ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY

TELECOM ParisTech

2017-ENST-0054

EDITE - ED 130

Doctorat ParisTech

T H È S E

pour obtenir le grade de docteur délivré par

TELECOM ParisTech

Spécialité « Signal et Images »

présentée et soutenue publiquement par

Leela Krishna GUDUPUDI

le 09 Novembre 2017

New Insights into Loudspeaker Nonlinearities : Application to Acoustic Echo Cancellation

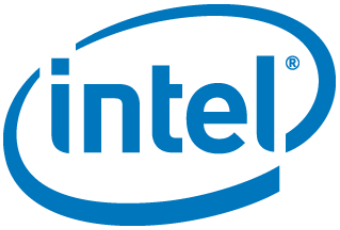
Directeur de thèse : **M. Nicholas EVANS**, Professeur, EURECOM, France

Jury

M. Dirk SLOCK, Professeur, EURECOM, France	President du Jury
Mme. Régine LE BOUQUIN-JEANNES, Professeur, Rennes University, France	Rapporteur
M. Phillip DE LEON, Professeur, New Mexico State University, USA	Rapporteur
M. Christophe BEAUGEANT, Docteur, Groupe Renault (Formerly INTEL), France	Examinateur

TELECOM ParisTech
école de l'Institut Télécom - membre de ParisTech

T H È S E



DISSERTATION

In Partial Fulfilment of the Requirements for the
Degree of Doctor of Philosophy from TELECOM ParisTech University
Specialization: Signal Processing

New Insights into Loudspeaker Nonlinearities: Application to Acoustic Echo Cancellation

Leela Krishna GUDUPUDI

Defense scheduled on 09/11/2017 before a committee composed of:

Thesis Advisor	**Prof. Nicholas EVANS**, EURECOM, France
Reviewers	**Prof. Régine LE BOUQUIN-JEANNES**, Rennes University, France
	Prof. Phillip DE LEON, New Mexico State University, USA
Examiners	**Dr. Christophe BEAUGEANT**, Groupe Renault (f.k.a INTEL), France
	Prof. Dirk SLOCK, EURECOM, France

Excellence is a continuous process and not an accident
— Dr. A.P.J. Abdul Kalam

This dissertation is respectfully dedicated to my Parents and my Teachers...

Acknowledgements

This thesis was carried out at EURECOM and INTEL Corporation from Feb. 2013 to Aug. 2016, located on the French Riviera (Côte d'Azur), one of the most beautiful places in the world. I couldn't have wished for a better place to live and work. This PhD thesis was completely sponsored by INTEL Corporation, Sophia Antipolis, France.

The research work presented in this thesis would not have been possible without my close association with many people from many countries who were always there when I needed them the most. I take this opportunity to acknowledge them and extend my sincere gratitude for helping me make this PhD thesis a possibility.

Foremost, I would like to express my deep sense of gratitude and ardent feeling of admiration to Dr. Christophe BEAUGEANT, my industrial supervisor, for believing in me and giving me the amazing opportunity to pursue PhD and to accept becoming my supervisor. He has guided me and constantly encouraged me to carry on through these years and has contributed to this thesis with a major impact. Thank you as well for cheerfully supporting, motivating, and helping me, often with big doses of patience, throughout my PhD journey.

Similar profound gratitude goes to Prof. Nick EVANS, my academic supervisor, for his enthusiasm, guidance and unrelenting support throughout my PhD process. He has gone beyond his duties to help me fight my worries, anxieties, and have worked to induce absolute confidence in both myself and my work. I appreciate you so much for giving me freedom and allowed me to explore my curiosities while keeping me on track to complete my PhD. Thank you as well for guiding me through the subtleties of scientific writing.

I am also very much grateful to my thesis committee: Prof. Régine LE BOUQUIN-JEANNES, Prof. Phillip DE LEON and Prof. Dirk SLOCK for insightful comments and valuable suggestions. I thoroughly enjoyed all the discussions with Prof. Phillip DE LEON, which helped me widen my research from various perspectives. Many thanks for sharing the knowledge, constant encouragement and being so accommodating.

A major share of my heart-felt gratitude I owe to my mentor and my friend Dr. Navin CHATLANI, who enlightened me with the first glance of research and kept me motivated since then. He has always stood with me, always willing to spend time for discussions, generously shared many insights, inspired every bit of me towards new possibilities in

life. Without a doubt his precious support has contributed enormously to the success of this work.

My special words of thanks should also go to Dr. Ludovick LEPAULOUX for giving me my first taste of industrial experience and also supporting my research during the initial days of my career at INTEL. I unwittingly laid the foundation for my PhD thesis in the summer of 2012 when I worked on the loudspeaker IR measurements under his guidance. I owe a lot of gratitude to him for always being there for me and I feel privileged to be associated with a person like him during my life.

I express my sincere word of thanks to my colleagues and friends Dr. Christelle and Dr. Moctar for many stimulating and entertaining discussions, for extending moral support and helping hands to me without fail over the last few years. Their assistance during the presented work is gratefully acknowledged.

In the same vein, I would like to extend great thanks to Gurhan, Peter, Kim, Fabrice, Guillaume, Philippe, Heide, Nicholas, and the other Audio DSP Group members at INTEL who offered their time, support, encouragement and many helpful discussions.

My acknowledgement will never be complete without the special mention of my fellow PhD colleagues and friends at EURECOM. Some very special words of gratitude go to Giovanni, Robin, Rajeev, and José who have always been a major source of support and for always being there and bearing with me the good and bad times during my wonderful days of PhD. I would like to acknowledge my other friends at EURECOM for their moral support and motivation, which drives me to give my best: Soumya, Pramod, Raj, Valeria, Chiara, Natacha, Kalyan, Sumit, Massimiliano, Héctor, Pepe, Rui, Ghislain, and the list goes on, all of whom were always ready for a coffee break. The long discussions we had around the coffee acted as a stress-buster in many cases. A big thanks also go to friends outside EURECOM, Sveta Kalyan, Roomi, Sakshi, Anikó and Ester. Thanks guys for making last few years a memorable experience.

Finally, but by no means least, I would like to acknowledge the people who mean world to me: my parents, my wife, my sisters, and my uncle K.V.S.N. Sarma. I extend my respect and many thanks to my parents for their constant love and support. I wouldn't be here without their unremitting encouragement and reassurance in pursuing my interests. The last word goes for my wife and love of my life, Vineela, for keeping things going and for always showing how proud she is of me. Words cannot express my gratitude for everything you have done. Thank you for accompanying me on this adventure, I look forward to our next one!! I consider myself the luckiest in the world to have such a supportive family, standing behind me with their love and support. They are the most important people in my world and I dedicate this thesis to them.

Antibes, 09 Nov. 2017 ***Leela Krishna Gudupudi***

Abstract

Portable electronic devices are increasingly becoming indispensable parts of everyday life. Driven by the digital revolution, consumer electronics are becoming increasingly smaller and less expensive. In this light, there is an enormous demand for low-cost transducers. Unfortunately, the low-cost/miniature loudspeakers are a major source of nonlinear distortion effects.

In a hands-free communication environment, nonlinear distortion not only impairs the speech intelligibility but also degrades the performance of speech enhancement algorithms like acoustic echo canceller which work on the assumptions of linearity. Acoustic Echo Cancellation (AEC) has been a very active area of research for many decades. Because of the nonlinearities in the acoustic echopath, the AEC problem is now become more challenging and reformulated as Nonlinear Acoustic Echo Cancellation (NAEC), which is today an active research area. This thesis focuses on the analysis, identification and characterisation of nonlinear distortion in loudspeakers and its application to NAEC.

This thesis is primarily divided into two parts. The first part aims at finding a reliable nonlinear model that emulates the loudspeaker response for the purpose of predicting and preventing the nonlinear distortion. First the nonlinear loudspeaker system identification problem is addressed. After discussing the exponential sine-sweep excitation based nonlinear convolution technique for identifying the nonlinear loudspeakers, we focused on empirical loudspeaker modeling. We compared the synthesized outputs of two loudspeaker models to empirically measured, real loudspeaker outputs. The work suggests that the generalized polynomial Hammerstein model (GPHM) approximates more reliable practical nonlinear loudspeaker behavior.

Another study reveals that the Echo Return Loss Enhancement (ERLE) performance of a NAEC algorithm can be inflated using the nonlinear echo signals synthesized using power-series model (PSM), a common practice followed in the literature for NAEC performance evaluation. In contrast, the results generated with the GPHM model better reflect practical measurements and is thus an appealing alternative model for future evaluations of NAEC performance.

The second part of the thesis is majorly devoted to the NAEC problem. After discussing the state-of-the-art NAEC solutions, we have presented a comprehensive performance and stability analysis of the widely used NAEC algorithms. The results demonstrated that the popular NAEC solutions perform better only in a few idealistic environments and are less competent in most of the practical acoustic environments. We then proposed a novel approach to NAEC based on Empirical Mode Decomposition (EMD), a recently developed technique for nonlinear and nonstationary signal analysis. EMD decomposes any signal into a finite number of time varying sub-band signals termed intrinsic mode functions (IMFs). The new approach to NAEC incorporates this multi-resolution analysis with conventional power filtering to estimate nonlinear echo in each IMF. Comparative experiments with a competitive baseline approach to NAEC (based on pure power filtering) show that the new EMD approach achieves greater nonlinear echo reduction and faster convergence. However, EMD induced delay and the computational complexity of this approach are major limitations.

The next part of the work is our first step to align the analysis of nonlinear distortion in loudspeakers to its physical origins. We consider the application of Hilbert-Huang Transform (HHT) to the analysis of nonlinear distortion in miniature loudspeakers. Based on EMD and the Hilbert Transform (HT), HHT offers a new time-frequency analysis method (referred as Hilbert-Huang spectrum) with instantaneous time and frequency resolution unlike the conventional methods. Instantaneous amplitude (IA) and frequency (IF) parameters give more detailed and enhanced representation of the underlying nonlinear behaviour of the distorted signals. On the basis of the results of this work, we reported an alternative interpretation of loudspeaker nonlinearities through the *cumulative* effects of harmonic content and intra-wave amplitude-and-frequency modulation. These new findings may stimulate and reshape the future direction of NAEC research.

Although this thesis mainly focuses on the nonlinear distortion in the context of hands-free telephone systems, similar techniques and practices can also be applicable to other hands-free consumer devices.

Résumé

Cette thèse porte sur l'analyse, l'identification et la caractérisation de la distorsion nonlinéaire dans les haut-parleurs et son application à l'annulation d'écho acoustique nonlinéaire (ou NAEC, pour "Nonlinear Acoustic Echo Cancellation").

La première partie de la thèse vise à la dérivation d'un modèle de haut-parleur plus précis et empirique. Celui-ci émule la réponse fréquentielle du haut-parleur dans le but de prédire et d'empêcher la distorsion nonlinéaire. Les travaux de recherche suggèrent que le modèle de Hammerstein généralisé se rapproche plus fiablement d'un comportement de haut-parleur nonlinéaire.

Dans la partie suivante, après avoir discuté les études avancées de development des algorithms de NAEC, nous présenterons l'analyse des performances des algorithmes les plus utilisés. Les résultats ont démontré que les solutions populaires n'obtiennent de meilleurs résultats que dans quelques conditions idéales et sont moins performants dans la plupart des environnements acoustiques réels. Nous proposons ensuite une nouvelle approche de NAEC basée sur la décomposition modale empirique (ou EMD, pour "Empirical Mode Decomposition"), une technique récemment développée pour l'analyse de signaux nonlinéaires et non-stationnaires. Des expériences comparatives sur des techniques de reference montrent que la nouvelle approche (NAEC basée sur la EMD) permet d'obtenir une plus grande réduction d'écho nonlinéaire et une convergence plus rapide.

Dans l'étape qui suit, les travaux mis en place sont le commencement sur l'établissement de la correspondence entre l'analyse de la distorsion non linéaire dans les haut-parleurs à ses origines physiques. Nous considérons l'application de la transformée d'Hilbert-Huang (ou HHT, pour "Hilbert-Huang Transform") à l'analyse de la distorsion nonlinéaire dans les haut-parleurs. Sur la base des résultats de cette étude, nous avons rapporté une interprétation alternative des nonlinéarités des haut-parleurs à travers les effets cumulatifs du contenu harmonique et de la modulation en amplitude et en fréquence. Ces nouvelles conclusions pourraient stimuler et renouveler la direction future de la recherche sur la NAEC.

Contents

Acknowledgements **i**

Abstract **iii**

List of Figures **xi**

List of Tables **xvii**

1 Introduction **1**
1.1 Acoustic Echo Cancellation 2
1.2 Nonlinear Acoustic Echo Cancellation 6
1.3 Goal of the Thesis 7
1.4 Thesis Outline and Contributions 8

I Nonlinear Systems: Modeling and Characterization **13**

2 Nonlinear Systems and Modeling **17**
2.1 Linear vs. Nonlinear Systems 17
2.1.1 Linear Systems 17
2.1.2 Nonlinear Systems 18
2.1.3 Memory Effects 19
2.2 Modeling Nonlinear Systems 19
2.2.1 Taylor Series 19
2.2.2 Volterra Series Expansion 21
2.2.3 Limitations of Volterra Series 23
2.2.4 Alternative Block-Oriented Models 25
2.2.5 Wiener Series Expansion 29
2.3 Quantifying Nonlinear Distortion 30
2.3.1 Total Harmonic Distortion 30
2.3.2 Linear-to-NonLinear-Ratio 31
2.4 Summary 32

3 Nonlinear Distortion in a LEMS **33**
3.1 Nonlinear Sources in the Acoustic Echopath . . . 33
3.1.1 ADC/DAC Distortions . . . 35
3.1.2 Power Amplifier Distortion . . . 36
3.1.3 Distortions in Loudspeakers . . . 37
3.2 Nonlinear Loudspeaker Modeling . . . 43
3.3 Loudspeaker System Identification . . . 46
3.3.1 Excitation Signal Generation . . . 47
3.3.2 Inverse Filter Generation . . . 49
3.3.3 Deconvolution: "Full" IR Measurement . . . 52
3.3.4 Computation of the Simplified Volterra kernels . . . 54
3.4 Summary . . . 55

4 Comparative Studies of Simulated and Real-Device Experiments **57**
4.1 Identification of a Real Mobile Phone Loudspeaker . . . 57
4.1.1 Experimental Setup . . . 57
4.1.2 Data Acquisition . . . 58
4.1.3 Deconvolution . . . 60
4.1.4 Equalization . . . 60
4.2 A Comparison of Loudspeaker Models . . . 62
4.2.1 Synthetic Signal Generation . . . 62
4.2.2 Assessment . . . 64
4.3 Validation of the GPHM . . . 67
4.3.1 Device Characterization . . . 67
4.3.2 Results . . . 69
4.4 Summary . . . 71

II Nonlinear Acoustic Echo Cancellation **73**

5 State-of-the-Art NAEC Solutions **77**
5.1 Hardware-based Solutions . . . 77
5.2 Software-based Solutions . . . 79
5.2.1 Nonlinear Pre-Filtering . . . 79
5.2.2 Nonlinear Post-Filtering . . . 82
5.2.3 Nonlinear Adaptive Filtering . . . 87
5.3 Influence of the simulated nonlinear signal models on NAEC evaluation . 94
5.3.1 Cascaded Model . . . 94
5.3.2 Experimental results . . . 96
5.4 Comprehensive performance analysis of NAEC algorithms . . . 97

5.4.1 Parallel Model . . . 98
5.4.2 Experimental work . . . 99
5.5 Summary . . . 104

6 Empirical Mode Decomposition 107
6.1 Why study EMD? . . . 107
6.2 Introduction to EMD . . . 108
6.2.1 EMD Analysis: The Sifting Process . . . 109
6.2.2 The Stopping Criteria . . . 114
6.2.3 EMD Synthesis . . . 115
6.2.4 EMD Applications . . . 116
6.3 Application of EMD to NAEC . . . 119
6.3.1 NAEC Structure . . . 121
6.3.2 Adaptive filtering . . . 123
6.3.3 Experimental work . . . 123
6.3.4 Experimental results . . . 125
6.4 Summary . . . 126

7 An Alternative Interpretation of Loudspeaker Nonlinearities 129
7.1 Time-Frequency Analysis . . . 129
7.2 Instantaneous frequency and The Hilbert transform . . . 131
7.3 The Hilbert-Huang Transform . . . 132
7.3.1 Hilbert-Huang Spectrum . . . 133
7.3.2 Relation to Fourier techniques . . . 134
7.4 Loudspeaker distortion analysis . . . 134
7.4.1 Experimental set-up . . . 135
7.4.2 HHT Analysis . . . 135
7.5 Validation of HHT . . . 138
7.6 Limitations of the HHT/EMD . . . 140
7.7 Recent advancements/extensions of the standard EMD . . . 141
7.7.1 EEMD . . . 141
7.7.2 MEMD . . . 142
7.7.3 Hilbert Spectral Analysis (HSA) . . . 143
7.8 Summary . . . 143

8 Conclusions and Future Directions 145
8.1 Contributions . . . 145
8.2 Future directions . . . 148

A Sommaire de la thèse en français **151**
A.1 Résumé 151
A.2 Introduction 152
A.2.1 Annulation d'écho acoustique 152
A.2.2 L'annulation d'écho acoustique nonlinéaire 155
A.3 Modélisation de distorsion nonlinéaire 157
A.3.1 Comparaison des modèles de haut-parleurs 160
A.3.2 Impact des signaux d'écho nonlinéaires simulés sur l'évaluation NAEC 163
A.3.3 Validation du GPHM 165
A.4 La Décomposition Modale Empirique 169
A.4.1 Introduction à EMD 170
A.4.2 Application de l'EMD à NAEC 175
A.5 Une interprétation alternative des nonlinéarités de haut-parleur 179
A.5.1 Fréquence instantanée et transformée de Hilbert 179
A.5.2 La transformation de Hilbert-Huang (HHT) 181
A.5.3 Analyse de distorsion de haut-parleur 183
A.5.4 Validation de HHT 186

Bibliography **201**

List of Figures

1.1 System model illustrating the acoustical coupling in the LEMS and a general approach to adaptive AEC. 3
1.2 ERLE test results to compare the performance of linear AEC algorithms . 4
1.3 ERLE test results to compare the performance of linear AEC algorithms in linear and nonlinear environments. 6

2.1 Homogeneity property of linear systems 18
2.2 Superposition property of linear systems 18
2.3 A block diagram representing the P^{th}-order Volterra kernel 23
2.4 A block diagram of the Hammerstein model 26
2.5 A block diagram of the Wiener model 27
2.6 A block diagram of the Hammerstein-Wiener model 27
2.7 A block diagram of the Wiener-Hammerstein model 28

3.1 System model illustrating a general approach to acoustic echo cancellation and the nonlinear sources in the LEMS. 34
3.2 A block diagram representing the process of ADC and DAC operation. . . 35
3.3 Different saturation curves for amplifiers 36
3.4 Internal diagram of a electro-dynamic loudspeaker 37
3.5 Different sizes of loudspeakers 39
3.6 The complete equivalent electrical circuit of the electro-dynamic loudspeaker 40
3.7 The generalized polynomial Hammerstein model. 47
3.8 Block diagram representing the process of nonlinear convolution technique. 48
3.9 Exponential sine-sweep signal in the time domain 49
3.10 Time-frequency representation of exponential sine-sweep signal 50
3.11 Time-reversed exponential sine-sweep signal 51
3.12 Inverse sine-sweep signal after amplitude modulation 51
3.13 An example of a "full" IR 53

4.1 Experimental setup used for the identification of a real mobile phone loudspeaker 58

4.2 Representation of the input and the output signals of a mobile phone loudspeaker 59
4.3 "Full" IR of a real mobile phone loudspeaker 61
4.4 Time domain representation of the linear IR (left) and the 3^{rd}-order IR of a mobile phone loudspeaker 61
4.5 Figure shows the variation of the frequency response curves of a mobile phone loudspeaker IR's before and after RIR equalization 63
4.6 The spectrogram of (a) Clean speech signal (b) A real mobile phone loudspeaker response (c) Synthesized speech signal using PSM (d) Synthesized speech signal using GPHM 65
4.7 (a) The PESQ scores between real measured loudspeaker signals and those synthesized with PSM and GPHM models. (b) The PESQ scores between clean speech signal and those synthesized with PSM and GPHM models along with a real measured loudspeaker signal. 65
4.8 An illustration of the cepstral distance between real measured loudspeaker signals and those synthesized with PSM and GPHM models. 67
4.9 Application of the nonlinear loudspeaker model. Input signals are processed according to the nonlinear loudspeaker response (LSR) and a room impulse response (RIR). The loudspeaker response (LSR) is synthesized using the GPHM model in Fig. 3.7 67
4.10 An illustration of nonlinear characterization and model performance. The first row illustrates the response of each of three devices to the exponential sine-sweep input signal. Rows two and three illustrate the performance of the resulting nonlinear model to sine-sweep and real-speech input signals respectively. Results shown for different orders of nonlinearity P (vertical axes) and Volterra kernel lengths L (horizontal axes). 68

5.1 Approaches to handle nonlinearities. 78
5.2 Loudspeaker linearisation system. 80
5.3 Block diagram of the Loudspeaker linearisation system for eliminating the second-order nonlinear distortion. 81
5.4 Handling nonlinearities in the LEMS using NRES. 83
5.5 Structure of NAEC. 87
5.6 Structure of parallel approach based NAEC. 88
5.7 Structure of cascaded approach based NAEC. 89
5.8 An illustration of the cascaded model NAEC. In the p^{th}-channel, the input signal vector passes through a low-pass filter (LPF) with cut-off frequency $f_s/2p$ to avoid aliasing. 95
5.9 NAEC performance in terms ERLE with either real recorded nonlinear echo signals or those synthesised with the PSM or the GPHM models. . . 96

5.10 An illustration of the parallel/power-filter model NAEC. In the p^{th}-channel, the input signal vector passes through a low-pass filter (LPF) with cut-off frequency $f_s/2p$ to avoid aliasing. 98

5.11 A performance comparison of the three AECs in terms of mean ERLE in the presence of different echo scenarios 101

5.12 A performance comparison of the three AECs in terms of ERLE in the presence of dynamically varying orders of nonlinear echo 102

5.13 A performance comparison of the three AECs in terms of mean ERLE in the presence of real recorded nonlinear echo as a function of pre-processor filter lengths 103

6.1 An illustration of the basic idea of EMD. Illustrated is a given parent data (Blue line in the left figure) and is considered as faster oscillation (top figure on the right) overlying to slower oscillation (bottom figure on the right). 109

6.2 The flowchart illustrating the sifting process procedure to decompose any complicated signal into a set of IMFs. 111

6.3 An illustration of the sifting process, which decomposes a test signal into a set of IMFs. 113

6.4 An illustration of EMD. Illustrated is a given parent signal (top) and the resulting 6 IMFs. 114

6.5 An illustration of EMD. Illustrated is a clean speech signal (top) and the first 6 IMFs. 117

6.6 IMF variance plots: indicates energy content in each IMF 118

6.7 A real mobile phone loudspeaker (in hands-free mode at maximum gain) excited by a clean speech signal (top) and simultaneously recorded using a high quality microphone (bottom) 120

6.8 A comparison of IMF variance plots of a clean speech signal and a real loudspeaker recorded signal 121

6.9 Structure of EMD based NAEC. 122

6.10 A performance comparison in terms of ERLE for the new EMD-based approach to NAEC and a baseline power filter approach. 124

7.1 Miniature loudspeaker (microspeaker) response to a pure sinusoidal input at 1kHz, sampled at 8kHz; 130

7.2 Experimental setup in an anechoic chamber to measure loudspeaker outputs. 135

7.3 (a) STFT spectrogram of a mobile phone loudspeaker response to a pure sinusoidal input at 1kHz, sampled at 48kHz; (b) Loudspeaker response to 1kHz sine tone is decomposed by the EMD, resulting in the 8 IMFs, first 4 IMFs are listed above and others are not displayed since they are almost zero; (c) IA profiles of the IMFs obtained by HHT; (d) IF profiles of the IMFs obtained by HHT 136
7.4 A real mobile phone loudspeaker response to 1kHz pure sine tone. The wave-profile deformation caused by the nonlinear distortion is not constant throughout the time. 137
7.5 Time-frequency-energy distributions: (a) STFT spectrogram of a mobile phone loudspeaker response to a pure sinusoidal input at 500Hz, sampled at 48kHz; (b) the IF profiles obtained by HHT; (c) IF profiles for a high-quality loudspeaker response to the same input; (d) the IF profiles of the same loudspeaker subject to an input excitation comprised of pure sinusoidal at 500Hz and its third harmonic. 138
7.6 (a) STFT spectrogram of a mobile phone loudspeaker response to a pure sinusoidal input at 2kHz, sampled at 48kHz; (b) Loudspeaker response to 2kHz sine tone (zoomed in) is decomposed by the EMD, resulting only a single IMF, meaning the loudspeaker response itself satisfy the EMD properties; (c) IA profile of the IMF obtained by HHT; (d) IF profile of the IMF obtained by HHT, indicating very low percentage of modulation 139

A.1 Modèle de système illustrant le couplage acoustique dans le système LEMS et une approche générale de l'AEC adaptatif. 153
A.2 Les résultats des tests ERLE pour comparer les performances des algorithmes linéaires AEC dans des environnements linéaires et nonlinéaires. . 155
A.3 Structure de l'approche parallèle basée NAEC. 156
A.4 Structure de l'approche en cascade basée NAEC. 157
A.5 Une illustration du modèle en cascade NAEC. Dans le canal p^{th}, le vecteur de signal d'entrée passe à travers un filtre passe-bas (LPF) avec une fréquence de coupure $f_s/2p$ pour éviter l'aliasing. 158
A.6 le modèle de Hammerstein généralisé polynomiale (GPHM). 159
A.7 Configuration expérimentale utilisée pour l'identification d'un réel haut-parleur de téléphonie mobile 160
A.8 Le spectrogramme de (a) signal de parole propre (b) une réponse de haut-parleur de téléphone mobile réel (c) signal de parole synthétisé utilisant PSM (d) signal de parole synthétisé utilisant GPHM 162
A.9 Une illustration de la distance cepstrale entre les signaux de haut-parleurs mesurés réels et ceux synthétisés avec les modèles PSM et GPHM. 162

A.10 Performances NAEC en termes ERLE avec des signaux d'écho nonlinéaires réels enregistrés ou ceux synthétisés avec les modèles PSM ou GPHM. . . 164

A.11 Une illustration de la caractérisation nonlinéaire et des performances du modèle GPHM. La première rangée illustre la réponse de chacun des trois dispositifs mobiles au signal d'entrée du balayage sinusoïdal exponentiel. Les lignes deux et trois illustrent les performances du modèle nonlinéaire résultant pour sinus balayent et signaux d'entrée de la parole réelle. Les résultats sont affichés pour différents ordres de nonlinéarité P (axes verticaux) et longueurs de noyau de Volterra L (axes horizontaux). . 166

A.12 Une illustration de l'idée de base d'EMD. Illustré est une donnée parent donnée (ligne bleue dans la figure de gauche) et est considérée comme une oscillation plus rapide (la figure du haut à droite) recouvrant une oscillation plus lente (figure du bas à droite). 170

A.13 L'organigramme illustre la procédure du processus de criblage pour décomposer tout signal compliqué en un ensemble de IMFs. 172

A.14 Une illustration d'EMD. Illustré est un signal parent donné (en haut) et les 6 IMFs résultants. 173

A.15 Une illustration d'EMD. Illustré est un signal de parole propre (en haut) et les 6 premiers IMFs. 175

A.16 Structure de NAEC basée sur EMD. 176

A.17 Une comparaison des performances en termes d'ERLE pour la nouvelle approche EMD basée sur NAEC et une approche de filtre de puissance de base. 178

A.18 Montage expérimental dans une chambre anéchoïque pour mesurer les sorties des haut-parleurs. 183

A.19 (a) Spectrogramme STFT de la réponse d'un haut-parleur de téléphone mobile à une entrée sinusoïdale pure à 1 kHz, échantillonnée à 48 kHz; (b) La réponse du haut-parleur au tonus sinusoïdal de 1 kHz est décomposée par le EMD, ce qui donne 8 IMFs, les 4 premiers IMF sont énumérés ci-dessus et les autres ne sont pas affichés puisqu'ils sont presque nuls; (c) Profils IA des IMF obtenus par HHT; (d) Profils IF des IMF obtenus par HHT 184

A.20 Une vraie réponse de haut-parleur de téléphone mobile à ton pur sinus 1kHz. La déformation du profil d'onde causée par la distorsion nonlinéaire n'est pas constante tout au long du temps. 186

A.21 Distributions temps-fréquence-énergie: a) spectrogramme STFT de la réponse d'un haut-parleur de téléphone mobile à une entrée sinusoïdale pure à 500 Hz, échantillonnée à 48 kHz; (b) les profils IF obtenus par HHT; (c) profils IF pour une réponse de haut-parleur de haute qualité à la même entrée; (d) les profils IF du même haut-parleur soumis à une excitation d'entrée composée de sinusoïdale pure à 500 Hz et de son troisième harmonique. 187

A.22 a) spectrogramme STFT de la réponse d'un haut-parleur de téléphone mobile à une entrée sinusoïdale pure à 2 kHz, échantillonnée à 48 kHz; (b) La réponse du haut-parleur au signal sinusoïdal de 2 kHz (zoom avant) est décomposée par le EMD, ce qui donne un seul IMF, ce qui signifie que la réponse du haut-parleur satisfait elle-même aux propriétés EMD; (c) profil IA du IMF obtenu par HHT; (d) Profil IF du IMF obtenu par HHT, indiquant un très faible pourcentage de modulation 188

List of Tables

5.1 Computational Complexity Comparison 100

6.1 Choice of parameters in each Filter Chamber (FC) 125

Chapter 1

Introduction

Mobile communications technology has evolved substantially in the past decade. Accordingly, the consumer electronic devices intended for mobile communications have been experiencing an explosive growth in the recent years. These devices range from smart-phones to tablet PCs to voice-activated speakers to wireless smart ear-buds to huge range of hands-free car kits, etc..

One of the most important reasons for this trend could be the evolution of hands-free technology that has yielded significantly better sound quality. If the recent growth of the hands-free communication devices market is any indication, many people understandably prefer hands-free communication. In any hands-free communication environment, acoustic echo cancellation (AEC) and noise cancellation play an increasingly important role in ensuring satisfactory speech quality. In this thesis, we focused on the acoustic echo cancellation problem. Many different devices are equipped with loudspeakers and microphones for a variety of different purposes and often these transducers are mounted in close proximity to one another. This acoustic coupling between the loudspeaker and the microphone along with subsequent additive reflections causes acoustic echo. In the case of mobile telephony, this acoustic echo will be transmitted to the far-end user and the conversation can be annoying to unbearable depending on the round-trip delay of the system. In the case of voice-activated assistant speakers (typical examples: Amazon Echo and Google Home), this acoustic echo is a source of interference for the automatic speech recognition engines affecting its performance (wake-word detection and/or speech recognition rate). Thus, the acoustic echo degrades the quality of the voice communication by degrading speech intelligibility and listening comfort. In order to combat the acoustic echo phenomenon, it is often necessary to employ an acoustic echo canceller.

1.1 Acoustic Echo Cancellation

Acoustic echo cancellation (AEC) is a decades-old problem in signal processing since the introduction of full-duplex voice communications, and it is still an active field of research. AEC is based on a well-established system identification approach. The acoustic echopath (from the loudspeaker to the microphone) is highly dynamic and subjected to variation in time, as a consequence of the modification of the acoustical characteristics of the Loudspeaker Enclosure Microphone System (LEMS). Hence, AEC generally uses a linear adaptive transversal filter to estimate the digital replica of the transfer function of the LEMS. A typical adaptive AEC system is illustrated in Fig. 1.1, where $d(n), s(n)$ and $v(n)$ represent the echo signal, near-end speech signal and noise respectively. Most of the scenarios in this thesis assume microphone signal contains echo only, that is $y(n) = d(n)$ and $s(n) = v(n) = 0$, unless it is specified. The far-end signal $x(n)$ is passed through the adaptive filter $\hat{h}(n)$ to synthesize the echo signal $\hat{y}(n)$, which is then subtracted from the microphone signal $y(n)$ to cancel the acoustic echo. If the adaptive filter impulse response, $\hat{h}(n)$, matches with that of LEMS, $h(n)$, (convergence) then the echo will be eliminated without any artefacts. However, achieving a perfect convergence is a quite challenging task especially while handling highly nonstationary signals like speech. Besides there are many other factors like near-end background noise, double-talk period (period at which both the near-end speech and the echo are present at the same time) and nonlinearities impair the performance of echo canceller. In the following, we examine the performance of few popular adaptive algorithms in terms of their convergence and Echo Reduction Loss Enhancement (ERLE). ERLE is a quantitative measure which represents the reduction in energy (in dB) of the microphone signal ($d(n)$) achieved by echo reduction. ERLE is given by:

$$ERLE = 10log\frac{E\{d^2(n)\}}{E\{e^2(n)\}} \tag{1.1}$$

where $e(n)$ is the AEC output signal to be transmitted to the far-end user.

Simulation Environment

The LEM system is characterized by the combination of a loudspeaker, the room and the microphone impulse responses. For the simulations considered in this chapter, the LEM system is modelled with the help of an empirically measured loudspeaker impulse response, $h_1(n)$ of size $L_1 = 128$ taps, and a room impulse response $h_{rir}(n)$ of size $L = 256$ taps, selected from the Aachen RIR database [1]. Microphone impulse response

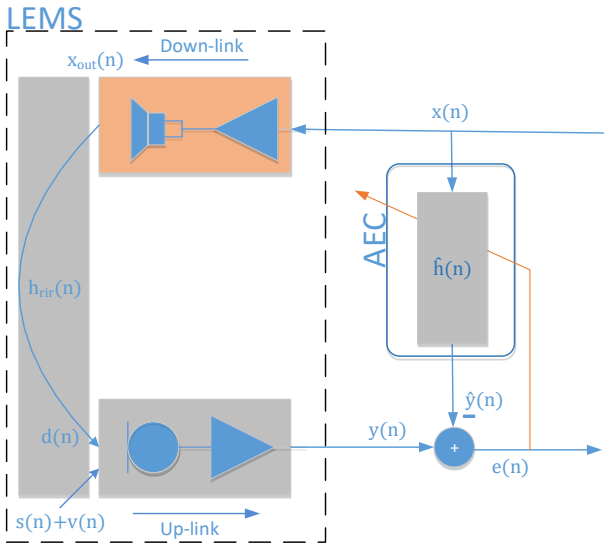


Figure 1.1 – System model illustrating the acoustical coupling in the LEMS and a general approach to adaptive AEC.

is ignored. A 5 second speech signal sampled at 8kHz is concatenated 12 times to produce the far-end test signal $x(n)$ of sufficient duration to ensure the convergence of each algorithm. The loudspeaker output signal, $x_{out}(n)$, is synthesized according to:

$$x_{out}(n) = \sum_{i=0}^{L_1-1} x(n-i)h_1(i) \tag{1.2}$$

The microphone output signal or the linear echo signal $y\,(n)$ is generated according to:

$$y\,(n) = \sum_{i=0}^{L-1} x_{out}(n-i)h_{rir}(i) \tag{1.3}$$

Following are the well-known linear adaptive algorithms considered for this study:

- Least Mean Square (LMS) algorithm with $\mu = 0.16$
- Normalized-LMS (NLMS) algorithm with $\mu = 1$
- Frequency Block-LMS (FBLMS) algorithm with $\mu = 0.5$ and the block length $B = 256$

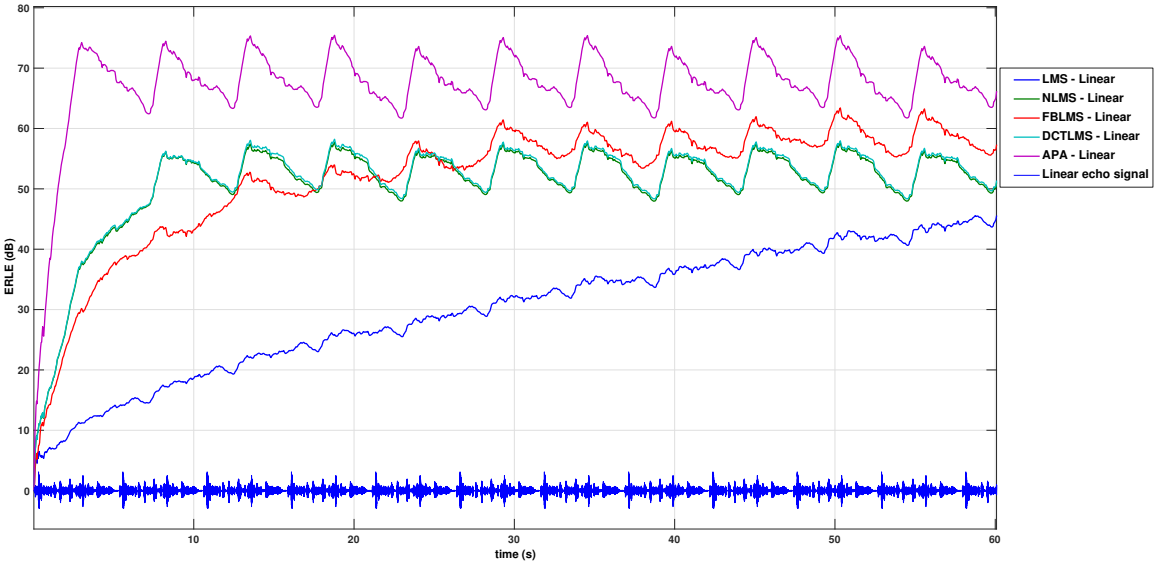


Figure 1.2 – ERLE test results to compare the performance of linear AEC algorithms

- Discrete Cosine Transform-LMS (DCTLMS) algorithm with $\mu = 0.5$
- Affine Projection Algorithm (APA) algorithm with $\mu = 1$ and order 2

Details of these adaptive algorithms are well described in the literature (for example, [2,3]), and shall not be repeated in this thesis. The step size μ of each algorithm is chosen such that it achieve maximum ERLE after convergence. The linear AEC was operated in equal modelling scenario, that is the length of the adaptive filter is considered as $L_1 + L - 1 = 383$ taps. Figure 1.2 illustrates the behaviour of the linear adaptive algorithms in terms of ERLE.

The results indicate that the APA algorithm clearly outperforms the rest, both in terms of the convergence rate and the maximum achievable ERLE. Despite being widely accepted as fast, simple in implementation and robust in harsh environments, the NLMS algorithm is inferior to APA of 2^{nd}-order. This is because the NLMS algorithm converges very slowly in the presence of highly correlated excitation signals like speech. One way to overcome this problem is to decorrelate (or pre-whiten) the incoming excitation signals [2]. Regarding the DCTLMS algorithm, there is no noticeable difference between the performance of DCTLMS and NLMS in terms of initial convergence and the maximum achievable ERLE. The performance behaviour of the FBLMS algorithm is however exceptional. Its convergence rate is remarkably high, because of the block-by-block processing. However, as the time progresses, the FBLMS algorithm became superior to the NLMS and the DCTLMS algorithms in terms of maximum achievable ERLE. As expected, the performance of LMS algorithm is poor as its behaviour is highly dependent

on the eigenvalue spread of the input signal's autocorrelation matrix. The larger the eigenvalue spread, the slower the convergence speed [2].

Most of the work in the literature assumes linearity of the electronic components in the LEMS. Under such linear conditions AEC algorithms generally perform well as seen in Fig. 1.2. However, the trend of miniaturization in electronic devices industry, particularly in the mobile devices, has forced the characteristic sizes of the electronic components to shrink accordingly. The miniaturization of the transducers and their associated electronic components along with the mobile devices enclosures often introduce non-negligible nonlinear distortion in the acoustic echopath. The linear AEC cannot handle the nonlinear echo in the LEMS and transmitted back to the far-end user. Further, the nonlinear distortion degrades the performance of linear AEC leading to high residual echo in the uplink signal [4, 5, 6]. Each of the linear adaptive algorithms employed by linear AEC behaves differently to nonlinear distortion in the LEMS. In order to witness the same, we examine the performance of above mentioned adaptive algorithms in the presence of nonlinear distortion. The simulation environment used is similar to the one above, except that the loudspeaker is assumed as a nonlinear device. To emulate the nonlinear loudspeaker output signal, $x_{out}(n)$, the linear ($h_1(n)$) and the harmonic impulse responses (the so called higher-order diagonal Volterra kernels) $h_p(n), p \in [2, P]$ of a real mobile-device loudspeaker are measured empirically as described in the next chapter using the nonlinear system identification method reported in [7,8]. The nonlinear loudspeaker output signal is synthesized according to:

$$x_{out}(n) = \sum_{p=1}^{P} \sum_{i=0}^{L_p-1} x^p(n-i) h_p(i) \tag{1.4}$$

where $x\,(n)$ is the downlink/reference signal and the diagonal Volterra kernel $h_p(n)$ is the $L_p = 128$-tap linear filter corresponding to the p^{th} harmonic. In order to avoid aliasing while generating the loudspeaker output signal, $x_{out}(n)$, the input vector, represented by $\mathbf{x}(n) = [x(n), \ldots, x(n - L_p + 1)]^T$, is passed through a low-pass filter (or anti-aliasing filter) with cut-off frequency $f_s/2p$ before taken to the p^{th} power. The microphone output signal with nonlinear echo is generated according to Eq. 1.3.

Figure 1.3 illustrates the behavior of the linear adaptive algorithms in terms of ERLE in both linear and nonlinear environments. These curves clearly demonstrate the impact of nonlinear distortion on each linear AEC algorithm. Downlink nonlinearities in the LEMS reduce the maximum achievable ERLE by each algorithm. The observations are consistent with the results published by Moctar et.al. in [5]. NLMS and DCTLMS perform similarly

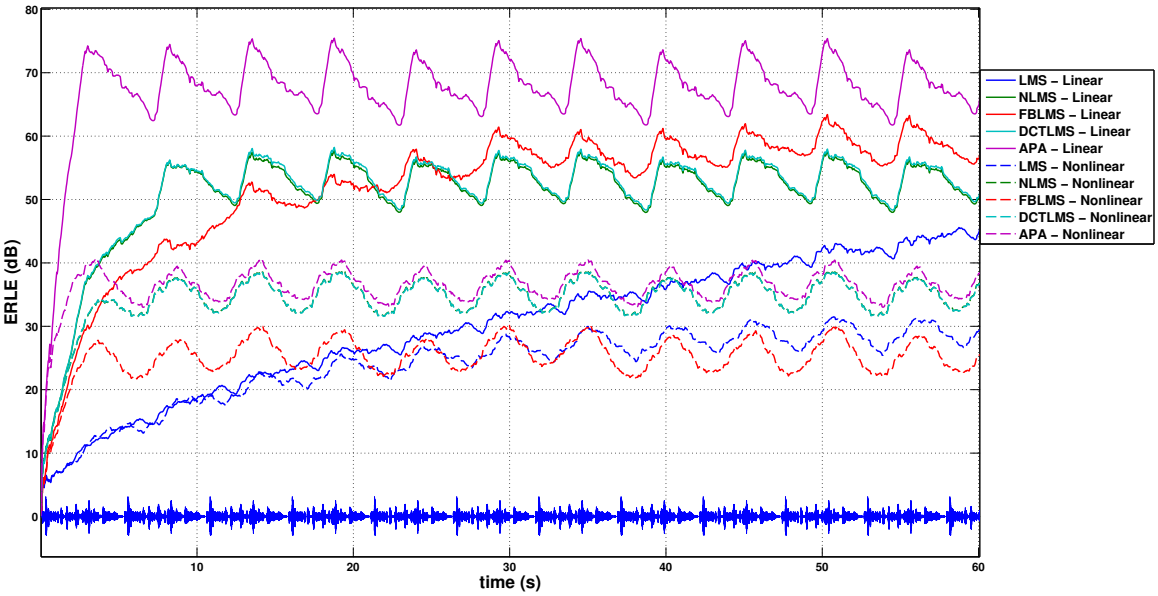


Figure 1.3 – ERLE test results to compare the performance of linear AEC algorithms in linear and nonlinear environments.

in both linear and nonlinear environments. FBLMS and APA algorithms are severely affected by nonlinearities. Despite the fact that the initial convergence is better even with the nonlinear distortion, the APA algorithm of 2^{nd}-order behaves almost similar to the NLMS algorithm in terms of ERLE. Non-intrusive tests confirmed that the performance of APA algorithm drastically decreases below the level of NLMS algorithm when nonlinear distortion increases. Even though the performance of LMS algorithm is poor, it stays robust to nonlinearities (in terms of difference in ERLE) compared to other algorithms. Refer [5, 6] for further details. From this discussion, it is plausible that conventional linear AEC algorithms alone are insufficient to tackle the nonlinear distortion in the LEMS. Accordingly, the age-old problem of AEC is now become more challenging and reformulated as Nonlinear Acoustic Echo Cancellation (NAEC), which is today an active research area.

1.2 Nonlinear Acoustic Echo Cancellation

Nonlinear distortion redistributes energy within the spectrum and attempting to cancel nonlinear echo using linear AEC leaves additional residual echo in the uplink signal, resulting in the degradation in ERLE. This highlights the need for advanced algorithms that tackles the nonlinear distortion. Nonlinear acoustic echo canceller must be capable of identifying and tracking not only the linear impulse response of the LEMS but also the nonlinearities associated with the device components. Loudspeaker and its associated

components (power amplifier and Digital-to-Analog Converter (DAC)) in the down-link path are considered the major sources of significant nonlinear distortion in the LEMS and are most studied in the literature. Much effort has been made by researchers in the last decade to characterize the nonlinearities of electrodynamic loudspeakers [9]. We summarize their main results in Chapter 1 of this thesis.

Unlike the conventional linear AEC system, the reference signal from the downlink channel alone is not sufficient for use in a NAEC system, but prior knowledge and models of the nonlinearities in the loudspeaker system are also required. There are three fundamental approaches to tackle the nonlinear distortion in the LEMS echopath:

- Nonlinear pre-filtering
- Nonlinear post-filtering
- Nonlinear adaptive filtering

The first approach aims at a linearisation of the loudspeaker and its associated components through nonlinear pre-filtering of the far-end signal. In case of ideal pre-filtering of the loudspeaker signal, the entire LEMS could be safely assumed as linear and thus a linear adaptive filter is sufficient to achieve better echo cancellation performance. Note that, in this case, an exact model of the nonlinear loudspeaker system is of primary requirement and any misapprehensions can result in major gradient noise. The second approach is based on nonlinear post-filtering to suppress the residual nonlinear echo using a post-filter preceded by a conventional linear adaptive filter. However, the nonlinearities have an adverse effect on linear filtering which impacts upon nonlinear post-filtering and thus degrades the global performance. The third approach, which is more general, is based on nonlinear adaptive filtering in the AEC in order to adaptively learn the behaviour of the nonlinearities along with the linear impulse response of the LEMS. The main drawback of this method is the lack of an alternative reference signal to adaptively determine the nonlinear loudspeaker model coefficients independent from the RIR.

1.3 Goal of the Thesis

Several methods have been proposed in the past to address NAEC problem as discussed further in this thesis. However, the current state-of-the-art solutions generally accomplish only modest reductions in nonlinear echo besides their limitations. First, the stability of nonlinear systems is not guaranteed. Next, low convergence rate and high computational complexity prevent these methods from being widely used in practical applications.

In this thesis, we particularly investigate the nonlinear adaptive filtering approach to achieve robust NAEC performance. Approaches to NAEC based on nonlinear adaptive filtering depend fundamentally upon a discrete-time model of the loudspeaker. Several nonlinear loudspeaker models have been reported in the literature [10, 11, 12, 13, 14, 15]. The most popular are based on Volterra series [16]. The more closely the model reflects the reality, the better the performance of NAEC in practice and better the simplifications and trade-offs that can be made. Thus the goal of this thesis can therefore be summarized as (i) the derivation of the more accurate and empirical loudspeaker model, (ii) to study the exact phenomenon of loudspeaker nonlinear distortion, and (iii) to propose new solutions to improve the performance of NAEC approaches.

Although this thesis mainly focuses on the nonlinear distortion in the context of hands-free telephone systems, however similar techniques and practices can also be applicable to other hands-free consumer devices.

1.4 Thesis Outline and Contributions

This thesis is mainly divided into two parts. Part 1 defines the main sources of nonlinear distortion in the LEMS, and presents an approach to their empirical identification and modeling. Overall, the first part aims at finding an accurate nonlinear model that emulates the loudspeaker response for the purpose of predicting and preventing the nonlinear distortion. Part 2 introduces the state-of-the-art NAEC solutions before proposing a novel NAEC algorithm. This work leads to an alternative interpretation of loudspeaker nonlinear distortion using a relatively new time-frequency analysis technique known as the Hilbert-Huang Transform (HHT). Thus the thesis begins with a traditional interpretation of nonlinear distortion in loudspeakers and ends with a novel and accurate interpretation of nonlinear distortion, which marks a new beginning of NAEC research. Both parts of the thesis contain original contributions to the field.

Part-1

Chapter-2

It is known that the linear and time invariant (LTI) systems are well studied, but those properties are generally not applicable to nonlinear systems. This chapter explains the theoretical background on the nonlinear systems. The first part of the chapter focuses

on the properties of the nonlinear systems. Next, the focus shifts toward understanding and adequately modeling the complex behaviour of nonlinear systems. The discussion includes the introduction to the popular Volterra series expansion and the block-oriented models (example: Hammerstein model). The last part of this chapter covers different ways of quantifying nonlinear distortion.

Chapter-3

After a brief overview of the nonlinear systems and modeling in the previous chapter, the emphasis in Chapter 3 switches to identify a suitable loudspeaker model to emulate the nonlinear behaviour of the loudspeakers. The chapter starts with an introduction to the general sources of nonlinearities in the LEMS, including a detailed study on electro-dynamic loudspeaker nonlinearities.

The next part of this chapter focuses on three different nonlinear models suitable for comprehensive modeling of a nonlinear loudspeaker: Volterra series (with memory), power series (Volterra series without memory) and generalized polynomial Hammerstein model (GPHM).

There is a significant necessity to know the nonlinear dynamics of the loudspeaker to achieve a better NAEC performance. Thus the last part of this chapter investigates a well-known nonlinear system identification technique, referred to as *nonlinear convolution*, first proposed in [7, 17].

Chapter-4

This chapter focuses on the empirical and experimental research and aims to complement the theoretical study covered in the previous chapters. First we report the experimental approach to identify the nonlinear behaviour of a real mobile phone loudspeaker. Then as a first contribution, we investigate the suitability of Volterra series based models in modeling the real nonlinear loudspeakers by comparing the synthesized outputs to empirically measured loudspeaker outputs. This work indicates that the GPHM model resembles (both statistically and perceptually) stable and reliable practical nonlinear dynamics of the loudspeaker. Hence throughout the thesis we used GPHM model to synthesize the nonlinear loudspeaker output wherever applicable.

After identifying an appropriate nonlinear model, determining its optimal model parameters is one of the challenging issues in real-time applications. Therefore, as a next contribution, we choose to investigate further the accuracy of the GPHM model as

a function of its key parameters, namely the number of filter taps and the order of nonlinearities. This work highlights the challenges involved to model accurately the distortion introduced by nonlinear loudspeakers.

Part of the work in this chapter has resulted in the following publications:

- L. K. Gudupudi, C. Beaugeant, N. W. D. Evans, M. I. Mossi, and L. Lepauloux, "A comparison of different loudspeaker models to empirically estimated nonlinearities," in *Proc. HSCMA*, May 2014.

- L. K. Gudupudi, C. Beaugeant, and N. W. D. Evans, "Characterization and modelling of nonlinear loudspeakers," in *Proc. IWAENC*, Sept. 2014.

Part-2

Chapter-5

The aim of this chapter is to present a state-of-the-art review of the existing methods for the nonlinear acoustic echo cancellation and/or suppression. Several NAEC algorithms have been proposed in the literature to handle nonlinearites in the acoustic echopath and to maintain stable echo cancellation performance. However, their evaluation methodologies are not as compelling as their key design idea because most of them had never been tested under both real nonlinear echoes and real mobile phone loudspeaker data.

Most of the NAEC algorithms are developed based on two different rationales, parallel and cascaded approaches, each possessed its own merit and claimed outperforming the other. The claim has prone to subjectivity because the algorithms are compared only in few idealistic situations. Therefore, in the last part of the chapter, we conduct an in-depth performance analysis and comparison of the two typical NAEC structures under various and more practical situations.

Part of the work in this chapter has resulted in the following technical report:

- L. K. Gudupudi, M. I. Mossi, C. Beaugeant, and N. W. D. Evans, "Comprehensive performance and stability analysis of NAEC algorithms," Technical report, EURECOM, Sophia Antipolis, France, 2015.

Chapter-6

The next part of the work is our first step to align the analysis of nonlinear distortion to its physical origins. Data analysis plays an integral role in scientific research and understanding any unknown system and/or signals. Traditional Fourier-based data analysis methods such as the discrete Fourier transform (DFT) and the short-time Fourier transform (STFT) dominate the signal analysis field. These methods all assume linear, (short-term) stationary signals. Wavelet analysis designed to handle nonstationary data still assumes linearity. Accordingly, Fourier and wavelet methods may not be the most suitable approaches for the analysis of nonlinear loudspeakers.

