage. The orthogonal algorithms equalize the power spectral density of the LMS input sequences, so they should provide significant benefits when it is necessary to track changes in the plant. Again, the advantages of the orthogonal filters are most obvious in the simulated environments, as shown in section 6.7.

When exposed to actual acoustic signals, the PWP-SOX-AF provides cancellation depths that are consistently deeper than those of the published algorithms, even when speech sequences are used to excite the system. During the tracking phase, the PWP-SOX-AF achieves about 1 dB

deeper cancellation than the AR-SOX-AF regardless of filter order. Its improvement over the MA-NLMS algorithm varies between 2 dB for $N=1024$ and 4 dB for $N=128$ on the average.

8.4 Suggestions for Further Research

8.4.1 Modified Structures

The method of weighted recombination in the transform domain discussed in section 4.3.2 appears to hold some promise for future researchers. Its intent is to reduce the order of the orthogonal adaptive filter without reducing the length of the observation vector. It will reduce the spectral resolution of the filter, but much of this resolution is redundant as typical inputs to AEC filters are highly correlated. Filters of lower order were shown to provide better tracking capability in chapter 7.

A few of the fundamental outstanding questions are:

- What is the analytical framework supporting such a structure?
- What is the preferred method to segregate the channels of recombination?
- How should the recombined channel step-sizes be chosen or estimated?

8.4.2 Linear Systems

A channelized adaptive- β scheme was outlined in section 4.4.3. The method is doubly-recursive in the sense that it varies the β in each channel according to the derived step-size in that channel, thereby equalizing the tracking properties in each stage. It may improve the responsiveness of the otherwise static linear estimator when the channel power is very low, which may address the divergence problem. On the other hand, it should reduce the step-size variance when the channel power is high, which should reduce misadjustment.

Such a structure was built during the course of this research, however it was found to be very temperamental. Very close attention must be paid to the algorithm's stability as the inversion of

the power estimate will produce very large step-sizes in channels with low power. Further, it appears that the analytical support of this structure is very complicated.

Some questions which should be asked early on in the research would include:

- Is there an analytical basis for equalizing the time constants in both stages?
- What are the stability bounds of the algorithm parameters?
- Is leakage in the transform an important consideration?
- What is the significance of the out-of-band aliased energy in the channel power sequences?

8.4.3 Non-Linear Systems

Now that the stability of the PWP estimator is known, the next objective should be to derive a set of analytical expressions which describe its convergence rate and misadjustment for broadband WSS excitation. It may be possible to extend these expressions to cover arbitrary distributions that are WSS, and then to non-stationary inputs, though the last step may require much more thought.

Two simpler and inviting topics might be:

- Is there an equivalent PWP-NLMS algorithm in the time-domain?
- Will the PWP estimator provide benefits of similar value in the BOX-AF and the GMDF?

Finally, with regard to the PWP-SOX-AF described herein,

- Should the past peak powers be time-weighted to make the estimator more responsive when the channel power is decreasing?
- What are the preferred values of $L1$ and $L2$? In particular,
 - 1) Should they be the same for all channels?
 - 2) Should they be static or time-varying?

These, and other questions I have not yet encountered, will be left to future researchers.

8.5 Summary of Contributions

- A set of preferred control parameter values for the AR-SOX-AF of general order;
 - The successful incorporation of this structure into an acoustic system identification problem;
 - A general method to accelerate the initial convergence of the AR- and MA-SOX-AF's;
 - The introduction of and justification for the non-linear Peak-Window-Power estimator which stabilizes the SOX-AF at all times;
 - Demonstration of the improved tracking capability and depth of cancellation afforded by the PWP-SOX-AF over both the MA-NLMS and AR-SOX-AF methods.
-

A1

A Review of Adaptive Filter Theory

This appendix provides a description of the mathematical foundations upon which the theory of adaptive filtering is based. Though it is not exhaustive, it should provide readers unfamiliar with the subject an introduction to the key properties of and differences between time- and transform-domain adaptive filters.

A1.1 Elements of Digital Filters

The building blocks of digital filters are numerical: adders (rather, accumulators), multipliers, and the element of unit delay. The unit delay in a digital system is the smallest time increment that can be resolved, and is equal to the sampling period of the input to the filter. Discrete time sequences can be filtered (through the process of *convolution*) or otherwise processed through repeated application of these functions. *Digital Signal Processors* (DSP's) are microprocessors optimized to multiply, accumulate or otherwise manipulate numbers from different data streams. Many modern DSP's are able to multiply two numbers together, add the product to a third, and load a fourth from external memory all in a single instruction. In fact, the advent of DSP was due largely to the need of an efficient convolution engine.

Digital filters fall into two general categories: A *Finite Impulse Response* (FIR) filter processes elements of the input sequence only, so its transfer function contains only zeros. Accordingly, the output of any FIR filter is guaranteed to be zero if all of its inputs are zero. The terms FIR and *Moving Average* (MA) are synonyms of one another, though the latter usually implies that the filter elements are unweighted (i.e. equal), making the filter a *rectangular window*.

An *Infinite Impulse Response* (IIR) filter is a recursive structure which synthesizes its output from a combination of at least one past output value, (thus its transfer function has at least one pole) and possibly a combination of the present or past input values (corresponding to transfer function zeros). An IIR filter with no zeros is said to be an *Auto-Regressive* (AR) process; one with one or more zeros is said to be an *Auto-Regressive Moving-Average* (ARMA) process. The output of any IIR filter need not be zero even if all elements of the input sequence are zero.

TABLE A1.1 Some characteristics of static FIR and IIR filters.

Property	FIR Structures	IIR Structures
Transfer Function Zeros	At least one not at the origin	Possibly one or more
Transfer Function Poles	None other than at the origin	At least one not at the origin
Stability	Always	If all poles inside unit circle
Computational Efficiency	Lower	Higher
Phase	Linear, if symmetrical	Not linear, in general

Adaptive filters may be FIR or IIR structures, though the latter are substantially more difficult to control because of the need to quantify the poles and to constrain all of them to lie within the unit circle. The causality of the filter must also be considered. Some features of both topologies are compared in table A1.1. Only FIR adaptive filters (or *adaptive linear combiners* [Widrow85]) will be considered in this manuscript, since they are inherently stable and easy to build and adapt. This is reflective of the adaptive filter literature in general, where applications requiring IIR filters can be very complicated and are quite rare. For those interested, descriptions of the theory and properties of IIR adaptive filters are available in [Mikhail86] [Mikhail89] [Nayeri89] [Johns91] [Mboup91] and [Kim95] for transform domain adaptive filters. FIR methods have become ubiquitous workhorses in time-varying digital filter applications.

TABLE A1.2 Some characteristics of adaptive FIR and IIR filters.

Property	FIR Structures	IIR Structures
Global Minimum	Exactly one	At least one
Local Minima	One - the global minimum	Possible in multi-pole systems
Adaptation Ease	Straight-forward	Complicated

A1.2 Adaptation of Digital Filters

Analog filters are ill-suited to adaptation because their elements are physical and static. Digital filters, on the other hand, are easy to adapt because their elements are purely numerical. In this section, the mechanisms of adaptive digital filters are outlined so as to provide the reader with an intuition for the mathematical development which follows. Here the objectives of the adaptive filtering algorithm are described in the context of the system identification problem.

A1.2.1 Objectives

The objectives of the system identification problem are simply stated, but not easily achieved:

- detect some input sequence which is incident on the plant to be identified;
- detect the response of the plant to the input sequence;
- find some linear combination of the input sequence which predicts the plant response;
- adjust the linear combination to track changes in the plant.

The linear combination sought is the optimal digital filter, w_{opt} . In most practical situations, the search for w_{opt} is constrained by the properties of the excitation signals, the unknown plant, and the requirements of the system as a whole. In general, it must be assumed that:

- the reference input signal is narrow-band and non-stationary;
- the impulse response of the plant is very (infinitely) long, and subject to change over time;
- the plant is composed of linear and non-linear elements;
- the signals are contaminated with time-varying noise that may or may not be correlated with the reference signal.

Further, part of the primary signal may be intended to pass through the echo canceller unaltered. In the case of AEC for example, the microphone signal may contain (unwanted) echos of far-end speech and (the desired) near-end speech simultaneously. When this happens, the system is said to be in *double-talk mode*. It must be assumed that that the near-end and far end signals occupy the same bandwidth, so the adaptive filter must be able to segregate and cancel the echo signals

and noise while transmitting unaltered the desired near-end speech. It is often assumed that the near-end and far-end signals are uncorrelated, so the objective of the adaptive filter, then, is to track and cancel any artifacts in the primary input that are correlated with the reference.

A1.3 Some Pertinent Results from Matrix Theory

A survey of the mathematical tools required to describe the operation of the adaptation analysis is presented in this section.

A *symmetric matrix* is one which is unchanged through transposition. For any symmetric matrix A , the function $f = \mathbf{h}^T A \mathbf{h}$ is *quadratic* in the N dimensions of vector $\mathbf{h}$. This ensures f has either a global maximum or a global minimum [Strang76] (pages 231-236) and that f is free of any local maxima or minima, because no $\mathbf{h}$ terms in f are of greater than second order.

If

$$f \geq 0 \quad \forall (\mathbf{h} \in \mathcal{R}^N), \quad (\text{A1.1})$$

then A is said to be *non-negative definite*¹, and its global minimum is located at a unique non-negative point or along a continuum of colinear points $\mathbf{h}_{opt} \in \mathcal{R}^N$. If

$$f > 0 \quad \forall (\mathbf{h} \neq \mathbf{0}) \in \mathcal{R}^N, \quad (\text{A1.2})$$

then A is *positive-definite* [Strang76] (page 236) and the minimum of f is a unique point, the coordinates of which are the values of $\mathbf{h}$ where

$$\frac{\partial f}{\partial \mathbf{h}} = \mathbf{0}. \quad (\text{A1.3})$$

All of the eigenvalues of a non-negative definite matrix are non-negative, and all the eigenvalues of a positive-definite matrix are strictly positive [Strang76] (page 238). If the eigenvalues of A are strictly positive and distinct, then all the eigenvectors of A are linearly independent of (and in

1. The terms “non-negative definite” and the less specific “positive semidefinite” used in some references (eg [Haykin96]) are synonymous, and therefore interchangeable.

fact orthogonal to) one another [Haykin96] (page 164). This means A can be *diagonalized* [Strang76] (page 182), such that the eigenvalues of A can be obtained from the equation

$$\Lambda = S^T A S \quad (\text{A1.4})$$

where Λ is a diagonal matrix whose non-zero elements are the eigenvalues of A , and S is an $N \times N$ matrix whose columns are the eigenvectors of A . S is not unique, in general. If S is made to be *orthonormal*¹, then $S^{-1} = S^T$, which will be assumed henceforth.

A1.3.1 Hyperelliptical Surfaces

If A is positive-definite, then the set of all possible vectors $\mathbf{h}$ for which

$$\mathbf{h}^T A \mathbf{h} + \mathbf{b}^T \mathbf{h} = c^2 \quad (\text{A1.5})$$

for some particular $c \in \mathcal{R}^1$ defines the principal axes of a *hyperellipse* in the N dimensions of $\mathbf{h}$. Each distinct value of c produces a distinct hyperellipse concentric and proportional to all the others. A *quadratic surface* can be constructed by forming a continuum of these hyperellipses, which are then recognized as the *contours* of a surface, each associated with a different constant c . The minimum value of c in equation A1.5 is the minimum of the surface, and is found at or below the (coincident) centres of the hyperellipses.

The hyperellipses are said to be in *standard position* [Anton91] (page 414) if and only if each is distributed evenly about all N coordinate axes defined by $\mathbf{h}$, which requires the minimum of the surface to be located at or above $\mathbf{h} = \mathbf{0}$. It is much more difficult to describe a trajectory on a quadratic surface that is not in standard position than for one that is. Accordingly, a procedure is sought to move such a surface from any position and orientation in $\mathbf{h}$ -space to standard position.

One can determine whether the hyperellipses are out of standard position by expanding equation A1.5 explicitly into its quadratic elements. For the case of $N=2$, this would produce

$$A h_0^2 + B h_1^2 + C h_0 h_1 + D h_0 + E h_1 + F = 0 \quad (\text{A1.6})$$

1. An orthogonal matrix is made orthonormal by dividing each of its eigenvectors by their euclidean norms such that the orthonormalized matrix has eigenvectors of unit magnitude.

The orientation and position of the hyperellipses can be determined easily:

- If $C \neq 0$, the hyperellipses contain *cross-terms* and are *rotated* out of standard position;
- If $D \neq 0$ or $E \neq 0$, the hyperellipses contain explicit *first-order terms* and are *translated* out of standard position in one or more directions.
- If $F \neq 0$ then the surface has been offset above or below the coordinate system of $\mathbf{h}$.

The hyperellipse is in standard position if and only if

$$C = D = E = 0 \quad (\text{A1.7})$$

If it is not, it can be represented as being there through translation and rotation.

A1.3.2 Translation to Standard Position

Assume for the time being that the axes of the surface to be translated are in alignment with the coordinate system $\mathbf{h}$. That is to say, $C = 0$ in equation A1.6 and the off-diagonal elements of $\mathbf{A}$ are zero. The centre of the hyperellipses can be translated with a simple substitution of variables that will dismiss D and E in equation A1.6. This is achieved by completing the squares of each $\mathbf{h}$ term in equation A1.5 and then substituting

$$\underline{\mathbf{h}} = \mathbf{h} - \mathbf{h}' \quad (\text{A1.8})$$

where $\mathbf{h}'$ is the set of values required to remove the offsets in each axis.

