

Efficient Subband Structures for Acoustic Echo Cancellation in Nonstationary and Nonlinear Environments

submitted by

Trevor G. Burton, B.Sc.E., M.A.Sc.

A thesis submitted to
the Faculty of Graduate and Postdoctoral Affairs
in partial fulfilment of
the requirements for the degree of

Doctor of Philosophy

Ottawa-Carleton Institute for Electrical and Computer Engineering
Faculty of Engineering and Design
Department of Systems and Computer Engineering
Carleton University
Ottawa, Ontario, Canada K1S 5B6
September 20, 2010

Copyright © 2010

Trevor G. Burton



Library and Archives
Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 978-0-494-70545-2
Our file Notre référence
ISBN: 978-0-494-70545-2

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

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

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

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

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.


Canada

Abstract

The focus of this thesis is the development of computationally efficient adaptive filter structures capable of operating in nonstationary and nonlinear environments, with application to the acoustic echo cancellation (AEC) problem in hands-free communication systems. The specific nonstationary environments considered are changing conditions in the acoustic environment resulting from the movement of people and objects, as well as changing operating conditions of hands-free devices. The particular nonlinear environments arise from the harmonic loudspeaker distortion in hands-free devices.

The linear and nonlinear echo path components for several hands-free systems are identified based on experimentally measured data collected under the nonstationary and nonlinear environments previously outlined. The subsequent analysis revealed that these components and their fluctuations are confined to specific time and frequency regions, which provides insight for achieving computational complexity reduction in echo cancellers (ECs) operating in these environments.

An oversampled subband EC structure based on a generalized gradient proportionate step size adaptive algorithm is proposed for AEC in fast changing environments. The structure's enhanced echo cancellation performance compared to fullband ECs is verified through computer simulations. Improvements in echo return loss enhancement (ERLE) of up to 7 decibels (dBs) are realized compared to an equivalent fullband structure, based on measured data from hands-free systems under changing conditions, while requiring decreased complexity.

AEC in the presence of nonlinear loudspeaker distortion is addressed by developing an oversampled subband EC structure based on adaptive Volterra filters. Experimental results validate the structure's equivalent nonlinear echo attenuation ability compared to its fullband counterpart, while requiring much lower complexity. Compared to a linear EC, the proposed structure achieves up to 7 dB higher ERLE with a similar computational cost.

For AEC in changing nonlinear environments, an oversampled subband EC structure with adaptive Volterra filters based on a generalized gradient proportionate variable step size adaptive algorithm is proposed. The structure improves echo cancellation performance by up to 5 dB higher ERLE with significantly reduced computational expense compared to an equivalent fullband structure, based on experimental results with measured hands-free system data.

Acknowledgments

First and foremost I would like to thank my supervisor, Dr. Rafik Goubran, for his guidance, motivation, and support during the course of my work. I would also like to express my appreciation to the Faculty of Graduate and Postdoctoral Affairs along with the Department of Systems and Computer Engineering (SCE) at Carleton University, the Natural Sciences and Engineering Research Council of Canada (NSERC), Ontario Graduate Scholarships in Science and Technology (OGS-ST), Ontario Centres of Excellence (OCE), and Mitel Networks for their financial support.

I would also like to thank the SCE technical staff for keeping the DSP lab running smoothly and for attending to my equipment needs. A special thanks to Dr. Franck Beaucoup for providing inspiration and valuable feedback on the subband Volterra structure in Chapter 5 of this thesis, and to my fellow graduate students in the DSP lab both past and present for their friendship and helpful discussions. Finally, I would like to thank my family, especially my parents, for their love and support.

Table of Contents

Abstract	ii
Acknowledgments	iv
Table of Contents	v
List of Figures	ix
List of Tables	xix
List of Symbols	xxi
List of Abbreviations	xxvii
Chapter 1 Introduction	1
1.1 Problem Statement	1
1.2 Thesis Objectives and Motivation	4
1.3 Thesis Contributions	5
1.4 Thesis Organization and Scope	8
Chapter 2 Review of Acoustic Echo Cancellation	11
2.1 The Acoustic Echo Cancellation Problem	12
2.1.1 Acoustic Echo Cancellation Structure	12
2.1.1.1 Adaptive Algorithms for Acoustic Echo Cancellation	15
2.1.1.2 Subband Structures for Acoustic Echo Cancellation .	18

2.2	Performance Limitations in Acoustic Echo Cancellation	23
2.2.1	Acoustic Echo Cancellation in the Presence of Loudspeaker Nonlinearity	25
2.2.1.1	Nonlinear Loudspeaker Modeling Using Volterra Fil- ters in Acoustic Echo Cancellation	27
2.2.2	Acoustic Echo Cancellation in Nonstationary Environments	33
2.2.2.1	Tracking Algorithms for Acoustic Echo Cancellation	35
2.3	Summary	37
Chapter 3 Experimental Measurements and Data Analysis		38
3.1	Measurement Methodology	38
3.2	Measured Linear Echo Path Impulse Responses	41
3.3	Measured Quadratic Volterra Kernels	44
3.4	Measured Harmonic Distortion	47
3.5	Summary	52
Chapter 4 Subband Structure for Acoustic Echo Cancellation in Nonstationary Environments		54
4.1	Generalized Proportionate Variable Step Size Subband Adaptive Filter Structure	55
4.1.1	Description	55
4.1.2	Computational Complexity Analysis	60
4.2	Experimental Results	67
4.2.1	Methodology	67
4.2.2	Simulation Results Based on Measured Impulse Responses	69
4.2.2.1	White Noise Excitation Signal	71
4.2.2.2	Speech Excitation Signal	85
4.2.3	Results Based on Measured Reference and Input Data	91
4.2.4	Computational Complexity Results	101

4.3	Summary	108
Chapter 5	Subband Volterra Structure for Acoustic Echo Cancel-	
	lation	111
5.1	Subband Volterra Filter Structure	112
5.1.1	Description	112
5.1.2	Computational Complexity Analysis	115
5.2	Experimental Results	118
5.2.1	Methodology	118
5.2.2	Simulation Results Based on Measured Volterra Kernels . . .	119
5.2.2.1	White Noise Excitation Signal	119
5.2.2.2	Speech Excitation Signal	127
5.2.3	Results Based on Measured Reference and Input Data	130
5.2.4	Computational Complexity Results	141
5.3	Summary	145
Chapter 6	Subband Volterra Structure for Acoustic Echo Cancel-	
	lation in Nonstationary Environments	147
6.1	Generalized Proportionate Variable Step Size Subband Adaptive Volterra Filter Structure	148
6.1.1	Description	148
6.1.2	Computational Complexity Analysis	151
6.2	Experimental Results	156
6.2.1	Methodology	156
6.2.2	Simulation Results Based on Measured Volterra Kernels . . .	157
6.2.2.1	White Noise Excitation Signal With a Changing Quadratic Volterra Kernel	157
6.2.2.2	White Noise Excitation Signal With a Changing Lin- ear and Quadratic Volterra Kernel	163

6.2.3	Results Based on Measured Reference and Input Data	170
6.2.4	Computational Complexity Results	179
6.3	Summary	183
Chapter 7	Conclusions and Future Research	186
7.1	Summary of Research	186
7.2	Summary of Contributions	191
7.3	Suggestions for Future Research	193
References	196

List of Figures

Figure 1.1:	Overview of how key thesis chapters are linked together.	10
Figure 2.1:	Acoustic echo in a hands-free communication system.	13
Figure 2.2:	Linear system model for acoustic echo cancellation.	13
Figure 2.3:	Acoustic echo cancellation using subband adaptive filters.	18
Figure 2.4:	Acoustic echo cancellation using oversampled subband adaptive filters.	21
Figure 2.5:	Cross sectional diagram of an electrodynamic loudspeaker.	25
Figure 2.6:	Nonlinear acoustic echo cancellation using a second order adaptive Volterra filter.	28
Figure 2.7:	Nonlinear acoustic echo cancellation using a fully coupled second order adaptive Volterra filter.	30
Figure 2.8:	Nonlinear acoustic echo cancellation using a partially decoupled second order adaptive Volterra filter.	31
Figure 3.1:	Office room acoustic echo cancellation measurement setup.	39
Figure 3.2:	Commercial phone sets.	40
Figure 3.3:	Measured linear impulse responses from the phone set configurations.	42
Figure 3.4:	Measured linear impulse responses from the unobstructed and obstructed conference phone configurations.	43
Figure 3.5:	Measured linear frequency responses from the unobstructed and obstructed conference phone configurations.	44

Figure 3.6:	Measured quadratic Volterra kernel and corresponding frequency response for the conference phone configuration.	45
Figure 3.7:	Measured quadratic Volterra kernel and corresponding frequency response for the smartphone configuration.	46
Figure 3.8:	Measured quadratic Volterra kernel difference and corresponding frequency response from the conference phone at low and high volume signal levels.	46
Figure 3.9:	Measured quadratic Volterra kernel difference and corresponding frequency response from the smartphone at low and high volume signal levels.	47
Figure 3.10:	Measured harmonic distortion for the conference phone.	49
Figure 3.11:	Measured harmonic distortion for the smartphone.	50
Figure 4.1:	Block diagram of the generalized proportionate variable step size subband adaptive filter structure.	57
Figure 4.2:	Real part of the unobstructed and obstructed complex subband impulse response estimates from the conference phone configuration.	70
Figure 4.3:	Subband NLMS versus fullband NLMS tracking.	72
Figure 4.4:	ERLE for FB-NLMS, FB-IPNLMS, FB-GPNLMS, and SB-GPNLMS with WGN input in a synthetic changing environment and a step size of 0.25.	73
Figure 4.5:	ERLE for FB-GPNLMS, SB-NLMS, SB-IPNLMS, and SB-GPNLMS with WGN input in a synthetic changing environment and a step size of 0.25.	75
Figure 4.6:	ERLE for FB-NLMS, FB-IPNLMS, FB-GPNLMS, and SB-GPNLMS with WGN input in a synthetic changing environment and a step size of 1.0.	76

Figure 4.7:	ERLE for FB-GPNLMS, SB-NLMS, SB-IPNLMS, and SB-GPNLMS with WGN input in a synthetic changing environment and a step size of 1.0.	77
Figure 4.8:	ERLE for SB-GPAP with WGN input in a synthetic changing environment and an increasing P for $\mu_i = 1.0$ in all subbands. . .	78
Figure 4.9:	ERLE for SB-GPNLMS with WGN input in a synthetic changing environment and the GPNLMS algorithm applied in only specific subbands.	79
Figure 4.10:	ERLE for SB-GPNLMS with WGN input in a synthetic changing environment and the GPNLMS algorithm applied to only the initial adaptive filter taps.	81
Figure 4.11:	ERLE for SB-GPNLMS with WGN input in a synthetic changing environment and an increasing number of subbands for the same OSR.	82
Figure 4.12:	ERLE from 6 to 10 seconds for SB-GPNLMS with WGN input in a synthetic changing environment and an increasing number of subbands for the same OSR.	83
Figure 4.13:	ERLE for FB-NLMS, FB-IPNLMS, FB-GPNLMS, and SB-GPNLMS with speech input in a synthetic changing environment and a step size of 1.0.	86
Figure 4.14:	Spectrogram of reference speech excitation signal.	86
Figure 4.15:	ERLE for FB-GPNLMS, SB-NLMS, SB-IPNLMS, and SB-GPNLMS with speech input in a synthetic changing environment and a step size of 1.0.	88
Figure 4.16:	ERLE for SB-GPAP with speech input in a synthetic changing environment and an increasing P for $\mu_i = 1.0$ in all subbands. . .	89

Figure 4.17: ERLE for SB-GPNLMS with speech input in a synthetic changing environment and the GPNLMS algorithm applied in only specific subbands.	90
Figure 4.18: ERLE for SB-GPNLMS with speech input in a synthetic changing environment and the GPNLMS algorithm applied to only the initial adaptive filter taps.	91
Figure 4.19: ERLE for FB-GPNLMS, SB-NLMS, SB-IPNLMS, and SB-GPNLMS with measured speech input in a changing environment and a step size of 1.0.	92
Figure 4.20: ERLE for SB-GPNLMS with measured speech input in a changing environment and the GPNLMS algorithm applied to only the initial adaptive filter taps.	93
Figure 4.21: ERLE from 0 to 3 seconds for FB-GPAP, SB-AP, SB-IPAP, and SB-GPAP with measured speech input in a changing environment and a step size of 1.0.	95
Figure 4.22: ERLE from 21 to 23 seconds for FB-GPAP, SB-AP, SB-IPAP, and SB-GPAP with measured speech input in a changing environment and a step size of 1.0.	96
Figure 4.23: ERLE from 11 to 12.5 seconds for SB-GPAP with measured speech input in a changing environment and the GPAPA applied to only the initial adaptive filter taps.	97
Figure 4.24: ERLE for SB-GPNLMS with measured speech input in a changing environment and an increasing OSR.	98
Figure 4.25: ERLE from 0 to 2 seconds for SB-GPNLMS with measured speech input in a changing environment and an increasing number of subbands for the same OSR.	99

Figure 4.26: ERLE for from 12 to 13 seconds SB-GPNLMS with measured speech input in a changing environment and an increasing number of subbands for the same OSR.	100
Figure 4.27: Computational complexity for SB-NLMS, SB-IPNLMS, SB-GPNLMS, FB-IPNLMS, and FB-GPNLMS.	102
Figure 4.28: Computational complexity for SB-GPNLMS with the GPNLMS algorithm applied in only specific subbands.	104
Figure 4.29: Computational complexity for SB-GPAP and FB-GPAP with an increasing projection order.	105
Figure 4.30: Computational complexity for SB-GPNLMS and SB-NLMS with an increasing number of subbands for the same OSR.	106
Figure 4.31: Computational complexity for SB-GPNLMS and SB-NLMS with an increasing OSR.	107
Figure 5.1: Block diagram of the subband adaptive second order Volterra filter structure.	113
Figure 5.2: ERLE for FB-NLMS, FB-NLMSVF, and SB-NLMSVF with WGN input under low volume loudspeaker distortion conditions and a step size of 0.5 for both the linear and quadratic adaptive filters.	120
Figure 5.3: ERLE for FB-NLMS, SB-NLMS, FB-NLMSVF, and SB-NLMSVF with WGN input under moderate volume loudspeaker distortion conditions and a step size of 0.5 for both the linear and quadratic adaptive filters.	122
Figure 5.4: ERLE for FB-NLMS, SB-NLMS, FB-NLMSVF, and SB-NLMSVF with WGN input under moderate volume loudspeaker distortion conditions and a step size of 1.0 for the linear adaptive filters and 0.5 for the quadratic adaptive filters.	124

Figure 5.5:	ERLE for FB-NLMS, SB-NLMS, FB-NLMSVF, and SB-NLMSVF with WGN input under moderate volume loudspeaker distortion conditions and a step size of 1.0 for both the linear and quadratic adaptive filters.	125
Figure 5.6:	ERLE for FB-NLMS, SB-NLMS, FB-NLMSVF, and SB-NLMSVF with WGN input under moderate volume loudspeaker distortion conditions and a step size of 1.0 for both the linear and quadratic adaptive filters and a decreasing N_{2, sb_1}	126
Figure 5.7:	ERLE for FB-NLMS, SB-NLMS, FB-NLMSVF, and SB-NLMSVF with speech input under high volume loudspeaker distortion conditions and a step size of 0.5 for both the linear and quadratic adaptive filters.	128
Figure 5.8:	ERLE for SB-NLMS, FB-NLMSVF, and SB-NLMSVF with speech input under high volume loudspeaker distortion conditions and a step size of 0.5 for both the linear and quadratic adaptive filters and a decreasing N_{2, sb_1}	129
Figure 5.9:	ERLE from 0 to 4 seconds for SB-NLMS, FB-NLMSVF, and SB-NLMSVF with measured speech input from the conference phone under high volume loudspeaker distortion conditions and a step size of 0.5 for both the linear and quadratic adaptive filters. . . .	131
Figure 5.10:	ERLE from 12 to 15 seconds for SB-NLMS, FB-NLMSVF, and SB-NLMSVF with measured speech input from the conference phone under high volume loudspeaker distortion conditions and a step size of 0.5 for both the linear and quadratic adaptive filters. . . .	132
Figure 5.11:	ERLE for SB-NLMS, FB-NLMSVF, and SB-NLMSVF with measured speech input from the conference phone under high volume loudspeaker distortion conditions and a step size of 0.5 for both the linear and quadratic adaptive filters and a decreasing N_{2, sb_1} . . .	134

Figure 5.12: ERLE from 12 to 15 seconds for SB-NLMSVF with measured speech input from the conference phone under high volume loudspeaker distortion conditions and an increasing number of subbands for the same OSR.	135
Figure 5.13: ERLE from 0 to 4 seconds for SB-NLMS, FB-NLMSVF, and SB-NLMSVF with measured speech input from the smartphone under high volume loudspeaker distortion conditions and a step size of 0.5 for both the linear and quadratic adaptive filters.	137
Figure 5.14: ERLE from 12 to 15 seconds for SB-NLMS, FB-NLMSVF, and SB-NLMSVF with measured speech input from the smartphone under high volume loudspeaker distortion conditions and a step size of 0.5 for both the linear and quadratic adaptive filters.	138
Figure 5.15: ERLE from 12 to 15 seconds for SB-NLMS, FB-NLMSVF, and SB-NLMSVF with measured speech input from the smartphone under high volume loudspeaker distortion conditions and a step size of 0.5 for both the linear and quadratic adaptive filters and a decreasing $N_{2, sb_{2-3}}$	139
Figure 5.16: ERLE from 0 to 3 seconds for SB-NLMSVF with measured speech input from the smartphone under high volume loudspeaker distortion conditions and an increasing $\mu_{1,i}$	140
Figure 5.17: ERLE from 0 to 3 seconds for SB-NLMSVF with measured speech input from the smartphone under high volume loudspeaker distortion conditions and an increasing $\mu_{2,i}$	141
Figure 5.18: Computational complexity for SB-NLMS, FB-NLMSVF, and SB-NLMSVF.	142
Figure 5.19: Computational complexity for SB-NLMSVF and SB-NLMS for an increasing number of subbands.	144

Figure 6.1:	Block diagram of the SB-GGNLMSVF structure.	149
Figure 6.2:	ERLE for FB-NLMS, FB-NLMSVF, FB-GPNLMSVF, SB-NLMS, SB-NLMSVF, and SB-GPNLMSVF with WGN input under changing nonlinear echo path conditions and a step size of 0.5 for both the linear and quadratic adaptive filters.	158
Figure 6.3:	ERLE from 6 to 10 seconds for FB-NLMS, FB-NLMSVF, FB-GPNLMSVF, SB-NLMS, SB-NLMSVF, and SB-GPNLMSVF with WGN input under changing nonlinear echo path conditions and a step size of 0.5 for both the linear and quadratic adaptive filters.	160
Figure 6.4:	ERLE for SB-GPNLMSVF with WGN input under changing nonlinear echo path conditions and the GPNLMS algorithm applied in only specific subbands.	161
Figure 6.5:	ERLE from 6 to 10 seconds for SB-GPNLMSVF with WGN input under changing nonlinear echo path conditions and the GPNLMS algorithm applied in only specific subbands.	162
Figure 6.6:	ERLE for FB-NLMSVF, FB-GPNLMSVF, SB-NLMS, SB-GPNLMS, SB-NLMSVF, and SB-GPNLMSVF with WGN input under changing linear and nonlinear echo path conditions and a step size of 0.5 for both the linear and quadratic adaptive filters.	165
Figure 6.7:	ERLE for SB-GPNLMSVF with WGN input under changing linear and nonlinear echo path conditions and the GPNLMS algorithm applied in only specific subbands.	166
Figure 6.8:	ERLE from 6 to 10 seconds for SB-GPNLMSVF with WGN input under changing linear and nonlinear echo path conditions and the GPNLMS algorithm applied in only specific subbands.	167
Figure 6.9:	ERLE for SB-GPNLMSVF with WGN input under changing linear and nonlinear echo path conditions and a decreasing $N_{2, sb1-2}$	169

Figure 6.10: ERLE from 0 to 3 seconds for FB-GPNLMSVF, FB-IPNLMSVF, SB-NLMS, SB-GPNLMS, SB-NLMSVF, and SB-GPNLMSVF with measured speech input under changing linear and nonlinear echo path conditions and a step size of 0.5 for both the linear and quadratic adaptive filters.	171
Figure 6.11: ERLE from 10 to 13 seconds for FB-GPNLMSVF, FB-IPNLMSVF, SB-NLMS, SB-GPNLMS, SB-NLMSVF, and SB-GPNLMSVF with measured speech input under changing linear and nonlinear echo path conditions and a step size of 0.5 for both the linear and quadratic adaptive filters.	172
Figure 6.12: ERLE from 10 to 13 seconds for SB-GPNLMSVF with measured speech input under changing linear and nonlinear echo path conditions and the GPNLMS algorithm applied in only specific subbands.	173
Figure 6.13: ERLE from 10 to 13 seconds for SB-GPNLMSVF with measured speech input under changing linear and nonlinear echo path conditions and a decreasing $N_{2, sb_{1\&3}}$	175
Figure 6.14: ERLE from 10 to 13 seconds for SB-GPNLMSVF with measured speech input under changing linear and nonlinear echo path conditions and an increasing number of subbands for the same OSR.	176
Figure 6.15: ERLE from 0 to 3 seconds for SB-GPNLMSVF with measured speech input under changing linear and nonlinear echo path conditions and an increasing $\mu_{1,i}$	177
Figure 6.16: ERLE from 0 to 3 seconds for SB-GPNLMSVF with measured speech input under changing linear and nonlinear echo path conditions and an increasing $\mu_{2,i}$	178
Figure 6.17: Computational complexity for FB-GPNLMSVF, SB-NLMS, and SB-GPNLMSVF with the GPNLMS algorithm applied in only specific subbands.	180

Figure 6.18: Computational complexity for SB-GPNLMSVF and SB-NLMS for
an increasing number of subbands for the same OSR. 182

List of Tables

Table 4.1:	Approximate computational complexity per iteration of the GG-PAPA applied to subband adaptive filters for real input data in terms of real multiplications, additions, and divisions.	63
Table 4.2:	Approximate computational complexity per iteration of the GG-PAPA applied to subband adaptive filters for complex input data in terms of real multiplications, additions, and divisions.	64
Table 4.3:	Approximate computational complexity per input sample period for the SB-GGPAP AEC structure in terms of real multiplications, additions, and divisions.	64
Table 4.4:	Approximate computational complexity per iteration of the GG-PAPA applied to the initial N_{f_i} subband adaptive filter taps and the APA applied to the remaining taps for complex input data in terms of real multiplications, additions, and divisions.	66
Table 4.5:	Approximate computational complexity per input sample period for the SB-GGPAP AEC structure with the GGPAPA applied to the initial N_{f_i} subband adaptive filter taps and the APA applied to the remaining taps in terms of real multiplications, additions, and divisions.	67
Table 4.6:	Summary of results for AEC in nonstationary environments.	110

Table 5.1:	Approximate computational complexity per iteration of the NLMS algorithm applied to subband adaptive second order Volterra filters for real input data in terms of real multiplications, additions, and divisions.	116
Table 5.2:	Approximate computational complexity per iteration of the NLMS algorithm applied to subband adaptive second order Volterra filters for complex input data in terms of real multiplications, additions, and divisions.	117
Table 5.3:	Approximate computational complexity per input sample period for the SB-NLMSVF AEC structure in terms of real multiplications, additions, and divisions.	118
Table 5.4:	Summary of results for AEC in nonlinear environments.	146
Table 6.1:	Approximate computational complexity per iteration of the GG-PNLMS algorithm applied to subband adaptive second order Volterra filters for real input data in terms of real multiplications, additions, and divisions.	153
Table 6.2:	Approximate computational complexity per iteration of the GG-PNLMS algorithm applied to subband adaptive second order Volterra filters for complex input data in terms of real multiplications, additions, and divisions.	155
Table 6.3:	Approximate computational complexity per input sample period for the SB-GGPLMSVF AEC structure in terms of real multiplications, additions, and divisions.	155
Table 6.4:	Summary of results for AEC in nonstationary and nonlinear environments.	185

List of Symbols

A boldface uppercase letter generally refers to a matrix of numbers or variables, while a boldface lowercase letter generally refers to a vector of numbers or variables. A non-boldface uppercase or lowercase letter is a scalar quantity.

Arabic Symbols

- $\mathbf{A}(n)$ Step size matrix at time n .
- $\mathbf{A}_i(m)$ Step size matrix for subband i at time m .
- $\mathbf{A}_{p,i}(m)$ Step size matrix for adaptive Volterra kernel in the i^{th} subband at time m .
- $c_l(n)$ Gradient estimate for the l^{th} coefficient of adaptive filter vector at time n .
- $\bar{c}_l(n)$ Average gradient estimate for the l^{th} coefficient of adaptive filter vector at time n .
- $c_{i_l}(m)$ Gradient estimate for the l^{th} coefficient of adaptive filter vector for subband i at time m .
- $c_{p,i_l}(m)$ Gradient estimate for the l^{th} coefficient of adaptive Volterra kernel vector for subband i at time m .
- $\bar{c}_{i_l}(m)$ Average gradient estimate for the l^{th} coefficient of adaptive filter vector for subband i at time m .
- $\bar{c}_{p,i_l}(m)$ Average gradient estimate for the l^{th} coefficient of adaptive Volterra kernel vector for subband i at time m .
- $C_{\sqrt{\cdot},a}$ Number of real addition operations required for computing the square root of a number.
- $C_{\sqrt{\cdot},d}$ Number of real division operations required for computing the square root

	of a number.
$C_{\sqrt{\cdot},m}$	Number of real multiplication operations required for computing the square root of a number.
D	Number of diagonals from the input signal matrix of a truncated second order Volterra series expansion.
$\mathbf{d}(n)$	Microphone signal vector at time n .
$\mathbf{d}_i(m)$	Microphone signal vector for subband i at time m .
$d(n)$	Microphone signal at time n .
$d_i(m)$	Microphone signal for subband i at time m .
$\mathbf{e}(n)$	Error signal vector at time n .
$\mathbf{e}_i(m)$	Error signal vector for subband i at time m .
$e(n)$	Error signal at time n .
$e_i(m)$	Error signal for subband i at time m .
$\mathbf{F}(z)$	Polyphase analysis filter bank matrix in the z domain.
$\mathbf{f}_k(n)$	Analysis filter for subband k at time n .
$\mathbf{G}(z)$	Polyphase synthesis filter bank matrix in the z domain.
$\mathbf{g}_k(n)$	Synthesis filter for subband k at time n .
$g_l(n)$	Adaptation step size for the l^{th} coefficient of adaptive filter vector at time n .
$g_{i_l}(m)$	Adaptation step size for the l^{th} coefficient of adaptive filter vector for subband i at time m .
$g_{p,i_l}(m)$	Adaptation step size for the l^{th} coefficient of adaptive Volterra kernel vector for subband i at time m .
$\mathbf{h}(n)$	Echo path impulse response vector at time n .
$\hat{\mathbf{h}}(n)$	Adaptive filter coefficient vector at time n .
$\mathbf{h}_p(n)$	Volterra kernel coefficient vector at time n .
$\hat{\mathbf{h}}_p(n)$	Adaptive Volterra kernel coefficient vector at time n .
$h_i(n)$	i^{th} coefficient of echo path impulse response vector at time n .

$\bar{h}_l(n)$	Average l^{th} coefficient of adaptive filter vector at time n .
$\hat{h}_i(n)$	i^{th} coefficient of adaptive filter vector at time n .
$\hat{h}_{p_i}(n)$	i^{th} coefficient of adaptive Volterra kernel vector at time n .
H_i	Power in the i^{th} harmonic.
$J(n)$	Adaptive filter cost function at time n .
K	Number of subbands in analysis/synthesis filter bank.
L	Echo path system length.
L_p	Volterra kernel length.
L_{pr}	Prototype filter length.
M	Filter bank downsampling factor.
N	Adaptive filter length.
N_{sb_i}	Subband adaptive filter length.
N_p	Adaptive Volterra kernel length.
N_{p,sb_i}	Subband adaptive Volterra kernel length.
$\mathbf{p}(n)$	Prototype filter at time n .
P	Projection order.
P_V	Order of a Volterra series.
p	Index with respect to the order of a Volterra series.
S_{max}	Maximum value of the power spectral density of a signal.
$T_{C,a}$	Number of real addition operations required for solving a system of equations with complex input data.
$T_{C,d}$	Number of real division operations required for solving a system of equations with complex input data.
$T_{C,m}$	Number of real multiplication operations required for solving a system of equations with complex input data.
$T_{R,a}$	Number of real addition operations required for solving a system of equations with real input data.

- $T_{R,d}$ Number of real division operations required for solving a system of equations with real input data.
- $T_{R,m}$ Number of real multiplication operations for solving a system of equations with real input data.
- $\mathbf{w}_i(m)$ Echo path impulse response vector for subband i at time m .
- $\hat{\mathbf{w}}_i(m)$ Adaptive filter coefficient vector for subband i at time m .
- $\mathbf{w}_{p,i}(m)$ Volterra kernel coefficient vector for subband i at time m .
- $\hat{\mathbf{w}}_{p,i}(m)$ Adaptive Volterra kernel coefficient vector for subband i at time m .
- $\bar{w}_{i_l}(m)$ Average l^{th} coefficient of adaptive filter vector for subband i at time m .
- $\bar{w}_{p,i_l}(m)$ Average l^{th} coefficient of adaptive Volterra kernel vector for subband i at time m .
- $\hat{w}_{i_j}(m)$ j^{th} coefficient of adaptive filter vector for subband i at time m .
- $\hat{w}_{p,i_j}(m)$ j^{th} coefficient of adaptive Volterra kernel vector for subband i at time m .
- $\mathbf{X}(n)$ Input signal matrix at time n .
- $\mathbf{X}_i(m)$ Input signal matrix for subband i at time m .
- $\mathbf{X}_{p,i}(m)$ Input signal matrix for adaptive Volterra kernel in the i^{th} subband at time m .
- $\mathbf{x}(n)$ Input signal vector at time n .
- $\mathbf{x}_i(m)$ Input signal vector for subband i at time m .
- $\mathbf{x}_p(n)$ Input signal vector of a Volterra kernel at time n .
- $\mathbf{x}_{2,j}(n)$ Input signal vector for the j^{th} diagonal of the input signal matrix of a truncated second order Volterra series expansion at time n .
- $\mathbf{x}_{p,i}(m)$ Input signal vector for adaptive Volterra kernel in the i^{th} subband at time m .
- $x(n)$ Input signal at time n .
- $x_p(n)$ Input signal of a Volterra kernel at time n .
- $x_{p,i}(m)$ Input signal for adaptive Volterra kernel in the i^{th} subband at time m .
- $x_i(m)$ Input signal for subband i at time m .

$y(n)$ Echo signal at time n .
 $\hat{y}(n)$ Echo signal estimate at time n .

Greek Symbols

β Weighting factor for generalized gradient proportionate variable step size adaptive filter.

β_i Weighting factor for generalized gradient proportionate variable step size adaptive filter in subband i .

$\beta_{p,i}$ Weighting factor for generalized gradient proportionate variable step size adaptive Volterra kernel in subband i .

δ Regularization parameter for adaptive filter.

δ_i Regularization parameter for adaptive filter in subband i .

δ_p Regularization parameter for adaptive Volterra kernel.

$\delta_{p,i}$ Regularization parameter for adaptive Volterra kernel in subband i .

ϵ, η Forgetting factors for generalized gradient proportionate variable step size adaptive filter.

ϵ_i, η_i Forgetting factors for generalized gradient proportionate variable step size adaptive filter in subband i .

$\epsilon_{p,i}, \eta_{p,i}$ Forgetting factors for generalized gradient proportionate variable step size adaptive Volterra kernel in subband i .

$\eta(n)$ Local noise signal at time n .

γ Correction factor for generalized gradient proportionate variable step size adaptive filter.

γ_i Correction factor for generalized gradient proportionate variable step size adaptive filter in subband i .

$\gamma_{p,i}$ Correction factor for generalized gradient proportionate variable step size adaptive Volterra kernel in subband i .

μ Adaptation step size parameter.

μ_i Adaptation step size parameter for subband i .

μ_p	Adaptive Volterra kernel step size parameter.
$\mu_{p,i}$	Adaptive Volterra kernel step size parameter for subband i .
$\nu(n)$	Near-end talker signal at time n .
$\sigma_d^2(n)$	Steady-state variance of signal $d(n)$ at time n .
$\sigma_e^2(n)$	Steady-state variance of signal $e(n)$ at time n .

List of Abbreviations

ADC	Analog-to-Digital Converter
AEC	Acoustic Echo Cancellation
AP	Affine Projection
APA	Affine Projection Algorithm
CPU	Central Processing Unit
DAC	Digital-to-Analog Converter
dB	Decibel
DCT	Discrete Cosine Transform
DFT	Discrete Fourier Transform
DSP	Digital Signal Processor
EC	Echo Canceller
ERLE	Echo Return Loss Enhancement
FB-GGPAPA	Fullband Generalized Gradient Proportionate Affine Projection Algorithm
FB-GPAP	Fullband Gradient Proportionate Affine Projection
FB-GPNLMS	Fullband Gradient Proportionate Normalized Least-Mean-Square
FB-GPNLMSVF	Fullband Gradient Proportionate Normalized Least-Mean-Square Volterra Filter
FB-IPNLMS	Fullband Improved Proportionate Normalized Least-Mean-Square
FB-IPNLMSVF	Fullband Improved Proportionate Normalized Least-Mean-Square Volterra Filter

FB-NLMS	Fullband Normalized Least-Mean-Square
FB-NLMSVF	Fullband Normalized Least-Mean-Square Volterra Filter
FIR	Finite Impulse Response
GDFT	Generalized Discrete Fourier Transform
GGPAPA	Generalized Gradient Proportionate Affine Projection Algorithm
GGPNLMS	Generalized Gradient Proportionate Normalized Least-Mean-Square
GPAPA	Gradient Proportionate Affine Projection Algorithm
GPNLMS	Gradient Proportionate Normalized Least-Mean-Square
IIR	Infinite Impulse Response
IPAPA	Improved Proportionate Affine Projection Algorithm
IPNLMS	Improved Proportionate Normalized Least-Mean-Square
LEMS	Loudspeaker-Enclosure-Microphone-System
LMS	Least-Mean-Square
mps	Multiplications Per Sample
MSE	Mean Square Error
NLMS	Normalized Least-Mean-Square
OSR	Oversampling Ratio
PAPA	Proportionate Affine Projection Algorithm
PNLMS	Proportionate Normalized Least-Mean-Square
RLS	Recursive Least-Squares
RMS	Root Mean Square
SB-AP	Subband Affine Projection
SB-APA	Subband Affine Projection Algorithm
SB-GGPAP	Subband Generalized Gradient Proportionate Affine Projection
SB-GGPAPA	Subband Generalized Gradient Proportionate Affine Projection Algorithm
SB-GGPAPVF	Subband Generalized Gradient Proportionate Affine Projection Volterra Filter

SB-GGNLMS	Subband Generalized Gradient Proportionate Normalized Least-Mean-Square
SB-GGNLMSVF	Subband Generalized Gradient Proportionate Normalized Least-Mean-Square Volterra Filter
SB-GPAP	Subband Gradient Proportionate Affine Projection
SB-GPAPA	Subband Gradient Proportionate Affine Projection Algorithm
SB-GPNLMS	Subband Gradient Proportionate Normalized Least-Mean-Square
SB-GPNLMSVF	Subband Gradient Proportionate Normalized Least-Mean-Square Volterra Filter
SB-IPAP	Subband Improved Proportionate Affine Projection
SB-IPAPA	Subband Improved Proportionate Affine Projection Algorithm
SB-IPNLMS	Subband Improved Proportionate Normalized Least-Mean-Square
SB-NLMS	Subband Normalized Least-Mean-Square
SB-NLMSVF	Subband Normalized Least-Mean-Square Volterra Filter
SB-PAPA	Subband Proportionate Affine Projection Algorithm
SB-PNLMS	Subband Proportionate Normalized Least-Mean-Square
SBVF	Subband Volterra Filter
SNR	Signal-to-Noise Ratio
THD	Total Harmonic Distortion
VAD	Voice Activity Detector
WGN	White Gaussian Noise

Chapter 1

Introduction

1.1 Problem Statement

The hands-free operation of communication devices has become essential as it allows parties to converse unencumbered and undistracted by the device itself. All hands-free communication systems are subject to the problem of acoustic and network echo which must be removed to ensure that a high quality full-duplex conversation takes place between parties. Network echo arises due to impedance mismatches between electrical components in the transmission network, while acoustic echo is caused by direct loudspeaker to microphone coupling and reflections within the enclosed environment. The cancellation of acoustic and network echo is typically performed by a digital echo canceller (EC) implemented as an adaptive filter, of which there is a rich field of literature dedicated to its design, implementation, and analysis [1]–[4].

In practice the performance of an EC is limited by many physical and environmental factors. Within the acoustic environment the presence of background noise, nonlinearities, nonstationarities, and doubletalk conditions, along with undermodeling of the loudspeaker-enclosure-microphone-system (LEMS) also referred to as the echo path, can all negatively impact the performance of an EC [5]–[8]. Furthermore,

residual echo can also impair the perceived effectiveness of an EC [9]. Thus, double-talk detector and post-filtering structures are typically required in addition to the EC itself in practical echo cancellation systems [10]. Ideally an EC will perform well under all conditions while requiring a low computational complexity in terms of the total number of elementary mathematical operations required to implement it. However, these requirements generally cannot be satisfied simultaneously resulting in compromises made in the implementation of the underlying adaptive filter structure and algorithm which govern the performance, complexity, and robustness of an EC.

The inherent nonstationarities within the acoustic environment, where the term nonstationarities refers to changes that occur such as the movement of persons and objects, are one of the limiting factors in the achievable echo attenuation of an EC. When a change occurs in the acoustic environment the EC will adjust its model of the LEMS to reflect the change. To avoid a degradation in the quality of the hands-free communication the EC must track these changes quickly. However, the tracking of echo path fluctuations is hindered by the generally long adaptive filter required to model the LEMS and the potentially slow convergence of the underlying adaptive algorithm, such as the normalized least-mean-square (NLMS) algorithm [6]. Thus, proportionate step size algorithms have been proposed that improve tracking performance compared to conventional NLMS by exploiting the time domain nature of a changing echo path [11]–[14]. The drawback to these proportionate algorithms is their increased computational complexity, which is especially undesirable for ECs implemented as part of an embedded system with stringent complexity budgets and energy requirements. A common method to reduce the complexity of an EC is to implement it with a subband adaptive filter structure [15]. A subband structure splits and decimates the input EC signals into separate frequency bands resulting in a shorter-length adaptive filter in each subband, which leads to an overall lower complexity compared to its fullband counterpart. Subband EC structures with adaptive filters based on proportionate step size algorithms would allow the frequency domain nature of echo

path changes to be exploited simultaneously with fast tracking of these changes, while achieving low computational complexity. However, this type of EC structure has not been thoroughly investigated in the literature.

Nonlinear distortion exhibited by small electrodynamic loudspeakers when operated at high volume is another factor that can limit the performance of an EC. The nonlinear echoes that arise due to loudspeaker distortion cannot be removed by a typical linear EC and can therefore degrade the hands-free communication between parties [16]. A common method for modeling loudspeaker distortion is with Volterra series expansions [17], [18]. Consequently, many nonlinear EC structures based on adaptive Volterra filters have been proposed to compensate for the loudspeaker distortion and improve echo cancellation performance [19]–[22]. However, these structures suffer from very high computational requirements which prohibit their practical application. Thus, there have been many reduced complexity nonlinear EC structures proposed based on adaptive Volterra filters [16], [23]–[26]. These reduced complexity nonlinear EC structures may still have large computational costs compared to a typical linear EC, especially when Volterra filters with long memory lengths are required to model the loudspeaker distortion. As a result, there is still a need for computationally efficient nonlinear ECs in hands-free systems that require them but have limited computational resources. One approach to mitigating this complexity issue is with a subband adaptive Volterra filter structure that exploits the frequency domain characteristics of loudspeaker distortion. However, limited literature exists for such nonlinear EC structures. Furthermore, nonlinear EC structures that allow for fast tracking of fluctuations in the nonlinear portion of the echo path, resulting from varying levels of loudspeaker distortion, are also lacking in the literature. A subband EC structure with proportionate step size adaptive Volterra filters is one potential method for fast tracking of both linear and nonlinear echo path components with low complexity.

1.2 Thesis Objectives and Motivation

To ensure a high quality hands-free conversation the EC implemented in a hands-free device must be able to operate effectively under adverse conditions. Moreover, the computational requirements of the EC need to be kept within a certain budget, despite the increasing computational capability of modern processing hardware, as it is typically only one of many systems that require processing on a hands-free device. For battery operated hands-free devices, such as cellular telephones, it is also necessary that the EC does not cause a large computational demand on the device's processor in order to minimize power consumption. Since these EC requirements are not generally satisfied simultaneously by a typically EC, there is a need for ECs capable of operating effectively under adverse conditions with low complexity. This thesis addresses both of these EC requirements through the development and analysis of computationally efficient signal processing structures and adaptive algorithms capable of operating effectively under adverse conditions. Specifically, this thesis proposes the following novel adaptive structures for application to the acoustic echo cancellation problem:

- An adaptive structure that can quickly model and track linear echo path components in nonstationary acoustic environments.
- An adaptive structure that can quickly model linear echo path components as well as nonlinear components resulting from electrodynamic loudspeaker distortion.
- An adaptive structure capable of quickly modeling and tracking both linear and nonlinear echo path components in nonstationary acoustic environments.

Furthermore, the following research questions are addressed by this thesis:

1. Based on experimentally measured data from LEMSs under changing conditions, what is the corresponding impact on the linear and nonlinear portions of the resulting echo path model in both the time and frequency domains?

2. Given the time domain nature of echo path changes, can the computational complexity, in terms of the total number of required elementary mathematical operations per input sample period, of ECs based on proportionate step size adaptive algorithms be reduced while maintaining fast tracking performance?
3. Is it possible to incorporate proportionate step size adaptive algorithms into a subband EC structure to reduce complexity and improve tracking performance, compared to an equivalent fullband structure? If so, can the complexity of this structure be further reduced by exploiting the frequency domain nature of echo path changes, without sacrificing tracking performance?
4. Can a nonlinear EC based on adaptive Volterra filters be implemented with a subband structure to decrease complexity and enhance echo cancellation performance, compared to a fullband Volterra EC? If so, can this structure realize further computational savings while maintaining performance, by considering the time and frequency domain characteristics of loudspeaker distortion?
5. Is it possible to incorporate proportionate step size adaptive algorithms into a subband Volterra filter EC structure, and achieve improved tracking performance with reduced complexity compared to an equivalent fullband structure?

1.3 Thesis Contributions

The contributions of this thesis are the development of echo cancellation structures, along with their underlying adaptation algorithms, that allow for improved performance in nonstationary and nonlinear acoustic environments compared to typical echo cancellation structures, while maintaining low computational complexity. Specifically, the main contributions of this thesis to the field of acoustic echo cancellation are as follows:

Contribution 1. Analysis of linear and nonlinear echo path components

– Based on experimental measurements from LEMSs with commercial hands-free sets the resulting linear and nonlinear echo path components are identified. It is shown that in response to a small obstruction introduced in close vicinity to the hands-free device, the majority of the corresponding changes in the linear echo path components are restricted to certain time and frequency regions. Furthermore, it is found that in response to an increase in the input signal level, the corresponding change in the nonlinear echo path components are also restricted to certain time and frequency regions. These contributions are described in Chapter 3 and offer insight into achieving complexity reduction in the EC structures presented in later chapters. A portion of these results are published in [27].

Contribution 2. Harmonic distortion characterization of commercial telephone sets

– An investigation is conducted into the harmonic distortion of a commercial telephone set along with a cellular telephone, based on experimental measurements for several fundamental test tones. The analysis reveals that the majority of harmonic distortion, under moderate to high input signal levels, is contained in the second harmonic and is limited to specific frequency regions. The results of this work also provide valuable insight for achieving computationally efficient implementations of the EC structures proposed in this thesis. These contributions are outlined in Chapter 3, and the results of this work are published in [28].

Contribution 3. Extension of a fullband generalized gradient proportionate step size adaptive filter to a subband echo cancellation structure

– In this work a subband EC structure with adaptive filters based on a generalized gradient proportionate variable step size adaptive algorithm is proposed. It is shown that this proposed structure is able to outperform its fullband counterpart during changing conditions in the acoustic environment, while requiring lower computational complexity. Moreover, the proposed structure is able to perform similarly to its fullband

counterpart with further reduced complexity, by exploiting the time and frequency domain nature of changing linear echo path components. This contribution is described in Chapter 4. The results of this work were submitted to *IEEE Transactions on Instrumentation and Measurement*.

Contribution 4. Design of a subband adaptive Volterra filter echo cancellation structure – This work proposes a subband EC structure based on adaptive second order Volterra filters for cancellation of nonlinear echoes that arise due to loudspeaker distortion. The proposed structure is shown to achieve equal or improved echo cancellation performance compared to its fullband counterpart at a much reduced computational cost. Furthermore, the proposed structure is able to take advantage of the time and frequency domain characteristics of loudspeaker distortion, resulting in increased complexity savings while maintaining performance. This contribution is described in Chapter 5, with the results of this work published in [27] and [29].

Contribution 5. Development of a generalized gradient proportionate step size subband adaptive Volterra filter echo cancellation structure – A subband EC structure with second order Volterra filters based on a generalized gradient proportionate variable step size adaptive algorithm is proposed in this work. The ability of the proposed structure to provide improved tracking performance, at a smaller computational expense, is demonstrated compared to its fullband counterpart during variations in both the linear and nonlinear echo path components. In addition, it is shown that the proposed structure is able to perform as well as its equivalent fullband structure at an even lower computational cost, by considering the time and frequency domain nature of linear and nonlinear echo path components and their changes. This contribution is described in Chapter 6. The results of this work were submitted to *IEEE Transactions on Audio, Speech, and Language Processing*.

1.4 Thesis Organization and Scope

There are four key chapters in this thesis. Chapter 3 analyzes experimentally measured linear and nonlinear echo path components obtained from several hands-free systems under various harsh conditions in the acoustic environment. A computationally efficient fast tracking subband EC structure based on proportionate step size adaptive filters is presented in Chapter 4. Chapter 5 outlines a low complexity nonlinear EC structure based on second order adaptive Volterra filters. A subband EC structure based on proportionate step size adaptive Volterra filters is introduced in Chapter 6, which is a combination of the structures presented in Chapters 4 and 5. Figure 1.1 depicts how these key chapters are linked together. Chapters 2 and 7 are supporting chapters that provide the necessary background material on acoustic echo cancellation including the related adaptive filter theory and structures, and summarize the main conclusions from this work respectively. A description of the contents of these chapters is as follows:

Chapter 2 outlines the echo cancellation problem, and reviews common EC structures and adaptation algorithms. In particular, the challenges of echo cancellation in nonstationary and nonlinear environments are addressed, along with the associated state-of-the-art EC structures.

Chapter 3 describes the experimental setup and equipment used for obtaining data from two hands-free systems. Based on the resulting measured data, linear and nonlinear echo path components are identified, and a time and frequency domain analysis of their characteristics during changing conditions in the hands-free environment is presented. Also, a harmonic distortion analysis is presented for the commercial phone sets used in the hands-free systems, based on measured data for several fundamental test tones with increasing signal levels.

Chapter 4 addresses the problem of echo cancellation in nonstationary environments by investigating the incorporation of proportionate step size adaptive algorithms into a subband EC structure. A mathematical description and detailed computational complexity analysis of the resulting structure are provided. The structure's improved echo cancellation performance at a reduced computational cost compared to its fullband equivalent is verified through computer simulations.

Chapter 5 investigates the application of adaptive Volterra filters into a subband EC structure to address the problem of nonlinear loudspeaker distortion in echo cancellation systems. The ensuing nonlinear subband EC structure is presented, including a mathematical description and complexity analysis. Simulation results are also presented verifying the structure's improved echo cancellation performance and reduced complexity compared to its fullband counterpart.

Chapter 6 addresses the problem of echo cancellation in nonstationary and nonlinear environments simultaneously. This is accomplished by investigating proportionate step size adaptive algorithms applied to Volterra filters in a subband EC structure. A mathematical description and computational cost analysis are given for the resulting structure. As well, simulation results demonstrating the improved performance and decreased complexity requirements of the structure compared to its fullband alternative are presented.

Chapter 7 outlines the main contributions originating from this thesis, and suggests potential areas of future research to build upon these contributions.

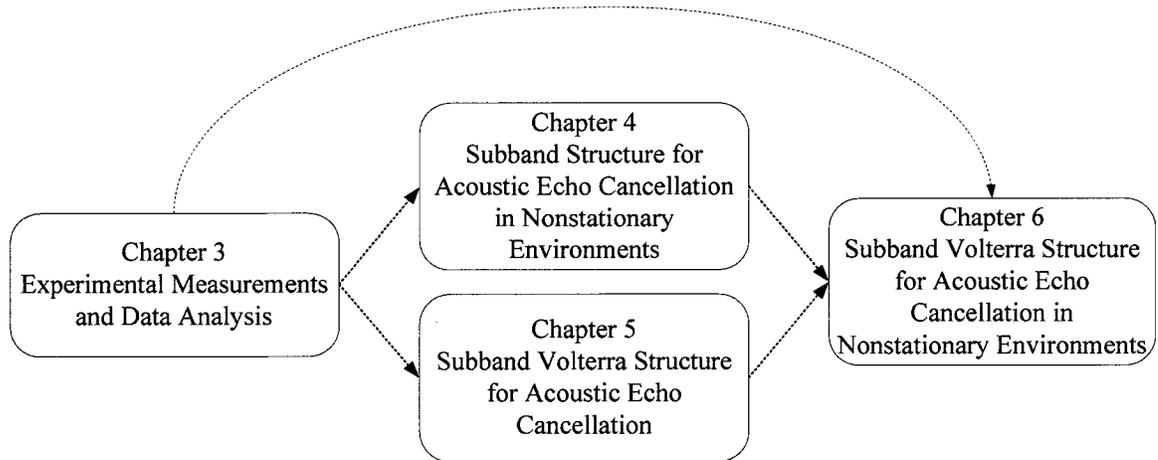


Figure 1.1: Overview of how key thesis chapters are linked together.

Chapter 2

Review of Acoustic Echo Cancellation

In this chapter a literature review of acoustic echo cancellation (AEC) structures along with some typically used adaptive algorithms is presented. As well, performance limitations in AEC are discussed along with methods to mitigate some of these limitations. The chapter sections are organized as follows: Section 2.1 outlines the problem of acoustic echo in hands-free communication systems along with the standard linear AEC model used to remove it. Also, detailed descriptions of some of the common adaptive filtering algorithms for AEC are given, along with a discussion of subband adaptive filtering structures. In Section 2.2 the performance limitations encountered in AEC are described in detail. Furthermore, AEC algorithms for dealing with the specific limitations of loudspeaker nonlinearities and acoustic environment nonstationarities are reviewed. In this thesis the term nonstationarities with respect to acoustic environments, refers to the changes that occur within these environments.

2.1 The Acoustic Echo Cancellation Problem

Acoustic echo is inherent in all hands-free communication systems and can corrupt the conversation between parties if it is not sufficiently suppressed. Hands-free systems are typically designed to be full-duplex, allowing for simultaneous communication between the near-end and far-end talkers. A block diagram showing one end of a typical hands-free communication system in the presence of acoustic echo is shown in Figure 2.1. Here the digitized far-end signal, $x(n)$, of the hands-free system is converted to a time domain signal, $x(t)$, by a digital-to-analog converter (DAC) and played through the loudspeaker into the enclosed environment. An echo signal, $y(t)$, is formed from $x(t)$ by the direct acoustic path between the loudspeaker and microphone along with reflections from enclosure boundaries and objects [1]. The signal captured by the microphone, $d(t)$, contains $y(t)$ as well as the near-end talker signal $\nu(t)$ and background noise signal $\eta(t)$, which is then digitized by an analog-to-digital converter (ADC) to produce $d(n)$. If left unabated the presence of acoustic echo will result in the annoyance of hearing a delayed copy of one's own voice during single-talk periods. In addition, during double-talk periods the presence of acoustic echo will contaminate the quality of the near-end speech signal. Network echo is another form of echo that can occur during transmission over the communication network. Network echoes occur due to impedance mismatches encountered in the transmission network, such as those found in network hybrids [2].

2.1.1 Acoustic Echo Cancellation Structure

The conventional approach to removing echo from a hands-free system is through the implementation of a digital EC [2]–[4]. Typically the LEMS or echo path (see Figure 2.1) is modeled as a linear system [2] where the acoustic EC can be implemented as an adaptive finite impulse response (FIR) filter. A block diagram of the linear system model for AEC is shown in Figure 2.2. The far-end signal $x(n)$ is applied to the LEMS

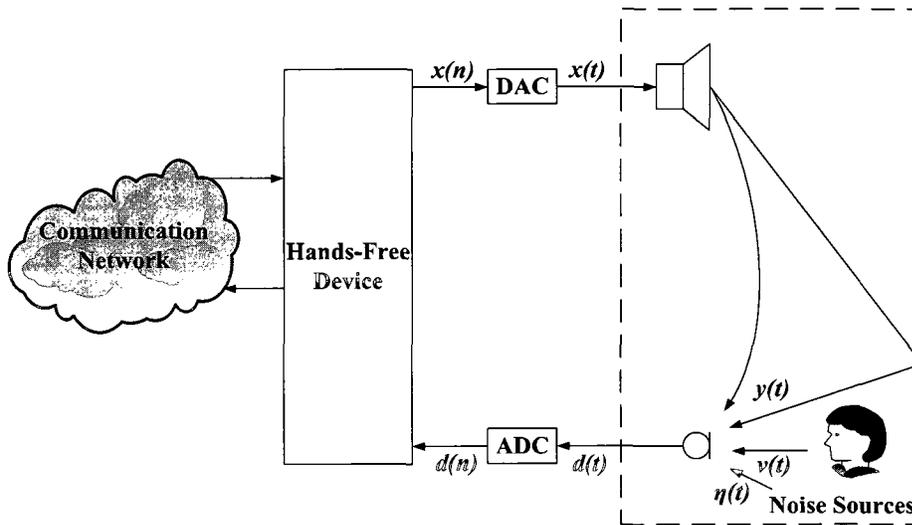


Figure 2.1: Acoustic echo in a hands-free communication system.

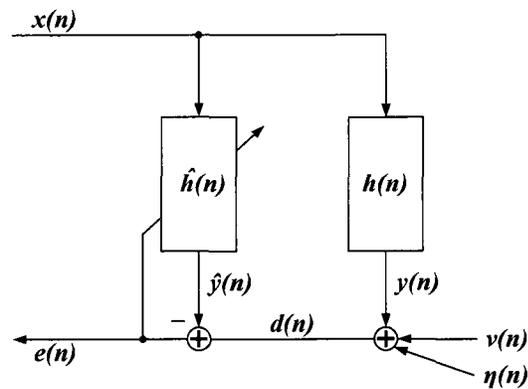


Figure 2.2: Linear system model for acoustic echo cancellation.

resulting in an echo signal $y(n)$ that is summed with the near-end talker signal $\nu(n)$, and the local noise signal $\eta(n)$ to produce the overall microphone signal $d(n)$, where the components of $d(n)$ are assumed to be uncorrelated with each other. The local noise signal represents the combined effects of background and measurement noise. The adaptive filter produces an estimate of the echo signal, $\hat{y}(n)$, which is removed from $d(n)$ resulting in an output error signal, $e(n)$, that ideally contains only $\nu(n)$ and $\eta(n)$:

$$d(n) = y(n) + \nu(n) + \eta(n) \quad (2.1)$$

$$e(n) = d(n) - \hat{y}(n) = y(n) - \hat{y}(n) + \nu(n) + \eta(n). \quad (2.2)$$

If the adaptive filter perfectly models the LEMS (i.e. $\hat{y}(n) = y(n)$) then $e(n)$ will contain only $\nu(n)$ and $\eta(n)$. In practice, $\hat{y}(n)$ is not a perfect replica of $y(n)$ and thus $e(n)$ will also contain a residual echo signal ($y(n) - \hat{y}(n)$). Generally an AEC system will contain a post-filtering stage to remove residual echo from $e(n)$ [30]. Also, practical AEC includes a voice activity detector (VAD) to distinguish between periods of far-end speech and near-end silence [31], since the EC will only operate effectively in the presence of far-end speech. A doubletalk detector is also typically found in practical AEC to determine if near-end speech is occurring simultaneously with far-end speech, and to stop adaptation of the EC during these periods [7], [32].

The echo path is an inherently time varying system due to the frequent changes occurring within it, such as movement of talkers and objects in the acoustic environment. Hence, an adaptive EC is not only required to determine an estimate of the unknown echo path impulse response or transfer function under stationary conditions, but also must track echo path variations that occur during changing conditions in the acoustic environment. The echo path impulse response is a finite length linear model of the unknown system consisting of L samples with its coefficients represented by a $L \times 1$ vector $\mathbf{h}(n)$. The corresponding echo signal $y(n)$ is determined from the convolution of $\mathbf{x}(n)$ with $\mathbf{h}(n)$ as follows:

$$y(n) = \mathbf{h}^T(n)\mathbf{x}(n) \quad (2.3)$$

$$\mathbf{h}(n) = [h_0(n), h_1(n), \dots, h_{L-1}(n)]^T \quad (2.4)$$

$$\mathbf{x}(n) = [x(n), x(n-1), \dots, x(n-L+1)]^T. \quad (2.5)$$

The echo path estimate is a FIR linear filter model of the true echo path with a coefficient vector $\hat{\mathbf{h}}(n)$ of length $N \leq L$ samples. The echo signal estimate $\hat{y}(n)$ at

the output of the adaptive filter is computed as follows:

$$\hat{y}(n) = \hat{\mathbf{h}}^H(n)\mathbf{x}(n) \quad (2.6)$$

$$\hat{\mathbf{h}}(n) = [\hat{h}_0(n), \hat{h}_1(n), \dots, \hat{h}_{N-1}(n)]^T \quad (2.7)$$

$$\mathbf{x}(n) = [x(n), x(n-1), \dots, x(n-N+1)]^T \quad (2.8)$$

where $[\cdot]^H$ is the conjugate transpose operator. The coefficients of $\hat{\mathbf{h}}(n)$ are determined by an adaptive algorithm that seeks to minimize some function of $e(n)$ with respect to $\hat{\mathbf{h}}(n)$. Adaptive filter theory and its applications are well established in the literature [33]. A discussion of the pertinent adaptive algorithms and filtering structures to this thesis are presented in the following sections.

2.1.1.1 Adaptive Algorithms for Acoustic Echo Cancellation

Criteria for the design and evaluation of adaptive algorithms for AEC include convergence speed, computational complexity, algorithmic delay, robustness, and tracking behaviour [34]. One of the most popular methods for filter adaptation in AEC are stochastic gradient-based algorithms such as the least-mean-square (LMS) and NLMS algorithms. These algorithms are popular due to their simplicity of implementation, robust performance, and low complexity [33]. Stochastic gradient-based algorithms minimize a cost function, $J(n)$, based on the expected value of the squared error signal of the EC:

$$J(n) = E[e^2(n)]. \quad (2.9)$$

The LMS algorithm arises when the gradient vector, $\nabla J(n)$, is based on instantaneous estimates of the input signal, $x(n)$, statistics [33]. The resulting recursive filter tap adaptation equation along with the stability bounds for its step size, μ , that minimizes

$J(n)$ in the mean-square sense is as follows:

$$\hat{\mathbf{h}}(n+1) = \hat{\mathbf{h}}(n) + \mu \mathbf{x}(n) e^*(n) \quad (2.10)$$

$$0 < \mu < \frac{2}{NS_{max}} \quad (2.11)$$

where S_{max} is the maximum value of the power spectral density of $x(n)$ [33] and $[\cdot]^*$ denotes the complex conjugate operation.

When the input signal vector $\mathbf{x}(n)$ is large in value the LMS algorithm can be impaired by gradient noise amplification since the filter tap update is directly proportional to $\mathbf{x}(n)$ [33]. To overcome this the filter tap weight adjustment can be normalized by the squared Euclidean norm of $\mathbf{x}(n)$ which gives rise to the NLMS algorithm [33]:

$$\hat{\mathbf{h}}(n+1) = \hat{\mathbf{h}}(n) + \frac{\mu}{\mathbf{x}^H(n)\mathbf{x}(n) + \delta} \mathbf{x}(n) e^*(n) \quad (2.12)$$

$$0 < \mu < 2 \quad (2.13)$$

where δ is a small positive constant that helps to avoid numerical difficulties when $\mathbf{x}(n)$ has small values.

One of the main advantages of the LMS and NLMS algorithms is their low complexity of approximately $2N$ multiplications per sample (mps). The biggest drawback of the LMS and NLMS algorithms is their slow rate of convergence when the correlation matrix of $x(n)$ has a large eigenvalue spread [33]. This is especially apparent for correlated input signals such as speech [35]. To increase the convergence speed a decorrelation filter can be used to whiten or reduce the eigenvalue spread of the input signal correlation matrix [36]. Other algorithms for AEC that increase convergence rate compared to LMS and NLMS are based on the recursive least-squares (RLS) adaptive filter [37]–[41]. The RLS filter whitens $x(n)$ through the inversion of the input signal correlation matrix in its filter tap update equation [33]. However, the increased convergence speed of these RLS methods come at the cost of a significant

increase in computational complexity compared to LMS and NLMS, and as a result they may be prohibitive for some AEC applications.

The affine projection algorithm (APA) offers a compromise between speed of convergence and complexity compared to the RLS and NLMS algorithms [33]. Here the filter tap update equation is based on the projection of the current and $P - 1$ previous input signal vectors onto the *a posteriori* error signal vector in an affine subspace as follows [42], [43]:

$$\mathbf{X}(n) = [\mathbf{x}(n), \mathbf{x}(n-1), \dots, \mathbf{x}(n-P+1)] \quad (2.14)$$

$$\mathbf{d}(n) = [d(n), d(n-1), \dots, d(n-P+1)] \quad (2.15)$$

$$\mathbf{e}(n) = \mathbf{d}(n) - \hat{\mathbf{h}}^H(n)\mathbf{X}(n) \quad (2.16)$$

$$\hat{\mathbf{h}}(n+1) = \hat{\mathbf{h}}(n) + \mu\mathbf{X}(n)(\mathbf{X}^H(n)\mathbf{X}(n) + \delta\mathbf{I})^{-1}\mathbf{e}^H(n) \quad (2.17)$$

where P is the projection order, μ is the step size parameter in the range $0 < \mu < 2$ [43], and δ is a small positive constant regularization parameter that helps to avoid numerical stability problems when inverting the $P \times P$ matrix product of $\mathbf{X}^H(n)\mathbf{X}(n)$. A method for selecting an optimal regularization parameter can be found in [44].

When $P = 1$ the APA described by (2.14) to (2.17) reduces to the NLMS algorithm described by (2.12). Thus, the APA can be considered as a generalization of the NLMS algorithm. As the projection order increases so does the rate of convergence for the APA as multiplication by the inverse of the $P \times P$ matrix in (2.17) helps to increasingly decorrelate $\mathbf{X}(n)$. However, as the projection order increases the computational complexity of the APA increases as well. The complexity of the APA is approximately $2NP + P_{inv}P^2$ mps, where P_{inv} is a constant associated with the complexity of computing the inverse of the $P \times P$ matrix in (2.17) [45]. To help reduce the complexity of the APA while still maintaining a faster rate of convergence compared the NLMS algorithm, “fast” versions of the algorithm for AEC have been

developed [45]–[47].

2.1.1.2 Subband Structures for Acoustic Echo Cancellation

A lower complexity alternative for AEC compared to standard fullband methods can be achieved via multirate signal processing and filter bank structures [48]. In this case a K channel analysis filter bank is used to filter and M fold downsample ($M \leq K$) the reference far-end signal $x(n)$, and input microphone signal $d(n)$ into K separate subband signals. The K subband reference and input signals are processed by a corresponding subband adaptive filter, where the resulting subband error signals are each upsampled, filtered, and combined by a synthesis filter bank to produce the fullband error signal. A block diagram of AEC implemented using subband adaptive filters is shown in Figure 2.3.