Huang et al. proposed a new approach called Empirical Mode Decomposition (EMD) which is well-suited to the analysis of nonlinear and nonastationary signals [18]. The work aims to provide an alternative approach to signal analysis which goes beyond Fourier-based approaches. Unlike traditional approaches, EMD adapts the bases to the signal itself and can therefore yield more physically relevant results. In this thesis, we have studied the application of EMD to the problem of nonlinear distortion in hands-free communications. The theory of the EMD and our novel solution to NAEC based on EMD are discussed in Chapter 6.

Part of the work in this chapter has resulted in the following publication:

- L. K. Gudupudi, N. Chatlani, C. Beaugeant, and N. W. D. Evans, "Non-linear acoustic echo cancellation using empirical mode decomposition," in *Proc. ICASSP*, Apr. 2015.

Chapter-7

Since the traditional data analysis methods rely on a priori defined bases for data representation, Fourier-based approaches are ill-suited to the analysis of nonlinear and nonstationary signals; they assume a *linear* superposition of different signal components. As a consequence, the energy of a nonlinear signal is spread across a number of harmonics. nonlinear distortion is then represented traditionally as harmonic distortion, even if the link to a physical source is questionable. Chapter 7 reports our first attempt to apply EMD in combination with Hilbert transform (Hilbert-Huang Transform) to the analysis of nonlinear distortion produced by mobile phone loudspeakers. The results demonstrate that the real nonlinear loudspeakers distortion is more complex. On the basis of the results, we reported an alternative interpretation of loudspeaker nonlinearities through the *cumulative* effects of harmonic content and intra-wave amplitude-and-frequency modulation. This work calls into question the interpretation of nonlinear distortion

through harmonic distortion and points towards a link between physical sources of nonlinearity and amplitude-and-frequency modulation.

Part of the work in this chapter has resulted in the following publication:

- L. K. Gudupudi, N. Chatlani, C. Beaugeant, and N. Evans, "An alternative view of loudspeaker nonlinearities using the Hilbert-huang transform," in *Proc. WASPAA*, Oct. 2015.

Chapter-8

This chapter presents a summary of the findings, the conclusions of the thesis and offers recommendations for further research.

Part I

Nonlinear Systems: Modeling and Characterization

Part 1 begins with a general treatment of nonlinear systems and modeling which is presented in Chapter 2. Chapter 3 discusses the major sources of nonlinearity in the acoustic echopath of hands-free telephones. The downlink path is shown to be more prone to nonlinear distortion, with miniature loudspeakers being a major source. Approaches to the nonlinear modeling of electrodynamic loudspeakers are then presented. Chapter 3 also reviews a well-known nonlinear system identification technique which provides an empirical approach to estimate model parameters. An experimental procedure to identify real mobile phone loudspeakers is discussed later in Chapter 4. The quality of different loudspeaker models is assessed through both objective and subjective tests. Finally, a reliable nonlinear model is thoroughly validated as a function of its key parameters.

Chapter 2

Nonlinear Systems and Modeling

The main problem tackled in this thesis involves nonlinear distortion in the context of acoustic echo cancellation. Before discussing this complex phenomenon, we present an overview of nonlinear systems and related concepts from the literature. We start with the fundamentals of nonlinear systems in Section 2.1, followed by a review of the trends in nonlinear systems modeling and characterization in Section 2.2. Section 2.3 presents the common distortion metrics which are used to measure/quantify nonlinear distortion. The material presented in this chapter also serves as an introduction to loudspeaker modeling which is discussed later in Chapter 3.

2.1 Linear vs. Nonlinear Systems

2.1.1 Linear Systems

A system is a machine that performs a transformation between the instantaneous and past inputs to yield an output. A system is either linear or nonlinear depending on the nature of the transformation. Mathematically we say that the transformation is linear if the system satisfies the following two properties:

1. Homogeneity: $f(\alpha x) = \alpha f(x), \forall \alpha \in \Re$

   A loudspeaker is driven with a pure tone sine wave and the output is exclusively a sine wave of the same frequency, and if the magnitude of the output sine wave is directly proportional to the magnitude of the input sine wave scaled by α, then the loudspeaker is said to be linear. See Fig. 2.1.

2. Superposition: $f(x_1 + x_2) = f(x_1) + f(x_2), \forall x_1, x_2 \in \Re^n$

   The input to a loudspeaker consists of two sine waves at different frequencies. The

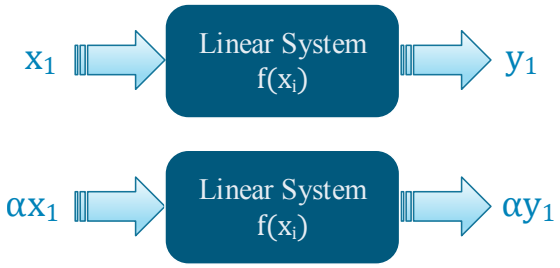


Figure 2.1 – Homogeneity property of linear systems

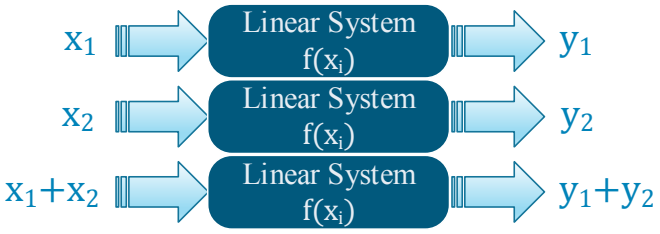


Figure 2.2 – Superposition property of linear systems

loudspeaker output due to each input is independent and the combined output is equal to their sum. See Fig. 2.2.

Both of these properties are necessary for a system to be linear. If the system is linear then it is said to be characterized uniquely by its Impulse Response (IR). The output of such a system is determined by convolving the input with the impulse response. Most signal processing applications are well understood within a uniform theory of discrete linear systems. However, many physical systems exhibit some nonlinear behavior, and in such situations the linear assumption is a poor approximation of the real system.

2.1.2 Nonlinear Systems

Nonlinear systems are those that do not satisfy the above mentioned properties. Alternative nonlinear models and methods have therefore been developed to design and analyze physical systems. Nonlinear systems are those whose outputs are a nonlinear function of their input [19]. Nonlinear systems are rarely well defined, and are an often misunderstood field of signal processing. As we move from linear to nonlinear systems, we shall face a more difficult situation. The superposition and the homogeneity properties no longer hold, and hence we can no longer relate the input and output of the system using simple linear convolution. Nonlinear system analysis tools necessarily involve more advanced and complex mathematics.

2.1.3 Memory Effects

A system (either linear or nonlinear) has memory when its output at any time $t = t_0$ depends not only on the input at that time, but also on past inputs and outputs ($t < t_0$). Equivalently, if a system has memory then such systems are called frequency dependent (or dynamic) systems, that is the system does not have a flat magnitude response[1] . A system is said to be memoryless (or static) if its output at time t_0 depends only on the input at time t_0, such systems exhibit a flat magnitude response.

Note that linearity does not mean the system magnitude response is flat. Linearity or nonlinearity is a property of the system (either dynamic or static), independent of the input signal type. Taking advantage of the superposition principle, linear systems can be modeled in time domain using Auto-Regressive Moving Average (ARMA) models, adaptive digital filters, and the like. On the other hand, nonlinear systems do not obey the principle of superposition, hence their response depends on the input signal amplitude. Amplitude-dependent systems are synonymous with nonlinear systems and are interpreted as dynamic or static nonlinear systems based on with or without frequency dependency respectively. Linear system models are inadequate to represent the complex behavior of nonlinear systems and hence we have to consider nonlinear system models in this case.

2.2 Modeling Nonlinear Systems

In this Section we describe the popular nonlinear models and illustrate the ways in which they can be integrated into the current research in order to gain a better understanding of the complex behavior of nonlinear systems.

2.2.1 Taylor Series

As discussed in the previous section, a linear dynamic system or a linear system with memory can be described by the convolution operation

$$y(n) = \sum_{i=0}^{L-1} h(i)x(n-i) \tag{2.1}$$

where $h(n)$ is the impulse response of the system of size L-taps and, $x(n)$ and $y(n)$ are

[1]The case of all-pass filters are an exception here. All-pass filters can have memory and still exhibits a flat magnitude response

input and output signals respectively. We initiate our discussion of nonlinear system models with the static nonlinear systems. The static nonlinear system or the memoryless nonlinear system is usually presented in the form of a *Taylor series* expansion [20]. The Taylor series is a representation of a function $F[x]$ as an infinite sum of terms that are calculated from the values of the function derivatives at a *single* point, $x = a$ (assuming $F[x]$ has derivatives of every order) [20]:

$$\begin{aligned} y &= F[x] = F_0[x] + F_1[x] + F_2[x] + \cdots + F_p[x] \\ &= f(a) + f^{'}(a)(x-a) + \frac{f^{''}(a)}{2!}(x-a)^2 + \cdots + \frac{f^p(a)}{p!}(x-a)^p + \cdots \\ &= \sum_{p=0}^{\infty} \frac{f^p(a)}{p!}(x-a)^p \end{aligned} \tag{2.2}$$

where $F[]$ denotes any unknown nonlinear functional operator taking input x to output y; $f^p(a)$ is the p^{th}-order derivative of $f(a)$ and $f^0(a)$ means $f(a)$. In Equation 2.2, if $a = 0$ and provided the function $f(x)$ can be differentiable at $x = a = 0$, then the expression is known as the *Maclaurin Series* [20]. In practice a memoryless nonlinear system is presented in more general form of Taylor/Maclaurin series referred to as the *Power series* expansion or the *Polynomial* expansion:

$$y = a_0 + a_1 x + a_2 x^2 + \cdots + a_p x^p + \cdots = \sum_{p=0}^{\infty} a_p [x]^p \tag{2.3}$$

where $a_0, a_1, a_2, \cdots$ are scalar coefficients of the Taylor series[2] . Since in practice we can never use the whole infinite series, we always truncate to a limited order, say P^{th}-order, thereby obtaining an approximate value. Such a truncated system is referred to as P^{th}-order memoryless (or static) nonlinear system. For example, a P^{th}-order memoryless nonlinear system with an input sinusoid of magnitude A, and frequency f will generate at the output a fundamental (first harmonic) as well as higher order harmonics (integer multiples of the fundamental frequency) and a constant component (a_0) which represents the DC offset of the system. This phenomenon is called *Harmonic Distortion*. The

[2]These coefficients, $a_p, p \in [0, \infty]$, should not be confused by "a" of Taylor series expansion in Eq.2.2

magnitude of each output component is dependent on the input signal magnitude A. If the excitation signal has multiple tones then the output contains not only the harmonics of all input tones but also the *subharmonics*, which are the combination tones with frequencies equal to the sum or difference of harmonics. This type of phenomenon is called *Inter-modulation Distortion*.

2.2.2 Volterra Series Expansion

Memoryless nonlinear systems are often considered as the simplest and most commonly implemented form of nonlinear system. Furthermore, complex nonlinear systems are the nonlinear systems with memory or dynamic nonlinear systems. A dynamic nonlinear system is difficult to describe. The most common method for modeling dynamic nonlinear systems is the *Volterra* series. The Volterra series is a generalization of the classical Taylor series expansion which includes a time dispersive element (memory). Since this thesis concerns dynamic systems, the output of the nonlinear system not only depends on the instantaneous input but also on the past inputs. To model such a dynamic nonlinear system, the Volterra series takes the following form [16, 21, 22]:

$$\begin{aligned} y(n) &= F[x(n), x(n-1), x(n-2), \cdots] = F[\mathbf{x}] = F_0[\mathbf{x}] + F_1[\mathbf{x}] + \cdots + F_P[\mathbf{x}] + \cdots \\ &= h_0 + \sum_{i_1=0}^{\infty} h_1(n; i_1)x(i_1) + \sum_{i_1=0}^{\infty}\sum_{i_2=0}^{\infty} h_2(n; i_1, i_2)x(i_1)x(i_2) + \cdots \\ &\quad \cdots + \sum_{i_1=0}^{\infty}\sum_{i_2=0}^{\infty}\cdots\sum_{i_P=0}^{\infty} h_P(n; i_1, i_2, \cdots, i_P)x(i_1)x(i_2)\cdots x(i_P) + \cdots \end{aligned} \tag{2.4}$$

where, $\mathbf{x} = [x(0), x(1), x(2), \cdots]$ is the input signal vector and $h_P(n; i_1, i_2, \cdots, i_P)$ are the coefficients of the P^{th}-order nonlinearity. Equation 2.4 provides the ability to capture the *memory* effect of the nonlinear systems. Again this Volterra series is an infinite series and for the practical purposes we always truncate to a limited order, say P^{th}-order, thereby obtaining an approximate value. Also, the summation limits in Equation 2.4 are limited to finite values, say N_p, which correspond to the memory length of the p^{th}-order nonlinearity. Often, nonlinear systems are considered as time-invariant, where the input-output relation does not change with time. While modeling a nonlinear time-invariant system with memory, the coefficients in the truncated Volterra series expansion only depend on time differences, so the expansion in Equation 2.4 then takes the form:

$$y(n) = h_0 + \sum_{p=1}^{P} \sum_{i_1=0}^{N_p-1} \sum_{i_2=0}^{N_p-1} \cdots \sum_{i_p=0}^{N_p-1} h_p(i_1, i_2, \cdots, i_p) x(n-i_1) \cdots x(n-i_p) \quad (2.5)$$

where $h_p(i_1, i_2, \cdots, i_p)$ are the p^{th}-order *Volterra Kernels*, which approximately characterize the nonlinear system. The constant term h_0 can be safely neglected without any loss of generality [22]. It is worth mentioning that the first order Volterra kernel, $h_1(i_1)$, corresponds to the linear impulse response of the system. The higher order Volterra kernels, $h_p(i_1, i_2, \cdots, i_p), p \in \{2, \cdots, P\}$, are p-dimensional matrices of size N_p, and are usually assumed symmetrical in the indices $i_1, i_2, \cdots, i_p$. For example, the second order Volterra kernel, $h_2(i_1, i_2)$, has two input arguments and can be expressed as a symmetrical function of its arguments:

$$h_2(i_1, i_2) \Leftrightarrow h_2(i_2, i_1)$$

Similarly $h_p(i_1, i_2, \cdots, i_p)$ is left unchanged for any of the possible $p!$ permutations of the p input arguments. By chance, if a nonlinear system has an asymmetric kernel, it can be symmetrized according to the techniques proposed by Wiener in [23]. It is worthwhile to note that, if the Volterra kernels are dirac delta functions ($h_p(i_1, i_2, \cdots, i_p) = a_p \delta(i_1) \cdots \delta(i_p)$) then the Volterra series in Equation 2.5 becomes ordinary power series (shown in Equation 2.3). We can also express the output of a causal P^{th}-order dynamic nonlinear system by the sum of the outputs of all Volterra functionals up to order P:

$$y(n) = \sum_{p=1}^{P} y_p(n) \quad (2.6)$$

By taking the advantage of the symmetric property for a causal nonlinear system, the output of the p^{th}-order Volterra functional, $y_p(n)$, can be approximated without any loss

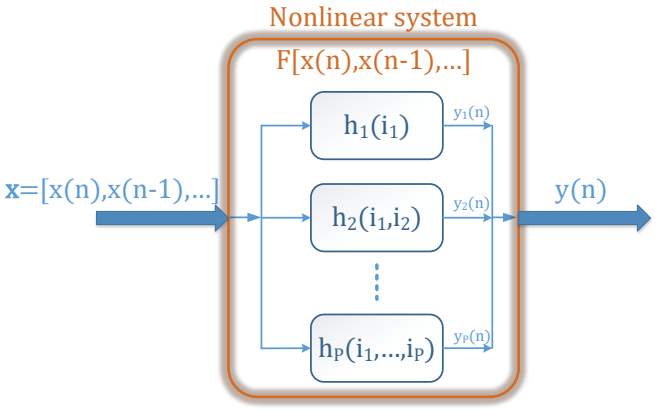


Figure 2.3 – A block diagram representing the P^{th}-order Volterra kernel

of generality by taking the triangular form:

$$y_p(n) = \sum_{i_1=0}^{N_p-1} \sum_{i_2=i_1}^{N_p-1} \cdots \sum_{i_p=i_{p-1}}^{N_p-1} h_p(i_1, i_2, \cdots, i_p) \prod_{k=1}^{p} x(n - i_k) \tag{2.7}$$

The symmetry property leads to a reduction in the number of coefficients required for a Volterra series representation [19]. Equation 2.6 emphasizes the parallel structure of the truncated Volterra series as shown in Fig. 2.3. Note that the output of a nonlinear system modeled by a Volterra series can be interpreted as a linear combination of the inputs to each Volterra kernel. Equivalently, the output of the Volterra nonlinear system is linear with respect to the kernel coefficients. This fact highly simplifies the theoretical performance analysis of dynamic nonlinear systems represented by Volterra series. Hence Volterra series have been successfully employed in a wide variety of applications including nonlinear system identification [8, 19, 24, 25], nonlinear detection and parameter estimation [21], communications [26], adaptive filtering [22], loudspeaker linearization [27, 28] and echo cancellation [4, 29, 30].

2.2.3 Limitations of Volterra Series

Although there are advantages of modeling nonlinear systems with Volterra series, there are a number of limitations. The most common limitations are [19]:

1. The number of coefficients required to model a system determine its computational complexity. Volterra series require many coefficients to model a nonlinear system.

The computational complexity increases exponentially with the increasing order of nonlinearity P even for modest memory length N_p. If $N_p = N, \forall p$ then a P^{th}-order Volterra kernel contains N^P coefficients. Taking account of symmetry, the computational complexity of a P^{th}-order Volterra kernel can be reduced to the combination:

$$\binom{N+P-1}{P} = \frac{(N+P-1)!}{P!(N-1)!} \tag{2.8}$$

In order to limit the computational complexity, the nonlinear system model is often truncated to 2^{nd} or 3^{rd} order. However, the number of coefficients can still pose a problem and 2^{nd} or 3^{rd} order models can only describe the system nonlinearities in a very limited operating range. For example, a truncated 3^{rd}-order nonlinear system with memory length of 50 requires 125000 coefficients without exploiting symmetry property and 19600 coefficients otherwise. These numbers illustrate the computational burden of employing Volterra series to model dynamic Volterra systems and is prohibitive for most practical applications.

2. Consider developing adaptive filtering algorithms using truncated Volterra series for nonlinear system identification applications. We call such adaptive nonlinear filters as *Volterra filters* throughout this thesis. In such applications, if a system is to be identified is "strongly nonlinear", then the truncated Volterra filters are impractical and often diverge. Besides, the Volterra filters involve cross products between the input signal elements (because of memory) which are not mutually orthogonal, even for white Gaussian inputs. This kind of situation is not uncommon and makes the eigenvalue spread of the auto-correlation matrix of the input signal very large. This leads to poor convergence.

3. Another major drawback of the Volterra series expansion is its inability to model nonlinear systems with subharmonics and/or nonlinear systems with discontinuities. The nonlinear phenomenon may take different forms and Volterra series cannot be applicable in all situations. In particular, Volterra series cannot model strong nonlinear systems that generate limit cycles, subharmonics and other perturbations. For example, the signum function ($sign(x)$) which is discontinuous at $x = 0$, and neither Volterra series nor Taylor series expansion exists for such types of nonlinear phenomenon.

2.2.4 Alternative Block-Oriented Models

Based on the limitations it is not surprising that there is always a trade-off between the Volterra model performance and computational complexity. The inclusion of higher order nonlinearity terms without the need for thousands of coefficients necessitates structural changes. Although the Volterra series is impractical for the modeling of strong nonlinear systems it has proven to be successful for the modeling of systems that exhibit weak nonlinearity [19, 31]. The term weak nonlinearity means that the off-diagonal values of the higher order Volterra kernels $(h_2, h_3, \cdots, h_P)$ are weak compared to the significant diagonal values. The nonlinear behavior of such weak dynamic nonlinear systems can be represented by using only the diagonal terms of higher order Volterra kernels (i.e., one dimensional vectors) instead of large multi-dimensional matrices. Such one dimensional Volterra kernels are referred to as simplified (or diagonal) Volterra kernels or higher order impulse responses.

Block-oriented structured approaches are popular nonlinear system modeling techniques that make use of simplified Volterra kernels to describe adequately the nonlinear behavior of the systems over the entire operating conditions. Block-oriented nonlinear models consist of the interaction of dynamic linear and static nonlinear time-invariant systems. In other words, block-oriented models contain multiples blocks connected in cascade form in which the nonlinear system is always memoryless and the linear system is with memory. The simplest and the most popular block-oriented approaches are classified into four nonlinear models based on the interconnection of the linear and nonlinear system blocks. They are the:

- Hammerstein model;
- Wiener model;
- Hammerstein-Wiener model;
- Wiener-Hammerstein model

Hammerstein Model

The Hammerstein model is composed of two blocks, a memoryless nonlinear block and a linear time-invariant block, as shown in Fig. 2.4. The input to the first memoryless nonlinear block is the signal $x(n)$. Its output, $w(n)$, is fed into the second linear block whose impulse response is $h(n)$, which has L-taps. Its output is $y(n)$. Since the nonlinear block is memoryless it can be represented using P^{th}-order power series expansion of

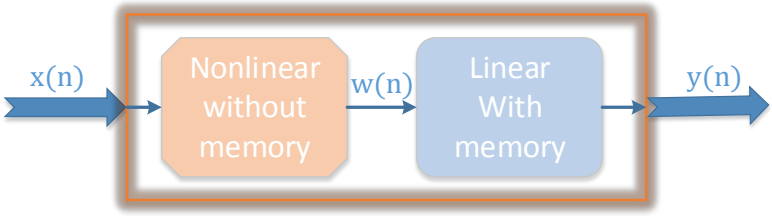


Figure 2.4 – A block diagram of the Hammerstein model

Equation 2.3.

$$w(n) = \sum_{p=1}^{P} a_p [x(n)]^p \tag{2.9}$$

According to Fig. 2.4, the output of the Hammerstein model is therefore given by:

$$\begin{aligned} y(n) &= \sum_{i=1}^{L} h(i) w(n-i) \\ &= \sum_{i=1}^{L} h(i) \sum_{p=1}^{P} a_p [x(n-i)]^p \\ &= \sum_{p=1}^{P} \sum_{i=1}^{L} h_p(i) [x(n-i)]^p \end{aligned} \tag{2.10}$$

where a_p are the scalar coefficients of the power series expansion and $h_p(i) = h(i).a_p$ $\forall p = 1, \cdots, P$ are the simplified (or diagonal) Volterra kernels. Equation 2.10 is popularly known as *Nonlinear Convolution* as proposed in [7]. The Hammerstein model can be seen as a special case of generalized Volterra series expansion with simplified (or diagonal) Volterra kernels in the place of multi-dimensional Volterra kernels. Despite its simplicity, the Hammerstein model is successful in modeling a variety of nonlinear systems such as power amplifiers [32, 33], resonant converters [34] and miniaturized loudspeakers [35].

Wiener Model

The transposition of the linear and nonlinear blocks in the Hammerstein model leads to what is commonly known as the Wiener model, illustrated in Fig. 2.5. In this case, the relation between the input $x(n)$ and the output $y(n)$ by means of an unknown

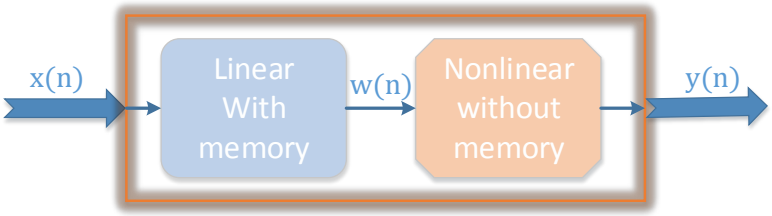


Figure 2.5 – A block diagram of the Wiener model

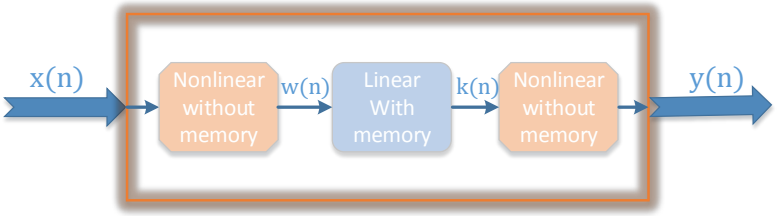


Figure 2.6 – A block diagram of the Hammerstein-Wiener model

intermediate signal $w(n)$ is given by:

$$\begin{aligned} y(n) &= \sum_{p=1}^{P} a_p [w(n)]^p \\ &= \sum_{p=1}^{P} a_p \left[\sum_{i=1}^{L} h(i)x(n-i) \right]^p \\ &= a_1[h(n) * x(n)] + a_2[h(n) * x(n)]^2 + \cdots + a_P[h(n) * x(n)]^P \end{aligned} \tag{2.11}$$

where

$$w(n) = \sum_{i=1}^{L} h(i)x(n-i) \tag{2.12}$$

and where $*$ denotes linear convolution and $h(n)$ is the L-tap impulse response of the linear block.

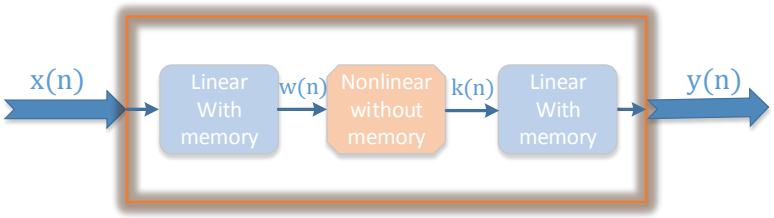


Figure 2.7 – A block diagram of the Wiener-Hammerstein model

Hammerstein-Wiener Model

The Hammerstein-Wiener model is composed of three blocks, where a linear block is sandwiched in between two memoryless nonlinear blocks as illustrated in Fig. 2.6. The first nonlinear block is represented using a $(P_1)^{th}$-order power series expansion with scalar coefficients a_p whereas the second nonlinear block is represented using a $(P_2)^{th}$-order power series expansion with scalar coefficients b_p. The input-output relationship of the Hammerstein-Wiener model is given by:

$$\begin{aligned} y(n) &= \sum_{p_2=1}^{P_2} b_{p_2}[k(n)]^{p_2} \\ &= \sum_{p_2=1}^{P_2} b_{p_2}\left[\sum_{p_1=1}^{P_1}\sum_{i=1}^{L} h_{p_1}(i)[x(n-i)]^{p_1}\right]^{p_2} \end{aligned} \tag{2.13}$$

where $k(n)$ is the output of the Hammerstein model as shown in Equation 2.10 and $h_{p_1}(i) = h(i).a_{p_1}\ \forall p_1 = 1, \cdots, P_1$ are the simplified (or diagonal) Volterra kernels of the first nonlinear block.

Wiener-Hammerstein Model

The Wiener-Hammerstein model, as depicted in Fig. 2.7, is composed of a memoryless nonlinear block sandwiched between two linear blocks. Let the impulse responses of the two linear blocks be $h_1(n)$ and $h_2(n)$ with memory lengths L_1 and L_2 respectively. The

output of the Wiener-Hammerstein model is represented by:

$$\begin{aligned} y(n) &= \sum_{j=1}^{L_2} h_2(j)k(n-j) \\ &= \sum_{j=1}^{L_2} h_2(j) \sum_{p=1}^{P} a_p \left[\sum_{i=1}^{L_1} h_1(i)x(n-i-j) \right]^p \\ &= \sum_{j=1}^{L_2} \sum_{p=1}^{P} h_p(j) \left[\sum_{i=1}^{L_1} h_1(i)x(n-i-j) \right]^p \end{aligned} \tag{2.14}$$

where $k(n)$ is the output of the Wiener model as shown in Equation 2.11 and $h_p(j) = h_2(j).a_p$ $\forall p = 1, \cdots, P$ are the simplified (or diagonal) Volterra kernels of the nonlinear block.

The block-oriented methods have all been employed successfully for characterizing different nonlinear systems in various areas, including signal processing and digital communications. This thesis skims over the foundation of Volterra series in favor of applying them to Nonlinear Acoustic Echo Cancellation (NAEC) problem.

2.2.5 Wiener Series Expansion

This chapter has thus far discussed the Volterra series expansion to represent nonlinear systems for all its simplicity and intuitive appeal. However, there are many limitations of Volterra series, as discussed in Section 2.2.3. Another major disadvantage of Volterra series expansion is its lack of orthogonality in the statistical sense. The output of two different Volterra functionals in Equation 2.6 is usually not orthogonal, therefore their respective output time series are correlated. This non-orthogonality makes the measurement of Volterra kernels very difficult as there is no exact method to separate the individual Volterra kernels [19]. Norbert Wiener resolved this limitation by using an orthogonal form of the Volterra series, called the Wiener series. The Wiener series expansion is another class of polynomial representation of nonlinear systems and is represented in the form:

$$y(n) = \sum_{p=0}^{\infty} G_p[k_p, x(n)] \tag{2.15}$$

where $G_p[k_p, x(n)]$ is a p^{th}-order Wiener operator, given by:

$$G_p[k_p, x(n)] = \left[\sum_{i_1=0}^{\infty} \sum_{i_2=0}^{\infty} \cdots \sum_{i_p=0}^{\infty} k_p(i_1, i_2, \cdots, i_p) \prod_{j=1}^{p} x(n - i_j) \right] + \sum_{m=0}^{int(\frac{p}{2})} \left[\sum_{i_1=0}^{\infty} \sum_{i_2=0}^{\infty} \cdots \sum_{i_p=0}^{\infty} k_{p-2m,p}(i_1, i_2, \cdots, i_{p-2m}) \prod_{j=1}^{p-2m} x(n - i_j) \right] \quad (2.16)$$

where $k_p(i_1, i_2, \cdots, i_p)$ is the p^{th}-order Wiener kernel and $k_{p-2m,p}(i_1, i_2, \cdots, i_{p-2m})$ is a $(p-2m)^{th}$-order Wiener kernel produced by the p^{th}-order Wiener kernel. Unfortunately, Wiener series expansion also requires thousands of coefficients to model a nonlinear system. Even if the Wiener functionals in Equation 2.16 are orthogonal and even if nonlinear systems may be identified more efficiently by a Wiener representation than Volterra representation, Wiener kernels are difficult to interpret. Hence in this thesis different versions of the Volterra series expansion are used to represent nonlinear systems. It is not the purpose of thesis to discuss more details about Wiener series expansion which has been analyzed in detail elsewhere [19].

2.3 Quantifying Nonlinear Distortion

The theoretical concepts of nonlinear systems and different approaches to their mathematical modeling have been discussed in the previous sections. In this section we examine the common distortion metrics to evaluate the signals with the same amount of nonlinear distortion. It is understood from the complex behavior of nonlinear systems that the wave-profile deformation caused by the nonlinear distortion is primarily the result of accumulated harmonic content. Stronger harmonic content leads to more distorted output signal waveform. Therefore it is important to gauge the total effect of the harmonic content.

2.3.1 Total Harmonic Distortion

Total Harmonic Distortion (THD) is a popular metric for measuring the degree of harmonic content in a nonlinear distorted signal. The THD of a given signal y is defined by the following equation:

$$THD = \frac{\sqrt{\sum_{p=2}^{P} y_i^2}}{y_1} * 100\% \quad (2.17)$$

where y_1 is the fundamental component and y_i is the root mean squares (RMS) of successive harmonics in the output signal. THD is generally expressed as a percentage. The higher the percentage, the more the harmonic distortion. The individual harmonic distortion can also be determined individually for each harmonic component p:

$$HD_p = \frac{y_p}{y_1}.100\% \qquad p = 2, 3, \cdots, P \tag{2.18}$$

Unfortunately, the correlation between the THD scores and the subjective evaluation of sound quality of nonlinear distortion is very poor and hence some other alternatives were proposed in [36, 37].

2.3.2 Linear-to-NonLinear-Ratio

Linear-to-NonLinear-Ratio (LNLR) gives the degree of nonlinear distortion in a signal. LNLR is defined as a ratio of the power of linear content to the power of all nonlinear content in the signal. Similar to the segmental signal-to-noise ratio (SNRseg), LNLR is computed over short frames during speech activity, and then averaged as shown below. Just like THD, we have considered two different LNLR's for a given nonlinear signal $x_{out}(n)$:

1. The linear-to-total-nonlinear-ratio ($LNLR_{tot}$), which is computed according to the following expression:

$$LNLR_{tot} = \frac{1}{J}\sum_{i=1}^{J} LNLR_{seg}(i) \tag{2.19}$$

where J is the number of segments and the segmental LNLR, $LNLR_{seg}(i)$, is given by:

$$LNLR_{seg}(i) = 10log_{10}\left(\frac{\sum_{n=0}^{L_s-1} x_{out,1,i}^2(n)}{\sum_{n=0}^{L_s-1} x_{out,p,i}^2(n)}\right) \tag{2.20}$$

where L_s is the length of each segment, which is generally 256 samples for a $8kHz$ sampling frequency signal. $x_{out,1,i}(n)$ and $x_{out,p,i}(n) = x_{out,2,i}(n) + \cdots + x_{out,P,i}(n)$, $p \in [2, P]$ are the linear and the nonlinear components respectively for segment i.

2. The linear-to-individual (p^{th}-order)-nonlinear-ratio ($LNLR_p$) is computed as:

$$LNLR_p = \frac{1}{J}\sum_{i=1}^{J} LNLR_{seg,p}(i) \tag{2.21}$$

where $LNLR_{seg,p}(i)$ is given by:

$$LNLR_{seg,p}(i) = 10log_{10}\left(\frac{\sum_{n=0}^{L_s-1} x_{out,1,i}^2(n)}{\sum_{n=0}^{L_s-1} x_{out,p,i}^2(n)}\right); p \in [2, P] \tag{2.22}$$

Unlike the previous case, here $x_{out,p,i}(n)$ is the individual p^{th}-order ($p \in [2, P]$) nonlinear component. We believe that if two signals have the same $LNLR_{tot}$ and $LNLR_p$ then they have equal amounts of nonlinear distortion. This criteria is especially useful when comparing the empirical and the mathematically modeled nonlinear signals as discussed in later chapters.

2.4 Summary

Most physical systems are nonlinear in their behavior and are vastly more difficult to analyse. For this reason, an ever increasing proportion of modern mathematical research is devoted to the analysis of nonlinear systems. This chapter has been a brief introduction to the theoretical concepts of nonlinear systems and nonlinear modeling. The concepts discussed in this chapter are an essential prerequisite to handle nonlinearities and can be widely applicable in the areas such as engineering, physics and biological systems. As discussed we are particularly interested in applying the underlying nonlinear foundations in the context of AEC.

Next, in Chapter 3, we discuss the sources of nonlinearities in the Loudspeaker Enclosure Microphone System (LEMS). We examine in more detail the topics of loudspeaker modeling and system identification.

Chapter 3

Nonlinear Distortion in a LEMS

This chapter provides valuable insights into nonlinear loudspeaker modeling and system identification. Section 3.1 discusses the general sources of nonlinearities in the acoustic echopath. Section 3.1 argues that miniaturized loudspeakers are prone to nonlinear distortion and hence the theory of loudspeaker nonlinearities is discussed in detail. Next, Section 3.2 introduces the concept of loudspeaker modeling for the purpose of predicting and preventing nonlinear distortion. Finally, in Section 3.3 we review a well-known approach to nonlinear system identification.

3.1 Nonlinear Sources in the Acoustic Echopath

As mentioned earlier, standard approaches to Acoustic Echo Cancellation (AEC) strongly rely on the assumption of linearity everywhere in the Loudspeaker Enclosure Microphone System (LEMS). Any deviation from the linearity assumption leads to significant performance loss [5]. Before discussing the impact of nonlinearity on the performance of AEC, this section discusses the more common sources of nonlinearity in the context of Acoustic Echo Cancellation (AEC). A general approach to acoustic echo cancellation together with the sources of nonlinearity in the acoustic echopath is illustrated in Fig. 3.1. The assumption of linearity in the acoustic echopath is a poor approximate of reality where nonlinearities are introduced in almost each and every block of the LEMS. The latter is divided into three main blocks as shown in Fig. 3.1 the downlink path, the acoustic channel and the uplink path.

- The major sources of nonlinearities in the downlink path are:
  - The Digital-to-Analog Converter (DAC), which converts the incoming far-end signal in the digital domain (binary data) into an analog signal. The analog

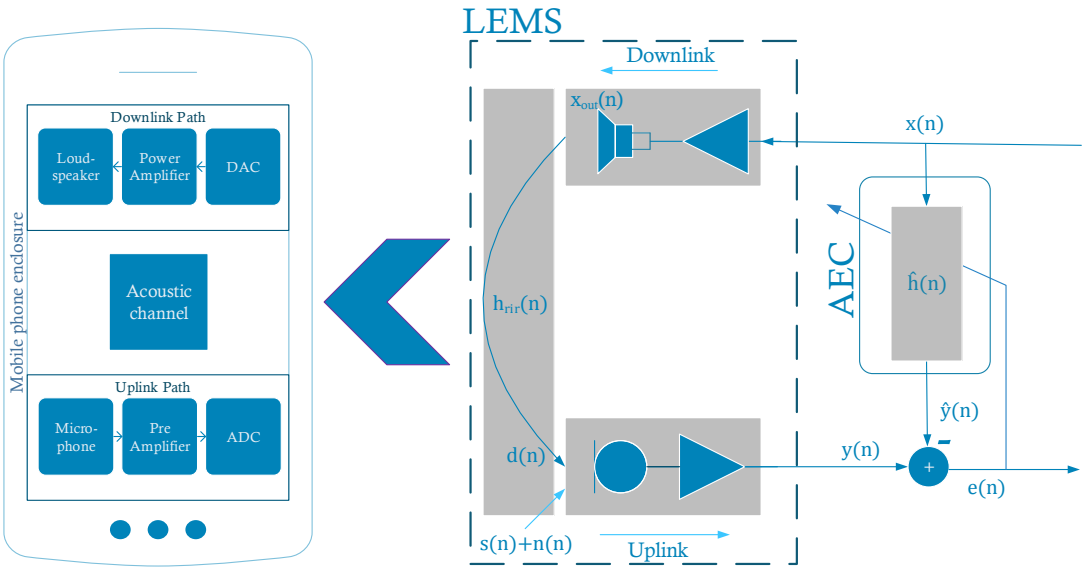


Figure 3.1 – System model illustrating a general approach to acoustic echo cancellation and the nonlinear sources in the LEMS.

signal can either be a current signal or a voltage signal which drives the analog power amplifier.

- The power amplifier, whose input is the low-power audio signal from the DAC, and whose output drives the loudspeaker.
- The loudspeaker is an electro-acoustic transducer, which converts the high-powered electrical audio signal into a corresponding sound signal.

- The acoustic channel or the acoustic space from the output of the loudspeaker to the input of the microphone is usually considered as linear time-invariant and is characterized by a Room Impulse Response (RIR), $h_{rir}(n)$.

- The nonlinear sources in the uplink path are:

  - The microphone is an electro-acoustic transducer, which converts the sound/acoustic signals into corresponding electrical signals.
  - The microphone pre-amplifier, whose input is a weak analog signal from the microphone and whose output amplifies to a desired input level (or line level) of the rest of the circuit.
  - The Analog-to-Digital Converter (ADC), which converts the amplified analog signal from the microphone pre-amplifier to digital data. After treatment by several speech enhancement blocks, this digital signal will be sent to the far-end user.

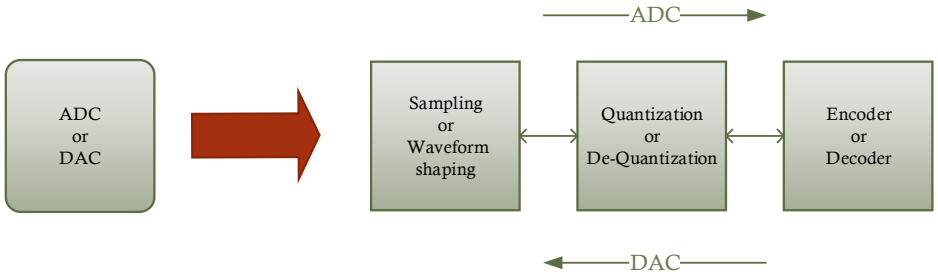


Figure 3.2 – A block diagram representing the process of ADC and DAC operation.

- Besides the ADC/DAC blocks, two transducers and their associated amplifiers, another subtle source of nonlinear distortion in the LEMS is the mobile phone enclosure. The mobile phone enclosure, which is often constructed with plastic or metal, acts as a physical coupling between the loudspeaker and the microphone. These enclosures sometimes cause chaotic rattling and vibration effects. This nonlinear phenomenon is extremely difficult to predict or model and often considered to be uncorrelated noise [3]. Enclosure distortions are not considered in this thesis.

3.1.1 ADC/DAC Distortions

The process of Analog-to-Digital Conversion consists of three main stages as illustrated in Fig. 3.2. The first step in the process is sampling, where the analog (or continuous) signal is sampled at discrete points in time. The time interval between any two successive points is usually constant, and is referred to as the sampling interval. The inverse of the sampling interval is the sampling rate. In the next stage, referred to as quantization, discrete sample values are typically rounded to their nearest integer. The last step in the ADC process is encoding, where typically quantized samples are transferred from integer values to binary codes. The binary representation of the discrete-time signal is referred to as the digital signal. An ideal ADC uniquely maps a large set of analog signals within a certain range to a smaller set of digital binary codes. A typical ADC exhibits many physical imperfections, for example the quantization process involves many irregularities which cause nonlinear distortion referred to as quantization noise [38]. For the ADC/DAC, the nonlinear distortion refers to the input-output functional relationship.

A functional diagram of the Digital-to-Analog Conversion process is also shown in Fig. 3.2. The first stage of the DAC process is the decoder, which takes the binary codes as input and which converts them to corresponding integers. The next stage is the more complicated de-quantization process which converts the discrete-time input signal to an analog (or continuous) signal. Equivalently, this stage creates the electrical signal (either current or voltage) that possess physical dimensions, i.e., amplitude and time [39].

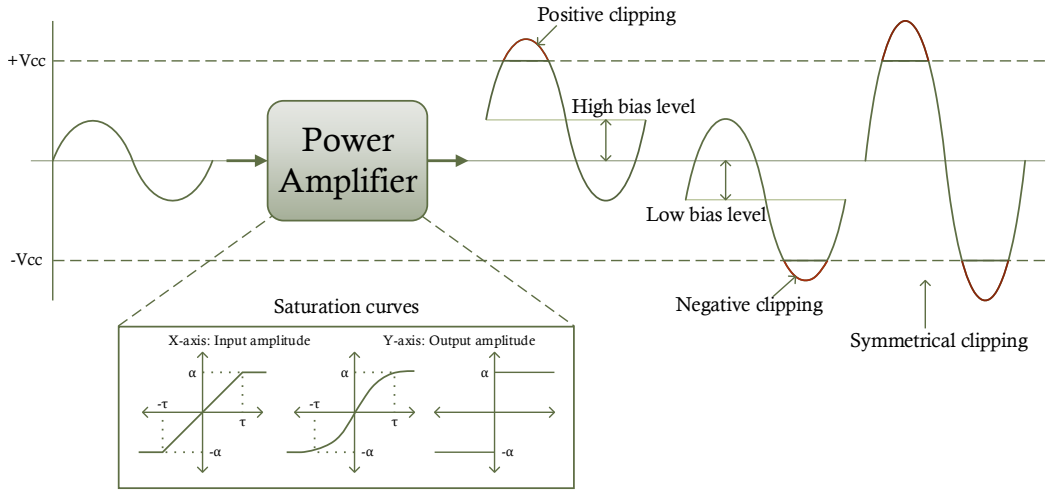


Figure 3.3 – Almost all power amplifiers exhibit some type of saturation effects. The figure shows different types of possible saturation curves for amplifiers.

The last stage of the process is waveform shaping, which uses interpolation filtering techniques to obtain the predetermined shape for the electrical signal. Digital-to-Analog Conversion is more susceptible to nonlinear distortion due to the more sensible de-quantization process. Increased DAC resolution inevitably tends to greater mismatches during fabrication which cause amplitude, pulse shape and timing errors in the DAC output signals thereby causing nonlinear distortion.

Additional nonlinear phenomena caused by the ADC/DAC blocks are discussed in [38, 40, 41]. The nonlinear distortions due to the imperfections in ADC/DAC blocks are generally static (or memoryless) and are modeled using a power series given by Equation 2.3 [39].

3.1.2 Power Amplifier Distortion

Audio power amplifiers are designed to amplify the power (voltage and current) of an input signal to a desirable level sufficient to drive a loudspeaker. The power amplifier uses the DC power from the mobile phone battery to produce the amplified output power. Ideally, the only difference between the input and the output signals of an amplifier is the energy of the signals. However, in reality, to achieve high output power levels with the low DC power of the mobile phone battery, often means that the power amplifier operates close to saturation, thus results in nonlinear distortion. The amplifier output signal then contains additional components that are not present in the input signal. Power amplifiers exhibit various forms of nonlinear distortion. Harmonic and intermodulation distortions are typically the most significant.

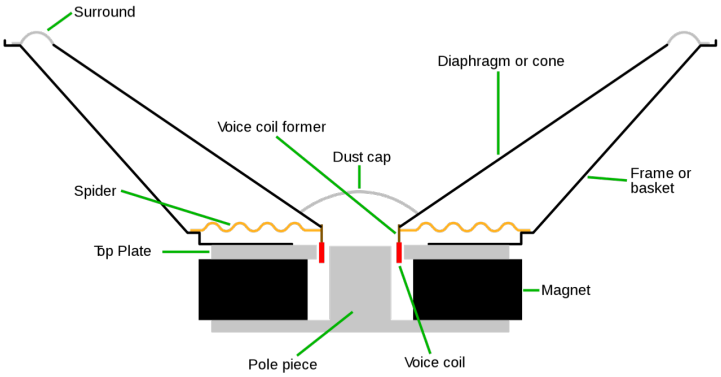


Figure 3.4 – Internal diagram of a electro-dynamic loudspeaker

Harmonic distortion occurs when a power amplifier operates close to saturation. The output signal is then clipped at the maximum capacity of the amplifier, as shown in Fig. 3.3. If the amplifier is wrongly biased this leads to asymmetrical clipping, where one polarity of the signal is clipped and the other remains clean. If there is too much bias then the output waveform exhibits positive clipping whereas low bias leads to negative clipping, as also illustrated in Fig. 3.3. Asymmetrical clipping produces both even and odd order harmonics in the output signal. Even with correct biasing, high input signal levels that can also lead to symmetrical clipping at the output. If the power amplifiers are highly distorted or over-driven by heavy clipping then the output waveform of the input sinusoid resembles a distorted square waveform as shown in Fig. 3.3. Symmetrical clipping creates strong odd order harmonics and weak even order harmonics. Fortunately, harmonic distortion due to saturation effects can be modeled adequately using memoryless Volterra series expansion (Equation 2.3).

3.1.3 Distortions in Loudspeakers

The loudspeaker is a major source of nonlinear distortion. This is because of the ever smaller in size to be sustainable, and operating beyond their natural action. Different types of loudspeakers are available and depend on different operating principles. Since they are most widely used in mobile devices, this thesis focuses on electro-dynamic loudspeakers. Electro-dynamic loudspeakers work on the principle of electro-magnetic induction. The anatomy of a typical electro-dynamic loudspeaker is shown in Fig. 3.4.

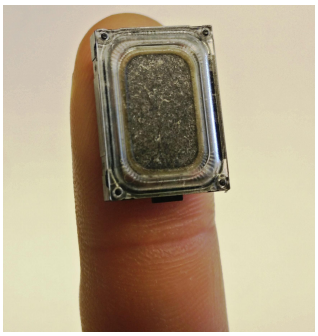
The loudspeaker is composed of a lightweight cone (or diaphragm) which is connected to a rigid frame via flexible suspension. The spider and the surround together make up the

loudspeaker suspension system. The spider, usually made of cotton, allows the moving parts of the loudspeaker to move vertically up and down but not horizontally. The top of the cone is attached to the surround and allows the cone to move freely. The bottom end of the cone is attached to the voice coil, usually made of copper, and is suspended in the circular or cylindrical gap between the poles of permanent magnet. The voice coil acts as an electromagnet when electricity flows through it. The loudspeaker suspension system helps to keep the voice coil centered in the gap and also ensures that the moving cone returns to a neutral position by providing a restoring force. The lower parts of the loudspeaker, including the magnet and the gap, the top plate and the pole piece, are responsible for producing motion and are generally referred to as the motor structure.

The loudspeaker is driven by the power amplifier. When the AC electrical (voltage and/or current) signal from the power amplifier is applied to the voice coil, it is magnetized and acts as an electromagnet upon the creation of a magnetic field around the coil (Lorentz force). The magnetic field intensity and the direction are controlled by the AC electrical signal. The electromagnet and the permanent magnet interact with each other as would any two magnets. The input AC signal in the voice coil causes the polarity of the electromagnet to change with respect to the frequency of the AC signal. This change in polarity of the electromagnet repels with the polarity of the permanent magnet, and pushes the voice coil back and forth rapidly. Since the narrow end of the cone is attached to the voice coil, the cone also moves in and out in accordance with the voice coil. This diaphragm excursion mechanism creates pressure waves in the air in front of the loudspeaker. The human ears perceive these as sound waves. The amplitude and the frequency of the input AC electrical signal dictate the moving rate and distance of the voice coil movement, which in turn determines the amplitude and the frequency of the sound waves produced by the cone (or diaphragm).

The basic operating principle of an electro-dynamic loudspeaker has not changed since its invention. Even miniaturized loudspeakers work in the same way. The quality of the loudspeaker depends on its frequency response. Humans can hear audio between $20Hz$ and $20,000Hz$. In order for a loudspeaker to produce a $50Hz$ sine tone (50cycles) the cone diaphragm must move in and out 50 times per second. On the one hand, if the diameter of the cone is large then it can be too heavy to vibrate quickly enough to produce high frequencies. Large loudspeakers are thus better suitable to the production of low frequencies. On the other hand, if the cone is small and light, it is harder to vibrate slowly enough to produce very low frequencies, at which the cone will deform. Thus the loudspeaker design (diameter of the cone, size of coil, etc.) determines its frequency response and dictates quality. The design of loudspeakers reproducing the entire audible frequency range was the greatest challenge in the loudspeaker. High quality loudspeakers like the one shown in Fig. 3.5a will typically use multiple speaker units with different

(a) Studio loudspeaker



(b) Microspeaker (or miniaturized loudspeaker) used in mobile phones

Figure 3.5 – Different sizes of loudspeakers

diaphragm sizes, each optimized to work in a subband of the overall audio frequency range. In such cases, a separate *crossover* electronic circuit is necessary in order to split the incoming audio signal into multiple bands. Such high quality loudspeakers are capable of reproducing the full sound spectrum and exhibit a near flat frequency response. Furthermore, the larger the loudspeaker the easier it is to design a high quality suspension system that can handle any setbacks and remain linear.

Since the beginning of this decade, the mobile device market (cellular phones, smartphones, tablets, laptops, etc.) has been the fastest growing category of any technology in the world [42]. Coupled with the rising demand, the drive towards miniaturization and convergence has led to the use of ever-smaller transducers. A miniaturized loudspeaker or a microspeaker is illustrated in Fig.3.5b. Researchers continue to advance the technology by making loudspeakers smaller, efficient and durable, however with limited success. Unfortunately, today's microspeakers are still incapable of producing better sound spectrum, especially at low frequencies. As loudspeakers get thinner, the diaphragm excursion becomes smaller, which explains microspeakers inefficiency in producing low frequency sounds (having longer wavelengths) as they need more space to push the volume of air. Hence the frequency response of a typical microspeaker appears like a high-pass filter response with cut-off frequencies close to $1000Hz$. This explains the

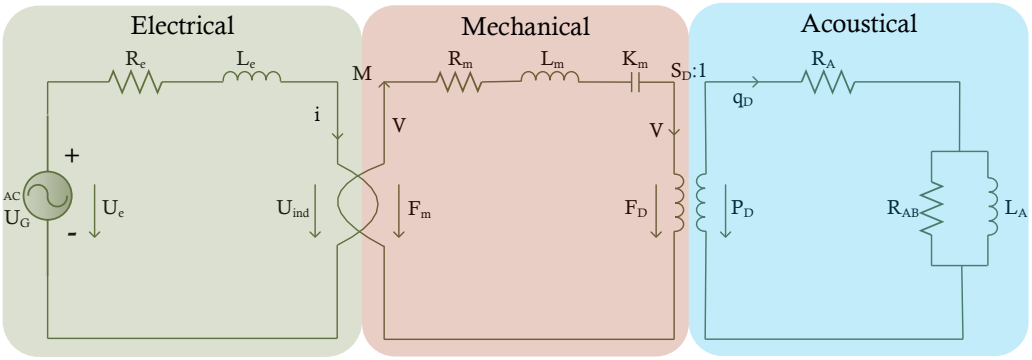


Figure 3.6 – The complete equivalent electrical circuit of the electro-dynamic loudspeaker

linear distortion phenomenon introduced by microspeakers.

The more critical problem entails nonlinear distortion. In order to understand the sources of nonlinear distortion, let us consider the complete equivalent electrical circuit of a typical electro-dynamic loudspeaker, illustrated in Fig.3.6. In this thesis, the terms microspeaker and loudspeaker are interchangeable.