A1.3.3 Rotation into Standard Position

An easy way to rotate the surface is to diagonalize the quadratic term $\mathbf{h}^T \mathbf{A} \mathbf{h}$ of equation A1.5 with some matrix $\mathbf{S}$ (as in equation A1.4). A pure rotation is guaranteed if $\det[\mathbf{S}] = 1$. Hence, equation A1.5 can be re-written as

$$(\mathbf{S}^T \mathbf{h})^T \mathbf{A} (\mathbf{S}^T \mathbf{h}) + \mathbf{b}^T (\mathbf{S}^T \mathbf{h}) = c^2, \quad (\text{A1.9})$$

which is simplified with the substitution

$$\bar{\mathbf{h}} = \mathbf{S}^T \mathbf{h} \quad (\text{A1.10})$$

to yield the same quadratic form of equation A1.5. The cross terms can be seen to be zero by rewriting equation A1.7 as

$$\mathbf{h}^T (\mathbf{S}^T \mathbf{A} \mathbf{S}) \mathbf{h} + \mathbf{b}^T \mathbf{S}^T \mathbf{h} = c^2 \quad (\text{A1.11})$$

The bracketed term is simply the diagonal eigenvalue matrix of equation A1.4, which is certain to be free of cross terms. Thus, an ellipse (or a surface) of general position and orientation has been moved into standard position, which was the desired objective.

If the hyperellipses are rotated out of standard position with respect to the coordinate space in which they reside, then their eigenmodes are not aligned with the base coordinate system. It will be shown shortly that when the surface is rotated out of standard position, the adaptive filter coefficients cannot be controlled independently, rendering such a system sub-optimal. Individual control of the convergence modes associated with each eigenvector can only be achieved when the cross-terms in the reference input are zero. An orthogonal transform such as the DCT will reduce the magnitude of the cross-terms substantially by arranging a coordinate system (i.e. a transformed set of filter coefficients) which is in alignment with the eigenvectors of the MSE hyperellipses.

A1.4 Mathematical Framework in the Time Domain

Many texts provide excellent introductions to linear algebra and matrix theory (eg [Strang76], [Anton91]), and a few relate these principles specifically to adaptive filtering [Widrow85] [Haykin96] [Treichler87], so only the key results will be presented here; no proofs of the results will be included. It is important to note that all adaptive filter algorithms mentioned in this work are derived for *stationary plants* and *wide-sense stationary input signals* only, averaged over a very long period of time. These conditions are very restrictive and rarely exist in practice, but they are crucial if the mathematical development is to be tractable.

A1.4.1 Input Crosscorrelation Vector

The *correlation* between the reference signal and the primary signal is given by

$$\underline{p} \equiv E[d_k x_k] . \quad (\text{A1.12})$$

where $E[\sim]$ is the expected-value operator. If $\underline{p} = \mathbf{0}$, the two inputs are uncorrelated with one another and no further decorrelation can be achieved. In AEC, however, d_k is derived from x_k through the impulse response of the unknown plant, so $\underline{p}$ is non-zero generally. Thus the *Input Crosscorrelation Vector* provides a measure of how much “work” remains to be done by the adaptive filter.

A1.4.2 Input Autocorrelation Matrix

The *Input Autocorrelation Matrix* is the average outer product of the windowed reference input over (an infinite) time index k :

$$\mathbf{R}_{xx} \equiv E[x_k x_k^T] \quad (\text{A1.13})$$

Note that $\mathbf{R}_{xx}$ is time-invariant, and therefore it is not a function of k . The definition is valid only if the reference input is WSS. All such reference inputs discussed in this research are real-valued stochastic processes. Some important results of these assertions are:

- $\mathbf{R}_{xx}$ is *Toeplitz*, which means that the elements along any vector parallel to its main diagonal are equal. This is the case whenever the excitation is WSS [Haykin96] (page 102).
- $\mathbf{R}_{xx}$ is *Hermitian* [Haykin96] (page 101), meaning $\mathbf{R}_{xx}$ is equal to its conjugate transpose.
- $\mathbf{R}_{xx}$ is *Symmetric* such that it also equals its (non-conjugated) transpose.

All the eigenvalues of $\mathbf{R}_{xx}$ are real numbers, regardless of their sign. This is a direct consequence of the Hermitian property [Haykin96] (page 164). The latter half of the statement is true since

$$\mathbf{x}_k \in \mathbb{R}^N \quad \forall k . \quad (\text{A1.14})$$

$\mathbf{R}_{xx}$ is positive-definite unless the excitation is exactly the sum of fewer than N pure sinusoids, in which case it is non-negative definite. The former property is assumed for the time-being, in

which case the eigenvalues of R_{xx} are strictly positive [Strang76] (page 217). Some examples of R_{xx} are shown in section A1.6.1 for various stochastic inputs. Autocorrelation matrices of real-valued WSS inputs have the following properties in the time domain:

- Elements that lie on the main diagonal of R_{xx} are all of equal value and they are proportional to the average power in the reference signal, so

$$R_{xx}(i, i) = E[x_k^2] \quad (A1.15)$$

Elements on any one diagonal parallel to the main diagonal are also equal to one another, and they are certain to be smaller than the elements on the main diagonal. This re-states the Toeplitz property, which holds for any excitation that is real-valued and WSS.

- Elements that lie off the main diagonal of R_{xx} are all equal to zero if and only if the PSD of the excitation is uniform, in which case there is no correlation between x_k and any delayed version of x_k .
- The perpendicular distance between the main diagonal and any off-diagonal of R_{xx} parallel to it is a direct measure of the amount of delay which produces a given amount of power in the autocorrelation function. Thus, the PSD of the excitation and the autocorrelation function are intimately related, and this relation is conveniently made through the DFT:

$$S_x(f) = \int_{-\infty}^{\infty} R_x(\tau) \exp[-j2\pi f\tau] d\tau \quad (A1.16)$$

where

$$R_x(\tau) = E[\langle x_{k+\tau} x_k \rangle_N] \quad (A1.17)$$

and

$$\langle x_{k+\tau} x_k \rangle_N = \frac{1}{2N} \sum_{i=k-N+1}^{k+N-1} x_{i+\tau} x_i, \quad k \geq N \quad (A1.18)$$

Equation A1.16 is the well-known *Weiner-Khinchin theorem*¹, where $S_x(f)$ is the PSD of the reference input.

The performance metrics for adaptive filters will now be developed, from which adaptation rules for the filter elements can be obtained.

A1.4.3 Mean Squared Error

From equation 2.5, the error signal is

$$e_k = d_k - \mathbf{w}_k^T \mathbf{x}_k \quad (\text{A1.19})$$

with instantaneous power of

$$e_k^2 = d_k^2 - 2d_k \mathbf{w}_k^T \mathbf{x}_k + \mathbf{w}_k^T \mathbf{x}_k \mathbf{x}_k^T \mathbf{w}_k. \quad (\text{A1.20})$$

The *Mean Squared-Error* (MSE) is

$$\xi = E[e_k^2] = E[d_k^2] - 2E[d_k \mathbf{x}_k] \mathbf{w}_k^T + \mathbf{w}_k^T E[\mathbf{x}_k \mathbf{x}_k^T] \mathbf{w}_k. \quad (\text{A1.21})$$

From the definitions of $\mathbf{p}$ and $\mathbf{R}_{xx}$ above, the MSE is written more compactly as

$$\xi = E[e_k^2] = E[d_k^2] - 2\mathbf{p}^T \mathbf{w}_k + \mathbf{w}_k^T \mathbf{R}_{xx} \mathbf{w}_k. \quad (\text{A1.22})$$

1. The theorem was actually presented about twenty years earlier by Albert Einstein, whose 1914 paper "Method for Determination of the Statistical Values of Observations Concerning Quantities Subject to Irregular Observations" that contained this result was re-printed in *IEEE ASSP Magazine*, 1987/10.

A1.4.4 Performance Surface

Equation A1.22 indicates the MSE is a quadratic function of $\mathbf{w}_k$, which is in turn a function of k and, until the filter has achieved its optimal value, $\mathbf{x}_k$ and d_k . It is useful, therefore, to construct a *graph* of the MSE¹ as a function of the elements of $\mathbf{w}$. This graph represents all possible values of MSE within the range of $\mathbf{w}_k$.

The MSE is a strictly positive quadratic surface with one global minimum and no other local minima. The gradient of the surface at any point is obtained by evaluating the partial derivative of the surface with respect to each element of $\mathbf{w}_k$:

$$\nabla = \frac{\partial \xi(\mathbf{w}_k)}{\partial \mathbf{w}_k} = 2\mathbf{R}_{xx}\mathbf{w}_k - 2\mathbf{p} . \quad (\text{A1.23})$$

From elementary calculus, the minimum of a concave-up quadratic surface is that point where the gradient is zero, so the values of $\mathbf{w}_k$ corresponding to the minimum MSE are revealed when

$$\mathbf{w}_k = \mathbf{w}_{opt} = \mathbf{R}_{xx}^{-1}\mathbf{p} \quad (\text{A1.24})$$

When this is the case, it is easy to show that the correlation between $\mathbf{x}_k$ and e_k approaches zero. Post-multiplying each term of equation A1.19 by $\mathbf{x}_k$ and applying the expected value operator yields

$$E[e_k \mathbf{x}_k] = E[d_k \mathbf{x}_k] - E[\mathbf{w}_k^T \mathbf{x}_k \mathbf{x}_k^T] \quad (\text{A1.25})$$

which is

$$E[e_k \mathbf{x}_k] = \mathbf{p} - \mathbf{R}_{xx}\mathbf{w}_k \quad (\text{A1.26})$$

and when $\mathbf{w}_k = \mathbf{w}_{opt}$, equation A1.26 can be re-written after substitution of equation A1.24 as

$$E[e_k \mathbf{x}_k] = \mathbf{p} - \mathbf{R}_{xx}\mathbf{R}_{xx}^{-1}\mathbf{p} = \mathbf{p} - \mathbf{p} = \mathbf{0} . \quad (\text{A1.27})$$

1. Here, the term “graph” is used in the strict sense, meaning it is an object in $N+1$ dimensions, where N is the number of independent variables in the coordinate system. In this case the coordinate space is defined by the N -dimensional vector $\mathbf{w}$.

This result shows that all artifacts of the reference signal are removed from the primary when $\mathbf{w}_k = \mathbf{w}_{opt}$. This is the purpose of the adaptive filter stated in section A1.2.1. It is now appropriate to refine that objective to include the desire to achieve $\mathbf{w}_k = \mathbf{w}_{opt}$ *in the shortest time possible*.

When relating the hyperellipses to the coordinate system of $\mathbf{w}$, the following are true in general:

- *The eigenvectors of the hyperellipses are not aligned with the axes of $\mathbf{w}$.*

This is to say that the hyperellipses are rotated from standard position, and it implies that the elements of the adaptive filter cannot be manipulated independently, which in turn means that there are fewer than N degrees of freedom in the adaptation process.

- *The shape of the performance surface is changed whenever the reference input is changed.*

This is because the principal axes of the performance surface are functions of $\mathbf{R}_{xx}$. Hence, the step-size bounds will be different when the reference PSD changes.

- *The centre of each hyperellipse is not coincident with $\mathbf{w} = \mathbf{0}$.*

Thus, $\mathbf{w}_{opt}$ is not at the origin of $\mathbf{w}$ -space, and the surface is translated from standard position.

- *The location of $\mathbf{w}_{opt}$ is changed when the unknown plant is changed.*

The filter must re-converge to track changes in the plant. The adaptation rate depends on the geometry of the performance surface and the direction of the new $\mathbf{w}_{opt}$ with respect to $\mathbf{w}_k$.

A1.4.5 Modal Convergence

It has been shown that the MSE (the function to be minimized in a gradient-based algorithm) constitutes a performance surface in $N-1$ dimensions. Following the development of section A1.4.3, points of equal MSE on a performance surface form concentric hyperellipses in $\mathbf{w}$ -space. Thus, the performance surface can be represented using a contour plot in N dimensions, where it is appropriate to use methods of eigenanalysis to describe the MSE geometry in terms of $\mathbf{R}_{xx}$.

The rate of convergence in the adaptive filter is best explained using the geometrical description underlying the performance surface. The description is simplified when a coordinate system is defined in which the performance surface is rendered in standard position. In this case, the cross terms are eliminated, leaving

$$\xi = \bar{\mathbf{w}}_k^T \Lambda \bar{\mathbf{w}}_k \quad (\text{A1.28})$$

where

$$\bar{\mathbf{w}} = S \mathbf{w} \quad (\text{A1.29})$$

and S is the matrix of eigenvectors of R_{xx} , Λ is the (diagonal) matrix whose non-zero elements are the eigenvalues of R_{xx} , $\bar{\mathbf{w}}$ is the newly created coordinate system, $\underline{\mathbf{w}} = \mathbf{w} - \mathbf{w}_{opt}$ is the translation vector, and $\mathbf{w}$ is the “natural” coordinate system of the adaptive filter. The first and second derivatives of ξ in the direction of each eigenvector are, respectively,

$$\frac{\partial \xi}{\partial \mathbf{w}(n)} = 2\lambda_n \overline{\mathbf{w}(n)} \quad (\text{A1.30})$$

$$\frac{\partial^2 \xi}{\partial \mathbf{w}(n)^2} = 2\lambda_n \quad (\text{A1.31})$$

Equation A1.30 shows that the principal axes of the performance surface are in alignment with the eigenvectors of R_{xx} , and these are aligned with the adaptive filter coordinate system *when the performance surface is in standard position*. Equation A1.31 shows that the second derivatives of the performance surface in the directions of its principal axes are proportional to the eigenvalues of R_{xx} , and this is the case regardless of the performance surface orientation. Thus, *the eigenvalues provide a measure of the steepness of the quadratic performance surface in the direction of their respective eigenvectors*. In practical terms, they represent the *power* in each eigenmode of the adaptive filtering system.

In time-domain systems, it is not possible to control the filter coefficients independently because the coordinate space of the adaptive filter is not aligned generally with the eigenvectors of R_{xx} . Accordingly, a single step-size is generally chosen for the entire filter, and its maximum value is chosen as a function of the largest eigenvalue of R_{xx} or, more conservatively, as the trace of Λ , the diagonal eigenvalue matrix of R_{xx} . The best choice of step-size depends on the algorithm chosen, amongst other factors. A more detailed discussion of the step-size bounds is provided in chapter 2.