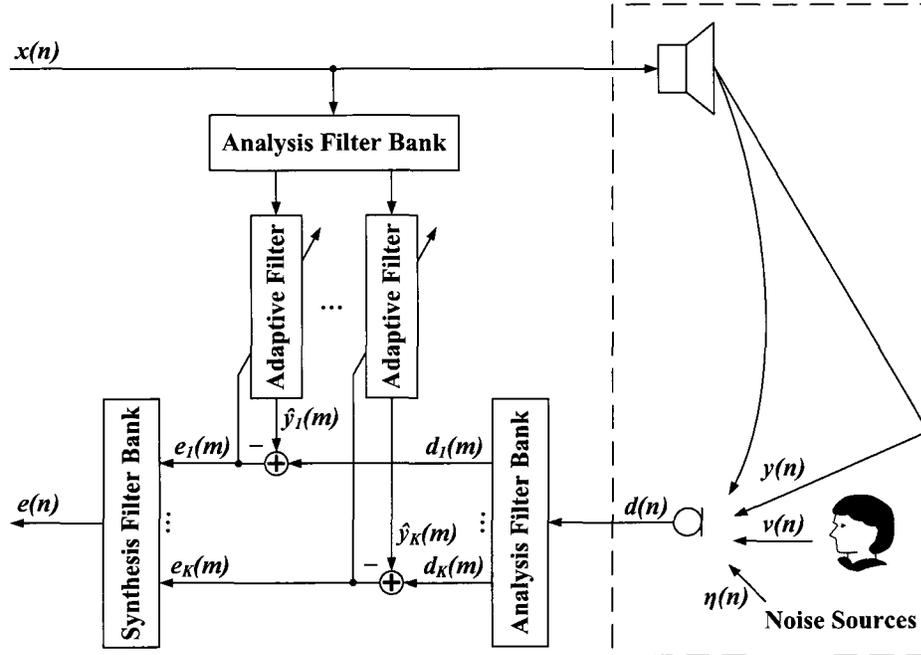


Figure 2.3: Acoustic echo cancellation using subband adaptive filters.

Compared to fullband adaptive filters, subband adaptive filter structures achieve

lower complexity by employing shorter-length adaptive filters per subband, which operate at an M fold reduced sampling rate. Furthermore, as the subband signals tend to be more spectrally flat, subband adaptive filters are able to converge faster than their fullband counterparts for correlated input signals such as speech [49]. Also, by using a subband structure to split the far-end reference and microphone signals into independent frequency bands, the adaptive algorithms can be tailored to exploit the signal characteristics within each subband. The most commonly used algorithms for subband adaptive filtering are the LMS and NLMS algorithms [50]. The APA has also been successfully used in subband adaptive filtering structures [51]–[54]. The main drawback of subband adaptive filtering structures for AEC is the delay introduced into the signal paths due to the analysis and synthesis filter banks, which can hinder the communication between parties. Thus, delayless subband structures have been proposed to circumvent the filter bank delay [55]–[57]. Another potential drawback is aliasing distortion in the subband signals due to the downsampling process and non-ideal filter banks, which increases the mean square error (MSE) of the subband adaptive filters [58]. As well, subband adaptive filters need to model the anti-causal subband LEMS impulse response taps that result from the temporal spreading of the fullband LEMS impulse response taps when they are translated to the subband domain, resulting in non-causal subband LEMS impulse responses [1]. An increase in MSE for the subband AEC structure will result if these anti-causal taps are not accounted for. Methods for implementing anti-causal subband LEMS impulse response taps are given in [1] and [50].

Subband adaptive filter structures achieve maximum computational savings when they are critically sampled ($M = K$). However, these maximally decimated filter banks are prone to aliasing in the subband signals due to non-ideal stopband attenuation at the cutoff frequencies [48]. One approach to deal with aliasing in critically sampled subband adaptive filter structures is to employ adaptive filters between adjacent subbands [59]. However, this approach tends to result in poor convergence

performance due to the extra adaptive filtering required to compensate for aliasing. Other critically sampled subband adaptive filter structures that mitigate subband aliasing without the need for extra adaptive filters between adjacent subbands, but still require extra computational resources to do so, are proposed in [52] and [60].

Oversampled filter banks arise when the downsampling factor is less than the number of filters used in the filter bank, $M < K$. These types of filter banks can mitigate aliasing in the subband signals if the analysis filters possess a high stopband attenuation [61]. The non-critical sampling also leads to analysis and synthesis filter banks that can be designed with shorter-length filters, which reduces signal path delay. Compared to critically sampled subband adaptive filter structures, oversampled structures tend to be more computationally complex due to the lower downsampling factor. This drawback is partially offset due to the extra processing required to prevent aliasing in critically sampled subband adaptive filter structures. Thus, oversampled subband adaptive filter structures provide a compromise between complexity reduction, aliasing distortion, and system delay.

A common method for the design of oversampled filter banks is through modulation of a prototype filter using discrete cosine transform (DCT) or discrete Fourier transform (DFT) operations [61]–[63]. Oversampled DFT modulated filter banks are preferential compared to oversampled DCT modulated filter banks as they tend to be more robust to aliasing [62]. Oversampled filter banks that possess the perfect reconstruction property, where the output is a delayed replica of the input, result in significant subband aliasing due to the relatively low stopband attenuation of the analysis filters, which is undesirable for subband AEC [50]. Thus, oversampled filter banks with near perfect reconstruction are more suitable for AEC as they pose better filter bank frequency characteristics resulting in lower subband aliasing, while introducing an acceptably small amount of reconstruction distortion [50], [62]. An efficient implementation of an oversampled filter bank can be achieved through polyphase factorization [63]. The polyphase implementation of the analysis and synthesis filters

allows them to operate at the M fold reduced sample rate, resulting in complexity savings compared to a standard filter bank implementation. A block diagram of an oversampled subband adaptive filter structure in polyphase form is shown in Figure 2.4 in the context of AEC. The polyphase analysis, $\mathbf{F}(z)$, and synthesis, $\mathbf{G}(z)$,

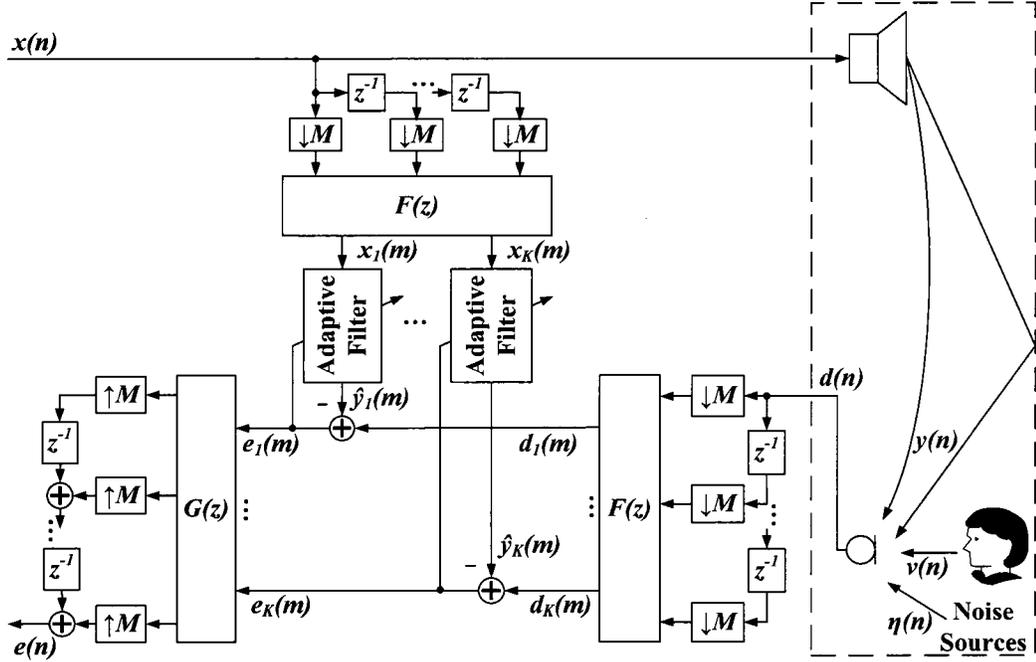


Figure 2.4: Acoustic echo cancellation using oversampled subband adaptive filters.

matrices can be created based on the modulation of a single prototype filter. The derivation of $\mathbf{F}(z)$ and $\mathbf{G}(z)$ along with the filter bank implementation details are given in [63]. One modulating transform that can be used in creating near perfect reconstruction oversampled filter banks is the generalized discrete Fourier transform (GDFT) [61]. The analysis, $\mathbf{f}_k(n)$, and synthesis, $\mathbf{g}_k(n)$, filters are determined by GDFT modulation of a prototype lowpass FIR filter $\mathbf{p}(n)$ as follows:

$$\mathbf{f}_k(n) = \mathbf{p}(n)e^{j\frac{2\pi}{K}(k+\frac{1}{2})(n+n_0)} \quad (2.18)$$

where $k = 0, \dots, K - 1$. The time offset $n_0 = -\frac{L_{pr}-1}{2}$ results in linear phase analysis and synthesis filters if the prototype filter is linear phase, where L_{pr} is the length of

$\mathbf{p}(n)$ which is assumed to be even. If $\mathbf{p}(n)$ is a real valued linear phase filter then selecting the synthesis filters to be time-reversed and complex conjugated version of the analysis filters results in $\mathbf{g}_k(n) = \mathbf{f}_k(n)$. A method for designing prototype filters under these conditions is outlined in [61]. Other methods for designing prototype filters for oversampled modulated filter banks can be found in [62].

The signal path delay introduced by analysis and synthesis filter banks based on the modulation of a single prototype FIR filter is determined by the group delay of the prototype filter. Linear phase prototype filters will have a constant group delay for all frequencies resulting in the desirable condition of no phase distortion and result in a delay of half of the filter order [64]. Thus, for analysis and synthesis filter banks based on the same prototype filter with length L_{pr} , the total delay introduced into the signal path is $L_{pr} - 1$ samples or the order of the prototype filter.

An alternative to performing adaptive filtering in the subband domain is frequency domain adaptive filtering [15], [65], [66]. These methods use block-based LMS adaptive filtering in the frequency domain to reduce complexity and improve convergence speed compared to their time domain counterparts. The main disadvantage of these techniques is their large algorithmic delay due to the block processing. Thus, many reduced delay frequency domain adaptive filtering schemes have been proposed [67]–[70]. As well, many of these reduced delay frequency domain adaptive filtering algorithms have been applied in the context of AEC [71]–[74]. It should also be noted that frequency domain adaptive filtering can be considered a special case of subband adaptive filtering as the DFT can be considered a critically sampled filter bank [15]. However, as the focus of this thesis involves subband adaptive filtering structures, standard frequency domain adaptive filtering is not discussed further.

2.2 Performance Limitations in Acoustic Echo Cancellation

The standard performance measures for AEC are echo return loss enhancement (ERLE) and MSE. For a digital EC, MSE is defined as the average error signal power and ERLE is the ratio of average microphone to error signal powers [10]. They are computed in decibels (dBs) as follows:

$$MSE(n) = 10 \log_{10} \left\{ E [e^2(n)] \right\} = 10 \log_{10} [\sigma_e^2(n)] \approx 10 \log_{10} \left[\frac{1}{K} \sum_{k=0}^{K-1} e^2(n-k) \right] \quad (2.19)$$

$$ERLE(n) = 10 \log_{10} \left\{ \frac{E [d^2(n)]}{E [e^2(n)]} \right\} = 10 \log_{10} \left[\frac{\sigma_d^2(n)}{\sigma_e^2(n)} \right] \approx 10 \log_{10} \left[\frac{\sum_{k=0}^{K-1} d^2(n-k)}{\sum_{k=0}^{K-1} e^2(n-k)} \right] \quad (2.20)$$

where $\sigma_d^2(n)$ and $\sigma_e^2(n)$ are the steady-state variances of $d(n)$ and $e(n)$ respectively, $E[\cdot]$ is the statistical expectation operator, and K in this case is the length of the window of samples over which the ERLE and MSE are calculated. As the variances of $d(n)$ and $e(n)$ are not typically known, MSE and ERLE can be estimated using their approximate forms from (2.19) and (2.20). ERLE can be interpreted as how much echo signal power has been removed by the EC, while MSE provides a measure of the remaining echo signal power. The results presented in later chapters of this thesis are based on the ERLE and MSE performance measures.

In a hands-free communication system the performance of an EC can be limited for many reasons. Noise sources are responsible for the upper bound on steady-state ERLE performance as they are uncorrelated from the echo signal $y(n)$ and cannot be suppressed by standard ECs. Background noise originating from objects within the LEMS as well as electronic noise from the hands-free device circuitry are the main noise sources. Undermodeling of the acoustic echo path, by employing an adaptive filter with a memory length that is insufficient to model the impulse response of the LEMS, will limit the ERLE performance of an EC [8]. Perceptual performance

limitations of an EC can also arise due to the psychoacoustic properties of the human auditory system [9]. In this case the standard ERLE and MSE performance measures may not reflect the perceivability of residual echo from an EC.

Nonlinearities in the LEMS also limit the ERLE performance of a linear EC. Major sources of nonlinearities in an AEC system include loudspeaker distortion, amplifier saturation, nonlinear DAC and ADC operation, and vibration effects of the hands-free device enclosure [5], [8]. Vibrations occurring in the hands-free device due to mechanical coupling between the microphone, loudspeaker, and their chassis, cause nonlinear rattling of components such as buttons. This results in another source of uncorrelated noise in the microphone signal that cannot be removed by conventional ECs. Nonlinearities that arise from DACs and ADCs are due to a nonlinear mapping relationship between the input and output signals caused by non-ideal hardware components [75]. The quantization errors produced by ADCs due to finite word lengths in digital signal processors (DSPs) is an inherently nonlinear process. The nonlinear behaviour of DACs and ADCs has been successfully modeled with a third order Taylor series in the context of nonlinear network echo cancellation [76]. However, the nonlinear effect of DACs and ADCs tends to be small especially with the high resolution converters used in modern DSPs. Another source of nonlinearity in the LEMS can result from the nonlinear operation of amplifiers when they are overdriven. The nonlinear behaviour of an amplifier is roughly memoryless and can be modeled by a saturation characteristic where soft clipping occurs at high amplitude values [34]. A method for nonlinear AEC that addresses the nonlinear amplifier problem is given in [77]. The preamplifier for the microphone signal is generally considered to be operating in its linear region due to the relatively small value of the microphone signal. Nonlinear behaviour of the loudspeaker can also reduce ERLE performance of a linear EC and cause annoying nonlinear echoes to the far-end talker [34]. AEC in the presence of loudspeaker distortion is discussed in detail in Section 2.2.1.

The operation of an EC will be disrupted whenever changes occur in the acoustic

environment such as movement of local talkers or objects [6]. These nonstationarities will cause the LEMS impulse response to deviate from its original state resulting in a performance drop for the EC. The EC will attempt to track these disturbances by adjusting its filter taps to match the new LEMS impulse response. As the EC reconverges its ERLE performance will increase back to a steady-state value. To minimize the impact on the quality of the hands-free communication it is essential that the EC tracks the LEMS impulse response variations as fast as possible. A detailed discussion of AEC in nonstationary environments is presented in Section 2.2.2.

2.2.1 Acoustic Echo Cancellation in the Presence of Loudspeaker Nonlinearity

When operated at high signal levels, electrodynamic loudspeakers in a hands-free communication system exhibit nonlinear distortion which inhibits the ERLE performance of a linear EC, and the corresponding quality of communication between parties [16]. Thus, to ensure that the hands-free communication between parties is not corrupted by loudspeaker distortion, the EC should model and remove this nonlinearity in addition to modeling and removing the linear echo path components. An overview of the different components within an electrodynamic loudspeaker and their nonlinear behaviour can be found in [78]. A cross sectional diagram of an electrodynamic loudspeaker is shown in Figure 2.5. The main sources of nonlinearities within a

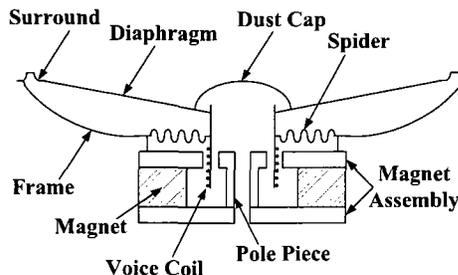


Figure 2.5: Cross sectional diagram of an electrodynamic loudspeaker.

loudspeaker are as follows [78]. The suspension system of the loudspeaker, consisting of the spider and surround, has a nonlinear stiffness associated with it resulting in a distortion of the output signal occurring mainly at low frequencies. Nonlinear vibrations can occur on the diaphragm at high frequencies due to high stress and strain in the diaphragm material. The driving force on the voice coil is a nonlinear function of its displacement due to a weakening magnetic field outside the magnetic air gap. Also, the inductance within the voice coil varies nonlinearly with its displacement. Nonlinear sound wave propagation can also occur, however this tends to occur in horn loudspeakers only. Other factors that can contribute to the nonlinear distortion of a loudspeaker include defects such as loose glue joints, the voice coil rubbing on the magnet assembly, and loose particles in the magnetic air gap. In addition, at very high signal levels mechanical clipping of the voice coil can occur if its displacement range is insufficient for the required movement [34].

One common way to model the nonlinear response of a loudspeaker is through Volterra series expansions [17], [18]. Other methods include loudspeaker linearization through preprocessing of the loudspeaker signal [79]. A major drawback to loudspeaker linearization schemes is that they generally require *a priori* knowledge of the loudspeaker nonlinearities which may not be available. Neural networks have also been used to model loudspeaker distortion in the context of AEC [80], [81]. However, the computational complexity of these methods may still be prohibitively high for practical AEC. For miniaturized loudspeakers, such as those found in cellular telephones, it has been found that a memoryless saturation characteristic can approximate the nonlinearity, which has been applied in the context of AEC in [82], and [83]. The nonlinear behaviour of an electrodynamic loudspeaker requires a model with memory, if it is not operated in its saturation region, due to the long time constant of its electro-mechanical system [77]. The application of Volterra series expansions for loudspeaker distortion modeling in AEC is discussed thoroughly in the next section.

2.2.1.1 Nonlinear Loudspeaker Modeling Using Volterra Filters in Acoustic Echo Cancellation

The general discrete time P_V -th order finite memory Volterra series expansion of an input signal, $x(n)$, is given by [84]:

$$y(n) = h_0 + \sum_{p=1}^{P_V} \left[\sum_{m_1=0}^{L_p-1} \cdots \sum_{m_p=0}^{L_p-1} h_p(m_1, \dots, m_p) x(n - m_1) \cdots x(n - m_p) \right] \quad (2.21)$$

where h_0 is a constant term and $h_p(m_1, \dots, m_p)$ are the p -th order Volterra kernels that characterize the nonlinear system with memory lengths L_p . An important property of (2.21) is the linear relationship between the output, $y(n)$, and the Volterra kernels, $h_p(m_1, \dots, m_p)$ [19]. This linear relationship allows direct application of adaptive algorithms (e.g. LMS) to Volterra filters. Another important property of (2.21) is the symmetry of the Volterra kernels due to the commutativity of $x(n - m_1) \cdots x(n - m_p)$. This allows a direct reduction of the kernel coefficients from L_p^p to $\binom{L_p+p-1}{p}$. Taking the symmetry of the Volterra kernel coefficients into consideration, the Volterra series expansion of $x(n)$ in triangular form becomes:

$$y(n) = h_0 + \sum_{p=1}^{P_V} \left[\sum_{m_1=0}^{L_p-1} \sum_{m_2=m_1}^{L_p-1} \cdots \sum_{m_p=m_{p-1}}^{L_p-1} h_p(m_1, m_2, \dots, m_p) \times x(n - m_1) x(n - m_2) \cdots x(n - m_p) \right]. \quad (2.22)$$

Second order Volterra filters have been found to effectively model the nonlinear harmonic distortion of small sized loudspeakers [16], [23]. With this in mind the truncated second order Volterra series expansion of input signal $x(n)$ is given in triangular form as:

$$y(n) = \sum_{m_1=0}^{L_1-1} h_1(m_1) x(n - m_1) + \sum_{m_1=0}^{L_2-1} \sum_{m_2=m_1}^{L_2-1} h_2(m_1, m_2) x(n - m_1) x(n - m_2) \quad (2.23)$$

where $h_1(m_1)$ is the linear Volterra kernel with memory length L_1 , representing the linear portion of the system. The nonlinear portion of the system is represented by the quadratic Volterra kernel, $h_2(m_1, m_2)$, with total memory length $L_2(L_2 + 1)/2$. The linear and quadratic Volterra kernels are denoted as vectors by $\mathbf{h}_1(n)$ and $\mathbf{h}_2(n)$, respectively. Also, the h_0 term has been dropped without loss of generality.

Figure 2.6 shows the nonlinear system model for AEC using second order adaptive Volterra filters. Here the microphone signal $d(n)$ and error signal $e(n)$ are determined

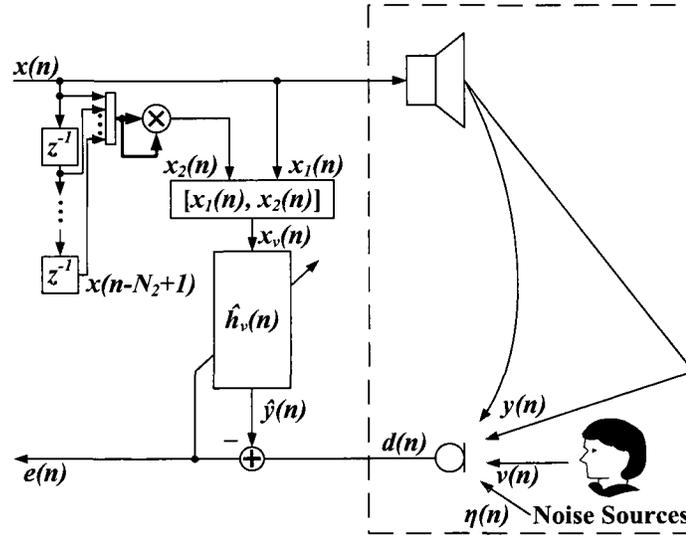


Figure 2.6: Nonlinear acoustic echo cancellation using a second order adaptive Volterra filter.

as in (2.1) and (2.2). The echo signal $y(n)$ is computed based on first and second order convolutions with the far-end input signal $x(n)$ as given in (2.23). The echo signal estimate $\hat{y}(n)$ at the output of the adaptive second order Volterra filter is determined as follows:

$$\hat{y}(n) = \hat{\mathbf{h}}_v^H(n) \mathbf{x}_v(n) \quad (2.24)$$

with the corresponding Volterra kernel coefficient vector estimate definitions:

$$\hat{\mathbf{h}}_v(n) = [\hat{\mathbf{h}}_1^T(n), \hat{\mathbf{h}}_2^T(n)]^T \quad (2.25)$$

$$\hat{\mathbf{h}}_1(n) = [\hat{h}_{1_0}(n), \hat{h}_{1_1}(n), \dots, \hat{h}_{1_{N_1-1}}(n)]^T \quad (2.26)$$

$$\hat{\mathbf{h}}_2(n) = [\hat{h}_{2_{0,0}}(n), \hat{h}_{2_{0,1}}(n), \dots, \hat{h}_{2_{0,N_2-1}}(n), \hat{h}_{2_{1,1}}(n), \dots, \hat{h}_{2_{N_2-1,N_2-1}}(n)]^T \quad (2.27)$$

and input signal vector definitions:

$$\mathbf{x}_v(n) = [\mathbf{x}_1^T(n), \mathbf{x}_2^T(n)]^T \quad (2.28)$$

$$\mathbf{x}_1(n) = [x(n), x(n-1), \dots, x(n-N_1+1)]^T \quad (2.29)$$

$$\mathbf{x}_2(n) = [x^2(n), x(n)x(n-1), \dots, x(n)x(n-N_2+1), x^2(n-1), \dots, x^2(n-N_2-1)]^T. \quad (2.30)$$

The length of the linear input signal vector $\mathbf{x}_1(n)$ and linear Volterra kernel coefficient vector estimate $\hat{\mathbf{h}}_1(n)$ is N_1 where $N_1 \leq L_1$. The total length of the quadratic input signal vector $\mathbf{x}_2(n)$ and quadratic Volterra kernel coefficient vector estimate $\hat{\mathbf{h}}_2(n)$ is $N_2(N_2+1)/2$ where $N_2 \leq L_2$. The length of the combined input signal vector $\mathbf{x}_v(n)$ and combined Volterra kernel coefficient vector estimate $\hat{\mathbf{h}}_v(n)$ is the sum of the lengths of the constituent parts.

The NLMS algorithm can be applied to determine the Volterra filter coefficients that minimize the error signal $e(n)$ in the mean-square sense analogously to the pure linear filter approach described in (2.12) to (2.13). In this case the NLMS filter tap update equation becomes:

$$\hat{\mathbf{h}}_v(n+1) = \hat{\mathbf{h}}_v(n) + \frac{\mu}{\mathbf{x}_v^H(n)\mathbf{x}_v(n) + \delta} \mathbf{x}_v(n)e^*(n) \quad (2.31)$$

where $e(n)$ is given by (2.2) and the step size bounds for stability have been shown to be the same as in (2.13) [85].

The linear and quadratic Volterra kernel estimates can also be adapted with separate step sizes using the NLMS algorithm as follows [86]:

$$\hat{y}(n) = \hat{\mathbf{h}}_1^H(n)\mathbf{x}_1(n) + \hat{\mathbf{h}}_2^H(n)\mathbf{x}_2(n) \quad (2.32)$$

$$\hat{\mathbf{h}}_1(n+1) = \hat{\mathbf{h}}_1(n) + \frac{\mu_1}{\mathbf{x}_v^H(n)\mathbf{x}_v(n) + \delta_1} \mathbf{x}_1(n)e^*(n) \quad (2.33)$$

$$\hat{\mathbf{h}}_2(n+1) = \hat{\mathbf{h}}_2(n) + \frac{\mu_2}{\mathbf{x}_v^H(n)\mathbf{x}_v(n) + \delta_2} \mathbf{x}_2(n)e^*(n) \quad (2.34)$$

where the stability bounds for the step sizes is $0 < \mu_1, \mu_2 < 2$ and $e(n)$ is given by (2.2). Also, δ_1 and δ_2 are both small positive constants that help to avoid numerical stability problems when \mathbf{x}_v is comprised of small values. Figure 2.7 reflects this fully coupled version of an adaptive second order Volterra filter for AEC.

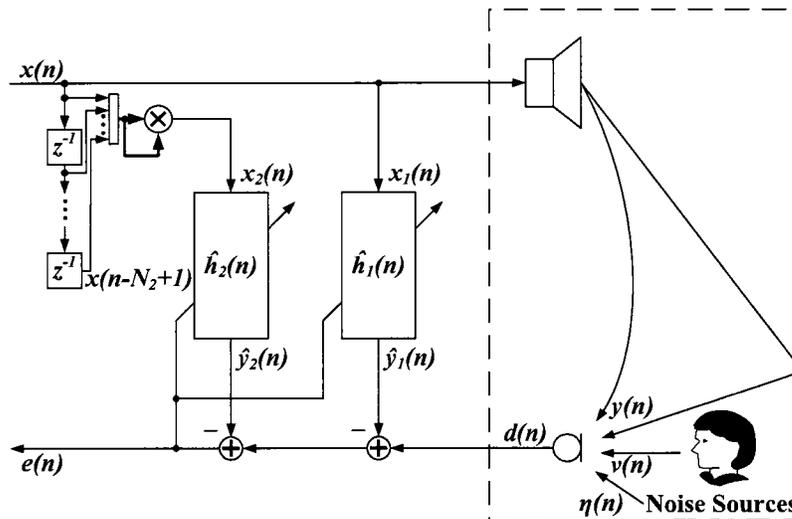


Figure 2.7: Nonlinear acoustic echo cancellation using a fully coupled second order adaptive Volterra filter.

Furthermore, the partially decoupled version of an adaptive second order Volterra filter [87] for AEC is shown in Figure 2.8. In this case the NLMS algorithm for

updating $\hat{\mathbf{h}}_1(n)$ and $\hat{\mathbf{h}}_2(n)$ becomes [88]:

$$e_1(n) = d(n) - \hat{\mathbf{h}}_1^H(n)\mathbf{x}_1(n) \quad (2.35)$$

$$\hat{\mathbf{h}}_1(n+1) = \hat{\mathbf{h}}_1(n) + \frac{\mu_1}{\mathbf{x}_1^H(n)\mathbf{x}_1(n) + \delta_1} \mathbf{x}_1(n)e_1^*(n) \quad (2.36)$$

$$\hat{\mathbf{h}}_2(n+1) = \hat{\mathbf{h}}_2(n) + \frac{\mu_2}{\mathbf{x}_2^H(n)\mathbf{x}_2(n) + \delta_2} \mathbf{x}_2(n)e_1^*(n) \quad (2.37)$$

where $\hat{y}(n)$ is given by (2.32), and $e(n)$ is given by (2.2). In general the partially decoupled structure can achieve faster convergence compared to the fully coupled version, at the expense of a slight increase in MSE [87]. Also, it should be noted that the separate adaptation of the linear and quadratic Volterra kernel estimates assumes that the elements of $\mathbf{x}_1(n)$ and $\mathbf{x}_2(n)$ are approximately orthogonal as discussed in [88].

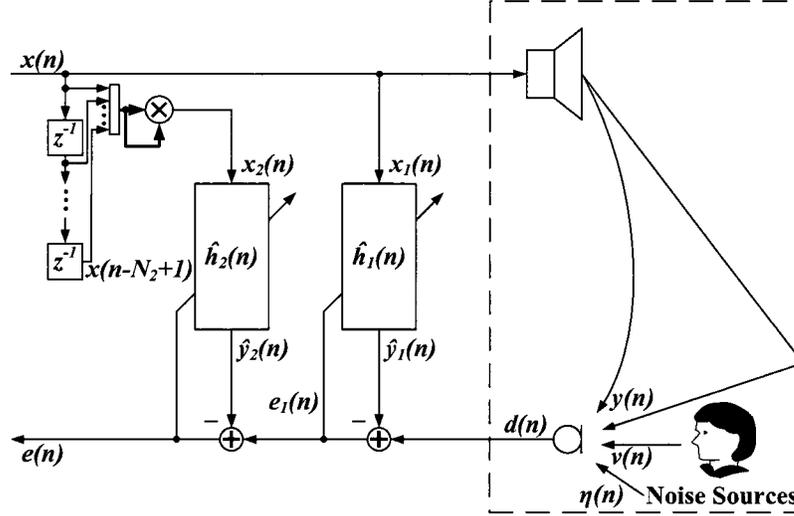


Figure 2.8: Nonlinear acoustic echo cancellation using a partially decoupled second order adaptive Volterra filter.

The use of adaptive Volterra filters to compensate for the nonlinear behaviour of electrodynamic loudspeakers in AEC systems has been well researched [19]–[22].

The biggest drawback to using adaptive Volterra filters is their very large complexity requirement, which can be prohibitive for practical AEC especially as the order of the Volterra series increases. Thus, there have been many algorithms proposed for reduced complexity nonlinear AEC using adaptive Volterra filters [16], [23]–[26], [86], [89], [90]. In general, these structures employ only second order adaptive Volterra filters. Also, for AEC the memory length of the second order Volterra kernel can be chosen much less than the first order Volterra kernel while still effectively modeling loudspeaker distortion, thus helping to reduce complexity. The time domain approaches (see [23], [86], [89]) reduce complexity by exploiting symmetries within the Volterra series expansion, taking into consideration the nature of the loudspeaker distortion, and using simplified filtering structures. Still, the complexity of these time domain approaches may be prohibitively high. The frequency domain methods (see [16], [24], [25]) reduce complexity further by exploiting fast convolution techniques in the discrete Fourier transform domain. However, these frequency domain structures can be complex to implement. The subband domain techniques (see [26], [90]) aim to decrease complexity by performing the adaptive filtering operations in multiple frequency bands at a reduced sampling rate, where shorter adaptive filter lengths are employed. The subband technique in [26] uses a delay-less subband structure that considers adaptive Volterra filters beyond the second order. However, as many second order channels may need to be considered by this structure for accurate modeling of loudspeaker distortion, the structure’s complexity will increase significantly. The authors in [90] present a subband parallel cascade structure which minimizes complexity by considering only a subset of the total number of eigenvalues in the second order Volterra filter matrix. In cases where second order Volterra filters with large memory lengths are required, the number of eigenvalues needed for adequate performance will increase, resulting in high complexity levels for this structure. Thus, there is still a need for reduced complexity nonlinear EC structures based on adaptive Volterra filters, especially when the loudspeaker distortion requires a filter

with a large memory length in order to be accurately modeled. An efficient subband adaptive second order Volterra filter structure for AEC is presented in Chapter 5.

2.2.2 Acoustic Echo Cancellation in Nonstationary Environments

The LEMS is an inherently nonstationary system as rapid changes within the environment, such as movement of persons and objects, occur often resulting in fluctuations in the LEMS impulse response. To ensure good AEC performance an EC needs to be able to track these changes quickly. However, tracking of echo path variations is inhibited by the typically long adaptive filters required to model the LEMS and the potentially slow convergence of the adaptive algorithm used, such as the NLMS algorithm [6]. Adaptive algorithms that achieve fast initial convergence, such as the APA and RLS algorithm, may not necessarily offer fast tracking performance in their direct form [12]. For the standard LMS adaptive algorithms the steady-state MSE and convergence rate are proportional to the step size parameter. Thus, large step sizes result in fast convergence and high MSE while small step sizes result in slow convergence and low MSE. To improve upon the convergence and tracking performance of standard LMS and NLMS algorithms, many variable step size LMS based adaptive algorithms have been proposed [91]–[96]. These algorithms aim to mitigate the trade off between convergence speed and MSE, caused by the fixed single step size LMS algorithms, by using a time varying step size parameter. In general the variable step size parameter is large when the filter taps are far from their optimal values, and is small when the taps are close to their optimal values [96]. The drawbacks to these variable step size algorithms is their potential sensitivity to the selection of certain algorithm parameters, and to the characteristics of the input signal [95]. As well, an increase in complexity compared to typical LMS algorithms is incurred. Thus, robust variable step size LMS algorithms have been proposed that are less sensitive to the

limitations of the standard approaches, examples include [95] and [96].

Another type of variable step size algorithm that exploits the time domain characteristics of the LEMS impulse response are exponentially weighted step size algorithms [97], [98]. These algorithms exploit the general exponentially decaying nature of the LEMS impulse response by applying a corresponding exponentially weighted time invariant update to each filter tap. This allows for significantly faster convergence and tracking compared to standard LMS algorithms at a potentially similar complexity. The disadvantage to the exponentially weighted step size algorithms is the dependence of the time invariant exponential weight update on the conditions of the LEMS, which may not be known *a priori*. Proportionate step size algorithms are presented in [11] and [12]–[14], that exploit the time domain features of the LEMS impulse response through individual filter tap adjustment in relation to the magnitude of the current tap estimates, allowing for fast tracking and convergence. As the filter taps of the LEMS impulse response model with large coefficients tend to fluctuate more significantly than taps with smaller coefficients, in response to a change within the acoustic environment, an adaptive algorithm with step-sizes proportional to the filter tap coefficients results in fast tracking. These algorithms avoid the drawback of the exponential step size algorithms while achieving fast convergence and tracking in LEMS with sparse or semi-sparse impulse responses, which are typical for most AEC systems. Compared to the variable step size algorithms above, that use a global time varying step size to adjust all the filter taps together, these individual filter tap step size algorithms can achieve faster convergence and tracking by exploiting the time domain nature of the LEMS impulse response. However, the downside to these proportionate step size algorithms is their increased complexity compared to the standard LMS and NLMS algorithms. Thus, there is a need for reduced complexity proportionate step size algorithms for echo cancellation in changing acoustic environments.

The application of the proportionate step size algorithm in [14] to second order

adaptive Volterra filters for nonlinear echo cancellation in the presence of loudspeaker distortion is presented in [88]. However, the performance of this structure in changing acoustic environments is not investigated and its computational cost is very high, especially for linear and quadratic Volterra filters with long memory lengths. Thus, there is a need for reduced complexity proportionate step size adaptive Volterra filter structures, capable of fast convergence and tracking of linear and nonlinear system components in the context of AEC.

2.2.2.1 Tracking Algorithms for Acoustic Echo Cancellation

The proportionate normalized least-mean-square (PNLMS) algorithm presented in [11] achieves fast convergence and tracking by using individual filter tap step sizes based on the magnitude of the current filter tap coefficient. Thus, large step sizes are applied to filter taps with large magnitudes and small step sizes to taps with small magnitudes. The PNLMS method achieves fast convergence and tracking for sparse impulse response environments but is impaired when the impulse response is dispersive [14]. The improved proportionate normalized least-mean-square (IPNLMS) algorithm in [14] enhances the PNLMS scheme by updating the filter taps based on the average magnitude of the current filter coefficients, along with the magnitude of the current filter tap coefficient itself. This allows the IPNLMS algorithm to converge and track faster when the LEMS impulse response is sparse or semi-sparse. The generalized gradient proportionate variable step size algorithm discussed in [12] can achieve faster convergence and tracking performance than the PNLMS and IPNLMS algorithms, since it bases its individual filter tap step sizes on the gradient estimate provided by the difference between the current filter tap coefficient and a delayed average filter tap coefficient. This gradient estimate does not have any time invariant components that lead to slower tracking performance. Also, the exponential step size algorithm in [97] as well as the PNLMS and IPNLMS algorithms arise as special cases of the generalized gradient proportionate affine projection algorithm (GGPAPA)

outlined in [12] with the appropriate setting of parameters. Using (2.8) and (2.14) to (2.16) in Section 2.1.1.1 from the APA, the GGPAPA is computed as follows [12]:

$$\hat{\mathbf{h}}(n+1) = \hat{\mathbf{h}}(n) + \mu \mathbf{A}(n) \mathbf{X}(n) \left(\mathbf{X}^H(n) \mathbf{A}(n) \mathbf{X}(n) + \delta \mathbf{I} \right)^{-1} \mathbf{e}^H(n) \quad (2.38)$$

$$\mathbf{A}(n) = \text{diag} [g_0(n), \dots, g_{N-1}(n)] \quad (2.39)$$

$$g_l(n) = \frac{1-\beta}{2} \frac{\bar{c}_l(n-1)}{\sum_{i=0}^{N-1} \bar{c}_i(n-1) + \xi} + \frac{\beta}{2N} \quad (2.40)$$

$$\bar{c}_l(n) = \epsilon \bar{c}_l(n-1) + (1-\epsilon) |c_l(n)| \quad (2.41)$$

$$c_l(n) = \hat{h}_l(n-1) - \gamma \bar{h}_l(n-1) \quad (2.42)$$

$$\bar{h}_l(n) = \eta \bar{h}_l(n-1) + (1-\eta) \hat{h}_l(n-1) \quad (2.43)$$

where $l = 0, \dots, N-1$. The gradient estimate and its average are represented by $c_l(n)$ and $\bar{c}_l(n)$ respectively. The average filter tap coefficient is denoted by $\bar{h}_l(n)$ and γ is a correction factor for $\bar{h}_l(n-1)$. The forgetting factors ϵ and η are used in computing $\bar{c}_l(n)$ and $\bar{h}_l(n)$ respectively. The $\text{diag}[\cdot]$ operator creates a square matrix with its arguments on the main diagonal and zero elements elsewhere. The overall step size parameter is denoted by μ and must be in the range $0 < \mu < 2$ for stability as for the standard APA. When ϵ is set to 0 the GGPAPA becomes a non-generalized version called gradient proportionate affine projection algorithm (GPAPA) as discussed in [12]. Furthermore, when γ is set to 0 the GPAPA reduces to an improved proportionate affine projection algorithm (IPAPA). In addition, setting β close to 0 leads to behaviour like a proportionate affine projection algorithm (PAPA), while setting $\beta = 1$ and dividing δ by $2N$ results in the APA. Thus, β controls the weighting between constant and proportionate step sizes. Moreover, under the same conditions with P set to 1 the GPAPA reduces to a gradient proportionate normalized least-mean-square (GPNLMS) algorithm, the IPAPA reduces to the IPNLMS algorithm, the PAPA reduces to the PNLMS algorithm, and the APA reduces to the NLMS

algorithm. The exponential step size algorithm of [97] can be derived from the GGPAPA as ϵ approaches 1 [12]. A computationally efficient subband EC structure based on the GGPAPA, for fast convergence and tracking of linear echo path components, is presented in Chapter 4. Also, a subband EC structure for fast convergence and tracking of both linear and nonlinear echo path components is outlined in Chapter 6.

2.3 Summary

This chapter presented a literature review of signal processing structures and adaptive algorithms in the context of AEC. Performance limitations in AEC were also discussed along with current state-of-the-art approaches to mitigate them, specifically in nonstationary and nonlinear acoustic environments. As the focus of this thesis is computationally efficient subband structures for AEC in nonstationary and nonlinear environments, several of the topics discussed previously fall outside the scope of the thesis. Specifically, network echo cancellation along with post-filtering, VAD, and doubletalk detector structures are not further considered. As well, only the nonlinear behaviour of electrodynamic loudspeakers is addressed by the AEC structures proposed in this thesis. While the nonlinear behaviour of signal converters and internal device amplifiers are not considered as they are assumed to be linear components. Furthermore, only second order Volterra filters are employed by the proposed AEC structures in this thesis for modeling loudspeaker distortion.

Chapter 3

Experimental Measurements and Data Analysis

This chapter outlines the experimental setup used for collecting data from practical AEC systems, along with the procedure used for estimating the corresponding echo path components. A time and frequency domain analysis is also presented for the estimated linear and nonlinear echo path components. Also, a harmonic distortion analysis is given for the hands-free devices used in each LEMS. The chapter sections are organized as follows: Section 3.1 describes the experimental setup, including the equipment used along with the method for modeling the LEMS components in question. The time and frequency domain interpretation of the measured linear and nonlinear LEMS components is discussed in Sections 3.2 and 3.3 respectively. Finally, harmonic distortion results for the hands-free devices used in the AEC system are presented in Section 3.4.

3.1 Measurement Methodology

To perform meaningful simulations for the AEC structures presented in this thesis, input signal data from several practical LEMS scenarios were recorded. This allows

simulations to be performed based on the experimentally measured data directly, as well as on synthetic input signals created using the LEMS components extracted from the measured data. The measurements were performed at Carleton University in a large office room (MC 6015) measuring approximately 9 meters by 9 meters. In all measurement scenarios the reference signal played through the loudspeaker, driven by a Samson Servo-150 amplifier, and resulting microphone signal were digitized simultaneously at a 48 kHz sampling rate via a MOTU 896HD recording system with 24-bit ADCs. Adobe Audition software running on a host computer was used to control the playback and recording of the measurement signals via a FireWire interface.

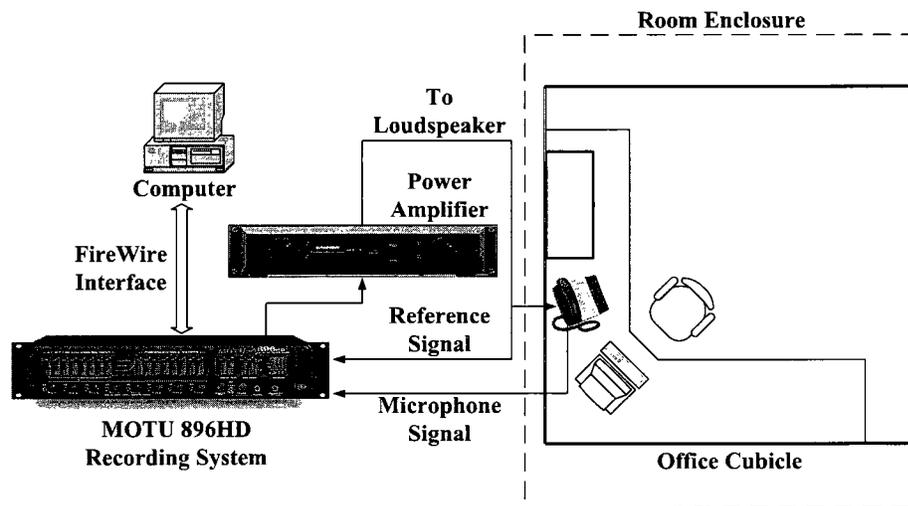


Figure 3.1: Office room acoustic echo cancellation measurement setup.

LEMS data were measured in the large office room from a conference phone and a smartphone operating in hands-free mode, that were located in an unobstructed position on an office desk. A block diagram of the LEMS measurement setup is shown in Figure 3.1. Each phone set has a microphone mounted inside the bottom lip of the chassis. The conference phone has a top centre mounted loudspeaker measuring approximately 6.4 centimeters in diameter with the bottom left mounted microphone located roughly 23 centimeters from the centre of the loudspeaker. The smartphone

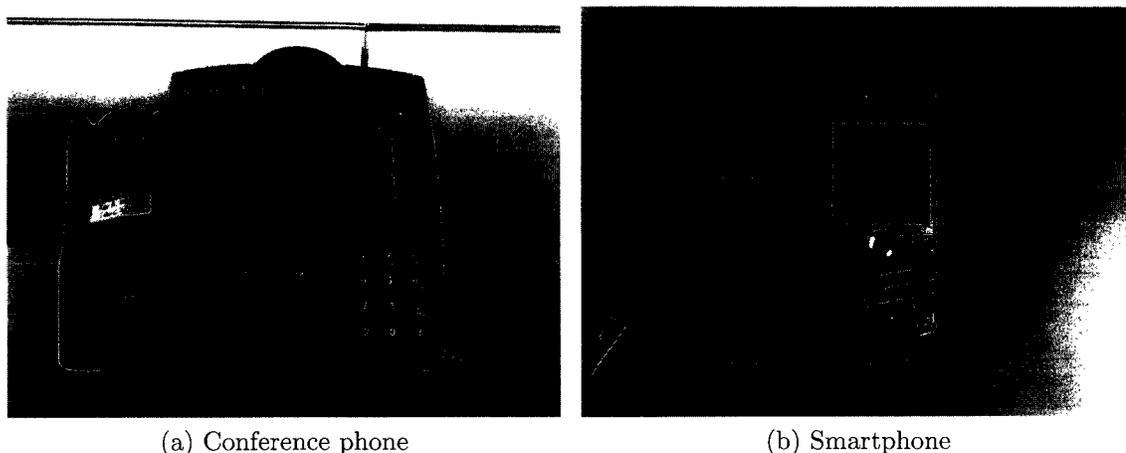


Figure 3.2: Commercial phone sets.

contains a miniaturized loudspeaker that is approximately 10 centimeters from the bottom mounted microphone. The internal amplifiers of both devices were not available for driving the loudspeaker within each respective device. Thus an external amplifier was used as shown in Figure 3.1. Photographs of the phone sets are shown in Figure 3.2. For both configurations data was recorded separately for 30 seconds of white Gaussian noise (WGN) and speech input loudspeaker signals. The speech signal consisted of concatenated speech segments from different male and female speakers within the TIMIT speech data base [99]. To investigate the impact of loudspeaker distortion on each LEMS, measured data was collected for several fundamental tones under increasing signal levels described in Section 3.4 below. As well, WGN and speech data were recorded under low, moderate, and high input loudspeaker signal levels. For the conference phone these input signal levels corresponded to amplifier output root mean square (RMS) levels of 0.6, 1.7, and 3.2 Volts respectively and to 0.5, 1.7, 2.9 Volts respectively for the smartphone, based on a 1 kHz test tone. Since the loudspeaker in both devices was driven by an external amplifier operating well within its linear region for all signal levels, any nonlinear amplifier saturation that

could have occurred by using the internal device amplifiers was avoided. Furthermore, to analyze the affect of an obstruction near each phone set, measured data was obtained under moderate input signal levels with a persons hand on the keypad of each phone set. The resulting measured linear and nonlinear echo path components for the various measurement setups are discussed in Sections 3.2 and 3.3 respectively.

3.2 Measured Linear Echo Path Impulse Responses

Estimates of the linear impulse responses for both LEMS configurations were determined based on the measured WGN input signal data. The recorded loudspeaker and microphone signals were decimated to an 8 kHz sampling rate to match that used in standard narrow band telephony, and then the NLMS algorithm was used to estimate the linear impulse response for each LEMS scenario. The linear NLMS filters were set to adapt to a 250 millisecond impulse response estimate, resulting in a filter length of $N = 2000$ in all cases. A small step size of $\mu = 0.1$ was used throughout the course of each adaptation to ensure a low steady-state modeling error. The resulting linear impulse responses for each of the hands-free systems under unobstructed conditions with moderate input signal levels, are shown in Figure 3.3. Both of the measured linear impulse responses share the same general exponentially decaying shape with the main differences occurring in the first part of the response. These measured linear impulse responses are typical of those for a LEMS found in the literature in the context of AEC [27], [52]. The initial part of each impulse response is due to the direct loudspeaker to microphone path and early reflections from the desk surface, which vary between each LEMS configuration. The remainder of each impulse response is a result of later reflections from walls and other more distant objects within the enclosure, which are similar for each configuration due to the common location of the hands-free device within the enclosure. The linear impulse responses obtained under the other measurement conditions resulted in the same set of observations.

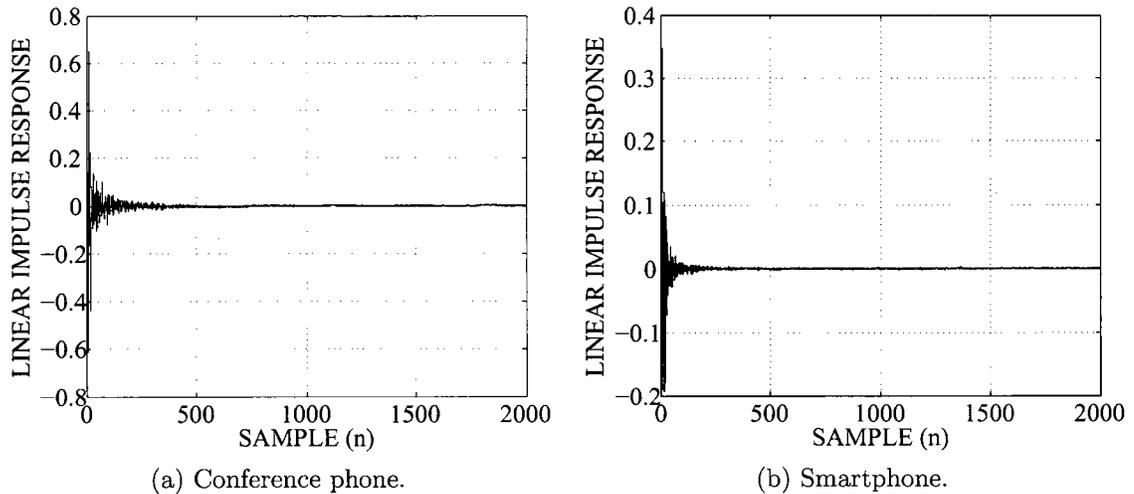


Figure 3.3: Measured linear impulse responses from the phone set configurations.

Figure 3.4 shows the unobstructed linear impulse response from Figure 3.3a with the associated obstructed linear impulse response obtained from the conference phone configuration for the first 100 taps. The first several impulse response coefficients, corresponding to the direct path and initial early reflections off the desk surface, are the same for both impulse responses due to the fixed location of the loudspeaker and microphone. The reflections represented by approximately taps 10 through 50 are noticeably different between the two responses, as a result of the new reflections introduced by the hand on keypad obstruction. The remainder of the linear impulse response taps caused by later reverberations from the walls, ceiling, and other distant objects are unchanged between the two responses. These findings are similar to those reported in [100] for impulse responses measured with and without a small object placed near a phone set in a LEMS.

The corresponding linear magnitude frequency responses for the unobstructed and obstructed LEMS setups with the conference phone are shown in Figure 3.5. For frequencies below roughly 1000 Hz the magnitude responses are nearly identical while they deviate considerably at higher frequencies. The low frequency similarities are a result of the corresponding longer wavelengths of sound being relatively unaffected by

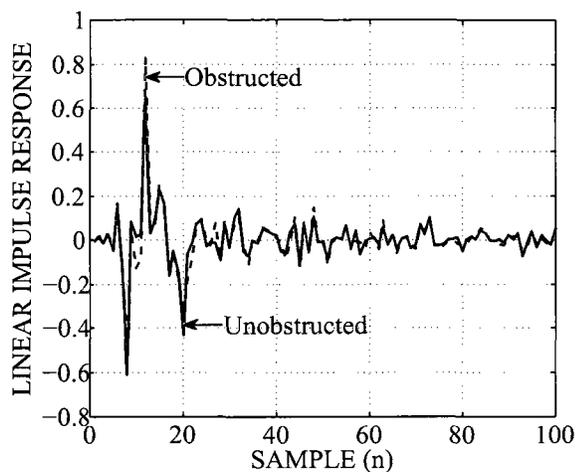


Figure 3.4: Measured linear impulse responses from the unobstructed and obstructed conference phone configurations.

the small hand obstruction. At higher frequencies the shorter wavelengths are more sensitive to the impediment, which is shown by the significant variations between the two responses. It should also be noted that at frequencies greater than approximately 1000 Hz, both responses tend to have a greater variability than in the lower frequency region. This is a result of the greater absorption of sound waves at higher frequencies by the objects and surfaces within the acoustic environment [101]. The obstructed linear impulse response measurements from the conference phone at the other signal levels, as well as those for the smartphone LEMS configuration at all signal levels, yielded similar trends in both the time and frequency domains when compared to the corresponding unobstructed impulse responses. These time and frequency domain observations provide the motivation for a subband AEC structure with adaptive filters based on a proportionate step size adaptive algorithm, to allow for fast tracking of echo path changes while requiring low computational complexity. Such an AEC structure is presented in Chapter 4.

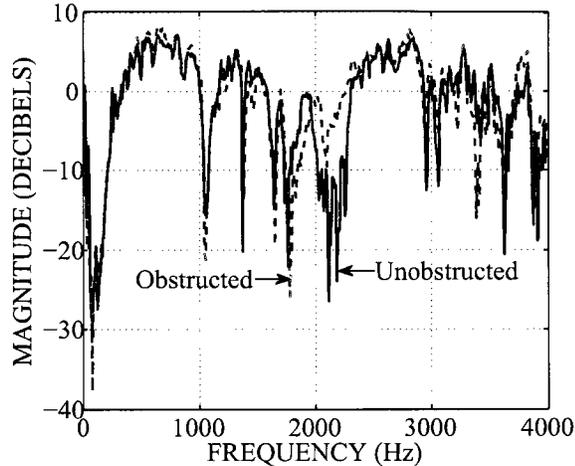


Figure 3.5: Measured linear frequency responses from the unobstructed and obstructed conference phone configurations.

3.3 Measured Quadratic Volterra Kernels

Quadratic Volterra kernels for each LEMS were determined based on the corresponding measured speech input signal data. The partially decoupled second order adaptive Volterra filter structure described in Section 2.2.1.1 of Chapter 2 was used to identify the quadratic Volterra kernel component of each LEMS using the NLMS algorithm given by (2.35) to (2.37). The quadratic NLMS filters were set to 100×100 tap models for the nonlinear portion of each LEMS, with a small step size of $\mu_2 = 0.1$ used for the duration of the adaptation process. As only the nonlinear portion of the echo paths were of interest in this case the linear portions were not identified, which was achieved by setting $\mu_1 = 0$. The resulting quadratic Volterra kernel for each LEMS configuration under unobstructed conditions with high input signal levels, are shown along with their corresponding frequency responses in Figures 3.6 and 3.7. In both cases the measured quadratic Volterra kernels have the same basic shape, where the majority of the significant coefficients are clustered near the origin and are located along the main and near adjacent diagonals of each response. These measured quadratic Volterra kernels are in accordance with other measured quadratic Volterra kernels

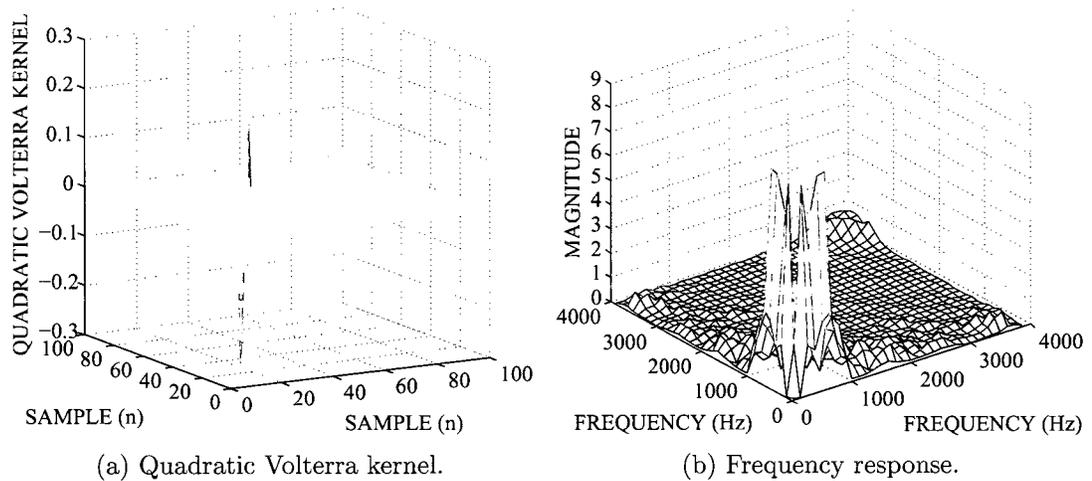


Figure 3.6: Measured quadratic Volterra kernel and corresponding frequency response for the conference phone configuration.

from LEMS setups with small loudspeakers [16], [23], [29], [89]. The corresponding quadratic frequency response from the conference phone configuration exists mainly at frequencies below approximately 1000 Hz. The quadratic frequency response from the smartphone configuration has the majority of its significant magnitude occurring at frequencies between roughly 1000 Hz and 2000 Hz and between 3000 Hz and 4000 Hz. The quadratic Volterra kernels obtained under the other signal level conditions resulted in the same set of time and frequency domain observations. In Chapter 5 a subband Volterra structure for AEC is presented that exploits these observations, by identifying only the important quadratic Volterra kernel taps along the main and adjacent diagonals in a subset of the total number of frequency bands.

The absolute difference between the measured quadratic Volterra kernels at low and high input signals levels for the conference phone LEMS configuration is shown in Figure 3.8(a). From this figure it is evident that as the input signal level changes, the quadratic Volterra kernel coefficients near the origin along the main and close adjacent diagonals experience the greatest magnitude changes. This is expected since these taps contain the greatest amount of energy originally. The corresponding difference

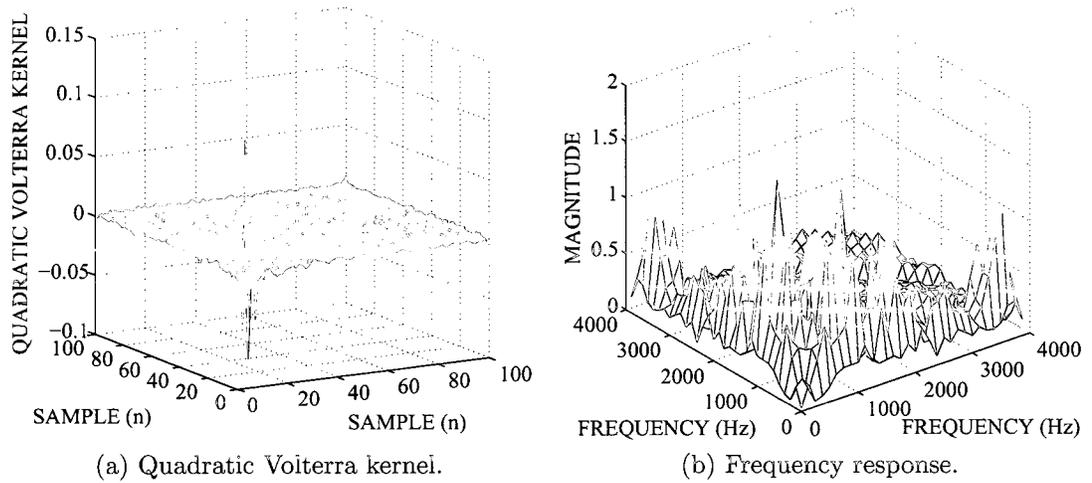


Figure 3.7: Measured quadratic Volterra kernel and corresponding frequency response for the smartphone configuration.

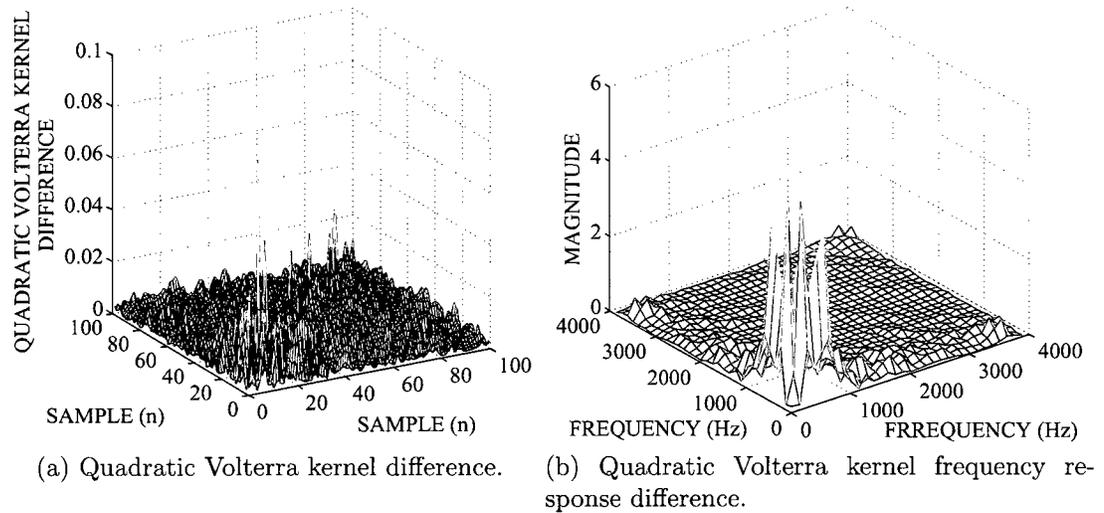


Figure 3.8: Measured quadratic Volterra kernel difference and corresponding frequency response from the conference phone at low and high volume signal levels.

in the quadratic frequency response for the conference phone LEMS configuration is shown in Figure 3.8(b). As the input signal level changes the associated change in the quadratic frequency response occurs mainly at frequencies below 1000 Hz, which is a result of the individual quadratic frequency responses being concentrated at these low frequencies. The same trend was observed for the absolute difference between

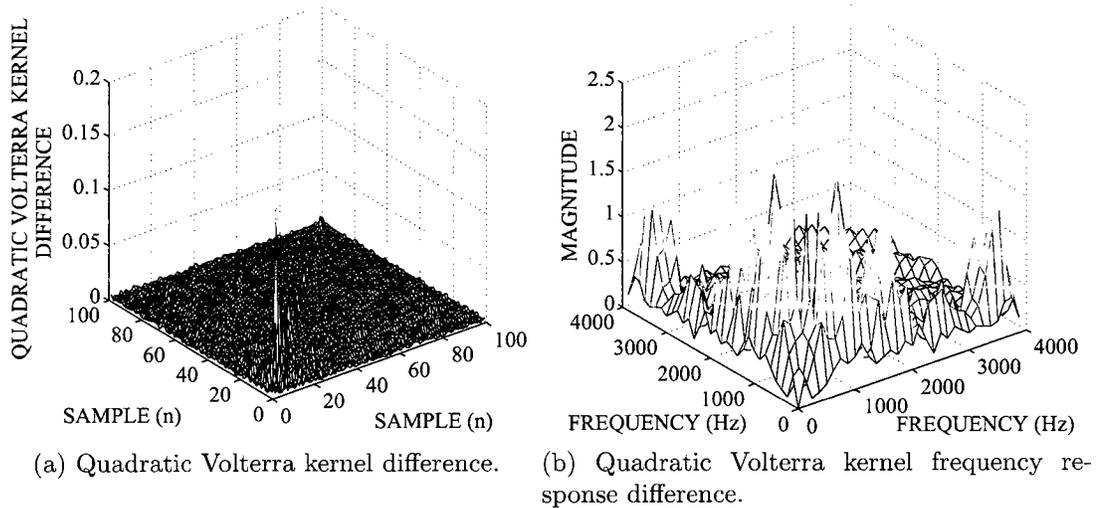


Figure 3.9: Measured quadratic Volterra kernel difference and corresponding frequency response from the smartphone at low and high volume signal levels.

the measured quadratic Volterra kernels at low and high input signals levels for the smartphone LEMS configuration, as shown in Figure 3.9(a). The corresponding difference in the quadratic frequency response for the smartphone device is shown in Figure 3.9(b) where the majority of its significant magnitude occurs at frequencies between roughly 1000 Hz and 2000 Hz and between 3000 Hz and 4000 Hz. This is a result of the individual quadratic frequency responses being concentrated in these frequency regions. A subband Volterra structure for AEC in nonstationary environments that takes advantage of these time and frequency domain observations is presented in Chapter 6.