The electrical domain consists of the voice coil representation with a DC resistance (R_e) and a self-inductance (L_e). U_{ind} is the voltage induced in the electrical domain by the mechanical domain, $U_{ind} = Blv = Bl\frac{dx}{dt}$, where B is the magnetic flux density in the permanent magnet air gap, l is the length of the voice coil conductor, v is the voice coil velocity and x is the voice coil displacement. The product Bl is called the force factor. The mechanical domain consists of the suspension system and the cone representation, where R_m indicates the total mechanical resistance due to the suspension system and the dissipation in the air load, L_m indicates the total mass of the voice coil, the cone and the air load and K_m indicates the stiffness of the suspension system which is given by the reciprocal of its compliance (C_m), $K_m = \frac{1}{C_m}$. The Lorentz force generated when the voice coil is traversed by an AC current i is given by $F_m = Bli$. The electrical and the mechanical parts interact through the magnetic field. The acoustic domain consists of the loudspeaker enclosure effects representation and the circuit design varies depending on the type of the loudspeaker enclosure, e.g., vented cabinet or closed cabinet. Referring to the electrical and mechanical domains circuits shown in Fig.3.6, when all the parameters are assumed to be linear, the coupling between the two domains leads to a system of two

linear differential equations:

$$\begin{aligned} U_e &= R_e i + L_e \frac{di}{dt} + Bl \frac{dx}{dt} \\ F_m &= Bli = L_m \frac{d^2x}{dt} + R_m \frac{dx}{dt} + K_m x \end{aligned} \tag{3.1}$$

where U_e is the voltage at the terminals as shown in Fig.3.6. In reality, it is well known that a microspeaker is a highly nonlinear system, hence Equation 3.1 does not hold. There are many causes of the nonlinear distortions in the microspeakers, starting from the operating principle of the loudspeaker itself and that extends to the main components of the loudspeaker [9]. The most regular nonlinearities are classified into four groups [9, 38]:

- Electrical nonlinearities: caused by the voice coil inductance
- Magnetic nonlinearities: caused by the permanent magnet
- Mechanical nonlinearities: caused by the suspension system and the diaphragm
- Acoustical nonlinearities: caused by the nonlinear propagation of the sound waves

All these type of nonlinearities should take into account in consideration of nonlinear distortion in microspeakers. Nonlinear distortions in the transducers are well studied by the researchers over the last decade [9, 36, 37, 38, 43, 44]. The following summarizes the main results: Most significant nonlinearities caused by the microspeaker depend on the displacement of the voice coil and/or the diaphragm.

Electrical Nonlinearities

Electrical nonlinearities arise mainly due to the nonlinear behavior of the voice coil inductance (L_e). Unlike in Equation 3.1, the self-inductance of the coil is not constant but varies with respect to the voice coil position, x_{po}. If the voice coil moves away from the air gap of permanent magnet then the self-inductance is lower and vice versa. This is due to the fact that, as the voice coil moves away from the magnet, its magnetic resistance increases and hence the current in the voice coil produces less magnetic field causing lower self-inductance. The same magnitude of current produces more magnetic field if the voice coil is inside the air gap of permanent magnet.

In addition to that, the self-inductance also varies with respect to the input AC current in the voice coil. As discussed above, the AC current in the voice coil produces a magnetic

field with flux density $B = \mu(i)H$, where H is the magnetic field strength and μ is the permeability. As the input current increases, the B also increases in proportionate with the H until the B reaches certain threshold value where it cannot increase anymore becoming constant as the H continues to increase. This nonlinear relationship between B and H is called magnetic saturation. If the relationship between B and H is nonlinear then the AC current creates hysteresis loop there by increasing inductance at higher frequencies [9]. Further, the eddy currents induced because of the changes in the B react with the current in the coil and cause the decrease of the inductance [43]. This non-uniformity of the voice coil inductance depicts nonlinear distortion.

According to Klippel in [9], the two curves representing the self-inductance of the voice coil versus the displacement and self-inductance versus current are highly likely asymmetrical for most of the loudspeakers, causing asymmetrical nonlinear distortion. As discussed in the previous section, asymmetrical nonlinear distortion deforms the output signal wave-profile by introducing strong odd-order harmonics and weak even-order harmonics. This type of nonlinear distortion can be modeled by power series expansion, given by Equation 2.3 [9].

However, Klippel states in [9] that in most cases, the electrical nonlinearities only has a minor influence.

Magnetic Nonlinearities

When the loudspeaker is driven with a constant current, the force on the voice coil ($F_v = Bl(x_{po})$) is not constant and is depends on its position (x_{po}). If the voice coil moves into the air gap of the permanent magnet then the magnetic flux density (B) increases and vice-versa. This non-uniform magnetic flux density affects the driving force on the voice coil causing nonlinear distortion. Since the length of the voice coil (l) is constant, the nonlinear function between the B and the F_v is static and hence they can be modeled as power series expansion [9].

Mechanical Nonlinearities

Loudspeakers use a suspension system, comprises of a spider and a surround to center the voice coil in the air gap of the permanent magnet. The suspension behaves like a normal spring and may be characterized by the force-displacement curve, which often show some hysteresis. This is because of the nonlinear stiffness (K_m) of the spider which is not constant but is a function of voice coil displacement (x_{po}) [9]. The geometry of the suspension system may also cause a significant asymmetry in the stiffness characteristic,

which leads to output waveform deformation. Since the nonlinear stiffness is also a function of voice coil displacement, one can safely model the approximate value of K_m using power series expansion. However, the exact modeling of the nonlinear stiffness is very complex as a lot of other parameters like the temperature and loudspeaker ageing may significantly influence the stiffness parameter.

Acoustical Nonlinearities

Acoustical nonlinearities arise due to the nonlinear wave propagation. The *Doppler effect* is the most dominant type of acoustical nonlinearities. As we know the diaphragm of the loudspeaker has to move relatively larger distances at low frequencies compared to high frequency signals. If a loudspeaker is simultaneously radiating both a low and a high frequency then the time taken by a low frequency signal is relatively more time to reach a fixed on-axis listening point compared to a high frequency signal. The difference in the time arrival is proportional to the diaphragm excursion difference for low and high frequency signals. Equivalently, this effect can be interpreted as high frequency signals are frequency modulated with respect to the low frequency signals. This type of acoustical nonlinearities are generally significant in horn loudspeakers and are potentially weak in electro-dynamic loudspeakers. Hence, acoustical nonlinearities are not considered in this thesis.

If the electric, magnetic, mechanic and acoustic nonlinearities are in-phase then one can use a single joint nonlinear system (either with memory or memoryless) to model these nonlinearities. If these nonlinearities are not in-phase then it is the worst case scenario, which makes the loudspeaker a more complex nonlinear system in the downlink path of the LEMS. In this thesis, we assume that the nonlinear distortion in the acoustic echopath is solely originated from the downlink path of the LEMS and the multiple sources of nonlinearities in the downlink path are totally in-phase.

3.2 Nonlinear Loudspeaker Modeling

Accurate and comprehensive modeling of a nonlinear loudspeaker is a very challenging task. Many attempts have been made during the last two decades to accurately model and identify the nonlinear loudspeakers. However, conventional loudspeaker models are often inadequate to represent nonlinear behavior over a wide range of audio frequencies, and/or at large amplitudes. In fact, nonlinear loudspeaker modeling and system identification are themselves major and distinct areas of research [35, 45].

During the early stages of research, dynamic nonlinear loudspeaker models were derived

from the first-principles of the lumped-parameter model of the loudspeaker [35]. The equivalent lumped element circuit diagram of loudspeaker and associated model equations are shown in Chapter 2, Fig.3.6 and Equations 3.1 respectively. The parameters like the inductance (L_e), the force factor (Bl) and the stiffness (K_m) are not constants as in the linear loudspeaker model, but are functions of the voice coil position, x_{po}. Such models derived from the first-principles are called *white-box* models; they reflect actual loudspeaker physics. However, white-box models are limited to linear and/or lower order nonlinear loudspeaker behavior.

On the other hand, *black-box* models are mathematical models that use measured input/output data to develop loudspeaker models without any physical insight. Such black-box models can be useful for simulation and prediction or for the design of loudspeaker systems. Black-box modeling typically involves a model structure which describes and/or emulates nonlinear system behavior. With a black-box model, we have a time-domain input/output mathematical relationship of the following type:

$$x_{out}(n) = f(x(n)) \tag{3.2}$$

where x and x_{out} are the input and output of the loudspeaker respectively and $f(.)$ represents a nonlinear function described by a model structure that maps $x(n)$ to $x_{out}(n)$. Hence, model structure selection is critical and yet there is no standard approach to its design. The model structure can be designed according to our knowledge of nonlinear distortion source. The most common nonlinear distortion for miniaturized loudspeakers is harmonic distortion [46, 47, 48, 49]. This makes the application of Volterra series expansions particularly well suited to the modeling of nonlinear loudspeakers. Volterra series expansions are discussed in Chapter 2.

Ambiguity over Memory

The modeling of nonlinear electrodynamic loudspeakers with Volterra series was first proposed in the early 1960s [50]. From the late 1980s, there has been a continuous effort in the application of Volterra series to obtain a comprehensive and precise nonlinear loudspeaker model [51]. The well-known (truncated) Volterra series expansion describing the modeling of nonlinear loudspeaker is given in Equation 3.3:

$$x_{out}(n) = \sum_{p=1}^{P} \sum_{i_1=0}^{N_p-1} \sum_{i_2=0}^{N_p-1} \cdots \sum_{i_p=0}^{N_p-1} h_p(i_1, i_2, \cdots, i_p) x(n-i_1) \cdots x(n-i_p) \tag{3.3}$$

where $h_1(i_1)$ is the linear impulse response, and where $h_p(i_1, i_2, \cdots, i_p)$ are the p^{th}-order *Volterra Kernels*, which represent the p^{th}-order nonlinearity with memory length N_p. Although the interpretation of the Volterra series is straightforward, the measurement of the Volterra kernels is extremely difficult. Few attempts reported in the literature were only moderately successful due to complexity and the over-parametrization problem [52].

Coming to scenario of miniaturized loudspeakers typically used in mobile devices, there is an unresolved ambiguity in the literature over frequency dependent (or memory) behavior of the nonlinearities. Some researchers considered such miniaturized loudspeakers as complex nonlinear systems and hence looked at frequency dependent nonlinear behavior, thereby modeling loudspeakers using Equation 3.3 [9, 12, 30, 53].

On the other hand, assuming the nonlinear distortion in the miniaturized loudspeakers is mostly due to saturation (or clipping), researchers concluded such nonlinear distortion is rather weak and frequency independent [46, 47, 48, 49]. Hence they can be modeled as memoryless Volterra series (or power series) expansions:

$$x_{out}(n) = \sum_{p=1}^{P} a_p [x(n)]^p \tag{3.4}$$

where a_p are scalar coefficients. However, this representation makes sense only when the frequency response of the loudspeaker is flat throughout the audible frequency band. In contrast, our experiments with the real mobile phones loudspeakers re-confirmed their inability to produce accurate sounds, especially at low frequencies (More details are given in the next Sections). Hence, miniaturized loudspeakers have a frequency-dependent response causing what is known as linear distortion.

Consequently, a practical alternative loudspeaker model that takes in to account both the memoryless nonlinear distortion and linear distortion with memory is the Hammerstein model, as discussed in Chapter 2. The input-output relationship of the Hammerstein

model reads:

$$x_{out}(n) = \sum_{p=1}^{P} \sum_{i=0}^{L-1} h_p(i)[x(n-i)]^p \tag{3.5}$$

where $h_p(i)$ are the simplified (or diagonal) Volterra kernels of the loudspeaker. If these simplified Volterra kernels of a loudspeaker are identified, then it is possible to reconstruct the loudspeaker's output for any given input signal $x(n)$.

Due to the different forms in which nonlinear characteristics occur in real loudspeakers we also have a diversity of nonlinear model structures accessible in the literature [35, 45]. We concentrate primarily on the above discussed Volterra series based models in this thesis. For the sake of computational simplicity, we assume that the loudspeaker is a weak or memoryless nonlinear system. Next Section focuses on the Hammerstein model system identification from measured input-output data of a real mobile phone loudspeaker.

3.3 Loudspeaker System Identification

In this Section, loudspeaker system identification process is discussed by evaluating the unknown simplified Volterra kernels ($\mathbf{h}_p$ for $p \in [1, P]$) in the Hammerstein model shown in Equation 3.5. The Hammerstein model is also referred to as the generalized polynomial Hammerstein model (GPHM) can be viewed as a structure illustrated in Fig. 3.7. The model is made up of P parallel branches, with each branch consisting of a p^{th} power static nonlinear function followed by a linear filter $\mathbf{h}_p = [h_p(0), h_p(1), \cdots, h_p(L-1)]^T$ for $p \in [1, P]$ of length L taps. Given both the input and output signals $x(n)$ and $x_{out}(n)$, the problem of system identification consists then of estimating the unknown linear filters $\mathbf{h}_p$, $p \in [1, P]$. $\mathbf{h}_1$ can be treated as linear Impulse Response (IR) and $\mathbf{h}_p$, $p \in [2, P]$ are (the so-called) higher-order impulse responses of the loudspeaker under test respectively.

For this purpose of loudspeaker system identification, there are a wide variety of identification techniques and excitation signals available in the literature. Several system identifications techniques are discussed in [45]. Further in this thesis we adapted a straight forward identification procedure called "nonlinear convolution technique" as first proposed in [7, 17]. The block diagram illustrating the procedure of the nonlinear convolution technique is shown in Fig. 3.8. The routine of the nonlinear convolution technique is:

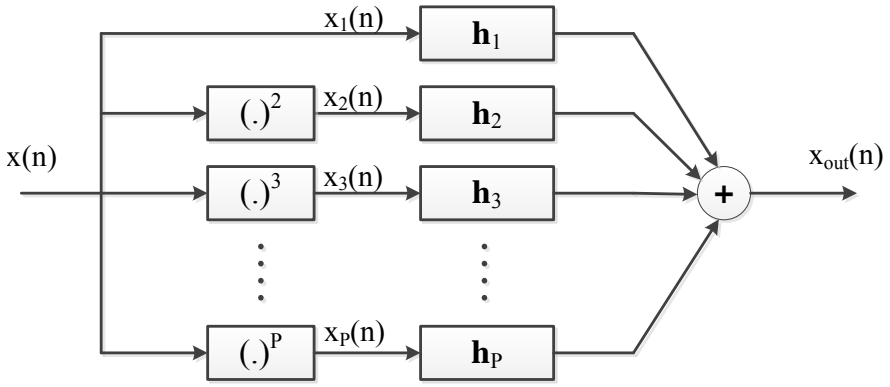


Figure 3.7 – The generalized polynomial Hammerstein model.

- First, an exponential sine-sweep signal and its inverse filter are generated.

- Use sine-sweep signal as an excitation signal to the loudspeaker under test and record its distorted output signal.

- The convolution between the distorted output signal and the inverse filter results a "full" IR, which is a set of impulse responses $\mathbf{g}_p$, $p \in [1, P]$ corresponding to the nonlinear distortion of the loudspeaker. (An example of "full" IR is shown in Fig. 3.13b)

- The measured impulse responses $\mathbf{g}_p$, $p \in [1, P]$ can be easily separated and post processed to estimate the desired simplified (diagonal) Volterra kernels $\mathbf{h}_p$, $p \in [1, P]$

The following sub-sections describe in detail each of the four blocks shown in Fig. 3.8.

3.3.1 Excitation Signal Generation

As the loudspeaker model identification is carried out from the input-output signals, the choice of excitation signal plays a key role on the quality of the identification. As discussed, the so called nonlinear convolution technique employs an exponential sine-sweep (or a chirp) signal as an excitation signal. An exponential sine-sweep (or a chirp) signal is a sinusoidal signal of length T with exponentially varying frequency, ranging from f_1 to f_2 is generated with the following analytical expression [17]:

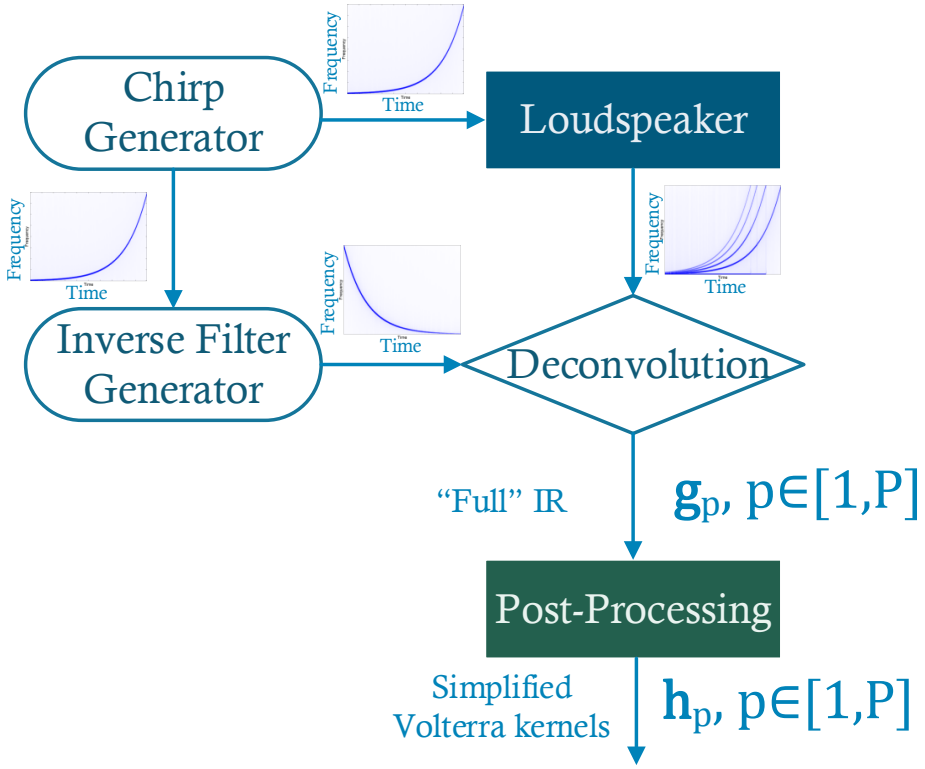


Figure 3.8 – Block diagram representing the process of nonlinear convolution technique.

$$\begin{aligned} x(n) &= a(n)sin[\Phi(n)] \\ &= a(n)sin\left[2\pi f_1.\frac{T}{ln\left(\frac{f_2}{f_1}\right)}.\left(e^{\frac{n}{T}.ln\left(\frac{f_2}{f_1}\right)}-1\right)\right] \end{aligned} \tag{3.6}$$

where $a(n)$ and $\Phi(n)$ are the amplitude and phase of an exponential sine-sweep signal respectively. The time-domain representation of an exponential sine-sweep signal is shown in Fig. 3.9. The spectrogram and the magnitude response of an exponential sine-sweep signal are shown in Fig. 3.10a and Fig. 3.10b respectively. These plots are generated in Matlab covering the frequency range between $f_1 = 20Hz$ and $f_2 = 20kHz$ in 10 seconds duration and is sampled at a frequency of $F_s = 48kHz$. As shown in the figure, an exponential sine-sweep signal does not have a flat spectrum, but the magnitude decreases by $3dB$ per octave. As the frequency of the sine-sweep signal increases exponentially with time, the time duration during which the signal oscillates at a particular frequency decreases. As the time duration decreases, the area under the signal also decreases, hence

the frequency response plot indicates a drop of $3dB$ in magnitude per octave.

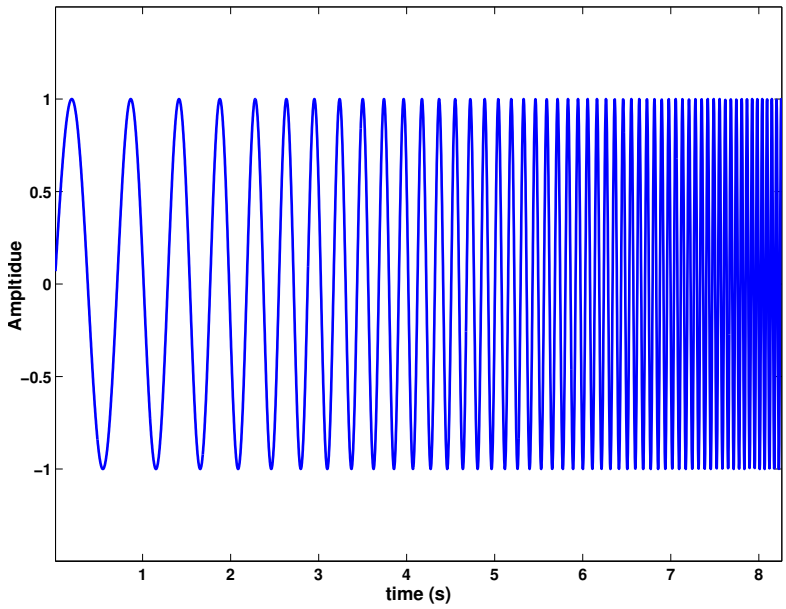


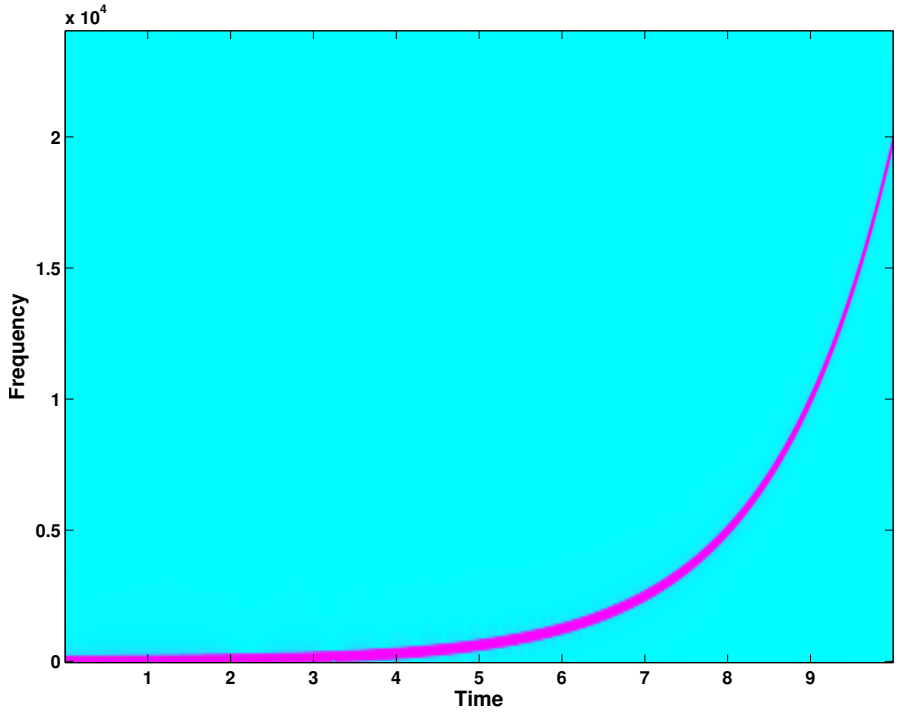
Figure 3.9 – Exponential sine-sweep signal in the time domain

3.3.2 Inverse Filter Generation

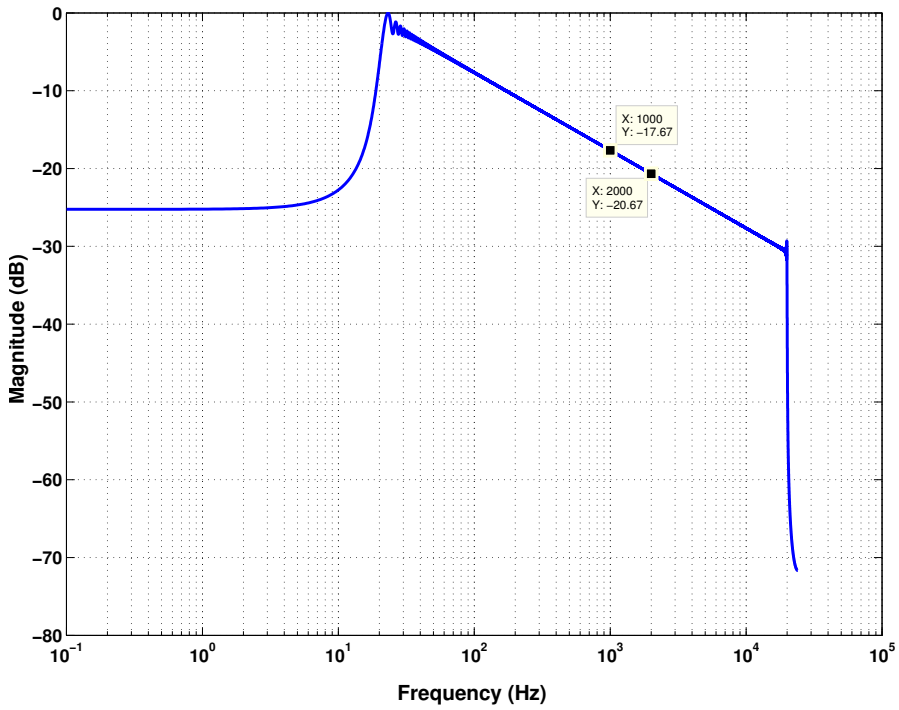
One of the advantages of employing an exponential sine-sweep signal as an excitation signal is its ease of generating an inverse filter for deconvolution. The inverse filter or the inverse sine-sweep signal, $x_{inv}(n)$, is simply the time-reversal of the input sine-sweep signal ($x(n)$), as shown in Fig. 3.11. However, the time reversed sine-sweep signal also exhibits a drop of $3dB$ per octave in its magnitude response, as shown in Fig. 3.11 and this would not result in a much needed flat spectrum after deconvolution with the sweep signal. To overcome this issue, an amplitude modulation of envelope $6dB$ per octave is applied to the time reversed exponential sine-sweep signal such that the convolution between the $x(n)$ and $x_{inv}(n)$ gives a Dirac delta function :

$$x_{inv}(n) = x(N_s - n - 1) * exp(-\frac{n}{ln(\frac{f_2}{f_1})}.T) \tag{3.7}$$

where $N_s = F_s * T$ is the length of the sweep signal. Fig. 3.12 shows the time domain representation of the inverse sine-sweep signal after amplitude modulation along with its magnitude response. This plot clearly demonstrates the uprising $3dB$ per octave spectrum.



(a) Spectrogram of an exponential sine-sweep signal



(b) Magnitude response of an exponential sine-sweep signal

Figure 3.10 – The sweep signal covering the frequency range between $f_1 = 20Hz$ and $f_2 = 20kHz$ in 10 seconds duration and is sampled at a frequency of $48kHz$

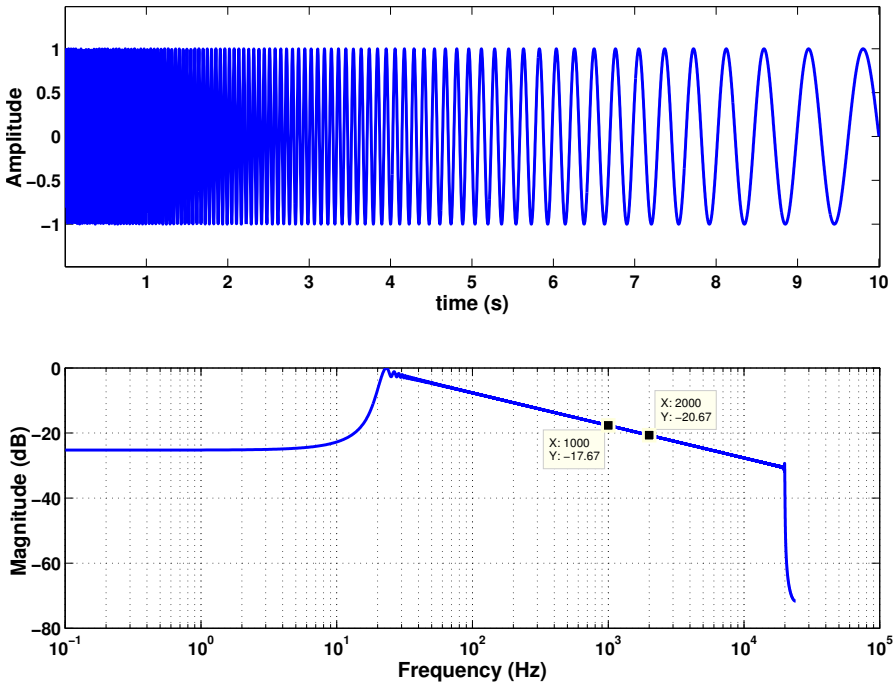


Figure 3.11 – Time reversed exponential sine-sweep signal (top) and its Magnitude response (bottom) indicating a drop of $3dB$ per octave

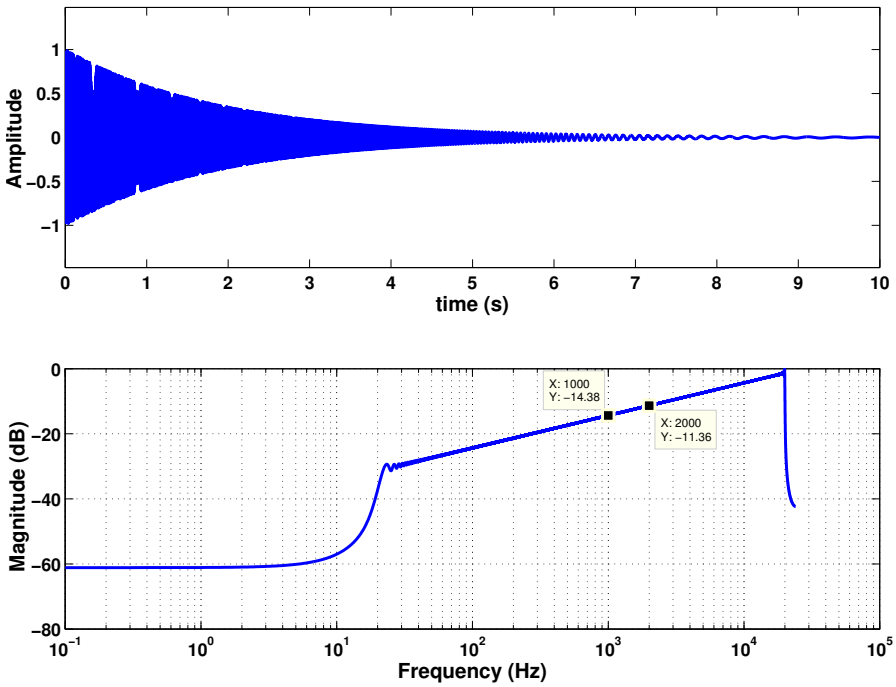


Figure 3.12 – Inverse filter used in nonlinear convolution technique (top) and its Magnitude response (bottom) indicating an uprising of $3dB$ per octave

3.3.3 Deconvolution: "Full" IR Measurement

The major advantage of employing an exponential sine-sweep signal as an excitation signal resides in the appearance of nonlinear distortion artifacts. Fig. 3.13a illustrates the spectrogram of a nonlinearly distorted sine-sweep signal, the right most curve is the fundamental sweep preceded by its 2^{nd}, 3^{rd} and 4^{th}-order harmonics respectively. In general IR measurement techniques, it is relatively difficult to separate the linear IR from the nonlinear distortion artifacts [54]. The nonlinear convolution technique overcomes such limitations. Deconvolution of a distorted sine-sweep signal with an inverse filter results a "full" IR, which is a combination of the linear IR and the higher-order IR's as shown in Fig. 3.13b. The rightmost IR is the linear IR, which is preceded by the 2^{nd}-order IR and so on. Hence, with a single experiment based on this technique, a near perfect linear IR is measured without the effect of nonlinearities while nonlinearities of various orders can be identified simultaneously.

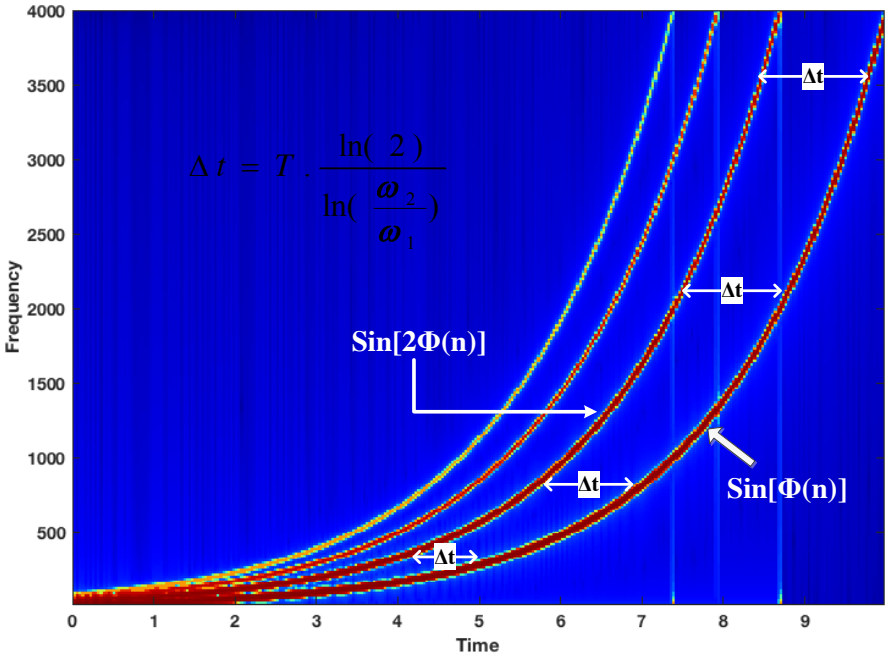
The "full" IR of the loudspeaker under test can be computed according to Equation 3.8, given by:

$$\tilde{x}(n) * x_{inv}(n) = \sum_{p=1}^{P} g_p(n + \triangle t_p) \tag{3.8}$$

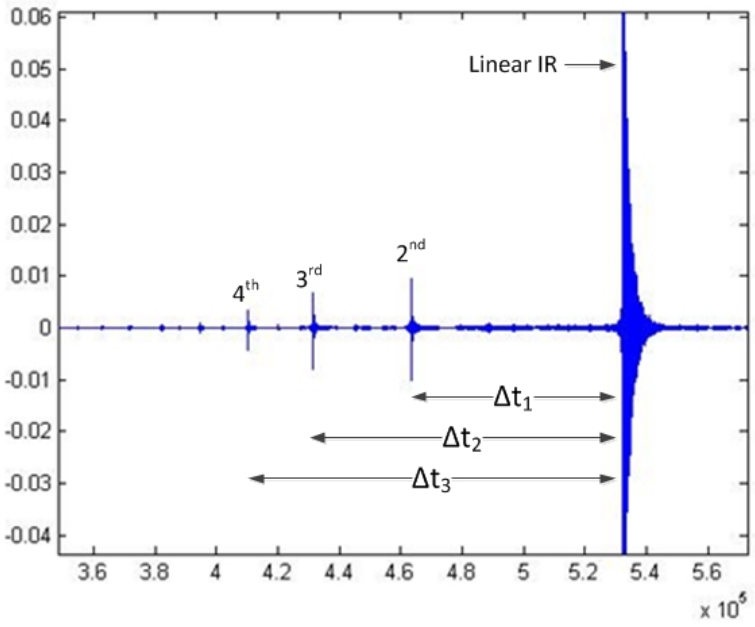
where $\tilde{x}(n)$ and $x_{inv}(n)$ are the distorted exponential sine-sweep signal from the loudspeaker and the inverse filter respectively. $\mathbf{g}_1$ is the linear IR and $\mathbf{g}_p = [g_p(0), g_p(1), \cdots, g_p(L-1)]^T, p \in [2, P]$ are the higher-order IR's occurred in the "full" IR because of any nonlinearities in the loudspeaker are accumulated with the time lag Δt_p from the linear IR:

$$\Delta t_p = T.\frac{ln(p)}{ln(\frac{f_2}{f_1})} \tag{3.9}$$

where $p \in [2, P]$ is the harmonic distortion order. For example, Δt_2 and Δt_3 indicate the time lag from the linear IR to the 2^{nd} and 3^{rd}-order IR's respectively. This time lag is an important property of the exponential sine-sweep signal and can be understood from Fig. 3.10. At one instance of frequency, say $1000Hz$ (see Fig. 3.10), the harmonics have lesser group delay than the fundamental i.e., the time taken by the fundamental to reach an instantaneous frequency (say 1000Hz again) is high compared to its harmonics. This time difference is constant irrespective of the frequency and is given by Δt_p [17].



(a) Spectrogram of a nonlinearly distorted exponential sine-sweep signal



(b) An example of a "full" IR showing the accumulation of higher-order IR's because of nonlinear distortion

Figure 3.13 – A "full" IR is resulted upon deconvolution of a nonlinearly distorted sine-sweep signal with an inverse filter.

It is important to note that, if the system under test is linear, for example a high quality loudspeaker, then $\tilde{x}(n) * x_{inv}(n)$ in Equation 3.8 results in the linear IR of the system and no nonlinear artefacts or distortion can be noticed in the "full" IR.

3.3.4 Computation of the Simplified Volterra kernels

The measured higher-order IR's ($\mathbf{g}_p = [g_p(0), g_p(1), \cdots, g_p(L-1)]^T, p \in [1, P]$) computed using Equation 3.8 are not directly the expected simplified (or diagonal) Volterra kernels ($\mathbf{h}_p, p \in [1, P]$) in Equation 3.5. For a given exponential sine-sweep excitation signal with amplitude $a(n) = 1$, the output of a loudspeaker can be emulated using the measured higher-order IR's, $\mathbf{g}_p, p \in [1, P]$:

$$\mathbf{x}_{out} = \sum_{p=1}^{P} \mathbf{g}_p * sin(p\Phi) \tag{3.10}$$

where $sin(\Phi)$ is an exponential sine-sweep signal generated using Equation 3.6 and $*$ represents linear convolution. However, the response of the polynomial Hammerstein model to the same exponential sine-sweep in terms of simplified Volterra kernels $\mathbf{h}_p = [h_p(0), h_p(1), \cdots, h_p(L-1)]^T, p \in [1, P]$, is given by:

$$\mathbf{x}_{out} = \sum_{p=1}^{P} \mathbf{h}_p * sin^p(\Phi) \tag{3.11}$$

The relation between Equations 3.10 and 3.11 is discussed in detail in [7, 8] which also describes a procedure to compute the simplified Volterra kernels $\mathbf{h}_p, p \in [1, P]$ from the measured higher-order IR's $\mathbf{g}_p, p \in [1, P]$.

Here we discuss the solution for a specific case of calculating the first 5 simplified Volterra kernels as discussed in [7]. The trigonometric formulas for a sine function can be written as:

$$
\begin{aligned}
sin^2(\omega n) &= \frac{1}{2} - \frac{1}{2}cos(2\omega n) \\
sin^3(\omega n) &= \frac{3}{4}sin(\omega n) - \frac{1}{4}sin(3\omega n) \\
sin^4(\omega n) &= \frac{3}{8} - \frac{1}{2}cos(2\omega n) + \frac{1}{8}cos(4\omega n) \\
sin^5(\omega n) &= \frac{5}{8}sin(\omega n) - \frac{5}{16}sin(3\omega n) + \frac{1}{16}sin(5\omega n)
\end{aligned} \tag{3.12}
$$

Substituting Equation 3.12 in Equation 3.11 for $P = 5$ and then transforming Equations 3.10 and 3.11 into frequency domain gives the following solution [7]:

$$
\begin{aligned}
H_1 &= G_1 + 3G_3 + 5G_5 \\
H_2 &= 2jG_2 + 8jG_4 \\
H_3 &= -4G_3 - 20G_5 \\
H_4 &= -8jG_4 \\
H_5 &= 16G_5
\end{aligned} \tag{3.13}
$$

where, H_p and G_p for $p \in [1, 5]$ are the Fourier transforms of $\mathbf{h}_p$ and $\mathbf{g}_p$ respectively.

After computing the simplified Volterra kernels from the measured "full" IR, the nonlinear convolution can be efficiently implemented using Equation 3.5 for mathematically reconstructing the loudspeaker nonlinear distortion. This nonlinear system identification technique has proven to be fast and robust approach for the emulation of nonlinear distorting systems [7]. Several extensions to the standard procedure of nonlinear convolution technique have been recently developed and interested readers can find additional details in [45, 55].

3.4 Summary

In this chapter, we have focused on nonlinear distortion in a LEMS, loudspeaker modeling and system identification. Sources of nonlinear distortion within a miniaturized loudspeaker have been discussed in detail. We also presented three nonlinear models

suitable for loudspeaker modeling: Volterra series (with memory), power series and generalized polynomial Hammerstein (GPHM) models. Assuming nonlinear signal distortion generated by the miniaturized loudspeakers as weak, we chose to ignore the nonlinear memory effects in this thesis. Hence, in the next chapters, we are going to use either power series or GPHM models for loudspeaker modeling. Unlike power series model, GPHM model considers the frequency dependent (memory) linear IR while modeling loudspeaker. Comparison between these two models and the limitations of potentially applying each model to the loudspeaker modeling are discussed in Chapter 4.

In addition, we have discussed a well-known approach to nonlinear (loudspeaker) system identification, referred to as nonlinear convolution, first proposed in [7, 17]. The method uses an exponential sine-sweep signal as an excitation signal and allows the simultaneous identification of both linear and higher-order impulse responses (called as "full" IR in this thesis) of a nonlinear system. The simplified or diagonal Volterra kernels, computed from the measured "full" IR, can be used for mathematically reconstructing the loudspeaker nonlinear distortion. The material in this chapter will be helpful in identifying a real mobile phone loudspeaker, as discussed in the next chapter.

Chapter 4

Comparative Studies of Simulated and Real-Device Experiments

This chapter aims to present comparative studies of real-device and synthesized loudspeaker signals. Section 4.1 discusses the experimental design used for the identification of a real mobile phone loudspeaker. Once the loudspeaker is identified, real-speech test signals are recorded using the same loudspeaker to compare the outputs of both the empirically estimated nonlinearities and the different nonlinear models.

Section 4.2 presents the objectively evaluated comparative results, which show that nonlinear distortion estimated with the GPHM better reflects that measured empirically. This work was published in [56].

Finally, Section 4.3 reports the validation of the GPHM model and the corresponding identification technique as a function of its key parameters. This later work was published in [57].

4.1 Identification of a Real Mobile Phone Loudspeaker

This section reports the application of nonlinear convolution technique to identify the simplified Volterra kernels of a real mobile phone loudspeaker.

4.1.1 Experimental Setup

The experimental setup used for the identification of mobile phone loudspeakers is illustrated in Fig. 4.1. A mobile device is placed before a head and torso mannequin at a distance of 32cm. The device is configured to operate in hands-free mode and at maximum volume for which nonlinear distortion is assured. A Personal Computer (PC) is used to store and record all audio data sent to, or received from a mobile device via

a high-quality external sound card and a network simulator [58]. As shown in Fig. 4.1, an excitation signal is played by the PC, transmitted through the network simulator to the mobile phone and then played by the mobile phone loudspeaker. The loudspeaker output is then simultaneously recorded with an independent, high-quality microphone mounted in the ear of a mannequin. The mannequin is connected to the PC via a high-quality external sound card and a software called ACQUA (Advanced Communication QUality Analysis) controls the signal flow between the mannequin and the PC [59]. Some additional non-intrusive tests confirmed that the nonlinear distortions are specifically introduced by the mobile phone and that all other elements in the acquisition chain are purely linear processing.

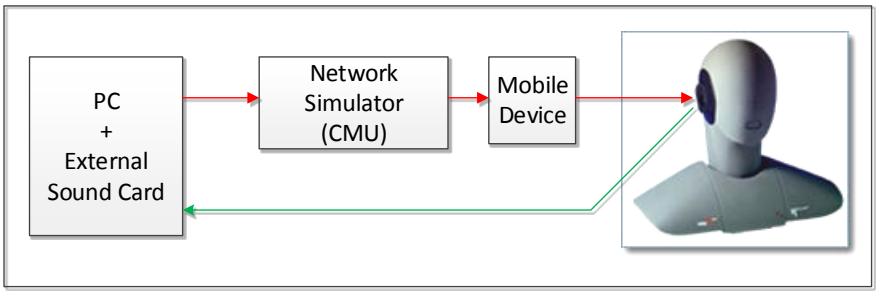
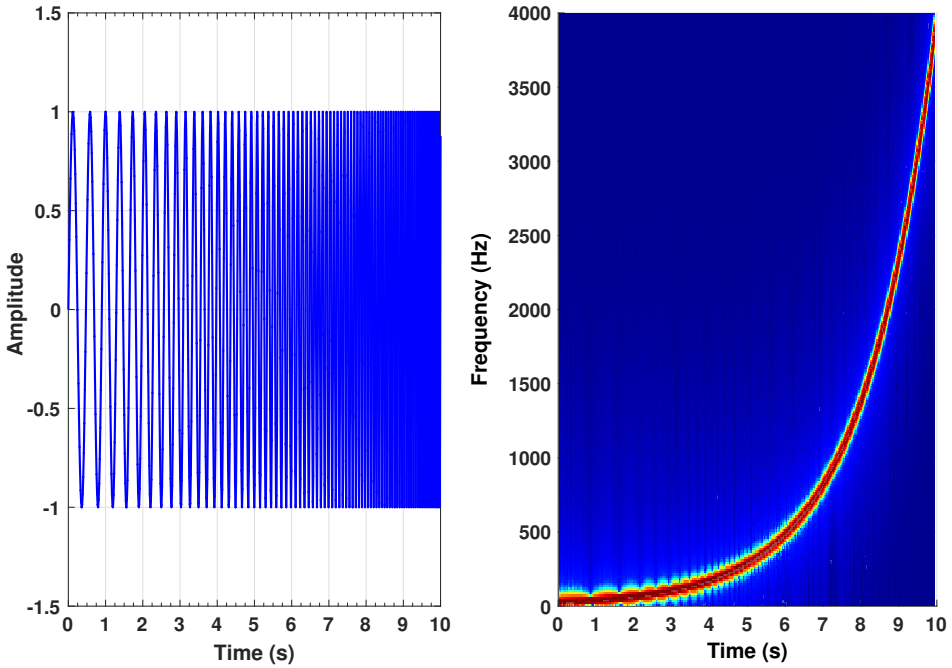


Figure 4.1 – Experimental setup used for the identification of a real mobile phone loudspeaker

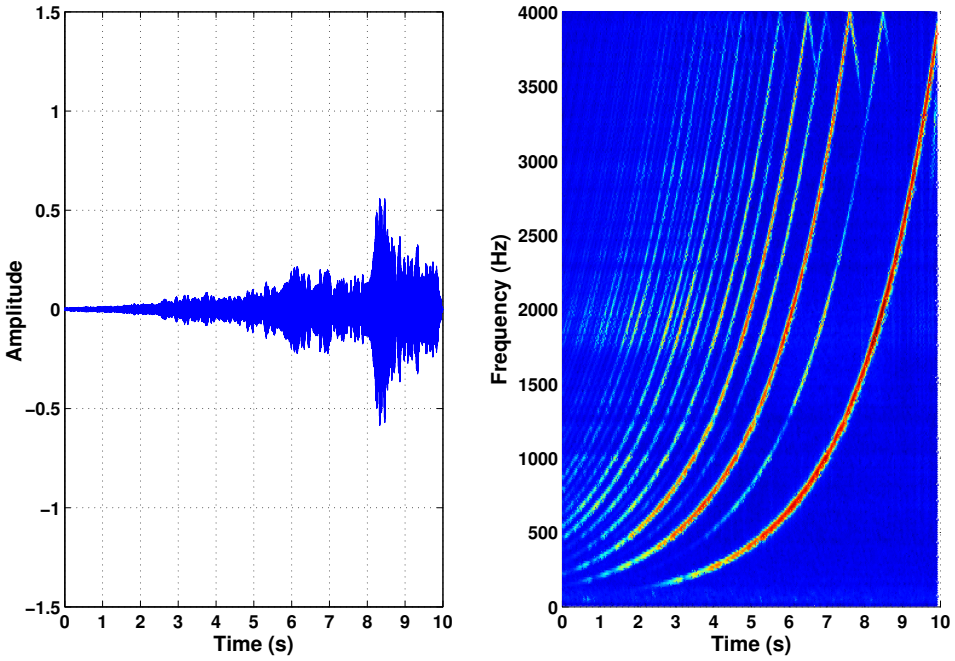
4.1.2 Data Acquisition

The measurements are performed using an exponential sine-sweep signal generated using Equation 3.6, with amplitude $a(n) = 1$, frequencies $f_1 = 20Hz$ and $f_2 = 4kHz$. The sine-sweep signal is $10s$ in duration and is sampled at a frequency of $8kHz$ as illustrated in Fig. 4.2a. In accordance with the sine-sweep signal, the inverse filter or the inverse sine-sweep signal is also generated using Equation 3.7. The sweep signal is played by the mobile phone loudspeaker, using the procedure discussed in Section 4.1.1, and recorded with the microphone mounted in the ear of a mannequin. The spectrograms of the input signal and the recorded signal are shown in Fig. 4.2.

Fig. 4.2b demonstrates the additional harmonics produced by the mobile phone loudspeaker and it is clear that the nonlinearities are dominant until 5^{th}-order. Further, examining Fig. 4.2b the spectrogram indicates the presence of strong odd order harmonics and weak even order harmonics, implying that the nonlinearity is asymmetric.



(a) Time domain and spectrogram representation input exponential sine-sweep signal



(b) Time domain and spectrogram representation of the mobile phone loudspeaker response

Figure 4.2 – Representation of the input and the output signals of a mobile phone loudspeaker

4.1.3 Deconvolution

The deconvolution process is realized in spectral domain by linear convolution[1] of the loudspeaker recorded signal with the inverse filter. The deconvolution leads to the "full" IR of the loudspeaker, a sequence of IR's clearly separated along the time axis as shown in Fig. 4.3. The time lag between the linear IR and the p^{th}-order IR can be computed using the Equation 3.9 and selectively separate each IR of length L from the "full" IR. The linear and the 3^{rd}-order IR are shown in Fig. 4.4. The separation between the IR's depends on the duration (T) of the sweep signal, if T is not sufficiently long that would result in tightly packed (inseparable) higher order IR's.

Further insight into nonlinear distortion of the loudspeaker may be gained by looking at the frequency responses of the IR's, illustrated in Fig. 4.5a. The frequency response of the linear IR is clearly not flat but frequency dependent, therefore the memory effect must be taken into consideration while modeling the linear response of the loudspeaker. On one hand the frequency response curves of the odd order (3^{rd} and 5^{th}-order) harmonic IR's appear similar to the one of the linear IR with only a change in the magnitude level, confirming memoryless nonlinear phenomenon. On the other hand the frequency response curves of even order (2^{nd} and 4^{th}-order) harmonic IR's exhibit a little variation at certain frequencies, indicating memory effect. However, considering their relative weak magnitude with respect to the odd order harmonics, one can safely ignore the memory effect.

4.1.4 Equalization

The recordings described above were collected in a non-anechoic acoustic booth. While reverberation is low, recordings reflect both loudspeaker behavior and room acoustic effects. Therefore, the measured higher-order IR's $\mathbf{g}_p$, $p \in [1, P]$ are thus equalized in order to suppress the influence of the room impulse response (RIR):

$$\mathbf{g}_{eq,p} = \mathbf{h}_{eq} * \mathbf{g}_p;\ p \in [1, P] \tag{4.1}$$

where $\mathbf{h}_{eq} = [h_{eq}(0), h_{eq}(1), \cdots, h_{eq}(L_{eq} - 1)]^T$ is an RIR equalization filter of length L_{eq}. It is estimated by inverting the RIR of the acoustic booth. The assumed linear IR of the acoustic booth was measured using a similar procedure to that described in Sections 3.3.3

[1] Multiplying an input signal and an impulse response in frequency domain implies circular convolution in time domain. In order to make it linear convolution, the signal has to be padded with sufficient zeros before multiplication in frequency domain.