A1.5 Mathematical Framework in the Transform Domain

Here, the mathematical basis just presented for time-domain filters is extended to describe the geometry of the *Sample-Oriented Transform-Domain Adaptive Filter* (SOX-AF) of chapter 3. Whereas the insertion of a transform operator in the LMS structure of figures 3.1 and 3.2 appears to be a minor change at first, it has profound implications on the theory of adaptation. In chapter 4, new SOX-AF mechanisms will be constructed using the results developed in this section.

The SOX-AF of figure 3.3 is similar to the time-domain LMS filter inasmuch as the predicted output y_k is the inner product of the adaptive filter vector and the filter inputs, and the error e_k is computed in both domains as the difference between y_k and d_k . On the other hand, the reference signal is conveyed to the adaptive filter very differently: in the SOX-AF, the data is fed to the adaptive filter *in parallel* whereas it is fed *in series* to time-domain adaptive filters. The effects of this fundamental difference on the performance surface geometry will be discussed once suitable terminology has been introduced to distinguish between corresponding quantities in the two domains. First of all, the adaptive filter vector and its optimal solution will be represented in the transform domain by $\mathbf{w}_k$ and $\mathbf{w}_{opt}$ respectively. The other quantities are developed as follows:

A1.5.1 Input Cross-Correlation Vector

The objective is to minimize the correlation between the reference input and the error signal. As before, it is convenient to define the input cross-correlation vector in the transform domain as

$$\bar{\mathbf{p}} = E[d_k \mathbf{u}_k] \quad (\text{A1.32})$$

where

$$\mathbf{u}_k = T[\mathbf{x}_k] \quad (\text{A1.33})$$

and the operator $T[\sim]$ is a real-valued orthogonal transform such as the DCT¹. Further,

$$\bar{\mathbf{p}} = T[\underline{\mathbf{p}}] \quad (\text{A1.34})$$

1. The development could be extended easily to the complex case, though it is not done so here because the DCT has proven to be a very popular transform in speech applications.

A1.5.2 Input Autocorrelation Matrix

Equation A1.33 invites the definition of the input autocorrelation matrix in the transform domain immediately:

$$\mathbf{R}_{uu} \equiv E[\mathbf{u}_k \mathbf{u}_k^T] . \quad (\text{A1.35})$$

As before, the reference signal is assumed to be a WSS real-valued stochastic process such that $\mathbf{R}_{uu}$ is time-invariant. It follows that $\mathbf{R}_{uu}$ is hermitian, and it is symmetric if a real transform is employed. Thus, all the eigenvalues of $\mathbf{R}_{uu}$ are real numbers, and will be assumed to be strictly positive.

The profile of $\mathbf{R}_{uu}$ is significantly different than that of $\mathbf{R}_{xx}$ for the same reference input. Recall that the excitation slides in sequence along the adaptive filter inputs in time-domain filters. In the SOX-AF, the excitation slides along the input to the transform, and the transform outputs are fed in parallel to the adaptive filter stage. Thus, each adaptive filter tap in the SOX-AF acts upon a different input sequence, whereas in the time domain, the taps act upon delayed versions of the same sequence. There are two major implications of this difference:

- Elements that lie on the main diagonal of $\mathbf{R}_{uu}$ are different from one another unless the PSD of the excitation is uniform. Thus, $\mathbf{R}_{uu}$ is *not Toeplitz in general*. In fact, elements on the main diagonal of $\mathbf{R}_{uu}$ are proportional to the average power in the corresponding channel of the reference signal. To be more precise,

$$\mathbf{R}_{uu}(i, i) = cE[u_k(i)u_k(i)] . \quad (\text{A1.36})$$

In the case of the unitary DFT, these elements represent the reference signal power at evenly distributed locations around the unit circle [Strang76], (page 217). The same is true for the DCT, however the distribution has twice the frequency resolution and is scaled differently than the DFT [Rao92]. Thus, the diagonal elements of $\mathbf{R}_{uu}$ represent the power distribution of the excitation in the domain of T .

- Elements that lie off the main diagonal of $\mathbf{R}_{uu}$ are zero, *regardless of the spectral distribution of the reference signal* if the transform is a perfect decorrelator. This is a direct consequence of the orthogonality of the transform, and it is true because there is no correlation between

sequences that are exactly orthogonal to one another. Therefore, the autocorrelation function in the transform domain involves shifts in frequency rather than in time. Specifically, the statement

$$R_{uu}(i,j) = cE[u_k(i)u_k(j)] \quad (\text{A1.37})$$

indicates the correlation is being measured between the outputs of different output channels of the transform. Frequency translation cannot occur in a purely linear transformation, so any residual energy in the off-diagonal elements of R_{uu} is due to *leakage* (arising from the lack of perfect decorrelation) between channels in the transform.

A1.5.3 Mean-Squared Error

The MSE derivation in the transform domain is similar to that provided for the time-domain adaptive filter. The derivation is repeated here to make explicit the change in variables. The error signal in the SOX-AF is

$$e_k = d_k - \mathbf{w}_k^T \mathbf{u}_k \quad (\text{A1.38})$$

The instantaneous error power is

$$e_k^2 = d_k^2 - 2d_k \mathbf{w}_k^T \mathbf{u}_k + \mathbf{w}_k^T \mathbf{u}_k \mathbf{u}_k^T \mathbf{w}_k \quad (\text{A1.39})$$

so the MSE in the transform domain can be written as

$$\xi = \mathbf{w}_k^T \mathbf{R}_{xx} \mathbf{w}_k - 2\bar{\mathbf{p}}^T \mathbf{w}_k + E[d_k^2] \quad (\text{A1.40})$$

Once again, the MSE is quadratic in $\mathbf{w}_k$.

A1.5.4 Performance Surface

The transform alters the excitation of the adaptive filter in a prescribed manner, so the location and orientation of the performance surface must be altered also. Since the transform of a sequence is numerically different than the sequence itself, $\mathbf{w}_{opt}$ must be translated away from $\mathbf{w}_{opt}$ in $\mathbf{w}$ -space. The goal now is to derive the location of $\mathbf{w}_{opt}$ with respect to $\mathbf{w}_{opt}$.

The coordinates of $\mathbf{w}_{opt}$ are

$$\mathbf{w}_{opt} = \mathbf{R}_{uu}^{-1} \bar{\mathbf{p}} \quad (\text{A1.41})$$

which can be expanded into

$$\mathbf{w}_{opt} = [\mathbf{T} \mathbf{R}_{xx} \mathbf{T}^T]^{-1} \mathbf{T} \underline{\mathbf{p}}. \quad (\text{A1.42})$$

Removing the inversion yields

$$\mathbf{T} \mathbf{R}_{xx} \mathbf{T}^T \mathbf{w}_{opt} = \mathbf{T} \underline{\mathbf{p}} \quad (\text{A1.43})$$

and premultiplying both sides in turn by $\mathbf{T}^{-1}$ and then by $\mathbf{R}_{xx}^{-1}$ leaves

$$\mathbf{T}^T \mathbf{w}_{opt} = \mathbf{R}_{xx}^{-1} \underline{\mathbf{p}}. \quad (\text{A1.44})$$

Now, assuming that the transformation is unitary, (so that $\mathbf{T}^T = \mathbf{T}^{-1}$), the desired relationship is revealed:

$$\mathbf{w}_{opt} = \mathbf{T}^{-1} [\mathbf{R}_{xx}^{-1} \underline{\mathbf{p}}] = \mathbf{T}^{-1} [\mathbf{w}_{opt}] \quad (\text{A1.45})$$

Equation A1.45 states that the optimal solution in the transform domain is merely a translation of $\mathbf{w}_{opt}$ whose value is revealed directly through the application of $\mathbf{T}$. It is convenient to view the elements of the translated filter as having defined a new coordinate system $\mathbf{w}$ which shares the same space as $\mathbf{w}$ but whose origin is located at $\mathbf{w}_{opt}$ rather than at $\mathbf{w}_{opt}$.

The $\mathbf{w}$ coordinate system need not be aligned to the original system $\mathbf{w}$. In fact, it is particularly useful that *the translated coordinate system is automatically aligned with the principal axes of the performance surface* if the transform is an effective decorrelator. Why is this so? Assume for

the time being that the unknown plant and the adaptive filter are of the same order. Recall that a truly orthogonal transform will produce N parallel output sequences that are uncorrelated with one another. Since the eigenvectors define the only orthogonal axes of the performance surface, it follows that each input sequence incident on the LMS stage of the transform-domain adaptive filter is associated with exactly one of them. This allows the adaptive filter coefficients to be controlled and adjusted independently of one another as each one is associated with a particular stream of distinct data.

A1.5.5 Modal Convergence

Using the arguments of the previous section, the transform domain filter can rightly be viewed as a set of N 1-dimensional filters. Some useful properties flow from this point of view:

- movement in a direction parallel to the n th eigenvector is achieved by varying $\mathbf{w}_{k(n)}$ only;
- the eigenvalue λ_n is proportional to the power in the n th eigenmode of $\mathbf{R}_{uu}$;
- λ_n is a measure of the slope of the performance surface in the direction of the n th eigenvector;
- the upper bound to the channel step-size is proportional to λ_n rather than λ_{max} or $\text{tr}[\Lambda]$.

It should now be apparent that in an ideal transform-domain system, the overall convergence of the filter should be equal in all eigenmodes of the filter regardless of the PSD of the reference input. The potential advantages of this structure over its time-domain counterpart are significant. The transformation can rotate and translate the MSE performance surface, but it cannot affect its shape [Marshall89]. As a result, no net improvement in convergence rate can be achieved due to the transform alone because the slopes of the surface in the directions of the eigenvectors remain unchanged. It is the alignment of the eigenvectors of the performance surface with the elements of the adaptive filter which allow the convergence rates in each channel to be normalized. To achieve this convergence normalization, it is essential to estimate the power in each eigenmode accurately and quickly in order to maintain the stability conditions prescribed for algorithms of the steepest-descent variety.

A1.6 Visualization of the Geometry

The key concepts of sections A1.4 and A1.5 are demonstrated graphically in this section. In the first part, comparisons between R_{xx} and R_{uu} are made for various WSS reference inputs, and the effect of power normalization on R_{uu} and the eigenvalue distribution will be demonstrated for a system of order 64. In the second part, the convergence rate improvement afforded by the power normalized SOX-AF will be demonstrated for a second order system.

A1.6.1 Autocorrelation Matrices

Figure A1.1 (a) shows a surface plot of R_{xx} for the AR1(.9) input. The zero delay term $R_x(0)$ is shown at the centre rear, and the maximum delay term $R_x(N-1)$ is in the near foreground. Surface plots of R_{uu} before and after power normalization are shown in figures A1.1 (b) and A1.1 (c) respectively. Here, the DC term is shown in the rearground, and the highest frequency term is shown in the foreground. The eigenvalues of the three autocorrelation matrices were calculated, and these are plotted together on a logarithmic scale in figure A1.1 (d).

Similar plots for the AR2(-1.9114,.95), AR2(-1.5955,.95), and AR12 processes are shown in figures A1.2, A1.3, and A1.4 respectively. The condition number (the ratio of the largest to smallest eigenvalues) of each matrix is listed in table A1.3. The condition number is often used as an indicator of the convergence rate which can be achieved for certain combinations of excitation and filter order.

TABLE A1.3 Peak eigenvalue ratios for various inputs with $N=64$.

Excitation	Time Domain	DCT Domain	Normalized DCT Domain
AR1(0.9)	331	646	3
AR2(-.195,.95)	954	954	34
AR2(-.975,.95)	2314	2314	29
AR2(-1.5955,.95)	8670	8680	26
AR2(-1.9114,.95)	78,280	78,280	24
AR12	7,336	7,336	10

FIGURE A1.1: Visualization of autocorrelation matrices and their eigenvalues for the AR1(0.9) process. In (a), R_{xx} is shown with the delay $\tau = 0$ at the rear and $\tau = N-1$ to the fore; in (b), R_{uu} is shown for the DCT with low frequencies in the rearground and high frequencies in the foreground; the power-normalized R_{uu} is shown in (c); the eigenvalues of R_{xx} (dotted), R_{uu} (dashed), and the normalized R_{uu} (solid) are shown in (d).

FIGURE A1.2: Visualization of autocorrelation matrices and their eigenvalues for the $AR2(-1.9114, 95)$ process. In (a), R_{xx} is shown with the delay $\tau = 0$ at the rear and $\tau = N-1$ to the fore; in (b), R_{uu} is shown for the DCT with low frequencies in the rearground and high frequencies in the foreground; the power-normalized R_{uu} is shown in (c); the eigenvalues of R_{xx} (dotted), R_{uu} (dashed), and the normalized R_{uu} (solid) are shown in (d).

FIGURE A1.3: Visualization of autocorrelation matrices and their eigenvalues for the AR2(-1.5955, .95) process. In (a), R_{xx} is shown with the delay $\tau = 0$ at the rear and $\tau = N-1$ to the fore; in (b), R_{uu} is shown for the DCT with low frequencies in the rearground and high frequencies in the foreground; the power-normalized R_{uu} is shown in (c); the eigenvalues of R_{xx} (dotted), R_{uu} (dashed), and the normalized R_{uu} (solid) are shown in (d).

FIGURE A1.4: Visualization of autocorrelation matrices and their eigenvalues for the AR12 process. In (a), R_{xx} is shown with the delay $\tau = 0$ at the rear and $\tau = N-1$ to the fore; in (b), R_{uu} is shown for the DCT with low frequencies in the rearground and high frequencies in the foreground; the power-normalized R_{uu} is shown in (c); the eigenvalues of R_{xx} (dotted), R_{uu} (dashed), and the normalized R_{uu} (solid) are shown in (d).