3.4 Measured Harmonic Distortion

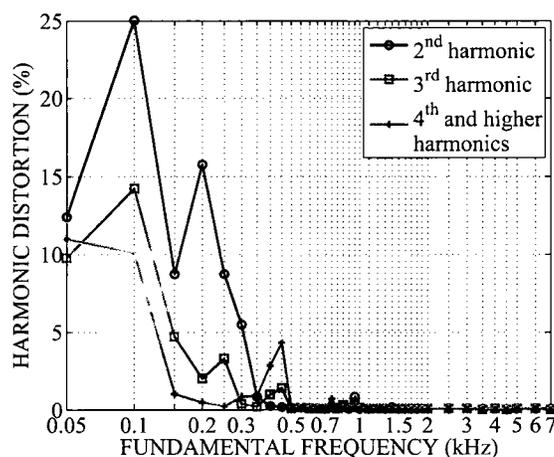
In this section a harmonic distortion analysis is presented for both the conference phone and the smartphone with both sets in hands-free mode. The distortion data was collected using the experimental setup shown in Figure 3.1. Microphone signals

were recorded, under maximum preamplifier gain, at a 48 kHz sampling rate for various test tones between 50 Hz and 7 kHz for the conference phone, and between 300 Hz and 7 kHz for the smartphone. Each test tone was applied to the conference phone loudspeaker at a RMS voltage of 1.4, 1.0, 0.8, and 0.5 Volts and to the smartphone loudspeaker at 1.1, 0.8, 0.5, and 0.2 Volts. For fundamental frequencies below 2.4 kHz ten harmonics were included in the distortion calculation, for higher fundamental frequencies all harmonics below the 24 kHz Nyquist rate were included in the distortion calculation. The total harmonic distortion (THD) for each fundamental frequency was computed as follows:

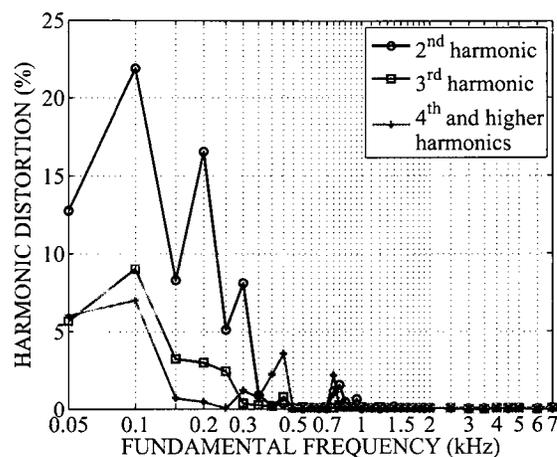
$$THD = \frac{\sum_{i=2}^n H_i}{H_1} \quad (3.1)$$

where H_i is the power in the i^{th} harmonic. The power of each harmonic was computed from the difference between the total power of the recorded microphone signal and the total power after notching out the harmonic in question. Signal powers were computed by integrating the corresponding Welch power spectral density estimates [75]. A 20th order Butterworth infinite impulse response (IIR) notch filter with a quality factor of 35 was designed [75] and used to remove the corresponding harmonic signal from the recorded microphone signal. The notch filter provided over 100 dB of attenuation for all notch frequencies. In Figures 3.10 and 3.11 the second, third, and fourth and higher harmonic distortion is presented where the THD can be computed by summing all three components for each fundamental frequency.

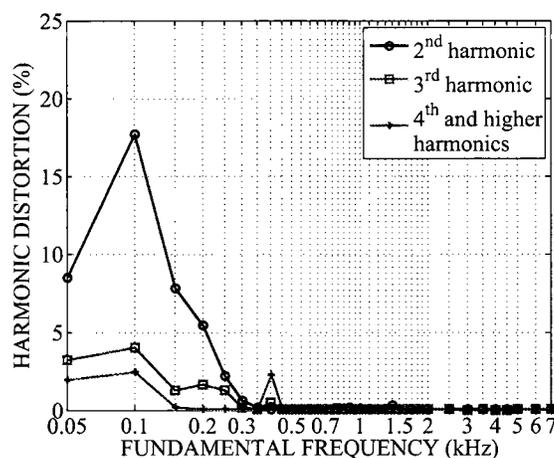
As shown in Figure 3.10(a) the conference phone exhibits significant second harmonic distortion of up to 25% for low fundamental frequencies between 50 and 300 Hz. At frequencies beyond this range the second harmonic distortion decreases rapidly to much smaller values. Significant third harmonic distortion of up to 15% occurs for frequencies between 50 and 150 Hz, before decaying towards zero at higher frequencies. Fourth and higher harmonic distortion of up to 10% occurs for fundamental frequencies of 50 and 100 Hz with very little occurring at higher frequencies. Similar



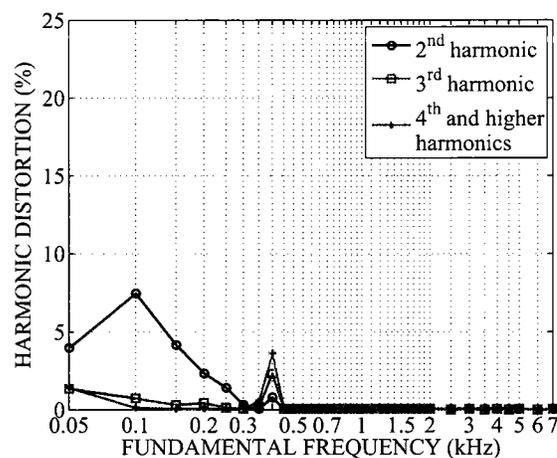
(a) Harmonic distortion of the conference phone at 1.4 Volts.



(b) Harmonic distortion of the conference phone at 1.0 Volt.



(c) Harmonic distortion of the conference phone at 0.8 Volts.

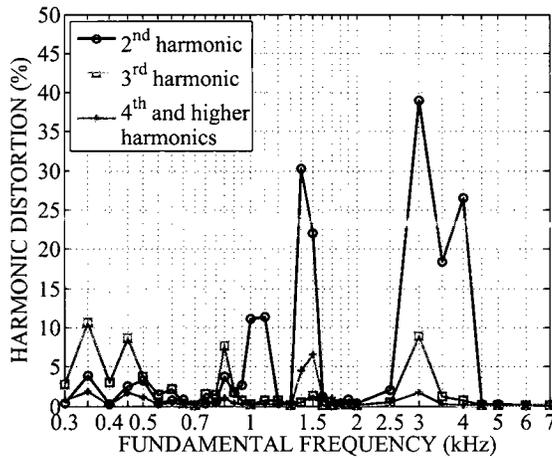


(d) Harmonic distortion of the conference phone at 0.5 Volts.

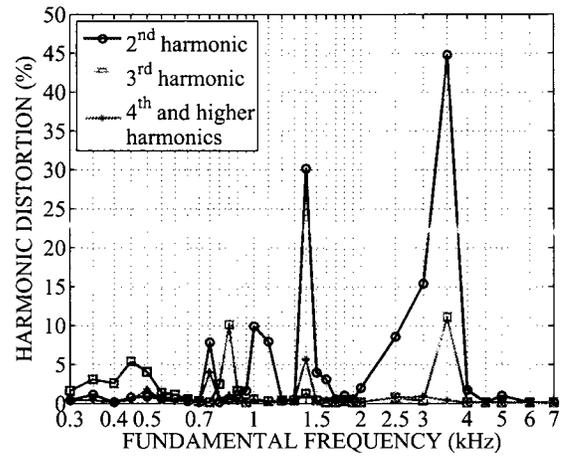
Figure 3.10: Measured harmonic distortion for the conference phone.

distortion results for the conference phone under high input signal level conditions are presented in [28]. As the signal level of the fundamental frequencies decreases so does the amount of harmonic distortion. However, significant second harmonic distortion of up to approximately 20% still occurs for fundamental frequencies between 50 and 300 Hz for input signal levels of 1.0 Volt as shown in Figure 3.10(b). At the same signal level a small amount of third and fourth and higher harmonic distortion of

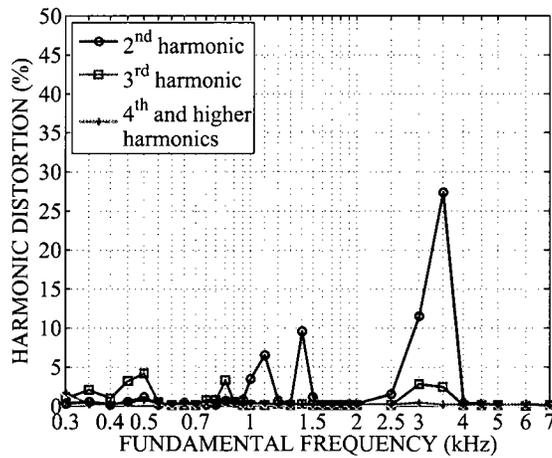
up to 10% occurs for fundamental frequencies of 50 and 100 Hz. As well, significant second harmonic distortion of up to 17% occurs for fundamental frequencies between 50 and 200 Hz for input signal levels of 0.8 Volts as shown in Figure 3.10(c). For 0.5 Volt fundamental test tones a small amount of second harmonic distortion of up to 7% occurs for fundamental frequencies between 50 and 200 Hz as shown in Figure 3.10(d).



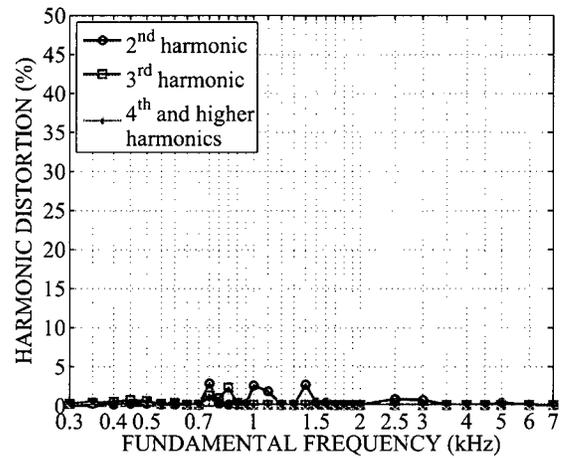
(a) Harmonic distortion of the smartphone at 1.1 Volts.



(b) Harmonic distortion of the smartphone at 0.8 Volts.



(c) Harmonic distortion of the smartphone at 0.5 Volts.



(d) Harmonic distortion of the smartphone at 0.2 Volts.

Figure 3.11: Measured harmonic distortion for the smartphone.

The smartphone reaches up to 10% third harmonic distortion for fundamental frequencies between 300 and 500 Hz, and for fundamental frequencies at 850 Hz and 3 kHz as shown in Figure 3.11(a). At other fundamental frequencies the third harmonic distortion is very small. Considerable second harmonic distortion of up to 30% occurs for fundamental frequencies near 1.5 kHz, and up to 40% between 3 and 4 kHz. Very little second harmonic distortion occurs at other fundamental frequencies with the exception of 11% at 1 and 1.1 kHz. Only a small amount of fourth and higher harmonic distortion of up to 6% is observed around 1.5 kHz. Similar distortion results for the smartphone under high input signal level conditions are presented in [28]. Again, as the signal level decreases the amount of THD decreases as well. However, significant second order distortion of 30% still occurs at 1.4 kHz along with 15% and 45% at 3 and 3.5 kHz respectively, as shown in Figure 3.11(b) for input signal levels of 0.8 Volts. Also, at the same signal level third harmonic distortion of 10% occurs at 850 Hz and 3.5 kHz, with very little fourth and higher harmonic distortion at all fundamental frequencies. With an input signal level of 0.5 Volts significant second harmonic distortion of 10%, 11%, and 27% occurs at 1.4, 3, and 3.5 kHz respectively as seen in Figure 3.11(c). At this signal level very little third and higher harmonic distortion exists at all fundamental frequencies. With an input signal level of 0.2 Volts very little harmonic distortion occurs for all fundamental frequencies as shown in Figure 3.11(d).

Based on these harmonic distortion results it is apparent that both the conference phone and smartphone experience high amounts of second harmonic distortion for fundamental frequencies in certain frequency regions. Also, at very high signal levels significant third and higher harmonic distortion occurs for the conference phone at low fundamental frequencies, and significant third harmonic distortion occurs for the smartphone set at various frequencies. It should also be noted that the frequency regions of significant second harmonic distortion, for both the conference phone and smartphone, correspond well with the significant magnitude regions of the measured

quadratic Volterra kernel frequency responses shown in Figure 3.6 and 3.7. Due to the diminished output of the miniaturized loudspeaker in the smartphone at lower frequencies the harmonic distortion is more significant at higher frequencies. The majority of harmonic distortion for the larger loudspeaker in the conference phone occurs at low frequencies as expected [78]. These distortion analysis observations provide valuable insight into achieving complexity reduction in the subband Volterra structures presented in Chapters 5 and 6. Also, it should be noted that for a speech input signal some of the harmonic distortion components will be masked by the speech signal itself, which may reduce the audibility of the distortion. Furthermore, with the transition to wideband telephony the impact of low frequency loudspeaker distortion on hands-free communication will become more problematic.

3.5 Summary

This chapter presented the experimental setup and equipment used for estimating the linear and nonlinear components of several hands-free systems. The time and frequency domain nature of the resulting components were investigated under various conditions in the hands-free environment. It was shown in Section 3.2 that in response to a small obstruction in the vicinity of the hands-free device the direct path, initial early reflection, and late reflection linear echo path impulse response coefficients remained relatively unchanged, while significant changes occurred to the remaining coefficients. Furthermore, the linear portion of the echo path was shown to vary more at high frequencies than at lower frequencies due to the small obstruction. In response to increased levels of excitation signal, the magnitude of the significant quadratic Volterra kernel coefficients increased the most, while the other insignificant coefficients experienced little change as shown in Section 3.3. Also, the nonlinear echo path components experienced the greatest variability in only certain frequency regions, in response to an increase in input signal level. A harmonic distortion analysis

of the hands-free devices in Section 3.4 revealed that significant second harmonic distortion occurred for specific fundamental frequency regions under high and moderate input signal levels.

Chapter 4

Subband Structure for Acoustic Echo Cancellation in Nonstationary Environments

As discussed in Section 2.2.2 of Chapter 2 one of the main limitations in the performance of an EC is caused by the inherent nonstationarities within the acoustic environment. Thus, an EC needs to be able to track the subsequent echo path changes quickly to maintain high quality communication between parties in the hands-free system. Proportionate variable step size adaptive algorithms have been introduced to improve the tracking performance of an EC, as discussed in Section 2.2.2.1. In this chapter a generalized proportionate variable step size subband adaptive filter structure is proposed, that exploits the time and frequency domain characteristics of the changes that occur within the acoustic environment. It is shown that this structure can achieve superior convergence and tracking performance for AEC compared to its fullband time domain counterpart, at a reduced computational complexity.

The following sections are presented in this chapter. In Section 4.1 the generalized proportionate variable step size subband adaptive filter structure is discussed in detail. This includes a description of the filter bank structure it employs along with the

underlying proportionate step size adaptive algorithm used in each subband, followed by a computational complexity analysis of the structure. Experimental results based on measured LEMS impulse responses, as well as on directly measured LEMS input signal data, are presented in Section 4.2 for the proposed structure. A summary of the main results and findings of this chapter are given in Section 4.3.

4.1 Generalized Proportionate Variable Step Size Subband Adaptive Filter Structure

4.1.1 Description

As shown in Section 3.2 of Chapter 3, the changes that occur in the linear impulse response of a LEMS due to a change in the acoustic environment are not uniform in both the time and frequency domains. The majority of the significant fluctuations in the impulse response tend to occur in the high magnitude taps that correspond to the early reflection section of the response. Furthermore, the impulse response variations are much more significant at higher frequencies than at lower ones. To exploit these trends in time and frequency during a change in the acoustic environment, a fast tracking subband structure for AEC is proposed. As discussed in Section 2.1.1.2 of Chapter 2, a subband structure uses shorter adaptive filters in multiple frequency bands operating at a reduced sampling rate, compared to a single long fullband adaptive filter. This allows for faster convergence and tracking at a reduced complexity as the subband signals tend to be more spectrally uniform, especially for highly correlated input signals such as speech. Furthermore, when a disturbance occurs in the hands-free environment only the filter taps in the correspondingly affected subbands need to be adjusted. In contrast, all of the taps of the equivalent fullband adaptive filter will need to be adjusted in response to a perturbation in the acoustic environment, resulting in diminished tracking performance.

To take advantage of the fact that the LEMS impulse response taps with the greatest magnitude experience the largest fluctuations due to a change in the hands-free environment, a subband AEC structure with adaptive filters based on the GGPAPA discussed in Section 2.2.2.1 of Chapter 2 is introduced. A block diagram of the subband generalized gradient proportionate affine projection (SB-GGPAP) AEC structure is shown in Figure 4.1. The subband structure is composed of near perfect reconstruction oversampled GDFT modulated analysis and synthesis filter banks with polyphase implementation, as discussed in Section 2.1.1.2 of Chapter 2. Oversampled GDFT modulated filter banks were employed instead of oversampled DCT modulated filter banks due to their aliasing robustness [62]. The K subband adaptive filters, $\hat{\mathbf{w}}_i(m)$, operate on the M ($M < K$) fold downsampled and filtered input reference, $x_i(m)$, and input microphone, $d_i(m)$, signals ($i = 1, \dots, K$) to produce estimates of the subband echo path impulse responses, $\mathbf{w}_i(m)$. The resulting subband error signals, $e_i(m)$, are passed through the polyphase synthesis filters to create the fullband error signal, $e(n)$. The complex GDFT modulation results in complex valued subband signals which increase the complexity of the subband adaptive filtering algorithm. However, this resulting complexity increase can be overcome for real valued input signals, which is the case for AEC, since the resulting subband signals will occur in complex conjugate pairs. Thus, only half of the subband signals need to be processed due to this redundancy. Since the subband adaptive filters are derived based on the GGPAPA, faster convergence and tracking performance can be achieved compared to equivalent subband structures based on the PAPA or the IPAPA. This is due to the individual filter tap step sizes being based on a gradient estimate given by the difference between the current and delayed average filter tap coefficients, which does not have any time invariant components that can slow tracking [12].

The subband input signal and adaptive filter vector definitions for the subband generalized gradient proportionate affine projection algorithm (SB-GGPAPA) are as

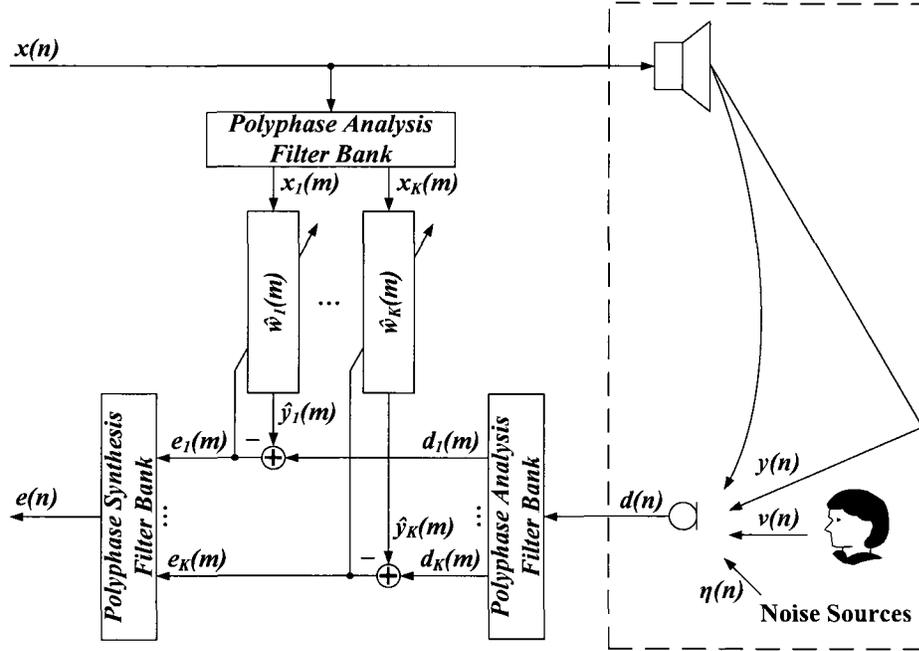


Figure 4.1: Block diagram of the generalized proportionate variable step size subband adaptive filter structure.

follows:

$$\hat{\mathbf{w}}_i(m) = [\hat{w}_{i_0}(m), \hat{w}_{i_1}(m), \dots, \hat{w}_{i_{N_{sb_i}-1}}(m)]^T \quad (4.1)$$

$$\mathbf{x}_i(m) = [x_i(m), x_i(m-1), \dots, x_i(m-N_{sb_i}+1)]^T \quad (4.2)$$

$$\mathbf{X}_i(m) = [\mathbf{x}_i(m), \mathbf{x}_i(m-1), \dots, \mathbf{x}_i(m-P+1)] \quad (4.3)$$

$$\mathbf{d}_i(m) = [d_i(m), d_i(m-1), \dots, d_i(m-P+1)] \quad (4.4)$$

$$\mathbf{e}_i(m) = \mathbf{d}_i(m) - \hat{\mathbf{w}}_i^H(m) \mathbf{X}_i(m) \quad (4.5)$$

where $\hat{\mathbf{w}}_i(m)$ are the subband filter coefficient vectors, $\mathbf{x}_i(m)$ are the subband reference input signal vectors of length N_{sb_i} , and P is the projection order. The subband adaptive filters are updated according to the GGPAPA, which is outlined in Section 2.2.2.1 of Chapter 2, as follows:

$$\hat{\mathbf{w}}_i(m+1) = \hat{\mathbf{w}}_i(m) + \mu_i \mathbf{A}_i(m) \mathbf{X}_i(m) \boldsymbol{\varepsilon}_i(m) \quad (4.6)$$

$$\boldsymbol{\varepsilon}_i(m) = \boldsymbol{\Gamma}_i(m)\mathbf{e}_i^H(m) \quad (4.7)$$

$$\boldsymbol{\Gamma}_i(m) = \left(\mathbf{X}_i^H(m)\mathbf{A}_i(m)\mathbf{X}_i(m) + \delta_i\mathbf{I} \right)^{-1} \quad (4.8)$$

$$\mathbf{A}_i(m) = \text{diag} \left[g_{i_0}(m), \dots, g_{i_{N_{sb_i}-1}}(m) \right] \quad (4.9)$$

$$g_{i_l}(m) = \frac{1 - \beta_i}{2} \frac{\bar{c}_{i_l}(m-1)}{\sum_{l=0}^{N_{sb_i}-1} \bar{c}_{i_l}(m-1) + \xi_i} + \frac{\beta_i}{2N_{sb_i}} \quad (4.10)$$

$$\bar{c}_{i_l}(m) = \epsilon_i \bar{c}_{i_l}(m-1) + (1 - \epsilon_i)|c_{i_l}(m)| \quad (4.11)$$

$$c_{i_l}(m) = \hat{w}_{i_l}(m-1) - \gamma_i \bar{w}_{i_l}(m-1) \quad (4.12)$$

$$\bar{w}_{i_l}(m) = \eta_i \bar{w}_{i_l}(m-1) + (1 - \eta_i)\hat{w}_{i_l}(m-1) \quad (4.13)$$

where ξ_i is a small positive number to avoid to avoid division by zero in (4.10), δ_i is a small positive regularization parameter, $l = 0, \dots, N_{sb_i} - 1$, and \mathbf{I} is a $P \times P$ identity matrix. The subband gradient estimates and their averages are represented by $c_{i_l}(m)$ and $\bar{c}_{i_l}(m)$ respectively. The average subband filter tap coefficient is denoted by $\bar{w}_{i_l}(m)$ where γ_i is a correction factor for $\bar{w}_{i_l}(m-1)$. The forgetting factors ϵ_i and η_i are used in computing $\bar{c}_{i_l}(m)$ and $\bar{w}_{i_l}(m)$ respectively. The overall subband step size parameter, μ_i , controls convergence, tracking speed, steady-state error [12]. The proportionate step size matrix, $\mathbf{A}_i(m)$, consists of the individual filter tap proportionate step sizes, $g_{i_l}(m)$, along its main diagonal. When ϵ_i is set to 0 in all subbands the SB-GGPAPA becomes a non-generalized version called subband gradient proportionate affine projection algorithm (SB-GPAPA) as discussed in [12] for the fullband GPAPA. Furthermore, when γ_i is set to 0 in all subbands the SB-GPAPA reduces to a subband improved proportionate affine projection algorithm (SB-IPAPA). In addition, setting β_i close to 0 in all subbands leads to behaviour like a subband proportionate affine projection algorithm (SB-PAPA), while setting $\beta_i = 1$ in all subbands and dividing δ_i by $2N_{sb_i}$ results in a subband affine projection algorithm (SB-APA). Moreover, under the same conditions with P set to 1 the SB-GPAPA reduces to a subband

gradient proportionate normalized least-mean-square (SB-GPNLMS) algorithm, the SB-IPAPA reduces to the subband improved proportionate normalized least-mean-square (SB-IPNLMS) algorithm given in [100], the SB-PAPA reduces to a subband proportionate normalized least-mean-square (SB-PNLMS) algorithm, and the SB-APA reduces to a subband normalized least-mean-square (SB-NLMS) algorithm. A subband version of the exponential step size algorithm of [97] can be derived from the SB-GGPAPA as ϵ_i approaches 1 in all subbands as in [12] for the fullband case.

Implementing the GGPAPA in a subband structure allows the fast tracking algorithm to be applied on a per subband basis depending on the nature of the nonstationarity that occurs within the AEC environment. As demonstrated in Section 3.2 of Chapter 3, the obstruction of a hand placed on the keypad of the phone set resulted in a change in the LEMS impulse response occurring mainly at frequencies above 1 kHz. Thus, the subband AEC structure of Figure 4.1 would only require the GGPAPA in the correspondingly affected subbands while the simpler APA would suffice in the unaffected subbands. This per subband tracking flexibility helps to minimize the increase in computational complexity required for implementing the subband adaptive filters with the GGPAPA compared to using the APA.

As shown experimentally in Figure 3.4 of Chapter 3, only the high magnitude linear echo path impulse response taps in the early reflection section of the response vary significantly due to the introduction of the hand obstruction. Thus, identifying and tracking the initial LEMS impulse response taps with the GGPAPA, and the remainder of the taps with the simpler APA will allow for a compromise between tracking performance during echo path fluctuations and overall computational complexity. With this in mind the subband adaptive filter update in (4.6) is replaced with the following:

$$\hat{\mathbf{w}}_{f_i}(m+1) = \hat{\mathbf{w}}_{f_i}(m) + \mu_{f_i} \mathbf{A}_{f_i}(m) \mathbf{X}_{f_i}(m) \varepsilon_{f_i}(m) \quad (4.14)$$

$$\boldsymbol{\varepsilon}_{f_i}(m) = \boldsymbol{\Gamma}_{f_i}(m) \mathbf{e}_i^H(m) \quad (4.15)$$

$$\boldsymbol{\Gamma}_{f_i}(m) = \left(\mathbf{X}_{f_i}^H(m) \mathbf{A}_{f_i}(m) \mathbf{X}_{f_i}(m) + \delta_{f_i} \mathbf{I} \right)^{-1} \quad (4.16)$$

$$\hat{\boldsymbol{w}}_{t_i}(m+1) = \hat{\boldsymbol{w}}_{t_i}(m) + \mu_{t_i} \mathbf{X}_{t_i}(m) \boldsymbol{\varepsilon}_{t_i}(m) \quad (4.17)$$

$$\boldsymbol{\varepsilon}_{t_i}(m) = \boldsymbol{\Gamma}_{t_i}(m) \mathbf{e}_i^H(m) \quad (4.18)$$

$$\boldsymbol{\Gamma}_{t_i}(m) = \left(\mathbf{X}_i^H(m) \mathbf{X}_i(m) + \delta_{t_i} \mathbf{I} \right)^{-1} \quad (4.19)$$

$$\hat{\boldsymbol{w}}_{f_i}(m) = \left[\hat{w}_{i_0}(m), \hat{w}_{i_1}(m), \dots, \hat{w}_{i_{N_{f_i}-1}}(m) \right]^T \quad (4.20)$$

$$\hat{\boldsymbol{w}}_{t_i}(m) = \left[\hat{w}_{i_{N_{f_i}}}(m), \hat{w}_{i_{N_{f_i}+1}}(m), \dots, \hat{w}_{i_{N_{sb_i}-1}}(m) \right]^T \quad (4.21)$$

$$\mathbf{X}_{f_i}(m) = \left[\mathbf{x}_{f_i}(m), \mathbf{x}_{f_i}(m-1), \dots, \mathbf{x}_{f_i}(m-P+1) \right] \quad (4.22)$$

$$\mathbf{x}_{f_i}(m) = \left[x_i(m), x_i(m-1), \dots, x_i(m-N_{f_i}+1) \right]^T \quad (4.23)$$

$$\mathbf{X}_{t_i}(m) = \left[\mathbf{x}_{t_i}(m), \mathbf{x}_{t_i}(m-1), \dots, \mathbf{x}_{t_i}(m-P+1) \right] \quad (4.24)$$

$$\mathbf{x}_{t_i}(m) = \left[x_i(m-N_{f_i}), x_i(m-N_{f_i}-1), \dots, x_i(m-N_{sb_i}+1) \right]^T \quad (4.25)$$

where N_{f_i} is the number of initial LEMS impulse response taps identified with the GGPAPA ($N_{f_i} < N_{sb_i}$) and the remaining $N_{sb_i} - N_{f_i}$ taps are identified with the APA. The equations in (4.9) to (4.13) can be applied with N_{sb_i} replaced with N_{f_i} for the initial LEMS impulse response taps adapted with the GGPAPA.

4.1.2 Computational Complexity Analysis

The computational complexity of the SB-GGPAP AEC structure shown in Figure 4.1 is comprised of the complexity of the filter banks along with the adaptive filtering expense within each subband. The computational cost of the AEC structures investigated in this thesis is expressed in terms of the total number of equivalent real elementary mathematical operations per input sample period. The cost of other operations such as memory transfers are not considered. All AEC structures are

assumed to be implemented on a general purpose floating point central processing unit (CPU). For complex input data one complex addition is realized with two real additions, and one complex multiplication is equivalent to four real multiplications with two real additions. Furthermore, a complex division is assumed to be computed by first multiplying the numerator and denominator by the conjugate of the denominator, resulting in six real multiplications, three real additions, and two real divisions overall. The absolute value of a complex number is assumed to be determined by multiplying by its conjugate and then taking the square root of the result. This requires two real multiplications and one real addition along with the complexity of computing the square root of a real valued number.

For the oversampled GDFT modulated polyphase analysis and synthesis filter banks used in the subband AEC structure of Figure 4.1 the computational cost, C_{FB} , is comprised of the same number of real multiplications and additions per input sample for each filter bank as follows [63]:

$$C_{FB} = \frac{1}{M} (L_{pr} + 4K \log_2 K + 4K) \quad (4.26)$$

where L_{pr} is the length of the filter bank prototype filter. The adaptive filtering cost is determined by the computational expense of adapting a subband filter at the reduced sampling rate times the number of subbands processed. As the complex GDFT modulated filter banks give rise to complex valued subband signals, which occur in complex conjugate pairs for real valued input signals, only up to $K/2$ subband signals need to be processed. The complexity of the GGPAPA described in (4.5) to (4.13) is summarized for real and complex input signal data in Tables 4.1 and 4.2 respectively. The terms $T_{R,m}$, $T_{R,a}$, and $T_{R,d}$ represent the required number of real multiplications, additions, and divisions for solving the system of equations in (4.7) for real input data, while $T_{C,m}$, $T_{C,a}$, and $T_{C,d}$ represent the required number for complex input data. If standard Gaussian elimination is used then $T_{R,m}$, $T_{R,a}$, $T_{C,m}$, and $T_{C,a}$ have

a complexity of $O(P^3)$ while $T_{R,d}$ and $T_{C,d}$ have a complexity of $O(P^2)$. The cost of computing a square root when determining the absolute value of a complex valued filter tap gradient estimate in (4.11) is denoted by $C_{\sqrt{\cdot},m}$, $C_{\sqrt{\cdot},a}$, and $C_{\sqrt{\cdot},d}$ for the required number of real multiplications, additions, and divisions respectively. Also, since the resulting $P \times P$ matrix from computing $\mathbf{X}_i^H(m)\mathbf{A}_i(m)\mathbf{X}_i(m)$ is a Hermitian matrix the required number of real multiplications and additions for real input data can be reduced from $P^2N_{sb_i}$ and $P^2(N_{sb_i} - 1)$ to $\frac{1}{2}(P^2 + P)N_{sb_i}$ and $\frac{1}{2}(P^2 + P)(N_{sb_i} - 1)$ respectively, given that the intermediate $N_{sb_i} \times P$ matrix resulting from $\mathbf{A}_i(m)\mathbf{X}_i(m)$ has already been determined. For real input signals, which is the case for AEC, the overall computational cost of the subband AEC structure with the GGPAPA used to adapt the $K/2$ unique complex subband filters operating at the M fold reduced sampling rate, is summarized in Table 4.3 in terms of the required number of real mathematical operations. Overall the complexity of the structure is dominated by the required number of real multiplication and addition operations, and is $O\left(P^2N_{sb_i}\frac{K}{M}\right)$ for both as P increases for equal N_{sb_i} in all subbands.

Term	\times	$+$	\div
$\hat{\mathbf{w}}_i^H(m)\mathbf{X}_i(m)$	PN_{sb_i}	$P(N_{sb_i} - 1)$	
$\mathbf{d}_i(m) - \hat{\mathbf{w}}_i^H(m)\mathbf{X}_i(m)$		P	
$\mathbf{A}_i(m)\mathbf{X}_i(m)$	PN_{sb_i}		
$\mathbf{X}_i^H(m)\mathbf{A}_i(m)\mathbf{X}_i(m)$	$\frac{1}{2}(P^2 + P)N_{sb_i}$	$\frac{1}{2}(P^2 + P)(N_{sb_i} - 1)$	
$\mathbf{X}_i^H(m)\mathbf{A}_i(m)\mathbf{X}_i(m) + \delta_i\mathbf{I}$		P	
$\boldsymbol{\varepsilon}_i(m) = \boldsymbol{\Gamma}_i(m)\mathbf{e}_i^H(m)$	$T_{R,m}$	$T_{R,a}$	$T_{R,d}$
$\mu_i\mathbf{A}_i(m)\mathbf{X}_i(m)\boldsymbol{\varepsilon}_i(m)$	$P(N_{sb_i} + 1)$	$(P - 1)N_{sb_i}$	
$\hat{\mathbf{w}}_i(m) + \mu_i\mathbf{A}_i(m)\mathbf{X}_i(m)\boldsymbol{\varepsilon}_i(m)$		N_{sb_i}	

Term	\times	$+$	\div
$s_i(m) =$ $\sum_{l=0}^{N_{sb_i}-1} \bar{c}_i(m-1) + \xi_i$		N_{sb_i}	
$((1 - \beta_i)/2)/s_i(m) \times$ $\bar{c}_i(m-1) + \beta_i/(2N_{sb_i})$	N_{sb_i}	N_{sb_i}	1
$\epsilon_i \bar{c}_i(m-1) +$ $(1 - \epsilon_i) c_i(m) $	$2N_{sb_i}$	N_{sb_i}	
$\hat{w}_i(m-1) -$ $\gamma_i \bar{w}_i(m-1)$	N_{sb_i}	N_{sb_i}	
$\eta_i \bar{w}_i(m-1) +$ $(1 - \eta_i)\hat{w}_i(m-1)$	$2N_{sb_i}$	N_{sb_i}	
Total per iteration	$\left(\frac{1}{2}P^2 + \frac{7}{2}P + 6\right)$ $\times N_{sb_i} + P + T_{R,m}$	$\left(\frac{1}{2}P^2 + \frac{5}{2}P + 5\right)$ $\times N_{sb_i} - \frac{1}{2}P^2 + \frac{1}{2}P$ $+ T_{R,a}$	$T_{R,d} + 1$

Table 4.1: Approximate computational complexity per iteration of the GGPAPA applied to subband adaptive filters for real input data in terms of real multiplications, additions, and divisions.

Term	\times	$+$	\div
$\hat{w}_i^H(m)\mathbf{X}_i(m)$	$4PN_{sb_i}$	$2P(2N_{sb_i} - 1)$	
$\mathbf{d}_i(m) - \hat{w}_i^H(m)\mathbf{X}_i(m)$		$2P$	
$\mathbf{A}_i(m)\mathbf{X}_i(m)$	$2PN_{sb_i}$		
$\mathbf{X}_i^H(m)\mathbf{A}_i(m)\mathbf{X}_i(m)$	$2(P^2 + P)N_{sb_i}$	$(P^2 + P)(2N_{sb_i} - 1)$	
$\mathbf{X}_i^H(m)\mathbf{A}_i(m)\mathbf{X}_i(m)$ $+ \delta_i \mathbf{I}$		P	
$\boldsymbol{\epsilon}_i(m) = \boldsymbol{\Gamma}_i(m)\mathbf{e}_i^H(m)$	$T_{C,m}$	$T_{C,a}$	$T_{C,d}$
$\mu_i \mathbf{A}_i(m)\mathbf{X}_i(m)\boldsymbol{\epsilon}_i(m)$	$P(4N_{sb_i} + 2)$	$(2P - 1)2N_{sb_i}$	

Term	×	+	÷
$\hat{w}_i(m) +$ $\mu_i \mathbf{A}_i(m) \mathbf{X}_i(m) \boldsymbol{\varepsilon}_i(m)$		$2N_{sb_i}$	
$s_i(m) =$ $\sum_{l=0}^{N_{sb_i}-1} \bar{c}_i(m-1) + \xi_i$		N_{sb_i}	
$((1 - \beta_i)/2)/s_i(m) \times$ $\bar{c}_i(m-1) + \beta_i/(2N_{sb_i})$	N_{sb_i}	N_{sb_i}	1
$\epsilon_i \bar{c}_i(m-1) +$ $(1 - \epsilon_i) c_i(m) $	$4N_{sb_i} +$ $N_{sb_i} C_{\sqrt{\cdot}, m}$	$2N_{sb_i} +$ $N_{sb_i} C_{\sqrt{\cdot}, a}$	$N_{sb_i} C_{\sqrt{\cdot}, d}$
$\hat{w}_i(m-1) -$ $\gamma_i \bar{w}_i(m-1)$	$2N_{sb_i}$	$2N_{sb_i}$	
$\eta_i \bar{w}_i(m-1) +$ $(1 - \eta_i) \hat{w}_i(m-1)$	$4N_{sb_i}$	$2N_{sb_i}$	
Total per iteration	$(2P^2 + 12P + 11)$ $\times N_{sb_i} + 2P + T_{C,m}$ $+ N_{sb_i} C_{\sqrt{\cdot}, m}$	$(2P^2 + 10P + 8)$ $\times N_{sb_i} - P^2 + T_{C,a}$ $+ N_{sb_i} C_{\sqrt{\cdot}, a}$	$T_{C,d} + 1 +$ $N_{sb_i} C_{\sqrt{\cdot}, d}$

Table 4.2: Approximate computational complexity per iteration of the GGPAPA applied to subband adaptive filters for complex input data in terms of real multiplications, additions, and divisions.

×	$\frac{3}{M} (L_{pr} + 4K \log_2 K + 4K) +$ $\frac{1}{M} \sum_{i=1}^{\frac{K}{2}} \left((2P^2 + 12P + 11) N_{sb_i} + 2P + T_{C,m} + N_{sb_i} C_{\sqrt{\cdot}, m} \right)$
+	$\frac{3}{M} (L_{pr} + 4K \log_2 K + 4K) +$ $\frac{1}{M} \sum_{i=1}^{\frac{K}{2}} \left((2P^2 + 10P + 8) N_{sb_i} - P^2 + T_{C,a} + N_{sb_i} C_{\sqrt{\cdot}, a} \right)$
÷	$\frac{1}{M} \sum_{i=1}^{\frac{K}{2}} (T_{C,d} + 1 + N_{sb_i} C_{\sqrt{\cdot}, d})$

Table 4.3: Approximate computational complexity per input sample period for the SB-GGPAP AEC structure in terms of real multiplications, additions, and divisions.

The computational complexity of the GGPAPA applied to the initial N_{f_i} subband adaptive filter taps and the APA applied to the remaining taps, as described in (4.14) to (4.25), is outlined for complex input signal data in Table 4.4. The total computational cost of the SB-GGPAP AEC structure with the GGPAPA applied to the initial N_{f_i} subband adaptive filter taps and the APA applied to the remaining taps is summarized in Table 4.5. The same terms and conditions as stated previously for Tables 4.2 and 4.3 are applicable to Tables 4.4 and 4.5. As N_{f_i} increases towards N_{sb_i} the complexity of the SB-GGPAP AEC structure increases towards the standard SB-GGPAP AEC structure given in Table 4.3. For small values of N_{f_i} the complexity of the SB-GGPAP AEC structure is similar to the simpler subband affine projection (SB-AP) AEC structure.

Term	×	+	÷
$\hat{\mathbf{w}}_i^H(m)\mathbf{X}_i(m)$	$4PN_{sb_i}$	$2P(2N_{sb_i} - 1)$	
$\mathbf{d}_i(m) - \hat{\mathbf{w}}_i^H(m)\mathbf{X}_i(m)$		$2P$	
$\mathbf{A}_{f_i}(m)\mathbf{X}_{f_i}(m)$	$2PN_{f_i}$		
$\mathbf{X}_{f_i}^H(m)\mathbf{A}_{f_i}(m)\mathbf{X}_{f_i}(m)$	$2(P^2 + P)N_{f_i}$	$(P^2 + P)(2N_{f_i} - 1)$	
$\mathbf{X}_{f_i}^H(m)\mathbf{A}_{f_i}(m)\mathbf{X}_{f_i}(m) + \delta_{f_i}\mathbf{I}$		P	
$\mathbf{X}_i^H(m)\mathbf{X}_i(m)$	$2(P^2 + P)N_{sb_i}$	$(P^2 + P)(2N_{sb_i} - 1)$	
$\mathbf{X}_i^H(m)\mathbf{X}_i(m) + \delta_{t_i}\mathbf{I}$		P	
$\boldsymbol{\varepsilon}_{f_i}(m) = \boldsymbol{\Gamma}_{f_i}(m)\mathbf{e}_i^H(m)$	$T_{C_f,m}$	$T_{C_f,a}$	$T_{C_f,d}$
$\boldsymbol{\varepsilon}_{t_i}(m) = \boldsymbol{\Gamma}_{t_i}(m)\mathbf{e}_i^H(m)$	$T_{C_t,m}$	$T_{C_t,a}$	$T_{C_t,d}$
$\mu_{f_i}\mathbf{A}_{f_i}(m)\mathbf{X}_{f_i}(m)\boldsymbol{\varepsilon}_{f_i}(m)$	$P(4N_{f_i} + 2)$	$(2P - 1)2N_{f_i}$	
$\mu_{t_i}\mathbf{X}_{t_i}(m)\boldsymbol{\varepsilon}_{t_i}(m)$	$P \times$ $(4(N_{sb_i} - N_{f_i}) + 2)$	$2(2P - 1) \times$ $(N_{sb_i} - N_{f_i})$	
$\hat{\mathbf{w}}_{f_i}(m) +$ $\mu_{f_i}\mathbf{A}_{f_i}(m)\mathbf{X}_{f_i}(m)\boldsymbol{\varepsilon}_{f_i}(m)$		$2N_{f_i}$	

Term	×	+	÷
$\hat{\mathbf{w}}_{t_i}(m) +$ $\mu_{t_i} \mathbf{X}_{t_i}(m) \boldsymbol{\varepsilon}_{t_i}(m)$		$2(N_{sb_i} - N_{f_i})$	
$s_{f_i}(m) =$ $\sum_{l=0}^{N_{f_i}-1} \bar{c}_{f_{i_l}}(m-1) + \xi_{f_i}$		N_{f_i}	
$((1 - \beta_{f_i})/2)/s_{f_i}(m) \times$ $\bar{c}_{f_{i_l}}(m-1) + \beta_{f_i}/(2N_{f_i})$	N_{f_i}	N_{f_i}	1
$\epsilon_{f_i} \bar{c}_{f_{i_l}}(m-1) +$ $(1 - \epsilon_{f_i}) c_{f_{i_l}}(m) $	$4N_{f_i} +$ $N_{f_i} C_{\sqrt{\cdot}, m}$	$2N_{f_i} +$ $N_{f_i} C_{\sqrt{\cdot}, a}$	$N_{f_i} C_{\sqrt{\cdot}, d}$
$\hat{w}_{f_{i_l}}(m-1) -$ $\gamma_{f_i} \bar{w}_{f_{i_l}}(m-1)$	$2N_{f_i}$	$2N_{f_i}$	
$\eta_{f_i} \bar{w}_{f_{i_l}}(m-1) +$ $(1 - \eta_{f_i}) \hat{w}_{f_{i_l}}(m-1)$	$4N_{f_i}$	$2N_{f_i}$	
Total per iteration	$(2P^2 + 10P)N_{sb_i} +$ $(2P^2 + 4P + 11)N_{f_i}$ $+ 4P + T_{C_{f,m}} +$ $T_{C_{t,m}} + N_{f_i} C_{\sqrt{\cdot}, m}$	$(2P^2 + 10P)N_{sb_i} +$ $(2P^2 + 2P + 8)N_{f_i}$ $- 2P^2 + T_{C_{f,a}} +$ $T_{C_{t,a}} + N_{f_i} C_{\sqrt{\cdot}, a}$	$T_{C_{f,d}}$ $+ T_{C_{t,d}}$ $+ 1$ $+ N_{f_i} \times$ $C_{\sqrt{\cdot}, d}$

Table 4.4: Approximate computational complexity per iteration of the GGPAPA applied to the initial N_{f_i} subband adaptive filter taps and the APA applied to the remaining taps for complex input data in terms of real multiplications, additions, and divisions.

\times	$\frac{3}{M} (L_{pr} + 4K \log_2 K + 4K) +$ $\frac{1}{M} \sum_{i=1}^{\frac{K}{2}} \left((2P^2 + 10P)N_{sb_i} + (2P^2 + 4P + 11)N_{f_i} + 4P + T_{C_f,m} + \right.$ $\left. T_{C_t,m} + N_{f_i} C_{\sqrt{\cdot},m} \right)$
$+$	$\frac{3}{M} (L_{pr} + 4K \log_2 K + 4K) +$ $\frac{1}{M} \sum_{i=1}^{\frac{K}{2}} \left((2P^2 + 10P)N_{sb_i} + (2P^2 + 2P + 8)N_{f_i} - 2P^2 + T_{C_f,a} + \right.$ $\left. T_{C_t,a} + N_{f_i} C_{\sqrt{\cdot},a} \right)$
\div	$\frac{1}{M} \sum_{i=1}^{\frac{K}{2}} (T_{C_f,d} + T_{C_t,d} + 1 + N_{f_i} C_{\sqrt{\cdot},d})$

Table 4.5: Approximate computational complexity per input sample period for the SB-GGPAP AEC structure with the GGPAPA applied to the initial N_{f_i} subband adaptive filter taps and the APA applied to the remaining taps in terms of real multiplications, additions, and divisions.

4.2 Experimental Results

4.2.1 Methodology

To perform meaningful AEC simulations under changing acoustic conditions, synthetic microphone input signals were created using (2.1) and (2.3) based on the measured unobstructed and obstructed LEMS impulse responses from the conference phone configuration, as shown in Section 3.2 of Chapter 3. This allows the performance of the AEC structures in question to be studied during a change in the acoustic environment under controlled input conditions. The change in the echo signal was created by linearly interpolating from the unobstructed impulse response coefficients to the obstructed impulse response coefficients, or vice versa, for a desired duration. This simulates a local talker's hand moving towards or away from the phone set and the impact it has on the performance of the AEC structure under test. Also, simulation results are presented based on experimentally measured input reference and microphone data from the conference phone. This allows the behaviour of the AEC structures to be investigated under real hands-free communication operating

conditions. The results for the proposed SB-GGPAPA AEC structure are compared to the fullband generalized gradient proportionate affine projection algorithm (FB-GGPAPA) and its variants outlined in [12], and to the SB-IPNLMS structure described in [100] in terms of ERLE performance and computational complexity. The simulation results for all AEC structures are based on an 8 kHz sampling rate to match that used in standard narrow band telephony.

Unless otherwise stated, the results presented in the following sections for the subband AEC structures were obtained using $K = 8$ band near perfect reconstruction oversampled GDFT modulated filter banks with a decimation factor of $M = 4$, as described in Section 4.1.1. This oversampling ratio (OSR) of 2 ($\text{OSR} = K/M$) was selected as it allowed for a low overall filter bank system delay, a high level of aliasing suppression within the subbands, and significant computational complexity savings compared to equivalent fullband structures. The prototype filter was a $L_{pr} = 64$ tap lowpass filter resulting in approximately 100 dB of stopband attenuation and was designed using the least squares method discussed in [61]. The high stopband attenuation of the analysis filters resulted in low aliasing levels within each subband. A 92 dB signal-to-noise ratio (SNR) due to aliasing was determined based on a WGN input. With a prototype filter of length $L_{pr} = 64$ the resulting delay introduced into the signal path by the analysis and synthesis filter banks, as discussed in Section 2.1.1.2 of Chapter 2, was 63 samples. Also, anti-causal subband LEMS impulse response taps were modeled by inserting a small delay [1] of 16 samples into the fullband microphone signal. This delay of 16 samples was chosen as it allowed for equal steady-state performance compared to the equivalent fullband AEC structures. The delay for implementing the anti-causal subband impulse response taps increased the overall delay of the subband AEC structures to 79 samples, or roughly 10 milliseconds for an 8 kHz sampling rate.

4.2.2 Simulation Results Based on Measured Impulse Responses

To demonstrate how the LEMS impulse response taps vary in time and frequency for the conference phone configuration, subband echo path impulse responses, $\mathbf{w}_i(m)$, were identified under the unobstructed and obstructed conditions outlined in Section 3.2 of Chapter 3. The recorded reference and microphone signals were decimated to an 8 kHz sampling rate before being processed by the filter bank described in Section 4.2.1. Within each subband the NLMS algorithm was used to estimate $\mathbf{w}_i(m)$, where each $\hat{\mathbf{w}}_i(m)$ was set to a length of $N_{sb_i} = 500$ to correspond to the fullband impulse responses of length $MN_{sb_i} = 2000$, as shown in Section 3.2 of Chapter 3. During the course of the adaptation process a small step size of 0.1 was used within each subband to ensure a low steady-state modeling error. The first 100 taps of the real part of each $\hat{\mathbf{w}}_i(m)$ for the unobstructed and obstructed conference phone configurations are shown in Figure 4.2. Only the 4 ($K/2$) $\hat{\mathbf{w}}_i(m)$ corresponding to the subbands covering frequencies up to half of the sampling rate are shown, as their complex conjugate partners are redundant as discussed in Section 4.1.1.

The unobstructed and obstructed subband impulse responses are nearly identical in the 0–1 kHz subband (Figure 4.2(a)), since the long wavelengths of sound at these low frequencies are less impeded by the small hand obstruction than the shorter sound wavelengths at higher frequencies. This is verified by the greater disparity between the unobstructed and obstructed subband impulse responses for the higher frequency subbands shown in Figures 4.2(b) to (d). Also, the greatest change in magnitude between the unobstructed and obstructed cases occurs for the high magnitude coefficients during the initial part of the subband impulse responses. These observations correspond well to those made for the equivalent fullband impulse responses shown in Figure 3.4 and Figure 3.5 in Section 3.2 of Chapter 3. Also, it should be noted that similar observations resulted for the imaginary parts of the complex subband

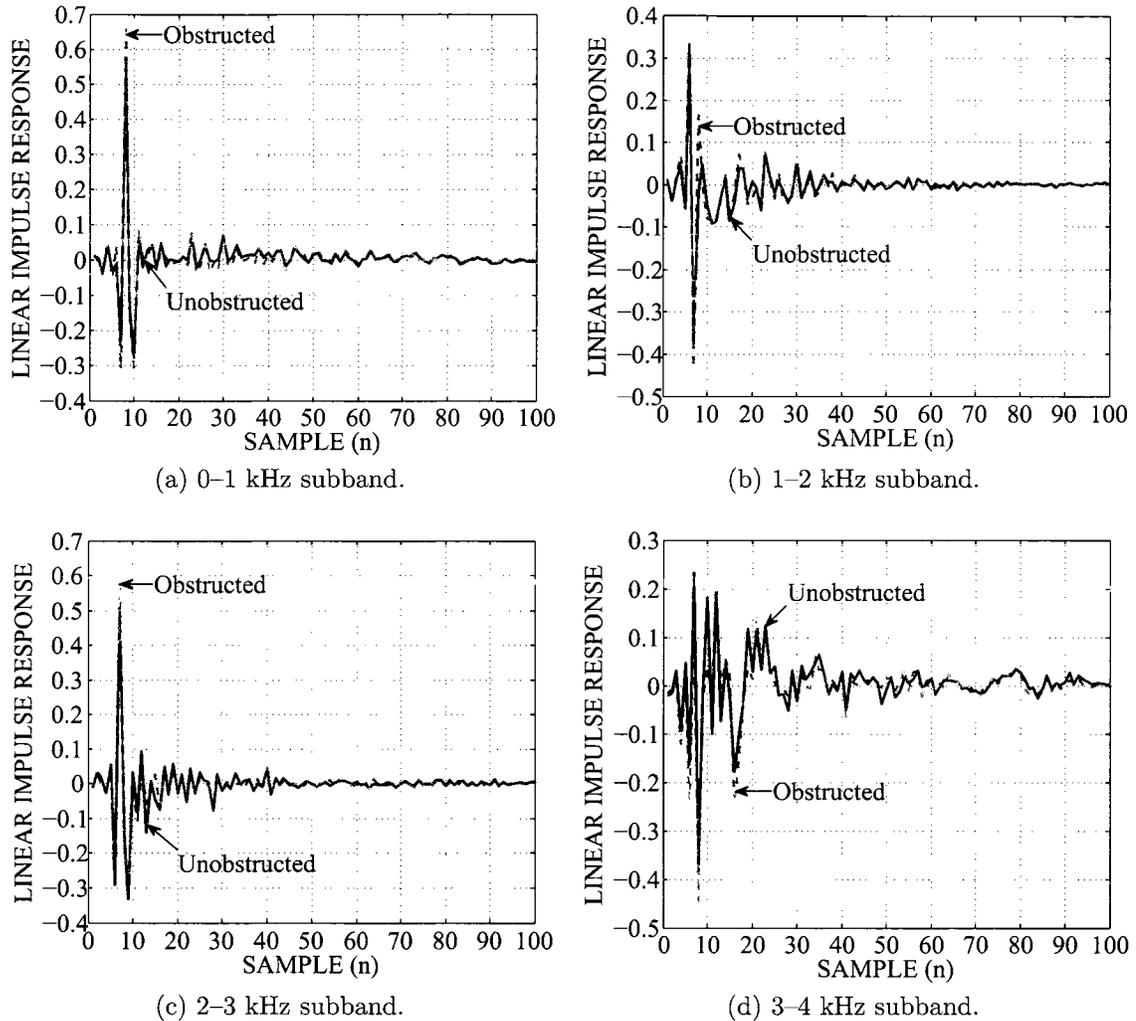


Figure 4.2: Real part of the unobstructed and obstructed complex subband impulse response estimates from the conference phone configuration.

impulse responses. In addition to the inherently faster tracking provided by a subband AEC structure compared to an equivalent fullband AEC structure, the fact that the LEMS impulse response fluctuations vary in frequency further motivates its use in a changing acoustic environment. In this case the adaptive filter in the 0–1 kHz subband would be impacted very little by the change in the acoustic environment and therefore maintain good AEC performance. The adaptive filters in the higher frequency subbands would be disrupted more by the echo path change, however the

degradation in AEC performance would be less compared the equivalent fullband AEC structure due to the faster reconvergence of the shorter subband adaptive filters, this was similarly noted in [100]. Furthermore, a subband AEC structure allows for fast tracking algorithms, such as the GGPAPA, to be applied in the subbands most affected by the disruption in the acoustic environment, while simpler slower tracking adaptive algorithms, such as the APA, can be used in the less impacted subbands. The application of fast tracking adaptive algorithms on a per subband basis can lead to improved overall AEC performance during changes in the acoustic environment, while helping to minimize the increase in complexity compared to using less complex slower tracking adaptive algorithms in all subbands.

4.2.2.1 White Noise Excitation Signal

In this section the AEC performance of the SB-GGPAP structure, discussed in Section 4.1.1, along with its variants are compared to each other as well as to their fullband counterparts under changing acoustic conditions. Simulations were performed for synthetic microphone input signals created based on the measured $L = 2000$ tap unobstructed and obstructed LEMS impulse responses as discussed in Section 4.2.1, with the results averaged over several trials. A 20 second WGN reference excitation signal, $x(n)$, with zero mean and unit variance was used along with the measured impulse responses to create a changing echo signal, $y(n)$, using (2.3). The change in $y(n)$ was achieved by linearly interpolating from the unobstructed impulse response coefficients to the obstructed coefficients over a desired duration, and then interpolating back over a desired time to the original unobstructed taps after a specified amount of time had passed. This change in $y(n)$ simulates a local talker moving their hand onto the keypad of the phone set, leaving it there for some time, and then moving their hand back to its original position. The input microphone signal, $d(n)$, was obtained by adding 20 seconds of uncorrelated WGN background noise, $\eta(n)$, to $y(n)$ to achieve a SNR of 30 dB. Since adaptation of the EC is only performed under

quiet local talker conditions, see Section 2.1.1 of Chapter 2, the local talker signal, $\nu(n)$, was set to 0 for the duration of the simulations.

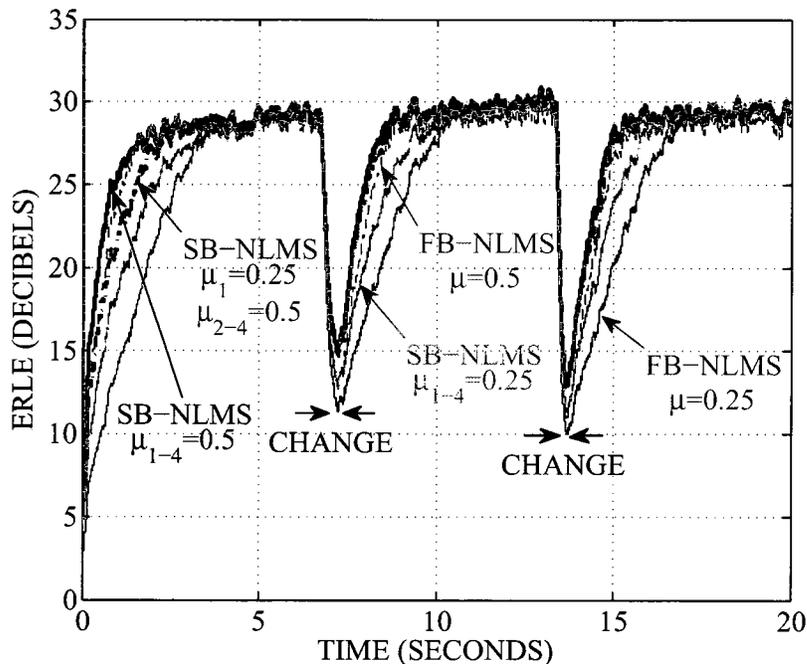


Figure 4.3: Subband NLMS versus fullband NLMS tracking.

Figure 4.3 compares the tracking performance of a SB-NLMS AEC structure to a fullband normalized least-mean-square (FB-NLMS) structure to demonstrate the inherently better tracking performance provided the subband structure. The changes in $y(n)$ took place at approximately 7 and 13 seconds into the simulation and lasted for 0.5 and 0.25 seconds respectively. The adaptive filter of the FB-NLMS structure was set to a length of $N = 2000$, while the adaptive filters of the SB-NLMS structure were set to $N_{sb_i} = N/M = 500$ in all subbands. The SB-NLMS structure was implemented using the filter bank characteristics discussed in Section 4.2.1 where the NLMS algorithm arose from the GGPAPA by setting $\epsilon_i = 0$, $\gamma_i = 0$, $\beta_i = 1$, and $P = 1$ as discussed in Section 4.1.1. With a step size of $\mu_i = 0.25$ used by all subband adaptive filters the SB-NLMS structure (SB-NLMS $\mu_{1-4} = 0.25$) suffers less of a performance degradation, and reconverges faster back to steady-state operation

while maintaining a higher ERLE than the equivalent FB-NLMS structure (FB-NLMS $\mu = 0.25$) for both echo path changes. The same trends occur for the SB-NLMS structure with an increased step size of $\mu_i = 0.5$ (SB-NLMS $\mu_{1-4} = 0.5$) compared to the FB-NLMS structure with the same step size (FB-NLMS $\mu = 0.5$). Also, at the higher step size both the fullband and subband structures outperform themselves in terms of convergence and tracking performance compared to the corresponding results at the lower step size. Moreover, the SB-NLMS structure with a step size of $\mu_i = 0.25$ in the 0–1 kHz subband and a step size of $\mu_i = 0.5$ in the 1–4 kHz subbands (SB-NLMS $\mu_1 = 0.25, \mu_{2-4} = 0.5$) attains nearly identical ERLE performance during the echo path changes as with a step size of $\mu_i = 0.5$ in all subbands (SB-NLMS $\mu_{1-4} = 0.5$). This verifies that faster tracking within the 0–1 kHz subband, via an increased step size, does not improve performance during the echo path changes, as the subband echo path impulse response taps in this band change very little as discussed in Section 4.2.2.

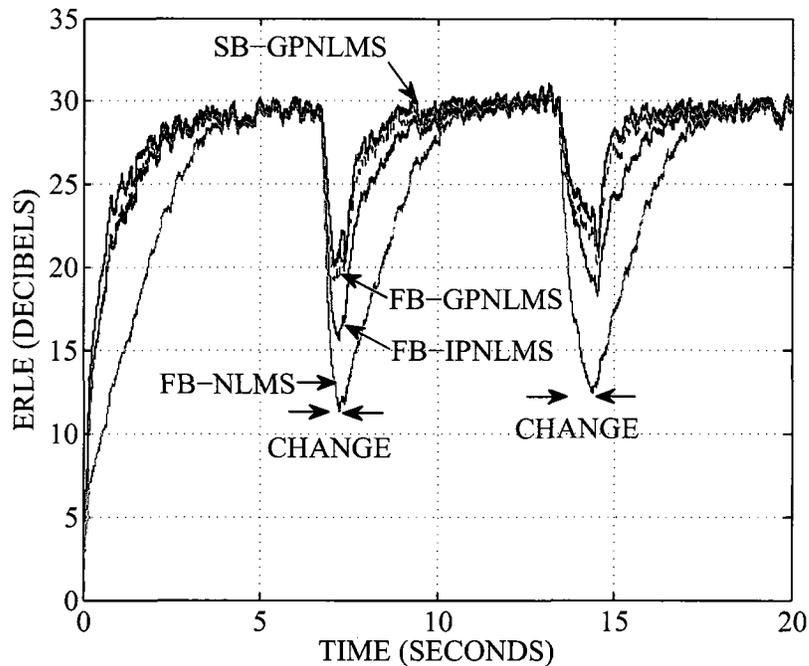


Figure 4.4: ERLE for FB-NLMS, FB-IPNLMS, FB-GPNLMS, and SB-GPNLMS with WGN input in a synthetic changing environment and a step size of 0.25.