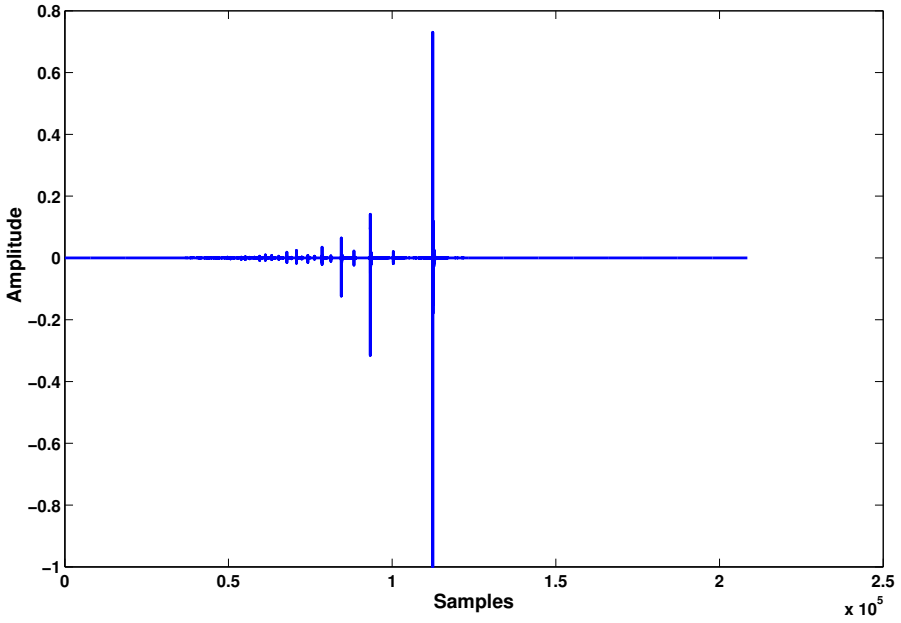


Figure 4.3 – "Full" IR of a real mobile phone loudspeaker

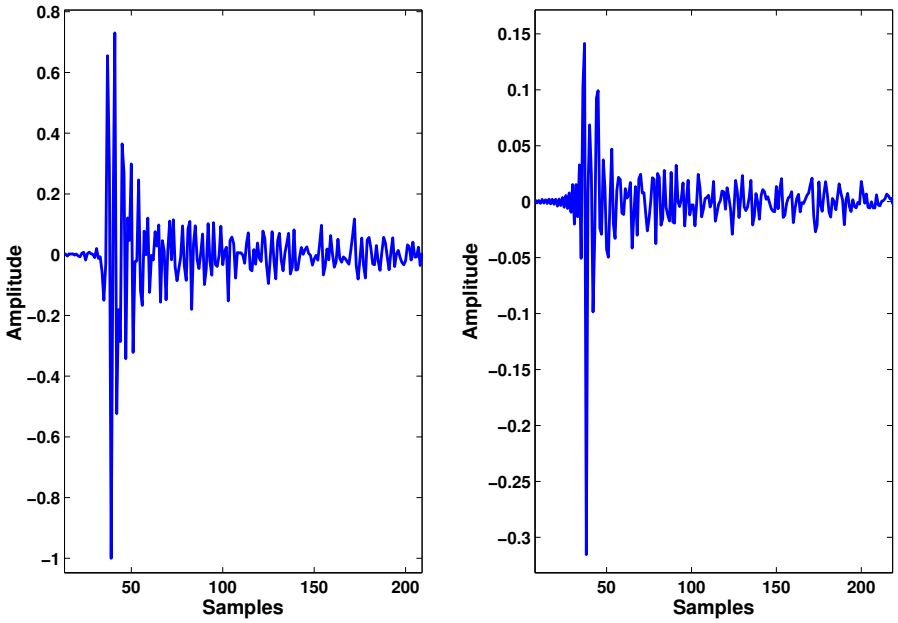


Figure 4.4 – Time domain representation of the linear IR (left) and the 3^{rd}-order IR of a mobile phone loudspeaker

and 4.1.1 . Here though, the mobile device is replaced with a high quality loudspeaker with a flat frequency response in the region of interest (audio frequency range). An equalization filter h_{eq}, which inverts the IR of the acoustic booth, was designed according to the approach described in [60]. Fig. 4.5 illustrates the frequency response curves

of the loudspeaker IR's before and after equalization. The acoustic booth where the experiments were conducted has a very low reverberation hence the impact of RIR equalization has minimal effect on the frequency response curves. Acoustically the low frequencies in the room/acoustic booth are the most difficult to absorb and hence the frequency response at low frequencies gets a boost [61]. Therefore, after equalization most noticeable change can be observed in the low frequency region of the curves. Note that the influence of room effect on the frequency response curves has been treated by considering only the magnitude response of the RIR but the phase response is beyond the reach of our Equalization.

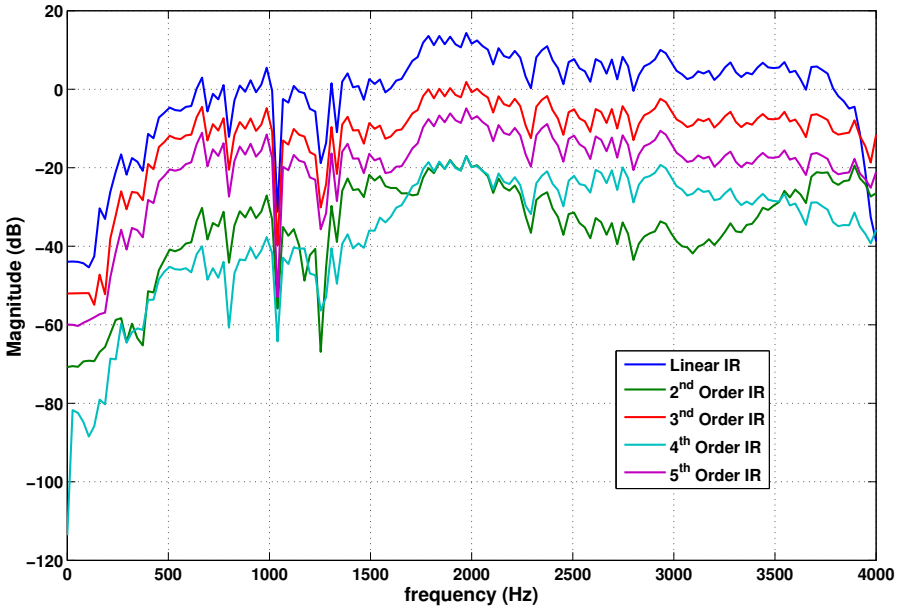
The simplified Volterra kernels $\mathbf{h}_p$, $p \in [1, P]$ of a mobile phone loudspeaker can be computed as a linear combination of the equalized higher-order IR's $\mathbf{g}_{eq,p}$, $p \in [1, P]$ using the method described in Section 3.3.4. Given the simplified Volterra kernels of a mobile phone loudspeaker, its output can be synthesized according to the GPHM using Equation 3.5. The choice of filter length L and order of nonlinearity P involves a trade-off between the accuracy and the computational complexity of the loudspeaker model.

4.2 A Comparison of Loudspeaker Models

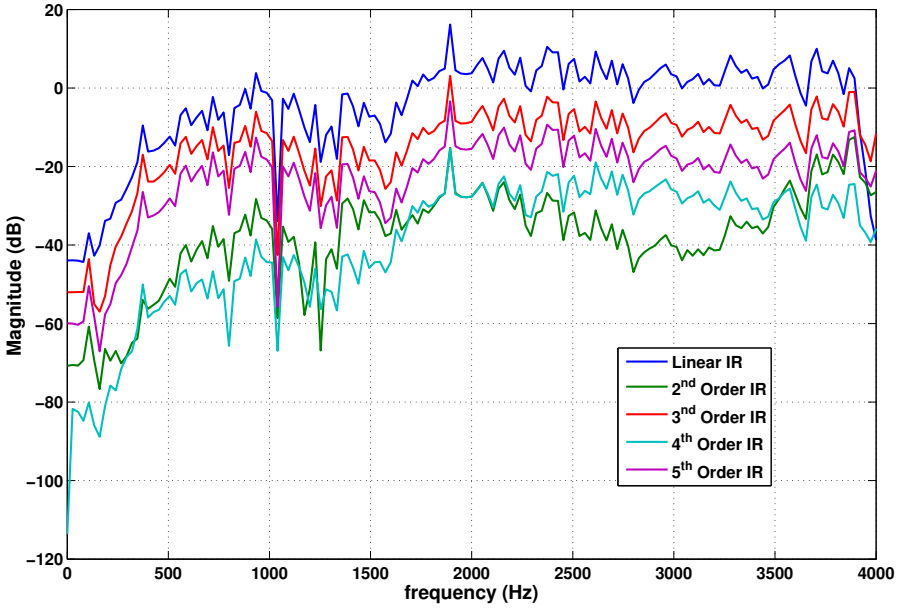
Most research in NAEC assumes loudspeaker as memoryless nonlinear system as reported in later chapters [46, 47, 48, 49]. We have discussed Volterra series based memoryless nonlinear models in the previous chapters. Modeling loudspeaker nonlinear distortion with lower complexity power series (or power filter or polynomial series) approach is today the most popular. Any research in NAEC depends on the accuracy of the loudspeaker model, be it used for NAEC itself, or to artificially synthesize nonlinear test signals. While the power series approach typically delivers efficient NAEC performance in well-controlled simulations, even slight model inaccuracies tend to degrade performance in real conditions. The generalized polynomial Hammerstein model has thus been investigated as an alternative model. A question now arises regarding the model accuracy: which model better reflects the real loudspeaker nonlinear distortion? This section investigates the suitability of modeling nonlinear loudspeaker distortion with power series approach. Also, the accuracy of the two models are compared in estimating the empirically measured, real loudspeaker outputs. The results are published in our first paper [56].

4.2.1 Synthetic Signal Generation

In the following we report the aspects of the synthetic loudspeaker output signal generation using the PSM and GPHM models. We compare loudspeaker output signals $x_{out}(n)$ synthesized from clean speech input signals $x(n)$ accordingly to Equations. 3.4 and 3.5



(a) Frequency responses of the linear IR and the higher order IR's of a mobile phone loudspeaker



(b) Frequency responses of a loudspeaker IR's after RIR equalization

Figure 4.5 – Figure shows the variation of the frequency response curves of a mobile phone loudspeaker IR's before and after RIR equalization

for the PSM and GPHM respectively to real empirically measured signals $x_{real}(n)$. These are obtained from the same clean speech signals played by the mobile device loudspeaker and subsequently recorded at the mannequin ear using the experimental set-up described in Section 4.1.1. This signal is similarly equalized according to $\mathbf{h}_{eq}$ (see Section 4.1.4) to remove room-effects.

For the GPHM, we used the simplified diagonal Volterra kernels $\mathbf{h}_p$, $p \in [1, P]$ empirically measured from a mobile phone loudspeaker as described in the previous section for the signal generation. We have considered the kernels of order $P = 5$ each of length $L = 256$ taps, as illustrated in Fig. 4.2b they are more dominating than the other higher order nonlinearities. Consequently, the GPHM structure has 5 parallel branches as in Fig. 3.7 and using a clean speech signal as input $x(n)$, the output signal has been generated according to Eq. 3.5.

For the PSM, we set the gain $a_1 = 1$. For the comparison purpose, we also used 5^{th} order PSM here. Weighting components a_p for $p \in [2, 5]$ are chosen such that the mean linear-echo-to-total-nonlinear-echo ratio ($LNLR_{tot}$) and the mean linear-echo-to-p^{th}-order-nonlinear-echo ratio ($LNLR_p$) (as discussed in Section 2.3.2) are the same as those of the GPHM. Once we get the weighting components $a_p, p \in [1, 5]$, using the same clean speech signal as input $x(n)$, the output signal has been generated according to Eq. 3.4.

4.2.2 Assessment

The speech signal recorded at the ear of the mannequin ($x_{real}(n)$) was compared to the results obtained according to the two models. The spectrograms of the input clean speech signal, a real mobile phone loudspeaker response, and the two synthesized signals are illustrated in Fig. 4.6. It is obvious from the figure that the signal synthesized with the GPHM is more identical to the real recorded (or measured) speech signal. The power series model assumes a flat frequency response which a loudspeaker linear IR does not have, that explains the difference in the distortion mechanism compared to the real recorded signal. Not surprisingly, the signal synthesized with the PSM has more energy at the low frequencies like the original clean speech signal which are actually not present in the real recorded signal. The real recorded signal has more energy in the high frequency region ($\sim\geq$ 1500) due to nonlinear distortion which the signal synthesized using the GPHM better reflects compared to PSM.

Further, these observations are absolutely correlated with the PESQ[2] scores illustrated in Fig. 4.7. In Fig. 4.7(a) the PESQ scores of the PSM and the GPHM synthesized

[2]Perceptual Evaluation of Speech Quality (PESQ) is the ITU-T P.862 standard objective metric to measure speech quality [62].

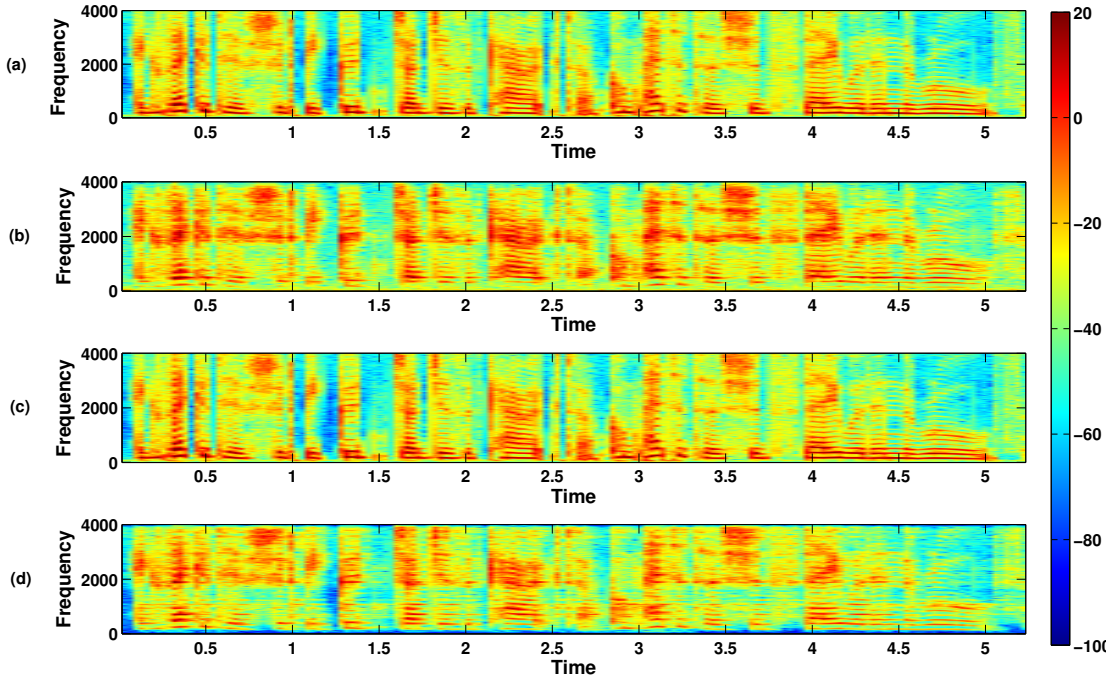


Figure 4.6 – The spectrogram of (a) Clean speech signal (b) A real mobile phone loudspeaker response (c) Synthesized speech signal using PSM (d) Synthesized speech signal using GPHM

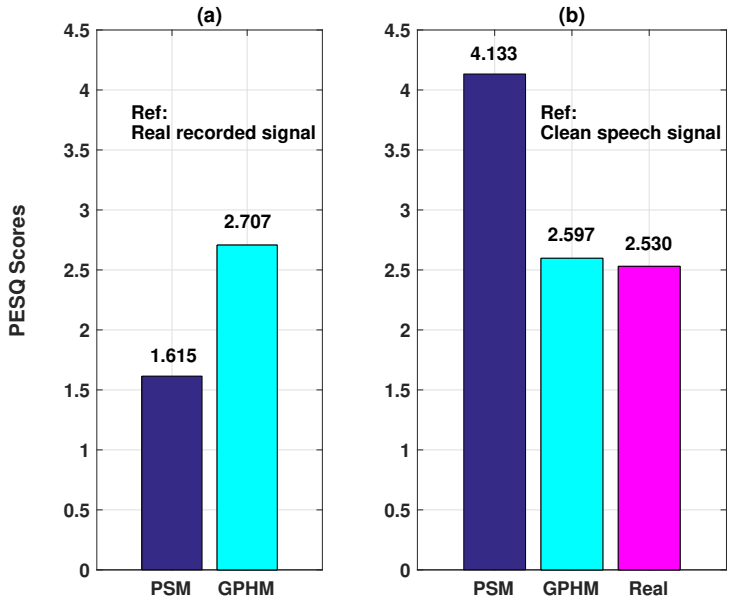


Figure 4.7 – (a) The PESQ scores between real measured loudspeaker signals and those synthesized with PSM and GPHM models. (b) The PESQ scores between clean speech signal and those synthesized with PSM and GPHM models along with a real measured loudspeaker signal.

signals are computed by comparing with the real recorded loudspeaker signal ($x_{real}(n)$) as a reference. The synthesized signal with GPHM attains a higher value, indicating it is a more accurate estimate of the nonlinearly distorted real recorded signal. On the other hand, Fig. 4.7(b) indicates the PESQ scores of the PSM, the GPHM and the real recorded signal by comparing with the input clean speech signal ($x(n)$) as a reference. The GPHM and the real recorded signal attains almost similar scores, indicating their similarity once again. The synthesized signal with PSM attains higher PESQ score when clean speech signal used as a reference, which indicates it is more close to clean speech signal than the distorted real recorded signal. Therefore, even though the GPHM and the PSM signals have equal amounts of nonlinear distortion, the GPHM more closely approximates the real nonlinear distortion of a loudspeaker.

Besides, the performance is also assessed objectively in terms of the Cepstral Distance (CD):

$$CD(m) = \sqrt{\sum_{L_f} \left[C_{x_{real}}(m) - C_{x_{model}}(m)\right]^2} \tag{4.2}$$

where L_f is the length of the frame. $C_{x_{real}}(m)$ and $C_{x_{model}}(m)$ are the column vectors of cepstral coefficients from the real recorded signal x_{real} and the model output x_{model} of the m^{th} frame respectively.

$$C_{x_{real}}(m) = IDFT\{ln\,|DFT[x_{real}(mL_f - 1)\cdots x_{real}((m+1)L_f)]|\} \tag{4.3}$$

In all cases measurements come from consecutive frames of $32ms$ ($L_f = 256$) in length. The reason why CD is that, it provides a more perceptually correlated assessment than alternative approaches based on energy or power differences. The CD profiles illustrated in Fig. 4.8 show that the difference between the measured signal and that synthesized with the GPHM model is consistently lower than that between the measured signal and the signal synthesized with the PSM model. The GPHM model thus better reflects the behavior of real nonlinear loudspeaker. This result was also confirmed with extensive informal listening tests which showed that signals synthesized with the GPHM model sound less artificial and are perceptually closer to the measured signal than those synthesized with the power series model.

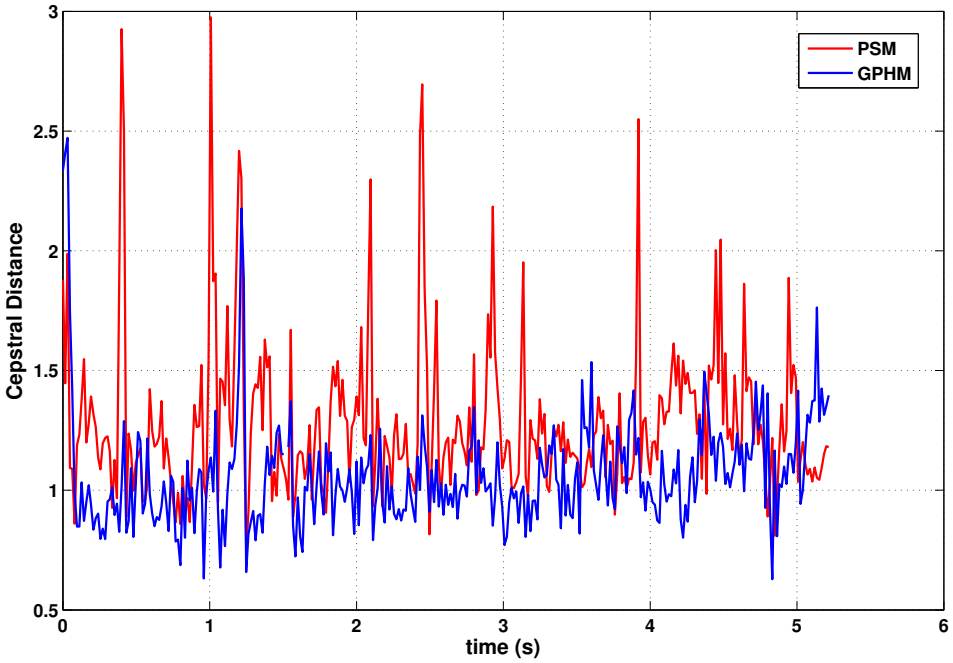


Figure 4.8 – An illustration of the cepstral distance between real measured loudspeaker signals and those synthesized with PSM and GPHM models.

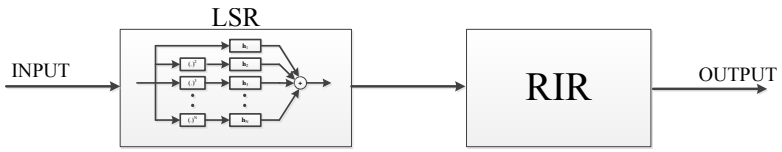


Figure 4.9 – Application of the nonlinear loudspeaker model. Input signals are processed according to the nonlinear loudspeaker response (LSR) and a room impulse response (RIR). The loudspeaker response (LSR) is synthesized using the GPHM model in Fig. 3.7 .

4.3 Validation of the GPHM

In this section, the GPHM model accuracy is investigated as a function of the key parameters, namely the number of filter taps L and the order of nonlinearities P. In this way, we can judge the influence of these parameters on the model performance. The results are published in our second paper [57].

4.3.1 Device Characterization

This work involves three different mobile phones (smart-phones). First, the simplified diagonal Volterra kernels $\mathbf{h}_p, p \in [1, P]$ for the three mobile phone loudspeakers are empirically computed using the procedure described in Section 4.1. Fig. 4.9 shows the

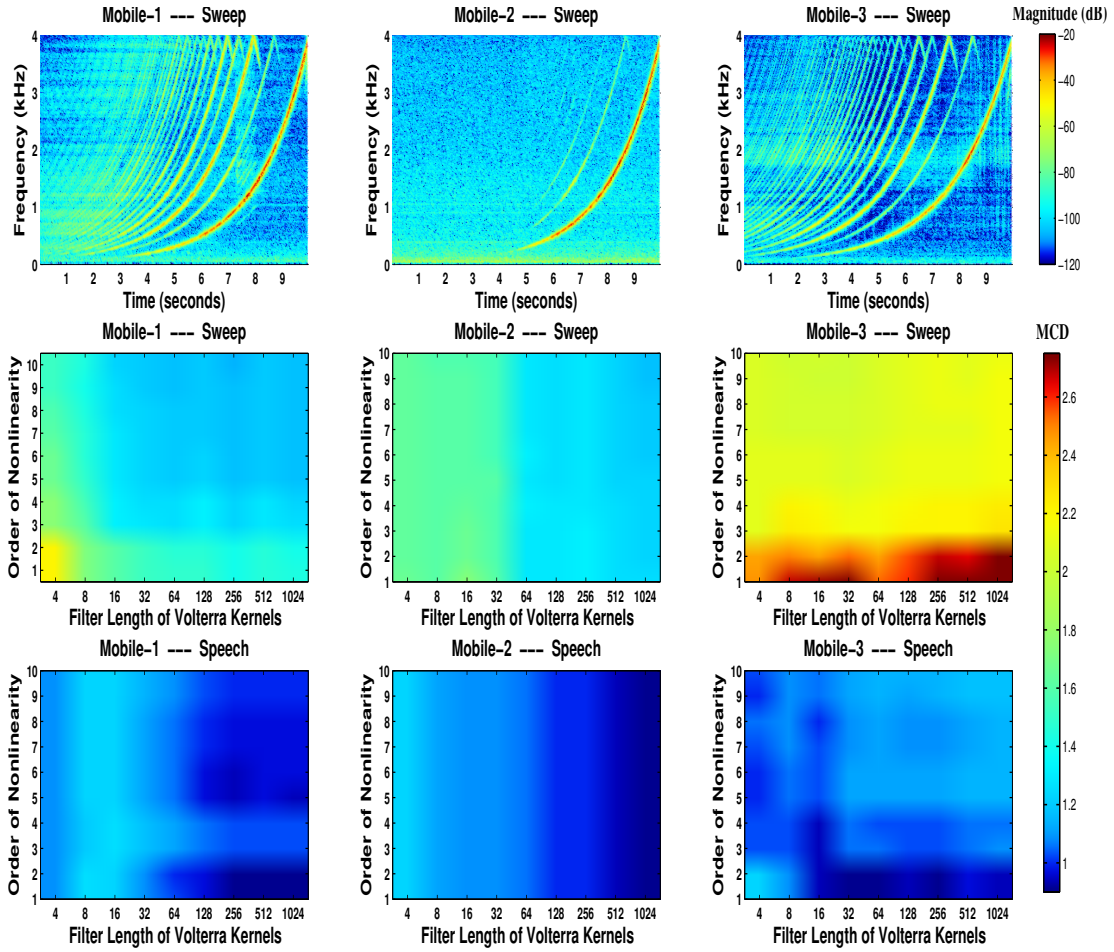


Figure 4.10 – An illustration of nonlinear characterization and model performance. The first row illustrates the response of each of three devices to the exponential sine-sweep input signal. Rows two and three illustrate the performance of the resulting nonlinear model to sine-sweep and real-speech input signals respectively. Results shown for different orders of nonlinearity P (vertical axes) and Volterra kernel lengths L (horizontal axes).

practical model topology. Input signals undergo two-fold filtering by: (i) a nonlinear loudspeaker response (LSR) and (ii) a room impulse response (RIR). Here we equalized all our responses as discussed earlier in Section 4.1.4 but the latter RIR section allows the application of the nonlinear model in any acoustic environment different to that used for practical measurements. In applying the nonlinear model, for all experiments reported below, we used the same acoustic booth RIR that have been measured in model estimation. It has a fixed length of 1024 taps at an $8kHz$ sampling frequency.

Model performance was assessed by comparing model and real loudspeaker outputs for a common input signal. Two different input signals were used: (i) the same exponential sine-sweep signal used in the experimental procedure and (ii) a real speech signal. Real

loudspeaker signals were recorded at the ear of the mannequin using the same experimental test-bed shown in Section 4.1.1. All signals are pulse code modulation signals sampled at $8kHz$.

Performance is assessed objectively in terms of the Mean Cepstral Distance (MCD) between the real recorded signals and model estimates:

$$\begin{aligned} CD(m) &= \sqrt{\sum_{L_f} \left[C_{x_{real}}(m) - C_{x_{model}}(m)\right]^2} \\ MCD &= mean(CD) \end{aligned} \tag{4.4}$$

Lower MCDs indicate that the model more accurately reflects the real measured outputs.

4.3.2 Results

The response of all three devices to the exponential sine-sweep input signal is shown in the form of spectrograms in the top row of Fig. 4.10. The 1st and particularly the 3rd device (left and right columns in Fig. 4.10) exhibit significant nonlinear distortion; spectrograms show additional higher order harmonics in addition to the input exponential sine-sweep input signal. The nonlinearity is furthermore asymmetric; odd-order harmonics are more significant than even-order nonlinearities. We note that some independent studies [9, 63] have reported similar observations. In contrast, the second device exhibits comparatively less nonlinear distortion.

Results for each of the three devices are also illustrated in Fig. 4.10. The middle row shows results for the exponential sine-sweep input signal whereas the lower row shows results for the real-speech input signal. In all cases, results are shown for different orders of nonlinearity P (vertical axes) and different Volterra kernel lengths L (horizontal axes). Blue colours illustrate lower MCDs whereas red colours indicate higher MCDs.

For satisfactory performance, the order of nonlinearity P should be high enough to capture the principal sources of nonlinear distortion, i.e. the most dominant harmonics. The simplified Volterra kernel filter length L should be sufficiently high so as to capture accurately both linear and nonlinear loudspeaker behavior. Both parameters are however a compromise between performance and computational efficiency.

Exponential sine-sweep input

The response of each device to the exponential sine-sweep input signal is illustrated in middle row of Fig. 4.10. For the 1st and 3rd devices, the MCD is higher for lower values of P, irrespective of the number of filter taps L. The MCD nonetheless decreases with increasing P. This behavior is not observed for the 2nd device where, in any case, the level of nonlinear distortion is comparatively low. It is nonetheless reassuring that there is negligible change in model accuracy for increasing (overestimated) P. For the 1st and 2nd devices, the MCD decreases as the kernel length L increases. However, for the 3rd device, with a value of $P > 2$ performance is relatively stable for varying L. One possible explanation for such behavior is that the highest order of significant nonlinearity exceeds that of the model ($P = 10$). Since the 3rd device exhibits nonlinearity greater than 10th order, P is not sufficient in this case to reduce the MCD. Accordingly, values of $P > 10$ would be needed where processing capacity allows.

Real-speech input

Results for real-speech inputs are illustrated in the last row of Fig. 4.10. Due to aliasing caused by the static nonlinearity modeling, MCD values are generally lower for speech than sine-sweep inputs. For the 1st device, the best performance is obtained for lower values of P and higher values of L. For the 2nd device, performance is best for higher values of L but is independent of P. For the 3rd device performance is best in the case of $P = 1$ and values of L around 64.

These results show that, for the two cases where nonlinearity is significant, the linear model ($P = 1$) outperforms the nonlinear model ($P > 1$) in the case of real-speech inputs. Despite the estimation of the nonlinearity is based on procedure that permits advanced analysis of nonlinear system [7, 8], our results show that this model does not match with approximation of nonlinearity observed with speech excitation signal for mobile devices. This leads to questions about the reasons of the observed mismatch. One explanation for this behavior lies in the wider variation in amplitude for speech signals compared to sine-sweep signals; lower amplitude speech signals may provoke significantly less nonlinear distortion. It is also possible that the model obtained from the system response to sine-sweep signals is overly simplistic. Whereas the sine-sweep signal consists in a single sinusoidal frequency at any instant, speech has a far more complex spectral density whereas the model neglects inter-spectral influences.

4.4 Summary

This chapter presents an overview of nonlinear loudspeaker modeling and an experimental way of identifying the nonlinearities of the loudspeakers used in mobile phones. This chapter also reports our work to assess the suitability of Volterra series derivatives in modeling the nonlinear distortion introduced by the mobile phone loudspeakers. We compared the synthesized outputs of two loudspeaker models to empirically measured, real loudspeaker outputs. The work suggests that the generalized polynomial Hammerstein model (GPHM) approximates more reliable practical nonlinear loudspeaker behavior.

This chapter also presents the key results of the GPHM model validation for the characterization and modeling of nonlinear loudspeakers. The simplified Volterra kernels, which characterize the nonlinear system, are empirically measured and then used to predict the response of three different mobile phones. Whereas validation with the same sine-sweep input signals used for characterization shows the potential, the model yields worse performance than a conventional linear model in the case of real-speech inputs. Benefits and limitations of the GPHM identification technique are discussed along with requirements for updating this technique to improve the ability to simulate behavior of complex loudspeaker systems.

The work highlights the challenge to model accurately the distortion introduced by nonlinear loudspeakers. Further refinements are thus necessary to achieve consistent practical performance, in particular with respect to inter-spectral influences. Future work should develop new modeling strategies based on real-speech input signals rather than specially-crafted, yet artificial inputs such as those used in this work. This will allow for the full consideration of intrinsic speech characteristics and the response of nonlinear systems to amplitude variations and the distribution of nonlinearities across the full spectrum.

Part II

Nonlinear Acoustic Echo Cancellation

After much of the study of loudspeaker nonlinearities in Part 1 of this thesis, Part 2 is majorly devoted to NAEC. Established state-of-the-art approaches to NAEC are described in Chapter 5 before a comprehensive performance evaluation and analysis of selected NAEC algorithms. This work highlights their strengths and weaknesses. Chapter 6 discusses a nonlinear and nonstationary signal analysis technique known as Empirical Mode Decomposition (EMD) and also reports a novel solution to NAEC based on EMD. The Hilbert-Huang Transform (HHT) allows spectral analysis of nonlinear and nonstationary signals by using EMD followed by the Hilbert transform. Chapter 7 reports our first attempt to apply the Hilbert-Huang Transform (HHT) to the analysis of nonlinear distortion produced by miniature loudspeakers. The work furthermore questions the suitability of traditional signal analysis approaches while giving weight to the use of HHT analysis in future work.

Chapter 5

State-of-the-Art NAEC Solutions

In this chapter, first we briefly review the existing techniques for the nonlinear acoustic echo cancellation and/or suppression, which have been developed within the last decade for mobile telephoney applications. Next, we have reported a comprehensive performance and stability analysis of widely used NAEC algorithms under various practical acoustic environments. A part of this work is presented as a technical report in [64].

Literature Survey

There is a considerable amount of literature in the last decade dedicated to the solutions that can handle nonlinear distortion and to maintain stable echo cancellation performance. First of all, we can divide these solutions into two main categories:

- Hardware-based solutions
- Software-based solutions

This is shown in Fig. 5.1. We should note that these two types of solutions are mutually exclusive in the literature till date and there are no solutions that can combine these two approaches within a single AEC system. In practice, however, AEC systems tend to end up in one or the other of the categories, for reasons of development and/or computational cost. We proceed then to the description of these solutions.

5.1 Hardware-based Solutions

Hardware-based solutions to handle nonlinear distortion are two fold:

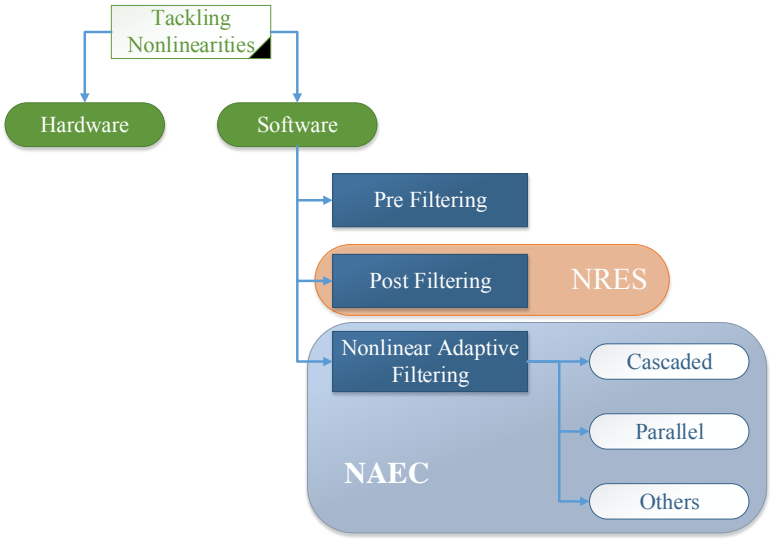


Figure 5.1 – Approaches to handle nonlinearities.

- Aim at furbishing up the loudspeaker's structural faults and resume normal linear performance levels.
- Choose to use additional hardware changes in the downlink path to get reference signals related to nonlinear distortion.

while the first option seems simple solution, hardware structural changes to miniature devices is not at all a cost effective solution. Using larger and high quality loudspeakers and its associated amplifiers can reduce the maximum amount of nonlinear distortion in the LEMS. But such a strategy will certainly increase the manufacturing cost and thereby price of the mobile devices significantly. The second option works on the idea of using nonlinearly distorted loudspeaker signal as reference signal to the linear AEC system. The advantage of this kind of method is that the echo canceler would have to model only the linear acoustic path from loudspeaker output to microphone output because the reference signal is nonlinearly distorted signal. Echo canceler doesn't have to model any nonlinearities. In order to do so, some additional hardware changes in the downlink path are mandatory to obtain the pure loudspeaker signal without room or any other acoustic artifacts. In [65], the authors used a sensor (an accelerometer) in the downlink path attached to the loudspeaker, expected to capture the nonlinear signal. This sensor signal is used as a reference signal to the NLMS based AEC to model the linear acoustic path. This approach provides $15dB$ ERLE improvement compared to conventional NLMS based AEC with no additional complexity. In [66, 67], the authors proposed to use either the voltage or the current signal that drives the loudspeaker as the reference signal to linear AEC. These signals capture some or all of the nonlinearities

in the downlink path. In order to obtain these signals to use as the reference signals, some hardware modifications based on RC filters are suggested in the loudspeaker cavity. This approach provides an average of $6dB$ ERLE improvement compared to conventional NLMS based AEC with no additional complexity.

Though the advantage of using nonlinearly distorted alternate reference signals is clearly seen, the usage of additional hardware changes in the downlink path is superfluous. Such additional hardware changes certainly increase the cost, and may cause additional distortions. For this reason, researchers believe that digital compensation of nonlinear distortion by means of Digital Signal Processing (DSP) algorithms may be a realistic solution.

5.2 Software-based Solutions

As shown in Fig. 5.1, the software-based solutions can be divided in to three categories depending on the way they tackle the nonlinear distortion in the LEMS:

- Nonlinear pre-filtering
- Nonlinear post-filtering
- Nonlinear adaptive filtering

Further, each category offered solutions in time, frequency and subband domain approaches in the literature. Hardware-based solutions do not necessarily depend on the exact nature of the nonlinear distortion, the alternate reference signals are expected to incorporate the required information. On the other hand, software-based solutions solely depend on the interpretation and analysis of the nonlinear distortion in the LEMS. Accordingly, several nonlinear models with and without memory are proposed in the literature.

5.2.1 Nonlinear Pre-Filtering

This approach aims at linearisation of the loudspeaker and its associated components in the downlink path through nonlinear pre-filtering of the far-end signal. Figure 5.2 illustrates the typical nonlinear pre-filtering scheme. The pre-filter is expected to compensate the nonlinear distortion introduced by the downlink path such that the united effect of this pre-filter and the loudspeaker behaves like a linear system. In case of ideal pre-filtering of the loudspeaker signal, the entire LEMS could be safely assumed as a linear system. This

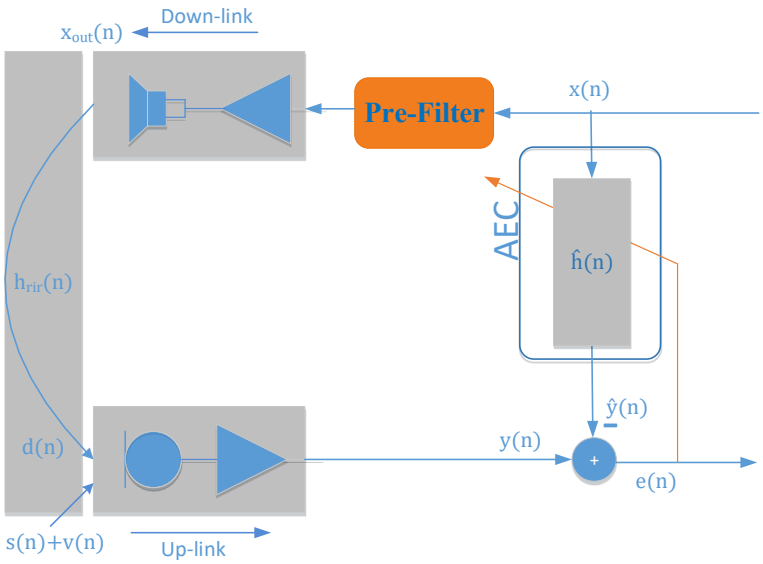


Figure 5.2 – Loudspeaker linearisation system.

assumption permits the use of conventional linear AEC and is sufficient to achieve better echo cancellation performance. As mentioned earlier, an exact model of the nonlinear loudspeaker system is of primary requirement here.

In [68,69], authors assumed the loudspeaker system as a 3^{rd}-order Volterra model with memory. So, they used a third order nonlinear pre-filter in series with the loudspeaker to compensate the nonlinear distortion. The kernels of the pre-filter were chosen such that the linear part of the overall transfer function (pre-filter + loudspeaker) must be equal to the linear transfer function of the loudspeaker. Since the complexity of the Volterra kernels increases exponentially with the increase in the model order, it may not be applicable for real-time processing. Hence, authors proposed a new kernel decomposition technique called Multi Memory Decomposition (MMD), to decompose the multi-dimensional quadratic Volterra kernels into a set of linear filters. Using MMD technique, authors achieved a good approximation of the quadratic Volterra kernels with less complexity. Further, authors proposed a MMD based pre-filter design with a flexibility of choosing the filters coefficients either by off-line or by real-time system identification techniques. This pre-filter design was implemented on a DSP chip followed by a real loudspeaker and the results are decent.

In [27], authors assumed the loudspeaker system as a 2^{nd}-order Volterra model with memory and proposed an off-line identification method to identify the linear and 2^{nd}-order Volterra kernel using multi-tone sinusoidal tones. Their pre-filter design shown in Fig. 5.3 was in frequency domain, $\hat{H}_2$ is the identified 2^{nd}-order Volterra kernel transfer function ($2D$ DFT) and H_1^{-1} is the inverse of the identified linear transfer function of

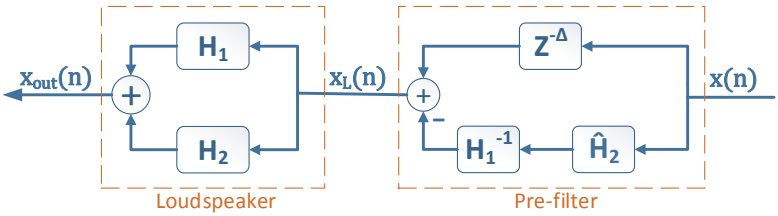


Figure 5.3 – Block diagram of the Loudspeaker linearisation system for eliminating the second-order nonlinear distortion.

the loudspeaker and satisfies the following equation:

$$H_1 H_1^{-1} = Z^{-\Delta} \tag{5.1}$$

where $Z^{-\Delta}$ indicates pure delay. This pre-filter design together with the loudspeaker model is expected to eliminate the 2^{nd}-order nonlinear distortion in the following way:

$$X_{out} = H_1.Z^{-\Delta}.X + Z^{-\Delta}.X.(H_2 - \hat{H}_2) - \hat{H}_2.H_1^{-1}.H_2.X \tag{5.2}$$

where X and X_{out} are the FFT of $x(n)$ and $x_{out}(n)$ respectively. If $\hat{H}_2 = H_2$ then the overall transfer function of the combined pre-filter and loudspeaker is approximately equal to the linear transfer function of the loudspeaker with some delay. However, this pre-filter structure introduces a new higher order nonlinear function $\hat{H}_2.H_1^{-1}.H_2$ but authors ignore this element. The simulation results presented are encouraging.

In [28], authors proposed a low complexity realization of the pre-filter design by proposing a subband version of the work in [27]. The structure is similar to the one shown in Fig. 5.3 but the only change is that the second path of pre-filter design contains the subband domain processing of the filters. Further, the identified 2^{nd}-order Volterra kernel in the pre-filter design is decomposed into canonical form using Eigen Value Decomposition (EVD) and represented as a parallel structure where the coefficients of each FIR filter are the elements of eigenvectors. Experimental results presented in the paper show that subband version of pre-filter design can produce the same compensation ability as the conventional method while reducing the computational complexity.

The solutions discussed so far may not be possible to deal with the variation of the loudspeaker nonlinear transfer function properties. In order to deal such an issue,

authors in [70] proposed an online loudspeaker linearisation approach with an adaptive pre-filter design. They assume the loudspeaker as a third-order memory-less nonlinear system followed by a linear system with memory, which can be realized as a Generalized Polynomial Hammerstein Model (GPHM) discussed in Chapter 2. Apparently the pre-filter design consists of a third-order parallel structure with a delay component in its first path and linear adaptive filters ($\hat{h}_2(n)$ and $\hat{h}_3(n)$) preceded by squaring and cubic terms in the second and third paths respectively. To minimize the second and third-order nonlinear distortion, the adaptive filters in the pre-filter design must converge to $-h_1^{-1}(n) * h_{p=2,3}(n)$ with $h_1^{-1}(n) * h_1(n) = \delta(n)$ (where $h_{p=1,2,3}(n)$ are the linear, second and third-order responses of the loudspeaker and $\delta(n)$ is the Dirac function). A NLMS based linear AEC is tested in the presence of nonlinear distortion with and without pre-filtering. Certainly, pre-filtering improved the ERLE performance by $5dB$ even under noisy conditions.

The major problem with the pre-filtering based loudspeaker linearisation systems is that the nonlinear distortion in the loudspeaker is not predictable. Model mismatches leads to great levels of gradient noise. The pre-filter designs involve huge serial arithmetic calculations which may distort the far-end signal, which is not acceptable in mobile devices. Moreover, most of the pre-filter designs involve inverse filtering without verifying the minimum-phase property of the loudspeaker linear transfer function. The pre-filter design often involves off-line calculation of its parameters which may not be suitable for dynamically varying speech signals. Most importantly, it is possible to compensate the nonlinear distortion using pre-filtering only by introducing new higher order nonlinearities. All these observations demonstrate the ineffectiveness of loudspeaker linearisation for mobile devices.

5.2.2 Nonlinear Post-Filtering

The second approach to handle the nonlinear distortion in the LEMS is by using nonlinear post-filtering. The linear and/or nonlinear post-filtering always works in combination with a linear echo canceler. The basic idea of nonlinear post-filtering is to suppress the nonlinear residual echo using a post-filter followed by a linear AEC. This scheme is an extension to the popular linear residual echo suppression and/or noise suppression techniques [2, 71]. The standard methods for designing a post-filter for the suppression of linear residual echoes also work under the assumptions of linear acoustic path and hence, they are not useful in the presence of strong nonlinear distortion in the LEMS. The structure of the post-processing scheme also known as Nonlinear Residual Echo Suppressor (NRES) is illustrated in Fig. 5.4. The microphone signal $y(n)$ comprises the desired near-end signal $s(n)$, nonlinearly distorted echo signal $d(n)$ and any unwanted

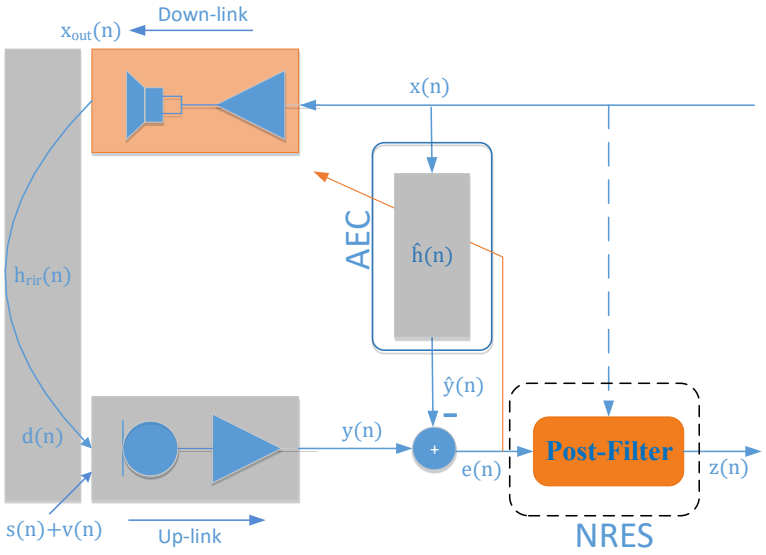


Figure 5.4 – Handling nonlinearities in the LEMS using NRES.

noise signal $v(n)$:

$$y(n) = s(n) + d(n) + v(n) \tag{5.3}$$

In the design of NRES system, we do not pose any specific constraints on the preceding linear AEC except its convergence, hence, any linear adaptive algorithm can be chosen for updating the linear AEC. After subtracting the linear echo estimate $\hat{y}(n)$, the output of the linear AEC $e(n)$ represents the near-end speech $s(n)$, the residual echo $d_r(n)$ and the noise $v(n)$.

$$e(n) = s(n) + d_r(n) + v(n) \tag{5.4}$$

The residual echo $d_r(n)$ comprises weak linear and strong nonlinear echo components. Under this assumption, we can further suppress the residual echo using a NRES system. Single-input post-filter designs are not applicable for mobile devices because of the strong nonlinear echo and it has common statistics from those of the near-end speech [72]. Therefore, most of the NRES systems proposed in the literature take in two inputs, one of them is obviously the output of the linear AEC $e(n)$ and the other input is either the far-end signal $x(n)$ [73, 74, 75, 76] or the linear echo estimate $\hat{y}(n)$ [72, 77, 78]. Irrespective

of the linear AEC processing domain, the NRES system always works in spectral envelop space (in frequency domain) and the functionality can be viewed as a famous spectral subtraction technique [79]. Accordingly, the output of the NRES system $Z(k, f)$ is obtained as a product of a real-valued spectral gain $G(k, f)$ and the output of linear AEC $E(k, f)$ as

$$Z(k, f) = G(k, f)E(k, f) \tag{5.5}$$

where k is the frame index, $Z(k, f)$ and $E(k, f)$ are the frequency domain representations of $z(n)$ and $e(n)$ respectively. It should be noted that, multiplying the output of the linear AEC $E(k, f)$ by the real-valued gain $G(k, f)$, it is expected to undo the variations of the magnitude/power spectrum caused by the nonlinear residual echo.

Just like spectral subtraction, the NRES system may be implemented in the power spectral domain [73, 74] or in the magnitude spectral domain [75, 76, 77, 78]. Accordingly, the real-valued spectral gain $G(k, f)$ depends on the power or the magnitude spectra of either the residual echo signal $D_r(k, f)$ or the near-end speech signal $S(k, f)$. However, they are not always readily available, hence the magnitude spectra of $D_r(k, f)$ or $S(k, f)$ must be estimated using recursive averages (first order digital low-pass filtering) [80] as follows

$$|\bar{D}_r(k, f)| = (1 - \alpha)|\bar{D}_r(k - 1, f)| + \alpha|\hat{D}_r(k, f)| \tag{5.6}$$

where $|\bar{D}_r(k, f)|$ is the smoothed magnitude estimate of the residual echo, $\hat{D}_r(k, f)$ is the estimated residual echo and α is the smoothing parameter. After the nonlinear residual echo suppression operation, the magnitude spectrum estimate $|Z(k, f)|$ is combined with the phase of the linear AEC output $e(n)$ and to inverse transform into the time domain.

In [77, 78], authors computed the spectral gain $G(k, f)$ as the ratio of $|S(k, f)|$ and $|E(k, f)|$. Since $|S(k, f)|$ is not readily available, it is estimated as follows (assuming $v(n) = 0$):

$$|S(k, f)| \simeq \sqrt{|\bar{S}(k, f)|^2} \simeq \sqrt{|\bar{E}(k, f)|^2 - |\bar{D}_r(k, f)|^2} \simeq \sqrt{|\bar{E}(k, f)|^2 - \hat{a_k}^2.|\hat{Y}(k, f)|^2} \tag{5.7}$$

where over-line $\bar{.}$ means recursive averaging operation and $|\hat{Y}(k,f)|$ is the smoothed magnitude spectrum of linear echo replica. The equation indicates that the smoothed magnitude spectrum of nonlinear residual echo $|\bar{D}_r(k,f)|$ is estimated using $|\hat{Y}(k,f)|$ as

$$\hat{a_k} = \frac{|\bar{D}_r(k,f)|}{|\hat{Y}(k,f)|} = \frac{|\bar{E}(k,f)|_{Single-talk}}{|\hat{Y}(k,f)|} \tag{5.8}$$

The parameters $\hat{a_k}$ for each frame index k is hardware dependent and have to compute by additional experimental measurements in quite environments for each and every mobile device after manufacturing. This parameter computation is the major drawback of this method; further this parameter $\hat{a_k}$ is static and does not have control over variations in the nonlinear distortion.

In [73, 74], authors derived the spectral gain $G(k,f)$ by minimizing the contribution of the nonlinear residual echo $D_r(k,f)$ to the post-filter output signal $Z(k,f)$ in the mean square error (MSE) sense:

$$G(k,f) = \frac{\bar{S}_E(k,f) - \beta.\bar{S}_{D_r}(k,f)}{\bar{S}_E(k,f)} \tag{5.9}$$

where β controls the *aggressiveness* of the NRES operation, $\bar{S}_E(k,f)$ and $\bar{S}_{D_r}(k,f)$ denote the recursively smoothed power spectral densities (PSDs) of $E(k,f)$ and $\hat{D}_r(k,f)$ respectively. In [73], in the place of conventional linear AEC, authors used multi-channel adaptive filters or *parallel/power filters* model Nonlinear Acoustic Echo Canceler (NAEC) [48] to estimate not only the linear echo signal but also higher-order nonlinear echo signal. The power filters based NAEC is discussed in detail in the next section of this chapter. Assuming the first channel of power filters NAEC (i.e., linear AEC) attains to its Wiener solution, we can compute the estimate of the nonlinear residual echo $\hat{d}_r(n)$ by summing the outputs of the other parallel channels $\hat{y}_{p\geq 2}$:

$$\hat{d}_r\,(n) = \sum_{p=2}^{P} \hat{y}_p\,(n) \tag{5.10}$$

Accordingly, the PSD of the nonlinear residual echo estimate can be written as

$$\bar{S}_{D_r}(k,f) = \sum_{p=2}^{P} \bar{S}_{\hat{y}_p}(n) \tag{5.11}$$

where $\bar{S}_{\hat{y}_p}$ is the smoothed power spectrum of the p^{th} parallel channel, $p = 2, \ldots, P$. Once we compute $\bar{S}_{D_r}(k,f)$, computing the $\bar{S}_E(k,f)$ is straight forward and correspondingly, the spectral gain can be found using Eq. 5.9. This structure seems compromising but the key assumption of linear AEC (first channel in the power filters) attaining Wiener solution is not convincing and it never happens in the presence of strong nonlinear distortion. More importantly the convergence of all the other adaptive filters is also mandatory to get the proper estimate of $\bar{S}_{D_r}(k,f)$ and it depends on the lengths of the adaptive filters involved. Further, this model is relatively too complex because of the multiple adaptive filters.

In order to solve these issues, authors in [74] estimated the PSD of the nonlinear residual echo $\bar{S}_{D_r}(k,f)$ independent of the length of the room impulse response (RIR) by modeling the loudspeaker nonlinearities as a linear combination of basis functions. This method outperforms the method in [73] in terms of both convergence rate and maximum achievable ERLE. However, the results presented involve theoretically synthesized echo signals and the performance may definitely vary under realistic situations. Moreover, the PSD estimation involve many assumptions which may not hold true in practice and also includes an auto-correlation matrix inversion which is computationally expensive.