It is clear from the plots of the raw R_{xx} and R_{uu} in this section that the transform has a profound effect on the shape of the autocorrelation matrix. The Toeplitz property of R_{xx} is clearly evident, and both R_{xx} and R_{uu} exhibit symmetry about their main diagonals. The off-diagonal elements of the power-normalized R_{uu} increase at higher frequencies, which confirms that the DCT is a much more efficient decorrelator at lower frequencies than at higher ones. These plots are in close agreement with the analytically derived plots of same presented in [Beaufays95b].

From table A1.3 it is also clear that the transform itself has no effect on the eigenvalue disparity. This is because the transform can only cause rotation and translation. These operations will change the eigenvectors, but not the eigenvalues. However, rotation of the performance into alignment with the adaptive filter's coordinate system facilitates independent control of channel step-sizes. The figures and table illustrate that the eigenvalue compression achievable through the proper power normalization of the DCT can be quite dramatic. A substantial increase in the overall convergence rate is bound to follow.

A1.6.2 Performance Surfaces and Convergence Rates

The importance of specifying equivalent starting points when comparing convergence rates in algorithms of different domains cannot be overstated. The primary purpose of this section is to demonstrate this importance graphically and to describe how fair comparisons can be assured in systems of higher order.

In the following diagrams $N=2$ has been chosen so as to allow the MSE to be drawn in three dimensions. This restriction is imposed because it is very difficult to display objects of greater than three dimensions on paper. The unknown plant is also chosen to be of order N to allow full convergence. This need not be the case in general, and it is often not the case in practice. The problem of *undermodeling* is discussed in section 6.5, and is not of concern here.

The reference input consisted of 250 samples drawn from the AR2(-1.5955,.95) process. The eigenvalues of the resulting autocorrelation matrix are 1.818 and 0.182, producing a condition number of 10. In each case, care was taken to specify initial adaptive filter coefficients which

allowed comparable convergence modes to be traversed in each domain. The following steps were taken to determine the equivalent starting points:

- compute the autocorrelation matrix of the sequence;
- compute the eigenvalues and eigenvectors of the autocorrelation matrix;
- rank the eigenvalues in ascending order and rearrange the eigenvector matrix accordingly;
- extract the eigenvector $S(n)$ associated with the desired eigenmode of convergence;
- compute the starting point as

$$\mathbf{w}_k = \mathbf{w}_{opt} - aS(n) \quad (\text{A1.46})$$

in the time-domain, or

$$\mathbf{w}_k = \mathbf{w}_{opt} - aS(n) \quad (\text{A1.47})$$

in the transform domain, where in each case a is a scale factor chosen to provide the desired MSE at point of reset.

The global stepsize was set to 0.015 for the time-domain LMS plots, and to half that value (since $N=2$) for the AR-SOX-AF plots. These step-sizes were made deliberately small so as to slow down the convergence and make the trajectories visible. The power estimation parameter β was set to 0.75 to provide reasonably smooth convergence trajectories in the transform domain plots.

The corresponding time- and transform-domain plots are shown for the fastest convergence mode in figure A1.5, for the slowest convergence mode in figure A1.6, and for a combination of modes in figure A1.7. All of the plots were corrected for proper aspect ratio. Each contour in a given plot represents a set of points with equal MSE that is twice the magnitude represented by the next inner contour. The same contour magnitudes were specified in each domain.

FIGURE A1.5: Comparative visualization of fast-mode convergence rates in time- and transform domain algorithms. The LMS algorithm is shown in (a) and the AR-SOX-AF is shown in (b).

FIGURE A1.6: Comparative visualization of slow-mode convergence rates in time- and transform domain algorithms. The LMS algorithm is shown in (a) and the AR-SOX-AF is shown in (b).

FIGURE A1.7: Comparative visualization of bi-modal convergence rates in time- and transform domain algorithms. The LMS algorithm is shown in (a) and the AR-SOX-AF is shown in (b).

It is clear from the plots that there is a significant disparity in convergence rates between the two domains. However, the degree of that disparity is accentuated if a slower convergence mode is traversed. Often, the slowest convergence mode is selected when making comparisons between algorithms in different domains because one is usually interested in the worst-case performance. An excellent pictorial example showing the effect of spectral distribution on surface geometry and the notion of equivalent eigenfilters is provided in section 9.9 of [Haykin96].

To summarize, the shape of the MSE performance surface is strictly a function of the reference input, while its position in the MSE coordinate space is specified by the unknown plant. Hence, the combination of the reference input and the unknown plant characterizes a particular adaptive filtering problem.

A1.7 Some Notes Concerning Computational Expense

Efficient algorithms minimize execution time and hardware costs simultaneously. Good coding practice is most important in AEC because the firmware is intended for use in mass-produced, cost-sensitive products whose signal processing must be undertaken in real-time.

Algorithm efficiency is very platform-dependent, however. Numerous DSP devices with very different architectures and instruction sets are available, so an algorithm tailored for one may prove to be very awkward or inefficient for another. Further, most DSP's provide some sort of multiply/accumulate instruction, so the total number of multiplications and/or additions is an insufficient metric of computational burden. Also, single-cycle multiplications are implemented entirely in hardware in most DSP's, making them no more onerous than any other instruction.

In this work, minimal theoretical calculations of computational expense will be provided, since the true burden will vary so widely from one platform to the next. All of the algorithms tested were coded using Matlab v4.2c, with double-precision floating point representation maintained at all stages.

A2

Signals and Systems Used in Simulations

This appendix provides further information about the reference input signals and measured impulse responses referred to in chapter 6, and room transfer functions observed in chapter 7.

A2.1 Excitation

The spectral characteristics of the AR2 and AR12 sequences are provided in figures A2.1 and A2.2. In each case, the PSD's were estimated using a sequence of length 100,000 that had been normalized to have zero mean and unit variance. The PSD plots for the two synthetic speech tracks SSGF2 and SSGM1 are provided in figure A2.3. These were also normalized, but the PSD was estimated over a single full-length speech burst with about 17,500 samples at 8 kHz.

FIGURE A2.1: Spectral characteristics of the four AR2 processes. In (a), the PSD's of the AR2(-.195,.95) (dotted), AR2(-.975,.95) (dash-dotted), AR2(-1.5955,.95) (dashed), and AR2(-1.9114,.95) (solid) are shown. In (b), the pole locations of the AR2 processes in the z-plane are shown. The pole pairs left to right from the imaginary axis correspond to the AR2(-.195,.95), AR2(-.975,.95), AR2(-1.5955,.95), and AR2(-1.9114,.95) processes.

FIGURE A2.2: Spectral characteristics of the AR12 process. In (a), the PSD is shown. In (b), the pole locations of the AR12 process in the z-plane are shown.

FIGURE A2.3: Power spectral density plots of the synthetic speech sequences. In (a), the SSGF2 PSD is shown, and the SSGM1 PSD is shown in (b). In each case, the window of observation contained one full-length speech burst of just over two seconds' duration.

A2.2 FIR Transfer Functions

The FIR impulse response IRN(1200) is shown in figure A2.4 (a) and (b), and its PSD in (c). The others are omitted as their visible characteristics are quite similar to those shown.

FIGURE A2.4: Time series and PSD plots of the FIR impulse response IRN(1200). In (a), the full 150 msec (1200 coefficients at 8.000 kHz) is shown; in (b), the first 20 msec are expanded; in (c), the PSD of the full-length sequence is shown.

The differences between the four FIR impulse responses are demonstrated in figure A2.5. Here, the TIP/TP ratio of IRN(1200) is shown alongside the TIP/TP *differences* (IR1-IRN), (IR2-IRN), and (IR3-IRN). A differential TIP/TP plot shows the location of the reflective surface much as a time-delay reflectometry study would. It should be apparent that substantially better cancellation can be achieved if the time delay associated with the major reflective surface is spanned by the order N of the adaptive filter.

FIGURE A2.5: Time-delay measurement of the reflective feature distinguishing four FIR impulse responses recorded in MC2014. In (a), the full 150 msec sequences are shown; in (b), the first 20 msec are expanded. In both cases, the TIP/TP ERLE limit for IRN (dotted) is shown along with (IR1-IRN) (dash-dotted), (IR2-IRN) (dashed), and (IR3-IRN) (solid).

A2.3 Acoustic Transfer Functions

MC2014 is a conference room with dimensions of 19 feet by 38 feet by 10 feet high, as shown in figure A2.6. All of the walls are drywalled except one which features a large window. The two doors remained closed at all times. Many large padded chairs were arranged around the table and the perimeter of the room. A tightly woven carpet covers the entire floor area.

ME4144 has dimensions of 10 feet by 12 feet by 10 feet high with cement block walls. The two tall and narrow windows remained closed, and the two work stations were turned off at all times. The floor is tiled, and its plan view is shown in figure A2.7.

FIGURE A2.6: Layout of conference room MC2014 at Carleton University. The room dimensions are approximately 19 feet by 38 feet by 10 feet high. The doors are represented by dark rectangles, and the windows by clear rectangles.

FIGURE A2.7: Layout of graduate office ME4144 at Carleton University. The room dimensions are approximately 10 feet by 12 feet by 10 feet high. The door is represented by a dark rectangle, and the windows by clear rectangles.

A3

List of References

- [Aboulnasr95] Tyseer Aboulnasr. l'Université d'Ottawa, related through course ELG5377 (1994) and through subsequent discussions
- [Aboulnasr97] Tyseer Aboulnasr and Khalid Mayyas: "A Robust Variable Step-Size LMS-Type Algorithm: Analysis and Simulations" *IEEE Trans. SP* Vol. 45 No. 3, 1997 March, pp 631-639
- [Akansu92] Ali N. Akansu and Richard A. Haddad: "Multiresolution Signal Decomposition - Transforms, Subbands, Wavelets" San Diego: Academic Press 1992
- [Alexander86] S.T. Alexander: "Fast Adaptive Filters: A Geometrical Approach" *IEEE ASSP*, 1986 October, pp 18-28
- [Algazi93] V. Ralph Algazi, Kathy L. Brown, Michael J. Ready, David H. Irvine, Christie L. Cadwell, and Sang Chung: "Transform Representation of the Spectra of Acoustic Speech Segments with Applications II: Speech Analysis, Synthesis, and Coding" *IEEE Trans. SAP* Vol. 1 No. 3, 1993 July, pp 277-286
- [Alkin95] Oktay Alkin and Hakan Caglar: "Design of Efficient M-Band Coders with Linear Phase and Perfect Reconstruction Properties" *IEEE Trans. SP* Vol. 43 No. 7, 1995 July, pp 1579-1590
- [Amand94] Fabrice Amand and Andre Gilloire: "Multi-Channel Acoustic Echo Cancellation" *Proc. IWAENC* 1994, pp 57-60
- [Amano91] Fumio Amano and Hector Perez: "A New Sub-Band Echo Canceller Structure" *Proc. ICASSP* 1991 pp 3585-3588
- [Amrane92] O. Ait Amrane, E. Moulines, M. Charbit, and Y. Grenier: "Low-Delay Frequency Domain LMS Algorithm: *Proc. ICASSP* 1992 pp IV-9 - IV-12
- [Anton91] Howard Anton: "Elementary Linear Algebra", Toronto: John Wiley and Sons, sixth edition, 1991
- [Antweiler95] Christiane Antweiler, Andrea Schiffer, and Matthias Dorbecker: "Acoustic Echo Control with Variable Individual Step-Size" *Proc. IWAENC* 1995 June, pp 15-18
- [Arce87] Gonzalo R. Arce and Robert L. Stevenson: "On the Synthesis of Median Filter Systems" *IEEE Trans. CAS* Vol. CAS-34 No. 4, 1987 April, pp 421-429
- [Ashari86] Mohammad Reza Asharif, Tomoyoshi Takebayashi, Takafumi Chujo and Kazuo Murano: "Frequency Domain Noise Canceller: Frequency Bin Adaptive Filtering" *Proc. ICASSP* 1986 pp 41.22.1 - 41.22.4
- [Ashari87] Mohammad Reza Asharif, Fumio Amano, Shigeyuki Unagami, and Kazuo Murano: "Acoustic Echo Canceller Based on Frequency Bin Adaptive Filtering" *IEEE Globecom* 1987 pp 49.2.1 - 49.2.4
- [Ayad95] Beghdad Ayad and Gérard Faucon: "Acoustic Echo and Noise Cancelling for Hands-Free Communication Systems" *Proc. IWAENC* 1995 June, pp 91-94
- [Baganne95] A. Baganne, E. Martin, P. Martineau, A. Gilloire, and J.P. Thomas: "Implementation Methods of an Acoustic Echo Cancellation Algorithm - GMDFA" *Proc. IWAENC* 1995 June, pp 99-102