The SB-GPNLMS AEC structure is compared to the fullband gradient proportionate normalized least-mean-square (FB-GPNLMS), fullband improved proportionate normalized least-mean-square (FB-IPNLMS), and FB-NLMS structures in Figure 4.4. The echo path changes occurred at the same times as in the previous simulation and lasted for 0.5 and 1.0 seconds respectively. The FB-GPNLMS, FB-IPNLMS, and FB-NLMS structures were derived from the GGPAPA according to the settings outlined in Section 2.2.2.1 of Chapter 2 with $\gamma = 1$, and $\eta = 0.9999$ for the GPNLMS algorithm, and $\beta = 0.5$, and $\xi = 0.00001$ for both the GPNLMS and IPNLMS algorithms, along with $\delta = 0.00001$ for all algorithms. The SB-GPNLMS structure was implemented based on the SB-GGPAPA, as discussed in Section 4.1.1, with the same GPNLMS parameter settings in all subbands as the FB-GPNLMS structure, along with the filter bank characteristics given in Section 4.2.1. The adaptive filters of the fullband structures were set to lengths of $N = 2000$ while the adaptive filters of the SB-GPNLMS structure were set to $N_{sb_i} = N/M = 500$ in all subbands. Furthermore, the GGPAPA parameters for the compared structures were set to the values outlined based on [12], and because they resulted in good ERLE performance for the given overall step size. With an overall step size of $\mu_i = 0.25$ used in all subbands of the SB-GPNLMS structure, a smaller drop in ERLE performance along with faster convergence back to steady-state operation is achieved for both echo path changes, compared to the fullband AEC structures with the same overall step size. The SB-GPNLMS structure achieves up to 10 dB, 5 dB, and 2 dB higher ERLE during the echo path changes as well as during reconvergence back to steady-state operation compared to the FB-NLMS, FB-IPNLMS, and FB-GPNLMS structures respectively. As minimum ERLE is important to the users' subjective quality of the hands-free conversation according to [12], the SB-GPNLMS has the greatest advantage compared to the other structures in this respect. The improved tracking performance of the SB-GPNLMS structure during the echo path fluctuations, as well as the faster reconvergence after the changes, can be attributed to the shorter length subband adaptive

filters and to the individual gradient proportionate subband filter tap step sizes, as discussed in Section 4.1.1. Also, it should be noted that during initial convergence before the echo path changes the SB-GPNLMS structure converges faster than the FB-NLMS, SB-IPNLMS, and FB-GPNLMS structures by up to 10 dB, 3 dB, and 2 dB respectively for the same reasons as previously mentioned. Furthermore, during stationary conditions within the acoustic environment the SB-GPNLMS structure achieves the same constant steady-state ERLE performance of approximately 30 dB as the other structures.

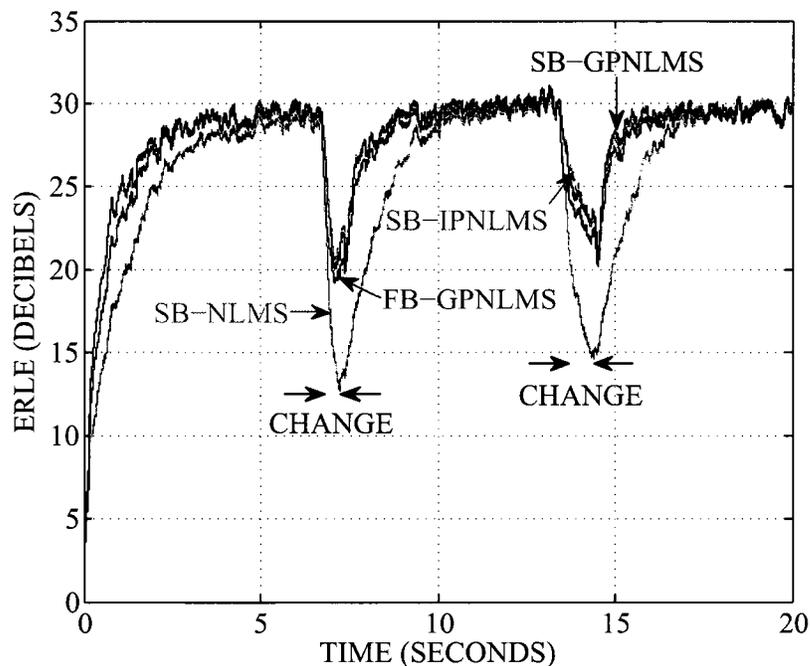


Figure 4.5: ERLE for FB-GPNLMS, SB-NLMS, SB-IPNLMS, and SB-GPNLMS with WGN input in a synthetic changing environment and a step size of 0.25.

Figure 4.5 compares the tracking performance of the SB-GPNLMS structure to the SB-IPNLMS, SB-NLMS, and FB-GPNLMS structures under the previous simulation conditions. The SB-IPNLMS and SB-NLMS AEC structures were realized from the SB-GGPAPA according to the parameter settings given in Section 4.1.1 with $\beta_i = 0.5$ in all subbands for SB-IPNLMS, and $\delta = 0.00001$ for all algorithms and in all subbands as in the previous simulation. All subband structures were implemented

based on the filter bank characteristics given in Section 4.2.1, with the same subband adaptive filter lengths and overall step sizes as for the SB-GPNLMS structure. The FB-GPNLMS structure is included in order to compare the best performing fullband AEC structure to the performance of the subband structures under the changing echo path conditions. The SB-GPNLMS and SB-IPNLMS structures perform nearly identically during the echo path changes as well as during reconvergence back to steady-state operation, while at the same time achieve up to 7 dB and 2 dB higher ERLE than the SB-NLMS and FB-GPNLMS structures respectively. During initial convergence the SB-GPNLMS and SB-IPNLMS structures converge faster than the SB-NLMS and FB-GPNLMS structures by the same amount as during reconvergence after the echo path changes.

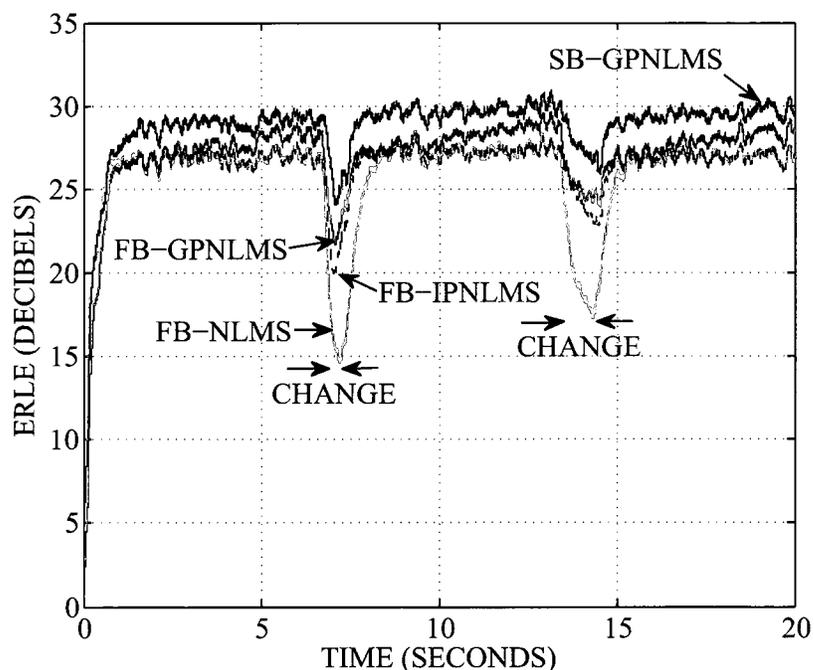


Figure 4.6: ERLE for FB-NLMS, FB-IPNLMS, FB-GPNLMS, and SB-GPNLMS with WGN input in a synthetic changing environment and a step size of 1.0.

Under the previous simulation conditions the tracking performance of the SB-GPNLMS AEC structure is compared to the FB-GPNLMS, FB-IPNLMS, and FB-NLMS structures in Figure 4.6 and to the SB-IPNLMS, SB-NLMS, and FB-GPNLMS

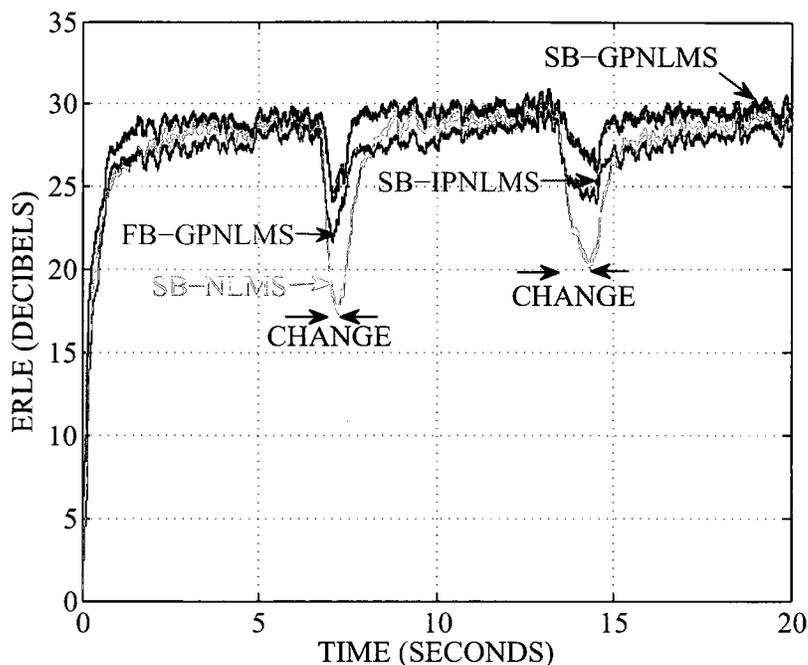


Figure 4.7: ERLE for FB-GPNLMS, SB-NLMS, SB-IPNLMS, and SB-GPNLMS with WGN input in a synthetic changing environment and a step size of 1.0.

structures in Figure 4.7, for an overall step size of 1.0 used in all adaptive filters of all structures. The same trends in ERLE performance for the SB-GPNLMS structure compared to the other AEC structures are shown in Figures 4.6 and 4.7 as for the smaller overall step size used previously, with the SB-GPNLMS structure outperforming the other AEC structures by slightly increased amounts during the echo path fluctuations. A notable difference in these results is the higher steady-state ERLE of approximately 3 dB achieved by the subband structures compared to the FB-IPNLMS and FB-NLMS structures. Moreover, the FB-GPNLMS structure converges towards the same steady-state value as the subband structures. At the smaller overall step size used in the previous simulations all AEC structures achieved the same steady-state ERLE level. The gap in steady-state ERLE performance can be attributed to the non-Wiener phenomenon that occurs for oversampled subband adaptive filters at sufficiently large step sizes for correlated inputs [102]. The non-Wiener phenomenon does not occur for lower step sizes used in oversampled subband adaptive filters as

shown in Figures 4.4 and 4.5 and [102]. Furthermore, as the OSR (K/M) decreases the subband signals become less correlated and the non-Wiener effect diminishes as shown in [102]. Also, it is interesting to note that the FB-GPNLMS structure appears to exhibit some non-Wiener phenomenon as its steady-state ERLE tends towards that of the subband structures, while the other fullband structures achieve a lower steady-state ERLE. This could be due to the individual gradient proportionate filter tap step sizes used in the FB-GPNLMS structure.

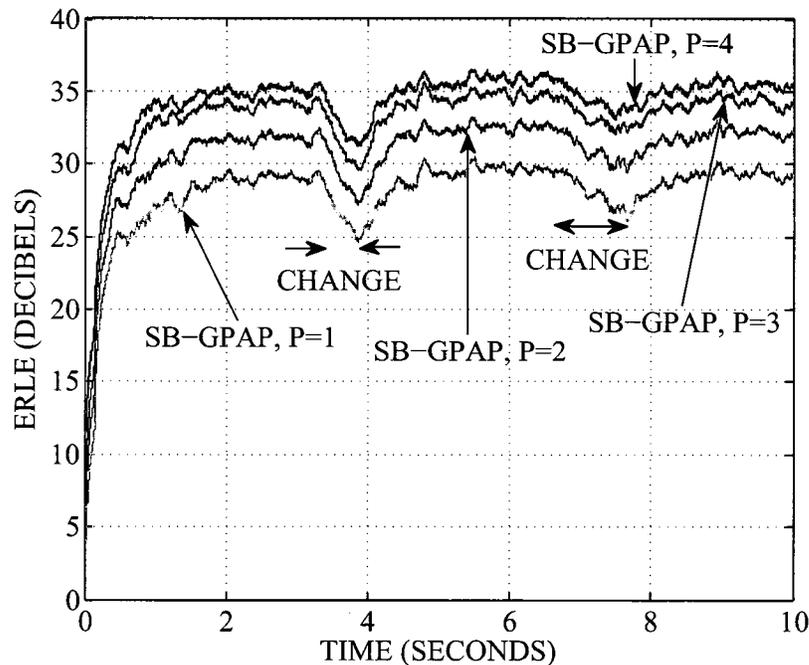


Figure 4.8: ERLE for SB-GPAP with WGN input in a synthetic changing environment and an increasing P for $\mu_i = 1.0$ in all subbands.

Figure 4.8 shows the ERLE performance of the subband gradient proportionate affine projection (SB-GPAP) AEC structure under the changing echo path conditions for an increasing projection order. The previous simulation conditions for the SB-GPNLMS structure (i.e. the SB-GPAP structure with $P = 1$) were used with the exception of P . Also, a 10 second WGN reference signal was used, where the changes in $y(n)$ took place at approximately 3 and 7 seconds into the simulation and lasted for 0.5 and 1 seconds respectively. As P increases the SB-GPAP structure is able

to track the echo path changes faster, as shown by the decreasing drop in ERLE performance, while converging back to steady-state faster after the echo path changes. The level of steady-state ERLE increases with P due to an increase in the non-Wiener phenomenon effect discussed previously. During steady-state operation before and after the echo path variations the SB-GPAP structure with $P = 2$, $P = 3$, and $P = 4$ reaches a higher ERLE by up to 3 dB, 5 dB, and 6 dB respectively, compared to the $P = 1$ case. Also, it should be noted that the improvement in performance tends to diminish as P increases, which agrees with the results of [52] for an affine projection (AP) based critically sampled subband AEC structure.

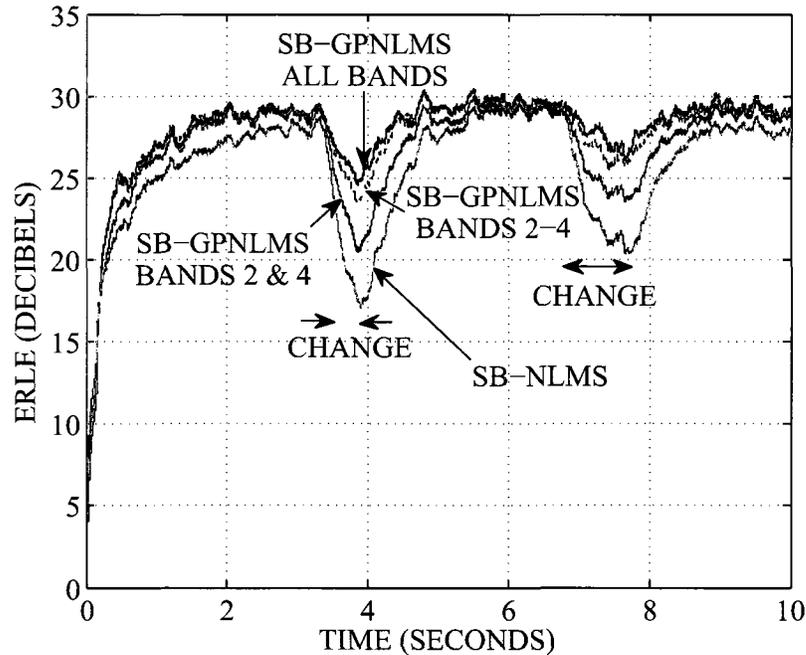


Figure 4.9: ERLE for SB-GPNLMS with WGN input in a synthetic changing environment and the GPNLMS algorithm applied in only specific subbands.

The impact on tracking performance by applying the GPNLMS algorithm to the subband filters on a per subband basis is investigated in Figure 4.9. The performance of the SB-GPNLMS structure with the GPNLMS algorithm implemented in all, the 1–4 kHz, and the 1–2 kHz and 3–4 kHz subbands, along with the performance of the SB-NLMS structure, are shown under the previous simulation conditions corresponding

to an overall step size of 1.0 with $P = 1$. Again, a 10 second WGN reference signal was used where the changes in $y(n)$ took place at approximately 3 and 7 seconds into the simulation and lasted for 0.5 and 1 seconds respectively. The simpler NLMS algorithm was used to adapt the subband filters that did not implement the GPNLMS algorithm. The SB-GPNLMS structure with the GPNLMS algorithm applied in all subbands (SB-GPNLMS ALL BANDS) outperforms the SB-NLMS structure by up to 7 dB higher ERLE during the echo path changes and during reconvergence to steady-state as discussed previously. With the GPNLMS algorithm applied in only the 1–2 kHz and 3–4 kHz subbands (SB-GPNLMS BANDS 2 & 4), the SB-GPNLMS structure is able to track and recover from the echo path changes in those subbands faster compared to the SB-NLMS structure by up to 3 dB higher ERLE. However, this structure still lags the same structure with the GPNLMS algorithm implemented in all subbands by up to 4 dB. When the GPNLMS algorithm is applied in the 1–4 kHz subbands (SB-GPNLMS BANDS 2-4), the SB-GPNLMS structure performs very similarly to the same structure with the GPNLMS algorithm in all subbands, with only up to 1 dB lower ERLE during and after the echo path fluctuations. This small ERLE performance difference is a result of the minor changes that occur in the 0–1 kHz subband echo path impulse response as outlined in Section 4.2.2. Thus, applying the slower tracking less complex NLMS algorithm in the 0–1 kHz subband provides almost the same performance as the GPNLMS algorithm.

The tracking ability of the SB-GPNLMS structure with the GPNLMS algorithm applied only to the initial adaptive filter taps is outlined in Figure 4.10. The previous simulation settings were used, where N_{f_i} was set to 50 in all subbands since the significant changes in the subband impulse response taps were shown to occur in this region, see Figure 4.2. The GPNLMS algorithm with $N_{f_i} = 50$ was implemented based on (4.14) to (4.25) with $\delta_{f_i} = 0.1$ and $\delta_{t_i} = 0.00001$. The SB-GPNLMS structure with $N_{f_i} = 50$ in all subbands (SB-GPNLMS ALL BANDS $N_{f_i} = 50$) attains almost

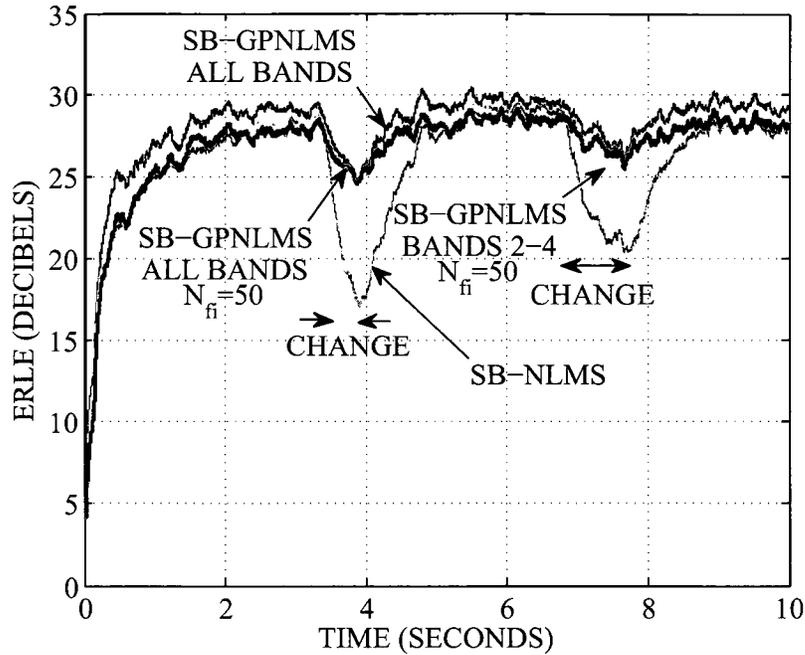


Figure 4.10: ERLE for SB-GPNLMS with WGN input in a synthetic changing environment and the GPNLMS algorithm applied to only the initial adaptive filter taps.

equivalent ERLE performance during the echo path changes as the standard SB-GPNLMS structure with $N_{sb_i} = 500$ in all subbands (SB-GPNLMS ALL BANDS). This demonstrates that tracking performance can be maintained by applying the GPNLMS algorithm on only the highly variable subband adaptive filter taps, with the simpler NLMS algorithm applied to the remainder of the taps that experience very small variations. This also results in significant complexity savings as discussed in Section 4.2.4. The SB-GPNLMS structure with $N_{f_i} = 50$ in only the 1–4 kHz subbands (SB-GPNLMS BANDS 2-4 $N_{f_i} = 50$) lags in tracking performance by only up to 1 dB lower ERLE compared to the SB-GPNLMS structure with $N_{f_i} = 50$ in all subbands. This verifies again that only minor changes occur to all coefficients in the 0–1 kHz subband echo path impulse response. Compared to the SB-NLMS structure, the SB-GPNLMS structure with $N_{f_i} = 50$ in all subbands achieves up to 7 dB higher ERLE during the echo path changes. In terms of initial convergence performance the SB-GPNLMS structures with $N_{f_i} = 50$ converge at approximately the same rate as

the SB-NLMS structure, since the majority of the subband filter taps are adapted by the NLMS algorithm. The SB-GPNLMS structures with $N_{f_i} = 50$ lag the standard SB-GPNLMS structure by up to 3 dB during the initial convergence period.

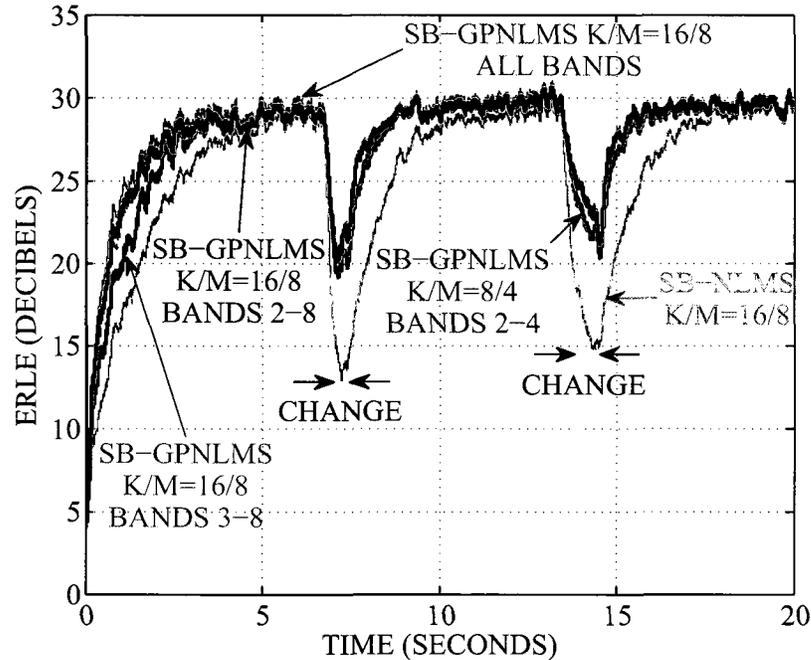


Figure 4.11: ERLE for SB-GPNLMS with WGN input in a synthetic changing environment and an increasing number of subbands for the same OSR.

The impact on tracking performance by applying the GPNLMS algorithm on a per subband basis is investigated in Figures 4.11 and 4.12 for an increasing number of subbands with the same OSR. A 20 second WGN reference signal was used with the echo path changes occurring at 7 and 13 seconds, and lasting for 0.5 and 1 second respectively. Also, the step sizes for all structures was set to $\mu_i = 0.25$ in all subbands. The SB-GPNLMS structures with $K = 16$ subbands and a decimation ratio of $M = 8$, were implemented based on a filter bank with a prototype lowpass filter of length $L_{pr} = 128$. This filter bank was designed in the same manner as the $K = 8$ $M = 4$ filter bank discussed in Section 4.2.1. The $K = 16$ $M = 8$ filter bank with the 128 tap prototype lowpass filter achieved approximately 100 dB of stopband attenuation, along with a 93 dB SNR due to aliasing based on a WGN input. The

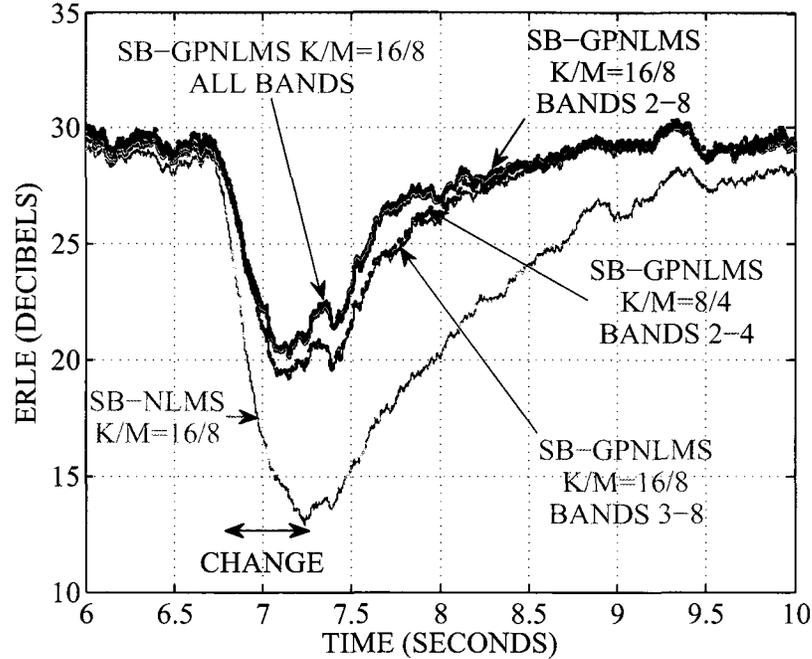


Figure 4.12: ERLE from 6 to 10 seconds for SB-GPNLMS with WGN input in a synthetic changing environment and an increasing number of subbands for the same OSR.

adaptive filters of the SB-GPNLMS and SB-NLMS structures based on the $K = 16$ $M = 8$ filter banks were set to a length of $N_{sb_i} = 250$ in all subbands. The adaptive filters of the $K = 8$ $M = 4$ SB-GPNLMS structure were set to $N_{sb_i} = 500$ in all subbands as in the previous simulations. Again, the simpler NLMS algorithm was used to adapt the subband filters that did not implement the GPNLMS algorithm. The $K = 16$ $M = 8$ SB-GPNLMS structure with the GPNLMS algorithm implemented in all subbands except the 0–500 Hz subband (SB-GPNLMS K/M=16/8 BANDS 2-8) achieves the same ERLE performance during and after the echo path changes as the same structure with the GPNLMS algorithm implemented in all the subbands (SB-GPNLMS K/M=16/8 ALL BANDS). This demonstrates that virtually no changes occur in the 0–500 Hz subband echo path impulse response. Thus, applying the fast tracking GPNLMS algorithm in the 0–500 Hz band results in no improvement in tracking performance. With the GPNLMS algorithm applied in all

subbands except the 0–1 kHz subbands, the $K = 16$ $M = 8$ SB-GPNLMS structure (SB-GPNLMS $K/M=16/8$ BANDS 3-8) and the $K = 8$ $M = 4$ SB-GPNLMS structure (SB-GPNLMS $K/M=8/4$ BANDS 2-4) attain the same ERLE performance during and after the echo path changes. These structures lag the $K = 16$ $M = 8$ SB-GPNLMS structure with the GPNLMS algorithm applied in all subbands (SB-GPNLMS $K/M=16/8$ ALL BANDS) by only up to 2 dB of ERLE during the echo path changes, as shown in Figure 4.12 for the first echo path change. This small decrease in ERLE performance is due to the minor changes that occur in the 0–1 kHz subband echo path impulse response as outlined in Section 4.2.2. The $K = 16$ $M = 8$ SB-GPNLMS structure with the GPNLMS algorithm applied in all subbands except the 0–1 kHz subbands (SB-GPNLMS $K/M=16/8$ BANDS 3-8), initially converges slower than its $K = 8$ $M = 4$ counterpart (SB-GPNLMS $K/M=8/4$ BANDS 2-4) by up to 3 dB. This is due to the increase in the number of band edges, and the slower asymptotic convergence that occurs near the band edges [35]. Compared to the SB-NLMS structure, the SB-GPNLMS structures achieve up to 7 dB higher ERLE during initial convergence as well as during the echo path changes. Implementing the SB-GPNLMS structure with a greater number of subbands allows the GPNLMS algorithm to be applied in subbands with a finer frequency range. This can allow for a better match between the affected frequencies during an echo path change and the subbands where the GPNLMS algorithm is applied. Furthermore, implementing the SB-GPNLMS structure with a greater number of subbands for the same OSR results in a lower computational cost, as discussed in Section 4.2.4, at the expense of a higher filter bank group delay due to the longer prototype filter. In this case the total group delay of the analysis and synthesis filter banks for the $K = 16$ $M = 8$ SB-GPNLMS structure is 127 samples, while the total group delay of the of the analysis and synthesis filter banks for the $K = 8$ $M = 4$ SB-GPNLMS structure is 63 samples.

4.2.2.2 Speech Excitation Signal

In this section the AEC performance of the SB-GGPAP structure along with its variants are again compared to each other, as well as to their fullband counterparts under the same changing acoustic conditions outlined in Section 4.2.1. A reference speech excitation signal, $x(n)$, of approximately 12 seconds in duration was used along with the measured impulse responses to create the changing echo signal, $y(n)$, in the same manner as described in Section 4.2.2.1. The speech signal consisted of 4 alternating female and male speech segments from different speakers concatenated together, obtained from the TIMIT speech data base [99]. The input microphone signal, $d(n)$, was created by adding uncorrelated WGN background noise, $\eta(n)$, to $y(n)$ to obtain a SNR of 30 dB. Again, the local talker signal, $\nu(n)$, was set to 0 for the duration of the simulations. The changes in $y(n)$ took place at approximately 4 and 8 seconds into the simulations and lasted for 0.5 and 2 seconds respectively. The performance of the AEC structures are shown during the 7 to 11 second part of the simulation in order to make the results more observable during the second echo path change. The trends in the results subsequently reported were similar for the first echo path change. The initial convergence performance under a speech excitation signal for the SB-GGPAP structure and its variants is investigated in Section 4.2.3.

Figure 4.13 compares the tracking performance of the SB-GPNLMS structure to that of the FB-GPNLMS, FB-IPNLMS, and FB-NLMS structures. All AEC structures were implemented with the same algorithm parameter and filter bank settings as those given in Section 4.2.2.1 for Figure 4.4. Furthermore, the GGPAPA parameters for the compared structures were set to the values outlined based on [12], and because they resulted in good ERLE performance for the given overall step size. The adaptive filters of the fullband structures were set to lengths of $N = 2000$ while the adaptive filters of the SB-GPNLMS structure were set to $N_{sb_i} = N/M = 500$ in all subbands. An overall step size of 1.0 was used by all the adaptive filters in all compared structures. During the echo path change the SB-GPNLMS structure

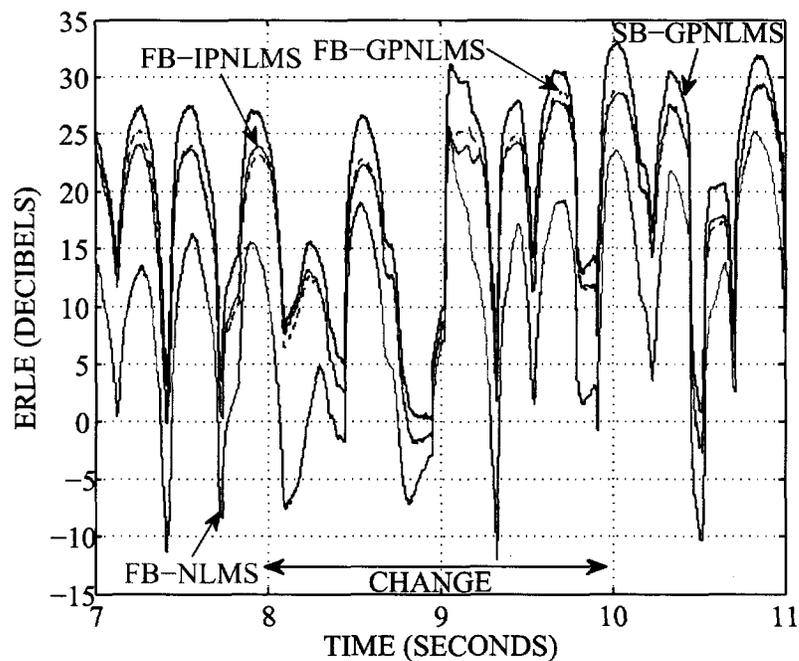


Figure 4.13: ERLE for FB-NLMS, FB-IPNLMS, FB-GPNLMS, and SB-GPNLMS with speech input in a synthetic changing environment and a step size of 1.0.

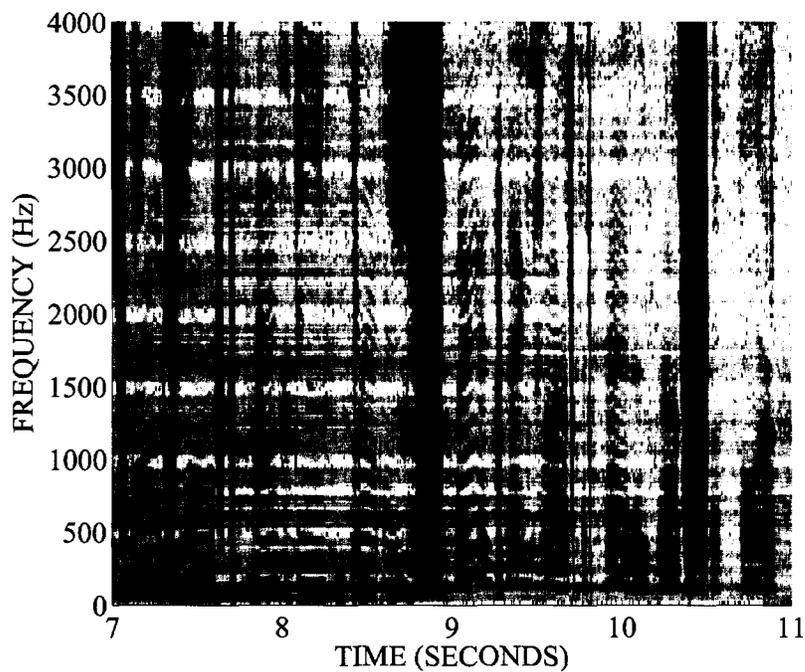


Figure 4.14: Spectrogram of reference speech excitation signal.

outperforms the FB-GPNLMS, FB-IPNLMS, and FB-NLMS structures by up to 5 dB, 7 dB, and 12 dB higher ERLE respectively. Furthermore, before and after the echo path change the SB-GPNLMS structure achieves up to 5 dB higher ERLE than the FB-GPNLMS and FB-IPNLMS structures, and up to 10 dB higher ERLE than the FB-NLMS structure. As the echo path variations are most significant at high frequencies due to the small hand obstruction (see Section 4.2.2) the performance improvement of the SB-GPNLMS structure is most apparent when high frequency speech coincides with the echo path change. This was similarly noted in [100] and can be seen around the 9 second mark in Figure 4.13. Figure 4.14 contains a spectrogram of the reference speech excitation signal that verifies the high frequency content around the 9 second mark. The improved performance of the SB-GPNLMS structure can again be attributed to the shorter length subband adaptive filters and to the individual gradient proportionate subband filter tap step sizes, as discussed in Section 4.1.1.

Under the previous simulation conditions the SB-GPNLMS structure is compared to the SB-IPNLMS, SB-NLMS, and FB-GPNLMS structures in Figure 4.15. The SB-IPNLMS and SB-NLMS AEC structures were realized with the same algorithm parameter and filter bank settings as those given in Section 4.2.2.1 for Figure 4.5. Also, both structures used the same subband adaptive filter lengths and overall step sizes as the SB-GPNLMS structure. The FB-GPNLMS structure is again included for reference purposes in order to compare the best performing fullband AEC structure to the performance of the subband structures in the changing acoustic environment. Compared to the SB-IPNLMS and SB-NLMS structures, performance improvements of up to 2 dB and 12 dB higher ERLE are realized by the SB-GPNLMS structure during the echo path change. Again, the improved performance of the SB-GPNLMS structure is most evident when high frequency speech coincides with the echo path change as seen around the 9 second mark of the simulation. The SB-GPNLMS structure also has similar performance improvements before and after the echo path change

compared to the other subband structures.

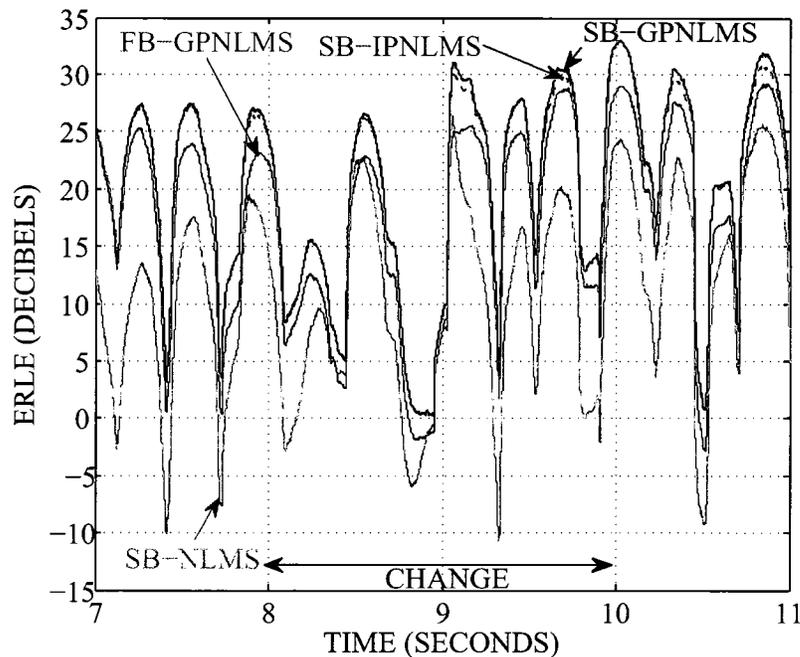


Figure 4.15: ERLE for FB-GPNLMS, SB-NLMS, SB-IPNLMS, and SB-GPNLMS with speech input in a synthetic changing environment and a step size of 1.0.

Figure 4.16 displays the ERLE performance of the SB-GPAP structure as the projection order increases. The parameters of the SB-GPAP structure were the same as for SB-GPNLMS structure (i.e. the SB-GPAP structure with $P = 1$) in the previous simulations except for P . As P increases from 1 to 2 the SB-GPAP structure attains up to 3 dB higher ERLE during the echo path change. However, increasing P to 3 provides only a marginal improvement in tracking performance during the echo path change compared to $P = 2$. Similar performance improvements also occur during reconvergence to steady-state after the echo path change. These diminishing performance improvements with an increasing P are consistent with the results reported for a WGN excitation signal in Section 4.2.2.1.

The tracking performance of the SB-GPNLMS structure with the GPNLMS algorithm applied in only specific subbands, and the NLMS algorithm applied in the other subbands is outlined in Figure 4.17. The performance of the SB-GPNLMS structure

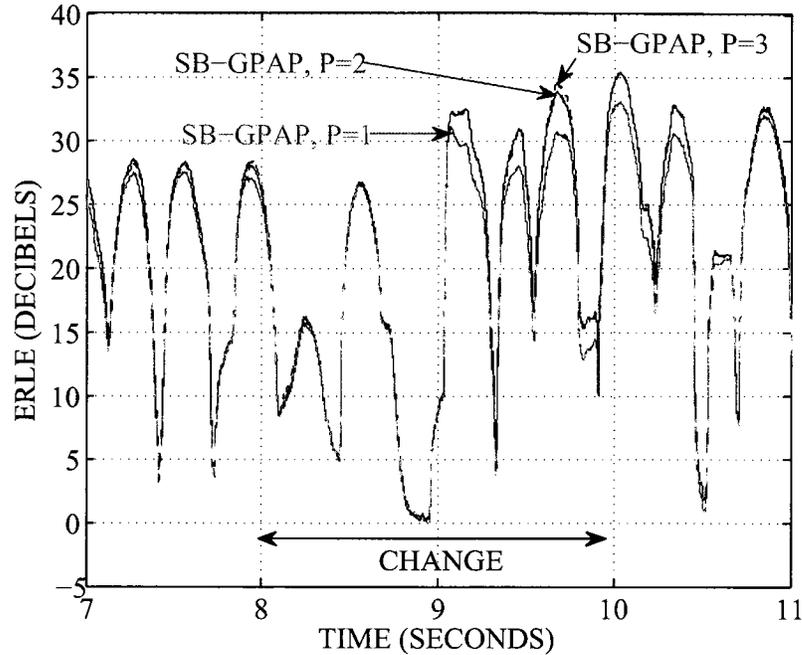


Figure 4.16: ERLE for SB-GPAP with speech input in a synthetic changing environment and an increasing P for $\mu_i = 1.0$ in all subbands.

with the GPNLMS algorithm implemented in all, the 1–4 kHz, and the 1–2 kHz and 3–4 kHz subbands, along with the performance of the SB-NLMS structure, are shown under the previous simulation conditions with an overall step size of $\mu_i = 0.25$ in all subbands and $P = 1$. The SB-GPNLMS structure with the GPNLMS algorithm used in the 1–4 kHz (SB-GPNLMS BANDS 2-4) along with the 1–2 kHz and 3–4 kHz subbands (SB-GPNLMS BANDS 2 & 4), attain greater ERLE performance than the SB-NLMS structure by up to 5 dB and 3 dB respectively during the echo path change. However, these SB-GPNLMS structures lag in ERLE performance by up to 5 dB during the echo path change compared to the standard SB-GPNLMS structure (SB-GPNLMS ALL BANDS). This performance discrepancy is greatest when the echo path change coincides with speech that is restricted to the subbands without the GPNLMS algorithm, as seen around the 8 and 9.5 second marks of the simulation. The lower ERLE performance occurs because of the slower tracking NLMS algorithm used in these subbands. When the echo path change occurs simultaneously

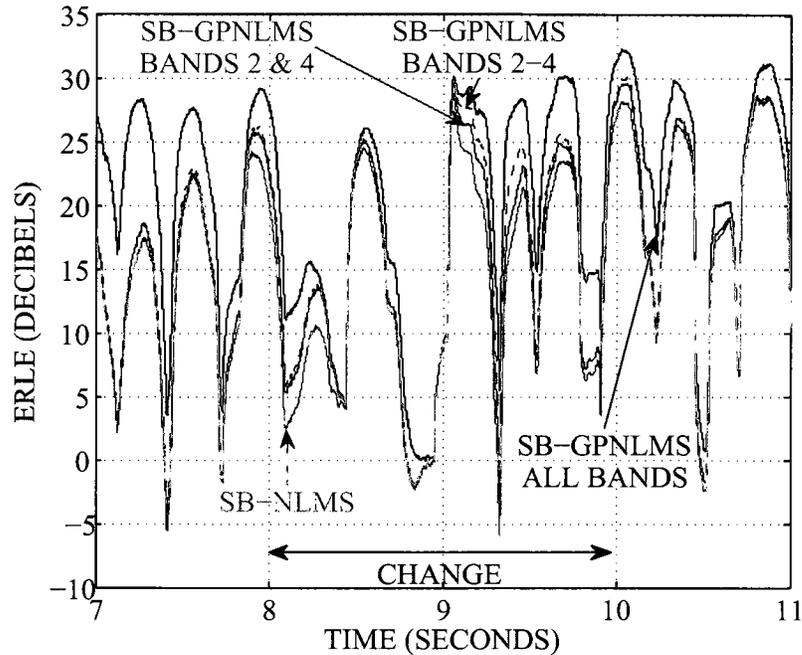


Figure 4.17: ERLE for SB-GPNLMS with speech input in a synthetic changing environment and the GPNLMS algorithm applied in only specific subbands.

with speech confined to the subbands with the GPNLMS algorithm applied, ERLE performance similar to having the GPNLMS algorithm in all subbands is achieved, as seen around the 8.5 and 9 second marks. Applying the GPNLMS algorithm in the subbands where little change occurs in the corresponding subband LEMS impulse response taps, results in only minor ERLE improvements compared to using the slower tracking less complex NLMS algorithm.

The tracking ability of the SB-GPNLMS structure with the GPNLMS algorithm applied only to the initial adaptive filter taps is shown in Figure 4.18. The previous simulation settings were used where N_{f_i} was set to 50 in all subbands, since the significant changes in the subband impulse response taps were shown to occur in this region, see Figure 4.2. The GPNLMS algorithm with $N_{f_i} = 50$ was implemented based on (4.14) to (4.25) with $\delta_{f_i} = 0.001$ and $\delta_{t_i} = 0.01$. The SB-GPNLMS structure with $N_{f_i} = 50$ in all subbands (SB-GPNLMS ALL BANDS $N_{f_i} = 50$) attains

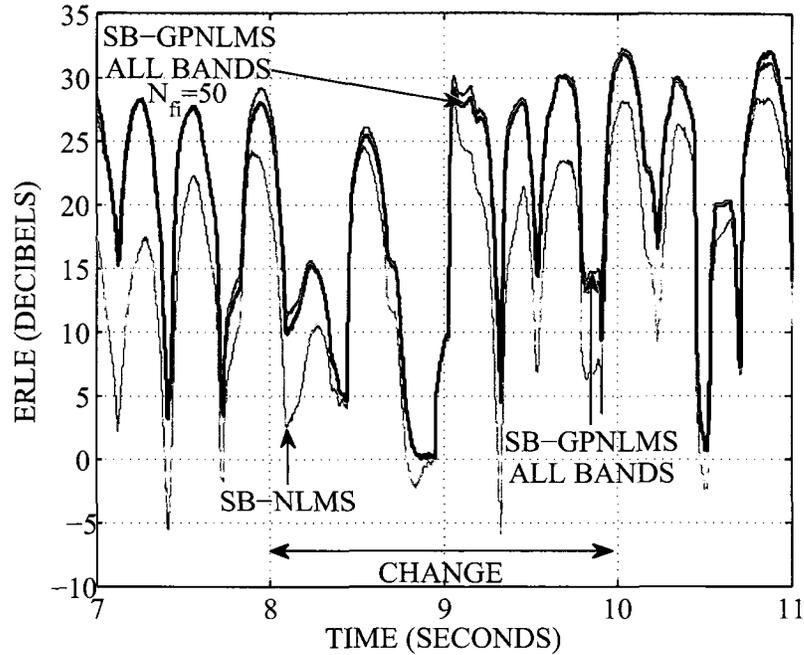


Figure 4.18: ERLE for SB-GPNLMS with speech input in a synthetic changing environment and the GPNLMS algorithm applied to only the initial adaptive filter taps.

nearly equivalent ERLE performance during the echo path change as the standard SB-GPNLMS structure with $N_{sb_i} = 500$ in all subbands (SB-GPNLMS ALL BANDS). Compared to the SB-NLMS structure the SB-GPNLMS structure with $N_{f_i} = 50$ in all subbands achieves up to 7 dB higher ERLE during the echo path change. By applying the fast tracking GPNLMS algorithm on only the initial highly fluctuating subband adaptive filter taps and the less complex NLMS algorithm on the rest of the taps, nearly equivalent tracking performance is obtained compared to the much more computationally expensive alternative of applying the GPNLMS algorithm on all of the subband adaptive filter taps.

4.2.3 Results Based on Measured Reference and Input Data

In this section the performance of the SB-GGPAP structure along with its variants are compared to each other, as well as to their fullband counterparts under experimentally

measured changing acoustic conditions. The experimental setup and methodology described in Section 3.1 of Chapter 3, was used to record microphone and loudspeaker data from the conference phone LEMS configuration for a 30 second speech excitation signal. During the recording an echo path fluctuation was introduced by continual hand waving in front of the phone set at a rate of approximately 1 Hz.

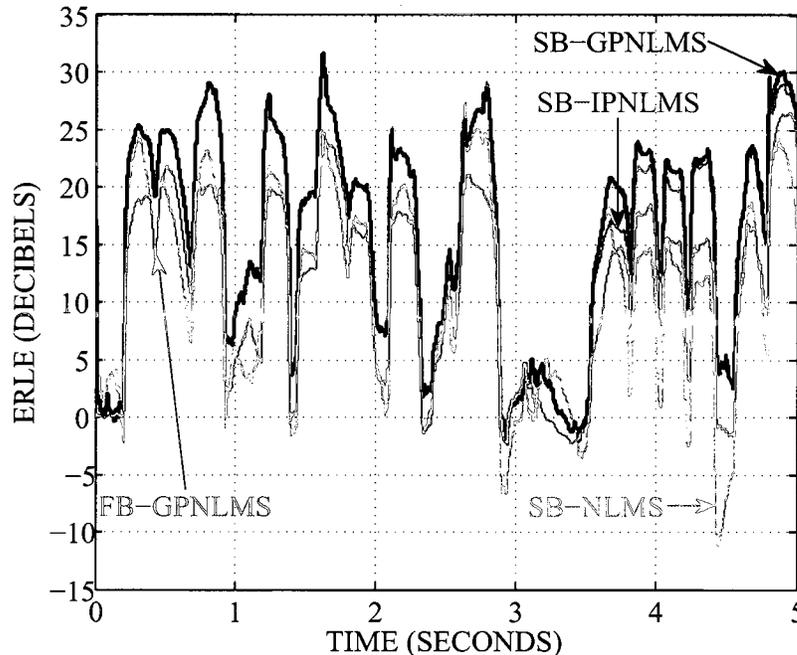


Figure 4.19: ERLE for FB-GPNLMS, SB-NLMS, SB-IPNLMS, and SB-GPNLMS with measured speech input in a changing environment and a step size of 1.0.

The tracking performance of the SB-GPNLMS, FB-GPNLMS, SB-IPNLMS, and SB-NLMS structures are outlined in Figure 4.19 for the measured changing echo path condition. All AEC structures were implemented with the same algorithm parameter and filter bank settings as those given in Section 4.2.2.1 for Figures 4.4 and 4.5. Furthermore, the GGPAPA parameters for the compared structures were set to the values outlined based on [12], and because they resulted in good ERLE performance for the given overall step size. The adaptive filter of the FB-GPNLMS structure was set to a length of $N = 2000$, while the adaptive filters of the subband structures were set to $N_{sb_i} = N/M = 500$ in all subbands. An overall step size of 1.0 was

used by all the adaptive filters in all compared structures. The SB-GPNLMS and SB-IPNLMS structures attain equivalent ERLE performance during initial convergence, while outperforming the FB-GPNLMS and SB-NLMS structures by up to 8 dB. During the changing echo path conditions the SB-GPNLMS structure achieves up to 3, 7, and 9 dB higher ERLE compared to the SB-IPNLMS, FB-GPNLMS, and SB-NLMS structures respectively, as seen near the 4 second mark in Figure 4.19. The improved convergence and tracking performance of the SB-GPNLMS structure can again be attributed to the shorter length subband adaptive filters compared to the fullband structure, and to the individual gradient proportionate subband filter tap step sizes, as discussed in Section 4.1.1.

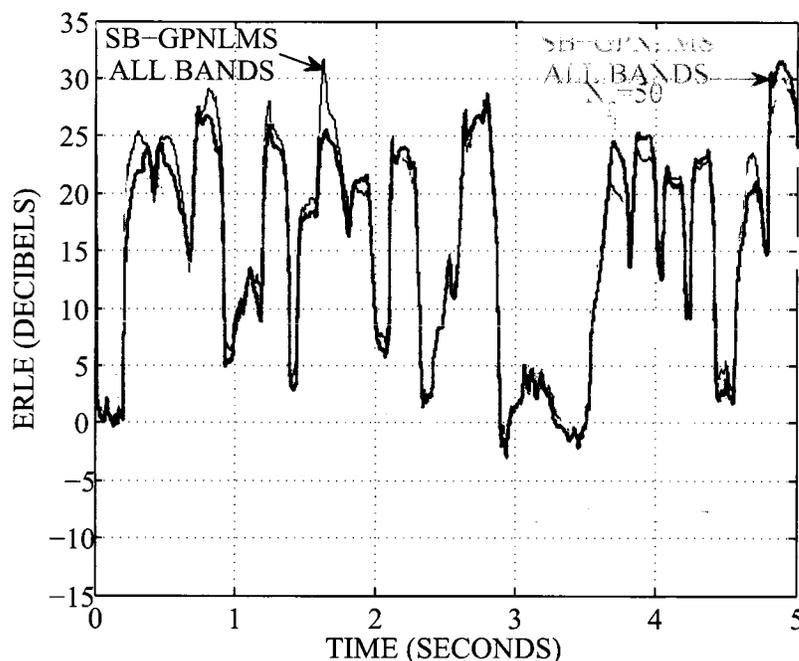


Figure 4.20: ERLE for SB-GPNLMS with measured speech input in a changing environment and the GPNLMS algorithm applied to only the initial adaptive filter taps.

The tracking ability of the SB-GPNLMS structure with the GPNLMS algorithm applied only to the initial adaptive filter taps is outlined in Figure 4.20. The previous simulation settings were used where N_{f_i} was set to 50 in all subbands since the

significant changes in the subband impulse response taps were expected to occur in this region for a small obstruction, see Figure 4.2. The GPNLMS algorithm with $N_{f_i} = 50$ was implemented based on (4.14) to (4.25) with $\delta_{f_i} = 0.001$ and $\delta_{t_i} = 0.01$. Also, the NLMS algorithm was applied in the subbands that did not implement the GPNLMS algorithm. During initial convergence the SB-GPNLMS structure with $N_{f_i} = 50$ in all subbands (SB-GPNLMS ALL BANDS $N_{f_i} = 50$) lags the standard SB-GPNLMS structure (SB-GPNLMS ALL BANDS) by up to 5 dB of ERLE, since the majority of its subband filter taps are adapted with the slower converging NLMS algorithm. The SB-GPNLMS structure with $N_{f_i} = 50$ in only the 1–4 kHz subbands (SB-GPNLMS BANDS 2-4 $N_{f_i} = 50$) converges slower than its counterpart with $N_{f_i} = 50$ in all subbands by up to 3 dB, due to the slower converging NLMS algorithm used to adapt the filter in the 0–1 kHz subband. The SB-GPNLMS structure with $N_{f_i} = 50$ in all subbands attains very similar ERLE performance during the echo path changes compared to the standard SB-GPNLMS structure with $N_{sb_i} = 500$ in all subbands, as seen between the 2 and 5 second period of the simulation in Figure 4.20. During the same period the SB-GPNLMS structure with $N_{f_i} = 50$ in only the 1–4 kHz subbands achieves roughly the same ERLE performance as the SB-GPNLMS structure with $N_{f_i} = 50$ in all subbands. This verifies that only minor changes occur in the 0–1 kHz subband, while the majority of the significant echo path fluctuations occur in the 1–4 kHz region. Compared to the SB-NLMS structure, the SB-GPNLMS structure with $N_{f_i} = 50$ in all subbands achieves up to 10 dB higher ERLE during the echo path changes. These results again demonstrate that applying the fast tracking GPNLMS algorithm on only the initial highly variable subband adaptive filter taps, and the less complex slower tracking NLMS algorithm on the rest of the taps, leads to similar tracking performance compared to the much more computationally expensive alternative of applying the GPNLMS algorithm on all of the subband adaptive filter taps.

The tracking performance of the SB-GPAP, fullband gradient proportionate affine

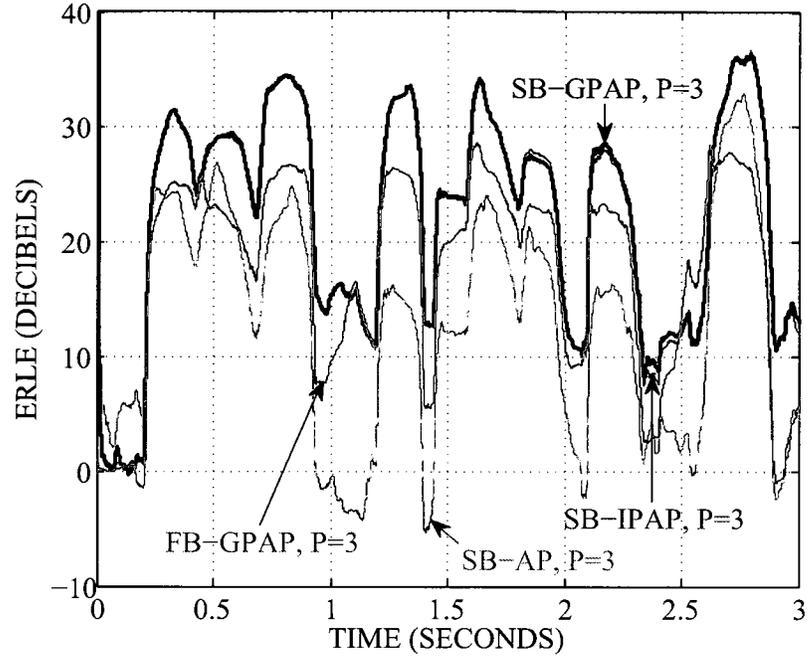


Figure 4.21: ERLE from 0 to 3 seconds for FB-GPAP, SB-AP, SB-IPAP, and SB-GPAP with measured speech input in a changing environment and a step size of 1.0.

projection (FB-GPAP), subband improved proportionate affine projection (SB-IPAP), and SB-AP structures are compared in Figures 4.21 and 4.22 for the measured changing echo path condition. Again, all AEC structures were implemented with the same algorithm parameter and filter bank settings as those given in Section 4.2.2.1 for Figures 4.4 and 4.5. The adaptive filter of the FB-GPAP structure was set to a length of $N = 2000$ while the adaptive filters of the subband structures were set to $N_{sb_i} = N/M = 500$ in all subbands. A step size of 1.0 was used by all the adaptive filters in all compared structures. The projection order was set to $P = 3$ for all structures as only marginal improvement in ERLE was observed for higher projection orders. During the initial convergence period the SB-GPAP and SB-IPAP structures attain equivalent ERLE performance, while outperforming the FB-GPAP and SB-AP structures by up to 5 and 15 dB respectively, as shown in Figure 4.21. During the changing echo path conditions the SB-GPAP structure achieves up to 7, 13, and 13

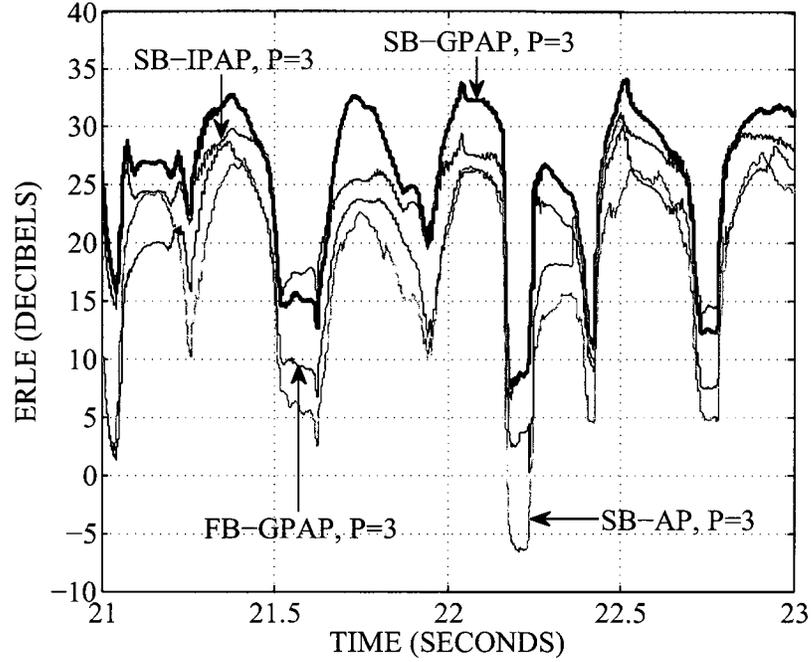


Figure 4.22: ERLE from 21 to 23 seconds for FB-GPAP, SB-AP, SB-IPAP, and SB-GPAP with measured speech input in a changing environment and a step size of 1.0.

dB higher ERLE compared to the SB-IPAP, FB-GPAP, and SB-AP structures respectively, as shown in Figure 4.22. As discussed previously the echo path variations are most significant at high frequencies due to the small hand obstruction, see Section 4.2.2. Thus, the performance improvement of the SB-GPAP structure is most apparent when high frequency speech coincides with the echo path change, as seen around the 21 and 22.25 second marks in Figure 4.22.

The tracking ability of the SB-GPAP structure with the GPAPA applied only to the initial adaptive filter taps is outlined in Figure 4.23. The previous simulation settings were used where N_{f_i} was set to 50 in all subbands as before. The GPAPA with $N_{f_i} = 50$ was implemented based on (4.14) to (4.25) with $\delta_{f_i} = 0.001$ and $\delta_{t_i} = 0.1$. Also, the APA was applied in the subbands that did not implement the GPAPA. The SB-GPAP structure with $N_{f_i} = 50$ in all subbands (SB-GPAP $P = 3$ ALL BANDS $N_{f_i} = 50$) attains similar ERLE performance during the echo path changes compared

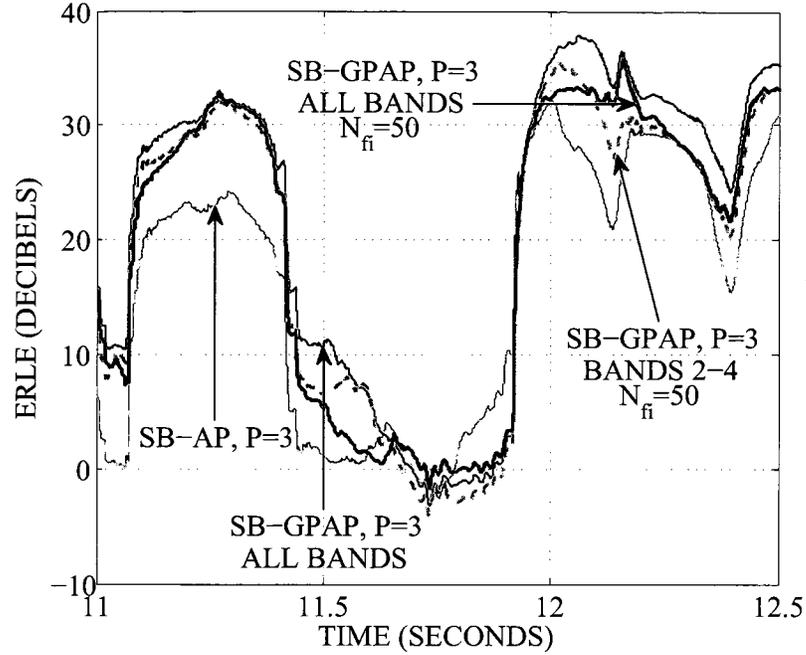


Figure 4.23: ERLE from 11 to 12.5 seconds for SB-GPAP with measured speech input in a changing environment and the GPAPA applied to only the initial adaptive filter taps.

to the standard SB-GPAP structure with $N_{sb_i} = 500$ in all subbands (SB-GPAP $P = 3$ ALL BANDS). The SB-GPAP structure with $N_{f_i} = 50$ in only the 1–4 kHz subbands (SB-GPAP $P = 3$ BANDS 2-4 $N_{f_i} = 50$) achieves roughly the same ERLE performance as the SB-GPAP structure with $N_{f_i} = 50$ in all subbands. This verifies again that only minor changes occur in the 0–1 kHz subband, while the majority of the significant echo path fluctuations occur in the 1–4 kHz region. Compared to the SB-AP structure the SB-GPAP structure with $N_{f_i} = 50$ in all subbands achieves up to 9 dB higher ERLE during the echo path changes. By applying the GPAPA on only the initial highly variable subband adaptive filter taps tracking performance is maintained while significant savings in complexity are realized, as discussed in Section 4.2.4.