Another frequency domain NRES approach based on spectral shaping, in particular to handle harmonic distortion, is presented in [75]. But in this method authors used the Modulated Complex Lapped Transform (MCLT) (refer [81]) instead of Fourier transform claiming MCLT allows for perfect reconstruction to transform the time domain signals to the frequency domain. This approach uses the regression coefficients to estimate the magnitude spectrum of the nonlinear residual echo $\bar{S}_{D_r}(k,f)$ from the far-end signal $X(k,f)$ but these regression coefficients are not fixed as in [77] but instead computed adaptively. The algorithm appears to be fairly robust especially in case of noisy environments but suffers a drop in ERLE if there is a mismatch in the assumed order of harmonic distortion.

The main attraction of nonlinear post-filtering or NRES is its relative simplicity, in that it only requires an estimate of the frequency dependent real-valued spectral gain. However, computing the estimates of the magnitude/power spectra of near-end speech

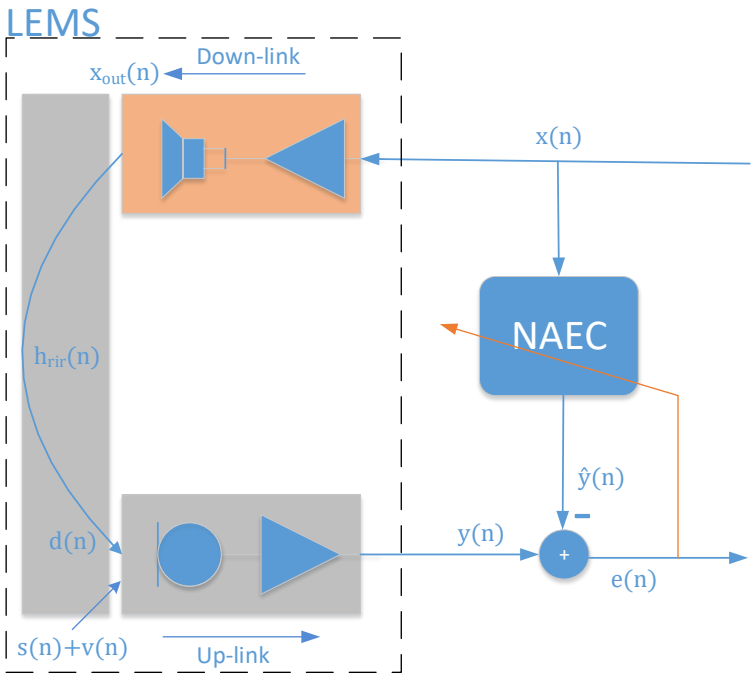


Figure 5.5 – Structure of NAEC.

and/or nonlinear residual echo are more prone to errors and often depends on many assumptions. These errors in the model leads to unwanted modulations and further causes over spectral subtraction leads to near-end speech distortion.

5.2.3 Nonlinear Adaptive Filtering

The third and the more general approach to handle the nonlinear distortion in the LEMS is by using Volterra series and nonlinear adaptive filtering in the AEC. The structure of Nonlinear Acoustic Echo Canceller (NAEC) is shown in Fig. 5.5. Using nonlinear adaptive filtering in the NAEC, one aims to estimate both the linear and nonlinear echo together and cancel from the microphone signal. Over the past decade, various approaches have been proposed for nonlinear acoustic system identification for echo cancellation. These techniques are classified into three types based on the domain of their implementation. There are time domain [4, 12, 30, 48, 49, 82, 83], frequency domain [84, 85] and subband domain [24, 29] solutions. As shown in Fig. 5.1, the NAEC solutions can be classified into three groups depending on the structure of nonlinear system identification involved:

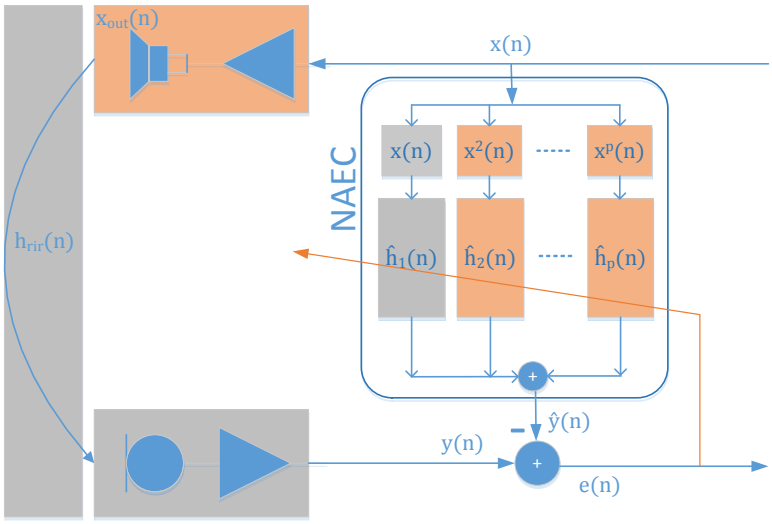


Figure 5.6 – Structure of parallel approach based NAEC.

- Parallel approach
- Cascaded approach
- Other approaches

Parallel Approach

The parallel approach (or the power filter approach) based NAEC takes into account the Hammerstein model discussed in Section 2.2.4 (Eq. 2.10) to estimate the overall LEM system. This involves the simultaneous tracking of both the nonlinear and linear impulse responses through the multi-channel adaptive filters as illustrated in Fig. 5.6. This approach is especially suitable for memoryless nonlinearities, where the first channel represents the overall linear impulse response of the LEMS. Concurrently, the other channels are used to adaptively track the higher order nonlinearities in the LEMS. The input signals to the different channels are accordingly $x(n), x^2(n), \cdots, x^P(n)$.

In [48], authors used a 3^{rd}-order parallel NAEC model and used an NLMS based approach to update all the three adaptive filters. Since the input signals to the multi-channels adaptive filters are the powers of the same reference signal, they are generally highly correlated. This strong inter-channel correlation has a profound impact on the convergence speed of the adaptive filters. Hence the authors proposed to use an additional orthogonalization stage prior to the adaptive filtering. The simulation results show the

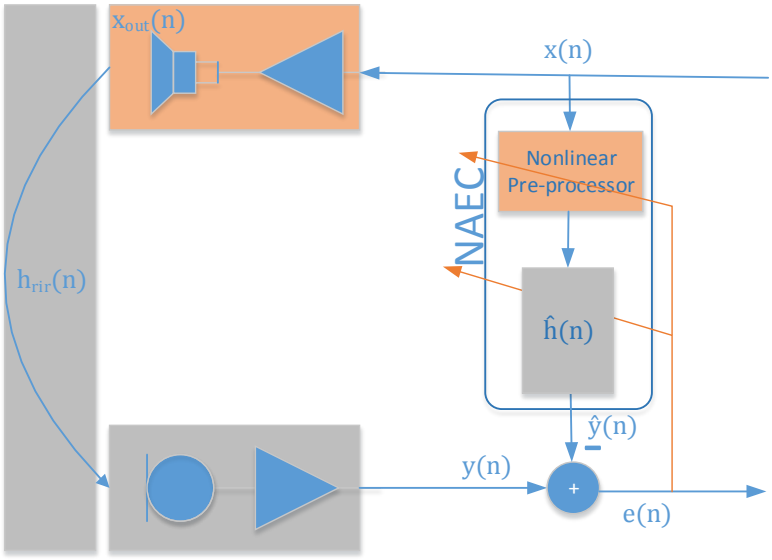


Figure 5.7 – Structure of cascaded approach based NAEC.

better ERLE improvement over linear AEC. This algorithm has been rigorously tested for validation and performance with empirical nonlinear signals and the results are presented in Section 5.4.

Adaptive filtering in frequency domain has an advantage of fast convergence speed and reduced complexity due to fast convolution property. On the other hand multi-channel adaptive filtering in frequency domain exploit the orthogonality properties of the DFT (discrete fourier transform) and hence do not suffer from inter-channel correlation. Making use of these benefits, authors in [85] proposed a parallel approach based NAEC in frequency domain. Simulation results demonstrate the superiority of frequency domain solution over time domain solution in [48] both in terms of convergence speed and maximum achievable ERLE.

Although this parallel approach seems to be simple and practically possible NAEC structure, the convergence speed is always slow compared to the linear AEC due to the multi-channel adaptive filtering and the increased number of filter coefficients. Also, there is an inherent disadvantage of estimating the linear acoustic path (RIR) multiple times through multiple channels (due to the effect of convolution between the loudspeaker parameters and the RIR). The stability and the robustness of parallel approach based NAEC during dynamic variation of acoustic echopaths and the order of nonlinearity is studied in the later sections of this chapter.

Cascaded Approach

The key idea behind the cascaded approach based NAEC was to decouple the identification of nonlinear loudspeaker parameters from the tracking of linear acoustic echopath. The structure of cascaded approach based NAEC is illustrated in Fig. 5.7. In this case, a nonlinear pre-processor is placed in cascade with a linear Finite Impulse Response (FIR) filter. The nonlinear pre-processor filters the reference signal from the far-end and aims to emulate the downlink path along with its nonlinearities. The linear FIR filter aims to accurately emulate the unwanted nonlinear echo in the microphone signal by filtering the incoming nonlinear signal. This cascaded approach involves joint parameter estimation for the nonlinear pre-processor and the FIR filter using the combined error signal $e(n)$ employing linear adaptive algorithms. The nonlinear pre-processor block requires an appropriate nonlinear model to accurately emulate the downlink nonlinear behaviour. Nonlinear models with memory [12, 30] and without memory [46, 47, 49] were proposed in the literature.

In [12], authors considered a 2^{nd}-order Volterra series with memory as a pre-processor and used NLMS adaptive algorithm to estimate the filters coefficients and achieved a 7dB ERLE improvement compared to the conventional linear AEC. In [30], authors assumed the loudspeaker model as a 3^{rd}-order Volterra series with memory and used a modified cascaded model to identify the nonlinear echopath. Assuming the invertibility of the loudspeaker linear impulse response ($h_1(n)$), in their modified cascaded approach the nonlinear pre-processor block adapts to the convolution between second order kernel $h_2(n)$ (respectively third order kernel $h_3(n)$) and the inverse of $h_1(n)$. The linear FIR filter adapts to the global linear impulse response of the LEMS, $h_{rir}(n) * h_1(n)$ (where $*$ indicates convolution). In accordance with the modifications in the cascaded structure, a modified NLMS algorithm was derived and used to adaptively estimate both of the blocks. Simulation results reported a 5dB ERLE over conventional NLMS based linear AEC and a 2dB ERLE improvement over parallel approach based NAEC respectively. This approach is conceptually viable but in practice the modified cascaded structure and the corresponding NLMS algorithm are computationally more complex in relative comparison with the normal cascaded structure. Further the primary assumption of loudspeaker linear impulse response invertibility may not be true and any mismatches leads to total system divergence and uncontrollable gradient noise in the uplink path.

Authors in [82] modeled the loudspeaker as a 2^{nd}-order Volterra series with memory and noticed that most of the energy in the second order kernel is concentrated around the diagonal and mainly contains the direct path with a very few early reflections. Hence, in order to control the computational complexity authors truncated the non-significant coefficients in the second order Volterra kernel. It has also been shown that the linear

kernel requires larger memory length compared to quadratic (or higher order) kernels in order to model the nonlinear behaviour of loudspeakers. The NLMS algorithm was used to track the overall nonlinear acoustic echopath and achieved a 7dB ERLE improvement over linear AEC. Since the NLMS algorithm usually exhibits low convergence and tracking rates when the input signals are coloured (especially for speech), authors in [86] used APA algorithm to improve the learning rate of the filters. However, algorithms like APA and RLS significantly increase the computational complexity of the system.

Authors in [84] considered a 2^{nd}-order Volterra series with memory as a pre-processor and proposed a frequency domain NAEC solution to achieve effective reduced complexity implementation. In general, frequency domain nonlinear adaptive filtering has a limitation to use the same memory length for all the kernels [87]. Authors presented a way to allow different memory lengths for the linear and 2^{nd}-order kernel by extending the partitioned block techniques to Volterra filters. Using NLMS algorithm with frequency dependent normalization authors achieved a better initial convergence and ERLE performance compared to partition block based frequency domain linear AEC and time domain adaptive Volterra filter algorithm. However, a comprehensive validation to prove the stability of the proposed method is lacking. Moreover, the study of nonlinear systems in frequency domain involves a range of uncertainty. Unlike linear systems, the connection between the input and the output spectra of nonlinear systems is more complicated. Certainly, the linear system frequency domain analysis techniques cannot easily be extended to the nonlinear systems. This phenomenon will be discussed at greater length in the next two chapters.

Several techniques based on memoryless nonlinear models have been proposed to cope with the overall system complexity and to improve the learning rate of the cascaded blocks. In [47], authors considered a memoryless nonlinear preprocessor and compared the performance of different adaptive algorithms, like NLMS, orthogonalised NLMS and Recursive Least-Squares (RLS) algorithms, by varying the order of nonlinearity (P). Authors also proposed appropriate step-size control mechanisms to enhance the convergence speed for the considered adaptive algorithms which then provide upto 10dB ERLE gain compared to the linear AEC case. In [49], authors proposed a low cost NAEC by using a memoryless preprocessor with fewer filter taps compared to the linear block and suggested to use higher step-size for the linear block to achieve good results. The simulation results with NLMS algorithm achieved better ERLE compared to both the parallel approach based NAEC and the linear AEC at different input SNR levels and at different orders of nonlinearity. The dynamic re-convergence and the maximum achievable ERLE also improved with the increasing input SNR. This algorithm has been rigorously tested for validation and performance with empirical nonlinear signals and the results are presented in Section 5.4.

We will see in the next section of this Chapter, there is the obvious question of the stability of most of the NAEC algorithms in the literature. In general, the signal statistics of the nonlinear echo are unknown and will be time-varying for nonstationary signals like speech. A stable NAEC algorithm avoids performance degradation in case of: nonlinear model mismatch, over/under modeling the nonlinear adaptive filters, dynamic variation of acoustic echopath, and dynamic variation of order of nonlinearity. The loudspeaker nonlinear behaviour changes over time due to structural deformation, overload, ageing, ambient temperature or climate effects [9]. Unfortunately, most of the published NAEC algorithms have never been rigorously tested to determine the solutions to manage these practical problems.

More importantly, for any NAEC algorithm to achieve a good ERLE (like a linear AEC) in the presence of linear echo (absence of nonlinearities in the echo) is a challenging task. Authors in [53, 88] attempted to achieve the same by using adaptive convex combination of a linear kernel and a 2^{nd}-order Volterra kernel. In this approach, the linear and the 2^{nd}-order Volterra kernels are independently adapted using their own error signals and their outputs are adaptively combined by means of a stochastic gradient algorithm in order to minimize the overall error. It has been presumed that the combined scheme performs at least as well as the best contributing filter. The simulation results presented are satisfactory, but the filter combinations schemes have not been thoroughly validated with real or empirical nonlinear echo signals. For a more detailed treatment of comparison results of filter combination techniques with the other NAEC algorithms, the reader is referred to [66, 67].

Most of the published NAEC solutions in cascaded approach are conceptually possible, but as a practical matter, they seem implausible, for at least two reasons. First, the cascaded model requires the nonlinear pre-processor and the linear filter adaptation using a single joint error signal. As a result the convergence of both the blocks (or all the adaptive filters) are interdependent, which leads to possible errors and also reduces the over-all convergence speed. One possible way to solve this drawback is to adapt the nonlinear pre-processor separately using a reference microphone signal by placing it very close to the loudspeaker cavity. However, in such scenarios there is no need to use a cascaded approach neither any NAEC solution. Since the reference microphone signal incorporates the nonlinearities and the downlink system properties, a conventional linear AEC would be sufficient to cancel the nonlinear echo. Second, most of the NAEC algorithms start with the assumption of the type of the nonlinear model suitable to a microspeaker (a miniature loudspeaker). A slight change in any of the assumptions drastically alters the performance of NAEC algorithms. Choosing the right order of nonlinearity and the memory size (or filter-length in case of memory-less nonlinearities) in the nonlinear pre-processor can be challenging. Under-modeling results in high residual

nonlinear echo where as over-modeling effects the convergence due to increased complexity, and the maximum achievable ERLE by adding gradient noise to the uplink signal.

Other Approaches

As discussed in Part 1 of this thesis, the complexity of the Volterra filters increases exponentially with the increase in the order of the Volterra model. The large number of filter taps leads to slow convergence, vast misadjustment, and increased computational complexity. In order to increase the effectiveness of nonlinear system identification, some of the authors deviated from the conventional adaptive filtering and other approaches have been devised and developed during the last few years..

In [83, 89] authors exploited the state-space modeling and Kalman filter recursions for the NAEC problem. In [89], authors first transformed a conventional parallel approach based NAEC model into a multi-channel state-space structure by augmenting with a multi-channel first-order Markov model. The state-space model in each channel consists of two model equations: an observation (or measurement) equation and a state (system) equation. Authors then proposed two different variants of updating the unknown filter coefficients in the model equations using a recursive Bayesian estimator in the form of frequency domain multi-channel Kalman filters. The proposed NAEC algorithm was tested under realistic situations like dynamically varying nonlinear distortion and acoustic echo path and in the presence of double-talk. The simulation results published are encouraging. In [83], authors proposed a time-domain Kalman filter solution to NAEC. Like in any state-space modeling, authors transformed a cascaded NAEC model to a cascaded state-space model based on mathematical manipulations. Authors then proposed to use Kalman filters operating sequentially to update the unknown filter coefficients in the model equations. However, the inherent problem here is the Kalman filter works on the basic assumption of the linearity and the Gaussianity of the state and the measurement models. This assumption does not hold in the NAEC problem. Also, the several mathematical assumptions made while transforming the conventional NAEC model to a state-space model may not valid in practice. Imprecise models can lead to a state-error covariance matrix not positive semi-definite and the smoothing errors can significantly impact the performance of NAEC.

In [90, 91, 92] authors make use of the artificial neural networks to solve the NAEC problem. In these approaches, the NAEC is carried out by composing the LEMS as a cascaded model. An artificial neural network (ANN) based nonlinear pre-processor is used in order to model the nonlinear behaviour of the loudspeakers followed by a conventional linear adaptive filter to model the linear echopath. Different algorithms were proposed in the literature to train the ANN. Despite having good simulation results

with the artificial data, there are many drawbacks associated with the usage of neural networks especially for the NAEC problem. Besides the general problems like higher computational complexity and biased convergence, a major drawback with the neural network based models is their extreme sensitivity to the model mismatches. Since there is no well-defined nonlinear loudspeaker model available, the behaviour of an ANN system largely depends on its training, which can also be problematic as it requires a huge amount of training data. Another major problem is that the ANN's introduce many local minima in the error surface which leads to higher probability to converge to those local minima while using the gradient descent based learning algorithms.

5.3 Influence of the simulated nonlinear signal models on NAEC evaluation

Most of the published NAEC algorithms have never been rigorously tested with empirical and/or real recorded nonlinear loudspeaker signals. As we have seen in Section 4.2, the correlation between the simulated nonlinear signals and the real nonlinear signals is always poor. The power series model (the PSM) shown in Eq. 3.4 is very often used in the literature to simulate the memoryless nonlinear signals. Some specific simulated nonlinear signals like the one in Eq. 3.4 have the potential to over-exaggerate the performance of NAEC algorithms. In this section, we assess the performance of a typical NAEC algorithm, for example the popular cascaded NAEC system proposed in [49], when exposed to the simulated (the PSM model given in Eq. 3.4), the empirically generated (the GPHM model give in Eq. 3.5) and the real recorded nonlinear signals.

5.3.1 Cascaded Model

The block diagram shown in Figure. 5.8 illustrates the structure of a P^{th} order cascaded model NAEC [49]. The input signal $x(n)$ with sampling frequency f_s is passed into a pre-processor that contains P different channels. In the p^{th} channel, the input vector $\mathbf{x}(n) = [x(n), \ldots, x(n-N_p+1)]^T$ is passed through a low-pass filter (or anti-aliasing filter) with cut-off frequency $f_s/2p$ before taken to the p^{th} power, and then passed through an estimated sub-filter whose impulse response is given by $\hat{\mathbf{h}}_p(n)$. The low-pass filters are used to avoid aliasing and to make sure that the frequency content of the input signal is limited before taken to the p^{th} power. The output $\hat{y}_s(n)$ of the pre-processor, which aims to model the loudspeaker output, is obtained by the summation of over-all channel

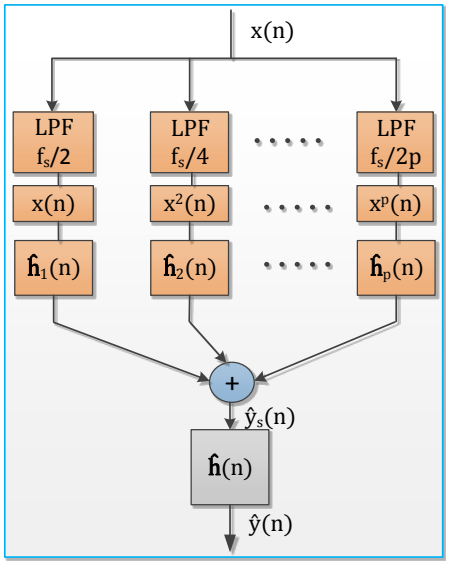


Figure 5.8 – An illustration of the cascaded model NAEC. In the p^{th}-channel, the input signal vector passes through a low-pass filter (LPF) with cut-off frequency $f_s/2p$ to avoid aliasing.

outputs:

$$\hat{y}_s(n) = \sum_{p=1}^{P} \hat{\mathbf{h}}_p^T(n)\mathbf{x}^p(n) \tag{5.12}$$

where $\hat{\mathbf{h}}_p(n)$ is the estimated sub-filter vector of length N_p. The pre-processor output $\hat{y}_s(n)$ is then passed through a linear filter $\hat{\mathbf{h}}(n)$, which aims to model the RIR, to get the overall output of the cascaded model $\hat{y}(n)$:

$$\hat{y}(n) = \hat{\mathbf{h}}^T(n)\hat{\mathbf{y}}_s(n) \tag{5.13}$$

The update equations based on NLMS algorithm for the pre-processor filters and the linear filter are given by [49]:

$$\hat{\mathbf{h}}_p(n+1) = \hat{\mathbf{h}}_p(n) + \frac{\mu_p}{\|\mathbf{X}_p(n)\mathbf{h}^T(n)\|_2^2 + \delta_p}\mathbf{X}_p(n)\mathbf{h}^T(n)\,e(n) \tag{5.14}$$

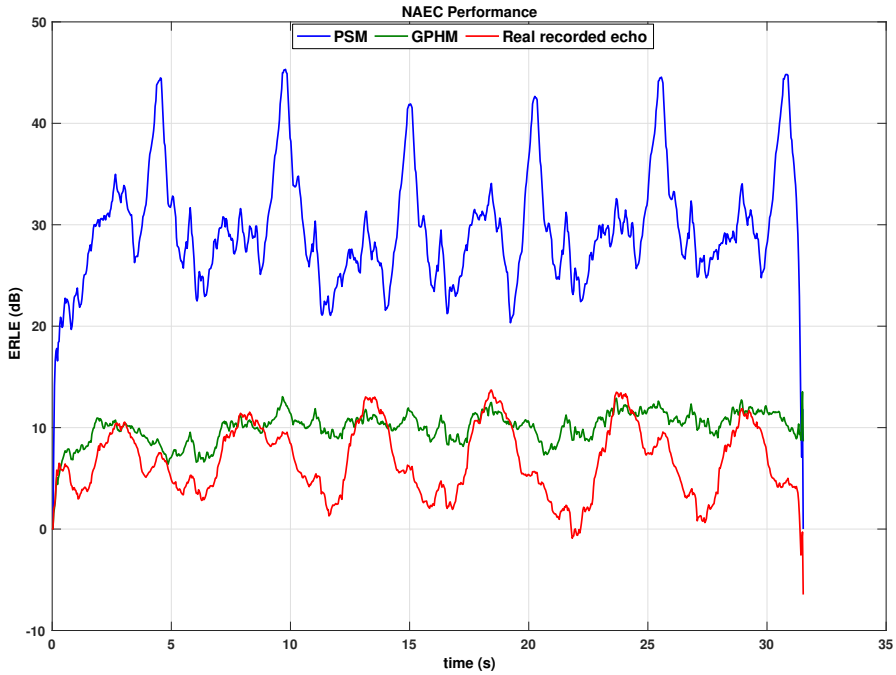


Figure 5.9 – NAEC performance in terms ERLE with either real recorded nonlinear echo signals or those synthesised with the PSM or the GPHM models.

$$\hat{\mathbf{h}}\,(n+1) = \hat{\mathbf{h}}\,(n) + \frac{\mu}{\|\hat{\mathbf{y}}_s(n)\|_2^2 + \delta}\hat{\mathbf{y}}_s(n)e\,(n) \tag{5.15}$$

where $\mathbf{X}_p(n) = [\mathbf{x}_p(n), \ldots, \mathbf{x}_p(n-N+1)]^T$ and δ is the regularization parameter.

5.3.2 Experimental results

The performance of the cascaded model NAEC has been investigated when exposed to the simulated, the empirically generated, and the real recorded nonlinear echo signals. First the loudspeaker nonlinearities ($x_{out}(n)$) are synthesized through the PSM and the GPHM models as described in Chapter 4.2. Then the microphone output signals with nonlinear echo are generated according to:

$$y\,(n) = \sum_{i=0}^{N-1} x_{out}(n-i)h_{rir}(i) \tag{5.16}$$

where $h_{rir}(n)$ is a room impulse response (RIR) and $x_{out}(n)$ is a loudspeaker output signal computed using either Eq. 3.4 (the PSM) or Eq. 3.5 (the GPHM). Experiments were performed with diagonal (one-dimensional) loudspeaker Volterra kernels $h_p(n)$ for values of $p \leq 5$ and with $N_p = 256$ taps. The acoustic channel was modeled with a fixed 256-tap room impulse response (RIR) $h_{rir}(n)$ selected from the Aachen RIR database [1]. All experiments were performed with a clean speech downlink/reference signal $x(n)$ of approximately 30 seconds duration with a sampling frequency of 8kHz. A set of common filter parameters were applied to all three test cases and were chosen to maintain stability and better performance. The test configurations are $N_{p=1,\cdots,5} = 256$, $N = 256$, $\mu_{p=1,\cdots,5} = 0.01$, $\delta_{p=1,\cdots,5} = 1e-4$, $N = 256$, $\mu = 0.5$, $\delta = 1e-7$. The ERLE results are illustrated in Fig. 5.9.

While NAEC performance in the case of loudspeaker signals synthesised with the PSM model is similar to that obtained in previous work [49], poorer performance is observed in the case of real recorded nonlinear echo signals. While NAEC performance in the case of signals synthesised empirically with the GPHM approach also differs from that with real recorded nonlinear echo signals, the difference is significantly reduced.

These observations confirm the significant, favourable bias in results generated with the popular PSM model and emphasise its potential influence on the evaluation of NAEC performance. Results generated with the GPHM model better reflect practical measurements and thus the empirically generated loudspeaker signals are an appealing alternative to be considered for future work. Results thus derived will exhibit less bias than those reported previously in the open literature, and provide a more realistic estimation of practical NAEC performance.

5.4 Comprehensive performance analysis of NAEC algorithms

As discussed in the first two sections of this Chapter, several NAEC algorithms have been proposed in the literature to handle nonlinearities in the acoustic echopath. However, their evaluation methodologies are not as compelling as their key design idea because most of the algorithms had never been tested under both real nonlinear echoes and real mobile phone loudspeaker data. We have learned in the previous section that the performance of the NAEC algorithms can be inflated using specific test signals. Further observed, many of the NAEC algorithms are developed based on two different rationales, Parallel and Cascaded approaches, each possessed its own merit and claimed outperforming other. The claim is prone to subjectivity because the algorithms are compared only in few idealistic situations. In this section, we conduct a deep performance analysis

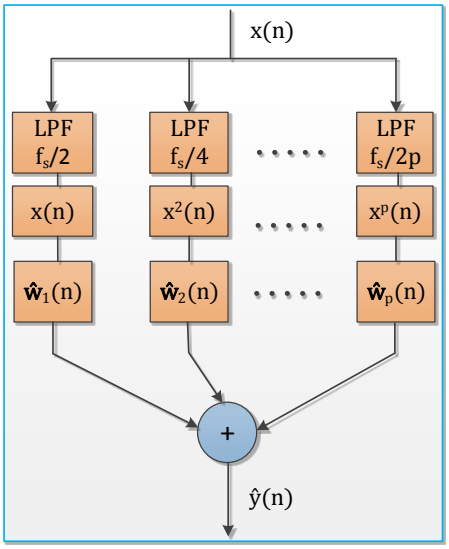


Figure 5.10 – An illustration of the parallel/power-filter model NAEC. In the p^{th}-channel, the input signal vector passes through a low-pass filter (LPF) with cut-off frequency $f_s/2p$ to avoid aliasing.

and comparison of these two typical NAEC structures under various and more practical situations. The cascaded model was described in the previous section and next we will briefly outline the parallel/power filter NAEC followed by the comprehensive analysis part.

5.4.1 Parallel Model

The block diagram shown in Fig. 5.10 illustrates the structure of a P^{th} order power-filter model NAEC [48]. The input signal $x(n)$ is passed into P different parallel channels. In the p^{th} channel, the input vector $\mathbf{x}(n) = [x(n), \ldots, x(n - L_p + 1)]^T$ is passed through a low-pass filter (or anti-aliasing filter) with cut-off frequency $f_s/2p$ before taken to the p^{th} power, and then passed through an estimated linear filter vector $\hat{\mathbf{w}}_p(n)$. The overall output $\hat{y}(n)$ of the power-filter model is obtained by the summation of all parallel channels outputs:

$$\hat{y}(n) = \sum_{p=1}^{P} \hat{\mathbf{w}}_p^T(n)\mathbf{x}_p(n) \tag{5.17}$$

where $\hat{\mathbf{w}}_p(n)$ is the estimated filter vector of length L_p. The update equation based on

NLMS algorithm for the power-filters is given by [48]:

$$\mathbf{w}_p(n+1) = \mathbf{w}_p(n) + \frac{\mu_p}{\|\mathbf{x}^p(n)\|_2^2 + \delta_p}\mathbf{x}^p(n)e(n) \tag{5.18}$$

In ideal case the linear filters of the power-filter model $\hat{\mathbf{w}}_p(n)$ are a combination of the pre-processor sub-filters $\hat{\mathbf{h}}_p(n)$ and the linear filter $\hat{\mathbf{h}}(n)$ of cascaded model:

$$\hat{\mathbf{w}}_p(n) = \hat{\mathbf{h}}(n) * \hat{\mathbf{h}}_p(n)$$

where $*$ represents linear convolution and leads to the equality $L_p = N_p + N - 1$.

In practice the cascaded and the power-filter NAEC are popular time-domain solutions because of their simple and relatively less complex structures. Hence, enhancing their efficiency is of vital importance for better NAEC performance. In order to highlight the advantages and drawbacks of each of them, authors in [49] have done a limited work in comparing these models but not with real loudspeaker data. Indeed, this evaluation is necessary as it will enable better understanding of the missing features in these two popular models and thus improved algorithms can be designed to deal with the complex nonlinear distortion in the acoustic echopath.

5.4.2 Experimental work

In the following we report a comprehensive performance comparison of the cascaded NAEC and the power-filter NAEC along with the linear AEC. In all cases tests were conducted using either real recorded nonlinear echo signals or echo signals synthesized empirically (the GPHM model as discussed in previous section) based on the real mobile phone loudspeaker responses. Performance is assessed in terms of the echo return loss enhancement (ERLE).

Test Set-up

Experiments were performed with diagonal (one-dimensional) loudspeaker Volterra kernels $h_p(n)$ for values of $p \leq 5$ and with $N_p = 64$ taps in all cases. The acoustic channel was modeled with a fixed 256-tap room impulse response (RIR) $h_{rir}(n)$ selected from the Aachen RIR database [1]. All experiments were performed with a clean speech

	Multiplications	Additions
Linear AEC with NLMS	3L+1	3L
Power Filter model NAEC	(3L+1)*P	3L*P
Cascaded model NAEC	[(3Np+1)*P]+2N_p^2	[(3Np+1)*P]+2(N_p-1)^2
	3N+1	3N

Table 5.1 – Computational Complexity Comparison

downlink/reference signal $x(n)$ of approximately 30 seconds duration with a sampling frequency of 8kHz. A set of common filter parameters were applied to all three AECs and were chosen to maintain stability and better performance. The three AECs and their configurations are: the linear NLMS AEC ($L = 319$, $\mu = 0.5$, $\delta = 1e-7$), a power-filter model NAEC without orthogonalization ($L_{p=1,\cdots,5} = 319$, $\mu_1 = 0.5$, $\mu_{p=2,\cdots,5} = 0.01$, $\delta_1 = 1e{-}7$, $\delta_{2,\cdots,5} = 1e{-}4$) and a cascaded model NAEC ($N_{p=1,\cdots,5} = 64$, $\mu_{p=1,\cdots,5} = 0.01$, $\delta_{p=1,\cdots,5} = 1e-4$, $N = 256$, $\mu = 0.5$, $\delta = 1e-7$)

Case 1: Computational complexity

Number of computations (multiplications and additions) required for each iteration of the NLMS algorithm based three AECs viz. linear AEC, power-filter model NAEC, cascaded model NAEC are compared in Table 5.1. Power-filter model contains P parallel channels each of length $L_p = L \,\forall\, p$ hence the computations increased P-times of linear AEC. Each iteration of the cascaded model NAEC of sub-filters length N_p and linear filter length N requires more computations because of the matrix-vector multiplication in its sub-filters update equation as shown in Eq. 5.14. So before examining the experimental work, one advantage of power-filter model over cascaded model is its relatively computationally simple structure.

Case 2: Echo cancellation with empirical nonlinear echo signals

Here we investigate the performance of the three AECs in the presence of fifth-order empirical nonlinear echo. The results are illustrated in Fig. 5.11 with the label 'Nonlinear Echo'. As shown in figure, the power-filter model NAEC outperforms both the linear AEC and the cascaded model NAEC. In contrast to the literature [49], the linear AEC outperforms cascaded model NAEC in the presence of empirical nonlinear echo. This is due to the fact that the cascaded model requires pre-processor and linear filter adaptation using a single joint error signal $e(n)$. As a result the convergence of both filters is interdependent, which leads to possible errors. Also, the cascaded model doesn't have the ability to prevent falling into local minima.

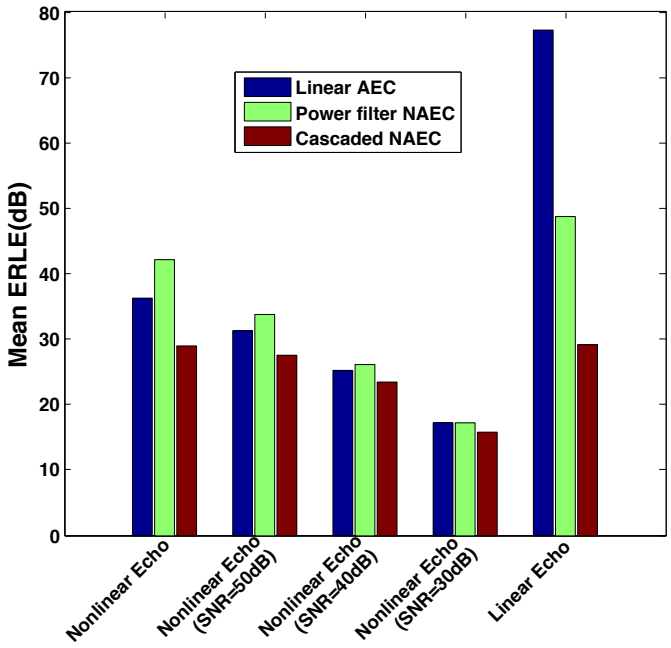


Figure 5.11 – A performance comparison of the three AECs in terms of mean ERLE in the presence of different echo scenarios

Case 3: Effect of linear echo on NAEC

The nonlinear distortion caused by the nonlinear systems like mobile phone loudspeaker is not constant and often time variant. How does a NAEC perform in the absence of nonlinearities (when the echopath is linear)? Therefore, in this case, the performance of the three AECs are analysed in the presence of linear echo. The results are illustrated in Fig. 5.11 with the label 'Linear Echo'. As expected, the linear AEC outperforms both NAEC models. In this situation, the NAEC models contain unnecessary coefficients (over-modeling) and that overestimating the linear echo signal results in suboptimal performance, as a consequence of the large amount of gradient noise introduced by the adaptation of the coefficients. However, The power-filter model is relatively better compared to the cascaded model.

Case 4: Effect of background noise

Here we analyse the performance of the three AECs in the presence of both the fifth-order nonlinear echo and the background noise. To simulate this scenario, a white Gaussian noise signal of different SNR levels has been added to the fifth-order nonlinear echo signal. The results are illustrated in Fig. 5.11 with the label 'Nonlinear Echo' and their corresponding SNR level. The echo cancellation performance of the three AECs decreases with increasing noise variance level.

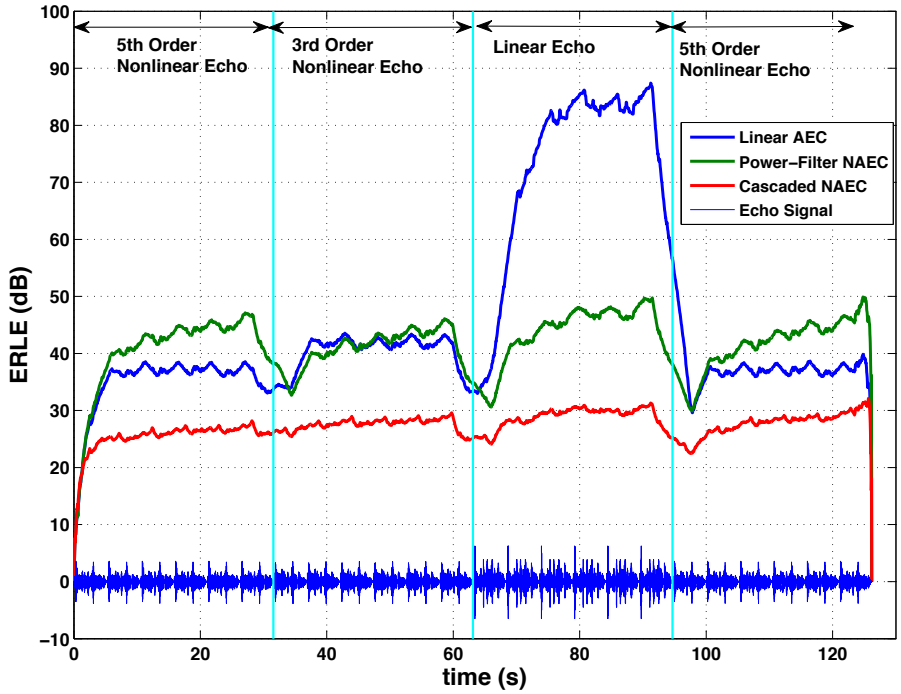


Figure 5.12 – A performance comparison of the three AECs in terms of ERLE in the presence of dynamically varying orders of nonlinear echo

Power-filter model NAEC is more sensitive to noise compared to the other two AECs. In moderate or high noisy environments (SNR $\leq$ 30), the linear AEC and the power-filter model NAEC show similar behavior. On the other hand, the cascaded NAEC is relatively less sensitive to the presence of noise.

Case 5: Performance with varying order-of-nonlinearity P in the echo signal

It has been observed while making many recordings using popular mobile phones that the waveform distortion caused by the loudspeaker's nonlinear distortion is not constant throughout the waveform but instead time-varying (more details on this are given in Chapter 7 of this thesis). Hence, in this case, the principle is to evaluate and compare the efficiency of the linear AEC and the two fifth-order NAECs under dynamically varying orders of nonlinear distortion in the acoustic echopath. Fig. 5.12 illustrates the performance comparison if the order of the nonlinear echo is changing from five to three after 30 seconds and then to linear after 60 seconds and going back to five after 90 seconds.

The power-filter NAEC clearly shows its ability to handle nonlinear echoes in the exact-modeling scenario (the model of the NAEC matches with the true echopath). However, it's performance is plagued by any changes in the echo signal. Upon each

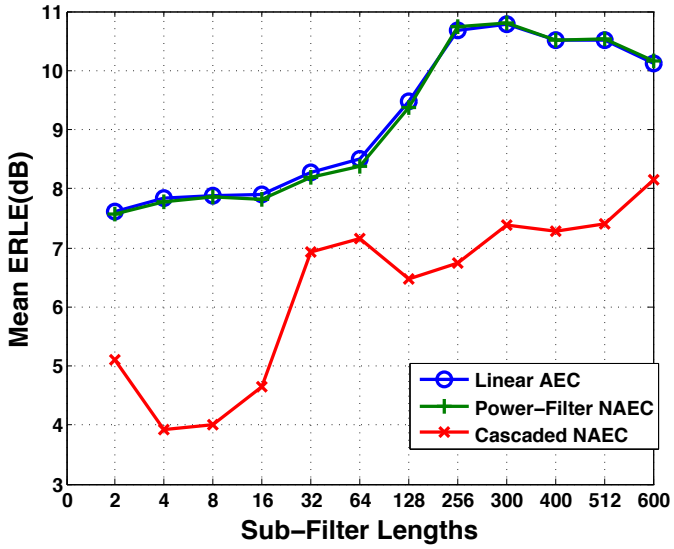


Figure 5.13 – A performance comparison of the three AECs in terms of mean ERLE in the presence of real recorded nonlinear echo as a function of pre-processor filter lengths

change in the order of nonlinear echo we observe that the power-filter NAEC exhibits poor convergence than the linear AEC. This is due to its large filter-tap lengths. Linear AEC is more robust to the changes and whenever the order of nonlinearity reduces it attains faster convergence and continuously provides an improvement of the echo cancellation performance. Linear AEC performance is obviously the best if the echopath is roughly linear, or contains negligible nonlinearities. Except for the initial convergence phase, the cascaded NAEC continuously provides poor performance compared to the other two algorithms. However, the cascaded NAEC is robust to changes in the order of nonlinear echo. In all the cases, upon initialization the convergence speed is faster for cascaded NAEC just because it has fewer taps in its linear filter.

Case 6: Impact of the pre-processor filter lengths

As discussed in earlier sections, choosing the right acoustic model parameters for any AEC is equally challenging like its design. Here, we investigate the impact of the pre-processor filter lengths on the three AECs in the presence of real-recorded nonlinear echo signal. In this setting, a response of a real mobile phone loudspeaker (in hands-free mode at maximum volume) to the speech signal is recorded using a very good quality microphone mounted in front of the mobile phone at a distance of 30cms. All the recordings were made in a very low-reverb Vocal booth. Ignoring the room effect on the loudspeaker recording, a nonlinear echo signal is generated by convolving the recorded speech and a 256 taps RIR. In this scenario, selection

of the order (P) of the NAEC and the length of each filter is subject to a trade-off involving performance and computational cost.

Fig. 5.13 illustrates the performance of the linear AEC and the two fifth-order NAECs as a function of the pre-processor filter lengths of the nonlinear channels. The linear filter length (N) is set at 256 taps. From the results, it is clear that increasing the pre-processor filters length improves the performance of the linear AEC and the power-filter NAEC up to certain extent and then the performance drops due to gradient noise produced by the over-modeling adaptation. Further, there is no noticeable difference between the linear AEC and the power-filter NAEC, this is because of the sensitivity to the background noise (noise floor) in the recorded speech signal. It is observed that the *Signal-plus-Noise to Noise Ratio* (SNNR) of the real recorded speech signal is around $25dB$. On the other hand, the performance of the cascaded NAEC is completely random, showing it's inability to achieve uniqueness in it's convergence. It can also be observed that the mean ERLE of both the NAEC algorithms are very low because in the presence of real recorded nonlinear echo, the NAEC models used to estimate the nonlinear distortion did not match the real distortion, thus rendering the poor performance.

Although so many efforts have been devoted to NAEC problem, the performance of most of the NAEC algorithms in practical acoustic environments still cannot meet expectation. Apparently, the dynamics of the nonlinear systems has not yet been fully realized as expected. The reasons behind this are very sophisticated and have not been fully understood thus far. Therefore, further research in this area is indispensable.

5.5 Summary

In this chapter we have provided a coherent and concise introduction to the state-of-the-art NAEC and/or NRES algorithms proposed to handle the nonlinear distortion in the LEMS. We outline an overview of both the hardware-based and the software-based solutions. In particular, we enumerated and described the software-based solutions by dividing them into three categories: nonlinear pre-filtering, nonlinear post-filtering and nonlinear adaptive filtering. We then analysed and discussed the technical issues involving the implementation of the available state-of-the-art solutions in each category.

Furthermore, we have identified that the convergence and the stability performance of the widely used NAEC algorithms have not been fully explored in the literature. Thus, we have provided a comprehensive performance analysis of the selected popular algorithms from the literature. The results from our experimental evaluations demonstrate that, while much valuable work has been accomplished in the literature, currently available NAEC

solutions are not suitable for most of the practical acoustic environments. Accordingly, there is a considerable potential for further development of NAEC techniques and a need for the utilisation of novel techniques to achieve a more effective nonlinear system analysis and design methodology.

Chapter 6

Empirical Mode Decomposition

Miniature loudspeakers, often used for mobile devices, are generally nonlinear systems associated with multiple nonlinear effects including electronic, magnetic, mechanical and sound. Traditional signal analysis techniques such as the Discrete Fourier Transform (DFT) and the Wavelet Transform (WT) were designed for linear signals (signals resulting from linear systems) and rely on a priori defined bases for data representation. These approaches are ill-suited to the analysis of nonlinear signals, and thus direct application of these approaches to nonlinear systems may lead to undesirable affects (like spreading of energy into high-frequency components) and unrelated physical interpretations.

Huang et.al. [18] proposed a novel engineering tool, known as Empirical Mode Decomposition (EMD), for systematic signal analysis and synthesis of nonlinear and nonstationary data. As an alternative to Fourier-based approaches, in this thesis we have studied the application of EMD to nonlinear signal processing. Theoretical aspects of the EMD are reviewed and its extensive field of contemporary applications are discussed in this chapter. This chapter also reports our novel solution to NAEC based on EMD. This work was published in [93].

6.1 Why study EMD?

Traditional data analysis techniques such as Fourier approaches are all based on assumptions of linearity and (short-term) stationarity. Wavelet analysis [94] was designed to handle nonstationary data, but still assumes linearity. Common to both of these techniques is the definition of standard and/or a priori defined bases for signal representation. The concept of *eigenfunctions* plays an extremely important role in the study of such traditional signal analysis techniques. In general, decomposition of signals are based upon linear combination of the eigenfunctions of linear systems [95]. In contrast, nonlinear

systems in general do not have a common set of eigenfunctions. Hence traditional signal analysis approaches are ill-suited to the analysis of nonlinear signals, and their direct application may lead to undesirable affects and unrelated physical interpretation.

The analysis of nonlinear and nonstationary data, however, necessitates data-dependent bases or, equivalently, adaptive bases [18]. Empirical mode decomposition (EMD) [18, 96] is one approach which meets this requirement of data-dependent basis functions necessary for adaptive data analysis. The motivation for performing EMD is to adaptively decompose nonlinear and/or nonstationary data into a set of elementary signals, referred to as Intrinsic Mode Functions (IMFs), in an ad-hoc manner without any a priori information. The IMFs retain the characteristics of the nonlinear input data and can reveal oscillatory trends that are not easily visible in the original input signal. This signal decomposition is not haphazard; the direct summation of the IMFs will (re)produce the original signal [97]. This allows the IMFs themselves to be used for processing and manipulation to effectively improve the input signal enhancement. As an alternative to the Fourier-based approaches, we apply this methodology to nonlinear loudspeaker analysis and to nonlinear echo signals.

6.2 Introduction to EMD

A relatively recently developed technique, empirical mode decomposition (EMD) assumes any signal is composed of different modes of oscillations. This may be considered as faster oscillations locally (in time) overlying to slow oscillations [18, 98]. Fig. 6.1 illustrates such an idea. The working principle of EMD is to iteratively break down a complex signal into finite and a usually very small number of empirical modes, referred to as intrinsic mode functions (IMFs), without leaving the time domain. IMFs are called empirical modes because they are neither pre-defined nor in a particular transform domain as is the case with traditional signal analysis techniques but are derived empirically based on the input signal. Accordingly, IMFs serve as adaptive basis functions of the EMD and each IMF represents a certain oscillatory trend (fast to slow) in the original signal. The original complex signal can be completely reconstructed by summing all the IMFs.

This section reviews EMD in a nutshell. There is abundant theoretical and empirical literature relating to the EMD and its use in applied sciences [18, 96]. All the details regarding the implementation of the EMD algorithm and the corresponding Matlab scripts are fully available in [99].

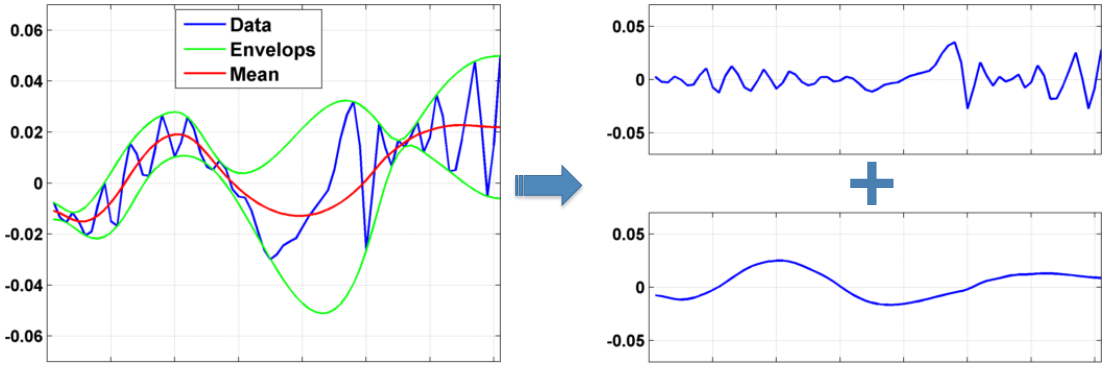


Figure 6.1 – An illustration of the basic idea of EMD. Illustrated is a given parent data (Blue line in the left figure) and is considered as faster oscillation (top figure on the right) overlying to slower oscillation (bottom figure on the right).

6.2.1 EMD Analysis: The Sifting Process

The ad-hoc adaptive data analysis procedure that EMD uses to extract IMFs from the original input signal is called *the Sifting process*. In representing and analyzing nonlinear and nonstationary signals, the basic approach of sifting process has been to decompose the input signals into a linear combination of empirical oscillatory modes (IMFs). The empirical modes or the IMFs help better understand the internal structure of the signal and the various components involved. For an elementary signal to be an IMF, it must satisfy the following two important properties [18]:

1. The number of extrema (maxima and minima) and the number of zero-crossings in the entire input signal (total duration of the signal) must either be equal or differ at most by one.

2. The mean value of the envelop defined by the local maxima and local minima is equal to zero at any point.

The EMD algorithm was originally proposed to overcome the limitations of the Hilbert transform, the latter will be introduced in the following chapter. The above two constraints admit the well-behaved Hilbert transforms. The first constraint eliminates the riding-waves[1] ensuring the local maxima of a signal are always positive and the local minima are negative, respectively. The second condition makes the waveform symmetric with respect to the origin by removing any unwanted fluctuations, which simplifies the data

[1]Riding-waves are the rapid oscillations with no zero crossings between the extrema. This causes positive local minima and negative local maxima in the signal. In general, riding waves are defined as transient signals that are interrupting the predominant pattern of the wave [97]

analysis by extracting the desired amplitude and frequency information without conflicting paradoxical results [100]. These conditions ensure that each IMF has a localized frequency content by preventing frequency spreading due to asymmetric waveforms [18].

The complete sifting process procedure to decompose a time series into a set of IMFs is illustrated schematically in Fig. 6.2 and is described below:

1. The sifting process starts by identifying all the local extrema points for a given parent (input) signal $x(n)$ as shown in Fig.6.3a.

2. Once all the local extrema points are identified, compute the upper envelope $e_{max}\,(n)$ and the lower envelope $e_{min}\,(n)$ by interpolating the local maxima and minima, respectively. The choice of interpolation method plays a key role in the decomposition. Different interpolation methods have been studied in detail and their effects on EMD algorithm have been compared in [101]. As recommended by the authors in the original work [18], we used a cubic-spline interpolation technique for all the work reported in this thesis.

3. Compute the local mean between the two envelopes as illustrated in Fig. 6.3b, given by:

$$m_1\,(n) = \frac{e_{min}\,(n) + e_{max}\,(n)}{2} \tag{6.1}$$

   where $m_1\,(n)$ is the local mean between the two envelopes after first sifting iteration.

4. Extract the residue, referred to as the detail signal ($d_1(n)$), defined by:

$$d_1\,(n) = x\,(n) - m_1\,(n) \tag{6.2}$$

   An example of such a detail signal is shown in Fig. 6.3c.