-
- [Beaufays95a] Françoise Beaufays: "Transform Domain Adaptive Filters: An Analytical Approach" *IEEE Trans. SP* Vol. 43 No. 2, 1995 February, pp 422-431
 - [Beaufays95b] Françoise Beaufays: "Two-Layer Linear Structures for Fast Adaptive Filtering" Ph.D. Dissertation, Stanford University, 1995 June
 - [Beaufays95c] Françoise Beaufays and Bernard Widrow: "On the Advantages of the LMS Spectrum Analyzer Over Non-Adaptive Implementations of the Sliding DFT" *IEEE Trans. CAS* Vol. 42 No. 4, 1995 April, pp 218-220
 - [Bellanger76] Maurice G. Bellanger, Georges Bonnerot, and Michel Coldreux: "Digital Filtering by Polyphase Network: Application to Sample Rate Alteration and Filter Banks" *IEEE Trans. ASSP* Vol. ASSP-24 No. 2, 1976 April, pp 109-114
 - [Berberidis96] Kostas Berberidis, and Jacques Palicot: "A Block Quasi-Newton Algorithm Implemented in the Frequency Domain" *Proc. ICASSP* 1996 CD-ROM
 - [Bershad79] N.J. Bershad and P.L. Feintuch: "Analysis of the Frequency-Domain Adaptive Filter" *Proc. IEEE* Vol. 67 No. 12, 1979 December, pp 1658-1659
 - [Bershad86] Neil J. Bershad and Paul L. Feintuch: "A Normalized Frequency Domain LMS Adaptive Algorithm" *IEEE Trans. ASSP* Vol. ASSP-34, No. 3, 1986 June, pp 452-461
 - [Binde95] Stephan Binde, Bernhard Nitsch, and Tobias Schneider: "Application of FRLS Algorithms for Real-Time Acoustic Echo Compensation" *Proc. IWAENC* 1995 June, pp 2-5
 - [Bitmead81] Robert R. Bitmead and Brian D.O. Anderson: "Adaptive Frequency Sampling Filters" *IEEE Trans. CAS* Vol. CAS-28 No. 6, 1981 June, pp 524-534
 - [Boudy95] Jerome Boudy, Francois Capman, and Philip Lockwood: "A Globally-Optimized Frequency-Domain Acoustic Echo Canceller for Adverse Environment Applications" *Proc. IWAENC* 1995 June, pp 95-98
 - [Bradley96] John S. Bradley: "Optimizing the Decay Range in Room Acoustics Measurements Using Maximum-Length Sequence Techniques" *JAES* Vol. 44 No. 4, 1996 April, pp 266-273
 - [Brigham88] E. Oran Brigham: "The Fast Fourier Transform and its Applications" Englewood Cliffs: Prentice Hall 1988
 - [Burger95] Thomas Burger: "Practical Applications of Adaptation Control for NLMS Algorithms used for Echo Cancellation with Speech Signals" *Proc. IWAENC* 1995 June, pp 87-90
 - [Burrus94] C. Sidney Burrus, James H. McClellan, Alan V. Oppenheim, Thomas W. Parks, Ronald W. Schaefer, Hans W. Schuessler: "Computer-Based Exercises for Signal Processing Using Matlab" Englewood Cliffs: Prentice-Hall 1994
 - [Chakra93] Chaitali Chakrabarti: "Sorting Network Based Architectures for Median Filters" *IEEE Trans. CAS-II* Vol. 40 No. 11, 1993 November, pp 723-727
 - [Chan90] S.C. Chan and K.L. Ho: "Direct Methods for Computing Discrete Sinusoidal Transforms" *IEE Proc. F* Vol. 137 No. 6, 1990 December, pp 433-442
 - [Chao90] Jinshui Chao, Hector Perez, and Shigeo Tsujii: "A Fast Adaptive Filter Algorithm Using Eigenvalue Reciprocals as Stepsizes" *IEEE Trans. ASSP* Vol. 38, No. 8, 1990 August, pp 1343-1352
 - [Chau94] Lap-Pui Chau and Wan-Chi Siu: "Efficient Formulation for the Realization of the DCT using Recursive Structure" *Proc. ICASSP* 1994, pp III-437 - III-440
 - [Chau95] Lap-Pui Chau and Wan-Chi Siu: "Direct Formulation for the Realization of the Discrete Cosine Transform using Recursive Structure" *IEEE Trans. CAS-II* Vol. 42 No. 1, 1995 January, pp 50-51
 - [Chen88a] Jiande Chen, Hugo Bes, Joos Vandewalle and Paul Janssens: "A New Structure for a Subband Acoustic Echo Canceller" *Proc. ICASSP* 1988 pp 2574-2577
-

-
- [Chen88b] Jiande Chen, Hugo Bes, Joos Vandewalle and Paul Janssens: "Design of a Zero-Delay Subband Acoustic Echo Canceller" *Proc. ISCAS* 1988 pp 1325-1328
 - [Chen96] Jiande Chen, Hugo Bes, Joos Vandewalle and Paul Janssens: "A Zero-Delay FFT-Based Subband Acoustic Echo Canceller for Teleconferencing and Hands-Free Telephone Systems" *IEEE Trans. CAS-II* Vol. 43 No. 10, 1996 October, pp 713-717
 - [Chiang96] Hwang-Cheng Chiang and Jie-Cherng Liu: "Fast Algorithm for FIR Filtering in the Transform Domain" *IEEE Trans. SP* Vol. 44 No. 1, 1996 January, pp 126-129
 - [Chu85] Peter L. Chu: "Quadrature Mirror Filter Design for an Arbitrary Number of Equal Bandwidth Channels" *IEEE Trans. ASSP* Vol. ASSP-33 No. 1, 1985 February, pp 203-218
 - [Clark81] Gregory A. Clark, Sanjit K. Mitra, and Sydney R. Parker: "Block Implementation of Adaptive Digital Filters" *IEEE Trans. CAS* Vol. CAS-28 No. 6, 1981 June, pp 584-592
 - [Clark83] Gregory A. Clark, Sydney R. Parker, and Sanjit K. Mitra: "A Unified Approach to Time and Frequency Domain Realization of FIR Adaptive Digital Filters" *IEEE Trans. ASSP* Vol. ASSP-31 No. 5, 1983 October, pp 1073-1083
 - [Cohen95] Leon Cohen: "Time-Frequency Analysis" Englewood Cliffs: Prentice-Hall 1995
 - [Cooley65] James W. Cooley and John W. Tukey: "An Algorithm for the Machine Computation of Complex Fourier Series" *Maths. Comp.* Vol. 19 No. 90, 1965, pp 297-301
 - [Copeland82] Greg C. Copeland: "Transmultiplexers Used as Adaptive Frequency Sampling Filters" *Proc. ICASSP* 1982 pp 319-323
 - [Cowan85] C.F.N. Cowan and P.M. Grant (editors): "Adaptive Filters" Englewood Cliffs: Prentice-Hall 1985
 - [Crochiere75] R.E. Crochiere and L.R. Rabiner: "Optimum FIR Digital Filter Implementations for Decimation, Interpolation, and Narrow-Band Filtering" *IEEE Trans. ASSP* Vol. ASSP-23 No. 5, 1976 October, pp 444-456
 - [Crochiere76] R.E. Crochiere, S.A. Webber, and J.L. Flanagan: "Digital Coding of Speech in Subbands" *Bell Sys. Tech. J.* Vol. 55 No. 8, 1976 October, pp 1069-1085
 - [Crochiere81] R.E. Crochiere and Lawrence R. Rabiner: "Interpolation and Decimation of Digital Signals - A Tutorial Review" *Proc. IEEE* Vol. 69 No. 3, 1981 March, pp 300-331
 - [Crochiere83] Ronald E. Crochiere and Lawrence R. Rabiner: "Multirate Digital Signal Processing" Englewood Cliffs: Prentice Hall 1983
 - [Deeba90] Amer S. Deeba and Sally L. Wood: "Convergence Rate Improvements for Frequency Domain Implementations of LMS Adaptive Filters" *Proc. ACSSC* 1990 pp 754-757
 - [DeCampos93] Marcellus L.R. De Campos, Paulo S.R. Diniz, and Andreas Antoniou: "Performance of LMS-Newton Adaptation Algorithms with Variable Convergence Factor in Non-Stationary Environments" *Proc. ICASSP* 1993, pp III-92 - III-95
 - [DeLeon95] Phillip L. De Leon II and Delores M. Etter: "Experimental Results with Increased Bandwidth Analysis Filters in Oversampled Subband Acoustic Echo Cancellers" *IEEE SP Ltrs.* Vol. 2 No. 1, 1995 January, pp 1-3
 - [Diniz92] P.S.R. Diniz and L.W.P. Biscainho: "Optimal Variable Step-Size Factor for Gradient-Based Adaptive Algorithms" *IEEE Trans. SP* Vol. 40 No. 11, 1992 November, pp 2825-2829
 - [Diniz95] Paulo S.R. Diniz, Marcellus L.R. de Campos, and Andreas Antoniou: "Analysis of LMS-Newton Adaptive Filtering Algorithms with Variable Convergence Factor" *IEEE Trans. SP* Vol. 43 No. 3, 1995 March, pp 617-627
 - [Dorbecker95] Matthias Dorbecker and Peter Vary: "Reducing the Delay of an Acoustic Echo Canceller with Subband Adaptation" *Proc. IWAENC* 1995 June, pp 103-106
-

-
- [Douglas92] Scott C. Douglas and Teresa H.-Y. Meng: "The Optimum Scalar Data Non-Linearity in LMS Adaptation for Arbitrary I.I.D. Inputs" *IEEE Trans. SP* Vol. 40 No. 6, 1992 June, pp 1566-1570
 - [Douglas94a] Scott C. Douglas and Teresa H.-Y. Meng: "Stochastic Gradient Adaptation Under General Error Criteria" *IEEE Trans. SP* Vol. 42 No. 6, 1994 June, pp 1335-1351
 - [Douglas94b] Scott Douglas and Teresa H.-Y. Meng: "Normalized Data Non-Linearities for LMS Adaptation" *IEEE Trans. SP* Vol. 42 No. 6, 1994 June, pp 1352-1365
 - [Duh90] W.J. Duh and J.L. Wu: "Two-Stage Circular-Convolution-Like Algorithm/Architecture for the Discrete Cosine Transform" *IEE Proc. F* Vol. 137 No. 6, 1990 December, pp 465-472
 - [Duhamel93] Pierre Duhamel, Mohsen Montazeri, and Katia Hilal: "Classical Adaptive Algorithms Seen as Recursive Structures" *Proc. ICASSP* 1993 pp III-496 - III-499
 - [Etter93] D.M. Etter, N.C. Carlson, and V.A. Margo: "Adaptive System Modeling with Critically Decimated Multirate Filters" *Proc. ISCAS* 1993 pp 68-71
 - [Evans93] Joseph B. Evans, Ping Xue, and Bede Liu: "Analysis and Implementation of Variable Step-Size Adaptive Algorithms" *IEEE Trans. SP* Vol. 41 No. 8, 1993 August, pp 2517-2535
 - [Ewada85] Ewada Ewada and Odile Macchi: "Tracking Error Bounds of Adaptive Nonstationary Filtering" *Automatica* Vol 21 No. 3, 1985 March, pp 293-305
 - [Farden81] D.C. Farden: "Tracking Properties of Adaptive Signal Processing Algorithms" *IEEE Trans. ASSP* Vol. ASSP-29 No. 6, 1981 June, pp 439-446
 - [Farhang91] B. Farhang-Boroujeny and S. Gazor: "Performance Analysis of Transform Domain Normalized LMS Algorithms" *Proc. ICASSP* 1991 July, pp 2133-2136
 - [Farhang92] B. Farhang-Boroujeny and S. Gazor: "Selection of Orthonormal Transforms for Improving Performance of the Transform Domain Normalized LMS Algorithm" *IEE Proc. -F* Vol. 139 No. 5, 1992 October, pp 327-335
 - [Farhang93] B. Farhang-Boroujeny: "Application of Orthonormal Transforms to Implementation of Quasi-LMS/Newton Algorithms" *IEEE Trans. SP* Vol. 41 No. 3, 1993 March, pp 1400-1405
 - [Farhang94a] B. Farhang-Boroujeny and S. Gazor: "Generalized Sliding FFT and Its Applications to Implementation of Block LMS Adaptive Filters" *IEEE Trans. SP* Vol. 42 No. 3, 1994 March, pp 532-538
 - [Farhang94b] B. Farhang-Boroujeny: "Variable Stepsize LMS Algorithm: New Developments and Experiments" *IEE Proc. VTSP* Vol. 141 No. 5, 1994 October, pp 311-317
 - [Farhang96a] B. Farhang-Boroujeny and Teng-Tiow Tay: "Transfer Function Identification with Filtering Techniques" *IEEE Trans. SP* Vol. 44 No. 11, 1996 November, pp 2865-2868
 - [Farhang96b] B. Farhang-Boroujeny: "Analysis and Efficient Implementation of Partitioned-Block LMS Adaptive Filters" *IEEE Trans. SP* Vol. 44 No. 6, 1996 June, pp 1334-1345
 - [Farhang96c] B. Farhang-Boroujeny and S. Gazor: "Performance of LMS-Based Adaptive Filters in Tracking a Time-Varying Plant" *IEEE Trans. SP* Vol. 44 No. 11, 1996 November, pp 2868-2871
 - [Feintuch93] Paul L. Feintuch, Neil J. Bershad, and Allen K. Lo: "A Frequency Domain Model for 'Filtered' LMS Algorithms - Stability Analysis, Design, and Elimination of the Training Mode" *IEEE Trans. SP* Vol. 41 No. 4, 1993 April, pp 1518-1531
 - [Ferrara80] Earl R. Ferrara: "Fast Implementation of LMS Adaptive Filters" *IEEE Trans. ASSP* Vol. ASSP-28 No. 4, 1980 August, pp 474-475
 - [Ferrara81] Earl R. Ferrara and Bernard Widrow: "The Time-Sequenced Adaptive Filter" *IEEE Trans. ASSP* Vol. ASSP-29 No. 3, 1981 June, pp 679-683
 - [Feuer85] A. Feuer: "Performance Analysis of the Block Least-Mean-Square Algorithm" *IEEE Trans. CAS*, Vol. CAS-32 No. 9, 1985 September, pp 960-963
-