The performance of the SB-GPNLMS structure for an increasing OSR is shown in Figure 4.24. The SB-GPNLMS structure was implemented based on filter banks

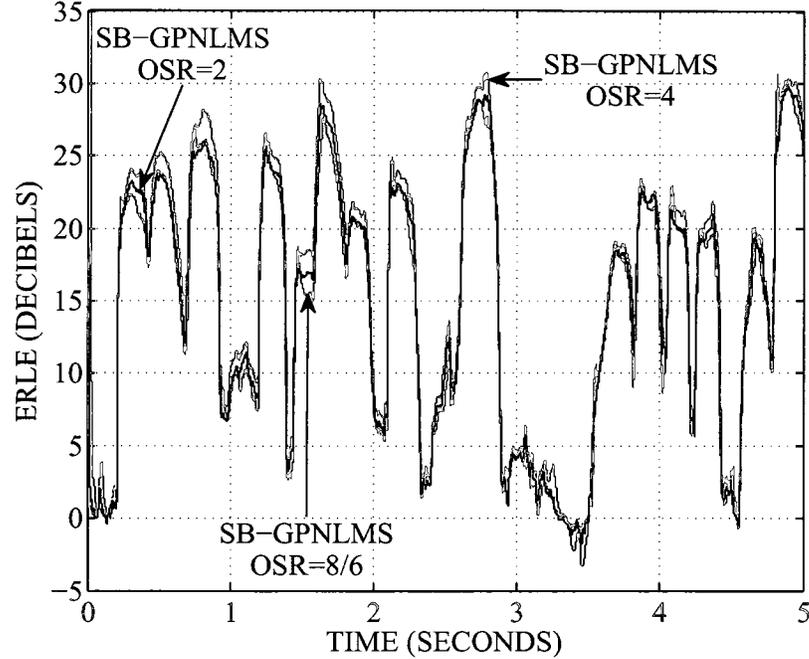


Figure 4.24: ERLE for SB-GPNLMS with measured speech input in a changing environment and an increasing OSR.

with OSRs of $8/6$, 2, and 4 with $K = 8$ subbands in all cases. The filter banks were designed as discussed in Section 4.2.1 with prototype filters of length 192, 64, and 32 taps for the corresponding OSRs of $8/6$, 2, and 4. In all cases the prototype filters provided a stopband attenuation of at least 100 dB and a SNR of at least 90 dB due to aliasing based on a WGN input. The SB-GPNLMS structure was implemented with the previous GPNLMS settings, where the adaptive filters in all subbands were set to lengths of 334, 500, and 1000 for the corresponding OSRs of $8/6$, 2, and 4. Also, a step size of 0.25 was used in all subbands of all structures. All of the SB-GPNLMS structures converge at approximately the same rate with the OSR= 4 structure converging slightly faster than the other two. During the changing echo path conditions all SB-GPNLMS structures attain roughly the same ERLE performance as seen around the 4 second mark. The increasing OSR results in a trade-off between computational complexity and group delay. As the OSR increases the group delay of the SB-GPNLMS structure decreases due to the shorter prototype filters, however

the computational cost of the structure increases due to the lower decimation ratio.

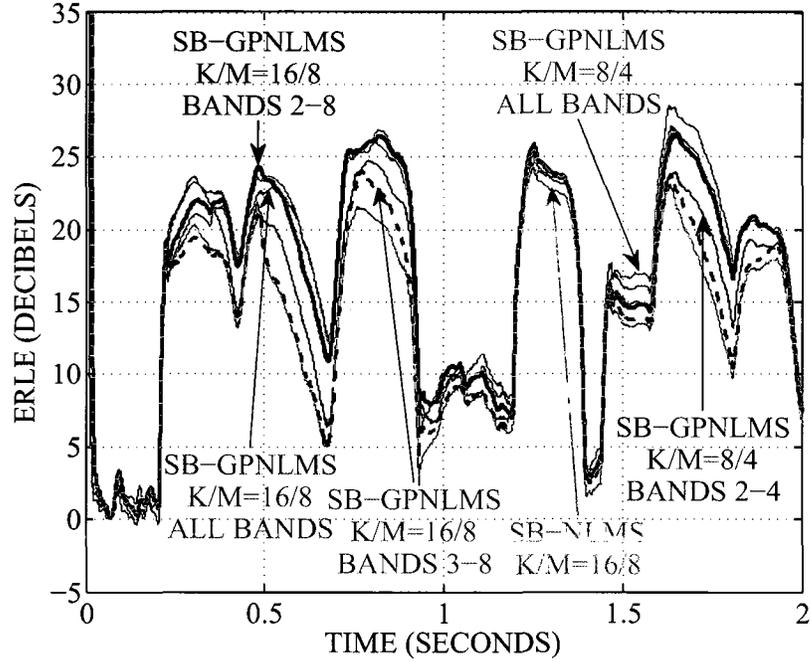


Figure 4.25: ERLE from 0 to 2 seconds for SB-GPNLMS with measured speech input in a changing environment and an increasing number of subbands for the same OSR.

The impact on tracking performance by applying the GPNLMS algorithm on a per subband basis is investigated in Figures 4.25 and 4.26 for an increasing number of subbands with the same OSR. The same filter bank settings as those used in Section 4.2.2.1 for Figure 4.11 were used along with the previous GPNLMS parameter settings. The adaptive filters of the SB-GPNLMS and SB-NLMS structures based on the $K = 16$ $M = 8$ filter banks were set to a length of $N_{sb_i} = 250$ in all subbands, while the adaptive filters of the $K = 8$ $M = 4$ SB-GPNLMS structure were set to $N_{sb_i} = 500$ in all subbands as in the previous simulations. Also, the step sizes for all structures was set to $\mu_i = 0.25$ in all subbands. Again, the simpler NLMS algorithm was used in the subbands of the SB-GPNLMS structure that did not implement the GPNLMS algorithm. The SB-GPNLMS structures with $K = 16$ subbands converge slightly slower than the equivalent SB-GPNLMS structures with $K = 8$ and the GPNLMS algorithm applied over the same frequency range (SB-GPNLMS

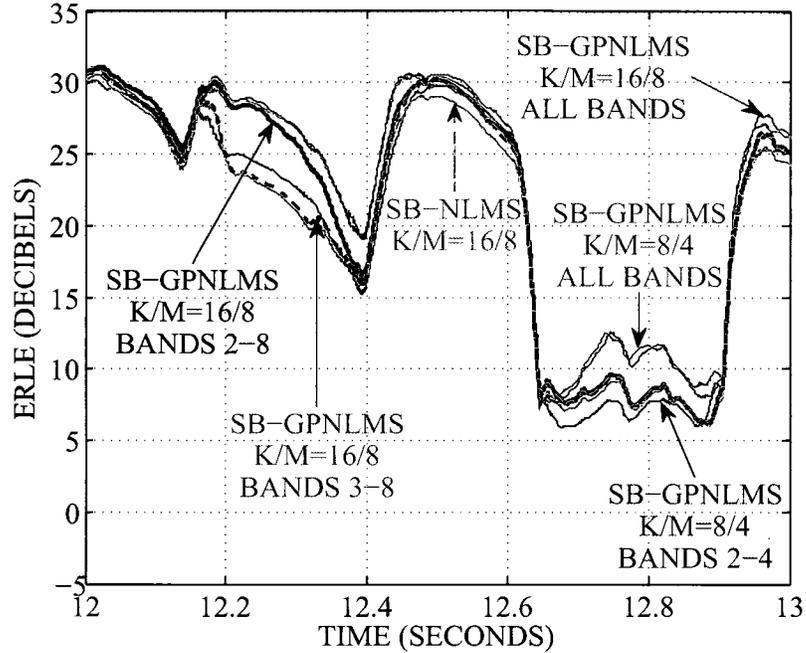


Figure 4.26: ERLE for from 12 to 13 seconds SB-GPNLMS with measured speech input in a changing environment and an increasing number of subbands for the same OSR.

$K/M=16/8$ ALL BANDS compared to SB-GPNLMS $K/M=8/4$ ALL BANDS, SB-GPNLMS $K/M=16/8$ BANDS 3-8 compared to SB-GPNLMS $K/M=8/4$ BANDS 2-4). This is due to the increase in the number of band edges and the slower asymptotic convergence that occurs near the band edges [35], as shown in Figure 4.25. The $K = 16$ $M = 8$ SB-GPNLMS structure with the GPNLMS algorithm implemented in all subbands except the 0–500 Hz subband (SB-GPNLMS $K/M=16/8$ BANDS 2-8), achieves approximately the same ERLE performance during the changing echo path conditions as the same structure with the GPNLMS algorithm implemented in all the subbands (SB-GPNLMS $K/M=16/8$ ALL BANDS), as shown in Figure 4.26 around the 12.2 second mark. This demonstrates that only minor changes occur in the 0–500 Hz subband echo path impulse response, since the low frequency long wavelength sound energy is less impeded by the small hand obstruction than the higher frequency sound energy, as discussed in Section 4.2.2. Again, implementing the SB-GPNLMS

structure with a greater number of subbands for the same OSR allows the GPNLMS algorithm to be applied in subbands with a finer frequency range, which can better match the affected frequency range during a change in the echo path impulse response. Also, implementing the SB-GPNLMS structure with a greater number of subbands for the same OSR results in further computational cost reduction, due to the higher decimation ratio and shorter subband adaptive filters. The drawback is the higher filter bank group delay, as discussed previously.

4.2.4 Computational Complexity Results

This section compares the computational complexity of the SB-GGPAP structure to its variants and corresponding fullband counterparts, in term of the total number of equivalent real mathematical operations required per input sample period. The complexity results presented assume the AEC structures were implemented on an Intel Core 2 Duo processor where a real addition and a real multiplication required approximately the same number of clock cycles to execute for single precision floating point data [103]. A real division and a real square root operation also required the same number of clock cycles to execute, and took approximately 6 times longer than a real addition or multiplication for single precision floating point data [103]. The complexity of the SB-GGPAP structure along with its variants and fullband counterparts were calculated based on Tables 4.1 – 4.3. The system of equations in (4.7), (4.15), and (4.18) were assumed to be solved with the Levinson-Durbin recursion requiring $T_{R,m} = 2P^2 - P$ multiplications, $T_{R,a} = 2P^2 - P$ additions, and $T_{R,d} = P$ divisions per iteration for real valued data as reported in [47]. Where $\mathbf{X}_i^H(m)\mathbf{A}_i(m)\mathbf{X}_i(m) + \delta_i\mathbf{I}$, $\mathbf{X}_{f_i}^H(m)\mathbf{A}_{f_i}(m)\mathbf{X}_{f_i}(m) + \delta_{f_i}\mathbf{I}$, and $\mathbf{X}_i^H(m)\mathbf{X}_i(m) + \delta_{i_i}\mathbf{I}$ were assumed to be approximately Toeplitz matrices. Thus, for complex valued data $T_{C,m} = T_{C_f,m} = T_{C_{i,m}} = 4(2P^2 - P) + 6P$ multiplications, $T_{C,a} = T_{C_{f,a}} = T_{C_{i,a}} = 4(2P^2 - P) + 3P$ additions, and $T_{C,d} = T_{C_{f,d}} = T_{C_{i,d}} = 2P$ divisions per iteration were required. Also, $C_{\sqrt{\cdot},m}$, $C_{\sqrt{\cdot},a}$, and $C_{\sqrt{\cdot},d}$ were set to 0, 0, 0,

and 1 respectively, to reflect the fact that a square root operation required the same amount of time to execute as a division. Furthermore, for $P = 1$ the cost of computing $\mathbf{X}_i^H(m)\mathbf{X}_i(m)$ for the SB-NLMS structure and for the SB-GPNLMS structure with the GPNLMS algorithm applied only to the initial N_{f_i} adaptive filter taps, was assumed to be determined in a reduced complexity fashion. This was achieved for the current iteration by adding the value of $\mathbf{x}_i^H(m)\mathbf{x}_i(m)$ from the previous iteration to the squared magnitude of the current input sample, and subtracting the squared magnitude of the last input sample in the input signal vector from the previous iteration [104]. This requires only 2 real multiplications and additions for real input data, and 4 real multiplications and additions for complex input data per iteration. For the complexity results subsequently presented, the GGPAPA parameter settings for all subband and fullband structures were as given in Sections 4.2.2 and 4.2.3.

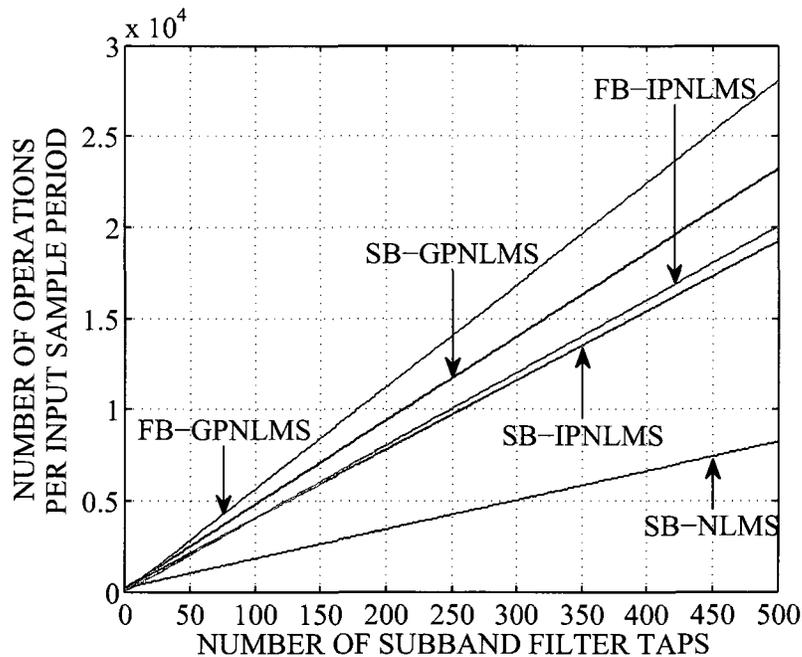


Figure 4.27: Computational complexity for SB-NLMS, SB-IPNLMS, SB-GPNLMS, FB-IPNLMS, and FB-GPNLMS.

Figure 4.27 presents computational complexity results for the SB-GPNLMS, SB-IPNLMS, SB-NLMS, FB-GPNLMS, and FB-IPNLMS structures as the length of the

subband filters increases. The results were computed based on the $K = 8$ $M = 4$ filter bank characteristics given in Section 4.2.1 with the length of each subband adaptive filter, N_{sb_i} , set equal in all subbands and the length of the corresponding fullband adaptive filters set to MN_{sb_i} . As the subband filter length increases the SB-NLMS structure achieves the lowest complexity, while the SB-GPNLMS and SB-IPNLMS structures attain a complexity less than their fullband counterparts. The complexity advantage of the SB-GPNLMS and SB-IPNLMS structures compared to their equivalent fullband structures increases as the number of subband filter taps increases. Compared to each other the SB-IPNLMS structure is slightly less complex than the SB-GPNLMS structure, due to the extra computations required by the GPNLMS algorithm in calculating the individual gradient proportionate subband filter tap step sizes. The significant ERLE performance improvements offered by the SB-GPNLMS structure compared to the SB-NLMS structure, as presented in Sections 4.2.2 and 4.2.3, require approximately 3 times the complexity of the SB-NLMS structure for $N_{sb_i} = 500$. The SB-GPNLMS structure requires roughly 20% less complexity than the FB-GPNLMS structure, and only 20% more complexity than the SB-IPNLMS structure to achieve the respective ERLE performance improvements presented in Sections 4.2.2 and 4.2.3 for $N_{sb_i} = 500$ and correspondingly $N = 2000$.

The complexity savings obtained by applying the GPNLMS algorithm in a subset of the available subbands with the NLMS algorithm in the rest, are given in Figure 4.28 for the SB-GPNLMS structure under the previous parameter settings. As the GPNLMS algorithm is applied in fewer subbands the complexity of the SB-GPNLMS structure naturally approaches that of the SB-NLMS structure. With the GPNLMS algorithm used in 2 and 3 subbands the SB-GPNLMS structure is approximately 2 and 2.4 times the complexity of the SB-NLMS structure for $N_{sb_i} = 500$, while providing improved ERLE tracking performance as shown in Section 4.2.2. Compared to applying the GPNLMS algorithm in all subbands with $N_{sb_i} = 500$, the SB-GPNLMS structure requires 30% and 20% less complexity with the GPNLMS

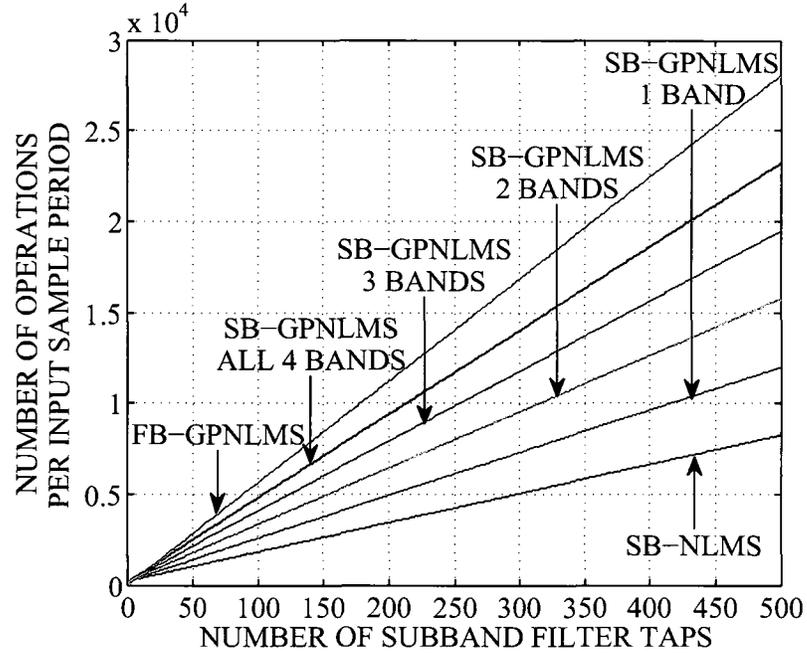


Figure 4.28: Computational complexity for SB-GPNLMS with the GPNLMS algorithm applied in only specific subbands.

algorithm implemented in 2 and 3 subbands respectively, while still achieving similar ERLE performance.

The computational complexity of the SB-GPAP structure compared to the FB-GPAP structure is shown in Figure 4.29 under the previous parameter settings, with the GPAPA applied in all subbands for an increasing P . As P increases the computational advantage of the SB-GPAP structure increases slightly compared to its equivalent fullband structure. However, the computational cost compared to the SB-NLMS structure climbs rapidly with an increasing P . The improved tracking performance of the SB-GPAP structure with $P = 2$ and $P = 3$, as shown in Sections 4.2.2 and 4.2.3, comes at the cost of 1.7 and 2.6 times the operations compared to SB-GPAP structure with $P = 1$ for $N_{sb_i} = 500$. It should also be noted that in all the cases presented the computational advantage of the subband AEC structures will increase compared their fullband equivalents as the OSR decreases.

Significant computational savings are realized with the GPAPA applied to the

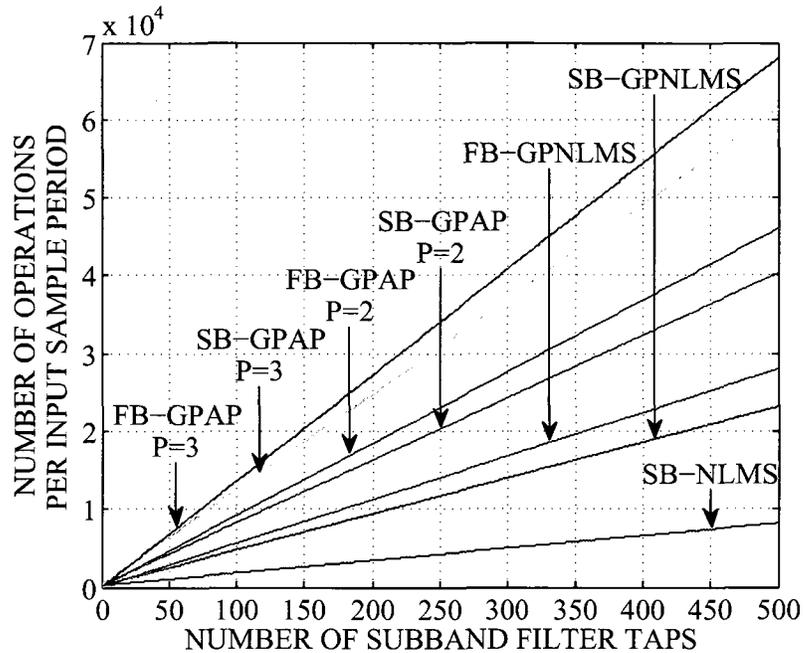


Figure 4.29: Computational complexity for SB-GPAP and FB-GPAP with an increasing projection order.

initial subband adaptive filter taps and the APA applied to the remainder in the SB-GPAP structure. The computational cost of the SB-GPNLMS structure (SB-GPAP with $P = 1$) computed from Table 4.5 under the previous parameter settings with $N_{f_i} = 50$ in all subbands, and an overall filter length of $N_{sb_i} = 500$ in each subband is 9817 operations per input sample period. This is an increase of only 18% compared to the SB-NLMS structure (see Figure 4.27), and results in the significantly improved tracking performance shown in Figures 4.10, 4.18, and 4.20. The SB-GPNLMS structure with $N_{f_i} = 50$ in all subbands is approximately 60% less computationally expensive compared to the standard SB-GPNLMS structure (see Figure 4.27), while still providing nearly the same tracking performance as shown in Figures 4.10, 4.18, and 4.20. Under the same settings with $P = 3$ the SB-GPAP structure with $N_{f_i} = 50$ achieves a complexity of 52353 operations per input sample period, which is only 8% higher than the SB-AP structure (48468 operations per input sample period) and results in the significantly improved tracking performance shown in

Figure 4.23. Compared to the standard SB-GPAP structure (see Figure 4.29), the SB-GPAP structure with $N_{f_i} = 50$ and $P = 3$ is approximately 15% less computationally expensive while attaining similar tracking performance as shown in Figure 4.23.

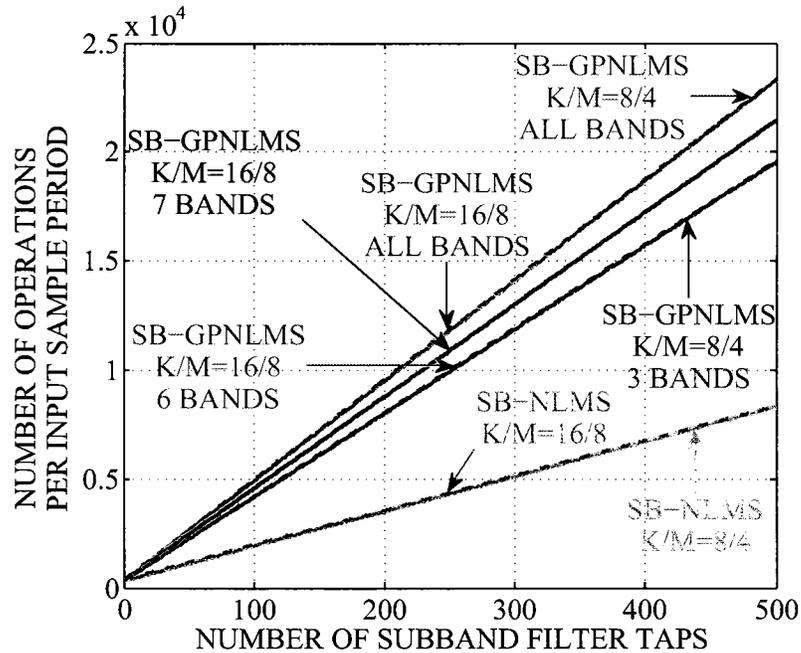


Figure 4.30: Computational complexity for SB-GPNLMS and SB-NLMS with an increasing number of subbands for the same OSR.

The computational cost of the SB-GPNLMS and SB-NLMS structures are shown in Figure 4.30 for an increasing number of subbands with the same OSR of 2. The results are based on the same filter bank settings as those outlined in Section 4.2.2.1 for Figure 4.11. The complexity of the SB-GPNLMS structure with $K = 16$ subbands is slightly greater than with $K = 8$ subbands due to the increased filter bank cost. However, the SB-GPNLMS structure with $K = 16$ subbands requires significantly fewer subbands adaptive filter taps than with $K = 8$ subbands to achieve the same ERLE performance, with the GPNLMS algorithm applied in the subbands covering the same frequency region. As shown in Figures 4.11, 4.25, and 4.26 the SB-GPNLMS structure with $K = 16$ subbands requires subband adaptive filters with a length of

250 to achieve the same performance as the same structure with $K = 8$ subbands and adaptive filters of length 500. Thus, the overall complexity of the SB-GPNLMS structure with $K = 16$ subbands is approximately half of the corresponding $K = 8$ subband structure. Compared to the SB-NLMS structure with $K = 16$ subbands, the SB-GPNLMS structure with $K = 16$ subbands requires slightly more than twice the computations for subbands filter lengths of 250 to achieve the significantly improved ERLE performance shown in the figures previously mentioned. With the GPNLMS algorithm applied in fewer subbands, the complexity of the SB-GPNLMS structure decreases further while providing improved ERLE performance compared to the SB-NLMS structure. With $K = 8$ subbands the SB-GPNLMS structure requires over 4 times the complexity of the SB-NLMS structure with $K = 16$ subbands, to attain the same improved ERLE performance as the SB-GPNLMS structure with $K = 16$ subbands.

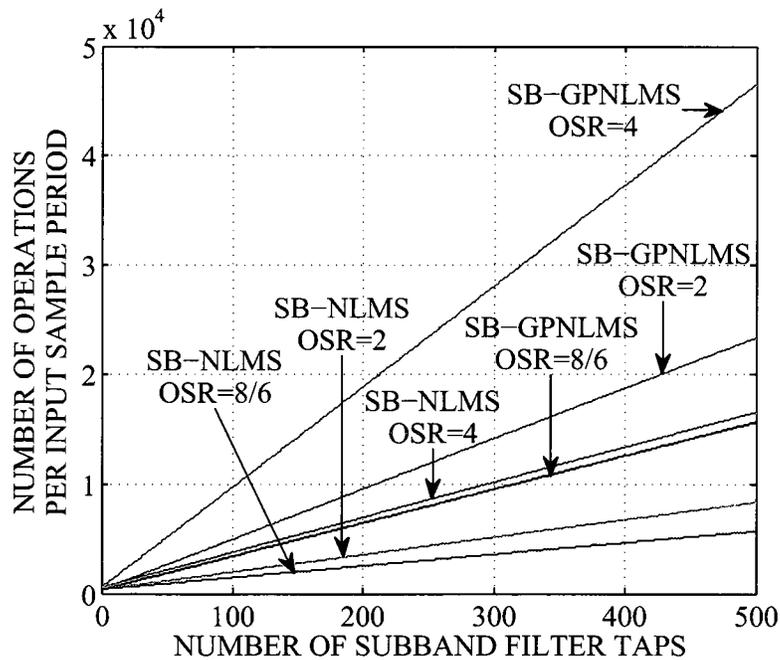


Figure 4.31: Computational complexity for SB-GPNLMS and SB-NLMS with an increasing OSR.

Figure 4.31 outlines the computation cost of the SB-GPNLMS and SB-NLMS

structures as the OSR increases for the same number of $K = 8$ subbands. The results are based on the same filter bank settings as those outlined in Section 4.2.3 for Figure 4.24. As the OSR increases so does the complexity of both structures. To achieve the same ERLE performance fewer subband adaptive filter taps are required for lower OSRs, as shown in Figure 4.24 for the SB-GPNLMS structure. In this case subbands filters of lengths of 334, 500, and 1000 were required for the corresponding OSRs of 8/6, 2, and 4. With subband adaptive filters of length 334, the SB-GPNLMS structure with an OSR of 8/6 is approximately half as complex as the SB-GPNLMS structure with an OSR of 2 and subband adaptive filters of length 500, while achieving equal ERLE performance. Compared to the SB-NLMS structure with an OSR of 2 and subband adaptive filters of length 500, the SB-GPNLMS structure with an OSR of 8/6 and subband adaptive filters of length 334 is only slightly more computationally complex while providing improved ERLE performance.

4.3 Summary

This chapter investigated the incorporation of the GGPAPA into an oversampled subband AEC system to achieve improved AEC performance during changing echo path conditions. Section 4.1 described the SB-GGPAP structure in detail and presented a thorough computational complexity analysis. This flexible SB-GGPAP structure allowed for a family of different proportionate variable step size adaptive filter algorithms to be implemented with only minor parameter modifications of the GGPAPA, resulting in a trade off between AEC tracking performance and overall structure complexity. Performance results for the SB-GGPAP structure, along with its variants and fullband equivalents, were outlined in Section 4.2 based on experimentally measured hands-free system data. The results demonstrated the ability of the SB-GGPAP structure to provide superior tracking performance during fluctuating echo

path conditions. The flexibility of the SB-GGPAP structure was also shown in Sections 4.2.2 and 4.2.3, where the GGPAPA was applied on a per subband basis to allow for fast tracking based on the frequency characteristics of the echo path change, while reducing computational cost. The time domain nature of echo path changes was successfully exploited by applying the GGPAPA on only the initial highly variable subband filter taps. This allowed for significant computational savings while maintaining good tracking performance compared to the standard SB-GGPAP structure, as shown in Sections 4.2.2 and 4.2.3. In Section 4.2.4 the computational advantage of the SB-GGPAP structure and its variants compared to their equivalent fullband structures was discussed.

Table 4.6 summarizes the main results of this chapter. The maximum improvement in ERLE for each AEC structure compared to the SB-AP structure is given based on the measured changing echo path conditions presented in Section 4.2.3. The corresponding computational complexity results are also presented in Table 4.6 and are applicable to the AEC structures in all sections under the conditions outlined in Section 4.2.4. The results for the subband structures are based on $K = 8$ $M = 4$ filter banks with $L_{pr} = 64$. Also, it should be noted that ERLE improvements of 6 dB or more attained by the proposed SB-GGPAP structure translates to a reduction by a factor of two or more in the output error signal. Furthermore, selected results from this chapter have been archived online [105] to demonstrate the perceived improvement in echo reduction that can be realized by the proposed SB-GGPAP structure.

AEC Structure	ERLE Improvement Compared to SB-AP (dB)	Complexity (Operations per input sample period)
SB-AP $P = 1$ $N_{sb_{1-4}} = 500$	–	8305
FB-GPAP $P = 1$ $N = 2000$	5	28012

AEC Structure	ERLE Improvement Compared to SB-AP (dB)	Complexity (Operations per input sample period)
SB-GPAP $P = 1$ $N_{sb_{1-4}} = 500$	9	23300
SB-GPAP $P = 1$ $N_{f_{1-4}} = 50$	10	9817
SB-AP $P = 3$ $N_{sb_{1-4}} = 500$	–	48468
FB-GPAP $P = 3$ $N = 2000$	7	68053
SB-GPAP $P = 3$ $N_{sb_{1-4}} = 500$	13	61386
SB-GPAP $P = 3$ $N_{f_{1-4}} = 50$	9	52353

Table 4.6: Summary of results for AEC in nonstationary environments.

Chapter 5

Subband Volterra Structure for Acoustic Echo Cancellation

Nonlinear electrodynamic loudspeaker harmonic distortion within a hands-free set can limit the performance of an EC, especially when it is driven at high signal levels as described in Section 2.2.1 of Chapter 2. To mitigate the performance impairments introduced by nonlinear loudspeaker distortion, an EC should take into consideration this nonlinearity in its model of the LEMS. As discussed in Section 2.2.1.1 of Chapter 2, ECs based on adaptive Volterra filters have been introduced to model the nonlinear harmonic distortion of small sized electrodynamic loudspeakers. In this chapter an efficient subband adaptive Volterra filter structure for AEC is proposed that exploits the time and frequency domain nature of electrodynamic loudspeaker harmonic distortion. It is shown that this structure can attain equivalent AEC performance compared to its fullband counterpart at a much reduced computational complexity, while outperforming standard linear ECs with only a minor increase in complexity.

The following sections are presented in this chapter. The subband Volterra AEC structure is described in detail in Section 5.1.1, along with a comprehensive computational complexity analysis. Section 5.2 presents experimental results for the proposed structure based on measured LEMS Volterra kernels as well as on directly measured

LEMS input signal data. A summary of the main findings and results follows in Section 5.3.

5.1 Subband Volterra Filter Structure

5.1.1 Description

To alleviate the computational requirements of a fullband second order adaptive Volterra filter applied to AEC, as outlined in Section 2.2.1.1 of Chapter 2, a subband approach with selective subband adaptive second order Volterra filtering is proposed. A block diagram of the subband Volterra filter (SBVF) structure implemented with oversampled GDFT modulated polyphase analysis and synthesis filter banks, as discussed in Section 2.1.1.2 of Chapter 2, is shown in Figure 5.1. The SBVF structure consists of K linear subband adaptive filters, $\hat{\mathbf{w}}_{1,i}(m)$, for modeling the linear part of the LEMS, and K quadratic subband adaptive filters, $\hat{\mathbf{w}}_{2,i}(m)$, for modeling the nonlinear portion of the LEMS, where $i = 1, \dots, K$ in all cases. The linear and quadratic subband filters provide estimates of the linear, $\mathbf{w}_{1,i}(m)$, and quadratic, $\mathbf{w}_{2,i}(m)$, LEMS subband Volterra kernels. The echo signal, $y(n)$, is determined based on a truncated second order Volterra series expansion of the input reference excitation signal, $x(n)$, according to (2.23). The reference input to the filter bank driving the nonlinear part of the SBVF structure is denoted by $x_2(n)$, and is determined by selecting D diagonals from the symmetric input signal matrix created from the double summation term in (2.23). These D input diagonals are selected with the main diagonal as the point of origin, in accordance with previously reported measurements of second order Volterra kernels having coefficients with the greatest energy lying along diagonals close to the main one [23], [86], [89]. This was also verified experimentally in [27]. The second order reference input vector for each diagonal, $\mathbf{x}_{2,j}(n)$, where

$1 \leq j \leq D$ is given by:

$$\begin{aligned} \mathbf{x}_{2,j}(n) = & \left[x(n)x(n-j+1), x(n-1)x(n-j), \right. \\ & \left. \dots, x(n-N_2+j)x(n-N_2+1) \right] \end{aligned} \quad (5.1)$$

where N_2 is the length of the equivalent fullband quadratic adaptive Volterra kernel and $D \leq N_2$. For each $\mathbf{x}_{2,j}(n)$ only the first element needs to be computed as the rest can be determined by the appropriate delay. The apparent doubling of the sampling rate of $x_2(n)$ does not require sampling $x(n)$ at a rate higher than its Nyquist frequency in order to model the nonlinear system, as discussed in [106].

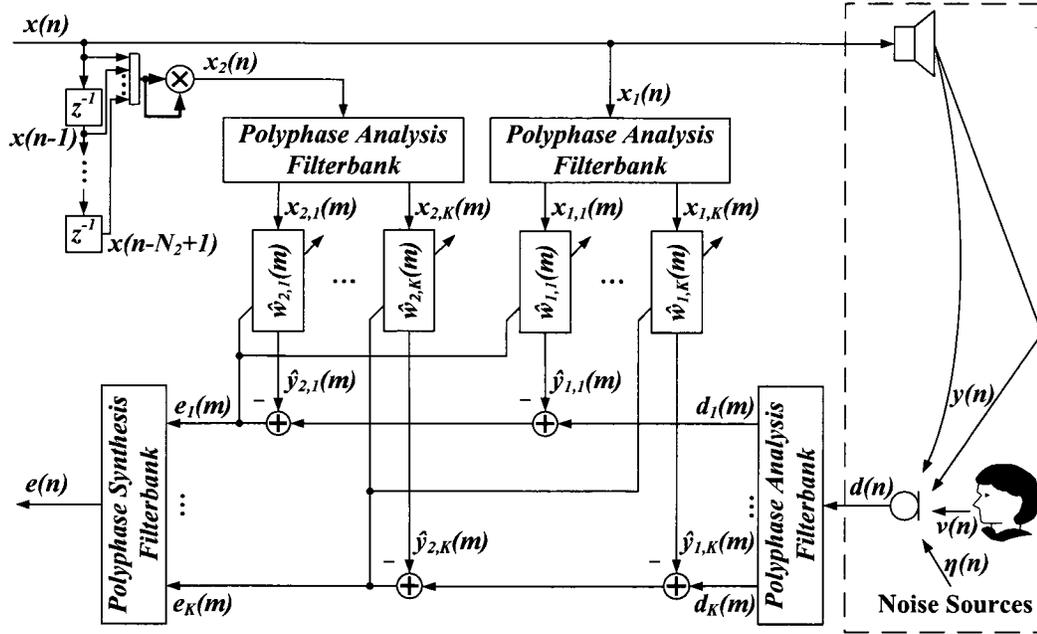


Figure 5.1: Block diagram of the subband adaptive second order Volterra filter structure.

The main feature of the SBVF structure is its ability to perform second order adaptive Volterra filtering in selective subbands. Ideally the nonlinear contribution of the microphone signal, $d(n)$, would be confined within a single subband, thus requiring nonlinear filtering in only that specific subband. In the case of loudspeaker distortion it is known that the nonlinearity generally occurs at low frequencies [78],

thus nonlinear processing should only be required in the low frequency subbands of the SBVF structure. With nonlinear filtering applied in the minimum number of subbands the lowest overall complexity for the structure will result. The explicit exploitation of the nonlinear portion of the LEMS's frequency response by the SBVF structure is an advantage compared to other subband Volterra structures where this is not taken into consideration, as in [90] and [106]. Thus, the SBVF structure can provide a more flexible and computationally trimmed alternative for nonlinear AEC. Also, by employing a subband structure and performing adaptive second order Volterra filtering on the decimated subband signals, AEC performance comparable to an equivalent fullband second order adaptive Volterra filter structure can be achieved with much reduced complexity.

The subband filters in the SBVF structure are adapted based on the M ($M < K$) fold downsampled and filtered input reference, $x_{1,i}(m)$ and $x_{2,i}(m)$, and input microphone, $d_i(m)$, signals. The subband error signals, $e_i(m)$, are computed based on the difference between $d_i(m)$ and the summation of the outputs of the associated linear and quadratic subband filters as depicted in Figure 5.1:

$$e_i(m) = d_i(m) - \hat{\mathbf{w}}_{1,i}^H(m)\mathbf{x}_{1,i}(m) - \hat{\mathbf{w}}_{2,i}^H(m)\mathbf{x}_{2,i}(m) \quad (5.2)$$

where $\hat{\mathbf{w}}_{1,i}(m)$ and $\hat{\mathbf{w}}_{2,i}(m)$ are the linear and quadratic subband adaptive Volterra filter coefficient vectors of length N_{1,sb_i} and N_{2,sb_i} respectively. Furthermore, $\mathbf{x}_{1,i}(m)$ and $\mathbf{x}_{2,i}(m)$ are the first and second order subband reference signal vectors with the same corresponding lengths as $\hat{\mathbf{w}}_{1,i}(m)$ and $\hat{\mathbf{w}}_{2,i}(m)$. The elements of $\mathbf{x}_{2,i}(m)$ correspond to the contribution of all $\mathbf{x}_{2,j}(n)$ to the i th subband at time m . The adaptation of $\hat{\mathbf{w}}_{1,i}(m)$ and $\hat{\mathbf{w}}_{2,i}(m)$ is performed using the NLMS algorithm as follows:

$$\hat{\mathbf{w}}_{1,i}(m+1) = \hat{\mathbf{w}}_{1,i}(m) + \frac{\mu_{1,i}}{\mathbf{x}_{1,i}^H(m)\mathbf{x}_{1,i}(m) + \delta_{1,i}} \mathbf{x}_{1,i}(m)e_i^*(m) \quad (5.3)$$

$$\hat{\mathbf{w}}_{2,i}(m+1) = \hat{\mathbf{w}}_{2,i}(m) + \frac{\mu_{2,i}}{\mathbf{x}_i^H(m)\mathbf{x}_i(m) + \delta_{2,i}} \mathbf{x}_{2,i}(m)e_i^*(m) \quad (5.4)$$

where $\mathbf{x}_i(m)$ is formed from the concatenation of $\mathbf{x}_{1,i}(m)$ and $\mathbf{x}_{2,i}(m)$. The adaptation step size parameters for $\hat{\mathbf{w}}_{1,i}(m)$ and $\hat{\mathbf{w}}_{2,i}(m)$ are denoted by $\mu_{1,i}$ and $\mu_{2,i}$ respectively, where $0 < \mu_{1,i}, \mu_{2,i} < 2$ is required for stability [86] and $[\cdot]^*$ denotes the complex conjugate operation. With the NLMS algorithm used to adapt the linear and quadratic subband Volterra filters the subband structure in Figure 5.1 will be referred to as the subband normalized least-mean-square Volterra filter (SB-NLMSVF) structure. The NLMS algorithm was chosen to adapt the subband filters due to its simplicity and ease of implementation [33]. Furthermore, the linear and quadratic subband adaptive Volterra filters were implemented in their fully coupled form as this resulted in improved steady-state performance compared to their separately adapted partially decoupled form.

5.1.2 Computational Complexity Analysis

The computational complexity of the SBVF structure shown in Figure 5.1 is comprised of the cost of implementing the four polyphase filter banks along with the linear and quadratic subband adaptive filtering expense. As in Section 4.1.2 of Chapter 4 the total computational complexity of the SBVF structure is given in terms of the number of equivalent real elementary mathematical operations per input sample period, where the cost of other operations such as memory transfers are not considered. The cost of one of the oversampled GDFT modulated polyphase analysis or synthesis filter banks is the same as reported in (4.26). The complexity of adapting the linear and quadratic subband filters with the NLMS algorithm, as described in (5.2) to (5.4), for real and complex input signal data is summarized in Tables 5.1 and 5.2 respectively. The computational cost is expressed in terms of the total number of real multiplication, addition, and division operations. The cost of computing $\mathbf{x}_i^H(m)\mathbf{x}_i(m)$ for the current iteration can be determined by adding the value from

the previous iteration to the squared magnitude of the current input sample, and subtracting the squared magnitude of the last input sample in the input signal vector from the previous iteration for both $\mathbf{x}_{1,i}(m)$ and $\mathbf{x}_{2,i}(m)$ [104]. This requires only 4 real multiplications and additions for real input data, and 8 real multiplications and additions for complex input data per iteration.

Term	\times	$+$	\div
$\hat{\mathbf{w}}_{1,i}^H(m)\mathbf{x}_{1,i}(m)$	N_{1,sb_i}	$N_{1,sb_i} - 1$	
$\hat{\mathbf{w}}_{2,i}^H(m)\mathbf{x}_{2,i}(m)$	N_{2,sb_i}	$N_{2,sb_i} - 1$	
$d_i(m) - \hat{\mathbf{w}}_{1,i}^H(m)\mathbf{x}_{1,i}(m) - \hat{\mathbf{w}}_{2,i}^H(m)\mathbf{x}_{2,i}(m)$		2	
$\mathbf{x}_i^H(m)\mathbf{x}_i(m)$	4	4	
$\mathbf{x}_i^H(m)\mathbf{x}_i(m) + \delta_{1,i}$		1	
$\mathbf{x}_i^H(m)\mathbf{x}_i(m) + \delta_{2,i}$		1	
$\mu_{1,i}e_i^*(m)/(\mathbf{x}_i^H(m)\mathbf{x}_i(m) + \delta_{1,i})$	1		1
$\mu_{2,i}e_i^*(m)/(\mathbf{x}_i^H(m)\mathbf{x}_i(m) + \delta_{2,i})$	1		1
$\mu_{1,i}e_i^*(m)/(\mathbf{x}_i^H(m)\mathbf{x}_i(m) + \delta_{1,i})\mathbf{x}_{1,i}(m)$	N_{1,sb_i}		
$\mu_{2,i}e_i^*(m)/(\mathbf{x}_i^H(m)\mathbf{x}_i(m) + \delta_{2,i})\mathbf{x}_{2,i}(m)$	N_{2,sb_i}		
$\hat{\mathbf{w}}_{1,i}(m) + \mu_{1,i}e_i^*(m)/(\mathbf{x}_i^H(m)\mathbf{x}_i(m) + \delta_{1,i})\mathbf{x}_{1,i}(m)$		N_{1,sb_i}	
$\hat{\mathbf{w}}_{2,i}(m) + \mu_{2,i}e_i^*(m)/(\mathbf{x}_i^H(m)\mathbf{x}_i(m) + \delta_{2,i})\mathbf{x}_{2,i}(m)$		N_{2,sb_i}	
Total per iteration	$2N_{1,sb_i} + 2N_{2,sb_i} + 6$	$2N_{1,sb_i} + 2N_{2,sb_i} + 6$	2

Table 5.1: Approximate computational complexity per iteration of the NLMS algorithm applied to subband adaptive second order Volterra filters for real input data in terms of real multiplications, additions, and divisions.

Term	\times	$+$	\div
$\hat{\mathbf{w}}_{1,i}^H(m)\mathbf{x}_{1,i}(m)$	$4N_{1,sb_i}$	$2(2N_{1,sb_i} - 1)$	
$\hat{\mathbf{w}}_{2,i}^H(m)\mathbf{x}_{2,i}(m)$	$4N_{2,sb_i}$	$2(2N_{2,sb_i} - 1)$	

$d_i(m) - \hat{\mathbf{w}}_{1,i}^H(m)\mathbf{x}_{1,i}(m) - \hat{\mathbf{w}}_{2,i}^H(m)\mathbf{x}_{2,i}(m)$		4	
$\mathbf{x}_i^H(m)\mathbf{x}_i(m)$	8	8	
$\mathbf{x}_i^H(m)\mathbf{x}_i(m) + \delta_{1,i}$		1	
$\mathbf{x}_i^H(m)\mathbf{x}_i(m) + \delta_{2,i}$		1	
$\mu_{1,i}e_i^*(m)/(\mathbf{x}_i^H(m)\mathbf{x}_i(m) + \delta_{1,i})$	2		1
$\mu_{2,i}e_i^*(m)/(\mathbf{x}_i^H(m)\mathbf{x}_i(m) + \delta_{2,i})$	2		1
$\mu_{1,i}e_i^*(m)/(\mathbf{x}_i^H(m)\mathbf{x}_i(m) + \delta_{1,i})\mathbf{x}_{1,i}(m)$	$4N_{1,sb_i}$	$2N_{1,sb_i}$	
$\mu_{2,i}e_i^*(m)/(\mathbf{x}_i^H(m)\mathbf{x}_i(m) + \delta_{2,i})\mathbf{x}_{2,i}(m)$	$4N_{2,sb_i}$	$2N_{2,sb_i}$	
$\hat{\mathbf{w}}_{1,i}(m) + \mu_{1,i}e_i^*(m)/(\mathbf{x}_i^H(m)\mathbf{x}_i(m) + \delta_{1,i})\mathbf{x}_{1,i}(m)$		$2N_{1,sb_i}$	
$\hat{\mathbf{w}}_{2,i}(m) + \mu_{2,i}e_i^*(m)/(\mathbf{x}_i^H(m)\mathbf{x}_i(m) + \delta_{2,i})\mathbf{x}_{2,i}(m)$		$2N_{2,sb_i}$	
Total per iteration	$8N_{1,sb_i} + 8N_{2,sb_i} + 12$	$8N_{1,sb_i} + 8N_{2,sb_i} + 10$	2

Table 5.2: Approximate computational complexity per iteration of the NLMS algorithm applied to subband adaptive second order Volterra filters for complex input data in terms of real multiplications, additions, and divisions.

The total computational cost for the SB-NLMSVF structure is comprised of the cost of the 4 oversampled GDFT modulated polyphase filter banks, the cost of computing $x_2(n)$, along with the NLMS adaptive filtering cost of the $K/2$ unique complex linear and quadratic subband filters operating at the reduced sampling rate. The cost of computing $x_2(n)$ is D real multiplications per sample period as discussed in Section 5.1.1. The overall computational complexity for the SB-NLMSVF structure is summarized in Table 5.3. The complexity of the SB-NLMSVF structure is dominated by the required number of real multiplication and addition operations and is $O\left((N_{1,sb_i} + N_{2,sb_i})\frac{K}{M}\right)$ for both when $N_{1,sb_i}, N_{2,sb_i} \gg L_{pr}$ and the length of each linear and quadratic subband adaptive filter is equal.

×	$\frac{4}{M} (L_{pr} + 4K \log_2 K + 4K) + D +$ $\frac{1}{M} \sum_{i=1}^{\frac{K}{2}} (8N_{1, sb_i} + 8N_{2, sb_i} + 12)$
+	$\frac{4}{M} (L_{pr} + 4K \log_2 K + 4K) +$ $\frac{1}{M} \sum_{i=1}^{\frac{K}{2}} (8N_{1, sb_i} + 8N_{2, sb_i} + 10)$
÷	$\frac{K}{M}$

Table 5.3: Approximate computational complexity per input sample period for the SB-NLMSVF AEC structure in terms of real multiplications, additions, and divisions.

5.2 Experimental Results

5.2.1 Methodology

Nonlinear AEC simulations based on synthetic and directly measured input signal data are presented in the following sections. To produce meaningful simulations based on synthetic input reference and microphone signals, the measured linear and quadratic LEMS Volterra kernels for the conference phone from Section 3.2 and 3.3 of Chapter 3 were used to create the input echo signal based on (2.23). Simulations based on synthetic input signals allows the performance of the AEC structures in question to be studied under controlled input conditions, with various levels of measured nonlinear loudspeaker distortion. The experimentally measured input reference and microphone data from the conference phone and smartphone, allows for the behaviour of the AEC structures to be investigated under real hands-free communication operating conditions.

The results presented in the subsequent sections for the subband AEC structures are based on the same filter bank settings given in Section 4.2.1 of Chapter 4, unless indicated otherwise. The results for the proposed SB-NLMSVF AEC structure are compared to the standard fullband normalized least-mean-square Volterra filter (FB-NLMSVF) structure outlined in [86], and to both a subband and fullband version of the NLMS algorithm [33] in terms of ERLE performance and computational

complexity for an 8 kHz sampling rate.

5.2.2 Simulation Results Based on Measured Volterra Kernels

5.2.2.1 White Noise Excitation Signal

This section presents the AEC performance of the SB-NLMSVF structure, discussed in Section 5.1.1, along with its fullband and linear only structure counterparts for a WGN input reference excitation signal. Simulations were performed and averaged over several trials for synthetic microphone signals created based on the measured linear and quadratic LEMS Volterra kernels from the conference phone configuration, with total lengths of $L_1 = 2000$ and $L_2(L_2 + 1)/2 = 1275$, respectively, where $L_2 = 50$. A 15 second WGN reference excitation signal, $x(n)$, with zero mean and unit variance was used along with the measured Volterra kernels to create the echo signal, $y(n)$, using (2.23). The input microphone signal, $d(n)$, was obtained by adding 15 seconds of uncorrelated WGN background noise, $\eta(n)$, to $y(n)$ to achieve a SNR of 30 dB. As discussed in Section 2.1.1 of Chapter 2 the local talker signal, $\nu(n)$, was set to 0 for the duration of the simulations since adaptation of the EC is only performed under quiet local talker conditions.

Figure 5.2 compares the ERLE performance of the SB-NLMSVF structure to a FB-NLMSVF structure as well as to a linear only FB-NLMS structure. The measured linear and quadratic LEMS Volterra kernels for the conference phone obtained under low volume conditions from Chapter 3 were used to create $y(n)$. The FB-NLMSVF structure was implemented based on the fully coupled NLMS adaptive second order Volterra filter described in (2.32) to (2.34). The linear and quadratic adaptive filters of the FB-NLMSVF structure were set to lengths of $N_1 = 2000$ and $N_2(N_2 + 1)/2 = 1275$ respectively, where $N_2 = 50$. The linear adaptive filter of the FB-NLMS structure, which arises from the FB-NLMSVF structure when $N_2 = 0$, was set to a length of

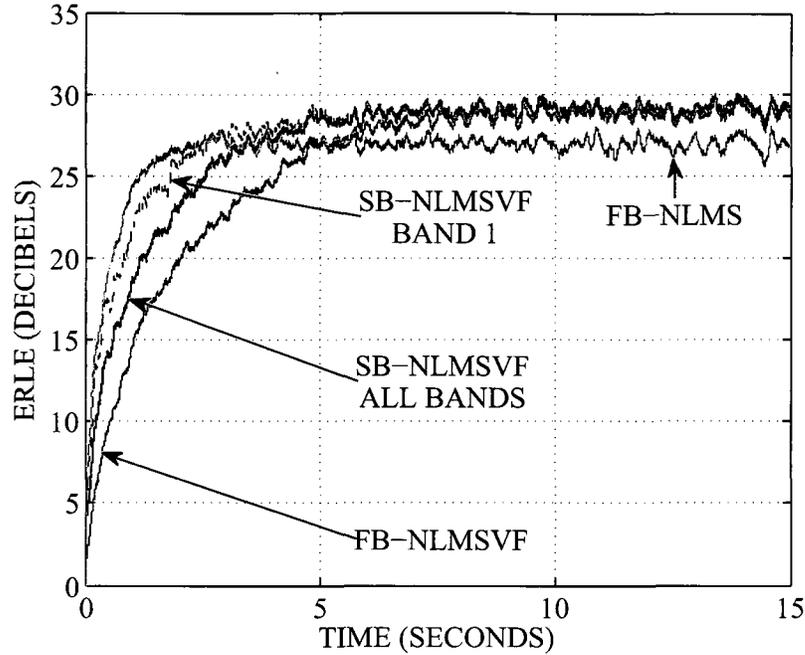


Figure 5.2: ERLE for FB-NLMS, FB-NLMSVF, and SB-NLMSVF with WGN input under low volume loudspeaker distortion conditions and a step size of 0.5 for both the linear and quadratic adaptive filters.

$N_1 = 2000$. The linear and quadratic adaptive filters of the SB-NLMSVF structure were set to $N_{1, sb_i} = N_1/M = 500$ and $N_{2, sb_i} = 120$ in all subbands respectively. A step size of 0.5 was used in adapting the linear and quadratic filters of all structures. Also, δ_1 and δ_2 were set to 0.00001 for the adaptation of the linear and quadratic filters of all structures with the same values used in all subbands. As shown in Figure 5.2, the SB-NLMSVF structure with second order adaptive Volterra filtering applied in all $K/2 = 4$ subbands (SB-NLMSVF ALL BANDS) achieves equivalent steady-state ERLE performance compared to the FB-NLMSVF structure. In addition, the SB-NLMSVF structure with second order adaptive Volterra filtering applied in all subbands converges faster than the FB-NLMSVF structure with up to 4 dB higher ERLE. This improvement in convergence can be attributed to the shorter length linear and quadratic subband adaptive filters. Furthermore, with second order adaptive Volterra filtering performed in only the 0–1 kHz subband (SB-NLMSVF BAND 1),

the SB-NLMSVF structure still achieves the same steady-state ERLE performance as with second order adaptive Volterra filtering in all subbands. This verifies that the SB-NLMSVF structure can successfully exploit the frequency domain nature of the loudspeaker distortion, by performing second order adaptive Volterra filtering in only the subbands where the distortion occurs, in this case the 0–1 kHz subband. Besides the inherent computational advantage of performing second order adaptive Volterra filtering in a subset of the total subbands, an increase in convergence rate also results for the SB-NLMSVF structure. With second order adaptive Volterra filtering performed in only the 0–1 kHz subband, the SB-NLMSVF structure attains up to 7 dB higher ERLE during initial convergence compared to the FB-NLMSVF structure and nearly equal convergence performance compared to the linear only FB-NLMS structure. The steady-state ERLE performance of the SB-NLMSVF structure with second order adaptive Volterra filtering applied in all and only the first subband, is up to 3 dB higher than the linear only FB-NLMS structure under these low volume loudspeaker distortion conditions. Also, the time domain nature of the loudspeaker distortion is exploited by the SB-NLMSVF structure through modeling only the highest energy second order Volterra kernel coefficients, and ignoring the insignificant coefficients as discussed in Section 5.1.1. These high energy second order Volterra kernel coefficients lie along its main and near adjacent diagonals as shown in Section 3.3 of Chapter 3, and allow the SB-NLMSVF structure to achieve equivalent ERLE performance compared to the FB-NLMSVF structure with less than the corresponding $(N_2(N_2 + 1)/2)/M = 319$ taps for the subband quadratic adaptive filters. In this case only $N_{2, sb_i} = 120$ taps were required by the subband quadratic adaptive filters to achieve the same steady-state ERLE performance as the FB-NLMSVF structure.

The ERLE performance of the SB-NLMSVF structure is compared to the FB-NLMSVF structure as well as to the linear only SB-NLMS and FB-NLMS structures in Figure 5.3 under moderate volume loudspeaker distortion conditions. The measured linear and quadratic LEMS Volterra kernels for the conference phone obtained

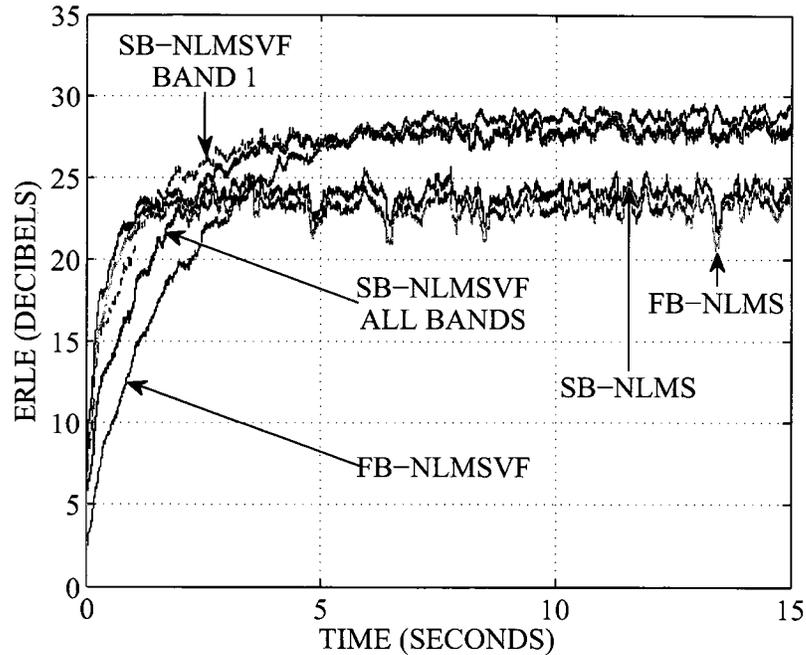


Figure 5.3: ERLE for FB-NLMS, SB-NLMS, FB-NLMSVF, and SB-NLMSVF with WGN input under moderate volume loudspeaker distortion conditions and a step size of 0.5 for both the linear and quadratic adaptive filters.

under moderate volume conditions from Chapter 3 were used to create $y(n)$. The simulation parameters for all structure were the same as previously used for the low volume distortion conditions, where the linear adaptive filters of the SB-NLMS structure were set to $N_{1, sb_i} = N_1/M = 500$ with a step size of 0.5 in all subbands. Also, it should be noted that the SB-NLMS structure can be realized from the SB-NLMSVF structure by setting $N_{2, sb_i} = 0$ in all subbands. Under these moderate signal level conditions the magnitude of the loudspeaker distortion has increased compared to the low volume case. This results in a greater disparity between the steady-state ERLE performance of the linear only SB-NLMS and FB-NLMS structures compared to the nonlinear SB-NLMSVF and FB-NLMSVF structures. The SB-NLMSVF structure with second order adaptive Volterra filtering in all (SB-NLMSVF ALL BANDS) and only the 0–1 kHz subband (SB-NLMSVF BAND 1) attain roughly equivalent steady-state ERLE performance compared to the FB-NLMSVF structure, while achieving

up to 3 and 7 dB higher ERLE respectively during initial convergence. This increase in convergence performance can again be attributed to the shorter length linear and quadratic subband adaptive filters. Compared to the linear only SB-NLMS and FB-NLMS structures, both versions of the SB-NLMSVF structure achieve up to 5 and 6 dB higher steady-state ERLE respectively. Also, the SB-NLMSVF structure with second order adaptive Volterra filtering in only the first subband is able to converge at approximately the same rate as the linear only SB-NLMS and FB-NLMS structures. The time and frequency domain nature of the loudspeaker distortion is again successfully exploited by the SB-NLMSVF structure in the same manner as under the previous simulation conditions. This is evident from the nearly equal steady-state ERLE performance of the SB-NLMSVF structure with second order adaptive Volterra filtering in only the 0–1 kHz subband with $N_{2, sb_1} = 120$ taps, compared to the FB-NLMSVF structure. The slight difference in steady-state ERLE between both versions of the SB-NLMSVF structure and the FB-NLMSVF structure is due to the non-modeling of the low energy quadratic subband Volterra kernel coefficients, which have correspondingly slightly higher magnitudes at these higher loudspeaker distortion conditions. The improvement in steady-state ERLE performance realized by the second order Volterra filter AEC structures compared to the linear only AEC structures, is due to their explicit modeling of the loudspeaker distortion.

Figure 5.4 compares the AEC structures from the previous simulation under the same conditions but with a step size of 1.0 used for the linear adaptive filters. The same trends in ERLE performance occur for the SB-NLMSVF structure compared to the other structures. All structures converge slightly faster to steady-state operation due to the higher step size used for the linear adaptive filters, and correspondingly achieve slightly lower steady-state ERLE due to the higher modeling error caused by the higher step size. Compared to the previous simulation the fullband structures experience a higher decrease in steady-state ERLE performance than the subband structures, due to the non-Wiener phenomenon that occurs for oversampled subband

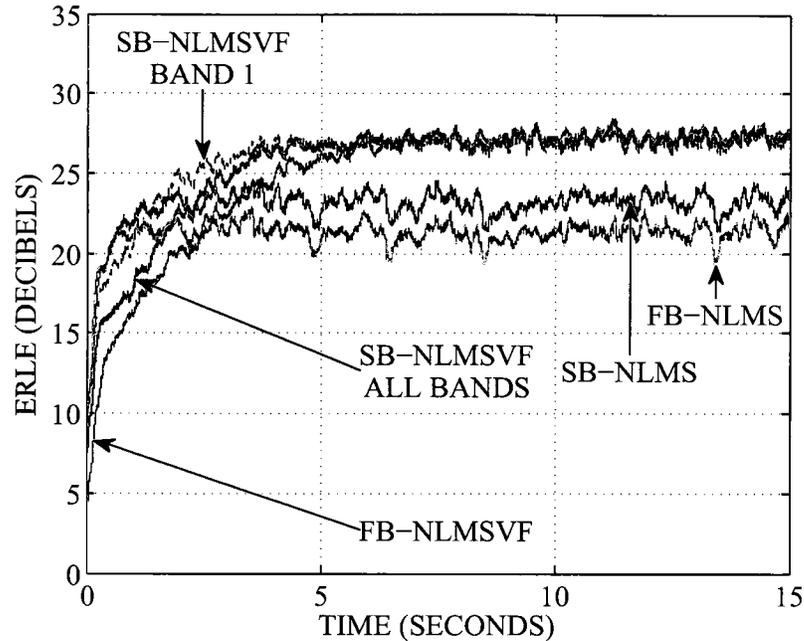


Figure 5.4: ERLE for FB-NLMS, SB-NLMS, FB-NLMSVF, and SB-NLMSVF with WGN input under moderate volume loudspeaker distortion conditions and a step size of 1.0 for the linear adaptive filters and 0.5 for the quadratic adaptive filters.

adaptive filters at sufficiently large step sizes [102]. This phenomenon was also observed and discussed in Section 4.2.2.1 of Chapter 4. Under these conditions the SB-NLMSVF structure with second order adaptive Volterra filtering in all (SB-NLMSVF ALL BANDS) and only the 0–1 kHz subband (SB-NLMSVF BAND 1) attains the same steady-state performance as the FB-NLMSVF structure, while achieving up to 2 and 5 dB higher ERLE respectively during initial convergence. In this case the non-Wiener phenomenon compensates for the undermodeling of the quadratic subband Volterra kernels in the SB-NLMSVF structure, allowing it to attain the same steady-state ERLE performance as the FB-NLMSVF structure. Both versions of the SB-NLMSVF structure achieve up to 4 and 6 dB higher steady-state ERLE compared to the linear only SB-NLMS and FB-NLMS structures respectively. Again, the SB-NLMSVF structure with second order adaptive Volterra filtering in only the first subband converges at approximately the same rate as the SB-NLMS and FB-NLMS

structures. The SB-NLMSVF structure with second order adaptive Volterra filtering in all subbands lags in initial convergence by up to 4 dB lower ERLE compared to the SB-NLMS and FB-NLMS structures, which is an improvement of 2 dB compared to the slower converging FB-NLMSVF structure.