5. $d_1\,(n)$ can be considered as a first IMF if it satisfies a stopping criterion, the two properties of the IMFs as discussed above. If not, take $d_1\,(n)$ in place of $x(n)$ and repeat the sifting process k times, until $d_{1k}\,(n)$ is an IMF:

$$d_{1k}\,(n) = d_{1(k-1)}\,(n) - m_{1k}\,(n) \tag{6.3}$$

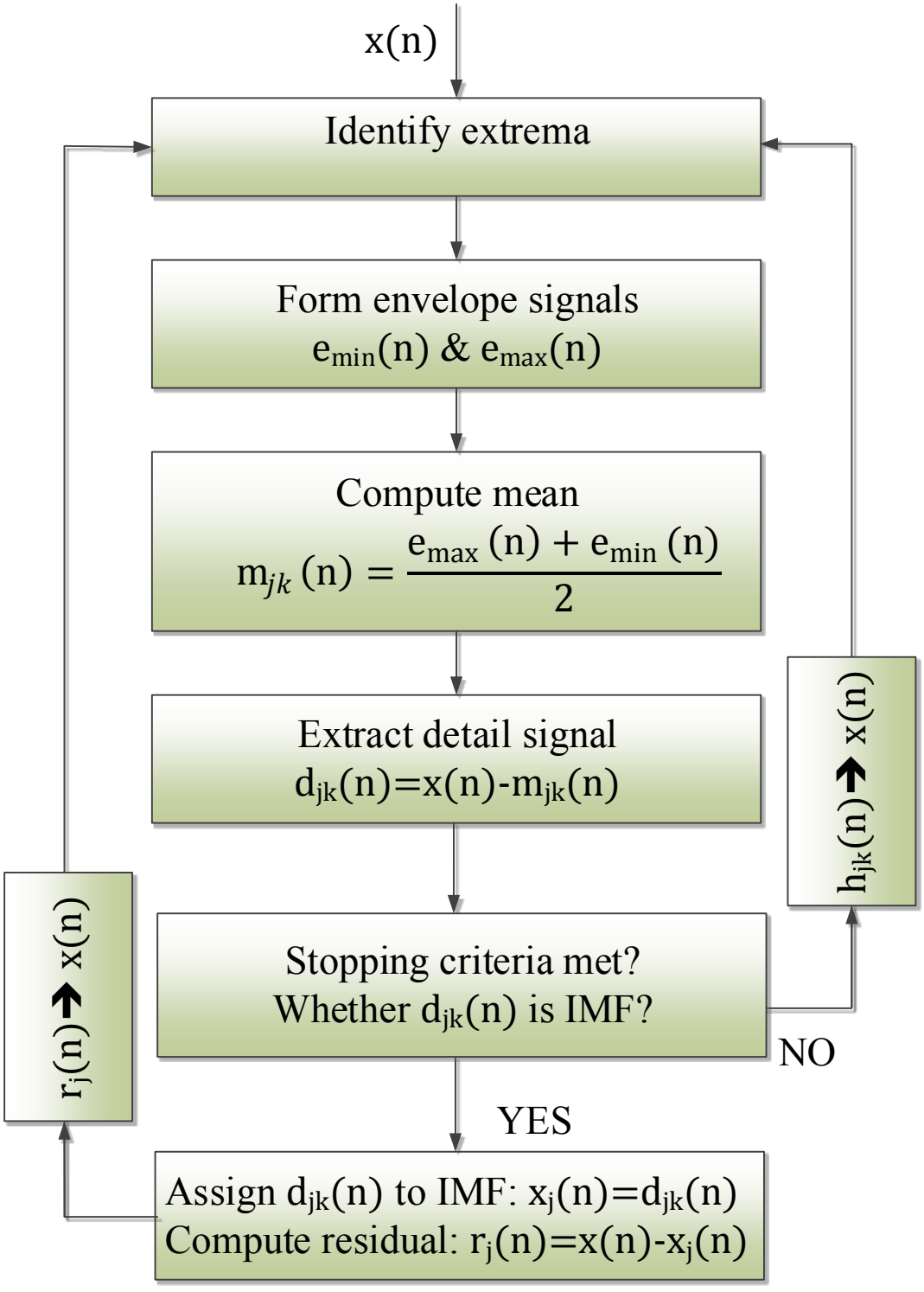


Figure 6.2 – The flowchart illustrating the sifting process procedure to decompose any complicated signal into a set of IMFs.

where $d_{1k}(n)$ and $m_{1k}(n)$ are the detail signal and the local mean of the envelopes, on the k^{th}-iteration of the sifting process respectively. Thus the first IMF component, $x_1(n)$, is given by:

$$x_1(n) = d_{1k}(n) \tag{6.4}$$

The first IMF component $x_1(n)$ is shown in Fig. 6.3d. It represents the fastest (or highest frequency) oscillatory mode in the parent signal, $x(n)$.

6. To obtain the next IMF, the first IMF, $x_1(n)$, is subtracted from the parent signal $x(n)$ and the difference signal $r_1(n)$ is used as a parent signal for a new sifting process.

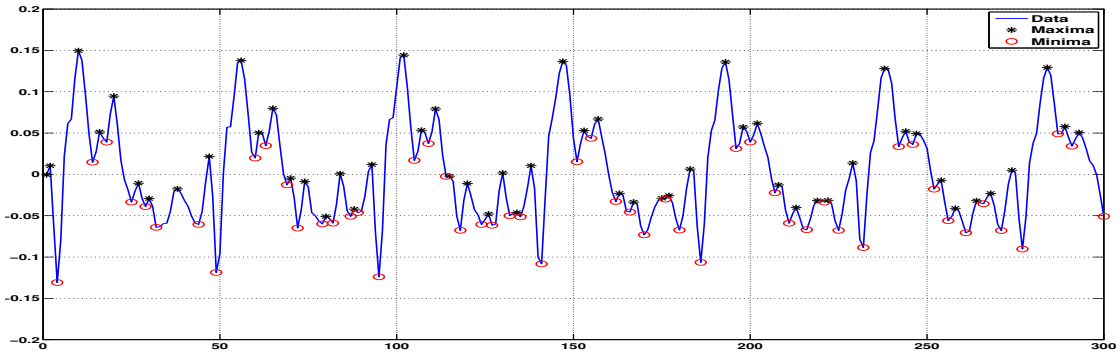
$$r_1(n) = x(n) - x_1(n) \tag{6.5}$$

7. The sifting process is then repeated on the difference signals until the residual signal $r_M(n)$ becomes a non-oscillatory monotonic function, i.e., it contains only two extrema:

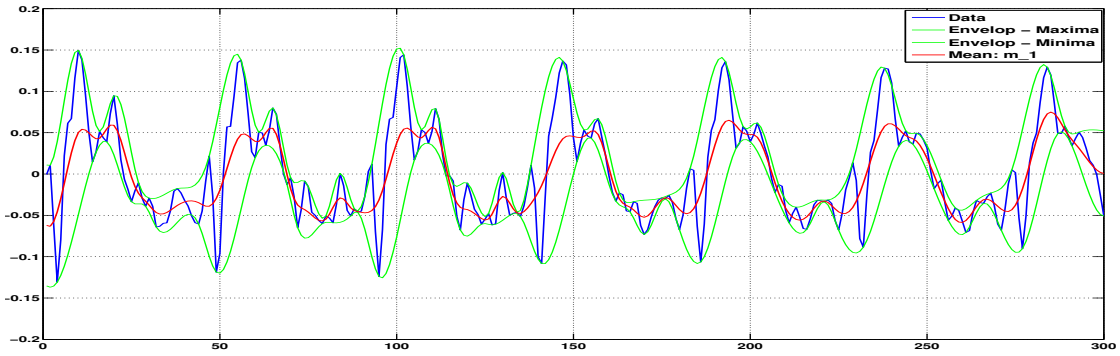
$$r_M(n) = x(n) - x_M(n) \tag{6.6}$$

where $x_M(n)$ is the M^{th}-IMF. Being a monotonic function, no envelops can be formed from $r_M(n)$ and no more IMFs can be extracted. We finally have M IMFs and a final residual signal $r_M(n)$.

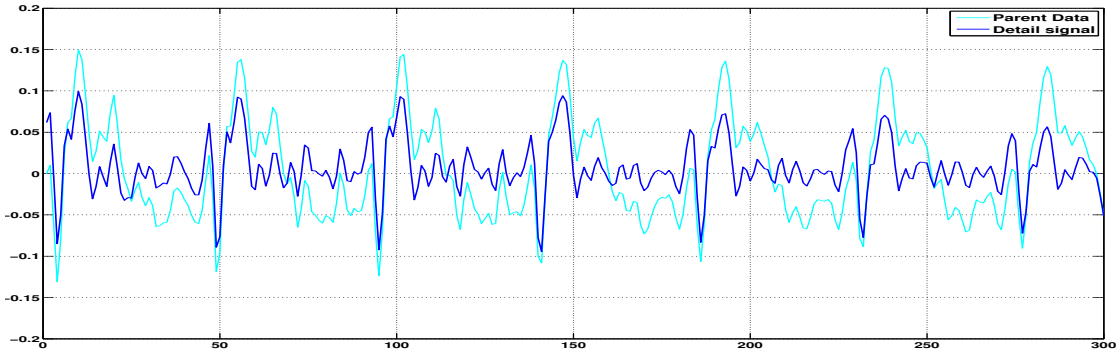
Fig. 6.4 illustrates a complete set of IMFs resulting upon application of EMD to the parent signal $x(n)$. The IMFs are iteratively derived starting from the highest frequency mode, IMF1, to the lowest frequency mode, IMF6. However, IMFs are not constant amplitude and frequency components like in Fourier analysis but may have amplitude modulation and also changing frequencies as shown in Fig. 6.4. The higher-order IMFs are subsequently smoother as we remove the high frequency components prior to their extraction. The residual signal (or the last IMF) represents the general trend in the input signal.



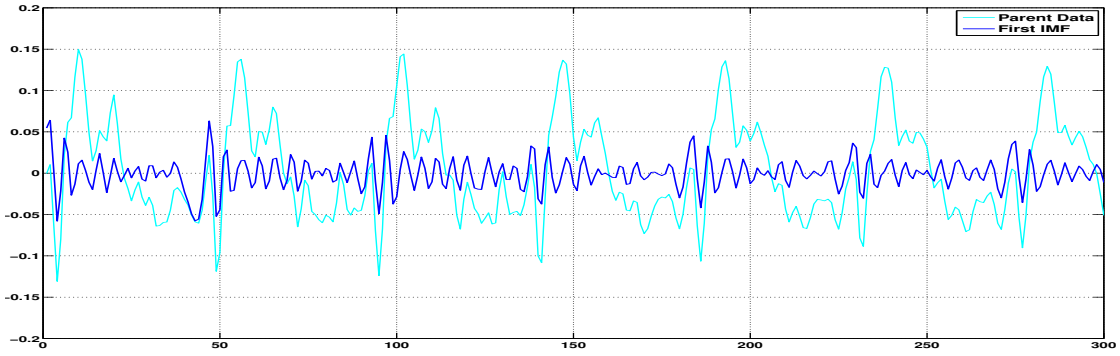
(a) Local maxima and local minima are identified for a given test signal



(b) Upper and lower envelopes are formed by interpolating (cubic-spline) local maxima and minima respectively. Illustrated also the local mean of the two envelopes.



(c) The parent data and the detail signal, $d_1\,(n)$



(d) The first IMF component is extracted from the parent data after multiple iterations of sifting process

Figure 6.3 – An illustration of the sifting process, which decomposes a test signal into a set of IMFs.

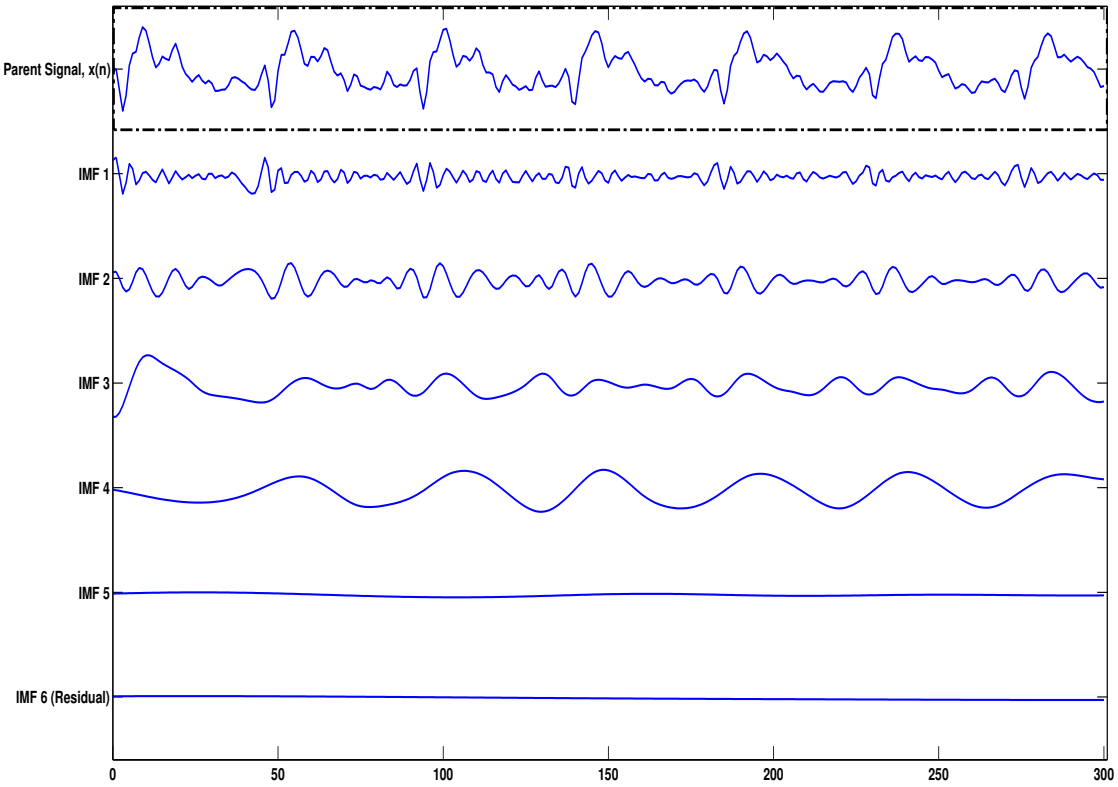


Figure 6.4 – An illustration of EMD. Illustrated is a given parent signal (top) and the resulting 6 IMFs.

6.2.2 The Stopping Criteria

The stopping criterion in Step 5 of the sifting process checks if the detail signal $d_1(n)$ satisfies the two IMF properties. In [102], it was argued that all types of signals cannot fit this IMF definition strictly, which often results in over-sifting. Over-sifting of IMFs leads to over-decomposition of signals causing spectral dispersion over adjacent oscillatory modes, whereas under-sifted IMFs tend to violate at least one of the IMF properties. Thus the choice of stopping criterion plays a key role in the accuracy of EMD. Analogous to system distance definition, the authors of the original EMD algorithm in [18] proposed a stopping criterion depending on the normalized Square Difference (SD_k) between two successive sifting operations:

$$SD_k = \sum_{n=0}^{L-1} \frac{|d_{1k-1}(n) - d_{1k}(n)|^2}{d_{1k-1}^2(n)} \tag{6.7}$$

where $d_{1k}(n)$ and $d_{1k-1}(n)$ are the two adjacent detail signals, of length L samples,

after k iterations of the sifting process. If this SD_k value is smaller than a pre-defined value then the sifting process can be stopped. Since this stopping criterion, SD_k, does not depend on the IMF conditions, choosing a threshold value lacks proper guidance. Imposing too low a threshold leads to over-sifting thereby causing over-decomposition. Hence, alternative stopping criteria for sifting process were proposed in [98, 103, 104]. In this thesis, we adopt a widely used stopping criterion first proposed in [98]. This stopping criterion compares the amplitudes of the mean of the envelopes and the corresponding detail signal. If the amplitude of mean of the envelopes is relatively small compared with the amplitude of the corresponding detail signal at all data points, then the sifting process is terminated. It is based on three parameter thresholds α, θ_1 and θ_2, which are purposed to ensure globally small variations in the mean, $m(n)$, while taking into account locally large excursions [98]:

$$a(n) = \frac{e_{max}(n) - e_{min}(n)}{2} \tag{6.8}$$

$$\sigma(n) = \left| \frac{m(n)}{a(n)} \right| \tag{6.9}$$

where $a(n)$ and $\sigma(n)$ are referred to as the *mode amplitude function* and the *evaluation function* respectively in [98]. The sifting process is iterated until $\sigma(n) < \theta_1$ for some prescribed fraction $1 - \alpha$ of the total duration, while $\sigma(n) < \theta_2$ for the remaining fraction. The typical values proposed by the authors are $\alpha \approx 0.05, \theta_1 \approx 0.05$ and $\theta_2 \approx 10\theta_1$ respectively (the default values in the EMD matlab scripts in [99]). If this stopping criterion is fulfilled, then the sifting process is terminated to give an IMF.

6.2.3 EMD Synthesis

The result of the sifting process produces M IMFs and a constant residue signal $r(n)$ ($=r_M(n)$, the subscript M is ignored for the sake of simplicity). The parent signal ($x(n)$) can be completely reconstructed through EMD synthesis process, which is simply a direct summation of physical domain IMFs generated by the EMD:

$$x(n) = \sum_{j=1}^{M} x_j(n) + r(n) \tag{6.10}$$

The sifting process has two postulates: 1) the input signal to the sifting process ($x(n)$) can be represented as a linear combination of its IMFs and 2) inputting an IMF to the sifting process results in just the input IMF with scaling factor 1. Hence, the IMFs can be remarked as eigenfunctions of the sifting process. Further, IMF components form a complete and "nearly" orthogonal basis for the input signal [18, 96, 98]. Thus, the fully data-driven and adaptability of the EMD method explains that it can be considered to be well accommodated for nonlinear and nonstationary data.

6.2.4 EMD Applications

EMD does the unsupervised signal decomposition based on local characteristic time scale of the data. Besides, EMD is adaptive, highly efficient and does not leave time domain. These properties of EMD were claimed to be well suited for nonlinear and nonstationary data and have prompted many researchers to investigate EMD method to various research fields. Accordingly, there have been hundreds of papers in the literature during the last decade dedicated to apply EMD technique to various engineering and non-engineering applications, for example, biomedical applications [100, 105, 106, 107], image processing and computer vision [101, 108, 109], meteorology and climate studies [110, 111], financial studies [112], ocean and seismic wave studies [113, 114], mechanical engineering [115, 116, 117] and many other diverse research areas.

Speech Processing using EMD

In the last few years, the application of EMD has also been extended to speech and audio signal processing. Before presenting our work on EMD-based NAEC, this section discusses other applications of EMD in speech and audio enhancement that are widely reported. Like many real-world signals, speech signals are also highly nonstationary, making traditional signal analysis dissatisfying due to dynamic variation of spectral content across the time. EMD is a better alternative suitable to analyse highly nonstationary signals like speech. An example of speech signal analysis using EMD is illustrated in Fig. 6.5. The input signal is a clean speech signal sampled at 8kHz. EMD decomposes the clean speech signal into 18 IMFs; the first 6 IMFs are shown in Fig. 6.5. The first IMF has a high-pass characteristic but also contains lower energy, low frequency content. The higher order IMFs have overlapping band-pass characteristics [118]. It is necessary to emphasize that the cut-off frequency between the consecutive IMFs is time varying and input signal dependent. Several efforts were recently made to use the IMFs for speech enhancement [119, 120, 121, 122]. In [121], linear predictive coding (LPC) analysis was computed on the IMFs in order to extract the resonant frequencies of the vocal tract (formants). The results are compared to the LPC analysis operated on original

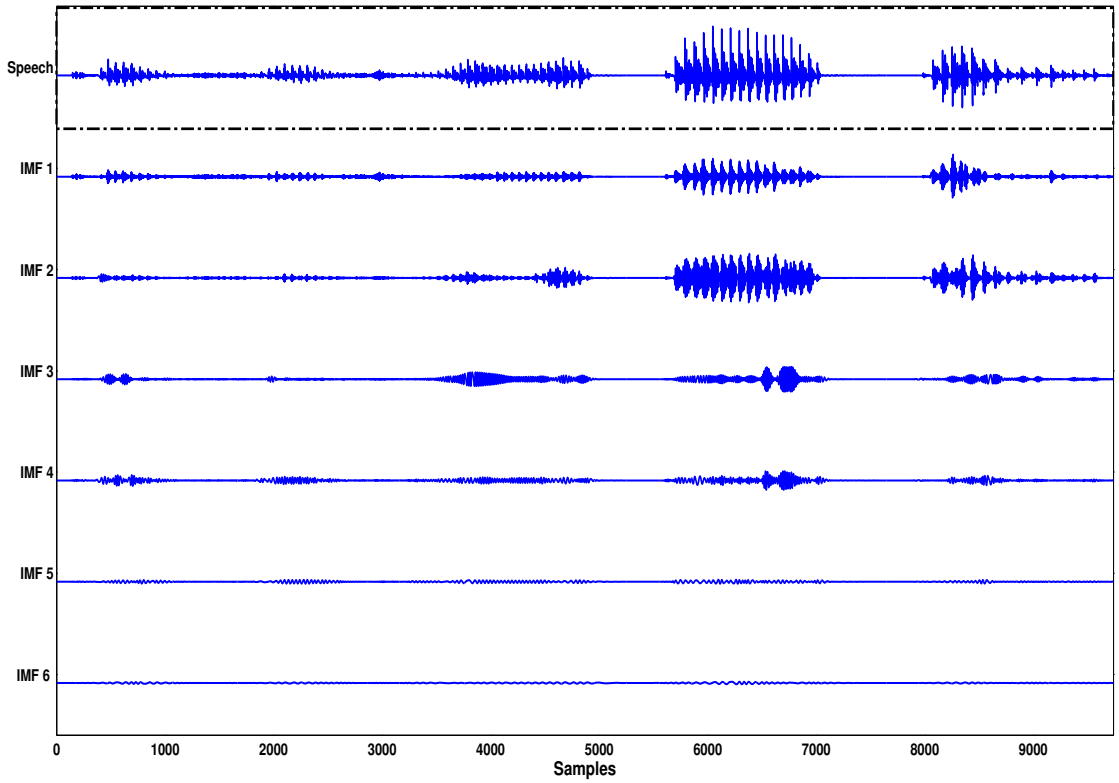
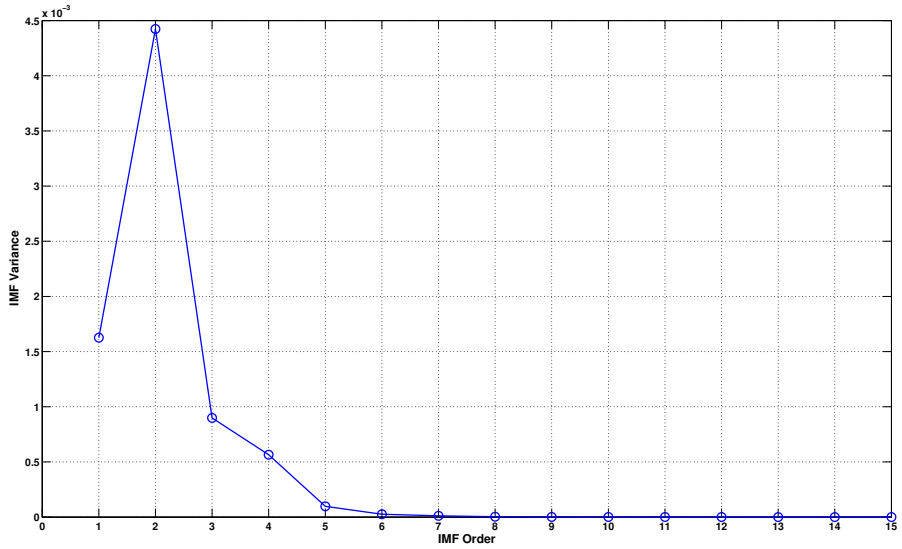


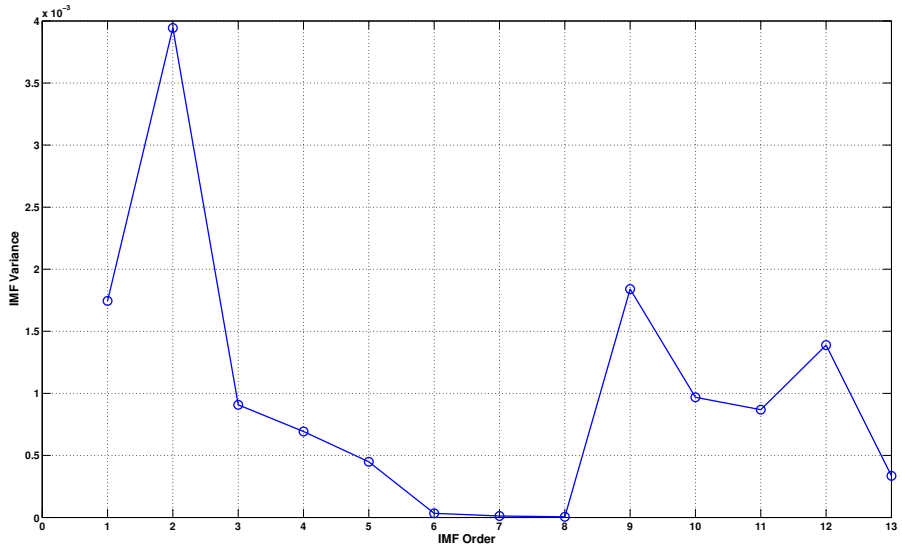
Figure 6.5 – An illustration of EMD. Illustrated is a clean speech signal (top) and the first 6 IMFs.

(full-band) speech signal and it was shown that the latter is more reprehensive than the same analysis operated on IMFs.

Further observing the IMFs in Fig. 6.5, it is clear that most of the energy content of the clean speech signal is concentrated in the first few IMFs. Fig. 6.6a illustrates the variance of IMFs with respect to the order of IMFs for a clean speech signal. The variance (or energy) of the IMFs decreases as the order of IMF increases. Similarly, if a speech signal is contaminated with a low-frequency noise (example: wind noise or car interior noise) then EMD has the capability to characterize the IMFs as either speech dominant or noise dominant. To illustrate such an example, a noisy speech signal was generated artificially by adding wind noise to the same clean speech signal shown in Fig. 6.5 at $0dB$ SNR level. Applying EMD to the noisy speech signal resulted in 12 IMFs. Fig. 6.6b illustrates the IMF variance plot of noisy speech signal. It can be seen that there is a sudden increase in the energy (variance) at higher order IMFs due to the low-frequency wind noise components. In [118], the authors showed that these observations hold true for most of the speech signals and low-frequency noisy signals. Thus, one way to reconstruct the speech signal from the noisy signal is by ignoring the higher-order noise-dominant IMFs and summing the first few IMFs composed of the



(a) IMF variance(energy) plot of a clean speech signal



(b) IMF variance(energy) plot of a speech signal contaminated with wind noise at 0 dB SNR

Figure 6.6 – IMF variance plots: indicates energy content in each IMF

desired speech signal. By taking advantage of this principle that separates the speech from the noise, approaches to EMD-based speech enhancement/noise cancellation are proposed in the literature [118, 119, 120].

Recently, an EMD-based sub-band approach to linear AEC is reported in [122]. In typical sub-band approaches to linear AEC, the far-end signal and the microphone output signal are divided into mutually exclusive multiple sub-band signals using identical analysis filter-banks[2]. The resulting sub-band signals are down-sampled by a known factor and

[2]A filter-bank is a bank of band-pass filters that divides the input signal into a set of sub-band signals

each down-sampled sub-band far-end signal serves as an input to an independent adaptive filter. The outputs of the sub-band adaptive filters are subtracted from the sub-band microphone signals forming the sub-band errors. These errors are then up-sampled and combined using synthesis filter-banks leading to the full-band echo-free output. This sub-band adaptive filtering technique allows for fast convergence and reduced complexity through the use of a robust NLMS algorithm, especially in longer reverberant acoustic environments [2]. However, in practice, the performance of sub-band adaptive filtering techniques is often degraded due to artifacts introduced by the filter-banks [123]. Authors in [122] presented an EMD-based sub-band adaptive filtering scheme to reduce the filter-banks artifacts. This structure uses EMD in the place of filter-banks, but there is no guarantee that the same number of IMFs with the same bandwidths will be produced at both downlink (far-end) and uplink (microphone) signals using EMD. Hence, the authors proposed to use an IMF separation process after EMD, where IMFs are grouped into separate bands according to the power spectral densities (PSD) of the IMFs. A detailed description about IMFs grouping is provided in [122]. The different IMF groups are treated with multiple adaptive filters similar to the traditional sub-band adaptive filtering techniques. The results presented show effective ERLE values with a faster convergence rate using the proposed EMD-based structure.

Other applications of EMD in speech and audio processing include speech analysis [124], source separation [125], voice activity detection [126] and pitch estimation [127].

6.3 Application of EMD to NAEC

This section reports the first EMD-based approach to NAEC. The work aims to demonstrate the application of EMD in the time domain as a potential solution to NAEC.

Before going into the details about EMD based NAEC, a real nonlinear echo signal is analysed using EMD, assuming that the downlink path is the only source of nonlinear distortion in the LEMS. A clean speech signal $x(n)$ sampled at 8kHz is played by a real mobile phone loudspeaker in hands-free mode at full gain. The output signal $x_{out}(n)$ is simultaneously recorded using a high-quality microphone. The experiment uses a setup similar to the one described in Section 4.1.1. The time series signals $x(n)$ and $x_{out}(n)$ (both normalized) are illustrated in Fig. 6.7. As a first step, EMD is applied to both $x(n)$ and $x_{out}(n)$. Consider the IMF variance plots shown in Fig. 6.8 for $x(n)$ and $x_{out}(n)$ respectively. The plots in Fig. 6.8 reveal that the first few IMFs of the real recorded signal, $x_{out}(n)$, have more energy than those of the clean speech signal, $x(n)$,

each of which corresponds to a dissimilar spectral region of the input signal. Filter-banks concept is discussed in detail in [80]

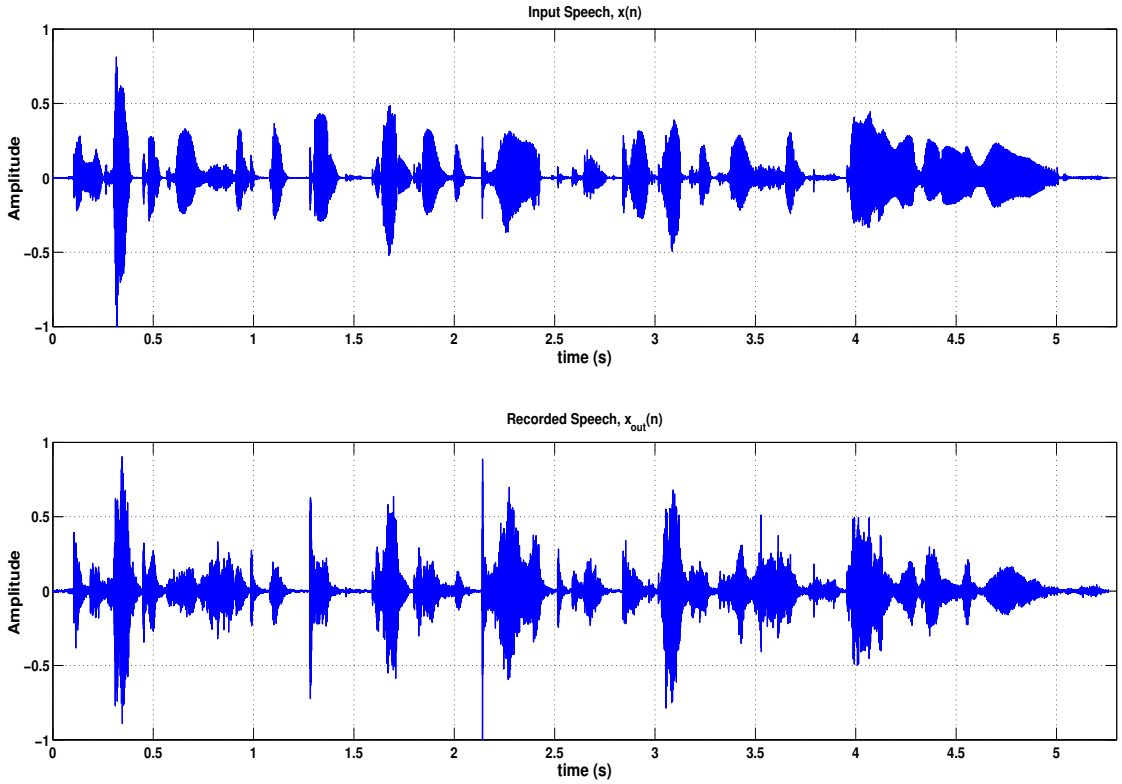


Figure 6.7 – A real mobile phone loudspeaker (in hands-free mode at maximum gain) excited by a clean speech signal (top) and simultaneously recorded using a high quality microphone (bottom)

this is because of the inherent loudspeaker nonlinear distortion. Also, the higher order IMFs of the recorded signal ($x_{out}(n)$) have its energy reduced because of the mobile device loudspeakers intrinsic inefficiency at low frequencies. Further, the first few IMFs of the recorded signal cover the higher-frequency band ranges from approximately 1kHz to 4kHz, the bandwidth which typically contains the majority of the higher-orders of nonlinear echo components. After first few IMFs, other IMFs are predominant with linear echo components.

Thus the data-adaptive EMD technique is more suitable to decompose a nonlinear distorted signal into a set of IMFs which can further be characterized as either nonlinear dominant or linear dominant. This IMF classification into nonlinear and linear dominants is one of the key factors of the nonlinear echo cancellation. It makes possible to eliminate the nonlinear processing for linear echo dominant signals (to avoid over-modeling) without degrading the linear AEC performance (due to gradient-noise). By taking advantage of this principle, we propose a novel scheme of NAEC.

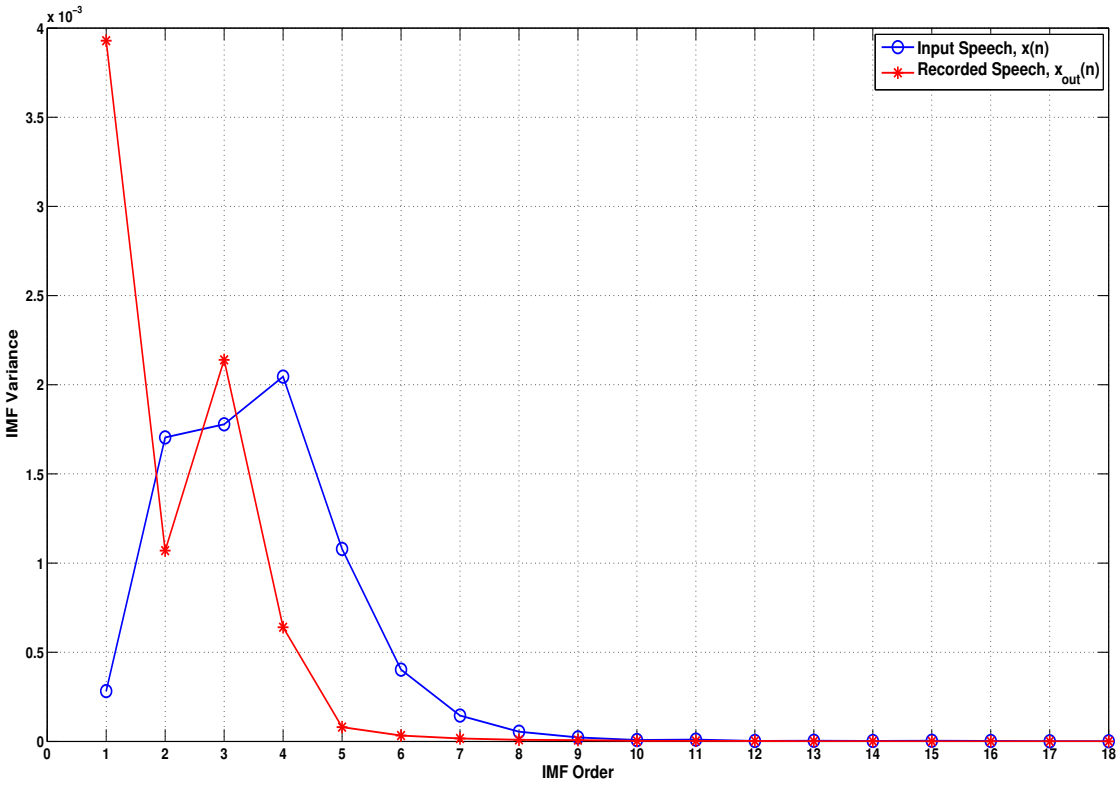


Figure 6.8 – A comparison of IMF variance plots of a clean speech signal and a real loudspeaker recorded signal

6.3.1 NAEC Structure

The approach is based on the decomposition of a full-band microphone signal into IMFs using EMD. NAEC is accomplished through the application of conventional, adaptive power filtering (parallel approach) to each IMF using a full-band reference signal $x(n)$. The structure of the new EMD-based NAEC scheme illustrated in Fig. 6.9 is essentially standard except for EMD decomposition, resynthesis and the use of multiple filter chambers (FCs). The downlink/reference signal is denoted by $x\,(n)$, the loudspeaker output signal by $x_{out}\,(n)$ and the uplink/microphone output signal by $y\,(n)$. In this first attempt to employ EMD for NAEC we suppose no near-end speech and no background noise. The uplink signal thus contains echo alone.

The microphone output $y\,(n)$ is decomposed by EMD into M IMFs according to the approach described in Section 6.2. Each IMF is then adaptively estimated from the full-band downlink/reference signal $x\,(n)$ by one of M filter chambers (FCs). Each FC contains the P^{th} order conventional power filter model [48] illustrated in Fig. 5.10. The power filter model is relatively an efficient approach to the identification of nonlinear acoustic echo paths as discussed in Chapter 5. The sub-filters adaptively estimate the

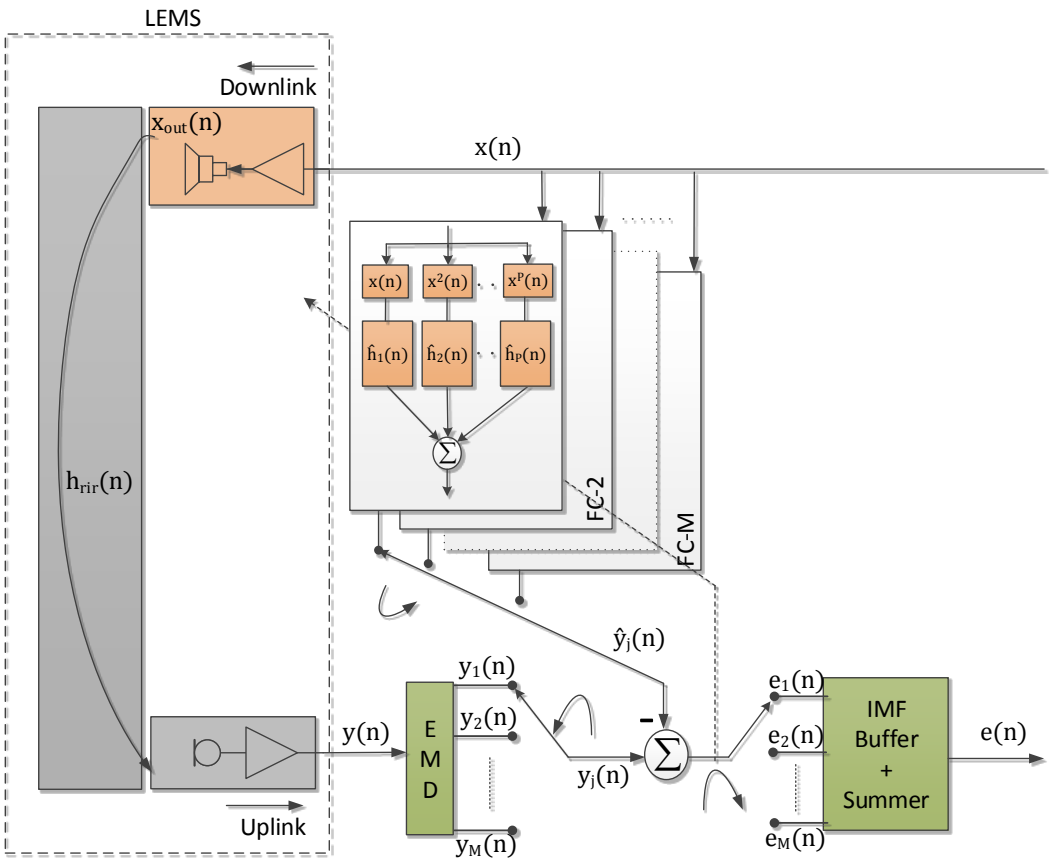


Figure 6.9 – Structure of EMD based NAEC.

acoustic channel and loudspeaker impulse response, collectively referred to as the LEMS illustrated in Fig. 6.9.

Decomposition of the microphone signal $y(n)$ produces M IMF signals $y_j; j = 1, \cdots, M$ where each IMF represents a distinct frequency range. Accordingly, each corresponding FC requires fewer filter taps than would otherwise be required in the case of a full-band signal. Featuring a frequency dependent control on IMFs, the order of the power filters, P, can be adjusted individually in each FC according to the spectral properties of the corresponding IMF. This structure also gives an additional degree of freedom to choose the system parameters of the power filter model (such as the order of nonlinearity P, length of sub-filters L_p, adaptive filters parameters, etc.) in each FC in accordance with the spectral range of the IMFs. The output of each FC, $\hat{y}_j(n)$, is subtracted from the corresponding IMF, $y_j(n)$, thereby generating individual error signals $e_j(n)$. Each error signal is used in the conventional manner to update FC sub-filter coefficients $\hat{h}_{p,j}(n); p = 1, \cdots, P$ and $j = 1, \cdots, M$. The parameter j in $\hat{h}_{p,j}(n)$ is ignored for the rest of the chapter for the sake of simplicity and will be referred to as $\hat{h}_p(n)$. Finally, the

individual error signals are summed together to reconstruct the full-band error signal:

$$e(n) = \sum_{j=1}^{M} e_j(n) \tag{6.11}$$

6.3.2 Adaptive filtering

EMD produces a total of M IMF signals $y_j; j = 1, \cdots, M$. Corresponding error signals $e_j; j = 1, \cdots, M$ are thus expressed by:

$$\begin{aligned} e_j(n) &= y_j(n) - \hat{y_j}(n) \\ e_j(n) &= y_j(n) - \sum_{p=1}^{P} \hat{\mathbf{h}}_p^T(n)\, \mathbf{x}^p(n) \end{aligned} \tag{6.12}$$

where $\hat{\mathbf{h}}_p(n)$ is the estimated sub-filter vector of length L_p, $\mathbf{x}^p(n) = [x^p(n), \ldots, x^p(n - L_p + 1)]^T$ is the reference signal vector and $\hat{y_j}(n) = \sum_{p=1}^{P} \hat{\mathbf{h}}_p^T(n)\, \mathbf{x}^p(n)$ is the output of the j^{th} FC. Due to its simplicity we used a normalized least mean square (NLMS) adaptive filtering algorithm within each FC. The NLMS algorithm for sub-filter $\hat{\mathbf{h}}_p(n)$ is derived using an approach similar to that given in [47]. Updates are applied in the usual manner according to:

$$\hat{\mathbf{h}}_p(n+1) = \hat{\mathbf{h}}_p(n) + \frac{\mu_p}{\|\mathbf{x}^p\|_2^2} \mathbf{x}^p e_j(n) \tag{6.13}$$

6.3.3 Experimental work

The following reports a performance comparison of the new EMD-based approach to NAEC to a baseline power filtering approach. All experiments were conducted with speech signals and the nonlinear echo signal is generated empirically (using the GPHM model by identifying a real mobile phone loudspeaker) as discussed in previous chapters (refer Section 4.1). Performance is assessed in terms of the echo return loss enhancement (ERLE).

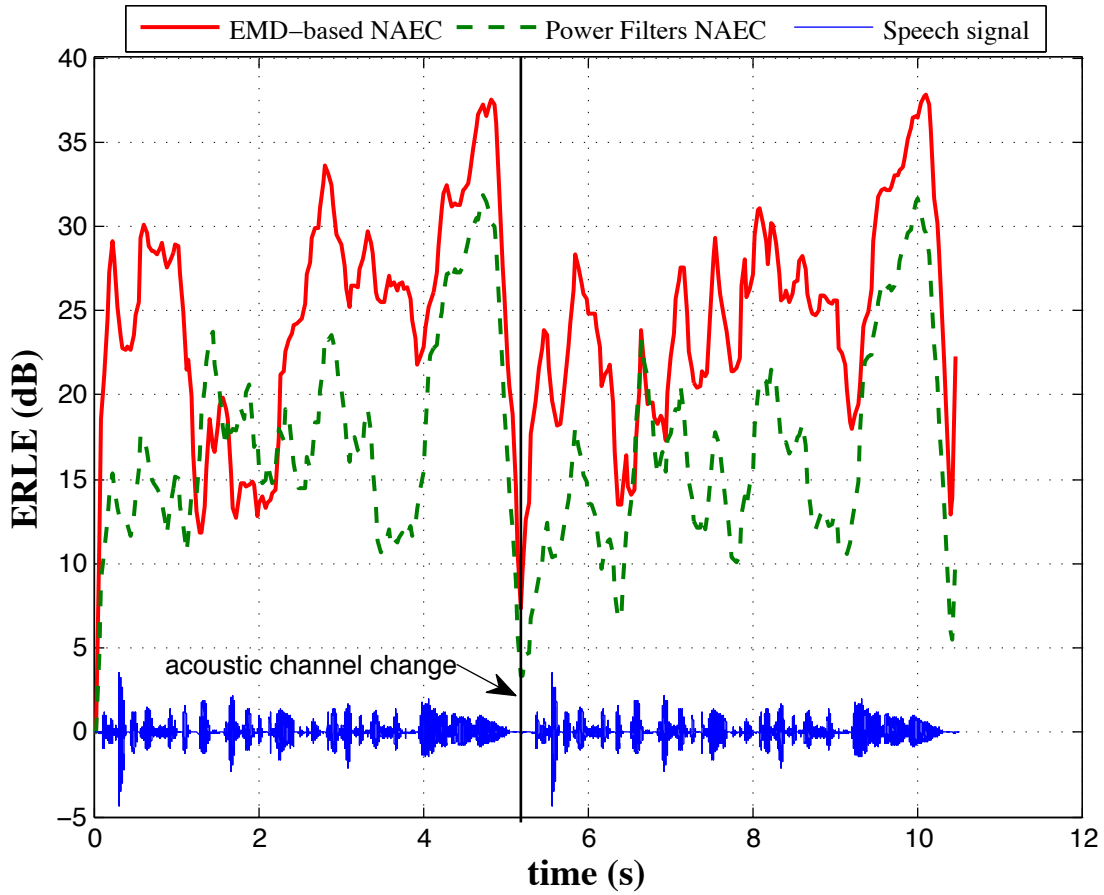


Figure 6.10 – A performance comparison in terms of ERLE for the new EMD-based approach to NAEC and a baseline power filter approach.

Experiments were performed with diagonal (one-dimensional) loudspeaker Volterra kernels $h_{p_L}(n)$ for values of $p \leq 5$ and with 32 taps in all cases. The acoustic channel was modeled with a fixed 256-tap room impulse response (RIR) $h_{rir}(n)$ selected from the Aachen RIR database [1]. All experiments were performed with a clean speech downlink/reference signal $x(n)$ of approximately 10 seconds duration with a sampling frequency of 8kHz. A change in the acoustic channel is introduced after approximately 5 seconds simply by delaying the RIR by 2.5 ms. This is done to compare the dynamic re-convergence performance of each algorithm.

We used the EMD routines available in [99] for decomposing the nonlinear echo signal into $M = 10$ IMFs. M varies for each speech signal; it depends on the stopping criteria used in the process outlined in Section 6.2. It is not the purpose of this thesis to address such issues which have been analyzed in detail elsewhere [18, 99]. Also, the energy content of the higher-order IMFs beyond 10^{th}-IMF are almost negligible. Accordingly, we have considered the 10^{th}-IMF is equivalent to the sum of all higher-order IMFs

FC	Sub-filter lengths				
	1(Linear)	2	3	4	5
1-4	128	32	32	32	32
5	128	32	32	32	X
6-7	128	32	32	X	X
8-10	128	X	X	X	X

Table 6.1 – Order of the power filters (P) and their associated sub-filter lengths (L_p, $p \in [1, P]$) in each Filter Chamber (FC)

from the 10^{th}-IMF ($IMF_{10} = \sum_{j=10}^{M} IMF_j;\ M > 10$). The order of the power filters (P) can be adjusted individually in each FC according to the spectral properties of the corresponding IMF. Over-modeling the order of power filters in the FCs increases computational complexity and the unnecessary degrees of freedom lead to noisy estimates $\hat{y}_j(n)$. For the test whose results are illustrated in Fig. 6.10, we have used different P for different FCs as illustrated in Table 6.1. Upon observing the spectral content of the IMFs, the first 4 IMFs covers the higher-frequency band ranges from approximately 1kHz to 4kHz, the bandwidth which typically contains significant nonlinear distortion. Hence, first 4 FCs each contain 5 adaptive sub-filters, $P = 5$. Similarly, the 5th FC has only 4 adaptive sub-filters whereas the 6th and 7th FCs have only 3 sub-filters[3]. The higher-order IMFs, from 8 to 10, correspond to the lower frequency range approximately less than 400Hz, which can be safely assume as linear. Therefore, nonlinear processing is not required for these IMFs and accordingly FCs 8–10 consist of a single, 287-tap linear transversal filter for linear echo cancellation. For all multi-channel FCs, the first sub-filter, which corresponds to the linear system response, has 128 taps. All other sub-filters have 32 taps.

Finally, the baseline power filter approach has $P = 5$ sub-filters, each with 287 taps. Neither the EMD-based nor the power filter approach uses orthogonalization since with the number of sub-filter taps used in these experiments, it does not improve performance [49].

6.3.4 Experimental results

ERLE results for the EMD and the baseline power filter approaches to NAEC are illustrated in Fig. 6.10 for a common excitation. The EMD approach is shown to outperform the baseline system; it attains a higher level of ERLE, around 8-10 dB more than the baseline. The use of different orders of power filters provides a convenient means of improving NAEC performance, thus minimizing gradient noise due to over-modelling.

[3]Assuming a miniatured loudspeaker low frequency cut-off at 200Hz, then the least possible 2^{nd} harmonic appears at 400Hz. Therefore, the frequency range below 400Hz is distortion free. Similarly, the frequency band below 600Hz could cause only a third order distortion

Fig. 6.10 also illustrates the response of each approach upon initialization and to a discrete change in the acoustic echo path which occurs at approximately 5 seconds. In both cases the EMD approach is shown to converge more rapidly than the baseline system. This is due to the lower spectral dynamic range in each IMF compared to the full-band signal in the baseline approach.

Notice the performance drop of the EMD approach at roughly between 1.5 and 2.5 seconds. The ERLE drop is a very good example and is due to one of the main limitations of the EMD sifting process, called "mode mixing". As discussed later in the thesis in Section 7.6, Orthogonality of the EMD is not guaranteed in some applications as often multiple IMFs are correlated, meaning different modes of oscillations (analogous to spectral content) coexist in multiple IMFs. This problem is called the "mode mixing" in the literature. The spectral content in that particular time period leaked into multiple IMFs (similarly to multiple FCs) which reduces the performance of those corresponding FCs leading to drop in the ERLE.

While the proposed EMD-based NAEC not only delivers greater average echo attenuation, faster convergence and thus better performance in the case of a dynamically changing acoustic path, it is not without cost. This entails increased computational complexity, principally due to the EMD decomposition and the use of multiple FCs. While there is scope to reduce the computational load via further optimization, the current system is approximately 1.8-times more demanding in terms of computation. While there is an on-line EMD algorithm [96], the work reported here was performed with an 'off-line' implementation, i.e. by application of EMD to entire signals. This was deliberate in order to demonstrate the application of EMD to nonlinear echo cancellation while avoiding additional problems inherent to on-line processing [96].

6.4 Summary

In hands-free telephony with low-cost transducers, the microphone signals are often nonlinear and nonstationary time series, hence their analysis is challenging. Linear signal analysis tools like Fourier analysis are ill-suited to analyse nonlinear signals. Empirical mode decomposition (EMD) was used in this thesis as a good alternative to analyse such complex nonlinear data. This chapter introduces and explains in detail the EMD technique. EMD is based on the local properties of the input signal and thus iteratively decomposes time series into different zero mean oscillations called intrinsic mode functions (IMFs). The IMFs themselves to be used for processing and manipulation to effectively improve the input signal enhancement. Various applications of EMD in the speech and audio signal processing are briefly discussed in this chapter.

This chapter also reports the first application of EMD to NAEC. The EMD solution entails the decomposition of the nonlinear echo signal from microphone into IMFs and their utilization in otherwise conventional echo cancellation using adaptive filtering. When compared to the power filter baseline system, experimental results demonstrate improved NAEC performance in terms of greater echo reduction and faster convergence. The proposed structure is also more robust to dynamic changes in the acoustic channel. While a modest increase in computational complexity is a drawback, there is scope to reduce this through further optimization.