-
- [Feuer93] Arie Feuer and Roberto Cristi: "On the Steady State Performance of Frequency Domain LMS Algorithms" *IEEE Trans. SP* Vol. 41 No. 1, 1993 January, pp 419-423
 - [Fitch85] J. Patrick Fitch, Edward J. Coyle, and Neal C. Gallagher, Jr.: "Root Properties and Convergence Rates of Median Filters" *IEEE Trans. ASSP* Vol. ASSP-33 No. 1, 1985 February, pp 230-240
 - [Florian88] Shaul Florian and Neil J. Bershad: "A Weighted Normalized Frequency Domain LMS Adaptive Algorithm" *IEEE Trans. ASSP* Vol. 36 No. 7 1988 July, pp 1002-1007
 - [Freeny80] Stanley L. Freeny: "TDM/FDM Translation as an Application of Digital Signal Processing" *IEEE Comm.* 1980 January, pp 5-15
 - [Frenzel92] Rudi Frenzel and Marcus E. Hennecke: "Using Prewhitening and Stepsize Control to Improve the Performance of the LMS Algorithm for Acoustic Echo Compensation" *Proc. ISCAS* 1992 pp 1930-1933
 - [Gallagher81] N.C. Gallagher and G.L. Wise: "A Theoretical Analysis of the Properties of Median Filters" *IEEE Trans. ASSP* Vol ASSP-29 No. 12, 1981 December, pp 1136-1141
 - [Gao91a] Frank Xiang Yang Gao: "Adaptive Linear and Non-Linear Filters" Ph.D. Dissertation, Carleton University, 1991 November
 - [Gao91b] Frank Xiang Yang Gao and W. Martin Snelgrove: "Adaptive Linearization of a Loudspeaker" *Proc. ICASSP* 1991 pp 3589-3592
 - [Gardner95] William G. Gardner: "Efficient Convolution Without Input-Output Delay" *JAES* Vol. 43 No. 3, 1995 March, pp 127-136
 - [Gilloire87] Andre Gilloire: "Experiments with Sub-Band Acoustic Echo Cancellers for Teleconferencing" *Proc. ICASSP* 1987 pp 2141-2144
 - [Gilloire88] Andre Gilloire and Martin Vetterli: "Adaptive Filtering in Sub-Bands" *Proc. ICASSP* 1988 pp 1572-1575
 - [Gilloire92] Andre Gilloire and Martin Vetterli: "Adaptive Filtering in Sub-bands with Critical Sampling: Analysis, Experiments, and Applications to Acoustic Echo Cancellation" *IEEE Trans. SP* Vol. 40 No. 8, 1992 August, pp 1862-1875
 - [Gilloire94] Andre Gilloire: "Performance Evaluation of Acoustic Echo Control: Required Values and Measurement Procedures" *Ann. Telecomm.* Vol. 49 No. 7-8, 1992, pp 368-372
 - [Gilloire95] Andre Gilloire: "Recent Advances in Adaptive Filtering Algorithms for Acoustic Echo Cancellation" *Proc. IWAENC* 1995 June, pp 115-134
 - [Gitlin77] Richard D. Gitlin and Francis R. Magee: "Self-Orthogonalizing Adaptive Equalization Algorithms" *IEEE Trans. Comm.* Vol. COM-25 No. 7, 1977 July, pp 667-672
 - [Giullemain96] Philippe Guillemain and Richard Kronland-Martinet: "Characterization of of Acoustic Signals Through Continuous Linear Time-Frequency Representations" *Proc. IEEE* Vol. 84 No. 4, 1996 April, pp 561-585
 - [Goubran96] Rafik A. Goubran, Carleton University, through course 94.563 (1995) and related conversation
 - [Gray72] Robert Molten Gray: "On the Asymptotic Eigenvalue Distribution of Toeplitz Matrices" *IEEE Trans. IT* Vol. IT-18 No. 6, 1972 November, pp 725-730
 - [Gudvangen95] S. Gudvangen and S.J. Flockton: "Modelling of Acoustic Transfer Functions for Echo Cancellers" *IEE Proc. VTSP* Vol. 142 No. 1, 1995 February, pp 47-51
 - [Guillemain96] Phillippe Guillemain and Richard Kronland-Martinet: "Characterization of Acoustic Signals Through Continuous Linear Time-Frequency Representations" *Proc. IEEE* Vol. 84 No. 4, 1996 April, pp 561-585
 - [Haddad88] Richard A. Haddad and Ali N. Akansu: "A New Orthogonal Transform for Signal Coding" *IEEE Trans. ASSP* Vol. 36 No 9, 1988 September, pp 1404-1411
-

-
- [Haddad95] Richard A. Haddad and Kyusik Park: "Modeling, Analysis, and Optimum Design of Quantized M-Band Filter Banks" *IEEE Trans. SP* Vol. 43 No. 11, 1995 November, pp 2540-2549
 - [Haneda94] Yoichi Haneda, Shoji Makino, and Yutaka Kaneda: "Common Acoustical Pole Modeling of Room Transfer Functions" *IEEE Trans. SAP* Vol. 2 No. 2, 1994 April, pp 320-328
 - [Haneda95] Yoichi Haneda, Shoji Makino, Junji Kojima, and Suehiro Shimauchi: "Implementation and Evaluation of an Acoustic Echo Canceller Using the Duo-Filter Control System" *Proc. IWAENC* 1995 June, pp 79-82
 - [Hansler91] Eberhard Hansler: "The Hands-Free Telephone Problem: An Annotated Bibliography" *Sig. Proc.* Vol. 27, 1992, pp 259-271
 - [Hansler94] Eberhard Hansler: "The Hands-Free Telephone Problem: An Annotated Bibliography Update" *Ann. Telecomm.* Vol. 49 No. 7-8, 1992, pp 360-367
 - [Hansler95] Eberhard Hansler: "The Hands-Free Telephone Problem: A Second Annotated Bibliography Update" *Proc. IWAENC* 1995 June, pp 107-114
 - [Haykin96] Simon Haykin: "Adaptive Filter Theory - Third Edition" Englewood Cliffs: Prentice-Hall 1996
 - [Hertz86] David Hertz, David Mansour, and Isaac Engel: "On Least-Squares Frequency Domain Adaptive Filters" *IEEE Trans. CAS* Vol. CAS-33 No. 3, 1986 March pp 335-337
 - [Higa95] Yoshito Higa, Hiroshi Ochi, and Shigenori Kinjo: "A Sub-Band Adaptive Filter with a Variable Analysis Filter Bank" *Proc. ACSSC* 1995
 - [Hirano92] Akihiro Hirano and Akihiko Sugiyama: "A Compact Multi-Channel Echo Canceller with a Single Adaptive Filter per Channel" *Proc. ISCAS* 1992 pp 1922-1925
 - [Huhn95] Tilman Huhn and Hans-Joachim Jentschel: "A Frequency-Sensitive and Time-Variant Stepsize Control for the Adaptive Echo Canceller" *Proc. IWAENC* 1995 June, pp 27-30
 - [Hull90] Andrew Hull and Kenneth Jenkins: "Transform Domain Adaptive Filtering with the Chirp-Z Transform" *Proc. ICASSP* 1990 pp 1437-1440
 - [Ikeda91] Shigeji Ikeda and Akihiko Sugiyama: "A Fast Convergence Algorithm for Adaptive FIR Filters with Coarsely Located Taps" *Proc. ICASSP* 1991 pp 1525-1529
 - [Iyer94] Umashankar Iyer, Majid Nayeri, and Hiroshi Ochi: "A Polyphase Structure for System Identification and Adaptive Filtering" *Proc. ICASSP* 1994 pp III-433 - III-436
 - [Iyer96] Umashankar Iyer, Majid Nayeri, and Hiroshi Ochi: "Polyphase-Based Adaptive Structure for Adaptive Filtering and Tracking" *IEEE Trans. CAS-II* Vol. 43 No. 3, 1996 March pp 220-233
 - [Jaggi90] S. Jaggi and A.B. Martinez: "Upper and Lower Bounds of the Misadjustment in the LMS Algorithm" *IEEE Trans. ASSP* Vol. 38, No. 1, 1990 January, pp 164-166
 - [Janssen86] A. J. E. M. Janssen, Raymond N. J. Veldhuis, and Lodewijk B. Vries: "Adaptive Interpolation of Discrete-Time Signals That Can Be Modeled as Autoregressive Processes" *IEEE Trans. ASSP* Vol. 34 No. 2, 1986 April, pp 317-330
 - [Jin96] Qu Jin, Zhi-Quan Luo and Kon Max Wong: "Optimum Filter Banks for Signal Decomposition and Its Application in Adaptive Echo Cancellation" *IEEE Trans. SP* Vol. 44 No. 7, 1996 July, pp 1669-1680
 - [Johns91] David A. Johns, W. Martin Snelgrove, and Adel S. Sedra: "Continuous-Time LMS Adaptive Recursive Filters" *IEEE Trans. CAS* Vol. 38 No. 7, 1991 July, pp 769-777
 - [Kellermann94] Walter Kellermann: "Some Properties of Echo Path Impulse Responses of Microphone Arrays and Consequences for Acoustic Echo Cancellation" *Proc. IWAENC* 1994, pp 39-43
 - [Kim95] Daeyoung Kim and Byeong Gi Lee: "Transform Domain IIR Filtering" *IEEE Trans. SP* Vol. 43 No. 10, 1995 October, pp 2431-2434
-

-
- [Knappe92] Michael Knappe: "Acoustic Echo Cancellation: Performance and Structures" M.Eng. Thesis, Ottawa-Carleton Institute for Electrical Engineering, 1992 September
 - [Knappe94] Michael Knappe and R.A. Goubran: "Steady-State Performance Limitations of Full-Band Acoustic Echo Cancellers" *Proc. ICASSP* 1994 pp II-73 - II-76
 - [Koipillai91] R.D. Koipillai and P.P. Vaidyanathan: "New Results on Cosine-Modulated FIR Filter Banks Satisfying Perfect Reconstruction" *Proc. ICASSP* 1991 pp 1793-1797
 - [Koizumi89] Nobuo Koizumi, Shoji Makino, and Hiroshi Oikawa: "Acoustic Echo Canceller with Multiple Echo Paths" *J. Acoust. Soc. Jpn. (E)* Vol. 10 No. 1, 1989, pp 39-44
 - [Kuo93a] Sen Kuo and Zhibing Pan: "Adaptive Acoustic Echo Cancellation Microphone" *J. Acoust. Soc. Am.* Vol. 93 No. 3, 1993 March, pp 1629-1636
 - [Kuo93b] Sen Kuo, and Zhibing Pan: "Distributed Acoustic Echo Cancellation System with Double-Talk Detector" *J. Acoust. Soc. Am.* Vol. 94 No. 6, 1993 December, pp 3057-3060
 - [Kurosawa87] Kaoru Kurosawa and Takuji Furusawa: "A Geometric Interpretation of Adaptive Algorithms" *Proc. Globecom* 1987 pp 49.7.1 - 49.7.5
 - [Kwan93] H.K. Kwan and P. Li: "High-Speed Realization of Adaptive Linear Phase FIR Digital Filters" *IEE Proc.-F* Vol. 140 No. 1, 1993 February, pp 48-54
 - [Kwong92] Raymond H. Kwong and Edward W. Johnston: "A Variable Step-Size LMS Algorithm" *IEEE Trans. SP* Vol. 40 No. 7, 1992 July, pp 1633-1642
 - [Lariviere97] Jeffery P. Lariviere: Related through conversation and tutorials at Carleton University, 1996 and 1997
 - [Lee84] Jae Chon Lee and Chong Kwan Un: "A Reduced Structure of the Frequency Domain Block LMS Adaptive Filter" *Proc. IEEE* Vol. 72 No.12, 1984 December, pp 1816-1818
 - [Lee85a] Jae Chon Lee, Chong Kwan Un, and D.K. Cho: "A Frequency-Weighted Block LMS Algorithm and Its Application to Speech Processing" *Proc. IEEE* Vol. 73 No. 6, 1985 June, pp 1137-1138
 - [Lee85b] Jae Chon Lee, Byong Ki Min, and Minsoo Suk: "Realization of Adaptive Digital Filters Using the Fermat Number Transform" *IEEE Trans. ASSP* Vol. ASSP-33 No. 4, 1985 August, pp 1036-1039
 - [Lee86a] Jae Chon Lee and Chong Kwan Un: "Block Realization of Multirate Adaptive Digital Filters" *IEEE Trans. ASSP* Vol. ASSP-34 No.1, 1986 February, pp 105-117
 - [Lee86b] Jae Chon Lee and Chong Kwan Un: "Performance of Transform Domain LMS Adaptive Digital Filters" *IEEE Trans. ASSP* Vol. ASSP-34, No. 3, 1986 June, pp 499-510
 - [Lee89] Jae Chon Lee and Chong Kwan Un: "Performance Analysis of Frequency Domain Block LMS Adaptive Digital Filters" *IEEE Trans. CAS* Vol. 36 No. 2, 1989 February, pp 173-189
 - [Lee93] Chang Long Lee and Chein-Wei Jen: "Binary Partition Algorithms and VLSI Architectures for Median and Rank-Order Filtering" *IEEE Trans SP* Vol. 41 No. 9, 1993 September, pp 2937-2942
 - [Leon89] Alberto Leon-Garcia: Probability and Random Processes for Electrical Engineering" Don Mills: Addison-Wesley 1989
 - [Lim89] Y.C. Lim and S.R. Parker: "Frequency Domain Modeling of Continuous Time Systems Using the Lattice Algorithms" *IEEE Trans. CAS* Vol. 36 No. 3, 1989 March, pp 429-433
 - [Lin95] Yuan-Pei Lin and P.P. Vaidyanathan: "Linear Phase Cosine-Modulated Maximally-Decimated Filter Banks with Perfect Reconstruction" *IEEE Trans. SP* Vol. 43 No. 11, 1995 November, pp 2525-2539
 - [Liu88a] J.-C. Liu and T.-P. Lin: "LMS-Based DHT Analyzer" *IEE Elec. Ltrs.* Vol. 24 No. 8, 1988 April, pp 483-485
-