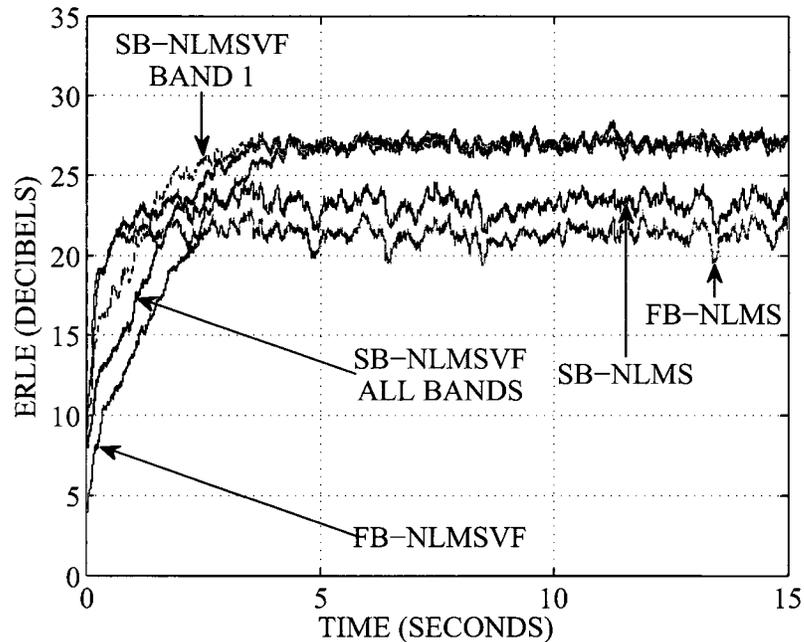


Figure 5.5: ERLE for FB-NLMS, SB-NLMS, FB-NLMSVF, and SB-NLMSVF with WGN input under moderate volume loudspeaker distortion conditions and a step size of 1.0 for both the linear and quadratic adaptive filters.

Under the previous simulation settings, Figure 5.5 compares the AEC structures with a higher step size of 1.0 used for the quadratic adaptive filters. The SB-NLMSVF and FB-NLMSVF structures experience a slight increase in convergence rate and a corresponding slight decrease in steady-state ERLE due to the higher step size. However, both the SB-NLMSVF and FB-NLMSVF structures experience approximately the same slight decrease in steady-state ERLE, which suggests that the non-Wiener phenomenon is not as prevalent for the quadratic subband adaptive filters as for the linear subband adaptive filters shown in Figure 5.4. Under these conditions the SB-NLMSVF structure attains roughly the same ERLE performance with respect to the

other AEC structures as for the previous simulation conditions.

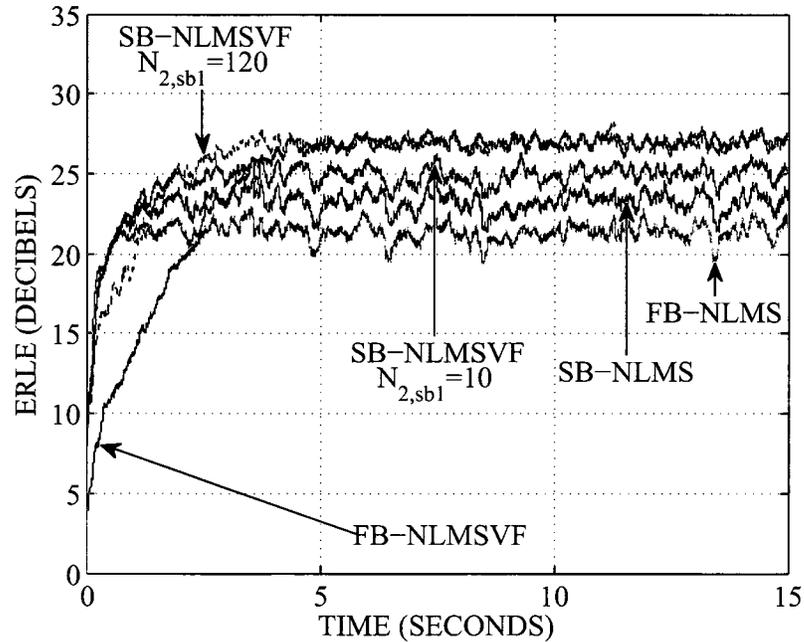


Figure 5.6: ERLE for FB-NLMS, SB-NLMS, FB-NLMSVF, and SB-NLMSVF with WGN input under moderate volume loudspeaker distortion conditions and a step size of 1.0 for both the linear and quadratic adaptive filters and a decreasing $N_{2, sb1}$.

The impact on ERLE performance due to a high degree of undermodeling of the loudspeaker nonlinearity by the SB-NLMSVF structure is investigated in Figure 5.6. The previous simulation settings were used, including the SB-NLMSVF structure with second order adaptive Volterra filtering in only the 0–1 kHz subband with $N_{2, sb1} = 10$ (SB-NLMSVF $N_{2, sb1} = 10$). Even with a very small number of quadratic subband adaptive filter taps the SB-NLMSVF structure is able to partially model the loudspeaker distortion, and provide an improvement in steady-state ERLE performance by up to 2 and 4 dB respectively compared to the linear only SB-NLMS and FB-NLMS structures. The very short length quadratic subband adaptive filter of the SB-NLMSVF structure also allows for convergence at the same rate as the linear only SB-NLMS and FB-NLMS structures. By exploiting the time domain nature of the second order Volterra kernel coefficients through modeling only some of the high

energy coefficients along its main diagonal, a considerable increase in ERLE performance can be realized by the SB-NLMSVF structure compared to the linear only AEC structures.

5.2.2.2 Speech Excitation Signal

In this section the performance of the SB-NLMSVF structure along with its fullband and linear only structure counterparts are investigated for a speech excitation signal. A 5 second reference speech signal, $x(n)$, was used in conjunction with the measured linear and quadratic LEMS Volterra kernels for the conference phone under high volume conditions to create the echo signal, $y(n)$, in the same manner as described in Section 5.2.2.1. The speech signal consisted of different speech utterances from the TIMIT speech data base [99] for the same male speaker concatenated together. The input microphone signal, $d(n)$, was obtained by adding 5 seconds of uncorrelated WGN background noise, $\eta(n)$, to $y(n)$ to achieve a SNR of 40 dB. As in Section 5.2.2.1 the local talker signal, $\nu(n)$, was set to 0 for the duration of the simulations.

Figure 5.7 displays the ERLE performance of the SB-NLMSVF structure compared to the FB-NLMSVF structure as well as to the linear only SB-NLMS and FB-NLMS structures. The linear and quadratic adaptive filters of the FB-NLMSVF structure were set to lengths of $N_1 = 2000$ and $N_2(N_2 + 1)/2 = 1275$ respectively, where $N_2 = 50$. The linear adaptive filter of the FB-NLMS structure was set to a length of $N_1 = 2000$, and correspondingly the linear adaptive filters of the SB-NLMS structure were set to $N_{1, sb_i} = N_1/M = 500$ in all subbands. The linear and quadratic adaptive filters of the SB-NLMSVF structure were set to $N_{1, sb_i} = N_1/M = 500$ and $N_{2, sb_i} = 120$ in all subbands respectively. A step size of 0.5 was used in adapting the linear and quadratic filters of all structures. Also, δ_1 and δ_2 were set to 0.00001 for the adaptation of the linear and quadratic filters of all structures with the same values used in all subbands. During initial convergence the SB-NLMS structure as well as the SB-NLMSVF structure with second order adaptive Volterra filtering in all

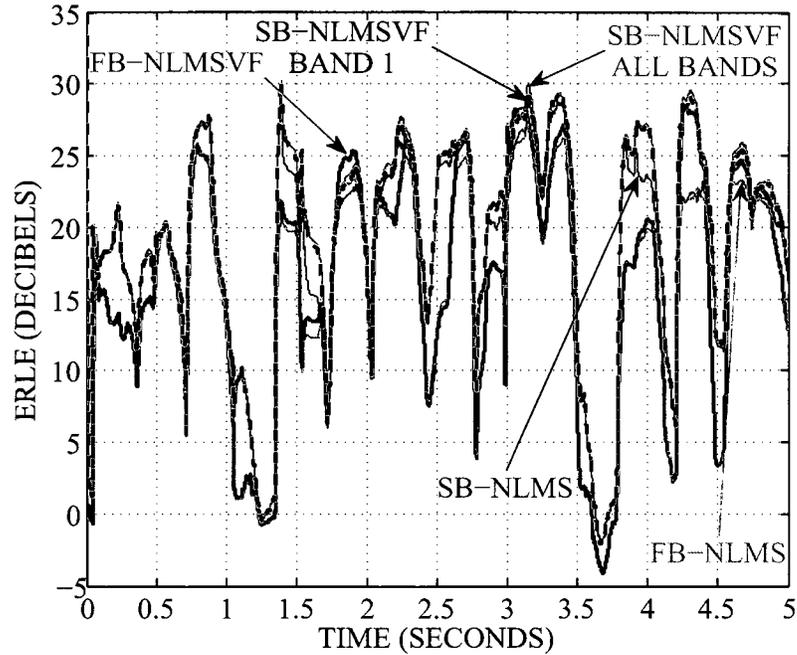


Figure 5.7: ERLE for FB-NLMS, SB-NLMS, FB-NLMSVF, and SB-NLMSVF with speech input under high volume loudspeaker distortion conditions and a step size of 0.5 for both the linear and quadratic adaptive filters.

(SB-NLMSVF ALL BANDS) and only the 0–1 kHz subband (SB-NLMSVF BAND 1) achieve up to 7 dB higher ERLE compared to the fullband structures. Again, the fast convergence of the subband structures can be attributed to the more spectrally flat subband signals and their shorter adaptive filters operating at the reduced sampling rate. As the frequency response of the quadratic Volterra kernel is concentrated below approximately 1 kHz (see Section 3.3 of Chapter 3) the quadratic portion of $y(n)$ will be highest during periods of speech with frequency content in this range. During these low frequency speech periods the Volterra AEC structures outperform the linear only AEC structures by up to 2 and 7 dB, as shown at the 3 and 4.25 second marks respectively in Figure 5.7. Also, during periods of minimum ERLE the subband structures provide up to 10 dB higher ERLE compared to the fullband structures, as seen at the 2.5 and 4.5 second marks of the simulation. The SB-NLMSVF structure is again able to exploit the frequency response nature of

the loudspeaker distortion. This is evident by the equivalent ERLE performance of the structure with second order adaptive Volterra filtering in only the 0–1 kHz subband compared to the performance with second order adaptive Volterra filtering in all subbands. Furthermore, the time domain nature of the loudspeaker distortion is again successfully exploited as only $N_{2, sb_i} = 120$ taps were required by the quadratic subband adaptive filters of the SB-NLMSVF structure, instead of the corresponding $(N_2(N_2 + 1)/2)/M = 319$ taps, to achieve equal or better ERLE performance than the FB-NLMSVF structure during periods of high echo path nonlinearity. As in Section 5.2.2.1, the quadratic subband adaptive filter coefficients of the SB-NLMSVF structure correspond to the high energy coefficients of the main and near adjacent diagonals of the fullband quadratic Volterra kernel.

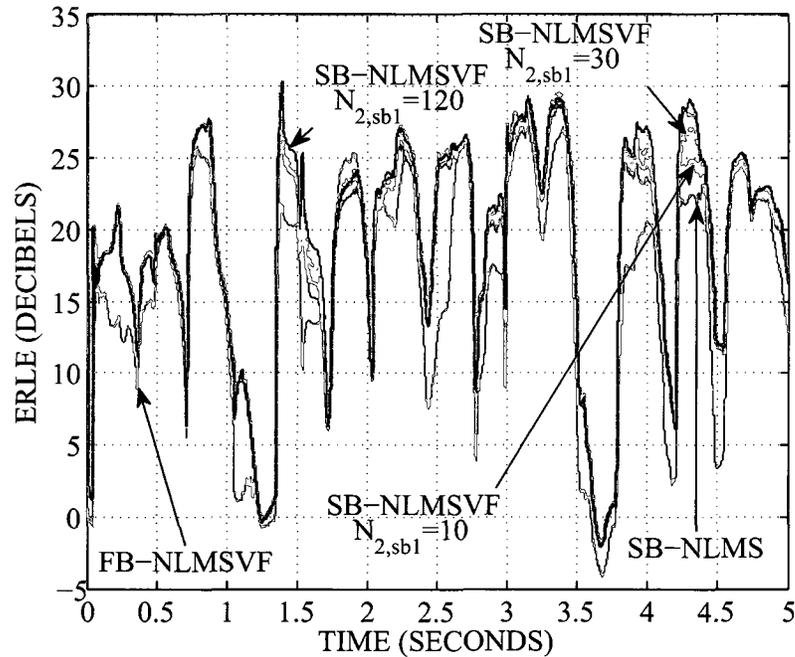


Figure 5.8: ERLE for SB-NLMS, FB-NLMSVF, and SB-NLMSVF with speech input under high volume loudspeaker distortion conditions and a step size of 0.5 for both the linear and quadratic adaptive filters and a decreasing N_{2, sb_1} .

The effect on the ERLE performance of the SB-NLMSVF structure due to a high amount of quadratic subband Volterra kernel undermodeling for a speech excitation

signal is investigated in Figure 5.8. The simulation was performed under the previous conditions for the SB-NLMSVF structure with second order adaptive Volterra filtering in only the 0–1 kHz subband with $N_{2, sb_1} = 120$, along with the FB-NLMSVF and SB-NLMS structures. As well, results for the SB-NLMSVF structure with second order adaptive Volterra filtering in only the 0–1 kHz subband with $N_{2, sb_1} = 30$ and $N_{2, sb_1} = 10$ coefficients are shown. With a very small number of quadratic subband adaptive filter taps the SB-NLMSVF structure is able to partially model the loudspeaker distortion, and achieve improved ERLE performance by up to 5 and 3 dB for $N_{2, sb_1} = 30$ and $N_{2, sb_1} = 10$ respectively, compared to the linear only SB-NLMS structure as seen at the 4.25 second mark of the simulation. Thus, by modeling only a small amount of the high amplitude quadratic Volterra kernel coefficients, a significant improvement in ERLE performance is realized by the SB-NLMSVF structure during periods of high echo path distortion compared to its linear only counterpart. With $N_{2, sb_1} = 30$ the performance of the SB-NLMSVF structure lags by at most 2 dB of ERLE compared to its performance when $N_{2, sb_1} = 120$. Compared to the FB-NLMSVF structure, the SB-NLMSVF structure with $N_{2, sb_1} = 30$ and $N_{2, sb_1} = 10$ achieves up to only 1 and 3 dB less ERLE respectively, during periods of high echo path distortion as seen near the 4.25 second mark in Figure 5.8. The SB-NLMSVF structure with the shorter quadratic subband adaptive filter lengths also achieves roughly the same initial convergence performance as with $N_{2, sb_1} = 120$.

5.2.3 Results Based on Measured Reference and Input Data

The performance of the SB-NLMSVF structure, along with its fullband and linear only structure counterparts, are investigated based on experimentally measured reference and microphone signals in this section. The recorded reference and microphone signals from the conference phone and smartphone under high volume speech excitation signal conditions, as discussed in Section 3.1 of Chapter 3, were used in the subsequent simulations.

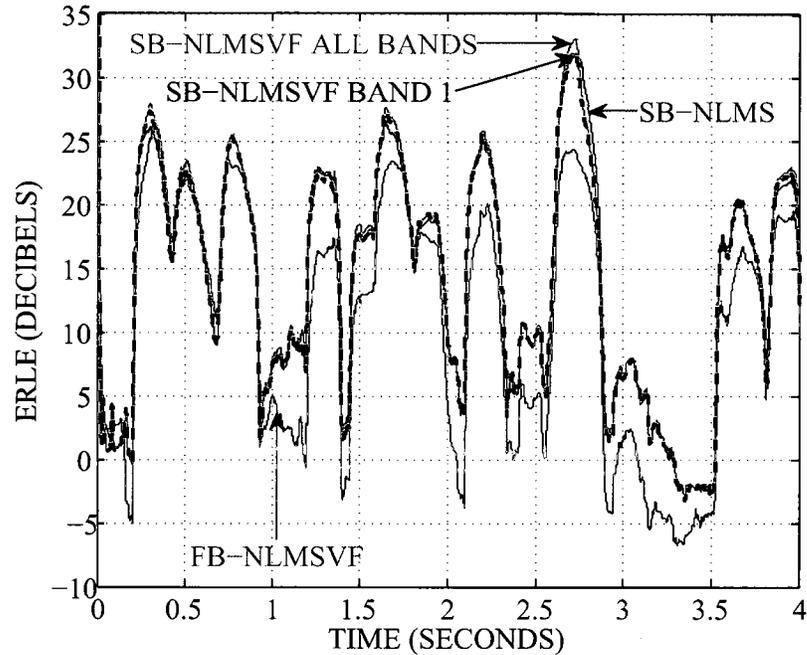


Figure 5.9: ERLE from 0 to 4 seconds for SB-NLMS, FB-NLMSVF, and SB-NLMSVF with measured speech input from the conference phone under high volume loudspeaker distortion conditions and a step size of 0.5 for both the linear and quadratic adaptive filters.

The ERLE performance of the SB-NLMSVF structure compared to the FB-NLMSVF structure as well as to the linear only SB-NLMS structure is displayed in Figures 5.9 and 5.10 for the conference phone. The results are presented in 0–4 and 12–15 second simulation sections in Figures 5.9 and 5.10 respectively, in order to differentiate between the performance of the AEC structures more easily during initial convergence and steady-state operation. The memory lengths of the adaptive filters in the FB-NLMSVF and SB-NLMS structures were the same as in Section 5.2.2.2. The length of the linear subband adaptive filters of the SB-NLMSVF structure were also the same as in Section 5.2.2.2, with the length of the quadratic subband adaptive filters set to $N_{2, sb_i} = 120$ in all subbands where second order adaptive Volterra filtering was performed. A step size of 0.5 was used in adapting the linear and quadratic filters of all structures. Also, δ_1 and δ_2 were set to 0.00001 for the adaptation of

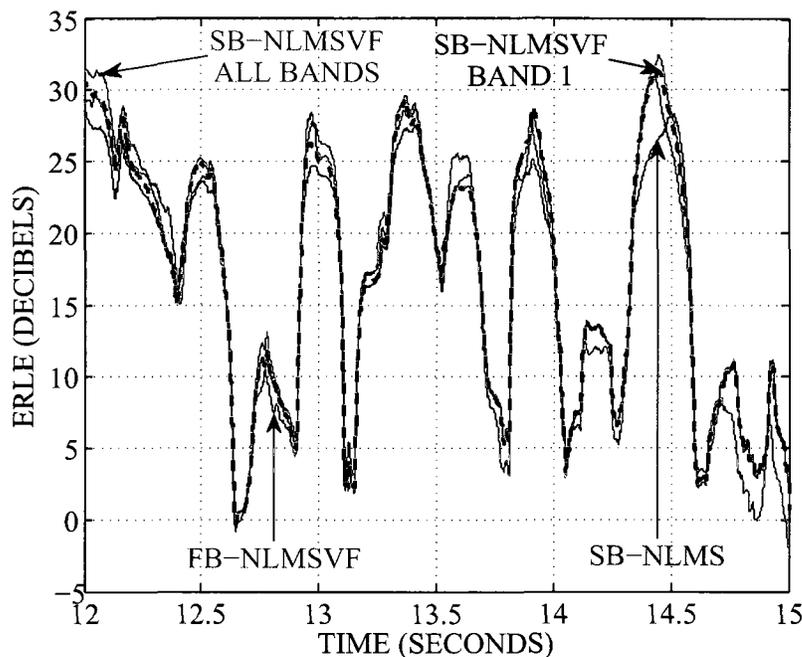


Figure 5.10: ERLE from 12 to 15 seconds for SB-NLMS, FB-NLMSVF, and SB-NLMSVF with measured speech input from the conference phone under high volume loudspeaker distortion conditions and a step size of 0.5 for both the linear and quadratic adaptive filters.

the linear and quadratic filters of all structures with the same values used in all subbands. As a result of the more spectrally uniform subband signals and their shorter adaptive filters the subband structures attain up to 7 dB higher ERLE compared to the FB-NLMSVF structure during initial convergence, as shown in Figure 5.9. During periods of high loudspeaker distortion in the echo path, as seen near the 12, 13, and 14.5 second mark in Figure 5.10, the SB-NLMSVF structure with second order adaptive Volterra filtering in all (SB-NLMSVF ALL BANDS) and only the 0–1 kHz subband (SB-NLMSVF BAND 1) outperforms the linear only SB-NLMS structure by up to 7 dB of ERLE. With second order adaptive Volterra filtering applied in only the 0–1 kHz subband, the SB-NLMSVF structure performs nearly identically compared to its performance with second order adaptive Volterra filter in all subbands. Thus, it is evident that the majority of distortion within the echo path is

due to low frequency speech occurring in the 0–1 kHz region. This low frequency echo path distortion corresponds well with the frequency response of the measured quadratic Volterra kernel for the conference phone shown in Figure 3.6(b), and to its harmonic distortion shown in Figure 3.10. The SB-NLMSVF structure with second order adaptive Volterra filter in all and only the 0–1 kHz subbands also performs as well as or slightly better than the FB-NLMSVF structure during the periods of high echo path nonlinearity. Also, during periods of minimum ERLE the subband structures provide up to 5 dB higher ERLE compared to the FB-NLMSVF structure, as seen between the 14.5 and 15 second marks of the simulation. Under these measured AEC conditions the SB-NLMSVF structure is able to take advantage of both the time and frequency domain nature of the loudspeaker distortion by performing second order adaptive Volterra filtering in only the low frequency subbands, and by modeling only the high amplitude quadratic Volterra kernel coefficients.

Figure 5.11 compares the ERLE performance of the SB-NLMSVF structure with second order adaptive Volterra filtering in only the 0–1 kHz subband to the FB-NLMSVF and SB-NLMS structures, as the length of the quadratic subband adaptive filters decreases during the 12–15 second simulation period for the conference phone. The previous simulation settings were used with the addition of the quadratic subband adaptive filters set to $N_{2, sb_1} = 30$ to study the impact of a high degree of undermodeling of the quadratic subband Volterra kernels. The performance of the SB-NLMSVF structure with $N_{2, sb_1} = 30$ during the 0–4 second simulation period was nearly identical to that obtained with $N_{2, sb_1} = 120$, as shown in Figure 5.9, and is therefore omitted. With only $N_{2, sb_1} = 30$ quadratic subband adaptive filter taps the SB-NLMSVF structure is able to partially model the loudspeaker distortion and perform within 2 dB of ERLE compared to $N_{2, sb_1} = 120$, as seen during the periods of high echo path distortion that occur near the 12, 14, and 14.5 second mark. Compared to the linear only SB-NLMS structure, ERLE improvements up to 5 dB are attained by the SB-NLMSVF structure during the periods of high loudspeaker

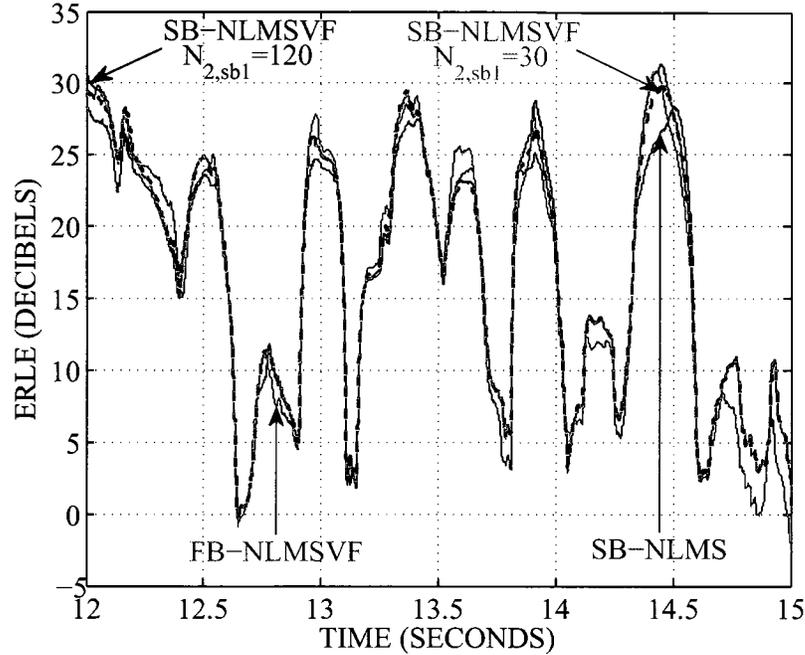


Figure 5.11: ERLE for SB-NLMS, FB-NLMSVF, and SB-NLMSVF with measured speech input from the conference phone under high volume loudspeaker distortion conditions and a step size of 0.5 for both the linear and quadratic adaptive filters and a decreasing $N_{2,sub1}$.

distortion with $N_{2,sub1} = 30$. As in Section 5.2.2.2, improved ERLE performance compared to the linear SB-NLMS structure is realized by the SB-NLMSVF structure by modeling only a small fraction of the high energy quadratic subband Volterra kernel coefficients. Furthermore, during these periods of high echo path nonlinearity the SB-NLMSVF structure with $N_{2,sub1} = 30$ performs about as well as the FB-NLMSVF structure.

The impact on ERLE performance of the SB-NLMSVF structure by applying second order adaptive Volterra filtering on a per subband basis is investigated in Figure 5.12, for an increasing number of subbands with the same OSR during the 12–15 second simulation period for the conference phone. The same filter bank settings as those used in Section 4.2.2.1 of Chapter 4 for Figure 4.11 were used for implementing

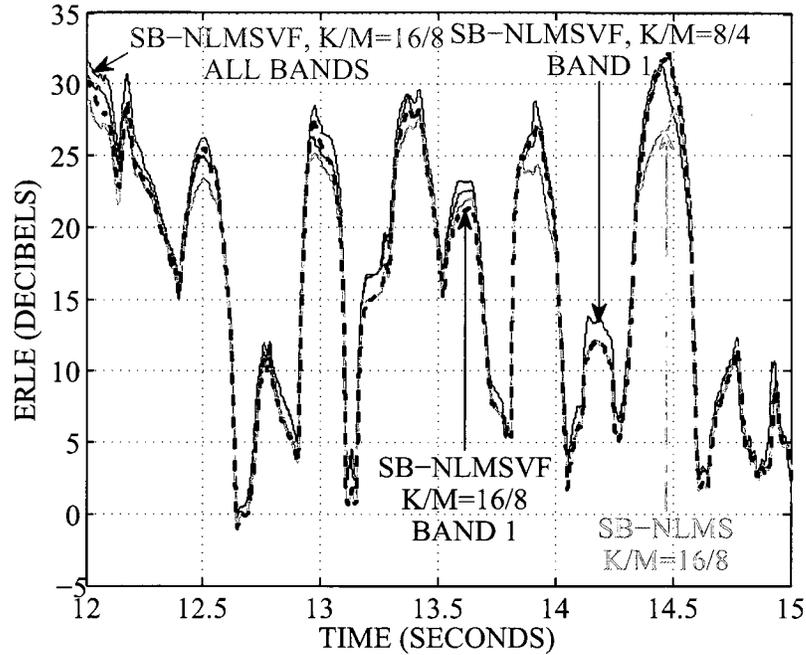


Figure 5.12: ERLE from 12 to 15 seconds for SB-NLMSVF with measured speech input from the conference phone under high volume loudspeaker distortion conditions and an increasing number of subbands for the same OSR.

the SB-NLMSVF and SB-NLMS structures. The linear adaptive filters of the SB-NLMSVF and SB-NLMS structures based on the $K = 16$ $M = 8$ filter banks were set to a length of $N_{1, sb_i} = 250$ in all subbands, while the linear adaptive filters of the $K = 8$ $M = 4$ SB-NLMSVF structure were set to $N_{1, sb_i} = 500$ in all subbands. The quadratic adaptive filters of the SB-NLMSVF structure based on the $K = 16$ $M = 8$ filter banks were set to a length of $N_{2, sb_i} = 30$ in all subbands and only the 0–500 Hz subband. Furthermore, the quadratic adaptive filters of the $K = 8$ $M = 4$ SB-NLMSVF structure were set to $N_{2, sb_i} = 120$ in only the 0–1 kHz subband. A step size of 0.5 was used in adapting the linear and quadratic filters of all structures. Also, δ_1 and δ_2 were set to 0.00001 for the adaptation of the linear and quadratic filters of all structures with the same values used in all subbands. During periods of high loudspeaker distortion the $K = 16$ SB-NLMSVF structure with quadratic filtering in only the 0–500 Hz subband (SB-NLMSVF $K/M=16/8$ BAND 1) attains

approximately equivalent ERLE performance as the same structure with quadratic filtering in all subbands (SB-NLMSVF $K/M=16/8$ ALL BANDS), as well as the $K = 8$ SB-NLMSVF structure with quadratic filtering in only the 0–1 kHz subband (SB-NLMSVF $K/M=8/4$ BAND 1), as seen near the 12, 14, and 14.5 second marks. This verifies that the majority of the loudspeaker distortion in the conference phone occurs for frequencies below 500 Hz, as shown in Figure 3.10. Furthermore, during the same periods of loudspeaker distortion the $K = 16$ SB-NLMSVF structure with quadratic filtering in only the 0–500 Hz subband outperforms the linear only $K = 16$ SB-NLMS structure by up to 7 dB of ERLE. Implementing the SB-NLMSVF structure with a greater number of subbands for the same OSR allows quadratic Volterra filtering to be performed in subbands with a finer frequency range, which can better match the frequency range of the loudspeaker distortion. Also, implementing the SB-NLMSVF structure with a greater number of subbands for the same OSR results in further computational cost reduction, due to the higher decimation ratio and shorter subband adaptive filters, at the expense of a higher filter bank group delay as discussed in Section 2.1.1.2 of Chapter 2. Although not shown, it should also be noted that all structures achieve approximately the same rate of initial convergence.

The ERLE performance of the SB-NLMSVF structure compared to the FB-NLMSVF structure, as well as to the linear only SB-NLMS structure, is displayed in Figures 5.13 and 5.14 for the smartphone under the same simulation settings as used for the conference phone in Figure 5.9 and 5.10. The subband structures attain up to 10 dB higher ERLE compared to the FB-NLMSVF structure during initial convergence, as shown in Figure 5.13. During periods of high loudspeaker distortion in the echo path, as seen near the 12 and 14.5 second mark in Figure 5.14, the SB-NLMSVF structure with second order adaptive Volterra filtering in all (SB-NLMSVF ALL BANDS) and only the 1–2 and 2–3 kHz subbands (SB-NLMSVF BANDS 2-3) outperforms the linear only SB-NLMS structure by up to 8 dB of ERLE. With second order adaptive Volterra filtering applied in only the 1–2 and 2–3 kHz subbands,

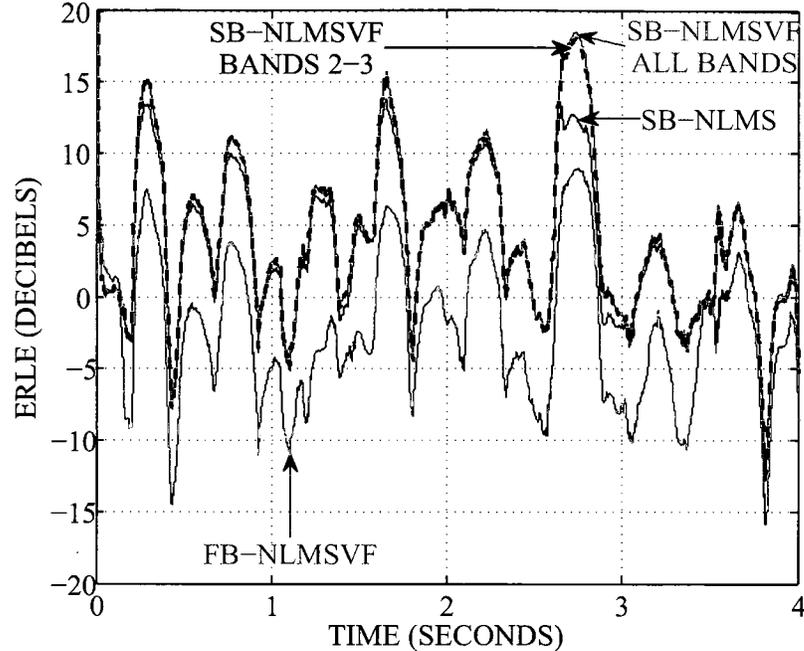


Figure 5.13: ERLE from 0 to 4 seconds for SB-NLMS, FB-NLMSVF, and SB-NLMSVF with measured speech input from the smartphone under high volume loud-speaker distortion conditions and a step size of 0.5 for both the linear and quadratic adaptive filters.

the SB-NLMSVF structure performs nearly identically compared to its performance with second order adaptive Volterra filter in all subbands. Thus, it is evident that the majority of distortion within the echo path is due to speech occurring in the 1–3 kHz region. This localized region of echo path distortion corresponds well with the frequency response of the measured quadratic Volterra kernel for the smartphone shown in Figure 3.7(b) and to its harmonic distortion shown in Figure 3.11. The SB-NLMSVF structure with second order adaptive Volterra filter in all and only the 1–2 and 2–3 kHz subbands also performs as well as or slightly better than the FB-NLMSVF structure during the periods of high echo path nonlinearity. Also, during periods of minimum ERLE the subband structures provide up to 3 dB higher ERLE compared to the FB-NLMSVF structure, as seen around the 14 and 15 second marks of the simulation. Under these measured AEC conditions the SB-NLMSVF structure

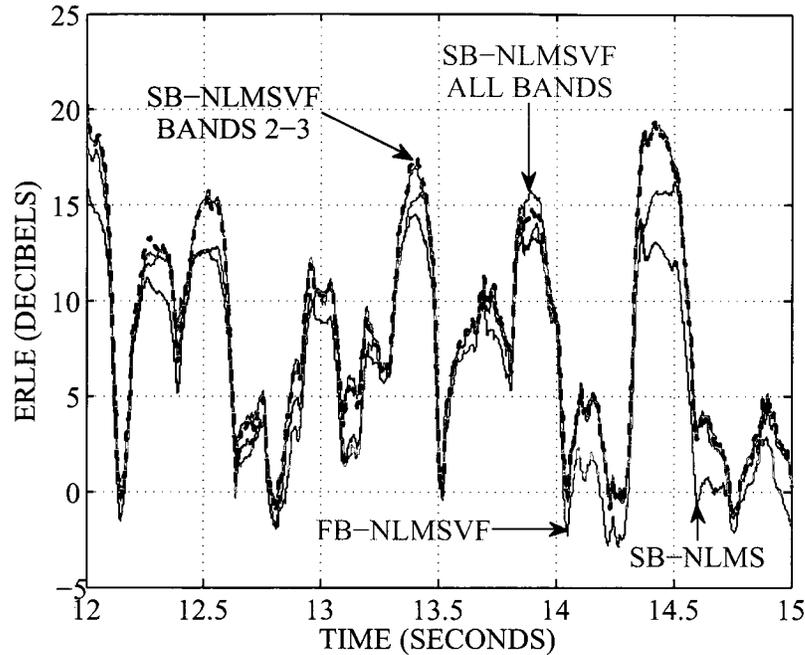


Figure 5.14: ERLE from 12 to 15 seconds for SB-NLMS, FB-NLMSVF, and SB-NLMSVF with measured speech input from the smartphone under high volume loud-speaker distortion conditions and a step size of 0.5 for both the linear and quadratic adaptive filters.

is able to take advantage of both the time and frequency domain nature of the loud-speaker distortion in the smartphone, by performing second order adaptive Volterra filtering in only the necessary subbands and by modeling only the high amplitude quadratic Volterra kernel coefficients.

Figure 5.15 compares the ERLE performance of the SB-NLMSVF structure with second order adaptive Volterra filter in only the 1–2 and 2–3 kHz subbands to the FB-NLMSVF and SB-NLMS structures, as the length of the quadratic subband adaptive filters decreases during the 12–15 second simulation period for the smartphone. The previous simulation settings were used with the addition of the quadratic subband adaptive filters set to $N_{2, sb_{2-3}} = 10$, to study the impact of a high degree of undermodeling of the quadratic subband Volterra kernels. The performance of the SB-NLMSVF structure with $N_{2, sb_{2-3}} = 10$ during the 0–4 second simulation period

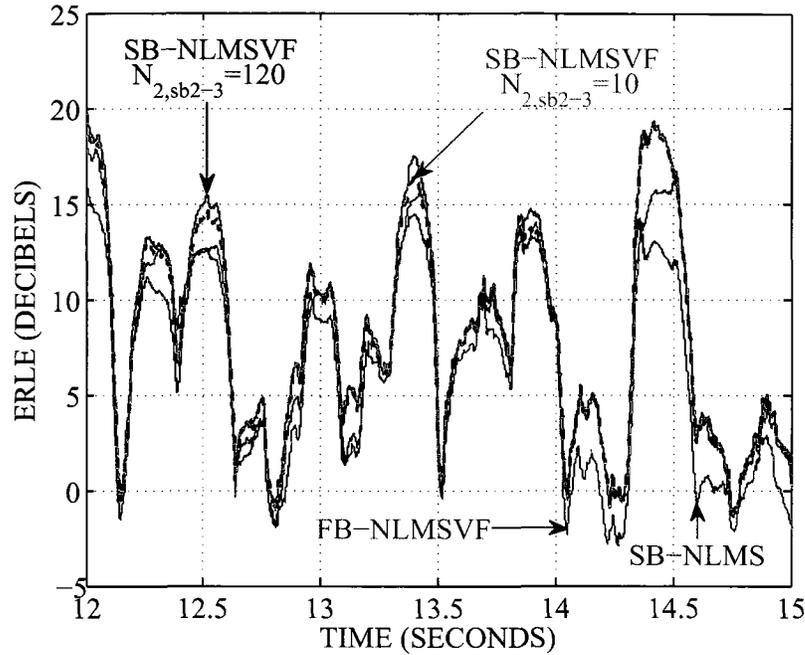


Figure 5.15: ERLE from 12 to 15 seconds for SB-NLMS, FB-NLMSVF, and SB-NLMSVF with measured speech input from the smartphone under high volume loudspeaker distortion conditions and a step size of 0.5 for both the linear and quadratic adaptive filters and a decreasing $N_{2, sb_{2-3}}$.

was nearly identical to that obtained with $N_{2, sb_{2-3}} = 120$, as shown in Figure 5.13, and is therefore omitted. With only $N_{2, sb_{2-3}} = 10$ quadratic subband adaptive filter taps the SB-NLMSVF structure is able to partially model the loudspeaker distortion, and perform within 2 dB of ERLE compared to $N_{2, sb_{2-3}} = 120$ during the periods of high echo path distortion, as seen near the 12 and 14.5 second marks. Compared to the linear only SB-NLMS structure, ERLE improvements up to 8 dB are attained by the SB-NLMSVF structure during the periods of high loudspeaker distortion with $N_{2, sb_{2-3}} = 10$. As in Section 5.2.2.2, improved ERLE performance compared to the linear SB-NLMS structure is realized by the SB-NLMSVF structure through modeling only a few of the high energy quadratic subband Volterra kernel coefficients. Furthermore, during the periods of high echo path nonlinearity the SB-NLMSVF structure with $N_{2, sb_{2-3}} = 10$ performs as well as or better than the FB-NLMSVF structure by

up to 4 dB of ERLE.

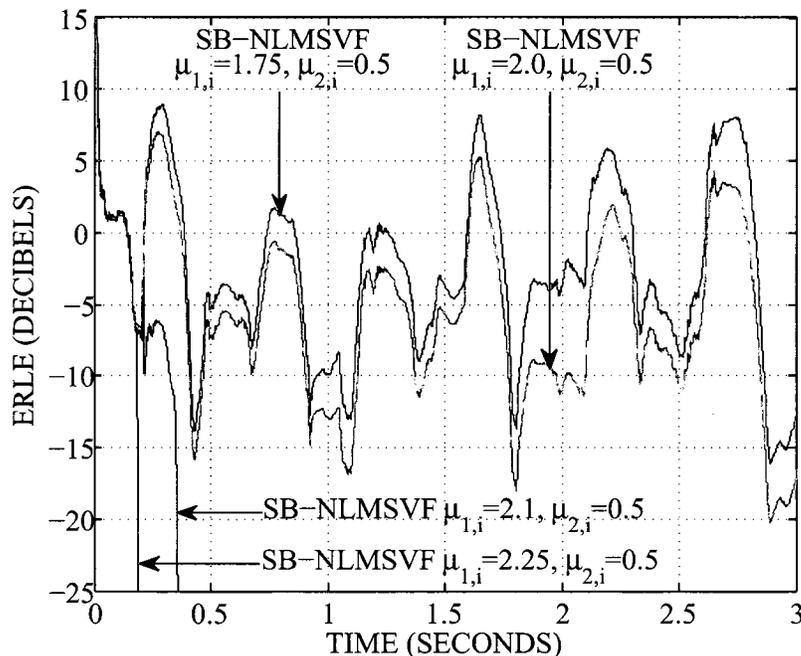


Figure 5.16: ERLE from 0 to 3 seconds for SB-NLMSVF with measured speech input from the smartphone under high volume loudspeaker distortion conditions and an increasing $\mu_{1,i}$.

The impact on convergence performance of the SB-NLMSVF structure for increasing step sizes applied to only the linear and to only the quadratic subband adaptive filters is shown in Figures 5.16 and 5.17. The previous parameter settings for the $K = 8$ SB-NLMSVF structure from Figures 5.13 and 5.14 were used. With the step sizes of the quadratic filters set to 0.5 in all subbands, increasing the step sizes of the linear filters beyond 2.0 in all subbands results in unstable behaviour as shown in Figure 5.16. Similarly, with the step sizes of the linear filters set to 0.5 in all subbands, increasing the step sizes of the quadratic filters above 2.0 in all subbands also results in unstable behaviour as seen in Figure 5.17. This demonstrates that $0 < \mu_{1,i}, \mu_{2,i} < 2$ are indeed reasonable bounds for stable operation of the SB-NLMSVF structure. Although not shown, when the step sizes of the quadratic filters are set to 0 in all

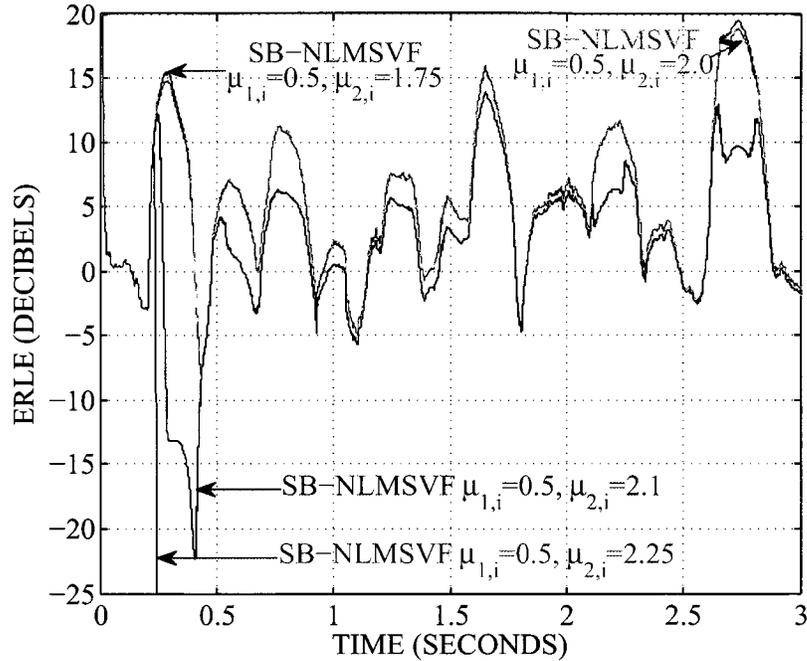


Figure 5.17: ERLE from 0 to 3 seconds for SB-NLMSVF with measured speech input from the smartphone under high volume loudspeaker distortion conditions and an increasing $\mu_{2,i}$.

subbands the SB-NLMSVF structure behaves like the linear only SB-NLMS structure. Furthermore, when the step sizes of the linear filters are set to 0 in all subbands the SB-NLMSVF structure behaves like a pure nonlinear structure by modeling and removing only the quadratic echo path components.

5.2.4 Computational Complexity Results

The computational complexity of the SB-NLMSVF structure compared to its fullband and linear only structure counterparts is presented in this section in terms of the total number of equivalent real mathematical operations required per input sample period. As in Section 4.2.4 of Chapter 4, the complexity results presented assume the AEC structures were implemented on an Intel Core 2 Duo processor where a real addition and a real multiplication required approximately the same number of clock cycles to execute, with a real division operation requiring approximately 6 times as many

clock cycles for single precision floating point data [103]. The complexity of the SB-NLMSVF structure along with its fullband and linear only variants were computed based on Tables 5.1 – 5.3.

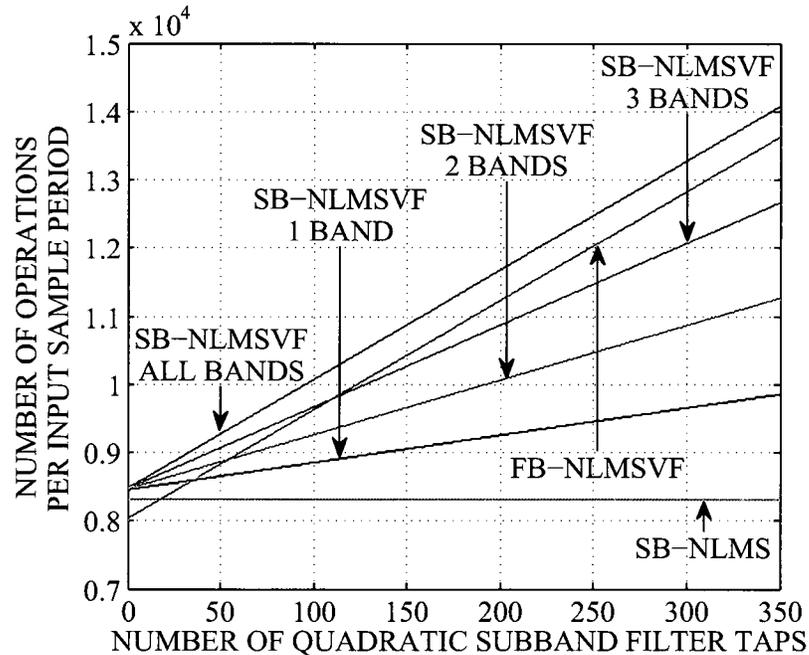


Figure 5.18: Computational complexity for SB-NLMS, FB-NLMSVF, and SB-NLMSVF.

The computational cost of the SB-NLMSVF structure along with the FB-NLMSVF and SB-NLMS structures is shown in Figure 5.18 as the length of the quadratic subband adaptive filters increase. The complexity cost for SB-NLMSVF structure was computed based on the filter bank settings outlined in Section 5.2.1, with the length of the linear subband adaptive filters set to $N_{1, sb_i} = 500$ in all subbands and the length of the quadratic subband adaptive filters, N_{2, sb_i} , set equal in all subbands as well. The corresponding linear and quadratic adaptive filter lengths of the fullband structures were set to $N_1 = MN_{1, sb_i} = 2000$ and MN_{2, sb_i} respectively, where it was assumed that equal ERLE performance could be achieved by the FB-NLMSVF and SB-NLMSVF structures for the given filter length relationships. With second order

adaptive Volterra filtering applied in all subbands the SB-NLMSVF structure constantly requires approximately 400 more operations than the FB-NLMSVF structure, as seen in Figure 5.18. However, as second order adaptive Volterra filtering is applied in fewer subbands of the SB-NLMSVF structure its computational complexity becomes increasingly less compared to the FB-NLMSVF structure, and approaches the complexity of the low cost linear only SB-NLMS structure. The significant improvement in ERLE performance of the SB-NLMSVF structure compared to the SB-NLMS structure, as presented in Sections 5.2.2 and 5.2.3, requires only approximately 8%, 13%, and 25% more operations with second order adaptive Volterra filtering in only the 0–1 kHz subband, the 1–2 and 2–3 kHz subbands, and all subbands respectively, for $N_{2, sb_i} = 120$. Compared to the FB-NLMSVF structure which required 1275 full-band quadratic filter taps, corresponding to 319 subband quadratic filter taps in Figure 5.18, the SB-NLMSVF structure required $N_{2, sb_i} = 120$ subband quadratic filter taps to attain the similar or better ERLE performance outlined in Sections 5.2.2 and 5.2.3. In this case the computational cost of the SB-NLMSVF structure requires roughly 30%, 26%, and 20% fewer operations with second order adaptive Volterra filtering in only the 0–1 kHz subband, the 1–2 and 2–3 kHz subbands, and all subbands respectively compared to the FB-NLMSVF structure. With second order adaptive Volterra filtering applied in only the 0–1 kHz subband and $N_{2, sb_1} = 30$, the SB-NLMSVF structure requires only roughly 3% more operations than the SB-NLMS structure to achieve the significantly improved ERLE performance shown in Sections 5.2.2 and 5.2.3 for the conference phone. As well, only approximately 3% more operations are required by the SB-NLMSVF structure with $N_{2, sb_{2-3}} = 10$ to attain the improved ERLE performance compared to the SB-NLMS structure shown in Figure 5.15 for the smartphone. Also, it should be noted that in general the computational advantage of the SB-NLMSVF structure will increase compared the FB-NLMSVF structure as the OSR decreases.

The computational cost of the SB-NLMSVF and SB-NLMS structures are shown

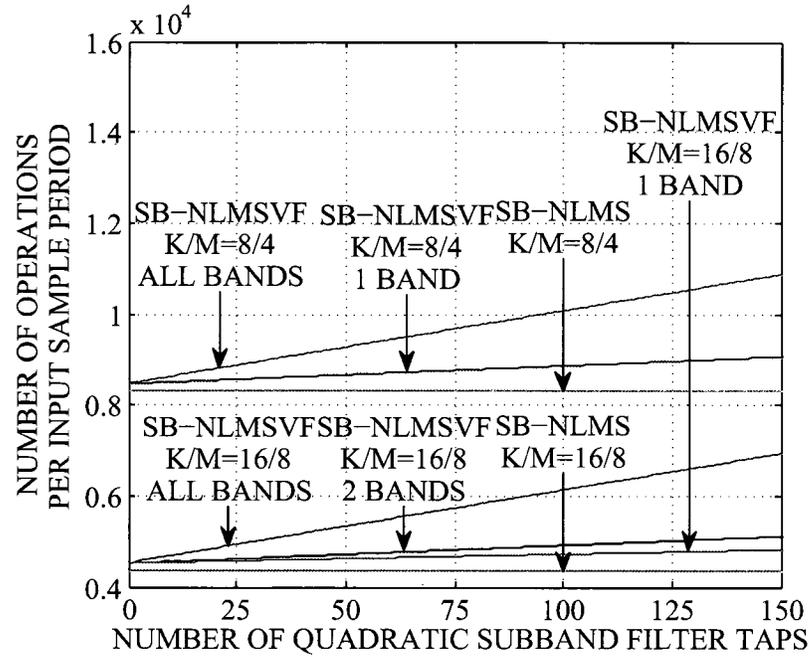


Figure 5.19: Computational complexity for SB-NLMSVF and SB-NLMS for an increasing number of subbands.

in Figure 5.19 for an increasing number of subbands with the same OSR of 2. The results are based on the same filter bank settings as those outlined in Section 5.2.3 for Figure 5.12. The length of the linear adaptive filters for the $K = 8$ structures were set to $N_{1, sb_i} = 500$ in all subbands and to $N_{1, sb_i} = 250$ in all subbands for the $K = 16$ structures, since equivalent ERLE performance was obtained by the SB-NLMSVF structures under these settings. With quadratic Volterra filtering applied in only one subband for the $K = 16$ SB-NLMSVF structure with $N_{2, sb_1} = 30$, a 50% reduction in complexity is achieved compared to the $K = 8$ SB-NLMSVF structure with quadratic Volterra filtering applied in only one subband with $N_{2, sb_1} = 120$. Moreover, the $K = 16$ SB-NLMSVF structure with a quadratic Volterra filter of length $N_{2, sb_1} = 30$ in only one subband provides roughly the same ERLE performance as the much more computationally expensive $K = 8$ SB-NLMSVF structure with quadratic Volterra filtering in only one band with $N_{2, sb_1} = 120$, as shown in Figure 5.12. Also, the $K = 16$ SB-NLMSVF structure with a quadratic Volterra filter of length $N_{2, sb_i} = 30$ in only

one subband requires only a slightly higher computational cost than the linear only $K = 16$ SB-NLMS structure, while achieving improved ERLE performance during periods of high echo path distortion as seen in Figure 5.12.

5.3 Summary

This chapter investigated the application of second order adaptive Volterra filtering in an oversampled subband AEC system to efficiently model electrodynamic loudspeaker nonlinearities and thereby improve AEC performance. An in depth discussion and computational complexity analysis of the SB-NLMSVF structure was given in Section 5.1. AEC performance results for the SB-NLMSVF structure along with its fullband and linear only counterparts were outlined in Sections 5.2.2 and 5.2.3 under experimentally measured LEMS conditions. These results demonstrated the ability of the SB-NLMSVF structure to attain equal or improved AEC performance compared to its fullband equivalent, and significantly better performance compared to the linear only AEC structures during periods of high loudspeaker harmonic distortion. Moreover, the SB-NLMSVF structure was able to take advantage of the time and frequency domain nature of the loudspeaker nonlinearity and achieve the same level of improved AEC performance. This was achieved by applying second order adaptive Volterra filtering in only those subbands impacted by the nonlinearity, and by modeling only the important high amplitude quadratic Volterra kernel coefficients. These exploits allowed the SB-NLMSVF structure to realize significant computational savings compared to its equivalent fullband Volterra AEC structure, along with requiring only minor increases in complexity compared to a linear fullband AEC structure, as described in Section 5.2.4.

Table 5.4 summarizes the main results of this chapter. The maximum improvement in ERLE for each AEC structure compared to the SB-NLMS structure is given based on the measured high volume loudspeaker conditions presented in Section 5.2.3

for the conference phone. The corresponding computational complexity results are also presented in Table 5.4 and are applicable to the AEC structures in all sections under the conditions outlined in Section 5.2.4. The results for the subband structures are based on $K = 8$ $M = 4$ filter banks with $L_{pr} = 64$. Also, it should be noted that ERLE improvements of 6 dB or more attained by the proposed SB-NLMSVF structure translates to a reduction by a factor of two or more in the output error signal. Furthermore, selected results from this chapter have been archived online [105] to demonstrate the perceived improvement in echo reduction that can be realized by the proposed SB-NLMSVF structure.

AEC Structure	ERLE Improvement Compared to SB-NLMS (dB)	Complexity (Operations per input sample period)
SB-NLMS $N_{1, sb_{1-4}} = 500$	–	8305
FB-NLMSVF $N_1 = 2000$ $N_2 = 50$	6	13128
SB-NLMSVF $N_{1, sb_{1-4}} = 500$ $N_{2, sb_{1-4}} = 120$	7	10388
SB-NLMSVF $N_{1, sb_{1-4}} = 500$ $N_{2, sb_1} = 120$	7	8935

Table 5.4: Summary of results for AEC in nonlinear environments.

Chapter 6

Subband Volterra Structure for Acoustic Echo Cancellation in Nonstationary Environments

As the performance of a typical EC can be adversely affected by changing echo path conditions along with nonlinear echo path components, it should consider both impairments in its model of the LEMS. In this chapter the generalized proportionate variable step size subband adaptive filter structure introduced in Chapter 4 is combined with the subband Volterra structure of Chapter 5, to create an AEC structure capable of quickly modeling and tracking changes within the linear portion of the LEMS, as well as the nonlinear portion resulting from electrodynamic loudspeaker harmonic distortion. It is shown that this generalized proportionate variable step size subband adaptive Volterra filter structure can achieve improved convergence and tracking performance for AEC compared to its fullband time domain counterpart, at a much reduced computational complexity.

The following sections are presented in this chapter. In Section 6.1 a detailed description of the generalized proportionate variable step size subband adaptive Volterra

filter structure is presented along with an in depth computational complexity analysis. Experimental results based on measured linear and quadratic LEMS Volterra kernels, as well as on directly measured LEMS input signal data, are given in Section 6.2. A summary of the main findings concludes the chapter in Section 6.3.

6.1 Generalized Proportionate Variable Step Size Subband Adaptive Volterra Filter Structure

6.1.1 Description

As with the changes that occur in the linear portion of the echo path due to a small obstruction, the changes that occur in the quadratic part due to an increase in loudspeaker input signal level also vary in specific time and frequency regions, as shown in Figures 3.8 and 3.9 of Chapter 3. The quadratic Volterra kernel taps with the highest magnitude experience the greatest change in value due to an increase in loudspeaker input signal level. Moreover, the quadratic Volterra kernels variations tend to occur in specific frequency regions only, due to the underlying nature of the loudspeaker distortion as described in Chapter 3. Thus, to exploit the trends in time and frequency during linear and quadratic echo path changes and allow for fast tracking of these changes the GGPAPA, discussed in Section 2.2.2.1 of Chapter 2, is applied to the second order subband Volterra filter AEC structure outlined in Chapter 5. Only the $P = 1$ version of the GGPAPA is considered, this will be referred to as the generalized gradient proportionate normalized least-mean-square (GGPNLMS) algorithm. A block diagram of the subband generalized gradient proportionate normalized least-mean-square Volterra filter (SB-GGPNLMSVF) structure implemented with the same oversampled GDFT modulated filter bank setup as in Figure 5.1, is shown again in Figure 6.1 for convenience. Since the SB-GGPNLMSVF structure shares the same filter bank configuration as the SB-NLMSVF structure from Chapter

5, a description of the fullband and corresponding subband input and output signals of the SB-GGPNLMSVF structure can be found in Section 5.1.1 of Chapter 5.

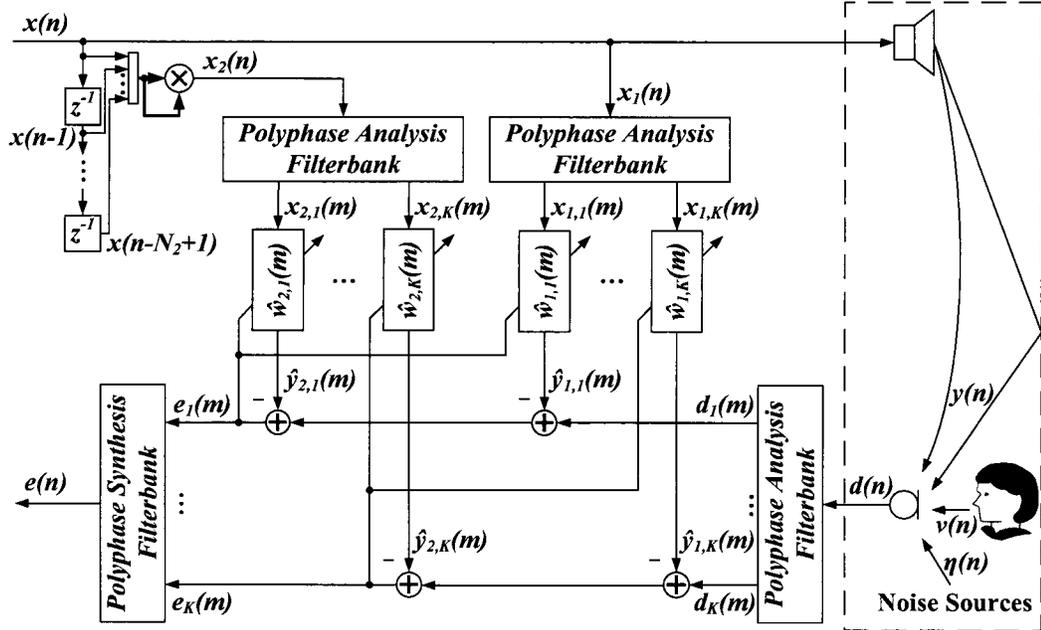


Figure 6.1: Block diagram of the SB-GGPNLMSVF structure.

As with the SB-GGPAP and SB-NLMSVF AEC structures discussed in Chapters 4 and 5 respectively, the key feature of the SB-GGPNLMSVF structure is its ability to perform second order adaptive Volterra filtering based on the fast tracking GGPNLMS algorithm or a simpler derivative algorithm, such as NLMS, on a per subband basis. This allows the GGPNLMS algorithm to be applied in only those linear and quadratic subbands that are impacted the most by fluctuations in the linear and quadratic echo path components, while the simpler NLMS algorithm can be used in the unaffected subbands. As shown in Chapter 3, the high frequency linear echo path components tend to be affected the most by the introduction of a small obstruction in close vicinity to the hands-free device. While a change in loudspeaker signal level affects the specific frequency regions that contain the majority of the original quadratic echo path energy. Thus, applying the GGPNLMS algorithm in only the correspondingly affected linear and quadratic subbands, will allow the SB-GGPNLMSVF structure to

maintain echo cancellation performance while reducing its computational cost.