Although the proposed EMD-based NAEC solution does not use Fourier analysis (or any linear signal analysis tools), the underlying interpretation of nonlinear distortion (as harmonic distortion) is still depends on the traditional Fourier based time-frequency analysis and the Volterra series. In continuation to the above described EMD algorithm, next chapter introduces a new method for time-frequency analysis called the Hilbert-Huang Transform (HHT). As an alternative method to the Fourier-based time-frequency analysis, we will apply this HHT methodology to analyse nonlinear loudspeakers in the next chapter.

Chapter 7

An Alternative Interpretation of Loudspeaker Nonlinearities

This chapter presents a new approach to nonlinear loudspeaker characterization using the Hilbert-Huang transform (HHT). Based upon the empirical mode decomposition (EMD) and the Hilbert transform, the HHT decomposes nonlinear signals into adaptive bases which reveal nonlinear effects in greater and more reliable detail than current approaches. Conventional signal decomposition techniques such as Fourier and Wavelet techniques analyse nonlinear distortion using linear transform theory. This restricts the nonlinear distortion to harmonic distortion. This work shows that real nonlinear loudspeaker distortion is more complex. HHT offers an alternate view through the *cumulative* effects of harmonics and intra-wave amplitude-and-frequency modulation. The work calls into question the interpretation of nonlinear distortion through harmonics and points towards a link between physical sources of nonlinearity and amplitude-and-frequency modulation. This work, published in [128], furthermore questions the suitability of traditional signal analysis approaches while giving weight to the use of HHT analysis in future work.

7.1 Time-Frequency Analysis

Of all commonly accepted practices for interpreting nonlinear distortion in loudspeakers, harmonic distortion is by far the most pervasive. Be it modeling, characterization or linearization of nonlinear loudspeakers, and/or nonlinear acoustic echo cancellation/suppression, most of all the concepts discussed in this thesis or published in the literature, presume nonlinear distortion as harmonic distortion. However, as Klippel mentioned in [9], harmonic distortion only provides a peculiar indication but not a comprehensive description of the nonlinear system.

Time-frequency-energy distribution (or the spectrogram) represents what frequencies are present in a signal, their energy and how they change over time. For example, Fig. 7.1

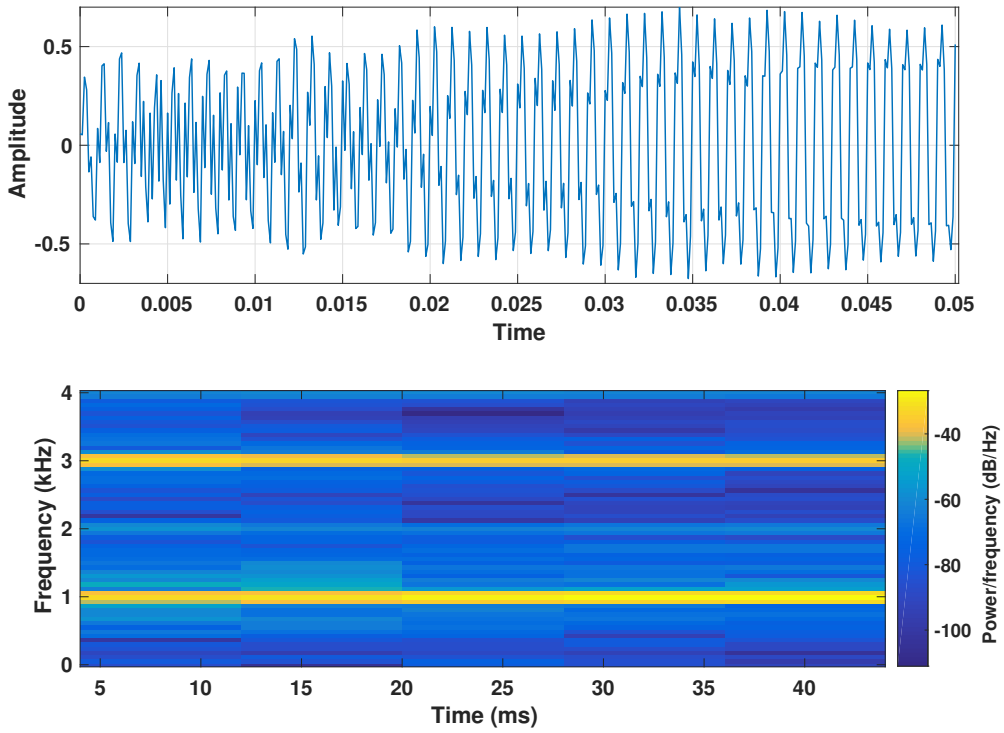


Figure 7.1 – Miniature loudspeaker (microspeaker) response to a pure sinusoidal input at 1kHz, sampled at 8kHz;

illustrates the Fourier-based spectrogram of the output of a miniature loudspeaker operating at max amplitude and excited with 1kHz sine tone. Harmonic distortion components appearing in the spectrogram indicate nonlinearities inherent in the loudspeaker under test. However, the conventional time-frequency analysis techniques are usually derived from a short-time Fourier transform (STFT) [80] or any generalized integral transforms like Gabor and wavelet transforms [94, 95].

Traditional Fourier-based signal analysis methods such as the discrete Fourier transform (DFT) and the STFT dominate the signal analysis field. These methods all assume linearity and stationarity, at least within the time window of observation. Since they rely on a priori defined orthogonal bases for data representation, Fourier-based approaches are ill-suited to the analysis of nonlinear signals; they assume the *linear* superposition of different signal components. As a consequence, the energy of a nonlinear signal is spread across a number of harmonics. Nonlinear distortion is then represented as harmonic distortion, even if the link to a physical source is questionable.

As an alternative to the STFT, Wavelet transform is developed for the improved visible localization of the frequency components in the analysed signals by means of variable width window (time-scale domain) [129]. However, Wavelet transform is also not unsupervised

or data driven, in that it relies very strongly on a parametric model of the input signal. Wavelet-based signal decomposition is characterized by the prior definition of the analyzing mother-wavelet (orthonormal basis function). Nonetheless, Wavelet analysis is designed to handle non-stationary data but still assumes linearity [98].

Although the two methods, Fourier and Wavelet analysis, are based on two different concepts, they both are designed for linear systems and/or signals. Therefore, direct application of these methods on nonlinear signals may lead to incorrect physical interpretation of the underlying nonlinear distortion. Accordingly, these mainstream methods may not be the most suitable approaches for the analysis of miniature loudspeakers. Alternative analysis methods are thus needed.

7.2 Instantaneous frequency and The Hilbert transform

To understand the complex behaviour of nonlinear systems, a thorough time-frequency analysis is necessary at the accuracy level of instantaneous frequency (IF) and instantaneous amplitude (IA). Instantaneous frequency is expected to reveal more precise details of the underlying phenomenon of the nonlinear distortion.

The Hilbert transform (HT) is a well-known technique in signal processing to compute instantaneous frequency and amplitude. The HT can be interpreted as a 90^o phase shifter. Reverting temporarily to continuous notation, for any arbitrary time series, $x(t)$, the Hilbert Transform (HT), $y(t)$, is obtained as follows [130]:

$$y(t) = \frac{1}{\pi} P \int_{-\infty}^{+\infty} \frac{x(\tau)}{t-\tau} d\tau \tag{7.1}$$

where P denotes the Cauchy principal value. With this definition, $x(t)$ and $y(t)$ form a complex conjugate pair leading to an analytic signal:

$$z(t) = x(t) + jy(t) = a(t)e^{j\theta(t)} \tag{7.2}$$

in which

$$a(t) = \sqrt{(x^2(t) + y^2(t))}$$
$$\theta(t) = \arctan \frac{y(t)}{x(t)} \tag{7.3}$$

Here, $a(t)$ is the IA and $\theta(t)$ is the instantaneous phase. The IF can be computed as:

$$\omega(t) = \frac{d\theta(t)}{dt} \tag{7.4}$$

This classical wave theory-styled definition of IF is computed through differentiation rather than integration. Hence the IF is local, not global, and reflects intra-wave frequency modulation [18]. Intra-wave frequency modulation represents the change of IF within one oscillation cycle (or within a period of a wave). However, this way of computing IF and sometimes the concept of IF itself has been subjected to controversies and limitations [124, 129, 131, 132]. Cohen [129] showed that the HT produces meaningful IF only for *monocomponent* signals while the Bedrosian and Nuttall theorems [131, 132] impose further constraints, e.g. non-overlapping amplitude spectra ($a(t)$) and the spectra of cosine term ($cos(\theta(t))$). If for a given function the spectra of $a(t)$ and $cos(\theta(t))$ are overlapped then that function cannot be expressed in the analytic signal form given in Eq. 7.2. Similarly any real signals with positive local minima and negative local maxima (the so called multicomponent signals) also cannot be expressed in the analytic signal form, meaning HT does not exist. Unfortunately, these conditions are too restrictive and most practical data do not meet these requirements. As a result, the full potential of the HT had to wait for the development of the empirical mode decomposition (EMD).

7.3 The Hilbert-Huang Transform

As discussed in the previous chapter, the EMD was first introduced to solve the limitations of the HT. Recalling what was said in the previous chapter, EMD decomposes any signal into a finite set of elementary signals through the Sifting process and for an elementary signal to be an IMF, it must satisfy the following two important properties [18]:

1. The number of extrema (maxima and minima) and the number of zero-crossings in the entire input signal (total duration of the signal) must either be equal or differ at most by one.
2. The mean value of the envelop defined by the local maxima and local minima is equal to zero at any point.

The IMFs generated by the EMD satisfy the constraints and/or the limitations to admit well-behaved HT and meaningful instantaneous local frequencies as a function of time. EMD together with the HT is what Huang et al. referred to as the Hilbert-Huang Transform (HHT) [18]. The Hilbert-Huang Transform (HHT) is a signal analysis approach which is well-suited to nonlinear, nonstationary signals [18,96]. The application of HHT involves two steps. The first decomposes a discrete time-domain signal $y(n)$ into a set of M intrinsic mode functions (IMFs), $y_j(n)\,;\ j = 1, \cdots, M$, using empirical mode decomposition (EMD) such that:

$$y(n) = \sum_{j=1}^{M} y_j(n) + r(n) \tag{7.5}$$

where $r(n)$ is the residue. The second step determines the instantaneous frequency (IF) and instantaneous amplitude (IA) of each IMF y_j using the Hilbert Transform. From these, one can construct straightforwardly the time-frequency-energy distribution referred to as the Hilbert spectrum [18, 96].

7.3.1 Hilbert-Huang Spectrum

The HT is readily applied to each IMF in order to determine the IA ($a_j(n)$; $j = 1, \cdots, M$) and IF ($\omega_j(n)$; $j = 1, \cdots, M$) according to Eqs. 7.3 and 7.4 respectively. The analytic representation of the input signal may then be expressed as:

$$y'(n) = \sum_{j=1}^{M} a_j(n) e^{i \int \omega_j(n) dn} \tag{7.6}$$

where, since it is constant, the residue $r(n)$ is omitted. The original input signal,

$y(n)$, is the real part of the analytic signal. The IAs ($a_j(n)$; $j = 1, \cdots, M$) and IFs ($\omega_j(n)$; $j = 1, \cdots, M$) then give a time-frequency-amplitude representation of the signal, termed the Hilbert-Huang Spectrum [18, 96]. A plot of the time-frequency distribution of IA^2 (square the amplitude) illustrates the energy density in similar fashion to a conventional spectrogram.

7.3.2 Relation to Fourier techniques

Expressed as a sum of sinusoids, the input signal is given by:

$$y'(n) = \sum_{j=1}^{\infty} a_j e^{i\omega_j n} \tag{7.7}$$

where a_j and ω_j are constant amplitude and frequency terms respectively. Because the frequency of each sinusoidal function is time-independent, Fourier analysis is able to construct stationary data only. Also, since the sine waves used to describe a signal are infinite in extent, Fourier analysis is considered a global analysis tool. The accuracy thus depends critically on data length and stationarity, yet practical data is generally short in existence and of arbitrary duration.

The comparison of Eqs. 7.6 and 7.7 show that the HHT is a generalised Fourier expansion but with time-varying amplitude and frequency which accommodate nonlinear, nonstationary data. The Fourier representation implies constant energy at a given frequency, i.e. a regular harmonic wave which persists unchanged throughout the full data record. HHT analysis, in contrast, reflects the *local* likelihood of energy at a given frequency. A brief description of HHT analysis is presented in this section and for more detailed presentations, readers are referred to [18, 96, 98].

7.4 Loudspeaker distortion analysis

This section reports our first attempt to apply HHT to the analysis of nonlinear distortion produced by miniature loudspeakers. The work is our first steps to align the analysis of nonlinear distortion to its physical origins. This work was performed using real mobile phone loudspeaker recordings.

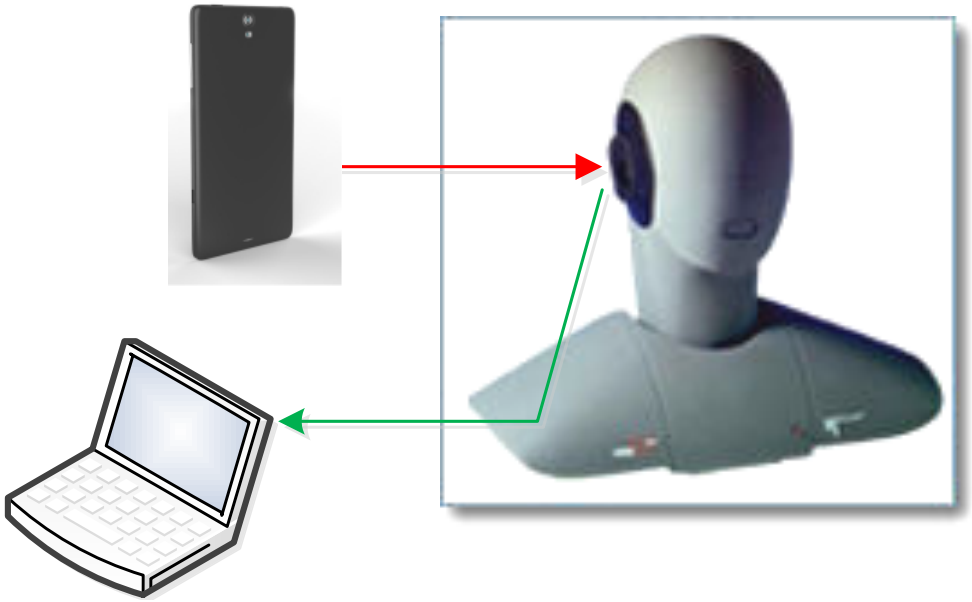


Figure 7.2 – Experimental setup in an anechoic chamber to measure loudspeaker outputs.

7.4.1 Experimental set-up

The nonlinear response of a loudspeaker is observed from its output to a single sinusoidal excitation signal. This approach was used to characterize a real mobile phone loudspeaker placed before a head and torso mannequin at a distance of 30cm in an anechoic chamber. The experimental set-up used is illustrated in Fig. 7.2. The device is configured to operate in hands-free mode and at maximum volume at which nonlinear distortion is greatest. Input signals sampled at 48kHz are pure sinusoids with frequencies between 100Hz and 3800Hz in 100Hz intervals. They are stored in mobile phone memory and played back using a pre-installed VLC player. Loudspeaker outputs are recorded with a high-quality (linear) microphone mounted in the mannequin ear. Recorded signals are stored on a PC at the same 48kHz sampling frequency.

7.4.2 HHT Analysis

As an example we consider a real mobile phone loudspeaker subjected to a single sinusoidal excitation of frequency 1kHz. Fig. 7.3(a) shows the results of STFT analysis. Several high-order harmonics are visible, representing the traditional view of nonlinear distortion.

Fig. 7.3(b) illustrates the four (out of eight) IMFs which result from decomposition of the loudspeaker signal using EMD and the routines available in [99]. Since EMD extracts the highest-frequency IMF first, IMF-1 is the distorted harmonic caused by loudspeaker nonlinearities. IMF-2 is the distorted natural frequency at 1kHz whereas the other IMFs

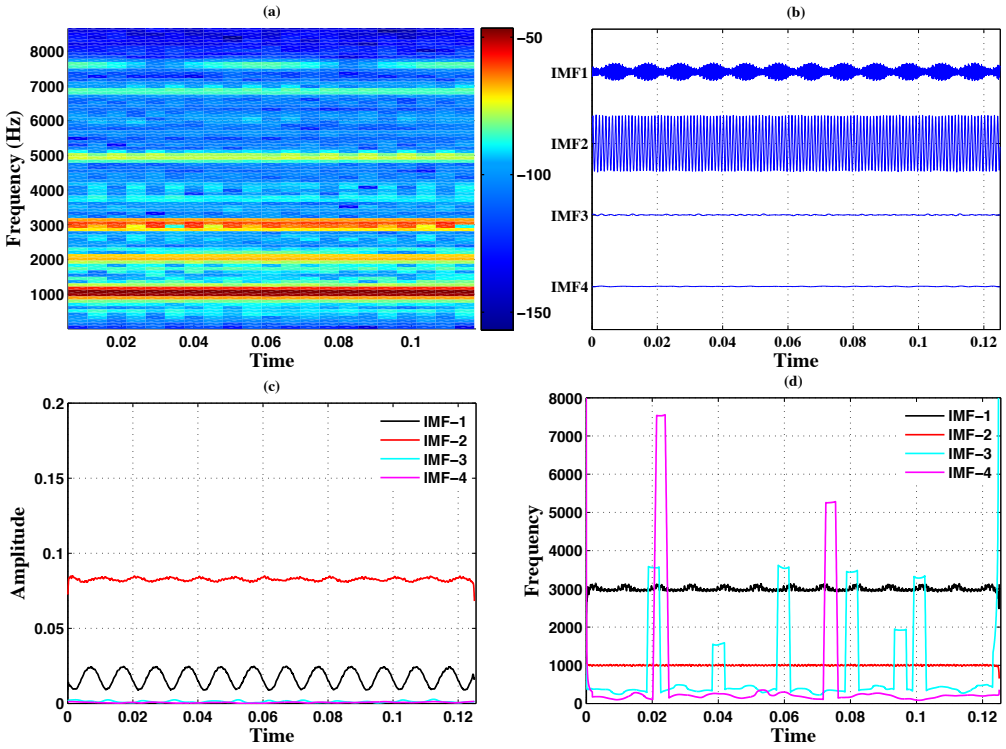


Figure 7.3 – (a) STFT spectrogram of a mobile phone loudspeaker response to a pure sinusoidal input at 1kHz, sampled at 48kHz; (b) Loudspeaker response to 1kHz sine tone is decomposed by the EMD, resulting in the 8 IMFs, first 4 IMFs are listed above and others are not displayed since they are almost zero; (c) IA profiles of the IMFs obtained by HHT; (d) IF profiles of the IMFs obtained by HHT

have negligible energy.

Fig. 7.3(c) illustrates the IA profiles of the four IMF components which exhibit intra-wave amplitude modulation, namely variation in amplitude across time. Fig. 7.3(d) illustrates the corresponding IF profiles which exhibit intra-wave frequency modulation. This is due to the displacement of the loudspeaker diaphragm which is no longer a pure sinusoidal function on account of nonlinear distortion. A relatively strong third-order harmonic is also generated as a result of asymmetrical loudspeaker nonlinearities.

The wave-profile deformation caused by the nonlinear distortion is the result of accumulated harmonic content and intra-wave amplitude-and-frequency modulation. This *cumulative* effect is observed in the time domain response of the loudspeaker shown in Fig. 7.4. The waveform deformation is not constant, but varies from high to low and vice versa in accordance with the IA profile in Fig. 7.3(c). The extent of the deformation depends on the magnitude of the additional harmonics and the strength of the intra-wave amplitude-and-frequency modulation. Close observation of IA and IF profiles in

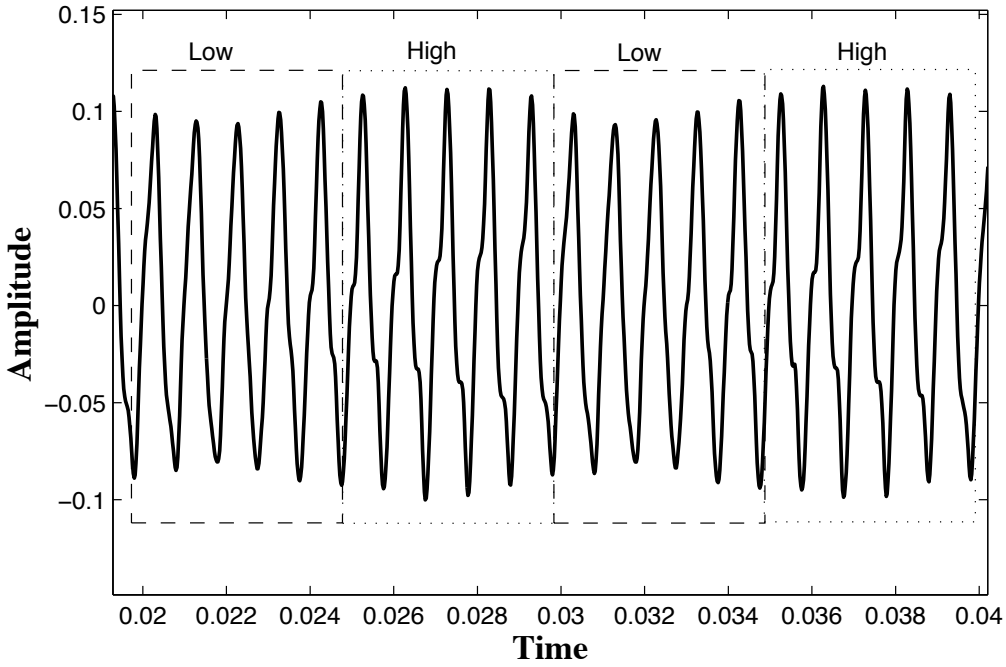


Figure 7.4 – A real mobile phone loudspeaker response to 1kHz pure sine tone. The wave-profile deformation caused by the nonlinear distortion is not constant throughout the time.

Figs. 7.3(c) & (d) respectively shows that the frequency variation of the IMF components increases when their amplitude decreases and vice versa. This is indicative of *softening* nonlinearity [133].

The effects described above are not reflected in the traditional STFT spectrogram which instead shows spurious harmonics. HHT-derived estimates may thus reflect more reliably nonlinear behavior than STFT-derived estimates. Huang et al. in [18] stated that the intra-wave frequency modulation is the hallmark of nonlinear distortion, where the frequency of the system changes with position even with-in one oscillation period. Besides, authors argue that a priori defined bases in the traditional signal analysis techniques impose numerous harmonics and that these are nothing more than a mathematical artifact, with no link to a physical source [18, 96]. Unlike traditional approaches, EMD adapts the bases to the signal itself and can therefore yield more physically relevant results. HHT analysis leads to a new physical interpretation of nonlinear distortion. In place of harmonic distortion is the concept of *cumulative* effect of harmonic content and intra-wave amplitude-and-frequency modulation. The real question, however, is whether the frequency and amplitude modulation illustrated in Fig. 7.3 have a real, physical source or whether they are simply an artefact of the HHT.

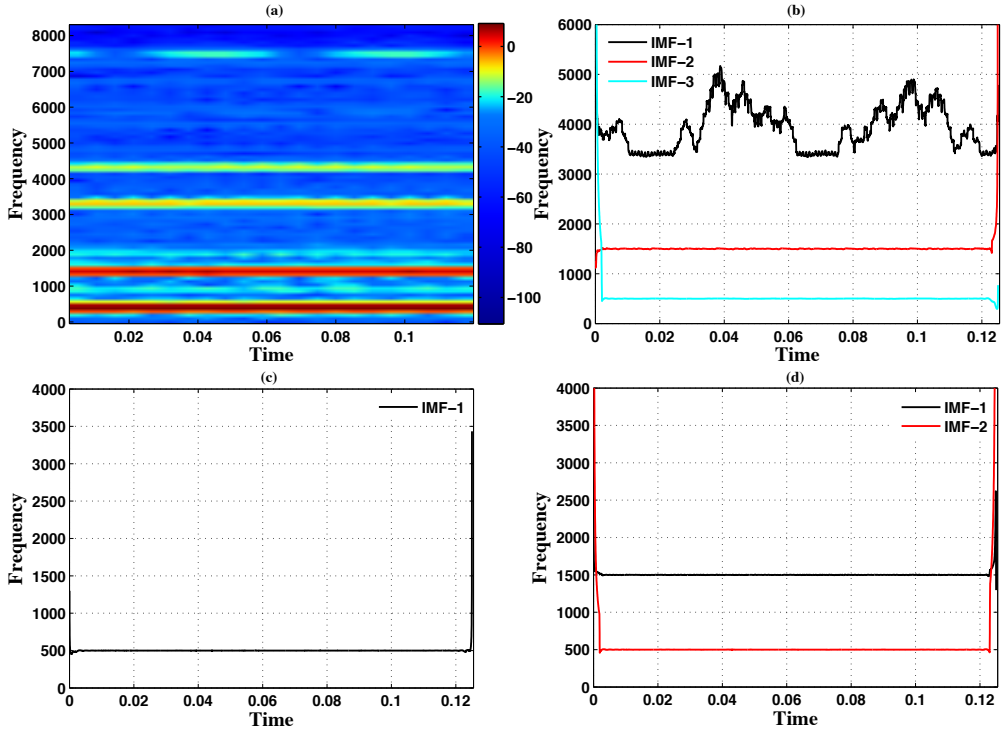


Figure 7.5 – Time-frequency-energy distributions: (a) STFT spectrogram of a mobile phone loudspeaker response to a pure sinusoidal input at 500Hz, sampled at 48kHz; (b) the IF profiles obtained by HHT; (c) IF profiles for a high-quality loudspeaker response to the same input; (d) the IF profiles of the same loudspeaker subject to an input excitation comprised of pure sinusoidal at 500Hz and its third harmonic.

7.5 Validation of HHT

The HHT is thoroughly validated in [18] with analytical examples. This section aims to validate the HHT technique as a means of characterizing nonlinear loudspeaker behavior. Figs. 7.5(a) and 7.5(b) illustrate the spectrogram and IF profiles of a mobile phone loudspeaker response to a pure sinusoidal input at 500Hz. The IF profiles show *cumulative* harmonic and modulation nonlinear distortion. There is only a weak third-order harmonic and significant intra-wave amplitude-and-frequency modulation, whereas the spectrogram shows a strong third order harmonic and several, weak harmonics. Fig. 7.5(c) shows the corresponding IF profiles when the mobile phone loudspeaker is replaced with a high-quality (linear) loudspeaker and shows a total absence of amplitude-and-frequency modulation. Fig. 7.5(d) shows the IF profiles of the (high-quality) loudspeaker output when a simulated, 3rd order harmonic distortion is added to the input. Once again, there is no amplitude and frequency modulation indicating that the distortion observed in (b) has physical origins and is not simply an artifact of HHT processing.

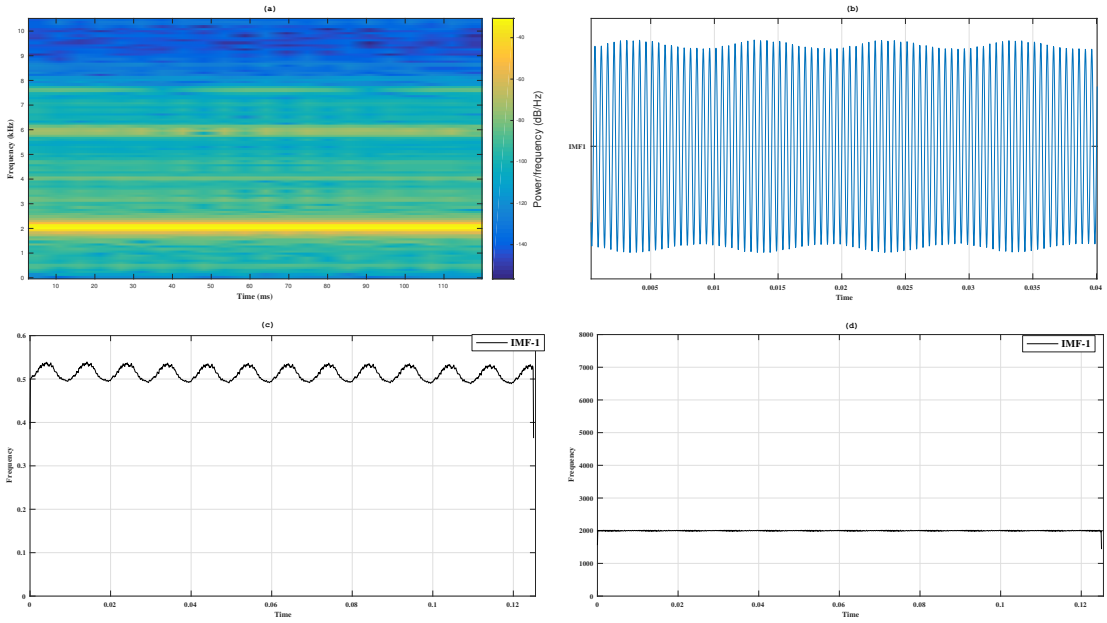


Figure 7.6 – (a) STFT spectrogram of a mobile phone loudspeaker response to a pure sinusoidal input at 2kHz, sampled at 48kHz; (b) Loudspeaker response to 2kHz sine tone (zoomed in) is decomposed by the EMD, resulting only a single IMF, meaning the loudspeaker response itself satisfy the EMD properties; (c) IA profile of the IMF obtained by HHT; (d) IF profile of the IMF obtained by HHT, indicating very low percentage of modulation

After HHT analysing the three different microspeakers outputs for pure sine tone excitations at different frequencies, we could determine that the nonlinear distortion is caused by the cumulative effect of the harmonic content and the intra-wave frequency-and-amplitude modulation. Further, the harmonics content is dependent on the input signal level. For an over-driven microspeaker, the harmonic content is stronger but is limited to the third-order distortion. At moderate level of excitation the modulation distortion is more detrimental than the harmonic content. On the other hand, the strength of the intra-wave frequency-and-amplitude modulation is nonlinearly dependent on the input signal level and frequency. The intra-wave frequency modulation is stronger even at moderate levels of excitation if the input signal frequency is close to the natural resonant frequency of the microspeaker compared to the other frequencies. The waveform deformation develops as soon as the intra-wave frequency modulation index (or the percentage of modulation) exceeds a certain threshold, which is different for different microspeakers. For example, Fig. 7.6 represents the STFT and HHT analysis of a microspeaker response to a 2kHz sine tone. In Fig. 7.6(a) traditional spectrogram shows a relatively stronger third harmonic and a weak second harmonic. The actual loudspeaker response shown as an IMF1 in Fig. 7.6(b) indicates no visual waveform distortion except the amplitude modulation. The same can be witnessed in the HHT analysis in terms of IA and IF

in Figs. 7.6(c) and 7.6(d) respectively. Since there is no waveform deformation in the signal, the intra-wave frequency modulation index is very weak and no sign of harmonic content. Hence it is safe to consider that the over-all degree of nonlinearity produced by a microspeaker is largely due to the modulation distortion with a limited impact due to harmonic content.

Expanding from these empirical findings, we believe that this new alternative interpretation of loudspeaker nonlinearities can be potentially applied to solve the NAEC problem. One possible solution to estimate the nonlinear echo is by incorporating the intra-wave amplitude-and-frequency modulation effect as a pre-processor in the NAEC system to model the loudspeaker distortion and a traditional linear adaptive filter in cascade to model the linear acoustic echopath. The major advantage over the traditional solutions is that the NAEC system does not require many orders of harmonics to model the downlink nonlinearities as the intra-wave modulations incorporate the inherent nonlinearities in the acoustic echopath. However, developing a probabilistic model that explicitly emulates the cumulative effect of the harmonic content and the intra-wave frequency-and-amplitude modulation effect in the microspeakers is a most challenging part of the NAEC design. This alternative interpretation can also be extended to other similar research areas like loudspeaker modeling and loudspeaker linearisation.

7.6 Limitations of the HHT/EMD

Despite having many characteristic advantages of employing EMD/HHT technique for studying nonlinear and nonstationary signals, current decomposition technique still suffers from several limitations. The major shortcoming of the EMD is that it is completely empirical by means it lacks a firm mathematical foundation and hence the performance of EMD can only be reviewed either empirically or numerically. Ideally, the decomposed elements or the IMFs must satisfy the properties like completeness, orthogonality and uniqueness to be considered as (adaptive) basis functions. In the case of standard EMD, orthogonality and uniqueness properties are highly dependent on the input data and the sifting process parameters (Eg. stopping criteria). Often a small change in the data or the sifting process parameters can result in a different series of IMFs. Orthogonality is also not guaranteed in many applications as often multiple IMFs are correlated, meaning different modes of oscillations (analogous to spectral content) coexist in multiple IMFs. This problem is called the "mode mixing" in the literature [18,124,134,135]. This so called mode mixing problem questions the physical meaning-fullness of the IMFs. However, Huang et al. argue that the orthogonality is not a necessary property for the IMFs for analysing the nonlinear and nonstationary signals and is required only to decompose linear signals [18, 135]. We have explored the tone masking method proposed in [136]

to get-rid of the mode mixing problem and to separate the closely spaced frequency components that would be inseparable with standard EMD technique.

Although the orthogonality and the uniqueness properties are questionable, the completeness is totally guranteed and can be easily verified by reconstructing the original signal as the sum of all IMFs. No loss of information is incurred. Despite perfect reconstruction, HHT still suffers from so-called *end effect* artifact's [18]. The cubic spline fitting to local extrema in the EMD process is error prone, especially due to discontinuities at signal extremities. As a result, Gibbs phenomenon is induced upon the application of the HT to each IMF [18, 96]. This effect is observed in Figs. 7.3 and 7.5.

Despite notable success of applying EMD in various fields, the standard EMD is limited to analyse real signals from a single channel data. Often in many applications the data usually comprise in multi-channel data sets. In such cases, applying standard EMD separately to each channel may not produce same number of IMFs for every channel and also the same-index IMFs may not contain the same spectral content across data channels. This is because of the data dependent adaptive basis nature of the EMD. Some solutions to avoid these problems are proposed in the literature and are briefly discussed in the next section before concluding the chapter.

7.7 Recent advancements/extensions of the standard EMD

To overcome the aforementioned drawbacks of the standard EMD, many extensions have been reported [100, 103, 104, 115, 124, 134, 137]. Popular among them with a full potential to be extended to study the nonlinear distortion problem in hands-free communications are briefly reported in this section.

7.7.1 EEMD

A significant property of the EMD algorithm has been deduced after studying the characteristics of the white noise using EMD in [138, 139]. It turns out that the EMD exhibits a dyadic filter bank structure of constant-Q bandpass filters for white Gaussian noise (WGN). For the intermittent data, which causes mode mixing in the standard EMD, this filter bank property is observed to be compromised. Adding white noise to such a data can provide a uniformly distributed reference frame in the time-frequency space, which helps standard EMD to repair the compromised filter bank property [104, 139]. Inspired from this filter bank property of EMD, authors in [104] proposed a noise assisted EMD, called the Ensemble Empirical Mode Decomposition (EEMD), to improve the robustness (in terms of orhtogonality and uniqueness properties) of EMD and to alleviate

the problem of mode mixing. The working principle of EEMD is explained in the following way [104]:

- For an original input signal $x(n)$ add WGN (zero mean and constant variance,σ^2) at different finite amplitudes and create an ensemble of $k = 1, \cdots, K$ signals.

$$\{x_k(n)\}_{k=1}^{K} = x(n) + \{w_k(n)\}_{k=1}^{K}$$

  where $\{w_k(n)\}_{k=1}^{K} \sim \mathcal{N}(0, \sigma^2)$ are K independent realizations of WGN.

- Decompose every noisy signal in the ensemble into a set of IMFs using the standard EMD. The noisy signals have a lot more local extrema and detailed envelopes than the original signal, due to the added noise, which render EMD to decompose more closely spaced frequencies, yielding enhanced IMFs.

- Take the average of same-index IMFs across the ensemble to obtain the resultant IMFs of the EEMD. In the process of averaging the WGN components will eventually cancel out (by an amount of $\frac{\sigma^2}{K}$) leaving only the true IMFs.

In this ensemble process, all the IMFs stay within the natural dyadic filter bank windows, reducing the spectral leakage into other IMFs and thus significantly reduce the chance of mode mixing. Although EEMD is a promising technique, the ensemble process makes it more computationally demanding than the standard EMD.

7.7.2 MEMD

Authors in [134] proposed a Multi-variate EMD (MEMD) and successfully extended the idea of EMD (rather EEMD) to multi-channel data. Authors claim that the MEMD has the ability to identify and align common oscillatory scales in different data sources and termed this phenomenon as mode alignment. Thus, MEMD generates the same number of IMFs for every input channel and same-index IMFs across the channels have almost the same spectral content. Authors further extended the work and advocated the usefulness of computing EMD using the MEMD in [137]. Taking advantage of the mode alignment and the filter bank properties of the MEMD, EMD via MEMD yields better signal decomposition and enhanced IMFs compared with the standard EMD and EEMD techniques. By better signal decomposition we mean resolving closely-spaced frequencies, no mode mixing, unique and reproducible decomposition.

7.7.3 Hilbert Spectral Analysis (HSA)

HSA proposed in [124] is a recent advancement in the time-frequency analysis methods and a more powerful signal analysis technique than the EMD/HHT to model nonlinear and nonstationary signals. We discussed EMD as a non-conventional signal analysis tool which does not depend on the linear transform theory and a priori defined basis functions like the traditional techniques. EMD uses HT to construct analytic signal (Eq. 7.2) and to compute IA and IF parameters of the IMFs to obtain the time-frequency-energy representation. Authors in [124] argued that the HT relies on the so called *Harmonic Correspondence* property ($\mathcal{H}\{\mathbf{a(t)cos}(\omega\mathbf{t}+\theta)\} = \mathbf{a(t)sin}(\omega\mathbf{t}+\theta)$) in the process of complex extending a real signal (that is constructing an analytic signal), which may lead to incorrect IA and IF parameters. For a given signal $x(t)$, by relaxing the harmonic correspondence assumption, there will be many choices for the quadrature components (imaginary part of the analytic signal). Accordingly there will be many complex extensions (thus many IA and IF pairs) for a given real signal $x(t)$, which authors in [124] termed as *Latent signals*. It has been shown that these latent signals can still maintain analyticity and are well-suited in representing/modeling the real physical phenomenon. It is worth mentioning that the analytic signals generated by HT (assuming harmonic correspondence) are usually confined to one particular region which is a subset of a larger set of latent signals.

Authors reformulated the time-frequency analysis problem as a latent signal analysis (LSA) problem where the uncertainty is not in the time or frequency localization but in choosing the right quadrature component. In this new framework, any nonlinear and/or nonstationary signal can be represented as a superposition of latent signals, where each latent signal is a complex amplitude-and-frequency modulated (AM-FM) component analogous to IMFs. Assuming IMFs are the perfect AM-FM components, authors incorporated EMD algorithm (with modifications to control its limitations) to decompose any signal into a set of IMFs and proposed a more sophisticated way to compute IA and IF parameters without using the HT. Further, authors proposed a 3-D visualization of the Hilbert spectrum by plotting IF vs. real-valued time-domain signal vs. time as a line in a 3-D space and coloring the line with respect to IA parameters. The benefits of HSA as opposed to the HHT and the traditional STFT analysis are illustrated in [124, 140].

7.8 Summary

In this chapter we have provided new insights into loudspeaker nonlinear distortion that can be potentially applied to solve the NAEC problem. This work is our first steps to align the analysis of nonlinear distortion to its physical origins. First we introduced

a relatively new time-frequency analysis method called the Hilbert-Huang Transform (HHT), which is well-suited to the analysis of nonlinear, nonstationary signals. The HHT is based on EMD and does not assume any particular linear transform theory prior to time-frequency decomposition of the signal. The HHT spectral analysis provides instantaneous time and frequency resolution unlike the conventional time-frequency analysis methods. Instantaneous amplitude (IA) and frequency (IF) parameters give more detailed and enhanced representation of the underlying nonlinear behaviour of the distorted signals.

Further, this chapter reports our first attempt to apply HHT to the analysis of nonlinear distortion produced by miniature loudspeakers. This approach gives an alternative interpretation of loudspeaker nonlinear behavior. The waveform deformation caused by the nonlinear distortion is the result of a *cumulative* effect, namely that of harmonics and intra-wave amplitude-and-frequency modulation, instead of the pure harmonic distortion interpretation which results from Fourier treatments. Besides, the HHT analysis supports the exploration of different nonlinear phenomena: quadratic, cubic or higher-order, softening and hardening effects, intra-wave amplitude-and-frequency modulation and distorted harmonic responses etc. This valuable information helps in designing more accurate nonlinear loudspeaker models.

Finally, we outlined the limitations of the EMD/HHT technique and briefly discussed the recent advancements/extensions to overcome those limitations.

Chapter 8

Conclusions and Future Directions

Acoustic echo cancellation (AEC) is an essential module in any hands-free communication device. AEC helps in creating smooth and comfortable full-duplex voice conversation in hands-free mobile telephony. AEC helps to improve the speech recognition rate in voice assistance systems. AEC greatly enhances the audio quality and prevents the listener fatigue in audio conferencing system. Currently, most of the hands-free communication devices use linear AEC algorithm to cancel the acoustic echoes. Linear AEC is well-studied as a system identification problem in the literature. In practice, the three main challenges of a linear AEC are background noise, reverberation and double-talk. There are many sophisticated linear AEC algorithms in the literature that are robust against these challenges and can demonstrate superior AEC performance. However, nonlinear distortion in the acoustic echopath has been the biggest threat to the performance of a linear AEC.

Current trend of portability and miniaturisation in consumer electronics have created an enormous demand for low-cost transducers. Particularly low-cost/miniature loudspeakers aka microspeakers are a major source of nonlinear distortion. We have shown in Chapter 1 that the conventional linear AEC algorithms are inadequate to tackle the nonlinear distortion in the LEMS and discussed the imperative need to implement nonlinear acoustic echo cancellation (NAEC) algorithms. In order to get the best NAEC performance, it is vital to accurately model the dynamic nonlinear behavior of the loudspeakers. Therefore, this dissertation concerns the analysis, identification and characterization of nonlinear distortion in loudspeakers and its application to NAEC.

8.1 Contributions

The contributions presented in this thesis are listed below:

- This thesis begins with the definition of nonlinear distortion in Chapter 2 followed by a brief discussion on the theoretical concepts of nonlinear systems and their modeling. Chapter 3 explores the possible sources of nonlinear distortion in the LEMS, highlighting the downlink path nonlinearities. This chapter also discusses different loudspeaker models to emulate its inherent nonlinear behaviour. Volterra series derivatives are more popular in the literature. This chapter also offers a theoretical framework of a nonlinear (loudspeaker) system identification approach based on exponential sine-sweep signals referred to as nonlinear convolution technique. Overall, Chapter 3 serves as a state-of-the-art for the identification and modeling of nonlinear loudspeakers.

- Our main contributions start in Chapter 4 with an assessment of the suitability of Volterra series derivatives in accurate modeling of nonlinear distortion in microspeakers. First we reported the identification of a real mobile phone loudspeaker using the nonlinear convolution technique. By identification we mean computing the linear and the higher-order impulse responses of a loudspeaker. We then compared the synthesized outputs of the power series model (PSM) and the generalized polynomial Hammerstein model (GPHM) to empirically measured, real loudspeaker outputs. This work suggests that the GPHM approximates more stable and reliable practical nonlinear behavior of a loudspeaker.

- After identifying a suitable loudspeaker model, determining its optimal model parameters is one of the challenging issues in real-time applications. Therefore, we investigated further the accuracy of the GPHM model as a function of the key parameters, namely the filter length (L) of the simplified Volterra kernels and the order of nonlinearities (P). This investigation involves the identification of loudspeakers from three different mobile phones. The results of this study demonstrate that the P should be high enough to capture the principle sources of nonlinear distortion where as moderate filter lengths of the higher-order kernels are sufficient to obtain reliable loudspeaker modeling. Even if the chosen order of nonlinearity (P) is greater than the true order, the model accuracy is still reliable provided the L is moderate. If both P and L are greater/lesser than the actual order then it leads to over/under-modeling scenario respectively and limits the model accuracy. This work also highlights the limitations of the GPHM model whose performance is inconsistent in accurate modeling the nonlinear loudspeaker behavior involving real-speech inputs.

- Another study[1] reveals that the ERLE performance of a NAEC algorithm can be inflated using the PSM model based synthesized nonlinear echo signals. For the same NAEC algorithm (for example cascaded approach) and under the identical test

[1] This study was actually presented in Chapter 5.3 of this thesis.

conditions, the ERLE results generated with the PSM model lead to favourably-biased indications of performance. In contrast, the results generated with the GPHM model better reflect practical measurements and is thus an appealing alternative model for future evaluations of NAEC performance.

- Following a review of the state-of-the-art NAEC and/or NAES solutions in Chapter 5, we have provided a comprehensive performance and stability analysis of the widely used NAEC algorithms, which have not been fully explored before. The results from our experimental evaluations demonstrate that the popular NAEC solutions such as the cascaded and the parallel approaches, which are shown to provide better performance in nonlinear environments in the literature, are less competent in many practical acoustic environments. This is a common problem related to most of the NAEC algorithms, if the nature of the nonlinear distortion varies then they will not perform as expected. Another open issue is that the stability of the NAEC algorithms is not guaranteed in the absence of nonlinear distortion (when the acoustic echopath is totally linear). Thus there is a lot of potential for further research in the NAEC domain.

- Thanks to the recent advances in the nonlinear signal processing field, empirical mode decomposition (EMD) technique developed in the recent past emerged as a dedicated nonlinear and nonstationary signal analysis tool. EMD has been successfully applied in various engineering and non-engineering applications involving nonlinear systems. In this thesis for the first time we have studied the application of EMD to combat the nonlinear distortion in the LEMS. Chapter 6 reports the first EMD-based approach to NAEC. EMD decomposes the nonlinear echo signal into a set of IMFs which can further be characterized as either nonlinear or linear dominant. NAEC is accomplished through the application of adaptive power filtering (parallel approach) to the nonlinear dominant IMFs and the conventional linear adaptive filtering to the linear dominant IMFs using the full-band reference signal. When compared to the power filter baseline system, experimental results demonstrate improved NAEC performance in terms of greater echo reduction and faster convergence. While the proposed solution is also robust to dynamic variations in the acoustic channel, computational complexity is a major drawback.

- It is noteworthy that all NAEC algorithms in the literature including our EMD-based NAEC solution share one common feature: the underlying interpretation of nonlinear distortion as harmonic distortion. This traditional interpretation of nonlinear distortion stemmed from the application of traditional (for example Fourier-based) time-frequency analysis to study the dynamic nonlinear systems. Traditional data analysis techniques are ill-suited to analyse nonlinear signals as they depend on linear transform theory. As an alternative, we introduce a relatively

new time-frequency analysis method called the Hilbert-Huang Transform (HHT). Based on EMD and the Hilbert transform (HT), HHT is well-suited to the analysis of nonlinear and nonstationary signals. Chapter 7 reports our first attempt to apply HHT to the analysis of loudspeaker nonlinearities. On the basis of the results of this work, this approach gives an alternative interpretation of loudspeaker nonlinear behavior. The waveform deformation caused by the nonlinear distortion is the result of a cumulative effect, namely that of harmonics and intra-wave amplitude-and-frequency modulation. The extent of deformation depends mostly on the magnitude of amplitude-and-frequency modulation. Besides, the HHT analysis supports the exploration of different nonlinear phenomenon which helps in designing more accurate nonlinear loudspeaker models.

Thus the thesis begins with a traditional interpretation of nonlinear distortion in loudspeakers and ends with a novel and accurate interpretation of nonlinear distortion, which marks a new beginning of NAEC research.

8.2 Future directions

While this thesis calls into question the interpretation of nonlinear distortion in the loudspeakers through harmonics and points towards a link between physical sources of nonlinearity and amplitude-and-frequency modulation, many opportunities for extending the scope of this thesis remain. The section presents some of these directions:

Amplitude-and-frequency modulation based NAEC

New findings suggest new approaches. As most of the NAEC solutions in the literature depends on the harmonic distortion and Volterra series based modeling, future research should move beyond the harmonic distortion and consider how the new physical interpretation of loudspeaker nonlinearities can be applied to solve the NAEC problem. Provided a probabilistic model that explicitly emulates the cumulative effect of the harmonic content and the intra-wave frequency-and-amplitude modulation, with a linear adaptive filter in cascade we believe that the best NAEC performance can be achieved at lower computational costs in the near future. Moreover, it is safe to assume the stability of such an algorithm when the nature of nonlinear distortion is varying and even in the absence of nonlinear distortion as this information is captured and incorporated in the amplitude-and-frequency modulation parameters.

Filter-bank property of EMD

As discussed in Chapter 7, EMD and its extensions (particularly, EEMD and MEMD) exhibit a dyadic filter-bank property. Exploiting such a filter-bank property of the EMD and its extensions, sub-band domain approaches may be explored to further lower the complexity and to obtain a combined reduction of noise and nonlinear echo disturbances.

Chapter A

Sommaire de la thèse en français

A.1 Résumé

Cette thèse porte sur l'analyse, l'identification et la caractérisation de la distorsion nonlinéaire dans les haut-parleurs et son application à l'annulation d'écho acoustique nonlinéaire (ou NAEC, pour "Nonlinear Acoustic Echo Cancellation").

La première partie de la thèse vise à la dérivation d'un modèle de haut-parleur plus précis et empirique. Celui-ci émule la réponse fréquentielle du haut-parleur dans le but de prédire et d'empêcher la distorsion nonlinéaire. Les travaux de recherche suggèrent que le modèle de Hammerstein généralisé se rapproche plus fiablement d'un comportement de haut-parleur nonlinéaire.

Dans la partie suivante, après avoir discuté les études avancées de development des algorithms de NAEC, nous présenterons l'analyse des performances des algorithmes les plus utilisés. Les résultats ont démontré que les solutions populaires n'obtiennent de meilleurs résultats que dans quelques conditions idéales et sont moins performants dans la plupart des environnements acoustiques réels. Nous proposons ensuite une nouvelle approche de NAEC basée sur la décomposition modale empirique (ou EMD, pour "Empirical Mode Decomposition"), une technique récemment développée pour l'analyse de signaux nonlinéaires et nonstationnaires. Des expériences comparatives sur des techniques de reference montrent que la nouvelle approche (NAEC basée sur la EMD) permet d'obtenir une plus grande réduction d'écho nonlinéaire et une convergence plus rapide.

Dans l'étape qui suit, les travaux mis en place sont le commencement sur l'établissement de la correspondence entre l'analyse de la distorsion nonlinéaire dans les haut-parleurs à ses origines physiques. Nous considérons l'application de la transformée d'Hilbert-Huang

(ou HHT, pour "Hilbert-Huang Transform") à l'analyse de la distorsion nonlinéaire dans les haut-parleurs. Sur la base des résultats de cette étude, nous avons rapporté une interprétation alternative des nonlinéarités des haut-parleurs à travers les effets cumulatifs du contenu harmonique et de la modulation en amplitude et en fréquence. Ces nouvelles conclusions pourraient stimuler et renouveler la direction future de la recherche sur la NAEC.

A.2 Introduction

Alors qu'aujourd'hui, l'interaction entre les humains et les machines se fait principalement par le toucher, la prochaine étape de l'interaction sera principalement par commande vocale. Si la croissance récente des dispositifs de communication mains libres marché est une indication, beaucoup de gens préfèrent naturellement la communication mains libres. Dans tous les environnements de communication mains libres, l'annulation de l'écho acoustique (AEC) et l'annulation du bruit jouent un rôle de plus en plus important pour assurer une qualité de communication (vocale) satisfaisante. Dans cette thèse, nous nous sommes concentrés sur le problème d'annulation de l'écho acoustique. De nombreux appareils différents sont équipés de haut-parleurs et de microphones pour une variété de buts différents et souvent ces transducteurs sont montés à proximité les uns des autres. Ce couplage acoustique entre le haut-parleur et le microphone ainsi que des réflexions d'additifs provoque l'écho acoustique. Dans le cas de la téléphonie mobile, cet écho acoustique sera transmis à l'utilisateur distant et la conversation peut être gênante voire insupportable en fonction du délai aller-retour du système. Dans le cas des «smart speakers» d'assistance vocale (exemples typiques: Amazon Echo et Google Home), cet écho acoustique est une source d'interférences pour les moteurs de reconnaissance vocale automatique affectant sa performance (détection de mots-clés et/ou taux de reconnaissance vocale). Ainsi, l'écho acoustique dégrade la qualité de la communication vocale en dégradant l'intelligibilité de la parole et le confort d'écoute. Afin de lutter contre le phénomène d'écho acoustique, il est souvent nécessaire d'utiliser un annuleur d'écho acoustique.