-
- [Liu88b] J.-C. Liu and T.-P. Lin: "Running DHT and Real-Time DHT Analyzer" *IEE Elec. Ltrs.* Vol. 24 No. 8, 1988 April, pp 762-763
 - [Liu93] J.-C. Liu and T.-P. Lin: "Short-Time Hartley Transform" *IEE Proc. -F* Vol. 140 No. 3, 1993 June, pp 171-174
 - [Makhoul75] John Makhoul: "Linear Prediction: A Tutorial Review" *Proc. IEEE* Vol. 63 No. 4, 1975 April, pp 561-580
 - [Makhoul78] John Makhoul: "A Class of All-Zero Lattice Digital Filters: Properties and Applications" *IEEE Trans. ASSP* Vol. ASSP-26 No. 4, 1978 August, pp 304-314
 - [Makhoul81] John Makhoul: "On the Eigenvectors of Symmetric Toeplitz Matrices" *IEEE Trans. ASSP* Vol. ASSP-29 No. 4, 1981 August, pp 868-872
 - [Makino90] Shoji Makino and Yutaka Kaneda: "Acoustic Echo Canceller Algorithm Based on the Variation Characteristics of a Room Impulse Response" *Proc. ICASSP* 1990, pp 1133-1136
 - [Makino93] Shoji Makino, Yutaka Kaneda, and Nobuo Koizumi: "Exponentially-Weighted Step-Size NLMS Adaptive Filter Based on the Statistics of a Room Impulse Response" *IEEE Trans. SAP* Vol. 1 No. 1, 1993 January, pp 101-108
 - [Makino95] Shoji Makino: "Relationship Between the ES Family of Algorithms and Conventional Adaptive Algorithms" *Proc. IWAENC* 1995 June, pp 11-14
 - [Mansour82] David Mansour and Augustine H. Gray, Jr.: "Unconstrained Frequency Domain Adaptive Filter" *IEEE Trans. ASSP* Vol. 30 No. 5, 1982 October, pp 726-734
 - [Mao93] M. Mao and Z.J. Gan: "Statistical Analysis for the Convergence Rate of Signals to Median Filter Roots" *IEEE Trans. SP* Vol. 41 No. 7, 1993 July, pp 2499-2502
 - [Margo93] V.A. Margo, D.M. Etter, N.C. Carlson: "Multiple Short Length Adaptive Filters for Time-Varying Echo Cancellation" *Proc. ICASSP* 1993, pp I-161 - I-164
 - [Marshall89] D. F. Marshall, W. K. Jenkins, and J.J. Murphy: "Use of Orthogonal Transforms for Improving Performance of Adaptive Filters" *IEEE Trans. CAS* Vol. 36 No. 4, 1989 April, pp 474-483
 - [Marshall96] L. Gerald Marshall: "Analysis Procedure for Room Acoustics and Sound Amplification Systems Based on the Early-to-Late Sound Energy Ratio" *JAES* Vol. 44 No. 5, 1996 May, pp 373-381
 - [Martin94] Eric Martin, Andre Gilloire, and Patrice Le Scan: "CAD of Signal Processing Architectures: An Application to Acoustic Echo Control" *Ann. Telecomm.* Vol. 49 No. 7-8, 1992, pp 447-459
 - [Mathews90] V. John Mathews and Zhenhua Xie: "Stochastic Gradient Adaptive Filters with Gradient Adaptive Step-Sizes" *Proc. ICASSP* 1990 pp 1385-1389
 - [Mathews93] V. John Mathews and Zhenhua Xie: "A Stochastic Gradient Adaptive Filter with Gradient Adaptive Step-Size" *IEEE Trans. SP* Vol. 41 No. 6, 1993 June, pp 2075-2087
 - [Mboup91] Mamadou Mboup and Madeleine Bonnet: "IIR Filtering for Acoustic Echo Cancellation" *Proc. ACSSC* 1991 pp 203-206
 - [Mikhael86] Wasfy B. Mikhael, Frank H. Wu, Leonid G. Kazovsky, George S. Kang, and Lawrence J. Fransen: "Adaptive Filters with Individual Adaptation of Parameters" *IEEE Trans. CAS* Vol. CAS-33 No. 7, 1986 July, pp 677-685
 - [Mikhail89] Wasfy B. Mikhael and Frank H. Wu: "A Fast Block FIR Adaptive Digital Filtering Algorithm with Individual Adaptation of Parameters" *IEEE Trans. CAS* Vol. 36 No. 1, 1989 January, pp 1-10
 - [Morgan95a] Dennis R. Morgan: "Slow Asymptotic Convergence of LMS Acoustic Echo Cancellers" *IEEE Trans. SAP* Vol. 3 No. 2, 1995 March, pp 126-136
 - [Morgan95b] Dennis R. Morgan and James C. Thi: "A Delayless Subband Adaptive Filter Architecture" *IEEE Trans. SP* Vol. 43 No. 8, 1995 August, pp 1819-1830
-

-
- [Moulines95] Eric Moulines, Omar Ait Amrane, and Yves Grenier: "The Generalized Multi-delay Adaptive Filter: Structure and Convergence Analysis" *IEEE Trans. SP* Vol. 43 No. 1, 1995 January, pp 14-28
 - [Murthy92] N. Rama Murthy and M.N.S. Swamy: "On the VLSI Implementation of Real-Time Order-Statistic Filters" *IEEE Trans. SP* Vol. 40 No. 5, 1992 May, pp 1241-1252
 - [Nagumo67] Jin-Ichi Nagumo and Atsuhiko Noda: "A Learning Method for System Identification" *IEEE Trans. AC* Vol. AC-12 No. 3, 1967 June, pp 282-287
 - [Narasimha79] Madihally J. Narasimha and Allen M. Peterson: "Design of a 24-Channel Transmultiplexer" *IEEE Trans. ASSP* Vol. ASSP-27 No. 6, 1979 December, pp 752-762
 - [Narayan81] S. Shankar Narayan, and A.M. Peterson: "Frequency Domain Least-Mean-Square Algorithm" *Proc. IEEE* Vol. 69 No. 1, 1981 January, pp 124-126
 - [Narayan83] S. Shankar Narayan, Allen M. Peterson, and Madihally J. Narasimha: "Transform Domain LMS Algorithm" *IEEE Trans. ASSP* Vol. ASSP-31 No. 3, 1983 June, pp 609-615
 - [Nayeri89] Majid Nayeri and W. Kenneth Jenkins: "Alternate Realizations to Adaptive IIR Filters and Properties of their Performance Surfaces" *IEEE Trans. CAS* Vol. 36 No. 4, 1989 April, pages 485-496
 - [Naylor94] Patrick Naylor, Jesus Alcazar, Jerome Boudy, and Yves Grenier: "Enhancement of Hands-Free Telecommunications" *Ann. Telecomm.* Vol. 49 No. 7-8, 1992, pp 373-379
 - [Nguyen88] Truong Q. Nguyen and P.P. Vaidyanathan: "Maximally Decimated PR-FIR Filter Banks with Pairwise Mirror-Image Analysis and Synthesis Filter Banks" *IEEE Trans. ASSP* Vol. 36 No. 5, 1988 May, pp 692-706
 - [Niedrist93] Georg Niedrist: "Echo Suppression for Loudspeaker-Microphone System Measurements" *J. AES* Vol. 41 No. 3, 1993 March, pp 143-153
 - [Ogue83] Juan Carlos Ogue, Tsuneo Saito, and Yukio Hoshiko: "A Fast Convergence Frequency Domain Adaptive Filter" *IEEE Trans. ASSP* Vol. ASSP-31 No. 5, 1983 October, pp 1312-1314
 - [Ogunfun90] Tokunbo Ogunfunmi and Allen Peterson: "Alternative Implementations for the Frequency Domain LMS Adaptive Filter" *Proc. ICASSP* Vol. 2 1990, pp1441-1444
 - [Oppenheim89] Alan V. Oppenheim and Robert W. Schaffer: "Discrete-Time Signal Processing" Englewood Cliffs: Prentice-Hall, 1989
 - [Petraglia93] Mariane R. Petraglia and Sanjit K. Mitra: "Performance Analysis of Adaptive Filter Structures Based on Sub-Band Decomposition" *Proc. ISCAS* 1993 pp 60-63
 - [Picci84] Giorgio Picci and Giancarlo Prati: "Self-Orthogonalizing Adaptive Equalization in the Discrete Frequency Domain" *IEEE Trans. Comm.* Vol. COM-32 No. 4, 1984 April pp 371-379
 - [Prado94] Jacques Prado and Eric Moulines: "Frequency Domain Filtering with Applications to Acoustic Echo Cancellation" *Ann. Telecomm.* Vol. 49 No. 7-8, 1992, pp 414-428
 - [Press90] William H. Press, Brian P. Flannery, Saul A. Teukolsky, and William T. Vetterling: "Numerical Recipes in C - The Art of Scientific Computing" Cambridge: Cambridge Press 1990
 - [Princen86] John P. Princen and Alan Bernard Bradley: "Analysis/Synthesis Filter Bank Design Based on Time Domain Alias Cancellation" *IEEE Trans. ASSP* Vol. ASSP-34 No. 5, 1986 October, pp1153-1161
 - [Princen95] John Princen: "The Design of Non-Uniform Modulated Filter Banks" *IEEE Trans. SP* Vol. 43 No. 11, 1995 November, pp 2550-2560
 - [Proakis88] John G. Proakis and Dimitris G. Manolakis: "Introduction to Digital Signal Processing" New York: MacMillan 1988
 - [Rabiner78] Lawrence R. Rabiner and Ronald W. Schaefer: "Digital Processing of Speech Signals" Englewood Cliffs: Prentice-Hall 1978
-

-
- [Ramachan96] Kannan Ramachandran, Martin Vetterli, and Cormac Herley: "Wavelets, Sub-band Coding, and Best Bases" *Proc. IEEE* Vol. 84 No. 4, 1996 April, pp541-560
 - [Ramstad91] Tor A. Ramstad and Joar P. Tanem: "Cosine-Modulated Analysis-Synthesis Filter Banks with Critical Sampling and Perfect Reconstruction" *Proc. ICASSP* 1991, pp 1789-1793
 - [Rao90] K.R. Rao and P. Yip: "Discrete Cosine Transform - Algorithms, Advantages, Applications" San Diego: Academic Press 1990
 - [Reed81] Francis A. Reed and Paul L. Feintuch: "A Comparison of LMS Cancellers Implemented in the Frequency Domain and the Time Domain" *IEEE Trans. CAS* Vol. CAS-28 No. 6, 1981 June, pp 610-615
 - [Restrepo94] Alfredo Restepo and Liliana Chacon: "A Smoothing Property of the Median Filter" *IEEE Trans. SP* Vol. 42 No. 6, 1994 June, pp 1553-1555
 - [Rife89] Douglas D. Rife and John Vanderkooy: "Transfer Function Measurement with Maximum-Length Sequences" *JAES* Vol. 37 No. 6, 1989 June, pp 419-443
 - [Rossi96] Michel Rossi, Jin-Yun Zhang, and Willem Steenaert: "Iterative Least-Squares Approach for the Design of Filter Banks" submitted for review to *IEEE Trans. SP* 1996 March
 - [Rothweiler83] Joseph H. Rothweiler: "Polyphase Quadrature Filters: A New Subband Coding Technique" *Proc. ICASSP* 1983 pp 1280-1283
 - [Ryan92] James Gregory Ryan: "Subband Adaptive Filters" M.Eng. Thesis, Ottawa-Carleton Institute for Electrical Engineering, 1992 March
 - [Satorius79] E. H. Satorius and S.T. Alexander: "Channel Equalization Using Adaptive Lattice Algorithms" *IEEE Trans. Comm.* Vol. COM-27 No. 6, 1979 June, pp 899-905
 - [Schaefer75] Ronald W. Schaefer and Lawrence R. Rabiner: "Digital Representation of Speech Signals" *Proc. IEEE* Vol. 63 No. 4, 1975 April, pp 662-677
 - [Schoenle93] M. Schoenle, N. Fliege, and U. Zolzer: "Parametric Approximation of Room Impulse Responses by Multirate Systems" *Proc. ICASSP* 1993 pp I-153 - I-156
 - [Shah95] S.A.H. Shah and C.F.N. Cowan: "Modified Stochastic Gradient Algorithm using Non-Quadratic Cost Functions for Data Echo Cancellation" *Proc. VTSF* Vol 142 No. 3, 1995 June, pp 187-191
 - [Shensa79] Mark J. Shensa: "The Spectral Dynamics of Evolving LMS Adaptive Filters" *Proc. ICASSP* 1979 pp 950-953
 - [Sherlock92] B.G. Sherlock and D.M. Monro: "Moving Discrete Fourier Transform" *IEE Proc. -F* Vol. 139 No. 4, 1992 August, pp 279-282
 - [Shoval95] Ayal Shoval, David A. Johns, and W. Martin Snelgrove: "Comparison of DC Offset Effects in Four LMS Adaptive Algorithms" *IEEE Trans. CAS-II* Vol. 42 No. 3, 1995 March, pp 176-185
 - [Shynk86] John J. Shynk and Bernard Widrow: "Bandpass Adaptive Pole-Zero Filtering" *Proc. ICASSP* 1986 pp 40.5.1-40.5.4
 - [Shynk92] John J. Shynk: "Frequency Domain and Multirate Adaptive Filtering" *IEEE.SP* 1992 January, pp 14-36
 - [Slock93] Dirk Slock: "On the Convergence Behaviour of the LMS and Normalized LMS Algorithms" *IEEE Trans. SP* Vol. 41 No. 9, 1993 September, pp 2811-2825
 - [Smith86] Mark J.T. Smith and Thomas P. Barnwell: "Exact Reconstruction Techniques for Tree-Structured Subband Coders" *IEEE Trans. ASSP* Vol. ASSP-34 No. 3, 1986 June, pp 434-441
 - [Smith87] Mark J.T. Smith and Thomas P. Barnwell: "A New Filter Bank Theory for Time-Frequency Representation" *IEEE Trans. ASSP* Vol. ASSP-35 No. 3, 1987 March, pp 314-327
 - [Snelgrove97] W. Martin Snelgrove, Carleton University, through conversation
-