The subband input signal and adaptive filter vector definitions for the SB-GGPNLMSVF algorithm are as follows:

$$\hat{\mathbf{w}}_{1,i}(m) = [\hat{w}_{1,i_0}(m), \hat{w}_{1,i_1}(m), \dots, \hat{w}_{1,i_{N_{1,sb_i}-1}}(m)]^T \quad (6.1)$$

$$\hat{\mathbf{w}}_{2,i}(m) = [\hat{w}_{2,i_0}(m), \hat{w}_{2,i_1}(m), \dots, \hat{w}_{2,i_{N_{2,sb_i}-1}}(m)]^T \quad (6.2)$$

$$\mathbf{x}_{1,i}(m) = [x_{1,i}(m), x_{1,i}(m-1), \dots, x_{1,i}(m-N_{1,sb_i}+1)]^T \quad (6.3)$$

$$\mathbf{x}_{2,i}(m) = [x_{2,i}(m), x_{2,i}(m-1), \dots, x_{2,i}(m-N_{2,sb_i}+1)]^T \quad (6.4)$$

$$\mathbf{e}_i(m) = \mathbf{d}_i(m) - \hat{\mathbf{w}}_{1,i}^H(m)\mathbf{x}_{1,i}(m) - \hat{\mathbf{w}}_{2,i}^H(m)\mathbf{x}_{2,i}(m) \quad (6.5)$$

where the elements of $\mathbf{x}_{2,i}(m)$ result from the corresponding fullband elements of $x_2(n)$, which is itself determined by (5.1) as discussed in Section 5.1.1 of Chapter 5. The linear and quadratic subband adaptive Volterra filters, $\hat{\mathbf{w}}_{1,i}(m)$ and $\hat{\mathbf{w}}_{2,i}(m)$, of the SB-GGPNLMSVF structure are updated based on the GGPNLMS algorithm as follows:

$$\hat{\mathbf{w}}_{1,i}(m+1) = \hat{\mathbf{w}}_{1,i}(m) + \mu_{1,i}\mathbf{A}_{1,i}(m)\mathbf{x}_{1,i}(m)e_i^*(m)\Gamma_{1,i}(m) \quad (6.6)$$

$$\hat{\mathbf{w}}_{2,i}(m+1) = \hat{\mathbf{w}}_{2,i}(m) + \mu_{2,i}\mathbf{A}_{2,i}(m)\mathbf{x}_{2,i}(m)e_i^*(m)\Gamma_{2,i}(m) \quad (6.7)$$

$$\Gamma_{1,i}(m) = (\mathbf{x}_{1,i}^H(m)\mathbf{A}_{1,i}(m)\mathbf{x}_{1,i}(m) + \mathbf{x}_{2,i}^H(m)\mathbf{A}_{2,i}(m)\mathbf{x}_{2,i}(m) + \delta_{1,i})^{-1} \quad (6.8)$$

$$\Gamma_{2,i}(m) = (\mathbf{x}_{1,i}^H(m)\mathbf{A}_{1,i}(m)\mathbf{x}_{1,i}(m) + \mathbf{x}_{2,i}^H(m)\mathbf{A}_{2,i}(m)\mathbf{x}_{2,i}(m) + \delta_{2,i})^{-1} \quad (6.9)$$

$$\mathbf{A}_{1,i}(m) = \text{diag} [g_{1,i_0}(m), \dots, g_{1,i_{N_{1,sb_i}-1}}(m)] \quad (6.10)$$

$$\mathbf{A}_{2,i}(m) = \text{diag} [g_{2,i_0}(m), \dots, g_{2,i_{N_{2,sb_i}-1}}(m)] \quad (6.11)$$

$$g_{1,i_k}(m) = \frac{1 - \beta_{1,i}}{2} \frac{\bar{c}_{1,i_k}(m-1)}{\sum_{k=0}^{N_{1,sb_i}-1} \bar{c}_{1,i_k}(m-1) + \xi_{1,i}} + \frac{\beta_{1,i}}{2N_{1,sb_i}} \quad (6.12)$$

$$g_{2,i_l}(m) = \frac{1 - \beta_{2,i}}{2} \frac{\bar{c}_{2,i_l}(m-1)}{\sum_{l=0}^{N_{2,sb_i}-1} \bar{c}_{2,i_l}(m-1) + \xi_{2,i}} + \frac{\beta_{2,i}}{2N_{2,sb_i}} \quad (6.13)$$

$$\bar{c}_{1,i_k}(m) = \epsilon_{1,i} \bar{c}_{1,i_k}(m-1) + (1 - \epsilon_{1,i}) |c_{1,i_k}(m)| \quad (6.14)$$

$$\bar{c}_{2,i_l}(m) = \epsilon_{2,i} \bar{c}_{2,i_l}(m-1) + (1 - \epsilon_{2,i}) |c_{2,i_l}(m)| \quad (6.15)$$

$$c_{1,i_k}(m) = \hat{w}_{1,i_k}(m-1) - \gamma_{1,i} \bar{w}_{1,i_k}(m-1) \quad (6.16)$$

$$c_{2,i_l}(m) = \hat{w}_{2,i_l}(m-1) - \gamma_{2,i} \bar{w}_{2,i_l}(m-1) \quad (6.17)$$

$$\bar{w}_{1,i_k}(m) = \eta_{1,i} \bar{w}_{1,i_k}(m-1) + (1 - \eta_{1,i}) \hat{w}_{1,i_k}(m-1) \quad (6.18)$$

$$\bar{w}_{2,i_l}(m) = \eta_{2,i} \bar{w}_{2,i_l}(m-1) + (1 - \eta_{2,i}) \hat{w}_{2,i_l}(m-1) \quad (6.19)$$

where $k = 0, \dots, N_{1,sb_i} - 1$, and $l = 0, \dots, N_{2,sb_i} - 1$. The GGPNLMS algorithm parameters for both the linear and quadratic parts of the SB-GGPNLMSVF structure are analogous to those described in Section 4.1.1 of Chapter 4. A non-generalized version of the SB-GGPNLMSVF AEC structure, called subband gradient proportionate normalized least-mean-square Volterra filter (SB-GPNLMSVF), results when $\epsilon_{1,i}$ and $\epsilon_{2,i}$ are both set to 0 in all subbands. The SB-NLMSVF structure, discussed in Section 5.1 of Chapter 5, arises as a special case of the SB-GGPNLMSVF structure when $\beta_{1,i} = \beta_{2,i} = 1$ in all subbands and $\delta_{1,i}$ and $\delta_{2,i}$ are divided by $2N_{1,sb_i}$ and $2N_{2,sb_i}$ respectively. Improved proportionate, proportionate, and exponential step size versions can result from the SB-GGPNLMSVF structure when the linear and quadratic subband adaptive filter parameters are set analogously to the values for the corresponding linear only structures, as outlined in Section 4.1.1 of Chapter 4.

6.1.2 Computational Complexity Analysis

The computational complexity of the SB-GGPNLMSVF structure shown in Figure 6.1 is comprised of the cost of implementing the four polyphase filter banks, along with the linear and quadratic subband adaptive filtering expense based on the GGPNLMS

algorithm. The computational complexity of the SB-GGPNLMSVF structure is determined under the same considerations as in Section 5.1.2 of Chapter 5 and Section 4.1.2 of Chapter 4. The computational cost of adapting the linear and quadratic sub-band Volterra filters with the GGPNLMS algorithm, as described in (6.5) to (6.19), is summarized in terms of the equivalent real elementary mathematical operations per input sample period for real and complex input data in Tables 6.1 and 6.2 respectively. The cost of computing a square root is denoted by $C_{\sqrt{\cdot},m}$, $C_{\sqrt{\cdot},a}$, and $C_{\sqrt{\cdot},d}$ for the required number of real multiplications, additions, and divisions respectively.

Term	\times	$+$	\div
$\hat{\mathbf{w}}_{1,i}^H(m)\mathbf{x}_{1,i}(m)$	N_{1,sb_i}	$N_{1,sb_i} - 1$	
$\hat{\mathbf{w}}_{2,i}^H(m)\mathbf{x}_{2,i}(m)$	N_{2,sb_i}	$N_{2,sb_i} - 1$	
$d_i(m) - \hat{\mathbf{w}}_{1,i}^H(m)\mathbf{x}_{1,i}(m) - \hat{\mathbf{w}}_{2,i}^H(m)\mathbf{x}_{2,i}(m)$		2	
$\mathbf{A}_{1,i}(m)\mathbf{x}_{1,i}(m)$	N_{1,sb_i}		
$\mathbf{A}_{2,i}(m)\mathbf{x}_{2,i}(m)$	N_{2,sb_i}		
$\mathbf{x}_{1,i}^H(m)\mathbf{A}_{1,i}(m)\mathbf{x}_{1,i}(m)$	N_{1,sb_i}	$N_{1,sb_i} - 1$	
$\mathbf{x}_{2,i}^H(m)\mathbf{A}_{2,i}(m)\mathbf{x}_{2,i}(m)$	N_{2,sb_i}	$N_{2,sb_i} - 1$	
$\mathbf{x}_{1,i}^H(m)\mathbf{A}_{1,i}(m)\mathbf{x}_{1,i}(m) + \mathbf{x}_{2,i}^H(m)\mathbf{A}_{2,i}(m)\mathbf{x}_{2,i}(m) + \delta_{1,i}$		2	
$\mathbf{x}_{1,i}^H(m)\mathbf{A}_{1,i}(m)\mathbf{x}_{1,i}(m) + \mathbf{x}_{2,i}^H(m)\mathbf{A}_{2,i}(m)\mathbf{x}_{2,i}(m) + \delta_{2,i}$		1	
$\mu_{1,i}\mathbf{A}_{1,i}(m)\mathbf{x}_{1,i}(m)e_i^*(m)\Gamma_{1,i}(m)$	$N_{1,sb_i} + 1$		1
$\mu_{2,i}\mathbf{A}_{2,i}(m)\mathbf{x}_{2,i}(m)e_i^*(m)\Gamma_{2,i}(m)$	$N_{2,sb_i} + 1$		1
$\hat{\mathbf{w}}_{1,i}(m) + \mu_{1,i}\mathbf{A}_{1,i}(m)\mathbf{x}_{1,i}(m)e_i^*(m)\Gamma_{1,i}(m)$		N_{1,sb_i}	
$\hat{\mathbf{w}}_{2,i}(m) + \mu_{2,i}\mathbf{A}_{2,i}(m)\mathbf{x}_{2,i}(m)e_i^*(m)\Gamma_{2,i}(m)$		N_{2,sb_i}	

Term	×	+	÷
$s_{1,i}(m) = \sum_{k=0}^{N_{1, sb_i} - 1} \bar{c}_{1, i_k}(m-1) + \xi_{1,i}$		N_{1, sb_i}	
$s_{2,i}(m) = \sum_{l=0}^{N_{2, sb_i} - 1} \bar{c}_{2, i_l}(m-1) + \xi_{2,i}$		N_{2, sb_i}	
$((1 - \beta_{1,i})/2)/s_{1,i}(m) \times$ $\bar{c}_{1, i_k}(m-1) + \beta_{1,i}/(2N_{1, sb_i})$	N_{1, sb_i}	N_{1, sb_i}	1
$((1 - \beta_{2,i})/2)/s_{2,i}(m) \times$ $\bar{c}_{2, i_l}(m-1) + \beta_{2,i}/(2N_{2, sb_i})$	N_{2, sb_i}	N_{2, sb_i}	1
$\epsilon_{1,i} \bar{c}_{1, i_k}(m-1) + (1 - \epsilon_{1,i}) c_{1, i_k}(m) $	$2N_{1, sb_i}$	N_{1, sb_i}	
$\epsilon_{2,i} \bar{c}_{2, i_l}(m-1) + (1 - \epsilon_{2,i}) c_{2, i_l}(m) $	$2N_{2, sb_i}$	N_{2, sb_i}	
$\hat{w}_{1, i_k}(m-1) - \gamma_{1,i} \bar{w}_{1, i_k}(m-1)$	N_{1, sb_i}	N_{1, sb_i}	
$\hat{w}_{2, i_l}(m-1) - \gamma_{2,i} \bar{w}_{2, i_l}(m-1)$	N_{2, sb_i}	N_{2, sb_i}	
$\eta_{1,i} \bar{w}_{1, i_k}(m-1) + (1 - \eta_{1,i}) \hat{w}_{1, i_k}(m-1)$	$2N_{1, sb_i}$	N_{1, sb_i}	
$\eta_{2,i} \bar{w}_{2, i_l}(m-1) + (1 - \eta_{2,i}) \hat{w}_{2, i_l}(m-1)$	$2N_{2, sb_i}$	N_{2, sb_i}	
Total per iteration	$10N_{1, sb_i} +$ $10N_{2, sb_i} + 2$	$8N_{1, sb_i} +$ $8N_{2, sb_i} + 1$	4

Table 6.1: Approximate computational complexity per iteration of the GGPNLMS algorithm applied to subband adaptive second order Volterra filters for real input data in terms of real multiplications, additions, and divisions.

Term	×	+	÷
$\hat{w}_{1,i}^H(m) \mathbf{x}_{1,i}(m)$	$4N_{1, sb_i}$	$4N_{1, sb_i} - 2$	
$\hat{w}_{2,i}^H(m) \mathbf{x}_{2,i}(m)$	$4N_{2, sb_i}$	$4N_{2, sb_i} - 2$	
$d_i(m) - \hat{w}_{1,i}^H(m) \mathbf{x}_{1,i}(m) -$ $\hat{w}_{2,i}^H(m) \mathbf{x}_{2,i}(m)$		4	
$\mathbf{A}_{1,i}(m) \mathbf{x}_{1,i}(m)$	$2N_{1, sb_i}$		
$\mathbf{A}_{2,i}(m) \mathbf{x}_{2,i}(m)$	$2N_{2, sb_i}$		
$\mathbf{x}_{1,i}^H(m) \mathbf{A}_{1,i}(m) \mathbf{x}_{1,i}(m)$	$4N_{1, sb_i}$	$4N_{1, sb_i} - 2$	

Term	\times	$+$	\div
$\mathbf{x}_{2,i}^H(m) \mathbf{A}_{2,i}(m) \mathbf{x}_{2,i}(m)$	$4N_{2, sb_i}$	$4N_{2, sb_i} - 2$	
$\mathbf{x}_{1,i}^H(m) \mathbf{A}_{1,i}(m) \mathbf{x}_{1,i}(m) +$ $\mathbf{x}_{2,i}^H(m) \mathbf{A}_{2,i}(m) \mathbf{x}_{2,i}(m) + \delta_{1,i}$		3	
$\mathbf{x}_{1,i}^H(m) \mathbf{A}_{1,i}(m) \mathbf{x}_{1,i}(m) +$ $\mathbf{x}_{2,i}^H(m) \mathbf{A}_{2,i}(m) \mathbf{x}_{2,i}(m) + \delta_{2,i}$		1	
$\mu_{1,i} \mathbf{A}_{1,i}(m) \mathbf{x}_{1,i}(m) e_i^*(m) \Gamma_{1,i}(m)$	$4N_{1, sb_i} + 2$	$2N_{1, sb_i}$	1
$\mu_{2,i} \mathbf{A}_{2,i}(m) \mathbf{x}_{2,i}(m) e_i^*(m) \Gamma_{2,i}(m)$	$4N_{2, sb_i} + 2$	$2N_{2, sb_i}$	1
$\hat{\mathbf{w}}_{1,i}(m) +$ $\mu_{1,i} \mathbf{A}_{1,i}(m) \mathbf{x}_{1,i}(m) e_i^*(m) \Gamma_{1,i}(m)$		$2N_{1, sb_i}$	
$\hat{\mathbf{w}}_{2,i}(m) +$ $\mu_{2,i} \mathbf{A}_{2,i}(m) \mathbf{x}_{2,i}(m) e_i^*(m) \Gamma_{2,i}(m)$		$2N_{2, sb_i}$	
$s_{1,i}(m) = \sum_{k=0}^{N_{1, sb_i} - 1} \bar{c}_{1, i_k}(m-1) + \xi_{1,1}$		N_{1, sb_i}	
$s_{2,i}(m) = \sum_{l=0}^{N_{2, sb_i} - 1} \bar{c}_{2, i_l}(m-1) + \xi_{2,i}$		N_{2, sb_i}	
$((1 - \beta_{1,i})/2)/s_{1,i}(m) \times$ $\bar{c}_{1, i_k}(m-1) + \beta_{1,i}/(2N_{1, sb_i})$	N_{1, sb_i}	N_{1, sb_i}	1
$((1 - \beta_{2,i})/2)/s_{2,i}(m) \times$ $\bar{c}_{2, i_l}(m-1) + \beta_{2,i}/(2N_{2, sb_i})$	N_{2, sb_i}	N_{2, sb_i}	1
$\epsilon_{1,i} \bar{c}_{1, i_k}(m-1) +$ $(1 - \epsilon_{1,i}) c_{1, i_k}(m) $	$4N_{1, sb_i} +$ $N_{1, sb_i} C_{\sqrt{\cdot}, m}$	$2N_{1, sb_i} +$ $N_{1, sb_i} C_{\sqrt{\cdot}, a}$	N_{1, sb_i} $\times C_{\sqrt{\cdot}, d}$
$\epsilon_{2,i} \bar{c}_{2, i_l}(m-1) +$ $(1 - \epsilon_{2,i}) c_{2, i_l}(m) $	$4N_{2, sb_i} +$ $N_{2, sb_i} C_{\sqrt{\cdot}, m}$	$2N_{2, sb_i} +$ $N_{2, sb_i} C_{\sqrt{\cdot}, a}$	N_{2, sb_i} $\times C_{\sqrt{\cdot}, d}$
$\hat{w}_{1, i_k}(m-1) - \gamma_{1,i} \bar{w}_{1, i_k}(m-1)$	$2N_{1, sb_i}$	$2N_{1, sb_i}$	
$\hat{w}_{2, i_l}(m-1) - \gamma_{2,i} \bar{w}_{2, i_l}(m-1)$	$2N_{2, sb_i}$	$2N_{2, sb_i}$	
$\eta_{1,i} \bar{w}_{1, i_k}(m-1) + (1 - \eta_{1,i}) \hat{w}_{1, i_k}(m-1)$	$4N_{1, sb_i}$	$2N_{1, sb_i}$	
$\eta_{2,i} \bar{w}_{2, i_l}(m-1) + (1 - \eta_{2,i}) \hat{w}_{2, i_l}(m-1)$	$4N_{2, sb_i}$	$2N_{2, sb_i}$	

Term	\times	$+$	\div
Total per iteration	$25N_{1,sb_i} +$	$20N_{1,sb_i} +$	$4 +$
	$N_{1,sb_i} C_{\sqrt{\cdot},m}$	$N_{1,sb_i} C_{\sqrt{\cdot},a}$	N_{1,sb_i}
	$+25N_{2,sb_i} +$	$+20N_{2,sb_i} +$	$\times C_{\sqrt{\cdot},d}$
	$N_{2,sb_i} C_{\sqrt{\cdot},m}$	$N_{2,sb_i} C_{\sqrt{\cdot},a}$	$+N_{2,sb_i}$
	$+4$		$\times C_{\sqrt{\cdot},d}$

Table 6.2: Approximate computational complexity per iteration of the GGPNLMS algorithm applied to subband adaptive second order Volterra filters for complex input data in terms of real multiplications, additions, and divisions.

\times	$\frac{4}{M} (L_{pr} + 4K \log_2 K + 4K) + D +$ $\frac{1}{M} \sum_{i=1}^{\frac{K}{2}} \left((25 + C_{\sqrt{\cdot},m})(N_{1,sb_i} + N_{2,sb_i}) + 4 \right)$
$+$	$\frac{4}{M} (L_{pr} + 4K \log_2 K + 4K) +$ $\frac{1}{M} \sum_{i=1}^{\frac{K}{2}} \left((20 + C_{\sqrt{\cdot},a})(N_{1,sb_i} + N_{2,sb_i}) \right)$
\div	$\frac{1}{M} \sum_{i=1}^{\frac{K}{2}} \left(4 + (N_{1,sb_i} + N_{2,sb_i}) C_{\sqrt{\cdot},d} \right)$

Table 6.3: Approximate computational complexity per input sample period for the SB-GGPNLMSVF AEC structure in terms of real multiplications, additions, and divisions.

The total computational cost for the SB-GGPNLMSVF AEC structure is comprised of the cost of the 4 oversampled GDFT modulated polyphase filter banks, the cost of computing $x_2(n)$, along with the GGPNLMS adaptive filtering cost of the $K/2$ unique complex linear and quadratic subband filters operating at the M fold reduced sampling rate. The cost of computing $x_2(n)$ is D real multiplications per sample period as discussed in Section 5.1.1 of Chapter 5. The overall computational complexity for the SB-GGPNLMSVF structure is summarized in Table 6.3. The complexity of the SB-GGPNLMSVF structure is dominated by the required number of real multiplication and addition operations and is $O\left((N_{1,sb_i} + N_{2,sb_i})\frac{K}{M}\right)$ for

both when $N_{1, sb_i}, N_{2, sb_i} \gg L_{pr}$ and the length of each linear and quadratic subband adaptive filter is equal.

6.2 Experimental Results

6.2.1 Methodology

To perform meaningful AEC simulations under changing acoustic conditions, synthetic microphone input signals were created using (2.23) based on the measured unobstructed and obstructed LEMS Volterra kernels obtained under low and high volume conditions for the conference phone configuration from Section 3.2 and 3.3 of Chapter 3. This allows the performance of the AEC structures in question to be studied during a change in both the linear and nonlinear portions of the echo path under controlled input conditions. The change in the echo signal was created by linearly interpolating from the measured linear and quadratic Volterra kernels coefficients obtained under one set of LEMS conditions, to those obtained under a different set of conditions over a desired duration. This allows for the simulation of a local talkers hand moving towards the phone set to change the loudspeaker volume level to be performed, and the impact it has on the performance of the AEC structure under test to be studied. Also, simulation results are presented based on experimentally measured input reference and microphone data from the smartphone. This allows for the behaviour of the AEC structures to be investigated under real hands-free communication operating conditions.

The results presented in the subsequent sections for the subband AEC structures are based on the same filter bank settings given in Section 4.2.1 of Chapter 4, unless indicated otherwise. The results for the proposed SB-GGPNLMSVF AEC structure are compared to the structures proposed in Chapters 4 and 5, along with the FB-NLMSVF structure outlined in [84], the fullband improved proportionate normalized

least-mean-square Volterra filter (FB-IPNLMSVF) structure described in [88], and to both a subband and fullband version of the NLMS algorithm [33], in terms of ERLE performance and computational complexity for an 8 kHz sampling rate.

6.2.2 Simulation Results Based on Measured Volterra Kernels

6.2.2.1 White Noise Excitation Signal With a Changing Quadratic Volterra Kernel

In this section the AEC performance of the SB-GGPNLMSVF structure, discussed in Section 6.1.1, along with its fullband and linear only structure counterparts are presented under changing conditions within only the nonlinear portion of the acoustic echo path. Simulations were performed and averaged over several trials for synthetic microphone signals created based on the measured linear and quadratic LEMS Volterra kernels, obtained under unobstructed low and high volume conditions, from the conference phone configuration with lengths $L_1 = 2000$ and $L_2(L_2 + 1)/2 = 1275$ respectively, where $L_2 = 50$. A 12 second WGN reference excitation signal, $x(n)$, with zero mean and unit variance was used along with the measured Volterra kernels to create the echo signal, $y(n)$, using (2.23). The change in $y(n)$ was achieved by linearly interpolating from the low volume quadratic Volterra kernels coefficients to the high volume coefficients over a 0.5 second duration starting at the 7 second mark of the simulation. The high volume linear Volterra kernel coefficients used in creating $y(n)$ were unchanged during the course of the simulations. This allows the effectiveness of applying the GGPNLMS algorithm to second order Volterra filters to be studied when only a change in the nonlinear part of the echo path occurs. The input microphone signal, $d(n)$, was obtained by adding 12 seconds of uncorrelated WGN background noise, $\eta(n)$, to $y(n)$ to achieve a SNR of 30 dB. As discussed in Section 2.1.1 of Chapter 2 the local talker signal, $\nu(n)$, was set to 0 for the duration of

the simulations since adaptation of the EC is only performed under quiet local talker conditions.

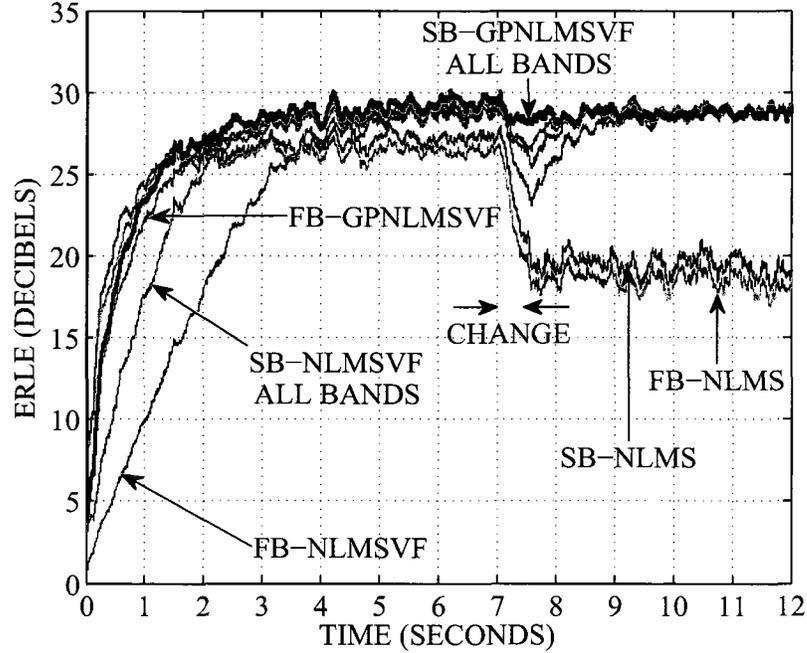


Figure 6.2: ERLE for FB-NLMS, FB-NLMSVF, FB-GPNLMSVF, SB-NLMS, SB-NLMSVF, and SB-GPNLMSVF with WGN input under changing nonlinear echo path conditions and a step size of 0.5 for both the linear and quadratic adaptive filters.

Figure 6.2 compares the tracking performance of the SB-GPNLMSVF structure to the SB-NLMSVF and SB-NLMS structures as well as to their fullband structure counterparts. The SB-GPNLMSVF structure was implemented based on the SB-GGNLMSVF structure as discussed in Section 6.1.1, with $\gamma_{1,i} = \gamma_{2,i} = 1$, $\eta_{1,i} = \eta_{2,i} = 0.9999$, $\beta_{1,i} = \beta_{2,i} = 0.5$, $\delta_{1,i} = \delta_{2,i} = 0.00001$, and $\xi_{1,i} = \xi_{2,i} = 0.00001$ in all subbands. The SB-NLMSVF structure was also implemented based on the SB-GGNLMSVF structure as discussed in Section 6.1.1, with $\delta_{1,i} = \delta_{2,i} = 0.00001$ in all subbands. The SB-NLMS structure arose from the SB-NLMSVF by setting $N_{2, sb_i} = 0$ in all subbands with $\delta_{1,i} = 0.00001$ in all subbands. The fullband structures were implemented with the same parameter values as their corresponding subband

counterparts. Furthermore, the GGNLMS algorithm parameters for the compared structures were set to the values outlined based on [12], and because they resulted in good ERLE performance for the given overall step size. The linear and quadratic adaptive filters of the fullband gradient proportionate normalized least-mean-square Volterra filter (FB-GPNLMSVF) and FB-NLMSVF structures were set to lengths of $N_1 = 2000$ and $N_2(N_2 + 1)/2 = 1275$ respectively, where $N_2 = 50$. The linear adaptive filter of the FB-NLMS structure, which resulted from the FB-NLMSVF structure by setting $N_2 = 0$, was set to a length of $N_1 = 2000$. The linear and quadratic adaptive filters of the SB-GPNLMSVF and SB-NLMSVF structures were set to $N_{1, sb_i} = N_1/M = 500$ and $N_{2, sb_i} = 120$ in all subbands respectively. The linear adaptive filters of the SB-NLMS structure were set to $N_{1, sb_i} = N_1/M = 500$ in all subbands. The length of the quadratic subband adaptive filters were set to $N_{2, sb_i} = 120$ because this filter length resulted in approximately equivalent steady-state performance with the corresponding fullband structures. A step size of 0.5 was used in adapting the linear and quadratic filters of all structures. The SB-GPNLMSVF structure with second order adaptive Volterra filtering applied in all $K/2 = 4$ subbands suffers a smaller drop in ERLE performance during the nonlinear echo path change, and converges back to steady-state operation after the change faster than the other nonlinear AEC structures. Figure 6.3 shows a closer view of the ERLE performance of all structures around the time of the nonlinear echo path fluctuation. The SB-GPNLMSVF structure achieves up to 1 dB, 2 dB, and 4 dB higher ERLE during the nonlinear echo path change as well as during reconvergence back to steady-state operation compared to the FB-GPNLMSVF, SB-NLMSVF, and FB-NLMSVF structures respectively. Compared to the linear only SB-NLMS and FB-NLMS structures, the SB-GPNLMSVF structure attains up to 7 dB and 8 dB higher ERLE respectively during the nonlinear echo path variation. After the change the nonlinear AEC structures perform at approximately 9 dB and 10 dB higher ERLE than the linear SB-NLMS and FB-NLMS structures respectively. This is due to the increase in magnitude of

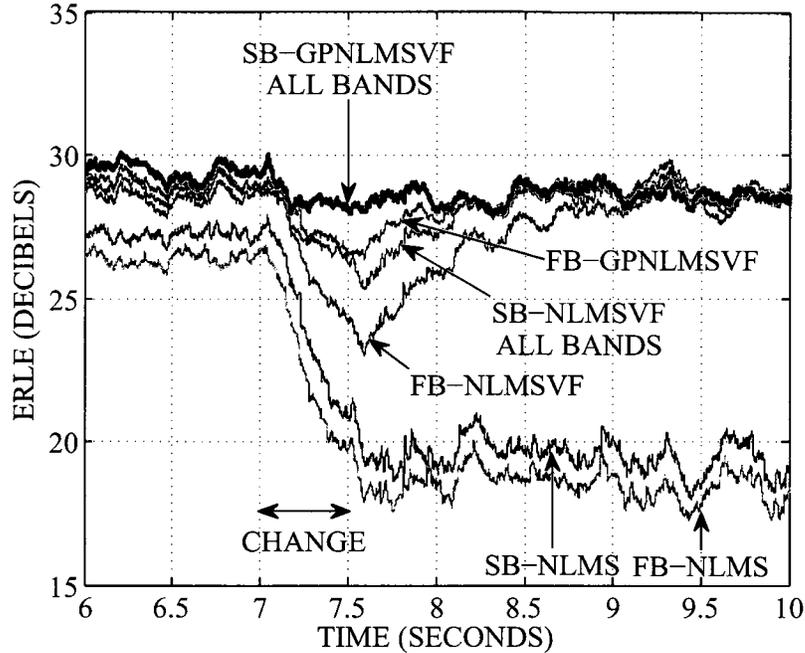


Figure 6.3: ERLE from 6 to 10 seconds for FB-NLMS, FB-NLMSVF, FB-GPNLMSVF, SB-NLMS, SB-NLMSVF, and SB-GPNLMSVF with WGN input under changing nonlinear echo path conditions and a step size of 0.5 for both the linear and quadratic adaptive filters.

the quadratic Volterra kernel taps. The improved tracking and reconvergence performance of the SB-GPNLMSVF structure can be attributed to the individual gradient proportionate quadratic subband filter tap step sizes, and to the shorter subband adaptive filters as discussed in Chapter 4 for the linear gradient proportionate step size structures. Also, during initial convergence at the beginning of the simulation the SB-GPNLMSVF structure converges significantly faster than the SB-NLMSVF and FB-NLMSVF structures by up to 5 dB and 12 dB higher ERLE respectively. Furthermore, the SB-GPNLMSVF structure initially converges slightly faster than the FB-GPNLMSVF structure by up to 1 dB higher ERLE, and achieves similar initial convergence performance compared to the linear only SB-NLMS and FB-NLMS structures which have much smaller overall filter memory requirements. The improvement in initial convergence performance of the SB-GPNLMSVF structure results for the

same aforementioned reasons as its improved tracking performance. As well, during stationary conditions within the acoustic environment the SB-GPNLMSVF structure achieves the same constant steady-state ERLE performance of approximately 30 dB as the other nonlinear structures.

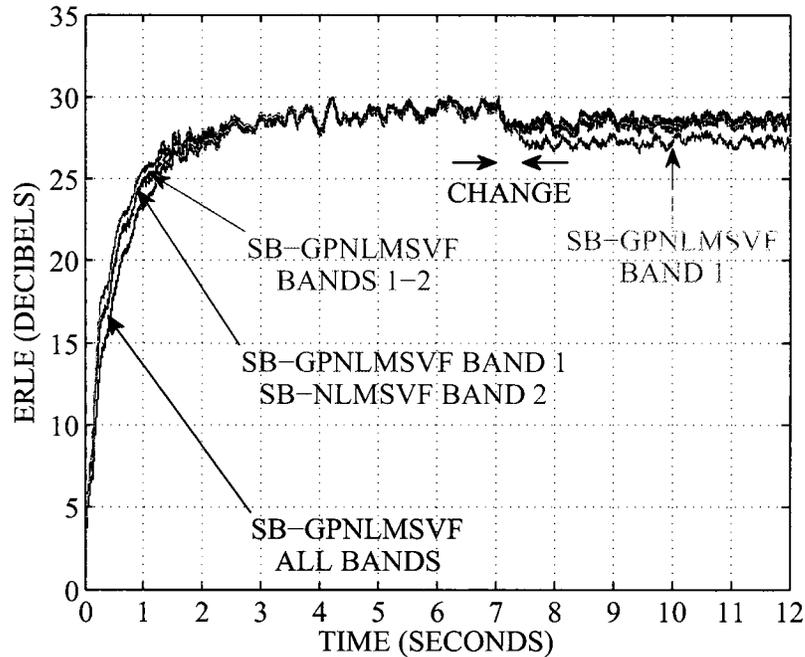


Figure 6.4: ERLE for SB-GPNLMSVF with WGN input under changing nonlinear echo path conditions and the GPNLMS algorithm applied in only specific subbands.

The affect on tracking performance by applying the GPNLMS algorithm to the quadratic filters of the SB-GPNLMSVF structure on an individual subband basis is investigated in Figure 6.4. The performance of the SB-GPNLMSVF structure with the GPNLMS algorithm applied to the quadratic adaptive filters in all, only the 0–1 kHz, and only the 0–1 kHz and 1–2 kHz subbands along with the GPNLMS algorithm applied to the linear adaptive filters in all subbands, is shown under the previous simulation conditions. As well, the tracking performance of the SB-GPNLMSVF structure is shown with the GPNLMS algorithm applied to the quadratic adaptive filter in only the 0–1 kHz subband and the NLMS algorithm applied to the quadratic adaptive filter in only the 1–2 kHz subband, with GPNLMS adaptive linear filters in all subbands.

A closer view of the ERLE performance of the SB-GPNLMSVF structure around the time of the nonlinear echo path fluctuation is shown in Figure 6.5. During the change

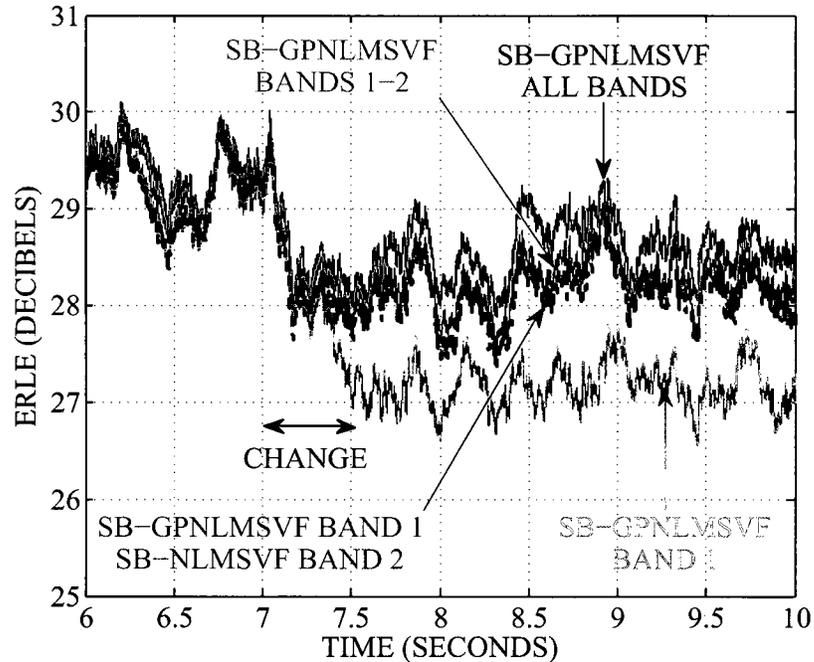


Figure 6.5: ERLE from 6 to 10 seconds for SB-GPNLMSVF with WGN input under changing nonlinear echo path conditions and the GPNLMS algorithm applied in only specific subbands.

in the quadratic Volterra kernel taps and the subsequent reconvergence to steady-state operation, the SB-GPNLMSVF structure with the GPNLMS algorithm applied to the quadratic filters in all (SB-GPNLMSVF ALL BANDS) and only the 0–1 kHz and 1–2 kHz subbands (SB-GPNLMSVF BANDS 1-2) achieves nearly identical ERLE tracking performance. This verifies that the quadratic Volterra kernel variations occur mainly at low frequencies and that the SB-GPNLMSVF structure is able to exploit this low frequency locality of the variation. Furthermore, the SB-GPNLMSVF structure with the GPNLMS algorithm used to adapt the quadratic filter in only the 0–1 kHz subband and a NLMS adaptive quadratic filter in only the 1–2 kHz subband (SB-GPNLMSVF BAND 1, SB-NLMSVF BAND 2) performs virtually the same as with GPNLMS adaptive quadratic filters in both the 0–1 kHz and 1–2 kHz subbands.

This demonstrates that the vast majority of the nonlinear echo path change occurs at frequencies below 1 kHz, and that the SB-GPNLMSVF structure can take advantage of this by applying the fast tracking GPNLMS algorithm in only the correspondingly affected subbands. With the GPNLMS algorithm applied to the quadratic filter in only the 0–1 kHz subband (SB-GPNLMSVF BAND 1), the SB-GPNLMSVF structure attains slightly worse ERLE performance by up to 2 dB compared to its other variants during and following the nonlinear echo path change. This is due to the non-modeling of the small amount of distortion that occurs for frequencies above 1 kHz. The SB-GPNLMSVF structure with quadratic adaptive Volterra filtering in all and only the 0–1 kHz and 1–2 kHz subbands achieves almost equivalent steady-state ERLE performance before and after the nonlinear echo path disturbance, which verifies that the nonlinear distortion itself occurs at frequencies below approximately 2 kHz. The majority of the distortion occurs below 1 kHz as shown by the slightly lower steady-state ERLE reached by the SB-GPNLMSVF structure with quadratic adaptive Volterra filtering in only the 0–1 kHz subband (SB-GPNLMSVF BAND 1), as initially discussed in Chapter 5. The SB-GPNLMSVF structure with GPNLMS adaptive quadratic filters in all subbands (SB-GPNLMSVF ALL BANDS) lags in initial convergence by up to 2 dB and 3 dB compared to the variants with quadratic adaptive filtering in only the 0–1 kHz and 1–2 kHz subbands (SB-GPNLMSVF BANDS 1-2 and SB-GPNLMSVF BAND 1, SB-NLMSVF BAND 2), and only the 0–1 kHz subband (SB-GPNLMSVF BAND 1) respectively. This lag in initial convergence performance is a result of applying quadratic adaptive filtering in all subbands compared to a subset of all subbands, which have a shorter overall filter memory length.

6.2.2.2 White Noise Excitation Signal With a Changing Linear and Quadratic Volterra Kernel

In this section the AEC performance of the SB-GPNLMSVF structure along with its fullband and linear only structure counterparts is outlined under changing conditions

within both the linear and nonlinear portion of the acoustic echo path. Simulations were performed for synthetic microphone signals created based on the measured linear and quadratic LEMS Volterra kernels, obtained under unobstructed and obstructed low and high volume conditions, from the conference phone configuration with lengths $L_1 = 2000$ and $L_2(L_2 + 1)/2 = 1275$ respectively, where $L_2 = 50$. The synthetic echo and microphone signals were created under the same conditions as in Section 6.2.2.1, where the change in the echo signal was created by linearly interpolating between the two experimentally measured sets of linear and quadratic Volterra kernels simultaneously, to simulate a local talkers hand moving towards the phone set to increase the loudspeaker volume level.

Figure 6.6 compares the tracking performance of the SB-GPNLMSVF structure to the SB-NLMSVF, SB-GPNLMS, and SB-NLMS structures as well as to their full-band structure counterparts. All structures were implemented with the same parameter settings as in Section 6.2.2.1 where the SB-GPNLMS structure resulted from the SB-GPNLMSVF structure by setting $N_{2, sb_i} = 0$ in all subbands. As for the change in only the nonlinear portion of the echo path in Section 6.2.2.1, the SB-GPNLMSVF structure again suffers a smaller drop in ERLE performance during the echo path change and converges back to steady-state operation after the change faster than the other AEC structures. The SB-GPNLMSVF structure achieves up to 2 dB, 5 dB, and 8 dB higher ERLE during the echo path change as well as during reconvergence back to steady-state operation compared to the FB-GPNLMSVF, SB-NLMSVF, and FB-NLMSVF structures respectively. The linear only SB-GPNLMS and SB-NLMS structures are unable to model the quadratic Volterra kernel, however the SB-GPNLMS structure provides up to 4 dB higher ERLE during the echo path change compared to the SB-NLMS structure due to its fast tracking of the linear echo path component. During steady-state operation after the echo path change, the nonlinear AEC structures outperform the linear structures by up to 9 dB higher ERLE due to the increase in loudspeaker distortion level. Prior to the echo path change, the nonlinear

AEC structures outperform the linear structures by only approximately 2 dB due to the small amount of loudspeaker distortion. The initial convergence performance of all the AEC structures is roughly the same as discussed in Section 6.2.2.1, with the SB-GPNLMSVF structure lagging its linear only SB-GPNLMS structure counterpart by at most 5 dB of ERLE. The improved tracking and convergence performance of the SB-GPNLMSVF structure can be attributed to its individual gradient proportionate linear and quadratic subband filter tap step sizes and shorter-length subband adaptive filters. Furthermore, during stationary conditions within the acoustic environment the SB-GPNLMSVF structure achieves the same constant steady-state ERLE performance of approximately 30 dB as the other nonlinear structures.

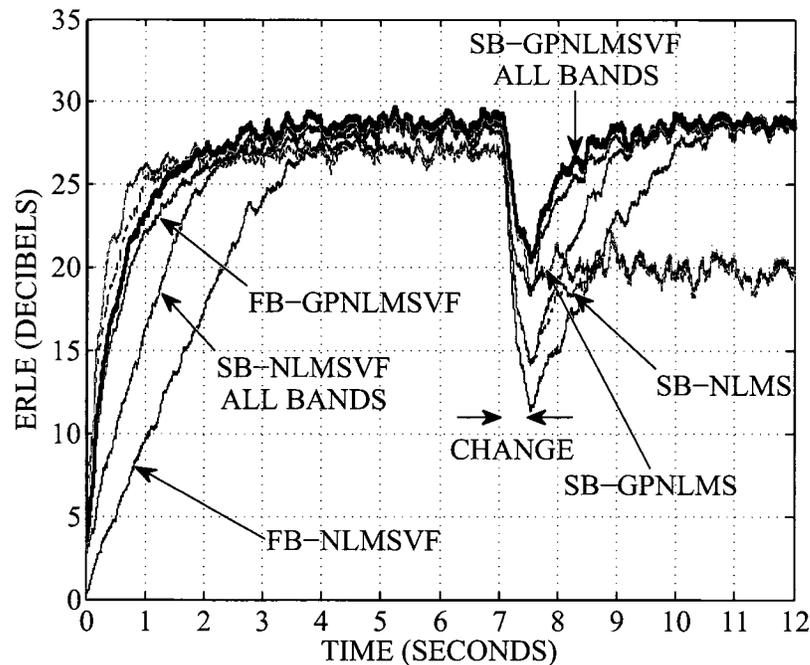


Figure 6.6: ERLE for FB-NLMSVF, FB-GPNLMSVF, SB-NLMS, SB-GPNLMS, SB-NLMSVF, and SB-GPNLMSVF with WGN input under changing linear and nonlinear echo path conditions and a step size of 0.5 for both the linear and quadratic adaptive filters.

The impact on tracking performance of the SB-GPNLMSVF structure with the GPNLMS algorithm implemented in specific linear and quadratic subband filters

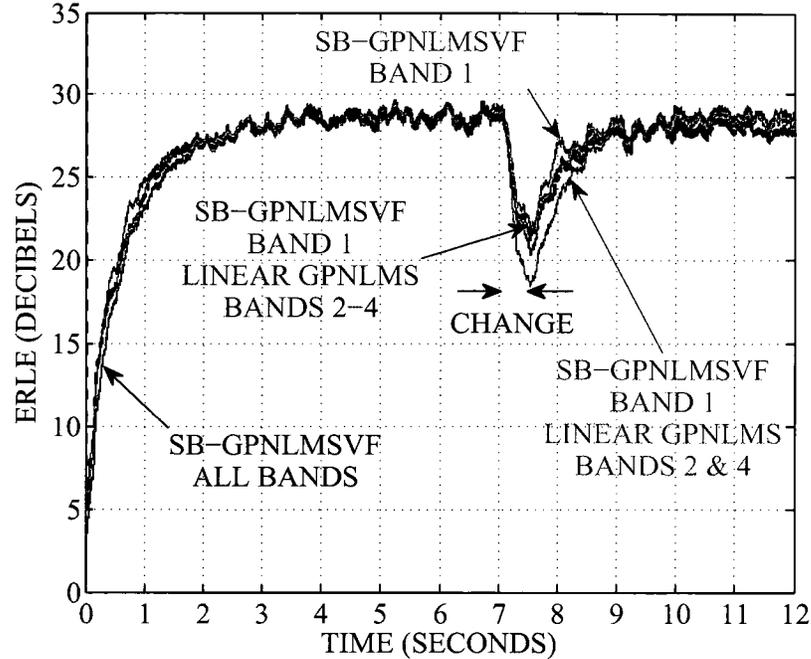


Figure 6.7: ERLE for SB-GPNLMSVF with WGN input under changing linear and nonlinear echo path conditions and the GPNLMS algorithm applied in only specific subbands.

is shown in Figure 6.7. The performance of the SB-GPNLMSVF structure with a GPNLMS adaptive quadratic filter in only the 0–1 kHz subband and a NLMS adaptive quadratic filter in only the 1–2 kHz subband along with GPNLMS adaptive linear filters in all subbands, in only the 1–4 kHz subbands, and in only the 1–2 kHz and 3–4 kHz subbands is shown under the previous simulation conditions. The performance of the standard SB-GPNLMSVF structure with GPNLMS adaptive linear and quadratic filters in all subbands is also shown. The NLMS algorithm was used to adapt the linear filters in the subbands that did not implement the GPNLMS algorithm. Figure 6.8 shows a closer view of the ERLE performance of the structures around the time of the echo path fluctuation. The SB-GPNLMSVF structure with a GPNLMS adaptive quadratic filter in only the 0–1 kHz subband and a NLMS adaptive quadratic filter in only the 1–2 kHz subband (SB-GPNLMSVF BAND 1) achieves nearly identical ERLE performance before and after the echo path change,

as well as slightly improved tracking performance during the change, as the same structure with the GPNLMS algorithm applied to the quadratic filters in all subbands (SB-GPNLMSVF ALL BANDS). This demonstrates again that the nonlinear echo path component is confined to frequencies below approximately 2 kHz, with the majority of its variations occurring below roughly 1 kHz as shown in Section 6.2.2.1. The slight improvement in tracking performance during the echo path change is due to the non-modeling of the insignificant loudspeaker distortion in the 2–4 kHz subbands, which avoids needless adaptation that slows tracking. The SB-GPNLMSVF

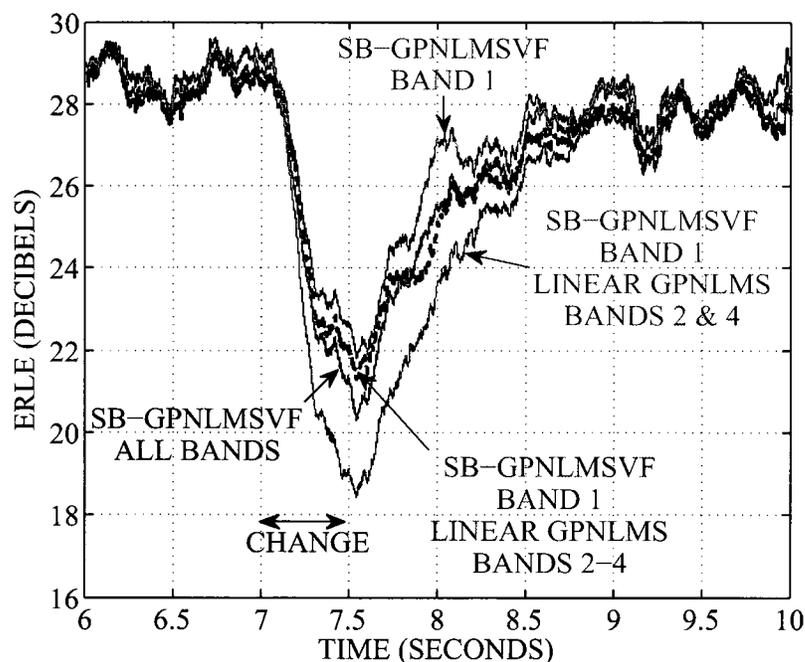


Figure 6.8: ERLE from 6 to 10 seconds for SB-GPNLMSVF with WGN input under changing linear and nonlinear echo path conditions and the GPNLMS algorithm applied in only specific subbands.

structure with a GPNLMS adaptive quadratic filter in only the 0–1 kHz subband and GPNLMS adaptive linear filters in only the 1–4 kHz subbands (SB-GPNLMSVF BAND 1, LINEAR GPNLMS BANDS 2-4), and in only the 1–2 kHz and 3–4 kHz subbands (SB-GPNLMSVF BAND 1, LINEAR GPNLMS BANDS 2&4) attains up to 1 dB, and 3 dB lower ERLE respectively during the echo path change compared

to the same structure with the GPNLMS adaptive linear filters in all subbands (SB-GPNLMSVF BAND 1). As initially shown in Chapter 4 the impact on the linear portion of the echo path is very small in the 0–1 kHz subband, as verified by the very similar ERLE performance of the SB-GPNLMSVF structure with GPNLMS adaptive linear filters in all (SB-GPNLMSVF BAND 1) and only the 1–4 kHz subbands (SB-GPNLMSVF BAND 1, LINEAR GPNLMS BANDS 2-4). This is again due to the correspondingly low frequency long wavelengths of sound being less impeded by the small hand obstruction than the shorter wavelengths at higher frequencies. By applying the fast tracking GPNLMS algorithm to the linear and quadratic filters in the subbands most affected by the echo path change and the simpler slower tracking NLMS algorithm to the filters in the unaffected subbands, the SB-GPNLMSVF structure can achieve equivalent AEC performance with reduced complexity compared to the same structure with GPNLMS based filters in all subbands. Furthermore, the per subband tracking ability of the SB-GPNLMSVF structure allows for slightly improved AEC performance at a much reduced computational cost compared to its corresponding fullband structure (FB-GPNLMSVF) shown in Figure 6.6. The reduced computational cost of the SB-GPNLMSVF structure compared to its fullband counterpart is discussed in Section 6.2.4.

Figure 6.9 displays the performance of the SB-GPNLMSVF structure under an increasing amount of undermodeling of the loudspeaker nonlinearity. The previous simulation settings were used including the SB-GPNLMSVF structure with a GPNLMS adaptive quadratic filter in only the 0–1 kHz subband and a NLMS adaptive quadratic filter in only the 1–2 kHz subband, as well as GPNLMS based linear filters in only the 1–4 kHz subbands with NLMS based linear filtering in the 0–1 kHz subband. The length of the quadratic subband filters for the SB-GPNLMSVF structure were decreased from $N_{2, sb_{1-2}} = 120$ to $N_{2, sb_{1-2}} = 60$, and $N_{2, sb_{1-2}} = 12$. The performance of the SB-NLMS structure is also included in Figure 6.9 to give a lower bound on the achievable AEC performance when only linear adaptive filtering is performed.

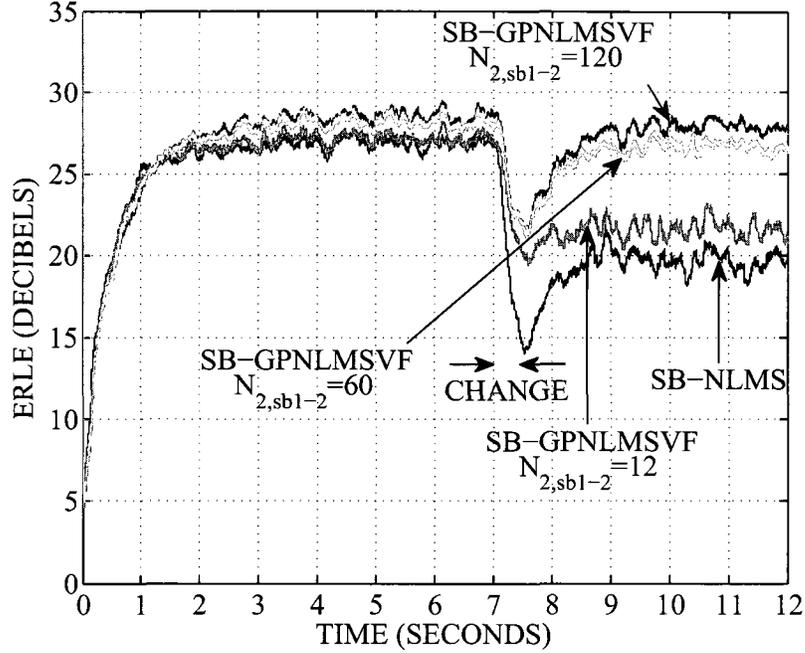


Figure 6.9: ERLE for SB-GPNLMSVF with WGN input under changing linear and nonlinear echo path conditions and a decreasing $N_{2, sb1-2}$.

With $N_{2, sb1-2} = 60$, and $N_{2, sb1-2} = 12$ the SB-GPNLMSVF structure achieves up to 1 dB, and 5 dB lower ERLE respectively, compared to the same structure with $N_{2, sb1-2} = 120$ during steady-state operation after the echo path change. During the echo path change the SB-GPNLMSVF structure with $N_{2, sb1-2} = 60$ perform nearly identically to the same structure with $N_{2, sb1-2} = 120$, while with $N_{2, sb1-2} = 12$ only up to 2 dB lower ERLE is attained. This verifies that good tracking performance can still be achieved by the SB-GPNLMSVF structure with undermodeled quadratic subband filters. The SB-GPNLMSVF structure with undermodeled quadratic subband filters attains tracking performance similar to the same structure with longer quadratic filters, by modeling only the high energy quadratic Volterra kernel taps which experience the greatest fluctuation during the echo path change, as shown experimentally in Section 3.3 of Chapter 3. Prior to the echo path fluctuation the SB-GPNLMSVF structure with $N_{2, sb1-2} = 60$ reaches the same level of ERLE performance as with $N_{2, sb1-2} = 120$, due to the small amount of loudspeaker distortion

resulting from the low volume input signal level. Even with very short quadratic filters of length $N_{2, sb_{1-2}} = 16$, the SB-GPNLMSVF structure is still able to provide partial loudspeaker distortion modeling and thus improved AEC performance by up to 5 dB during and immediately after the echo path change, and up to 2 dB during steady-state operation after the echo path change compared to the linear only SB-NLMS structure. As initially discussed in Chapter 5 for the SB-NLMSVF structure, by modeling only the highest energy second order Volterra kernel coefficients that lie along its main and adjacent diagonals and ignoring the insignificant coefficients, the SB-GPNLMSVF structure with very short quadratic adaptive filter lengths is able to achieve significantly improved ERLE performance compared to a linear only structure.

6.2.3 Results Based on Measured Reference and Input Data

The performance of the SB-GPNLMSVF structure along with its variants are compared to each other as well as to their fullband counterparts under experimentally measured changing acoustic conditions. The experimental setup and methodology described in Section 3.1 of Chapter 3 was used to record microphone and loudspeaker data from the smartphone LEMS configuration for a 30 second speech excitation signal. During the recording an echo path fluctuation was introduced by continual hand waving in front of the phone set along with volume changes between high and low levels at a rate of approximately 1 Hz.

The AEC performance of the SB-GPNLMSVF structure is compared to the FB-GPNLMSVF, SB-NLMSVF, SB-GPNLMS, and SB-NLMS structures, along with the FB-IPNLMSVF structure outlined in [88], in Figures 6.10 and 6.11. All structures were implemented with the same parameter settings as in the beginning of Section 6.2.2.1 for Figure 6.2, with $N_{2, sb_i} = 120$ in all subbands of the appropriate structures. Also, the linear and quadratic filters of the FB-IPNLMSVF structure were implemented based on the IPNLMS algorithm, which was derived from the GPNLMS

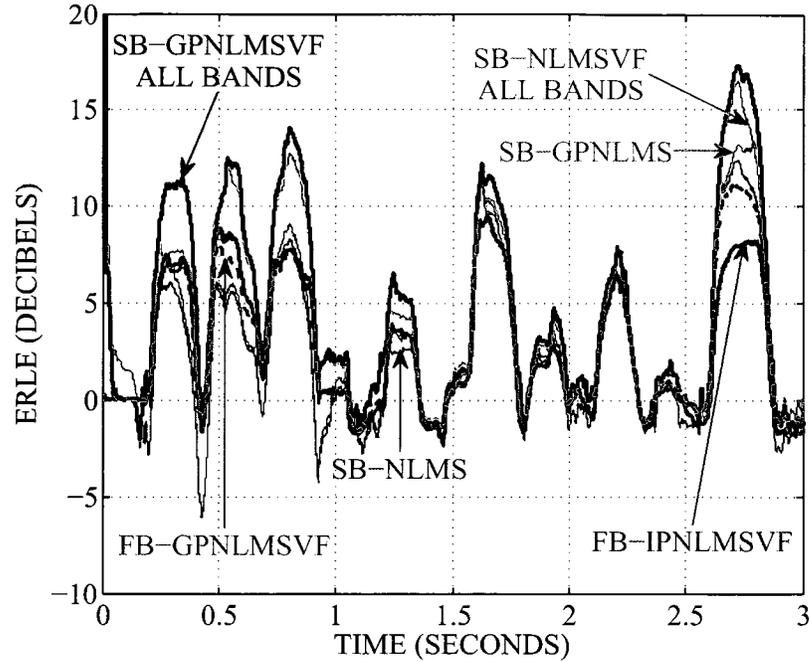


Figure 6.10: ERLE from 0 to 3 seconds for FB-GPNLMSVF, FB-IPNLMSVF, SB-NLMS, SB-GPNLMS, SB-NLMSVF, and SB-GPNLMSVF with measured speech input under changing linear and nonlinear echo path conditions and a step size of 0.5 for both the linear and quadratic adaptive filters.

algorithm by setting $\gamma_1 = \gamma_2 = 0$. During initial convergence the SB-GPNLMSVF structure achieves up to 5 dB, 6 dB, and 8 dB higher ERLE during the echo path changes compared to the FB-IPNLMSVF, FB-GPNLMSVF, and SB-NLMSVF structures respectively, as shown in Figure 6.10. The linear only SB-GPNLMS and SB-NLMS structures, which are unable to model the quadratic Volterra kernel and its variations, achieve up to 4 dB and 8 dB lower ERLE respectively compared to the SB-GPNLMSVF structure during the initial convergence period. During the echo path fluctuations between 10 and 13 seconds the SB-GPNLMSVF structure outperforms the FB-GPNLMSVF, FB-IPNLMSVF, SB-NLMSVF, SB-GPNLMS and SB-NLMS structures by up to 5 dB, 5 dB, 13 dB, 6 dB, and 18 dB higher ERLE respectively, as seen in Figure 6.11. The improved performance of the SB-GPNLMSVF structure is most apparent during the echo path change between 11 and 12 seconds compared to

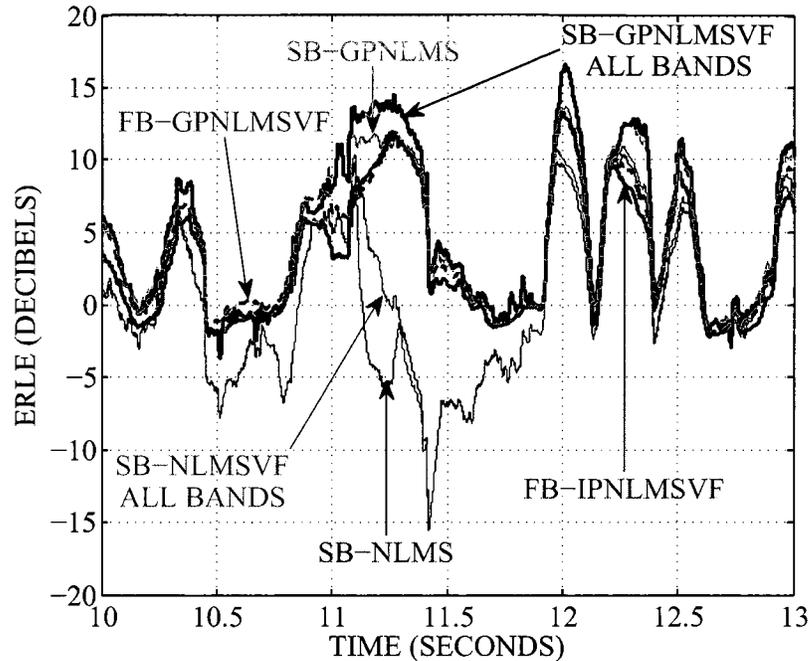


Figure 6.11: ERLE from 10 to 13 seconds for FB-GPNLMSVF, FB-IPNLMSVF, SB-NLMS, SB-GPNLMS, SB-NLMSVF, and SB-GPNLMSVF with measured speech input under changing linear and nonlinear echo path conditions and a step size of 0.5 for both the linear and quadratic adaptive filters.

its fullband counterpart. During this time the SB-GPNLMSVF structure provides improved ERLE performance of up to 5 dB compared to the FB-GPNLMSVF structure due to its shorter-length linear and quadratic subband adaptive filters. Also, during this time the SB-GPNLMSVF structure attains up to 3 dB higher ERLE than the linear only SB-GPNLMS structure due to its modeling and tracking of the nonlinear loudspeaker distortion. Furthermore, large improvements in ERLE of 13 and 18 dB are achieved by the SB-GPNLMSVF structure compared to the SB-NLMSVF and SB-NLMS structures, due to the fast tracking of linear and nonlinear echo path components by the gradient based linear and quadratic subband adaptive filters. Also, it should be noted that the FB-GPNLMSVF structure achieves improved ERLE performance by up to 3 dB compared to the FB-IPNLMSVF structure due again to the gradient based linear and quadratic subband adaptive filters, as seen in Figure 6.11.

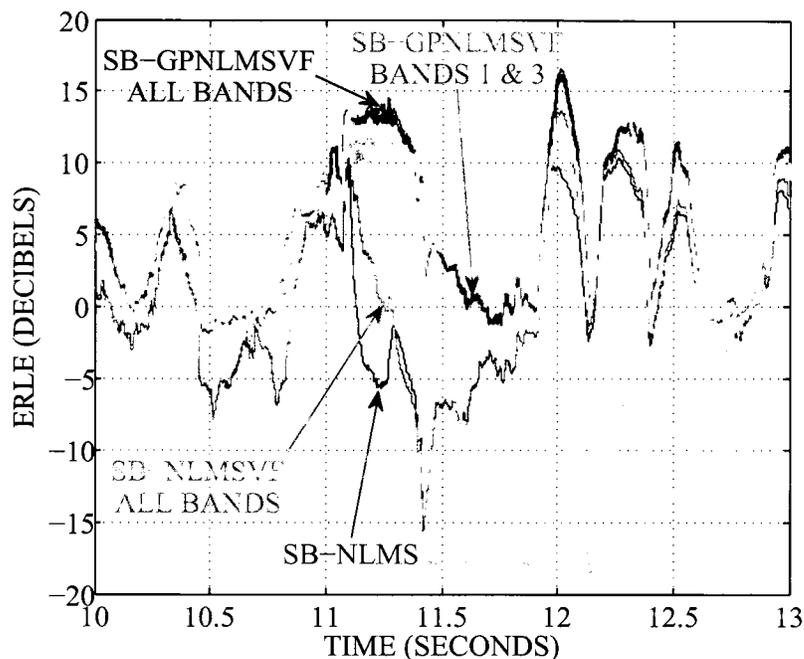


Figure 6.12: ERLE from 10 to 13 seconds for SB-GPNLMSVF with measured speech input under changing linear and nonlinear echo path conditions and the GPNLMS algorithm applied in only specific subbands.

Figure 6.12 compares the performance of the SB-GPNLMSVF structure with the GPNLMS algorithm applied in specific linear and quadratic subbands only. The performance of the SB-GPNLMSVF structure with the GPNLMS algorithm applied to the quadratic filters in only the 0–1 kHz and 2–3 kHz subbands and in only the 2–3 kHz subband, along with the GPNLMS algorithm applied to the linear filters in all subbands and in only the 1–4 kHz subbands is shown under the previous simulation conditions. When the GPNLMS algorithm was used to adapt the quadratic filter in only the 2–3 kHz subband, the NLMS algorithm was applied to the quadratic filter in only 0–1 kHz subband. Again, the simpler slower tracking NLMS algorithm was used to adapt the linear filters in the subbands that did not implement the GPNLMS algorithm. The SB-GPNLMSVF structure with the GPNLMS algorithm applied to the quadratic filters in only the 0–1 kHz and 2–3 kHz subbands (SB-GPNLMSVF BANDS 1&3), achieves nearly equivalent ERLE performance as the

same structure with the GPNLMS algorithm applied to the quadratic filters in all subbands (SB-GPNLMSVF ALL BANDS). This verifies that the majority of the echo path nonlinearity and variations in its magnitude occur for frequencies between 0–1 and 2–3 kHz. During the echo path changes the SB-GPNLMSVF structure with the GPNLMS algorithm applied to the quadratic filter in only the 2–3 kHz subband and to the linear filters in only the 1–4 kHz subbands (SB-GPNLMSVF BAND 3, LINEAR GPNLMS BANDS 2-4), attains up to 3 dB lower ERLE respectively compared to the same structure with the GPNLMS algorithm applied to the quadratic filters in only the 0–1 kHz and 2–3 kHz subbands and to the linear filters in all subbands (SB-GPNLMSVF BANDS 1&3). This verifies that nonlinear echo path variations are mainly confined to the 2–3 kHz region, and that the impact of the linear portion of the echo path variations due to the small hand waving obstruction is very small in the 0–1 kHz subband, as discussed previously in Section 6.2.2.2 as well as in Chapter 4. By applying the fast tracking GPNLMS algorithm to the linear and quadratic filters in the subbands most affected by the echo path changes, and the simpler slower tracking NLMS algorithm to the filters in the unaffected subbands, the SB-GPNLMSVF structure can achieve similar AEC performance with reduced complexity compared to the same structure with GPNLMS based filters in all subbands. As well, the per subband tracking ability of the SB-GPNLMSVF structure allows for similar or slightly improved AEC performance at a much reduced computational cost compared to its corresponding fullband structure (FB-GPNLMSVF) shown in Figure 6.11. Again, the reduced computational cost of the SB-GPNLMSVF structure compared to its fullband counterpart is discussed in Section 6.2.4.

The performance of the SB-GPNLMSVF structure with an increasing amount of undermodeling of the loudspeaker nonlinearity is shown in Figure 6.13. The previous simulation settings were used for the SB-GPNLMSVF structure with the GPNLMS algorithm applied to the quadratic filter in only the 2–3 kHz subband as well as GPNLMS based linear filters in only the 1–4 kHz subbands, with NLMS based linear

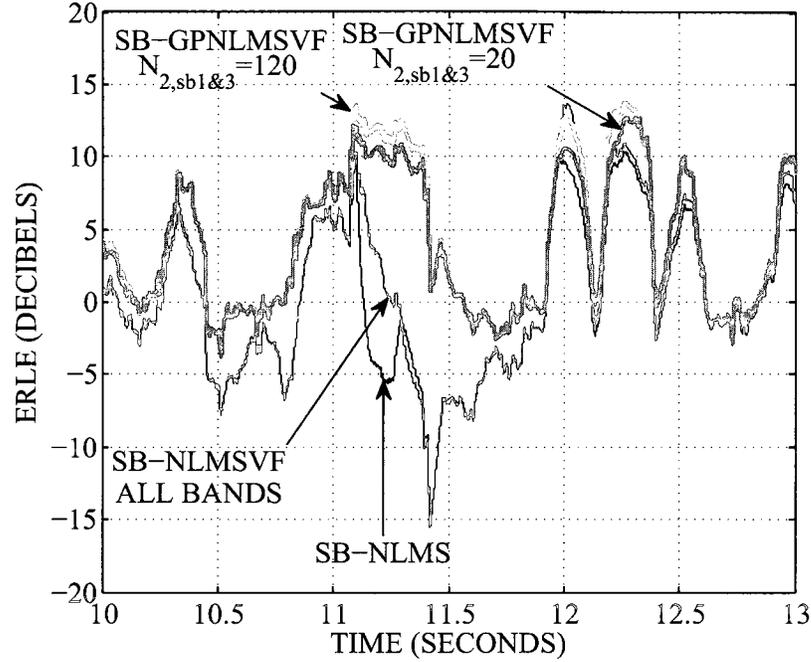


Figure 6.13: ERLE from 10 to 13 seconds for SB-GPNLMSVF with measured speech input under changing linear and nonlinear echo path conditions and a decreasing $N_{2, sb1\&3}$.

and quadratic filtering in only the 0–1 kHz subband. The length of the quadratic subband filters for the SB-GPNLMSVF structure were decreased from $N_{2, sb1\&3} = 120$ to $N_{2, sb1\&3} = 20$. The SB-GPNLMSVF structure with $N_{2, sb1\&3} = 20$ performs nearly as well as the same structure with $N_{2, sb1\&3} = 120$, achieving up to only 2 dB lower ERLE. This indicates that the SB-GPNLMSVF structure with $N_{2, sb1\&3} = 20$ is modeling the majority of the significant high amplitude quadratic Volterra kernel coefficients that experience the greatest variation in magnitude during the echo path change.

The impact on ERLE performance of the SB-GPNLMSVF structure by applying the GPNLMS algorithm in specific linear and quadratic subbands is investigated in Figure 6.14 for an increasing number of subbands with the same OSR. The same filter bank settings as those used in Chapter 5 for Figure 5.12 were used along with the previous GPNLMS parameter settings for the linear and quadratic subband filters. The linear adaptive filters of the SB-GPNLMSVF and SB-NLMS structures based on the

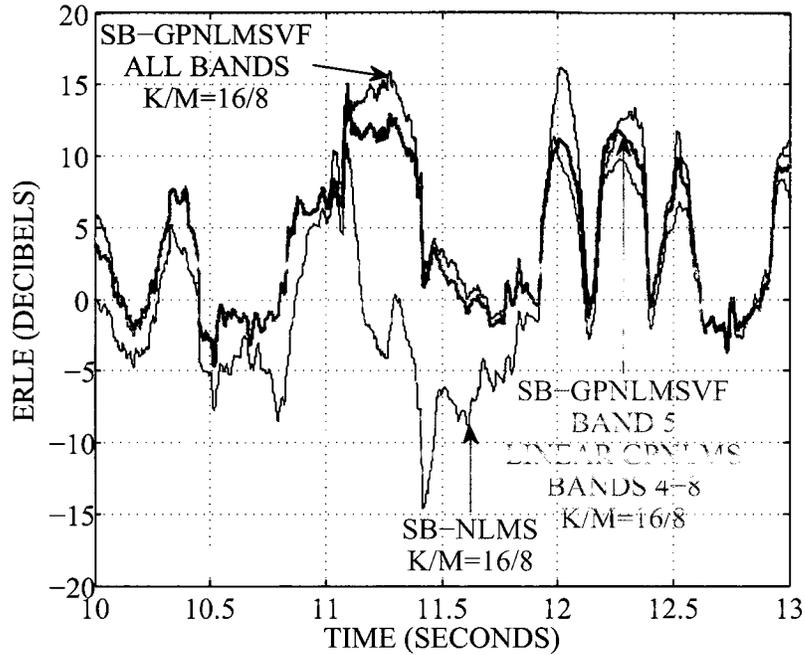


Figure 6.14: ERLE from 10 to 13 seconds for SB-GPNLMSVF with measured speech input under changing linear and nonlinear echo path conditions and an increasing number of subbands for the same OSR.

$K = 16$ $M = 8$ filter banks were set to a length of $N_{1, sb_i} = 250$ in all subbands, while the linear adaptive filters of the $K = 8$ $M = 4$ SB-GPNLMSVF structure were set to $N_{1, sb_i} = 500$ in all subbands. The quadratic adaptive filters of the SB-GPNLMSVF structure based on the $K = 16$ $M = 8$ filter banks were set to a length of $N_{2, sb_i} = 30$, while the quadratic adaptive filters of the $K = 8$ $M = 4$ SB-GPNLMSVF structure were set to $N_{2, sb_i} = 120$. Also, the simpler slower tracking NLMS algorithm was used to adapt the linear filters in the subbands that did not implement the GPNLMS algorithm. Again, a step size of 0.5 was used in adapting the linear and quadratic filters of all structures. The $K = 16$ SB-GPNLMSVF structure with the GPNLMS algorithm applied to the linear filters in only the 1.5–4 kHz subbands, and to the quadratic filter in only the 2–2.5 kHz subband with an NLMS based quadratic filter in only the 0–500 Hz subband (SB-GPNLMSVF BAND 5, LINEAR GPNLMS

BANDS 4-8, $K/M=16/8$) attains very similar ERLE performance as the $K = 8$ SB-GPNLMSVF structure with the GPNLMS algorithm applied to the linear filters in only the 1–4 kHz subbands, and to the quadratic filter in only the 2–3 kHz subband with an NLMS based quadratic filter in only the 0–1 kHz subband (SB-GPNLMSVF BAND 3, LINEAR GPNLMS BANDS 2-4, $K/M=8/4$). Also, these two structures achieve up to 15 dB higher ERLE compared to the $K = 16$ SB-NLMS structure during the changing echo path conditions. This demonstrates the advantage of implementing the SB-GPNLMSVF structure with an increasing number of subbands as the GPNLMS algorithm and quadratic Volterra filtering can be implemented in subbands with a finer frequency range, which can better match the affected frequency range during echo path variations and loudspeaker distortion. Also, implementing the SB-GPNLMSVF structure with a greater number of subbands for the same OSR results in further computational cost reduction, due to the higher decimation ratio and

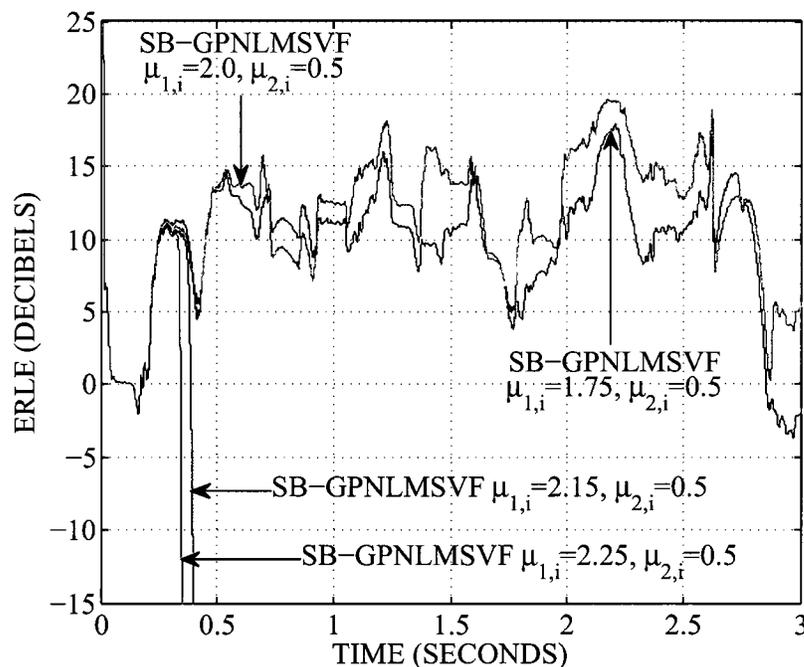


Figure 6.15: ERLE from 0 to 3 seconds for SB-GPNLMSVF with measured speech input under changing linear and nonlinear echo path conditions and an increasing $\mu_{1,i}$.

shorter subband adaptive filters, at the expense of a higher filter bank group delay, as discussed in Section 2.1.1.2 of Chapter 2. Although not shown, the $K = 16$ and $K = 8$ SB-GPNLMSVF structures with the GPNLMS algorithm implemented in the various linear and quadratic subbands as previously discussed, attain approximately the same rate of initial convergence.

The affect on convergence performance of the SB-GPNLMSVF structure for increasing step sizes applied to the linear or quadratic subband adaptive filters is shown in Figures 6.15 and 6.16. The previous parameter settings for the $K = 8$ SB-GPNLMSVF structure from Figures 6.10 and 6.11 were used. With the step sizes of the quadratic filters set to 0.5 in all subbands, increasing the step sizes of the linear filters beyond 2.0 in all subbands results in unstable behaviour as shown in Figure 6.15. Similarly, with the step sizes of the linear filters set to 0.5 in all subbands, increasing the step sizes of the quadratic filters above 2.0 in all subbands also results in

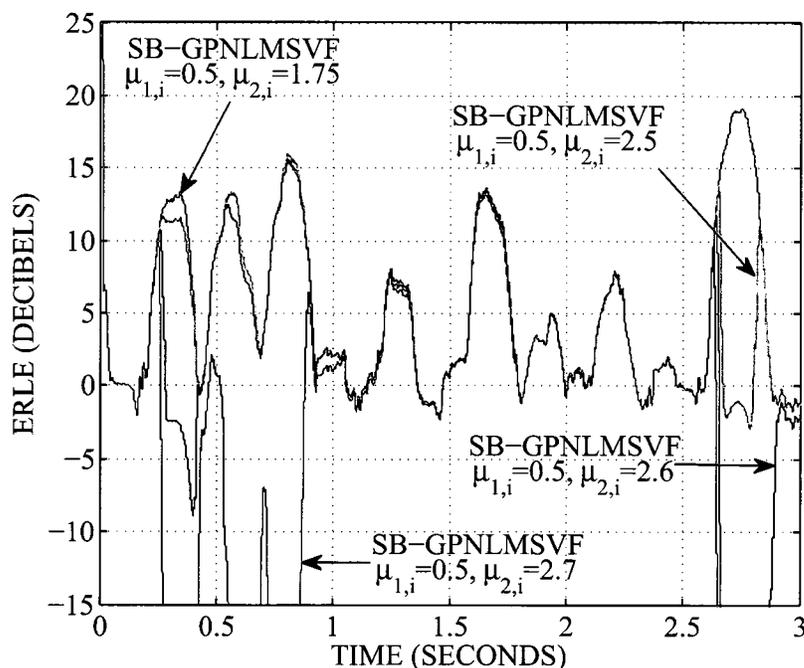


Figure 6.16: ERLE from 0 to 3 seconds for SB-GPNLMSVF with measured speech input under changing linear and nonlinear echo path conditions and an increasing $\mu_{2,i}$.

unstable behaviour as seen in Figure 6.16. This demonstrates that $0 < \mu_{1,i}, \mu_{2,i} < 2$ are reasonable bounds for stable operation of the SB-GPNLMSVF structure. Although not shown, when the step sizes of the quadratic filters are set to 0 in all subbands the SB-GPNLMSVF structure behaves like the linear only SB-GPNLMS structure. Furthermore, when the step sizes of the linear filters are set to 0 in all subbands the SB-GPNLMSVF structure behaves like a pure nonlinear structure by modeling and removing only the quadratic echo path components.

6.2.4 Computational Complexity Results

This section presents the computational complexity requirements of the SB-GGNLMSVF structure in comparison to its fullband and linear only counterparts under the same assumptions and conditions as in Section 5.2.4 of Chapter 5. The computational cost of the SB-GGNLMSVF structure along with its fullband and linear only counterparts were determined based on Tables 6.1 – 6.3. For the complexity results subsequently presented, the GGNLMS parameter settings for all subband and fullband structures were as given in Sections 6.2.2 and 6.2.3.

Figure 6.17 outlines the computational cost for the FB-GPNLMSVF structure, the SB-NLMS structure, and the SB-GPNLMSVF structure with the GPNLMS algorithm applied in specific subbands as the length of the quadratic subband filters increase. The complexity cost of the subband structures were computed based on the filter bank settings outlined in Section 6.2.1, with the length of the linear subband adaptive filters set to $N_{1, sb_i} = 500$ in all subbands and the length of the quadratic subband adaptive filters, N_{2, sb_i} , set equal in all subbands as well. The corresponding linear and quadratic adaptive filter lengths of the FB-GPNLMSVF structure were set to $N_1 = MN_{1, sb_i} = 2000$ and MN_{2, sb_i} respectively, where it was assumed that equal ERLE performance could be achieved by the FB-GPNLMSVF and SB-GPNLMSVF structures for the given filter length relationships. The computational requirements of the SB-GPNLMSVF structure are presented with the GPNLMS algorithm applied

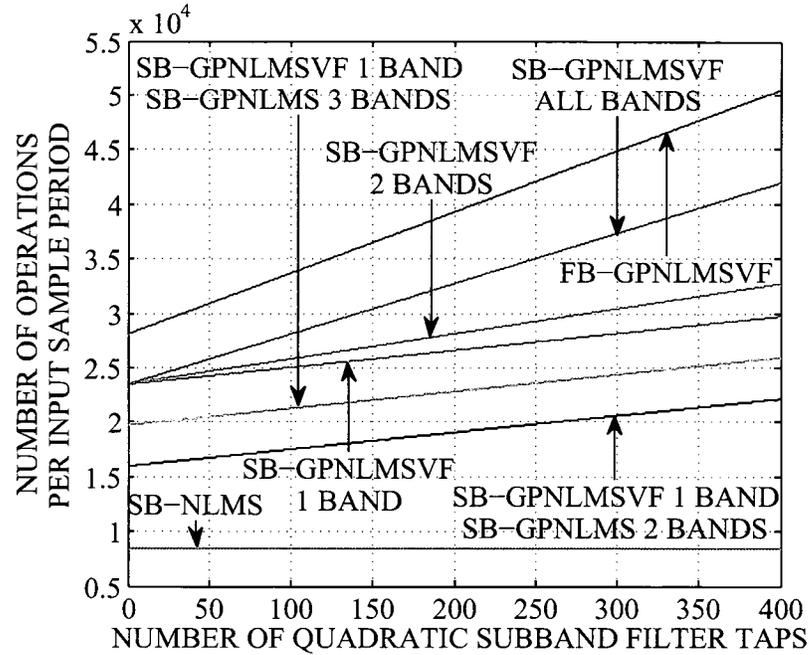


Figure 6.17: Computational complexity for FB-GPNLMSVF, SB-NLMS, and SB-GPNLMSVF with the GPNLMS algorithm applied in only specific subbands.

to the linear filters in all, three, and two subbands with corresponding GPNLMS based quadratic filters in all, only two, and only one subband to illustrate the complexity savings that can be achieved by applying the GPNLMS algorithm in only specific subbands. In determining the computational cost of the SB-GPNLMSVF structure, the complexity of the NLMS algorithm was used for the linear filters that did not implement the GPNLMS algorithm. Furthermore, when the GPNLMS algorithm was applied to a quadratic filter in only one subband, the cost of applying the NLMS algorithm to a quadratic filter in a second subband was also included to represent quadratic Volterra filtering in only two subbands. With the GPNLMS algorithm applied to the linear and quadratic filters in all subbands (SB-GPNLMSVF ALL BANDS), the SB-GPNLMSVF structure requires roughly 15% to 16% fewer operations than the equivalent FB-GPNLMSVF structure as the length of N_{2, sb_i} increases. As the GPNLMS algorithm is used on the linear and quadratic filters in fewer subbands, the complexity of the SB-GPNLMSVF structure reduces significantly and

tends towards that of the SB-NLMS structure. With $N_{2, sb_{1-2}} = 120$ or $N_{2, sb_{1\&3}} = 120$ and the GPNLMS algorithm applied to linear filters in three subbands and to a quadratic filter in one subband (SB-GPNLMSVF 1 BAND, SB-GPNLMS 3 BANDS), the SB-GPNLMSVF structure requires approximately 53% fewer operations than the FB-GPNLMSVF structure to achieve improved ERLE performance. This can be seen by comparing the performance of the SB-GPNLMSVF structure in Figure 6.7 and 6.8 to the performance of the FB-GPNLMSVF structure in Figure 6.6 for a WGN excitation signal. Also, for a measured speech excitation signal the improved performance of the SB-GPNLMSVF structure compared to the FB-GPNLMSVF structure, under these reduced complexity conditions, is evident by comparing Figure 6.12 to Figure 6.11. The FB-GPNLMSVF structure required 1275 fullband quadratic filter taps corresponding to 319 subband quadratic filter taps in Figure 6.17 to achieve the ERLE performance shown in Sections 6.2.2 and 6.2.3. With $N_{2, sb_{1-2}} = 120$ or $N_{2, sb_{1\&3}} = 120$ and the GPNLMS algorithm applied as before (SB-GPNLMSVF 1 BAND, SB-GPNLMS 3 BANDS), the SB-GPNLMSVF structure requires roughly 2.6 times the operations as the SB-NLMS structure to attain the vastly improved ERLE performance presented in Sections 6.2.2.2 and 6.2.3. It should also be noted that in general the computational advantage of the SB-GPNLMSVF structure will increase compared to its fullband counterpart as the OSR decreases.

The computational cost of the SB-GPNLMSVF and SB-NLMS structures are shown in Figure 6.18 for an increasing number of subbands with the same OSR of 2. The results are based on the same filter bank settings as those outlined in Section 6.2.3 for Figure 6.14. The length of the linear adaptive filters for the $K = 8$ structures were set to $N_{1, sb_i} = 500$ in all subbands and to $N_{1, sb_i} = 250$ in all subbands for the $K = 16$ structures, since equivalent ERLE performance was obtained by the SB-GPNLMSVF structure under these settings. The computational requirements of the SB-GPNLMSVF structure are presented with the GPNLMS algorithm applied to the linear filters in all and three subbands, with corresponding GPNLMS based

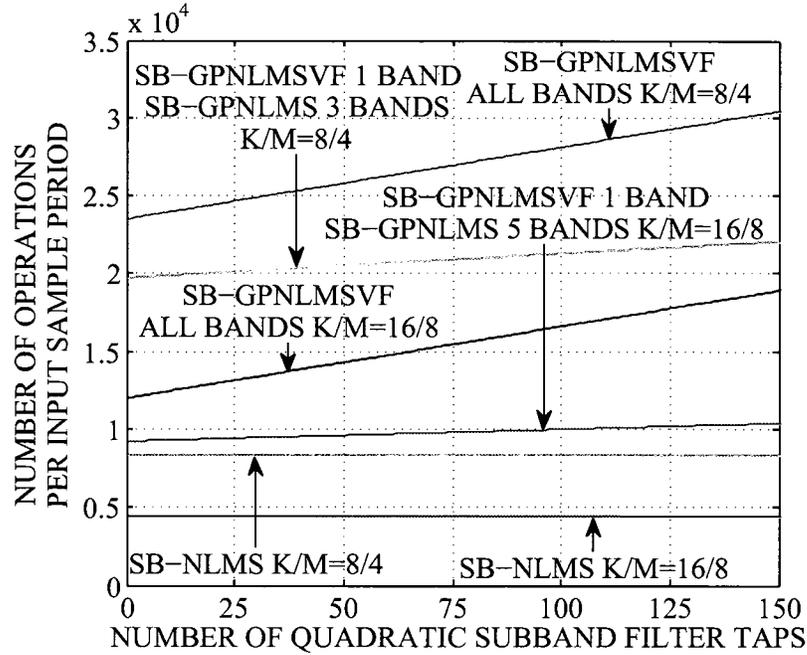


Figure 6.18: Computational complexity for SB-GPNLMSVF and SB-NLMS for an increasing number of subbands for the same OSR.

quadratic filters in all, and only one subband for the $K = 8$ structure. For the $K = 16$ SB-GPNLMSVF structure the complexity results are presented with the GPNLMS algorithm applied to the linear filters in all and five subbands, with corresponding GPNLMS based quadratic filters in all, and only one subband. Also, in determining the computational cost of the SB-GPNLMSVF structure, the complexity of the NLMS algorithm was used for the linear filters that did not implement the GPNLMS algorithm. Furthermore, when the GPNLMS algorithm was applied to a quadratic filter in only one subband, the cost of applying the NLMS algorithm to a quadratic filter in a second subband was also included to represent quadratic Volterra filtering in only two subbands. With the GPNLMS algorithm applied to the linear filters in five subbands and to the quadratic filter in one subband for the $K = 16$ SB-GPNLMSVF structure (SB-GPNLMSVF 1 BAND, SB-GPNLMS 5 BANDS, $K/M=16/8$) with a length of $N_{2, sb_i} = 30$, a reduction in computational complexity of over 50% is achieved

compared to the $K = 8$ SB-GPNLMSVF structure with the GPNLMS algorithm applied to the linear filters in three subbands and to the quadratic filter in one subband (SB-GPNLMSVF 1 BAND, SB-GPNLMS 3 BANDS, $K/M=8/4$) with $N_{2,sb_i} = 120$. Moreover, this $K = 16$ SB-GPNLMSVF structure provides roughly the same ERLE performance as the much more computationally expensive $K = 8$ SB-GPNLMSVF structure as shown in Figure 6.14. Also, the $K = 16$ SB-GPNLMSVF structure with the GPNLMS algorithm applied in the same manner (SB-GPNLMSVF 1 BAND, SB-GPNLMS 5 BANDS, $K/M=16/8$) and $N_{2,sb_i} = 30$, requires only a slightly higher computational cost than the linear only $K = 8$ SB-NLMS structure and approximately twice the computational cost compared to the $K = 16$ SB-NLMS structure, while attaining the significantly improved ERLE performance during changing echo path conditions shown in Figure 6.14.

6.3 Summary

This chapter investigated the application of the GGNLMS algorithm to linear and quadratic adaptive Volterra filters in an oversampled subband AEC system. This allowed for efficient modeling of linear and nonlinear echo path components with simultaneous fast tracking of these components during changing echo path conditions. A detailed discussion and computational complexity analysis of the SB-GGNLMSVF structure was given in Section 6.1. AEC performance results for the SB-GGNLMSVF structure along with its fullband and linear only counterparts were outlined in Sections 6.2.2 and 6.2.3, under changing echo path conditions based on experimentally measured LEMS data. The results demonstrated the ability of the SB-GGNLMSVF structure to achieve equal or improved AEC performance compared to its fullband equivalent. As well, significantly better performance compared to the linear only AEC structures was observed for the SB-GGNLMSVF structure during periods of high loudspeaker harmonic distortion and changing echo path

conditions. Furthermore, the SB-GGPNLMSVF structure was able to exploit the frequency domain nature of the echo path changes and loudspeaker nonlinearity, by applying the GGPNLMS algorithm to the linear and quadratic filters in only the correspondingly affected subbands. This allowed the SB-GGPNLMSVF structure to realize the same level of improved AEC performance at a much reduced computational cost compared to its equivalent fullband structure, as shown in Section 6.2.4. The SB-GGPNLMSVF structure was also able to take advantage of the time domain nature of the loudspeaker nonlinearity and its variations by modeling only the important high amplitude quadratic Volterra kernel coefficients, which allowed for further computational savings with equivalent or similar AEC performance compared to its fullband counterpart.

Table 6.4 summarizes the main results of this chapter. The maximum improvement in ERLE for each AEC structure compared to the SB-NLMS structure is given based on the measured changing linear and nonlinear echo path conditions presented in Section 6.2.3. The corresponding computational complexity results are also presented in Table 6.4 and are applicable to the AEC structures in all sections under the conditions outlined in Section 6.2.4. The results for the subband structures are based on $K = 8$ $M = 4$ filter banks with $L_{pr} = 64$. The complexity of the SB-GGPNLMSVF structure with $N_{2, sb_{1\&3}} = 120$ is based on GGPNLMS adapted linear filters in only three subbands with a NLMS linear filter in one subband as discussed in Section 6.2.4. Also, it should be noted that ERLE improvements of 6 dB or more attained by the proposed SB-GGPNLMSVF structure translates to a reduction by a factor of two or more in the output error signal. Furthermore, selected results from this chapter have been archived online [105] to demonstrate the perceived improvement in echo reduction that can be realized by the proposed SB-GGPNLMSVF structure.

AEC Structure	ERLE Improvement Compared to SB-NLMS (dB)	Complexity (Operations per input sample period)
SB-NLMS $N_{1, sb_{1-4}} = 500$	–	8305
FB-GPNLMSVF $N_1 = 2000 N_2 = 50$	15	45889
SB-GPNLMSVF $N_{1, sb_{1-4}} = 500 N_{2, sb_{1-4}} = 120$	18	28978
SB-GPNLMSVF $N_{1, sb_{1-4}} = 500 N_{2, sb_{1\&3}} = 120$	16	21564

Table 6.4: Summary of results for AEC in nonstationary and nonlinear environments.

Chapter 7

Conclusions and Future Research

7.1 Summary of Research

For hands-free telecommunication devices it is paramount that its EC operate effectively under adverse conditions in the acoustic environment to ensure a high quality conversation between parties. As well, it is necessary that the EC be implemented within a certain computational complexity budget as it is typically only one of many systems that require processing on the hands-free device. A low complexity EC is also highly desirable for portable battery operated hands-free devices in order to minimize its computational demand on the device's processor, thus minimizing its power consumption. These EC requirements are not generally satisfied simultaneously by a typically EC, thus exemplifying the need for ECs capable of operating effectively under adverse conditions with low complexity. This thesis has addressed these EC requirements through the investigation and development of computationally efficient signal processing structures and adaptive algorithms capable of operating under harsh conditions, with specific application to acoustic echo cancellation in nonstationary and nonlinear environments. The following relevant research questions were identified in Chapter 1:

1. Based on experimentally measured data from LEMSs under changing conditions, what is the corresponding impact on the linear and nonlinear portions of the resulting echo path model in both the time and frequency domains?
2. Given the time domain nature of echo path changes can the computational complexity, in terms of the total number of required elementary mathematical operations per input sample period, of ECs based on proportionate step size adaptive algorithms be reduced while maintaining fast tracking performance?
3. Is it possible to incorporate proportionate step size adaptive algorithms into a subband EC structure to reduce complexity and improve tracking performance, compared to an equivalent fullband structure? If so, can the complexity of this structure be further reduced by exploiting the frequency domain nature of echo path changes, without sacrificing tracking performance?
4. Can a nonlinear EC based on adaptive Volterra filters be implemented with a subband structure to decrease complexity and enhance echo cancellation performance, compared to a fullband Volterra EC? If so, can this structure realize further computational savings while maintaining performance, by considering the time and frequency domain characteristics of loudspeaker distortion?
5. Is it possible to incorporate proportionate step size adaptive algorithms into a subband Volterra filter EC structure, and achieve improved tracking performance with reduced complexity compared to an equivalent fullband structure?

In Chapter 2 a review of echo cancellation structures and associated adaptive algorithms revealed that these research questions have not been thoroughly addressed in the literature. Specifically, it was found that:

- Limited literature exists for incorporating proportionate step size adaptive algorithms into subband EC structures.

- Many reduced complexity nonlinear EC structures have been proposed based on adaptive Volterra filters. However, little work in the literature exists on computationally efficient subband EC structures based on adaptive Volterra filters.
- There is very little literature on proportionate step size adaptive Volterra filtering algorithms. Furthermore, there is no literature on incorporating proportionate step size adaptive Volterra filters into subband EC structures.

Chapter 3 examined the time and frequency domain nature of experimentally measured linear and nonlinear echo path components under various conditions for two hands-free systems. The hands-free system data was recorded in a large office room environment at Carleton University. Also, a harmonic distortion analysis of the two commercial phone sets was performed for several increasing input signal levels. Based on these experimental hands-free system measurements a number of conclusions can be made:

- In response to changes in the acoustic environment, the corresponding changes in the linear and nonlinear echo path components are limited to specific time and frequency regions. In particular, the linear echo path components corresponding to early reflections from nearby surfaces were demonstrated to change significantly in response to a small object introduced in the vicinity of the hands-free device, while the direct path and late reflection echo path components remained relatively unchanged.
- In the frequency domain it was demonstrated that the linear echo path components were relatively unchanged at low frequencies and varied considerably at high frequencies due to the small obstructing object.
- As the nature of loudspeaker distortion was shown to be confined mainly to specific frequency regions, the corresponding changes in the nonlinear echo path

components were also limited to these frequency regions in response to an increase in loudspeaker input signal level.

- The majority of harmonic distortion at high input signal levels was contained in the second harmonic and was limited to specific frequency regions for both phone sets. As the input signal level decreased so did the amount of harmonic distortion, however significant second harmonic distortion still occurred for both phone sets.

Chapter 4 investigated the incorporation of proportionate step size adaptive algorithms into a subband structure for echo cancellation in changing acoustic environments. An oversampled subband EC structure for fast convergence and tracking of echo path changes based on the GGPAPA outlined in Chapter 2 was proposed. The improved convergence and tracking ability of this structure compared to its fullband equivalent was shown, along with its reduced computational complexity. A method was also proposed to reduce the computational complexity of the GGPAPA by exploiting the time domain nature of the changes that occur in the linear portion of the echo path, due to a small obstruction introduced in close proximity to the hands-free device. In addition, a number of conclusions can be made:

- The SB-GGPAP EC structure allows the frequency domain characteristics of linear echo path fluctuations to be exploited by applying the GGPAPA in only the correspondingly affected subbands and the APA in the unaffected subbands, resulting in increased complexity savings while maintaining fast convergence and tracking performance.
- Simulation results showed that the computational complexity of the SB-GGPAP structure can be significantly reduced while maintaining fast tracking performance by adapting only the first few subband filter taps, corresponding to the highly variable section of the subband LEMS impulse response, with the GGPAPA and the rest of the filter taps with the APA.

- The GGPAPA allows for a family of fast tracking subband EC structures to be implemented by modifying only a few parameters, resulting in a trade-off between tracking performance and overall structure complexity.

Chapter 5 considered the application of adaptive Volterra filters into a subband structure for echo cancellation in the presence of nonlinear loudspeaker distortion. An oversampled subband EC based on second order NLMS adaptive Volterra filters was presented, with simulation results that verified its improved ERLE performance and reduced complexity compared to its fullband counterpart. Furthermore, the following conclusions can be made:

- A subband version of a nonlinear EC based on second order adaptive Volterra filters allows for an inherent improvement in convergence performance and reduction in computational complexity compared to an equivalent fullband structure.
- The subband adaptive second order Volterra filter EC structure allows for further complexity reduction while maintaining ERLE performance, compared to its fullband counterpart, by applying second order Volterra filtering in only the necessary subbands based on the harmonic distortion of the hands-free device.
- The subband adaptive second order Volterra filter EC structure can achieve improved ERLE performance compared to a linear subband EC structure, while requiring only a very small increase in computational complexity, by implementing second order Volterra filtering in only the necessary subbands and modeling only the significant nonlinear echo path components.

Chapter 6 focused on the application of proportionate step size algorithms to second order Volterra filters in a subband structure, for fast convergence and tracking of both linear and nonlinear echo path components simultaneously. An oversampled subband EC structure with second order Volterra filters based on the GGPAPA was

proposed. Simulation results demonstrated the proposed structure's superior tracking ability with reduced complexity compared to its equivalent fullband structure. Moreover, the following conclusions can be made:

- Applying the GGPAPA to second order Volterra filters in a subband structure results in an improvement in convergence and tracking performance, along with a reduction in computational complexity compared to an equivalent fullband structure.
- Simulation results revealed that subband EC structure with second order Volterra filters based on the GGPAPA can realize increased computational savings, while maintaining tracking performance during changing echo path conditions, by applying the GGPAPA in only the correspondingly affected linear and nonlinear subbands.
- A family of fast tracking linear and nonlinear subband EC structures can be implemented by altering only a few parameters of the GGPAPA, resulting in a trade-off between ERLE performance and overall structure complexity.

7.2 Summary of Contributions

The focus of this thesis was the development of computationally efficient subband EC structures capable of operating in nonstationary and nonlinear acoustic environments. The subsequent investigations, analyzes, and simulation results provided several contributions to the field of acoustic echo cancellation. These contributions are summarized briefly as follows:

1. An experimental analysis of the characteristics of linear and nonlinear echo path components revealed that changes in these components are limited to certain time and frequency regions. The trends from this analysis led to complexity savings in the proposed subband EC structures of this thesis.

2. A harmonic distortion analysis of two commercial phone sets demonstrated that the majority of the total harmonic distortion was contained in the second harmonic and was limited to specific frequency regions for both devices. These results also led to complexity savings in the proposed subband EC structures of this thesis.
3. An oversampled subband EC structure based on the GGPAPA was proposed for fast tracking of linear echo path changes. This structure achieves improved convergence and tracking performance at a reduced complexity compared to its equivalent fullband structure, by considering the time and frequency domain characteristics of linear echo path variations.
4. An oversampled subband EC structure based on second order NLMS adaptive Volterra filters was proposed for acoustic echo cancellation in the presence of nonlinear loudspeaker distortion. This structure attains improved ERLE performance at a significantly reduced computational expense compared to its fullband counterpart, by exploiting the time and frequency domain characteristics of loudspeaker distortion.
5. An oversampled subband EC structure with second order Volterra filters based on the GGPAPA was proposed for fast tracking and modeling of linear and nonlinear echo path changes simultaneously. This structure achieves improved convergence and tracking performance at a reduced complexity compared to its equivalent fullband structure.
6. Publication of the refereed conference and journal papers in [27], [28], and [29] respectively, which report on some of the research results outlined above.

7.3 Suggestions for Future Research

Although the proposed signal processing structures and adaptive algorithms proposed in this thesis perform well in adverse environments and are computationally efficient they have several limitations which warrant further research, specifically with their application to the acoustic echo cancellation problem in nonstationary and nonlinear environments. These are summarized as follows:

1. Application of the APA to the subband second order Volterra filter EC structures presented in Chapters 5 and 6. Implementing these EC structures with AP based linear and quadratic subband filters would result in further improved convergence and tracking performance compared to the current NLMS based filters. The inherent increase in complexity of the APA can be partially offset in the subband second order Volterra filter structures, compared to the equivalent fullband AP Volterra structures, as they can exploit the time and frequency domain nature of the linear and quadratic echo path components. Methods for AP based second order adaptive Volterra filters that could be applied to the proposed subband EC structures are outlined in [107] and [108].
2. Extension of the subband adaptive second order Volterra filter EC structures to higher orders. As shown in Chapter 3 the harmonic distortion of the two hands-free devices investigated were mainly confined to the second harmonic. However, under high input signal level conditions there was still some fairly significant third and higher harmonic distortion in certain frequency regions. Since the proposed subband adaptive Volterra filter EC structures of Chapters 5 and 6 are limited to the second order, incorporating third and higher order adaptive Volterra filters would allow for modeling of these higher order echo path components and thus improved EC performance under these high input signal level conditions. A computationally efficient time domain method for adaptive third order Volterra filtering applied to acoustic echo cancellation is outlined

in [20]. This method could be incorporated with the proposed subband adaptive Volterra filter EC structures to allow for improved nonlinear echo suppression with a minimal increase in complexity.

3. Adaptation control for proposed subband EC structures. Currently the EC structures proposed require *a priori* knowledge of the nature of echo path changes and distortion characteristics of the hands-free device, to determine which subset of all subbands the fast tracking and nonlinear adaptive algorithms should be applied to help reduce computational complexity. Thus, it would be beneficial to control adaptation in the proposed subband EC structures automatically for reduced computational complexity. One approach to controlling the adaptation could be based on the residual error within each subband. When the residual error is small enough the lower complexity APA could be used to adapt the subband filter or adaptation could be stopped. When the residual error becomes higher the fast tracking GGPAPA could be used to adapt the filters within the corresponding subbands. A potentially applicable method for determining if the residual error is small enough is given in [109].
4. Implementation of delayless subband structures. The proposed subband EC structures incur a delay in the signal path due to the group delay of the analysis and synthesis filter banks. For hands-free systems with strict delay budgets this additional delay can be detrimental to the quality of the communication. To avoid this potential limitation in the proposed subband EC structures, it would be beneficial to implement them in a delayless manner. Applicable approaches for delayless subband adaptive filtering that could be used are discussed in [55] and [57].
5. Incorporation of partial updating schemes to further reduce complexity in the proposed subband EC structures. Partial updating methods involve updating only a subset of the total filter taps at each iteration thereby reducing the

overall computational complexity. Thus, it would be interesting to incorporate partial updating in the proposed EC structures, especially as only a subset of the adaptive filter taps are significantly impacted in response to changes in the echo path. Partial updating methods that could be applied to the proposed subband EC structures are discussed in [110] and [111].

6. Incorporation of the proposed subband EC structures into a complete echo cancellation system. Although the proposed subband EC structures perform well on their own in changing and nonlinear acoustic environments, it is expected that they would also perform well in a complete echo cancellation system. Thus, it would be useful to verify their performance in a complete echo cancellation system consisting of a doubletalk detector, post-filter, VAD, along with the EC itself. Potentially applicable methods for doubletalk detection are given in [7] and [112], while approaches for post-filtering and voice activity detection are outlined in [30] and [31].
7. Extension of the proposed subband EC structures to multi-microphone systems. Multi-microphone hands-free systems involve combining beamforming and echo cancellation techniques to improve the quality of communication over traditional single microphone systems. Application of the proposed subband EC structures to multi-microphone hands-free systems would result in ECs that are capable of removing nonlinear echoes, and allow for fast tracking of variations in both the echo path and beamformer while requiring low computational complexity. A combined time domain microphone array beamforming and echo cancellation system is described in [113]. This structure could provide the foundation for applying the proposed EC structures into a multi-microphone hands-free system.

References

- [1] E. Hänsler and G. Schmidt, *Acoustic Echo and Noise Control: A Practical Approach*. Hoboken, NJ: Wiley, 2004.
- [2] M. M. Sondhi and D. A. Berkley, “Silencing echoes on the telephone network,” *Proceedings of the IEEE*, vol. 68, no. 8, pp. 948–963, Aug. 1980.
- [3] D. L. Duttweiler, “A twelve-channel digital echo canceler,” *IEEE Trans. Commun.*, vol. 26, no. 5, pp. 647–653, May 1978.
- [4] D. Messerschmitt, “Echo cancellation in speech and data transmission,” *IEEE J. Sel. Areas Commun.*, vol. 2, no. 2, pp. 283–297, Mar. 1984.
- [5] A. N. Birkett and R. A. Goubran, “Limitations of handsfree acoustic echo cancellers due to nonlinear loudspeaker distortion and enclosure vibration effects,” in *IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics*, Oct. 1995, pp. 103–106.
- [6] E. Hänsler, “The hands-free telephone problem – an annotated bibliography,” *Signal Processing*, vol. 27, no. 3, pp. 259–271, Jun. 1992.
- [7] T. Gänsler, S. L. Gay, M. M. Sondhi, and J. Benesty, “Double-talk robust fast converging algorithms for network echo cancellation,” *IEEE Trans. Speech Audio Process.*, vol. 8, no. 6, pp. 656–663, Nov. 2000.

- [8] M. E. Knappe and R. Goubran, "Steady-state performance limitations of full-band acoustic echo cancellers," in *Proc. IEEE ICASSP*, vol. 2, Apr. 1994, pp. 73–76.
- [9] J. D. Gordy and R. A. Goubran, "On the perceptual performance limitations of echo cancellers in wideband telephony," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, no. 1, pp. 33–42, Jan. 2006.
- [10] *Digital network echo cancellers*, International Telecommunication Union Std. ITU-T G.168, 2007.
- [11] D. L. Duttweiler, "Proportionate normalized least-mean-squares adaptation in echo cancelers," *IEEE Trans. Speech Audio Process.*, vol. 8, no. 5, pp. 508–518, Sep. 2000.
- [12] O. Hoshuyama, R. A. Goubran, and A. Sugiyama, "A generalized proportionate variable step-size algorithm for fast changing acoustic environments," in *Proc. IEEE ICASSP*, vol. 4, May 2004, pp. 161–164.
- [13] S. L. Gay, "An efficient, fast converging adaptive filter for network echo cancellation," in *Proc. Asilomar Conference on Signals, Systems and Computers*, vol. 1, Nov. 1998, pp. 394–398.
- [14] J. Benesty and S. L. Gay, "An improved PNLMS algorithm," in *Proc. IEEE ICASSP*, vol. 2, May 2002, pp. 1881–1884.
- [15] J. J. Shynk, "Frequency-domain and multirate adaptive filtering," *IEEE Signal Process. Mag.*, vol. 9, no. 1, pp. 14–37, Jan. 1992.
- [16] F. Kuech and W. Kellermann, "Partitioned block frequency-domain adaptive second-order Volterra filter," *IEEE Trans. Signal Process.*, vol. 53, no. 2, pp. 564–575, Feb. 2005.

- [17] A. J. M. Kaizer, "Modeling of the nonlinear response of an electrodynamic loudspeaker by a Volterra series expansion," *J. Audio Eng. Soc.*, vol. 35, pp. 421–433, Jun. 1987.
- [18] W. Klippel, "Dynamic measurement and interpretation of the nonlinear parameters of electrodynamic loudspeakers," *J. Audio Eng. Soc.*, vol. 38, no. 12, pp. 944–955, Dec. 1990.
- [19] T. Koh and E. Powers, "Second-order Volterra filtering and its application to nonlinear system identification," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 33, no. 6, pp. 1445–1455, Dec. 1985.
- [20] A. Guérin, G. Faucon, and R. L. Bouquin-Jeannès, "Nonlinear acoustic echo cancellation based on Volterra filters," *IEEE Trans. Speech Audio Process.*, vol. 11, no. 6, pp. 672–683, Nov. 2003.
- [21] J. Chen and J. Vandewalle, "Study of adaptive nonlinear echo canceller with Volterra expansion," in *Proc. IEEE ICASSP*, vol. 2, May 1989, pp. 1376–1379.
- [22] G. Sicuranza, A. Bucconi, and P. Mitri, "Adaptive echo cancellation with nonlinear digital filters," in *Proc. IEEE ICASSP*, vol. 9, Mar. 1984, pp. 130–133.
- [23] A. Stenger, L. Trautmann, and R. Rabenstein, "Nonlinear acoustic echo cancellation with 2nd order adaptive Volterra filters," in *Proc. IEEE ICASSP*, vol. 2, Mar. 1999, pp. 877–880.
- [24] M. J. Reed and M. O. Hawksford, "Efficient implementation of the Volterra filter," in *IEE Proc. - Vis., Image Signal Processing*, vol. 147, Apr. 2002, pp. 109–114.
- [25] S. Im and E. J. Powers, "A fast method of discrete third-order Volterra filtering," *IEEE Trans. Signal Process.*, vol. 44, no. 9, pp. 2195–2208, Sep. 1996.

- [26] D. Zhou, Y. Wang, V. DeBrunner, and L. DeBrunner, "Sub-band implementation of adaptive nonlinear filter for adaptive nonlinear echo cancellation," *Journal of Multimedia*, vol. 2, no. 2, pp. 77–82, Apr. 2007.
- [27] T. G. Burton, R. A. Goubran, and F. Beaucoup, "Nonlinear system identification using a subband adaptive Volterra filter," *IEEE Trans. Instrum. Meas.*, vol. 58, no. 5, pp. 1389–1397, May 2009.
- [28] T. Burton and R. Goubran, "An analysis of loudspeaker distortion in the context of acoustic echo cancellation," in *Proc. Acoustics Week in Canada, Canadian Acoustics*, vol. 37, no. 3, Sep. 2009, pp. 60–61.
- [29] T. Burton, F. Beaucoup, and R. Goubran, "Nonlinear system identification using a subband adaptive Volterra filter," in *Proc. IEEE I²MTC*, May 2008, pp. 939–944.
- [30] S. J. Park, C. Lee, and D. H. Youn, "A residual echo cancellation scheme for hands-free telephony," *IEEE Signal Process. Lett.*, vol. 9, no. 12, pp. 397–399, Dec. 2002.
- [31] E. Nemer, R. Goubran, and S. Mahmoud, "Robust voice activity detection using higher-order statistics in the lpc residual domain," *IEEE Trans. Speech Audio Process.*, vol. 9, no. 3, pp. 217–231, Mar. 2001.
- [32] J. D. Gordy and R. A. Goubran, "Statistical analysis of doubletalk detection for calibration and performance evaluation," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 3, pp. 1035–1043, Mar. 2007.
- [33] S. Haykin, *Adaptive Filter Theory*, 4th ed. Upper Saddle River, NJ: Prentice-Hall, 2002.

- [34] F. Küch, “Adaptive polynomial filters and their application to nonlinear acoustic echo cancellation,” Ph.D. dissertation, Friedrich-Alexander University Erlangen-Nuremberg, Nov. 2005.
- [35] D. R. Morgan, “Slow asymptotic convergence of LMS acoustic echo cancelers,” *IEEE Trans. Speech Audio Process.*, vol. 3, no. 2, pp. 126–136, Mar. 1995.
- [36] C. Breining, P. Dreiscitel, E. Hansler, A. Mader, B. Nitschand, H. Puder, T. Schertler, G. Schmidt, and J. Tilp, “Acoustic echo control: An application of very-high-order adaptive filters,” *IEEE Signal Process. Mag.*, vol. 16, no. 4, pp. 42–69, Jul. 1999.
- [37] M. Montazeri and P. Duhamel, “A set of algorithms linking NLMS and block RLS algorithms,” *IEEE Trans. Signal Process.*, vol. 43, no. 2, pp. 444–453, Feb. 1995.
- [38] S. L. Gay, “Dynamically regularized fast RLS with application to echo cancellation,” in *Proc. IEEE ICASSP*, vol. 2, May 1996, pp. 957–960.
- [39] F. Capman, J. Boudy, and P. Lockwood, “Acoustic echo cancellation using a fast QR-RLS algorithm and multirate schemes,” in *Proc. IEEE ICASSP*, vol. 2, May 1995, pp. 969–972.
- [40] M. G. Siqueira and A. Alwan, “New adaptive-filtering techniques applied to speech echo cancellation,” in *Proc. IEEE ICASSP*, vol. 2, Apr. 1994, pp. 265–268.
- [41] T. Petillon, A. Gilloire, and S. Theodoridis, “Complexity reduction in fast RLS transversal adaptive filters with application to acoustic echo cancellation,” in *Proc. IEEE ICASSP*, vol. 4, Mar. 1992, pp. 37–40.

- [42] K. Ozeki and T. Umeda, "An adaptive filtering algorithm using an orthogonal projection to an affine subspace and its properties," *Electronics and Communications in Japan*, vol. 67-A, no. 5, pp. 19–27, 1984.
- [43] P. S. R. Diniz, *Adaptive Filtering: Algorithms and Practical Implementation*, 3rd ed. New York, NY: Springer, 2008.
- [44] H. Rey, L. R. Vega, S. Tressens, and J. Benesty, "Variable explicit regularization in affine projection algorithm: Robustness issues and optimal choice," *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 2096–2109, May 2007.
- [45] S. L. Gay and S. Tavathia, "The fast affine projection algorithm," in *Proc. IEEE ICASSP*, vol. 5, May 1995, pp. 3023–3026.
- [46] F. Albu and C. Kotropoulos, "Modified Gauss-Seidel affine projection algorithm for acoustic echo cancellation," in *Proc. IEEE ICASSP*, vol. 3, Mar. 2005, pp. 121–124.
- [47] H. Ding, "Fast affine projection adaptation algorithms featuring stable symmetric positive-definite linear system solvers," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, Oct. 2005, pp. 166–169.
- [48] P. P. Vaidyanathan, *Multirate Systems And Filter Banks*. Englewood Cliffs, NJ: Prentice-Hall, 1992.
- [49] W. Kellermann, "Analysis and design of multirate systems for cancellation of acoustical echoes," in *Proc. IEEE ICASSP*, vol. 5, Apr. 1988, pp. 2570–2573.
- [50] K. Eneman, "Subband and frequency-domain adaptive filtering techniques for speech enhancement in hands-free communication," Ph.D. dissertation, Katholieke Universiteit Leuven, Mar. 2002.

- [51] H. R. Abutalebi, H. Sheikhzadeh, R. L. Brennan, and G. H. Freeman, "Affine projection algorithm for oversampled subband adaptive filters," in *Proc. IEEE ICASSP*, vol. 6, Apr. 2003, pp. 209–212.
- [52] J. D. Gordy and R. A. Goubran, "Fast system identification using affine projection and a critically sampled subband adaptive filter," *IEEE Trans. Instrum. Meas.*, vol. 55, no. 4, pp. 1242–1249, Aug. 2006.
- [53] V. Parsa and D. G. Jamieson, "Adaptive modelling of digital hearing aids using a subband affine projection algorithm," in *Proc. IEEE ICASSP*, vol. 2, May 2002, pp. 1937–1940.
- [54] M. Wirtzfeld and V. Parsa, "On subband adaptive modeling of compression hearing aids," in *Proc. IEEE ICASSP*, vol. 3, Mar. 2005, pp. 45–48.
- [55] D. R. Morgan and J. C. Thi, "A delayless subband adaptive filter architecture," *IEEE Trans. Signal Process.*, vol. 43, no. 8, pp. 1819–1830, Aug. 1995.
- [56] V. DeBrunner, L. S. DeBrunner, and L. Wang, "Sub-band adaptive filtering with delay compensation for active control," *IEEE Trans. Signal Process.*, vol. 52, no. 10, pp. 2932–2941, Oct. 2004.
- [57] M. R. Petraglia, R. G. Alves, and M. N. S. Swamy, "A new open loop delayless subband adaptive filter structure," in *Proc. IEEE ICASSP*, vol. 2, May 2002, pp. 1345–1348.
- [58] S. Weiss, A. Stenger, R. W. Stewart, and R. Rabenstein, "Steady-state performance limitations of subband adaptive filters," *IEEE Trans. Signal Process.*, vol. 49, no. 9, pp. 1982–1991, Sep. 2001.
- [59] A. Gilloire and M. Vetterli, "Adaptive filtering in subbands with critical sampling: analysis, experiments, and application to acoustic echo cancellation," *IEEE Trans. Signal Process.*, vol. 40, no. 8, pp. 1862–1875, Aug. 1992.

- [60] M. R. Petraglia, R. G. Alves, and P. S. R. Diniz, “New structures for adaptive filtering in subbands with critical sampling,” *IEEE Trans. Signal Process.*, vol. 48, no. 12, pp. 3316–3327, Dec. 2000.
- [61] M. Harteneck, S. Weiss, and R. W. Stewart, “Design of near perfect reconstruction oversampled filter banks for subband adaptive filters,” *IEEE Trans. Circuits Syst. II*, vol. 46, no. 8, pp. 1081–1085, Aug. 1999.
- [62] K. Eneman and M. Moonen, “DFT modulated filter bank design for oversampled subband systems,” *Signal Processing*, vol. 81, no. 9, pp. 1947–1973, Sep. 2001.
- [63] S. Weiss, “Analysis and fast implementation of oversampled modulated filter banks,” in *Mathematics in Signal Processing V*, J. G. McWhirter and I. K. Proudler, Eds. New York: Oxford University Press, 2002, pp. 263–274.
- [64] J. D. Z. Chen, H. Bes, J. Vandewalle, I. Evers, and P. Janssens, “A zero-delay fft-based subband acoustic echo canceller for teleconferencing and hands-free telephone systems,” *IEEE Trans. Circuits Syst. II*, vol. 43, no. 10, pp. 713–717, Oct. 1996.
- [65] M. Dentino, J. McCool, and B. Widrow, “Adaptive filtering in the frequency domain,” *Proc. IEEE*, vol. 66, no. 12, pp. 1658–1659, Dec. 1978.
- [66] E. Ferrara, “Fast implementations of LMS adaptive filters,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 28, no. 4, pp. 474–475, Aug. 1980.
- [67] M. Asharif, T. Takebayashi, T. Chujo, and K. Murano, “Frequency domain noise canceller: Frequency bin adaptive filtering (FBAF),” in *Proc. IEEE ICASSP*, vol. 11, Apr. 1986, pp. 2219–2222.
- [68] O. A. Amrane, E. Moulines, M. Charbit, and Y. Grenier, “Low-delay frequency domain LMS algorithm,” in *Proc. IEEE ICASSP*, vol. 4, Mar. 1992, pp. 9–12.

- [69] E. Moulines, O. A. Amrane, and Y. Grenier, "The generalized multidelay adaptive filter: structure and convergence analysis," *IEEE Trans. Signal Process.*, vol. 43, no. 1, pp. 14–28, Jan. 1995.
- [70] J.-S. Soo and K. K. Pang, "Multidelay block frequency domain adaptive filter," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 2, pp. 373–376, Feb. 1990.
- [71] J. Lariviere and R. Goubran, "GMDF for noise reduction and echo cancellation," *IEEE Signal Process. Lett.*, vol. 7, no. 8, pp. 230–232, Aug. 2000.
- [72] Y. Bendel, D. Burshtein, O. Shalvi, and E. Weinstein, "Delayless frequency domain acoustic echo cancellation," *IEEE Trans. Speech Audio Process.*, vol. 9, no. 5, pp. 589–597, Jul. 2001.
- [73] J. Benesty and D. R. Morgan, "Frequency-domain adaptive filtering revisited, generalization to the multi-channel case, and application to acoustic echo cancellation," in *Proc. IEEE ICASSP*, vol. 2, Jun. 2000, pp. 789–792.
- [74] J. M. P. Borrallo and M. G. Otero, "On the implementation of a partitioned block frequency domain adaptive filter (PBFDAF) for long acoustic echo cancellation," *Signal Processing*, vol. 27, no. 3, pp. 301–315, Jun. 1992.
- [75] J. G. Proakis and D. G. Manolakis, *Digital Signal Processing: Principles, Algorithms and Applications*, 3rd ed. Upper Saddle River, NJ: Prentice-Hall, 1996.
- [76] O. Agazzi, D. Messerschmitt, and D. Hodges, "Nonlinear echo cancellation of data signals," *IEEE Trans. Commun.*, vol. 30, no. 11, pp. 2421–2433, Nov. 1982.

- [77] A. Stenger and W. Kellermann, "Adaptation of a memoryless preprocessor for nonlinear acoustic echo cancelling," *Signal Processing*, vol. 80, no. 9, pp. 1747–1760, Sep. 2000.
- [78] W. Klippel, "Tutorial: Loudspeaker nonlinearities—Causes, parameters, symptoms," *J. Audio Eng. Soc.*, vol. 54, no. 10, pp. 907–939, Oct. 2006.
- [79] F. X. Y. Gao and W. M. Snelgrove, "Adaptive linearization of a loudspeaker," in *Proc. IEEE ICASSP*, vol. 5, Apr. 1991, pp. 3589–3592.
- [80] A. N. Birkett and R. Goubran, "Nonlinear loudspeaker compensation for hands free acoustic echo cancellation," *IEE Electronics Letters*, vol. 32, no. 12, pp. 1063–1064, Jun. 1996.
- [81] A. N. Birkett and R. A. Goubran, "Acoustic echo cancellation using NLMS-neural network structures," in *Proc. IEEE ICASSP*, vol. 5, May 1995, pp. 3035–3038.
- [82] F. Kuech, A. Mitnacht, and W. Kellermann, "Nonlinear acoustic echo cancellation using adaptive orthogonalized power filters," in *Proc. IEEE ICASSP*, vol. 3, Mar. 2005, pp. 105–108.
- [83] F. Kuech and W. Kellermann, "Nonlinear residual echo suppression using a power filter model of the acoustic echo path," in *Proc. IEEE ICASSP*, vol. 1, Apr. 2007, pp. 73–76.
- [84] V. J. Mathews and G. L. Sicuranza, *Polynomial Signal Processing*. New York, NY: Wiley, 2000.
- [85] S. Kalluri and G. R. Arce, "A general class of nonlinear normalized adaptive filtering algorithms," *IEEE Trans. Signal Process.*, vol. 47, no. 8, pp. 2262–2272, Aug. 1999.

- [86] F. Kuech and W. Kellermann, "Nonlinear line echo cancellation using a simplified second order Volterra filter," in *Proc. IEEE ICASSP*, vol. 2, May 2002, pp. 1117–1120.
- [87] D. W. Griffith, Jr. and G. R. Arce, "Partially decoupled Volterra filters: formulation and LMS adaptation," *IEEE Trans. Signal Process.*, vol. 45, no. 6, pp. 1485–1494, Jun. 1997.
- [88] F. Kuech and W. Kellermann, "Proportionate NLMS algorithm for second-order Volterra filters and its application to nonlinear echo cancellation," in *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Sep. 2003, pp. 75–78.
- [89] A. Fermo, A. Carini, and G. L. Sicuranza, "Simplified Volterra filters for acoustic echo cancellation in GSM receivers," in *Proc. of X European Signal Processing Conference*, Tampere, Finland, Sep. 2000.
- [90] H. Furuhashi, Y. Kajikawa, and Y. Nomura, "Realization of nonlinear acoustic echo cancellation by subband parallel cascade Volterra filter," in *Int. Symp. on Intelligent Signal Processing and Communication Systems (ISPACS)*, Dec. 2006, pp. 837–840.
- [91] R. Harris, D. Chabries, and F. Bishop, "A variable step (VS) adaptive filter algorithm," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 34, no. 2, pp. 309–316, Apr. 1986.
- [92] T. J. Shan and T. Kailath, "Adaptive algorithms with an automatic gain control feature," *IEEE Trans. Circuits Syst.*, vol. 35, no. 1, pp. 122–127, Jan. 1988.
- [93] R. H. Kwong and E. W. Johnston, "A variable step size LMS algorithm," *IEEE Trans. Signal Process.*, vol. 40, no. 7, pp. 1633–1642, Jul. 1992.

- [94] J. B. Evans, P. Xue, and B. Liu, "Analysis and implementation of variable step size adaptive algorithms," *IEEE Trans. Signal Process.*, vol. 41, no. 8, pp. 2517–2535, Aug. 1993.
- [95] V. J. Mathews and Z. Xie, "A stochastic gradient adaptive filter with gradient adaptive step size," *IEEE Trans. Signal Process.*, vol. 41, no. 6, pp. 2075–2087, Jun. 1993.
- [96] T. Aboulnasr and K. Mayyas, "A robust variable step-size LMS-type algorithm: analysis and simulations," *IEEE Trans. Signal Process.*, vol. 45, no. 3, pp. 631–639, Mar. 1997.
- [97] S. Makino, Y. Kaneda, and N. Koizumi, "Exponentially weighted stepsize NLMS adaptive filter based on the statistics of a room impulse response," *IEEE Trans. Speech Audio Process.*, vol. 1, no. 1, pp. 101–108, Jan. 1993.
- [98] H. Yuan, "Dynamic behavior of acoustic echo cancellation," Master's thesis, Carleton University, Jan. 1994.
- [99] J. Garofolo *et al.*, *DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus CD-ROM*. Gaithersburg, MD: NIST, 1990.
- [100] B. N. M. Laska, R. A. Goubran, and M. Bolic, "Improved proportionate sub-band NLMS for acoustic echo cancellation in changing environments," *IEEE Signal Process. Lett.*, vol. 15, pp. 337–340, 2008.
- [101] M. Kahrs and K. Brandenburg, Eds., *Applications of Digital Signal Processing to Audio and Acoustics*. New York, NY: Kluwer Academic Publishers, 2002.
- [102] B. N. M. Laska, F. Beaucoup, and R. A. Goubran, "Non-wiener phenomenon in oversampled subband adaptive filters," in *Proc. IEEE ICASSP*, vol. 3, Apr. 2007, pp. 1337–1340.

- [103] *Intel 64 and IA-32 Architectures Optimization Reference Manual*, Intel, Dec. 2008.
- [104] A. H. Vaidyanathan, *Adaptive Filters*. Hoboken, NJ: John Wiley and Sons, 2008, ch. 11.
- [105] T. Burton. (2010) Trevor Burton Ph.D. Thesis Audio Files. [Online]. Available: http://www.sce.carleton.ca/dsp_labs/trevor_burton/
- [106] D. Zhou, V. DeBrunner, Y. Zhai, and M. Yeary, "Efficient adaptive nonlinear echo cancellation using sub-band implementation of the adaptive Volterra filter," in *Proc. IEEE ICASSP*, vol. 5, May 2006, pp. 277–280.
- [107] A. Fermo, A. Carini, and G. L. Sicuranza, "Low-complexity nonlinear adaptive filters for acoustic echo cancellation in GSM handset receivers," *European Transactions on Telecommunications*, vol. 14, no. 2, pp. 161–169, Jun. 2003.
- [108] R. Niemisto, T. Makela, and V. Mlyla, "Robust fast affine projection algorithm for nonlinear acoustic echo cancellation," in *Proc. European Signal Processing Conference (EUSIPCO)*, vol. 1, Sep. 2002, pp. 523–526.
- [109] T. Gänsler, J. Benesty, M. M. Sondhi, and S. L. Gay, "Dynamic resource allocation for network echo cancellation," in *Proc. IEEE ICASSP*, vol. 5, May 2001, pp. 3233–3236.
- [110] T. Aboulnasr and K. Mayyas, "Complexity reduction of the NLMS algorithm via selective coefficient update," *IEEE Trans. Signal Process.*, vol. 47, no. 5, pp. 1421–1424, May 1999.
- [111] K. Doğançay and O. Tanrikulu, "Adaptive filtering algorithms with selective partial updates," *IEEE Trans. Circuits Syst. II*, vol. 48, no. 8, pp. 762–769, Aug. 2001.

- [112] J. Gordy and R. A. Goubran, "A subband doubletalk detector for echo cancellation in hands-free environments," in *Proc. IEEE Vehicular Technology Conference*, vol. 3, Sep. 2005, pp. 1397–1401.
- [113] T. Burton and R. Goubran, "A new structure for combining echo cancellation and beamforming in changing acoustical environments," in *Proc. IEEE ICASSP*, vol. 1, Apr. 2007, pp. 77–80.