-
- [Sodagar95] Iraj Sodagar, Kambiz Nayebi, Thomas P. Barnwell, and Mark J.T. Smith: "Time-Varying Analysis-Synthesis Systems Based on Filter Banks and Post Filtering" *IEEE Trans. SP* Vol. 43 No. 11, 1995 November, pp 2512-2524
 - [Sommen87] Piet C.W. Sommen, Piet J. Van Gerwen, Henk J. Kotmans, and A.J.E.M. Janssen: "Convergence Analysis of a Frequency-Domain Adaptive Filter with Exponential Power Averaging and Generalized Window Function" *IEEE Trans. CAS* Vol. CAS-34 No.7, 1987 July, pp 788-798
 - [Sommen89] Piet C.W. Sommen: "Partitioned Frequency Domain Adaptive Filters" *Proc. ACSSC* 1989 pp 677-680
 - [Sommen92] Piet C.W. Sommen and E. De Wilde: "Equal Convergence Conditions for Normal- and Partitioned-Frequency Domain Adaptive Filters" *Proc. ICASSP* 1992 pp IV-69 - IV-72
 - [Soo90] Jia-Sien Soo and Khee K. Pang: "Multidelay Block Frequency Domain Adaptive Filter" *IEEE Trans. ASSP* Vol. 38 No. 2 1990 February, pp 373-376
 - [Stanley84] William D. Stanley, Gary R. Dougherty, and Ray Dougherty: "Digital Signal Processing" Englewood Cliffs: Prentice-Hall (Reston) 1984
 - [Stark94] Henry Stark and John W. Woods: "Probability, Random Processes, and Estimation Theory for Engineers" Englewood Cliffs: Prentice-Hall 1994
 - [Strang76] Gilbert Strang: "Linear Algebra and its Applications" New York: Academic Press 1976
 - [Tanrikulu96] O. Tanrikulu and A. G. Constantinides: "The LMK Algorithm with Time-Varying Forgetting Factor for Adaptive System Identification in Additive Noise" *Proc. ICASSP* 1996 CD-ROM
 - [Treichler87] John C. Treichler, Richard Johnson, and Michael Larimore: "Theory and Design of Adaptive Filters" John Wiley, 1987
 - [Usevitch96] Bryan E. Usevitch and Michael T. Orchard: "Adaptive Filtering Using Filter Banks" *IEEE Trans. CAS-II* Vol. 43 No. 3, 1996 March, pp 255-265
 - [Vaidyan87a] P.P. Vaidyanathan: "Theory and Design of M-Channel Maximally Decimated Quadrature Mirror Filters Having the Perfect Reconstruction Property" *IEEE ASSP* Vol. 35 No. 4, 1987 April, pp 476-492
 - [Vaidyan87b] P.P. Vaidyanathan: "Quadrature Mirror Filter Banks, M-Band Extensions and Perfect Reconstruction Techniques" *IEEE ASSP* 1987 July, pp 4-20
 - [Vaidyan90] P.P. Vaidyanathan: "Multirate Digital Filters, Filter Banks, Polyphase Networks, and Applications: A Tutorial" *Proc. IEEE* Vol. 78 No. 1, 1990 January, pp 56-93
 - [Vaidyan93] P.P. Vaidyanathan: "Multirate Systems and Filter Banks" Englewood Cliffs: Prentice-Hall 1993
 - [Vanderkooy94] John Vanderkooy: "Aspects of MLS Measuring Systems" *J.AES* Vol. 42 No. 4, 1994 April, pp 219-231
 - [VanSchyn96] André Van Schyndel, Bell-Northern Research Ltd., through correspondence and discussion
 - [Vetterli87] Martin Vetterli: "A Theory of Multirate Filter Banks" *IEEE Trans. ASSP* Vol. ASSP-35 No. 3, 1987 March, pp 356-372
 - [Vetterli92] Martin Vetterli and Cormac Herley: "Wavelets and Filter Banks: Theory and Design" *IEEE Trans. SP* Vol. 40 No. 9, 1992 September, pp 2207-2232
 - [Vetterli95] Martin Vetterli and Jelena Kovacevic: "Wavelets and Subband Coding" Englewood Cliffs: Prentice-Hall 1995
 - [Wada95] Sigeo Wada: "Design of Nonuniform Division Multirate FIR Filter Banks" *IEEE Trans. CAS-II* Vol. 42 No. 2, 1995 February, pp 115-124
 - [Walach84] Eugene Walach and Bernard Widrow: "The Least-Mean-Fourth Adaptive Algorithm and Its Family" *IEEE Trans. IT* Vol. 30 No. 3, 1984 March, pp 275-283
-

-
- [Webster95] Tim Webster: "Tracking Performance of Acoustic Echo Cancellers" M.Eng. Thesis, Ottawa-Carleton Institute for Electrical Engineering, 1995 July
 - [Wendt86a] Peter D. Wendt, Edward J. Coyle, and Neal C. Gallagher, Jr.: "Some Convergence Properties of Median Filters" *IEEE Trans. CAS* Vol. CAS-33 No. 3, 1986 March, pp 276-286
 - [Wendt86b] Peter D. Wendt, Edward J. Coyle, and Neal C. Gallagher, Jr.: "Stack Filters" *IEEE Trans. ASSP* Vol. ASSP-34 No. 4, 1986 August, pp 898-911
 - [Weslowski92] K. Weslowski, C.M. Zhao, and W. Rupprecht: "Adaptive LMS Transversal Filters with Controlled Length" *IEE Proc. -F* Vol. 139 No. 3, 1992 June, pp 233-238
 - [Widrow75] Bernard Widrow, John McCool, and Michael Ball: "The Complex LMS Algorithm" *Proc. IEEE* Vol. 63 No. 4, 1975 April, pp 719-720
 - [Widrow76] Bernard Widrow and John M. McCool: "Stationary and Nonstationary Learning Characteristics of the LMS Filter" *Proc. IEEE* Vol. 64 No. 8, 1976 August, pp 1151-1162
 - [Widrow84] Bernard Widrow and Eugene Walach: "On the Statistical Efficiency of the LMS Algorithm with Stationary and Non-Stationary Inputs" *IEEE Trans. IT* Vol. IT-30 No. 2, 1984 March, pp 211-221
 - [Widrow85] Bernard Widrow and Samuel D. Stearns: "Adaptive Signal Processing" Englewood Cliffs: Prentice-Hall, 1985
 - [Widrow87] Bernard Widrow, Phillippe Beaudrenghien, Martin Vetterli, and Paul F. Titchener: "Fundamental Relations Between the LMS Algorithm and the DFT" *IEEE Trans. CAS* Vol. CAS-34 No. 7, 1987 July, pp 814-820
 - [Xu89] Min Xu and Yves Grenier: "Time-Frequency Domain Adaptive Filters" *Proc. ICASSP* 1989 pp 1154-1157
 - [Yang93] Yoon Gi Yang, Nam Ik Cho, and Sang Uk Lee: "On the Performance Analysis of the Sub-Band Adaptive Digital Filter" *Proc. ISCAS* 1993 pp 64-67
 - [Yasukawa93] Hiroshi Yasukawa and Shoji Shimada: "An Acoustic Echo Canceller Using Sub-Band Sampling and Decorrelation Methods" *IEEE Trans. SP* Vol. 41 No. 2, 1993 February, pp 926-930
 - [Yim92] W.H. Yim, F.P. Coakley, and B.G. Evans: "Extended Polyphase Structures for Multirate DSP" *IEE Proc. -F* Vol. 139 No. 4, 1992 August, pp 273-277
 - [Youn83] Dae Hee Youn, Nasir Ahmed, and G. Clifford Carter: "Magnitude-Squared Coherence Function Estimation: An Adaptive Approach" *IEEE Trans. ASSP* Vol. ASSP-31 No. 1, 1983 February, pp 137-142
 - [Yuan94] Hai Yan Yuan: "Dynamic Behaviour of Acoustic Echo Cancellation" M.Eng. Thesis, Ottawa-Carleton Institute for Electrical Engineering, 1994 May
 - [Zelinski77] Rainer Zelinski and Peter Noll: "Adaptive Transform Coding of Speech Signals" *IEEE Trans. ASSP* Vol. ASSP-25 No. 4, 1977 August, pp 299-309
-

A4

Bibliography

This appendix provides lists of abbreviations, ISBN/ISSN numbers, and library catalogue codes for all periodicals and texts cited in the references.

TABLE A4.1 Catalogue Numbers for cited texts.

Author	Title	ISBN	Call Number
Akansu and Haddad	Multiresolution Signal Decomposition - Transforms, Subbands, Wavelets	0-12-047140-X	TK5102.5A414
Anton	Elementary Linear Algebra - Sixth Edition	0-471-50900-0	QA184.A57
Brigham	The Fast Fourier Transform and Its Applications	0-13-307505-2	QA403.B75
Burrus et al.	Computer Based Exercises for DSP using Matlab	0-13-219825-8	TK5102.9C57
Cohen	Time-Frequency Analysis	0-13-594532-1	TK5102.9C557
Cowan and Grant	Adaptive Filters	0-13-004037-1	TK7872.F5A33
Crochiere & Rabiner	Multirate Digital Signal Processing	0-13-605162-6	TK5102.5C76
Haykin	Adaptive Filter Theory - Third Edition	0-13-322760-X	TK7872.F5H368
Leon-Garcia	Probability & Random Processes for Elec. Engineering	0-201-12906-X	TK153.L425
Press, Flannery, et al.	Numerical Recipes in C	0-521-35465-X	QA76.73.C15N86
Proakis & Manolakis	Introduction to Digital Signal Processing	0-02-396810-9	TK5102.5P677
Rao and Yip	Discrete Cosine Transform - Algorithms, Applications	0-12-580203-X	TK5102.5R335
Stanley, Dougherty...	Digital Signal Processing - Second Edition	0-8359-1321-X	TK5102.5S69
Stark and Woods	Probability, Random Processes, & Estimation Theory	0-13-728791-7	TA340.S7
Strang	Linear Algebra and Its Applications	0-12-673650-2	QA184.S8
Treichler et al.	Theory and Design of Adaptive Filters	0-471-83220-0	TK7872.F5T65
Vaidyanathan	Multirate Systems and Filter Banks	0-13-605718-7	TK5102.5V24
Vetterli/Kovacevic	Wavelets and Subband Coding	0-13-097080-8	TK5102.9V48
Widrow and Stearns	Adaptive Signal Processing	0-13-004029-0	TK5102.5W537

TABLE A4.2 Titles and indices for serial publications.

Abbreviation	Full Title of Publication	ISSN	Call Number
Ann. Telecomm.	Annals des Telecommunications	0003-4347	-
Automatica	Automatica (Int'l Federation of Automatic Control)	-	-
Bell Sys. Tech. J.	Bell Systems Technical Journal	0005-8580	TK1.B425
IEE Elec. Ltrs.	I.E.E. Electronics Letters	-	-
IEE Proc. -F	I.E.E. Proceedings of Radar and Signal Processing	0956-375X	-
IEE.Proc. VISIP	I.E.E. Proceedings of Vision, Image, and Signal Proc.	1350-245X	TK5102.9I43
IEEE.ASSP	I.E.E.E. Acoustics, Speech, Signal Proc. Magazine	0096-3518	TK5981.I2042
IEEE.COMM	I.E.E.E. Communications Magazine	0163-6804	TK5101.A1I526
IEEE.SPECTRUM	I.E.E.E. Spectrum Magazine	0018-9235	TK1.I15
IEEE.SP	I.E.E.E. Signal Processing Magazine	1053-5888	QC221.I22
IEEE Trans. ASSP	I.E.E.E. Trans on Acoustics, Speech, and Signal Proc.	1053-587X	TK5981.I2042
IEEE Trans. AC	I.E.E.E. Transactions on Automatic Control	-	-
IEEE Trans. CAS	I.E.E.E. Transactions on Circuits and Systems	1057-7122	TK3226.A1I1822
IEEE Trans. CAS-I	I.E.E.E. Transactions on Circuits and Systems Part I: Fundamental Theory and Applications	1057-7122	TK3226.A1I1822
IEEE Trans. CAS-II	I.E.E.E. Transactions on Circuits and Systems Part II: Analog and Digital Signal Processing	1057-7130	TK3226.A1I1823
IEEE Trans. Comm.	I.E.E.E. Transactions on Communications	0090-6778	TK5101.I204
IEEE Trans. IT	I.E.E.E. Transactions on Information Theory	0018-9448	Q175.I2
IEEE Trans. SP	I.E.E.E. Transactions on Signal Processing	1053-587X	TK5981.I20422
IEEE Trans. SAP	I.E.E.E. Transactions on Speech and Audio Processing	1063-6676	TK7882.I185
IEEE SP Ltrs.	I.E.E.E. Signal Processing Letters	1070-9908	TK5101.A1I35
Proc. IEEE	Proceedings of the I.E.E.E.	0018-9219	TK5700.I61
Proc. ACSSC	I.E.E.E. Asilomar Conf Signals, Sys, and Computers	-	TK7801.A83
Proc. GLOBECOM	I.E.E.E. Globecom	-	-
Proc. ICASSP	I.E.E.E. Int'l Conf. on Acoustics, Speech, Sig. Proc.	-	TK7882.S65I16
Proc. ISCAS	I.E.E.E. Int'l Symposium on Circuits and Systems	-	TK454.I2
Proc. IWAENC	I.E.E.E. Int'l W'shop on Acoustic Echo/Noise Ctrl	-	-
Proc. MSCAS	I.E.E.E. Midwest Symposium on Circuits and Systems	-	TK3226.M553
Maths. Comp.	Mathematics of Computation	0025-5718	QA1.M3
J.AES	Journal of the Audio Engineering Society	0004-7554	TK5981.I83
J. Acoust. Soc. Am	Journal of the Acoustic Society of America	0001-4966	QC221.A4
Sig. Proc.	Signal Processing - Elsevier Science BV	0165-1684	-

IMAGE EVALUATION TEST TARGET (QA-3)

APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989

© 1993, Applied Image, Inc., All Rights Reserved