A.2.1 Annulation d'écho acoustique

L'annulation d'écho acoustique (AEC) est un problème vieux de plusieurs décennies dans le traitement du signal depuis l'introduction des communications vocales en duplex intégral, et il s'agit toujours d'un domaine de recherche actif. L'AEC repose sur une approche d'identification de système bien établie. Le trajet de l'écho acoustique (du haut-parleur au microphone) est très dynamique et soumis à des variations dans le

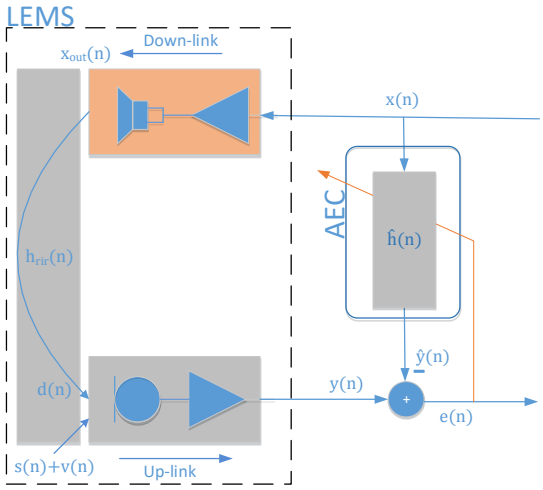


Figure A.1 – Modèle de système illustrant le couplage acoustique dans le système LEMS et une approche générale de l'AEC adaptatif.

temps, en raison de la modification des caractéristiques acoustiques du système LEMS (Loudspeaker Enclosure Microphone System). Par conséquent, AEC utilise généralement un filtre transversal adaptatif linéaire pour estimer la réplique numérique de la fonction de transfert du LEMS. Un système AEC adaptatif typique est illustré à la Fig. A.1, où $d(n)$, $s(n)$ et $v(n)$ représentent respectivement le signal d'écho, le signal de parole à proximité et le bruit.

La plupart des scénarios de cette thèse supposent que le signal du microphone contient uniquement l'écho, c'est-à-dire $y(n) = d(n)$, $s(n) = v(n) = 0$, sauf si spécifié. Le signal de distante $x(n)$ est passé à travers le filtre adaptatif $\hat{h}(n)$ pour synthétiser le signal d'écho $\hat{y}(n)$, qui est ensuite soustrait du signal de microphone $y(n)$ pour annuler l'écho acoustique. Si la réponse impulsionnelle du filtre adaptatif, $\hat{h}(n)$, correspond à celle de LEMS, $h(n)$, (convergence), alors l'écho sera éliminé sans aucun artefact. Cependant, parvenir à une convergence parfaite est une tâche assez difficile, en particulier lors de la manipulation de signaux hautement nonstationnaires comme la parole. En outre, il existe de nombreux autres facteurs tels que le bruit de fond à proximité, la période de double conversation (la période à laquelle signal de parole à proximité et l'écho sont présents en même temps) et les nonlinéarités nuisent aux performances de l'annuleur d'écho. La distorsion nonlinéaire dans le trajet d'écho acoustique est particulièrement gênante. Dans les appareils mobiles actuels, la tendance vers la portabilité et la miniaturisation a conduit à l'utilisation de transducteurs de plus en plus petits. Les contraintes sur la taille du haut-parleur entraînent souvent une sortie nonlinéaire.

Dans ce qui suit, nous examinons l'impact de la distorsion nonlinéaire sur quelques algorithmes adaptatifs populaires associés à l'AEC linéaire. La performance est évaluée

en termes de convergence et d'une mesure standard appelée ERLE (Echo Reduction Loss Enhancement). ERLE est une mesure quantitative qui représente la réduction de l'énergie (en dB) du signal du microphone ($d(n)$) obtenue par réduction d'écho. ERLE est donné par:

$$ERLE = 10log\frac{E\{d^2(n)\}}{E\{e^2(n)\}} \tag{A.1}$$

où $e(n)$ est le signal de sortie AEC à transmettre à l'utilisateur distant. Voici les algorithmes adaptatifs linéaires bien connus considérés pour cette étude:

- L'algorithme LMS (Least Mean Square) avec $\mu = 0.16$
- L'algorithme NLMS (Normalized Least Mean Square) avec $\mu = 1$
- L'algorithme FBLMS (L'algorithme LMS par bloc dans le domaine fréquentiel) avec $\mu = 0.5$ et la taille du bloc est $B = 256$
- L'algorithme DCTLMS (Discrete Cosine Transform-LMS) algorithm avec $\mu = 0.5$
- L'algorithme APA (Affine Projection Algorithm) avec $\mu = 1$ and l'ordre 2

Les détails de ces algorithmes adaptatifs sont bien décrits dans la littérature (par exemple, [2, 3]), et ne seront pas répétés dans cette thèse. La taille de pas μ de chaque algorithme est choisie de telle sorte qu'elle atteigne le maximum ERLE après convergence. Le processus de génération d'écho linéaire et nonlinéaire et l'environnement de simulation complet sont expliqués en détail à la Section 1.1. La figure A.2 illustre le comportement des algorithmes adaptatifs linéaires en termes d'ERLE dans des environnements linéaires et nonlinéaires. La plupart des travaux d'AEC dans la littérature supposent la linéarité des composants électroniques dans le LEMS. Dans de telles conditions linéaires, les algorithmes AEC fonctionnent généralement bien, comme le montre la Fig. A.2. Les nonlinéarités de liaison descendante dans le LEMS réduisent le maximum d'ERLE réalisable par chaque algorithme. Ces courbes démontrent clairement l'impact de la distorsion nonlinéaire sur chaque algorithme AEC linéaire. Les algorithmes FBLMS et APA sont gravement affectés par les nonlinéarités. En dépit du fait que la convergence initiale est meilleure même avec la distorsion nonlinéaire, l'algorithme APA du deuxième ordre se comporte presque comme l'algorithme NLMS en termes d'ERLE. Les algorithmes NLMS et DCTLMS fonctionnent de manière similaire dans les environnements linéaires et nonlinéaires. Même si la performance de l'algorithme LMS est faible, elle reste robuste

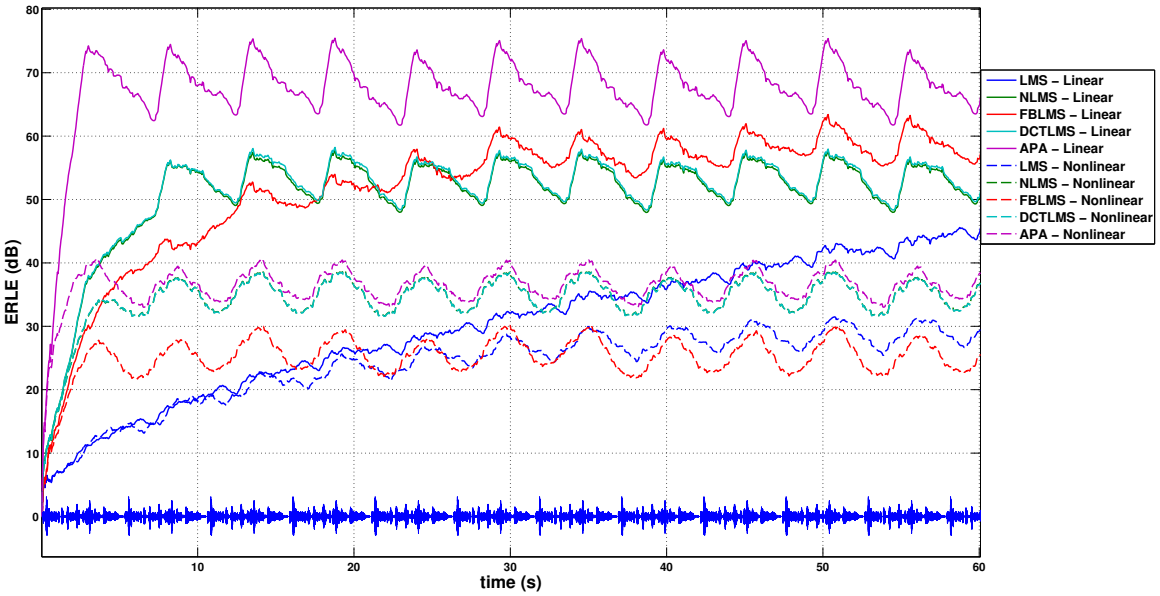


Figure A.2 – Les résultats des tests ERLE pour comparer les performances des algorithmes linéaires AEC dans des environnements linéaires et nonlinéaires.

aux nonlinéarités (en termes de différence dans ERLE) par rapport aux autres algorithmes. Reportez-vous à [5, 6] pour plus de détails.

De cette discussion, il est plausible que les algorithmes AEC linéaires conventionnels soient insuffisants pour traiter la distorsion nonlinéaire dans le LEMS. En conséquence, le très vieux problème de l'AEC est devenu plus difficile et reformulé comme NAEC (Nonlinear Acoustic Echo Cancellation), qui est aujourd'hui un domaine de recherche actif.

A.2.2 L'annulation d'écho acoustique nonlinéaire

La distorsion nonlinéaire redistribue l'énergie dans le spectre et tenter d'annuler l'écho nonlinéaire en utilisant un AEC linéaire laisse un écho résiduel supplémentaire dans le signal de liaison montante, ce qui entraîne la dégradation de l'ERLE. Cela met en évidence le besoin d'algorithmes avancés qui s'attaquent à la distorsion nonlinéaire. L'annuleur d'écho acoustique nonlinéaire doit être capable d'identifier et de suivre non seulement la réponse impulsionnelle linéaire du LEMS, mais également les nonlinéarités associées aux composants du dispositif. Au cours de la dernière décennie, les chercheurs ont déployé beaucoup d'efforts pour résoudre le problème du NAEC. L'état de l'art NAEC est bien décrite dans le Chapitre 5 de cette thèse. La plupart des algorithmes NAEC sont développés sur la base de deux logiques différentes, l'approche parallèle et l'approche en cascade.

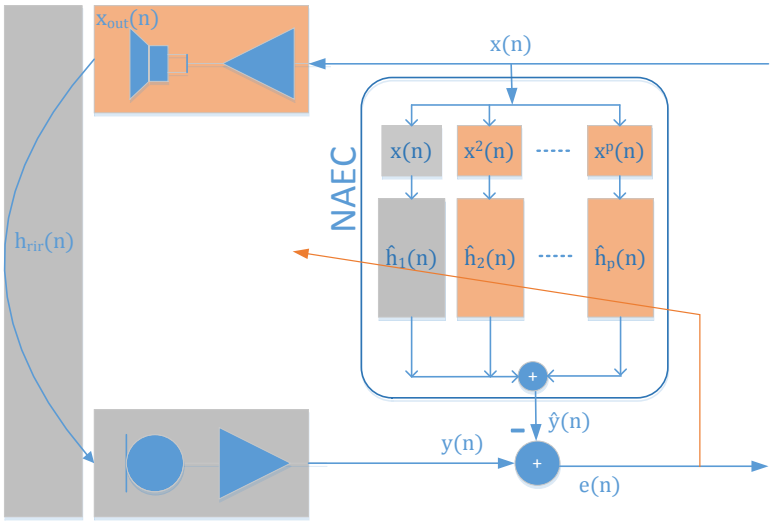


Figure A.3 – Structure de l'approche parallèle basée NAEC.

l'approche parallèle

L'approche parallèle (ou l'approche par filtre de puissance) basée sur NAEC implique le suivi simultané des réponses impulsionnelles nonlinéaires et linéaires à travers les filtres adaptatifs multicanaux, comme illustré à la Fig. A.3 [48]. Cette approche est particulièrement adaptée aux nonlinéarités sans mémoire, où le premier canal représente la réponse impulsionnelle linéaire globale du LEMS. Simultanément, les autres canaux sont utilisés pour suivre de manière adaptative les nonlinéarités d'ordre supérieur dans le LEMS. Bien que cette approche parallèle semble être une structure NAEC simple et pratiquement possible, la vitesse de convergence est toujours lente par rapport à l'AEC linéaire en raison du filtrage adaptatif multicanal et du nombre accru de coefficients de filtrage. En outre, il existe un inconvénient inhérent à l'estimation du trajet acoustique linéaire (RIR) plusieurs fois à travers de multiples canaux (en raison de l'effet de la convolution entre les paramètres de haut-parleur et le RIR).

l'approche en cascade

L'idée clé de l'approche NAEC en cascade était de découpler l'identification des paramètres de haut-parleurs nonlinéaires du suivi du trajet d'écho acoustique linéaire. La structure de NAEC basée sur l'approche en cascade est illustrée à la Fig. A.4. Dans ce cas, un pré-processeur nonlinéaire est placé en cascade avec un filtre linéaire à réponse impulsionnelle finie (FIR). Le pré-processeur nonlinéaire filtre le signal de référence à partir de l'extrémité distante et vise à émuler le chemin de liaison descendante avec ses nonlinéarités. Le filtre FIR linéaire vise à émuler avec précision l'écho nonlinéaire non désiré dans le signal du

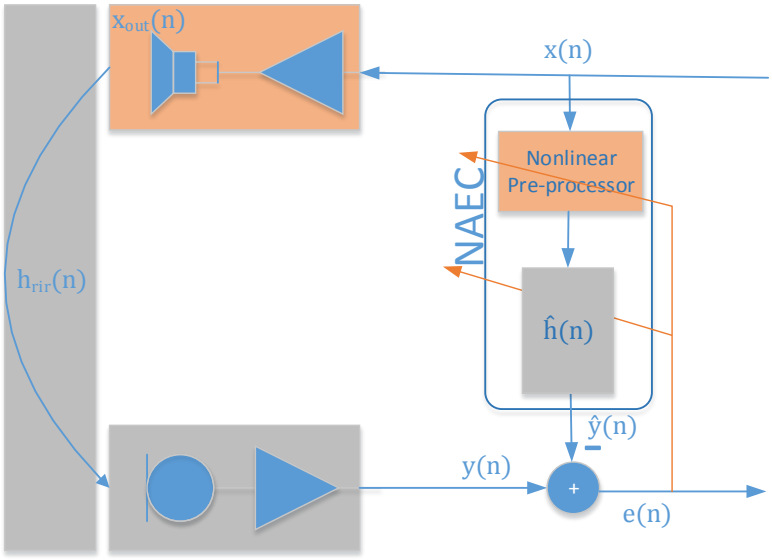


Figure A.4 – Structure de l'approche en cascade basée NAEC.

microphone en filtrant le signal nonlinéaire entrant.

Une NAEC en cascade typique proposée dans [49] est représentée sur la Fig. A.5. Cette approche en cascade implique une estimation de paramètre conjointe pour le pré-processeur nonlinéaire et le filtre FIR en utilisant le signal d'erreur combiné $e(n)$ employant des algorithmes adaptatifs linéaires. Par conséquent, la convergence des deux blocs (ou de tous les filtres adaptatifs) est interdépendante, ce qui entraîne des erreurs éventuelles et réduit également la vitesse de convergence globale.

Plus d'informations sur ces deux approches NAEC sont données au Chapitre 5. Ces algorithmes ont été rigoureusement testés pour la validation et la performance avec des signaux empiriques nonlinéaires et les résultats sont présentés dans la Section 5.4.

A.3 Modélisation de distorsion nonlinéaire

Les approches pour gérer la distorsion nonlinéaire dépendent fondamentalement d'un modèle à temps discret du haut-parleur. Dans cette section, nous décrivons les modèles nonlinéaires populaires et illustrons les façons dont ils peuvent être intégrés dans la recherche actuelle afin de mieux comprendre le comportement complexe des systèmes nonlinéaires.

Les systèmes nonlinéaires complexes sont les systèmes nonlinéaires avec mémoire ou systèmes dynamiques nonlinéaires. La méthode la plus courante pour modéliser les systèmes dynamiques nonlinéaires est la série Volterra. La série Volterra est une généralisation

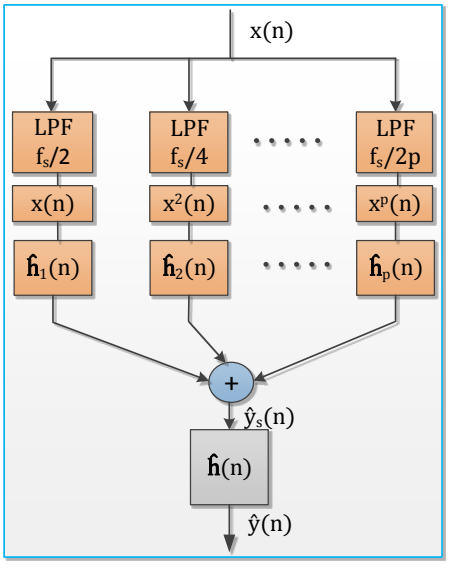


Figure A.5 – Une illustration du modèle en cascade NAEC. Dans le canal p^{th}, le vecteur de signal d'entrée passe à travers un filtre passe-bas (LPF) avec une fréquence de coupure $f_s/2p$ pour éviter l'aliasing.

de l'expansion classique de la série Taylor qui comprend un élément dispersif temporel (mémoire). Puisque cette thèse concerne les systèmes dynamiques, la sortie du système nonlinéaire dépend non seulement de l'entrée instantanée mais aussi des entrées passées. Pour modéliser un tel système nonlinéaire dynamique, la série Volterra tronquée prend la forme suivante [16, 21, 22]:

$$x_{out}(n) = h_0 + \sum_{p=1}^{P} \sum_{i_1=0}^{N_p-1} \sum_{i_2=0}^{N_p-1} \cdots \sum_{i_p=0}^{N_p-1} h_p(i_1, i_2, \cdots, i_p) x(n-i_1) \cdots x(n-i_p) \quad \text{(A.2)}$$

où $h_p(i_1, i_2, \cdots, i_p)$ sont les p^{th}-order *Noyaux de Volterra*, qui caractérisent approximativement le système nonlinéaire . Le terme constant h_0 peut être négligé sans aucune perte de généralité [22]. Il est à noter que le noyau Volterra du premier ordre, $h_1(i_1)$, correspond à la réponse impulsionnelle linéaire du système. Les noyaux de Volterra d'ordre supérieur, $h_p(i_1, i_2, \cdots, i_p), p \in \{2, \cdots, P\}$, sont p matrices-dimensionnelles de taille N_p, et sont généralement supposées symétriques dans les indices $i_1, i_2, \cdots, i_p$. Bien qu'il y ait des avantages à modéliser des systèmes nonlinéaires avec des séries Volterra, il y a un certain nombre de limitations. La limitation la plus commune est sa complexité de calcul. La complexité de calcul augmente exponentiellement avec l'ordre croissant de

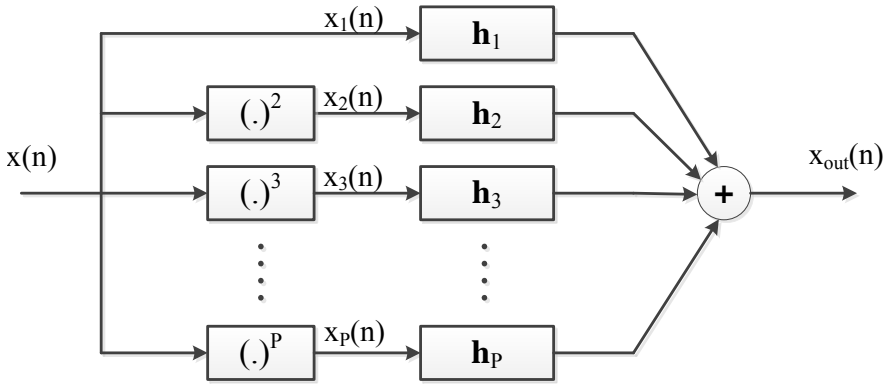


Figure A.6 – le modèle de Hammerstein généralisé polynomiale (GPHM).

nonlinéarité P même pour une longueur de mémoire modeste N_p. Si $N_p = N, \forall p$ alors un noyau Volterra P^{th} order contient les coefficients N^P.

Les systèmes nonlinéaires sans mémoire sont souvent considérés comme la forme la plus simple et la plus couramment implémentée d'un système nonlinéaire. Compte tenu de la nonlinéarité sans mémoire d'un haut-parleur et/ou d'un amplificateur de puissance, la distorsion peut être modélisée à l'aide de l'expansion de série de puissance tronquée:

$$x_{out}(n) = a_0 + a_1 x + a_2 x^2 + \cdots + a_P x^P = \sum_{p=0}^{P} a_p [x]^p \tag{A.3}$$

où $a_0, a_1, a_2, \cdots$ sont des coefficients scalaires. L'expansion de série de puissance est également appelée l'expansion polynomiale ou le modèle Power Series (PSM). La principale limitation avec le PSM est que le modèle suppose que la réponse en fréquence du haut-parleur est plate. Cette hypothèse n'est pas valable dans la pratique, en particulier avec les haut-parleurs miniatures. Par conséquent, une autre façon de modéliser un haut-parleur est de considérer les effets de mémoire (dépendance de fréquence) dans la partie linéaire, un tel système est appelé le modèle de Hammerstein généralisé polynomiale (GPHM), comme le montre la Fig. A.6. La relation entrée-sortie du GPHM est la suivante:

$$x_{out}(n) = \sum_{p=1}^{P} \sum_{i=0}^{L-1} h_p(i) [x(n-i)]^p \tag{A.4}$$

où L est la longueur du filtre linéaire et $h_p(i)$ sont les noyaux Volterra simplifiés (ou

diagonaux) du haut-parleur. Si ces noyaux Volterra simplifiés d'un haut-parleur sont identifiés, il est alors possible de reconstruire la sortie du haut-parleur pour un signal d'entrée donné $x(n)$. Dans cette thèse, nous avons adopté une procédure d'identification simple appelée "technique de convolution nonlinéaire" comme proposé dans [7, 17].

A.3.1 Comparaison des modèles de haut-parleurs

Il y a un manque de théorie unique pour modéliser et caractériser un haut-parleur nonlinéaire. Toute recherche dans NAEC dépend de la précision du modèle de haut-parleur, qu'il soit utilisé pour NAEC lui-même, ou pour synthétiser artificiellement des signaux de test nonlinéaires. Alors que l'approche série-puissance (PSM) offre généralement des performances NAEC efficaces dans des simulations bien contrôlées, même de légères inexactitudes de modèle ont tendance à dégrader les performances dans des conditions réelles. Le modèle polynomial généralisé de Hammerstein (GPHM) a donc été étudié comme un modèle alternatif. Une question se pose maintenant en ce qui concerne la précision du modèle: quel modèle reflète le mieux la distorsion nonlinéaire réelle du haut-parleur? Cette section étudie la possibilité de la modélisation de distorsion de haut-parleur nonlinéaire avec PSM. En outre, la précision des deux modèles est comparée dans l'estimation des sorties de haut-parleurs réels mesurés empiriquement. Les résultats sont publiés dans notre premier article [56].

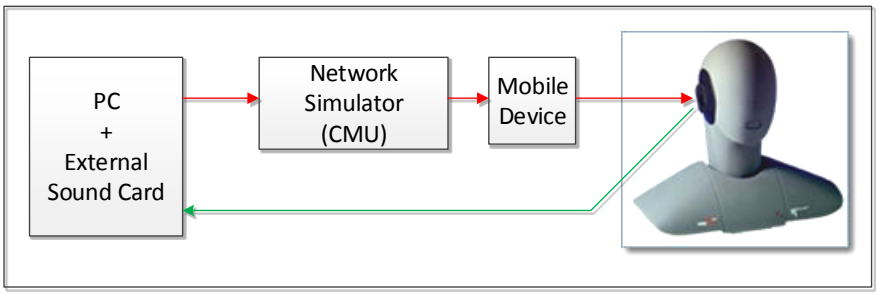


Figure A.7 – Configuration expérimentale utilisée pour l'identification d'un réel haut-parleur de téléphonie mobile

La configuration expérimentale utilisée pour l'identification des haut-parleurs de téléphones mobiles est illustrée sur la figure A.7. Un appareil mobile est placé devant un mannequin de tête et de torse à une distance de 32 cm. L'appareil est configuré pour fonctionner en mode mains libres et au volume maximal pour lequel une distorsion nonlinéaire est assurée. Un signal sinusoïdal exponentiel (en utilisant l'équation 3.6 avec amplitude $a(n) = 1$, fréquences $f_1 = 20Hz$ et $f_2 = 4kHz$ échantillonné à $8kHz$) est joué par le téléphone mobile haut-parleur et enregistré avec le microphone monté dans l'oreille d'un mannequin. Les noyaux Volterra simplifiés $\mathbf{h}_p$, $p \in [1, P]$ d'un véritable haut-parleur

de téléphone portable sont calculés expérimentalement en utilisant le signal de balayage sinusoïdal exponentiel (chirp) basé sur la technique de convolution nonlinéaire [7, 17]. Les détails sur la technique de convolution nonlinéaire et la procédure expérimentale d'identification des noyaux Volterra simplifiés sont décrits dans la section 3.3 et la section 4.1 respectivement.

Génération de signaux synthétiques

Tout d'abord, un signal vocal propre $x(n)$ est joué par le haut-parleur de l'appareil mobile et ensuite enregistré à l'oreille du mannequin en utilisant la même configuration expérimentale représentée sur la figure [A]. Nous appelons le signal vocal enregistré empiriquement $x_{real}(n)$. Maintenant, en utilisant le même signal de parole propre $x(n)$, les signaux de sortie de haut-parleur synthétisés $x_{out}(n)$ sont calculés pour PSM et GPHM en utilisant les équations A.3 et A.4 respectivement.

Pour le GPHM, nous avons utilisé les noyaux de Volterra diagonaux simplifiés $\mathbf{h}_p$, $p \in [1, P]$ mesurés empiriquement à partir d'un haut-parleur de téléphone portable comme décrit dans la section précédente pour la génération de signal . Nous avons considéré les noyaux d'ordre $P = 5$ chacun de longueur $L = 256$ taps, car ils sont plus dominants que les autres nonlinéarités d'ordre supérieur.

Pour le PSM, nous définissons le gain $a_1 = 1$. À des fins de comparaison, nous avons également utilisé 5^{th} order PSM ici. Les composantes de pondération a_p pour $p \in [2, 5]$ sont choisies de sorte que la quantité totale de distorsion nonlinéaire soit la même que celle du GPHM.

Évaluation

Le signal vocal enregistré à l'oreille du mannequin ($x_{real}(n)$) a été comparé aux résultats obtenus selon les deux modèles. Les spectrogrammes du signal de parole propre d'entrée, une réponse de haut-parleur de téléphone mobile, et les deux signaux synthétisés sont illustrés sur la Fig. A.8. Il est évident à partir de la figure que le signal synthétisé avec le GPHM est plus identique au signal vocal réel enregistré (ou mesuré). Le modèle de série de puissance suppose une réponse en fréquence plate qu'un haut-parleur linéaire IR n'a pas, ce qui explique la différence dans le mécanisme de distorsion par rapport au signal réel enregistré. Sans surprise, le signal synthétisé avec le PSM a plus d'énergie aux basses fréquences comme le signal de parole propre original qui ne sont pas réellement présents dans le signal enregistré réel. Le signal enregistré réel a plus d'énergie dans la région haute fréquence ($\sim\geq 1500$) en raison de la distorsion nonlinéaire que le signal

synthétisé en utilisant le GPHM reflète mieux par rapport au PSM.

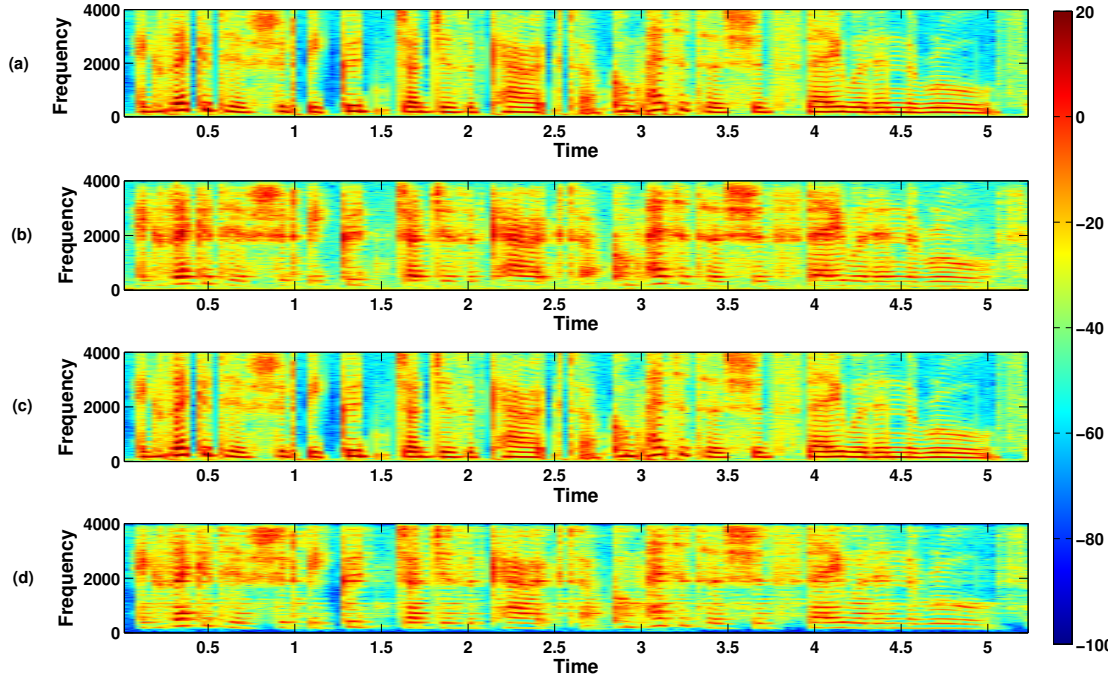


Figure A.8 – Le spectrogramme de (a) signal de parole propre (b) une réponse de haut-parleur de téléphone mobile réel (c) signal de parole synthétisé utilisant PSM (d) signal de parole synthétisé utilisant GPHM

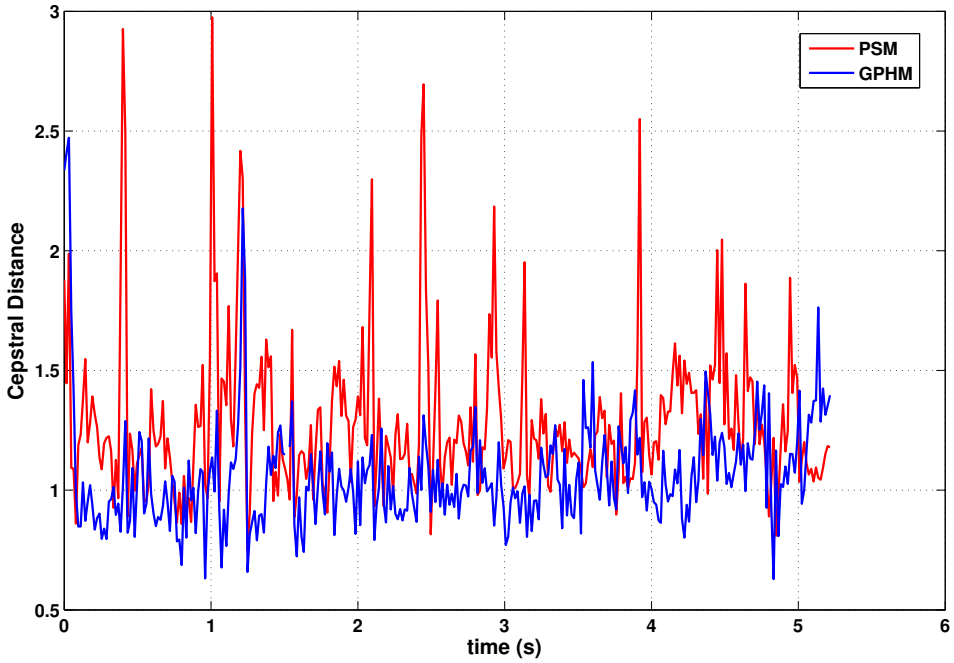


Figure A.9 – Une illustration de la distance cepstrale entre les signaux de haut-parleurs mesurés réels et ceux synthétisés avec les modèles PSM et GPHM.

En outre, la performance est également évaluée objectivement en termes de distance Cepstral (CD):

$$CD(m) = \sqrt{\sum_{L_f} [C_{x_{real}}(m) - C_{x_{model}}(m)]^2} \tag{A.5}$$

où L_f est la longueur de la trame. $C_{x_{real}}(m)$ et $C_{x_{model}}(m)$ sont les vecteurs de colonne des coefficients cepstraux à partir du signal réel enregistré x_{real} et du modèle x_{model} de la frame m^{th} respectivement.

$$C_{x_{real}}(m) = IDFT\{ln\,|DFT[x_{real}(mL_f - 1)\cdots x_{real}((m+1)L_f)]|\} \tag{A.6}$$

Dans tous les cas, les mesures proviennent de trames consécutives de $32ms$ ($L_f = 256$) de longueur. La raison pour laquelle CD est cela, il fournit une évaluation plus corrélée perceptuellement que les approches alternatives basées sur les différences d'énergie ou de puissance. Les profils CD illustrés sur la figure Fig. A.9 montrent que la différence entre le signal mesuré et celui synthétisé avec le modèle GPHM est toujours inférieure à celle entre le signal mesuré et le signal synthétisé avec le modèle PSM. Le modèle GPHM reflète donc mieux le comportement du vrai haut-parleur nonlinéaire.

Ce résultat a également été confirmé par de nombreux tests d'écoute informels qui ont montré que les signaux synthétisés avec le modèle GPHM semblent moins artificiels et sont perceptivement plus proches du signal mesuré que ceux synthétisés avec le modèle de série de puissance.

A.3.2 Impact des signaux d'écho nonlinéaires simulés sur l'évaluation NAEC

Le modèle de série de puissance (PSM) est très souvent utilisé dans la littérature à la fois pour synthétiser les signaux nonlinéaires et pour évaluer une performance de l'algorithme NAEC. Cependant, nous avons vu dans la section précédente que la corrélation entre le PSM et les signaux nonlinéaires réels enregistrés est faible. Dans cette section, nous évaluons la performance d'un algorithme NAEC typique en présence de signaux d'écho nonlinéaires réels et simulés, ce qui est rarement discuté dans la littérature.

Dans cette évaluation, nous avons considéré un système NAEC en cascade populaire

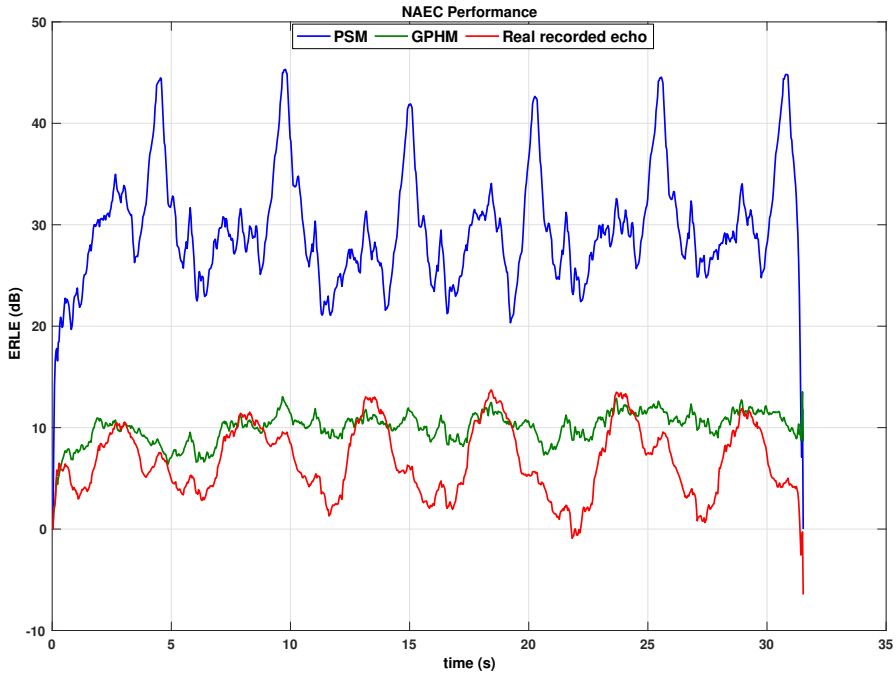


Figure A.10 – Performances NAEC en termes ERLE avec des signaux d'écho nonlinéaires réels enregistrés ou ceux synthétisés avec les modèles PSM ou GPHM.

proposé dans [49]. Les performances du modèle NAEC en cascade ont été étudiées lorsqu'elles ont été exposées aux signaux simulés (PSM et GPHM) et aux signaux d'écho nonlinéaires réels enregistrés. D'abord, les nonlinéarités du haut-parleur ($x_{out}(n)$) sont synthétisées à travers les modèles PSM et GPHM comme décrit dans la section précédente. Ensuite, les signaux de sortie du microphone avec écho nonlinéaire sont générés selon:

$$y\left(n\right) = \sum_{i=0}^{N-1} x_{out}(n-i)h_{rir}(i) \tag{A.7}$$

où $h_{rir}(n)$ est une réponses d'impulsions d'espace (RIR). Les expériences ont été réalisées avec des haut-parleurs diagonaux (unidimensionnels) Volterra kernels $h_p(n)$ pour les valeurs de $p \leq 5$ et avec $L = 256$ taps.

Le canal acoustique a été modélisé avec une réponses d'impulsions d'espace fixe (RIR) de 256-tap $h_{rir}(n)$ sélectionnée à partir de Aachen la base de données RIR [1]. Toutes les expériences ont été effectuées avec un signal de liaison descendante / signal de référence propre $x(n)$ d'une durée d'environ 30 secondes avec une fréquence d'échantillonnage de 8 kHz.

Un ensemble de paramètres de filtre communs a été appliqué aux trois cas de test et a été choisi pour maintenir la stabilité et une meilleure performance. Les configurations de test sont $L_{p=1,\cdots,5} = 256$, $N = 256$, $\mu_{p=1,\cdots,5} = 0.01$, $\delta_{p=1,\cdots,5} = 1e-4$, $N = 256$, $\mu = 0.5$, $\delta = 1e-7$. Les résultats ERLE sont illustrés sur la figure A.10.

Alors que les performances NAEC dans le cas de signaux de haut-parleurs synthétisés avec le modèle PSM sont similaires à celles obtenues dans le travail précédent [49], les performances sont moins bonnes dans le cas de signaux d'écho nonlinéaires réels enregistrés. Alors que les performances NAEC dans le cas de signaux synthétisés de manière empirique avec l'approche GPHM diffèrent également de celles avec des signaux d'écho nonlinéaires réels enregistrés, la différence est significativement réduite.

Ces observations confirment le biais significatif et favorable des résultats générés avec le modèle populaire de PSM et soulignent son influence potentielle sur l'évaluation de la performance NAEC. Les résultats générés avec le modèle GPHM reflètent mieux les mesures pratiques et, par conséquent, les signaux de haut-parleurs générés de manière empirique constituent une alternative attrayante à envisager pour des travaux futurs. Les résultats ainsi obtenus présenteront moins de biais que ceux rapportés précédemment dans la littérature ouverte, et fourniront une estimation plus réaliste de la performance pratique NAEC.

A.3.3 Validation du GPHM

Dans cette section, la précision du modèle GPHM est étudiée en fonction des paramètres clés, à savoir le nombre de prises de filtre L et l'ordre des nonlinéarités P. De cette manière, nous pouvons juger de l'influence de ces paramètres sur les performances du modèle. Les résultats sont publiés dans notre deuxième article [57].

Caractérisation des dispositif

Ce travail implique trois téléphones mobiles différents (téléphones intelligents). Tout d'abord, les noyaux de Volterra diagonaux simplifiés $\mathbf{h}_p, p \in [1, P]$ pour les trois haut-parleurs de téléphones mobiles sont calculés empiriquement en utilisant la procédure décrite dans la section 4.1.

Les performances du modèle ont été évaluées en comparant les sorties du modèle et du haut-parleur réel pour un signal d'entrée commun. Deux signaux d'entrée différents ont été utilisés: (i) le même signal sinusoïdal exponentiel utilisé dans la procédure expérimentale et (ii) un signal vocal réel. De véritables signaux de haut-parleurs ont été enregistrés à l'oreille du mannequin en utilisant le même banc d'essai expérimental

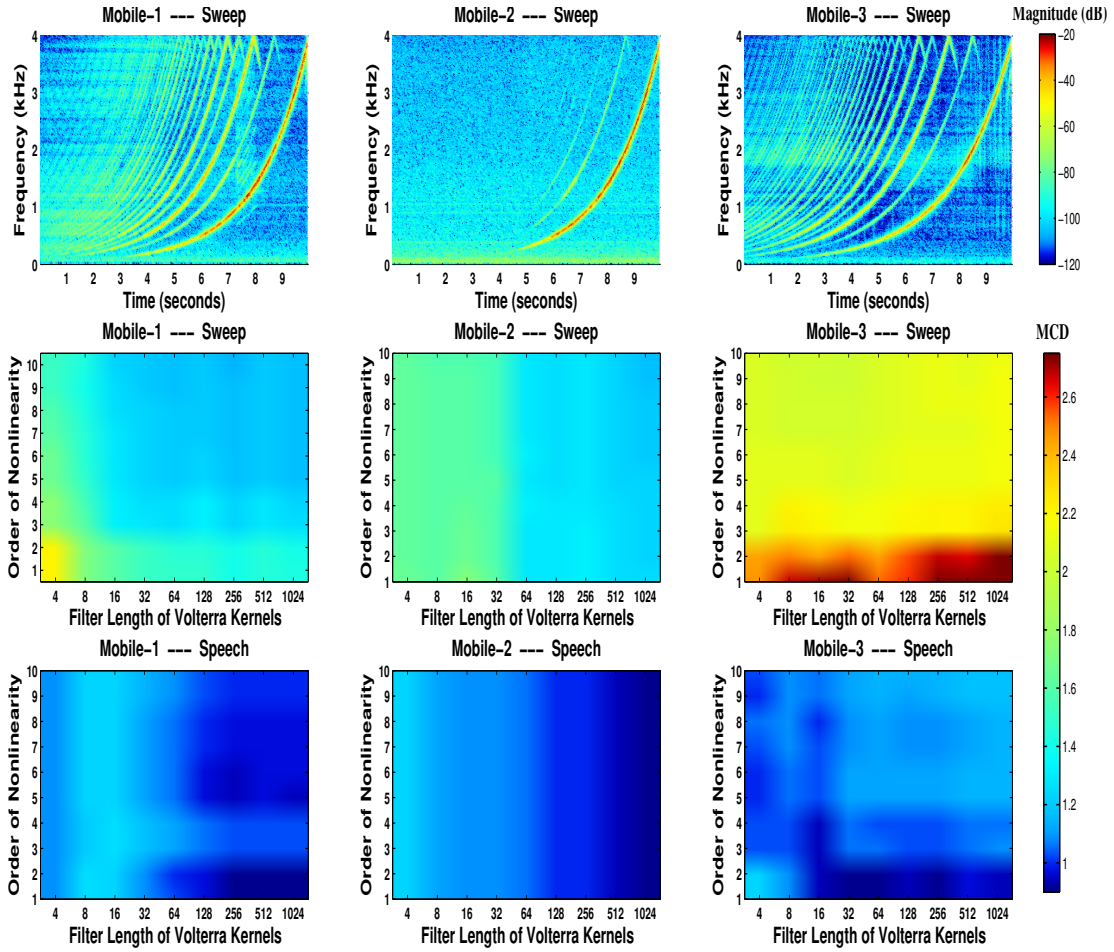


Figure A.11 – Une illustration de la caractérisation nonlinéaire et des performances du modèle GPHM. La première rangée illustre la réponse de chacun des trois dispositifs mobiles au signal d'entrée du balayage sinusoïdal exponentiel. Les lignes deux et trois illustrent les performances du modèle nonlinéaire résultant pour sinus balayent et signaux d'entrée de la parole réelle. Les résultats sont affichés pour différents ordres de nonlinéarité P (axes verticaux) et longueurs de noyau de Volterra L (axes horizontaux).

que celui présenté dans la section précédente. Tous les signaux sont des signaux de modulation par impulsions codées échantillonnées à $8kHz$.

La performance est évaluée objectivement en termes de moyenne distance cepstrale (MCD) entre les signaux réels enregistrés et les estimations du modèle:

$$\begin{aligned} CD(m) &= \sqrt{\sum_{L_f} \left[C_{x_{real}}(m) - C_{x_{model}}(m)\right]^2} \\ MCD &= mean(CD) \end{aligned} \tag{A.8}$$

Les MCD inférieurs indiquent que le modèle reflète plus fidèlement les résultats mesurés réels.

Résultats

La réponse des trois dispositifs au signal d'entrée du balayage sinusoïdal exponentiel est montrée sous la forme de spectrogrammes dans la rangée supérieure de la Fig. A.11. Le premier et en particulier le troisième dispositif (colonnes gauche et droite de la Fig. A.11) présentent une distorsion nonlinéaire significative; les spectrogrammes montrent des harmoniques supplémentaires d'ordre supérieur en plus du signal de balayage sinusoïdal exponentiel d'entrée. La nonlinéarité est en outre asymétrique; les harmoniques d'ordre impair sont plus significatifs que les nonlinéarités d'ordre pair. Nous notons que quelques études indépendantes [9, 63] ont rapporté des observations similaires. En revanche, le second dispositif présente relativement moins de distorsion nonlinéaire.

Les résultats pour chacun des trois dispositifs sont également illustrés sur la Fig. A.11. La rangée du milieu montre les résultats pour le signal d'entrée du balayage sinusoïdal exponentiel alors que la rangée inférieure montre les résultats pour le signal d'entrée du véritable discours. Dans tous les cas, les résultats sont affichés pour différents ordres de nonlinéarité P (axes verticaux) et différentes longueurs de noyau de Volterra L (axes horizontaux). Les couleurs bleues illustrent les MCD inférieurs tandis que les couleurs rouges indiquent des MCD plus élevés.

Pour une performance satisfaisante, l'ordre de nonlinéarité P devrait être suffisamment élevé pour capturer les principales sources de distorsion nonlinéaire, c'est-à-dire les harmoniques les plus dominantes. La longueur de filtre de noyau de Volterra simplifiée L devrait être suffisamment élevée pour capturer avec précision le comportement des haut-parleurs linéaires et nonlinéaires. Les deux paramètres sont cependant un compromis entre performance et efficacité de calcul.

Entrée de balayage sinus exponentiel

La réponse de chaque dispositif au signal d'entrée du balayage sinus exponentiel est illustrée dans la rangée du milieu de la Fig. A.11. Pour les 1er et 3ème appareils, le MCD est plus élevé pour les valeurs inférieures de P, quel que soit le nombre de prises de filtre L. La MCD diminue néanmoins avec l'augmentation de P. Ce comportement n'est pas observé pour le 2ème appareil où, dans tous les cas, le niveau de distorsion nonlinéaire est relativement faible. Il est néanmoins rassurant de constater qu'il y a un changement négligeable dans la précision du modèle pour l'augmentation (surestimée) de P. Pour les 1er et 2ème appareils, le MCD diminue à mesure que la longueur du noyau L augmente. Cependant, pour le troisième périphérique, avec une valeur de $P > 2$, la performance est relativement stable pour varier L. Une explication possible d'un tel comportement est que l'ordre le plus élevé de nonlinéarité significative dépasse celui du modèle ($P = 10$). Puisque le troisième appareil présente une nonlinéarité supérieure au 10ème ordre, P n'est pas suffisant dans ce cas pour réduire le MCD. Par conséquent, des valeurs de $P > 10$ seraient nécessaires lorsque la capacité de traitement le permet.

Entrée de la parole réelle

Les résultats pour les entrées de la parole réelle sont illustrés dans la dernière rangée de la Fig. A.11. En raison de l'aliasing causé par la modélisation de nonlinéarité statique, les valeurs de MCD sont généralement plus faibles pour les entrées vocales que pour les balayages sinusoïdaux. Pour le premier appareil, les meilleures performances sont obtenues pour des valeurs inférieures à P et des valeurs plus élevées de L. Pour le deuxième périphérique, la performance est meilleure pour les valeurs supérieures de L mais est indépendante de P. Pour la troisième performance de l'appareil est le meilleur dans le cas de $P = 1$ et des valeurs de L autour de 64.

Ces résultats montrent que, pour les deux cas où la nonlinéarité est significative, le modèle linéaire ($P = 1$) surpasse le modèle nonlinéaire ($P > 1$) dans le cas des entrées de parole réelle. Malgré l'estimation de la nonlinéarité basée sur une procédure permettant une analyse avancée du système nonlinéaire [7, 8], nos résultats montrent que ce modèle ne correspond pas à l'approximation de nonlinéarité observée avec le signal d'excitation de la parole. Cela conduit à des questions sur les raisons de la discordance observée. Une explication de ce comportement réside dans la plus grande variation d'amplitude pour les signaux de parole par rapport aux signaux de balayage sinusoïdal; les signaux de parole de plus faible amplitude peuvent provoquer une distorsion nonlinéaire significativement moindre. Il est également possible que le modèle obtenu à partir de la réponse du système aux signaux de balayage sinusoïdal soit trop simpliste. Alors que le signal de balayage sinusoïdal consiste en une seule fréquence sinusoïdale à un instant donné, la parole

a une densité spectrale bien plus complexe alors que le modèle néglige les influences inter-spectrales.

Afin de concevoir une solution robuste de NAEC et de traiter la distorsion nonlinéaire dans les haut-parleurs par le traitement du signal, il est apparemment nécessaire de comprendre la dynamique des nonlinéarités du haut-parleur par son origine physique. Nous avons fait une tentative similaire dans la partie suivante de cette thèse en engageant la décomposition modale empirique.

A.4 La Décomposition Modale Empirique

Les techniques traditionnelles d'analyse des données, telles que les approches de Fourier, reposent toutes sur des hypothèses de linéarité et de stationnarité (à court terme). L'analyse en ondelettes [94] a été conçue pour gérer les données non-stationnaires, mais suppose toujours la linéarité. La définition de bases standard et/ou a priori définies pour la représentation du signal est commune à ces deux techniques. Le concept de *fonctions propres* joue un rôle extrêmement important dans l'étude de ces techniques traditionnelles d'analyse des signaux. En général, la décomposition des signaux est basée sur la combinaison linéaire des fonctions propres des systèmes linéaires [95]. En revanche, les systèmes nonlinéaires n'ont généralement pas un ensemble commun de fonctions propres. Par conséquent, les approches traditionnelles d'analyse des signaux sont mal adaptées à l'analyse des signaux nonlinéaires, et leur application directe peut conduire à des effets indésirables et à une interprétation physique sans rapport.

L'analyse des données nonlinéaires et non-stationnaires nécessite cependant des bases dépendant des données ou, de manière équivalente, des bases adaptatives [18]. La Décomposition Modale Empirique (ou EMD, pour "Empirical Mode Decomposition")) [18, 96] est une approche qui répond à cette exigence de fonctions de base dépendant des données nécessaires pour l'analyse adaptative des données. La motivation pour effectuer un EMD est de décomposer de manière adaptative des données nonlinéaires et/ou non-stationnaires en un ensemble de signaux élémentaires, appelés Fonctions Modale Intrinsèque (ou IMFs, pour "Intrinsic Mode Functions"), de manière ad-hoc sans aucune information a priori. Les IMF conservent les caractéristiques des données d'entrée nonlinéaires et peuvent révéler des tendances oscillatoires qui ne sont pas facilement visibles dans le signal d'entrée d'origine. Cette décomposition du signal n'est pas hasardeuse; la sommation directe des IMFs va (re)produire le signal original [97]. Cela permet aux IMFs eux-mêmes d'être utilisés pour le traitement et la manipulation afin d'améliorer efficacement l'amélioration du signal d'entrée. Comme alternative aux approches basées sur Fourier, nous appliquons cette méthodologie à l'analyse de haut-

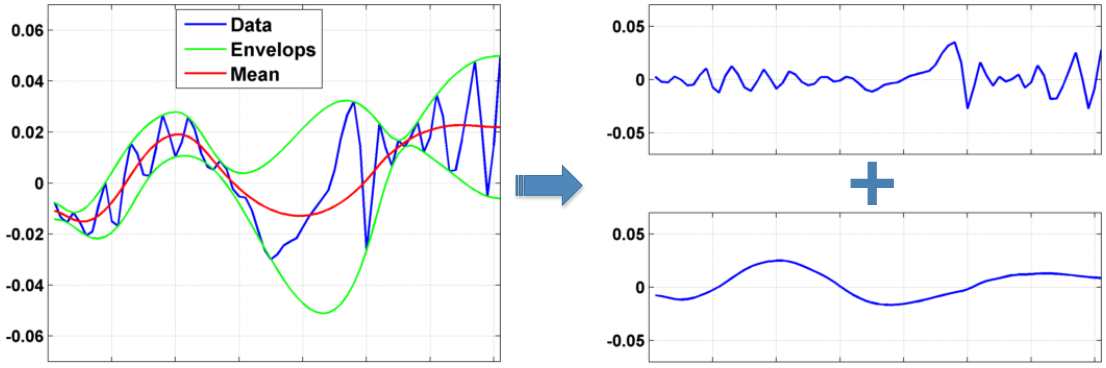


Figure A.12 – Une illustration de l'idée de base d'EMD. Illustré est une donnée parent donnée (ligne bleue dans la figure de gauche) et est considérée comme une oscillation plus rapide (la figure du haut à droite) recouvrant une oscillation plus lente (figure du bas à droite).

parleur nonlinéaire et aux signaux d'écho nonlinéaires.

A.4.1 Introduction à EMD

Une technique relativement récente, la décomposition modale empirique (EMD) suppose que tout signal est composé de différents modes d'oscillations. Cela peut être considéré comme des oscillations plus rapides localement (dans le temps) recouvrant des oscillations lentes [18, 98]. La figure A.12 illustre une telle idée. Le principe de fonctionnement d'EMD consiste à décomposer itérativement un signal complexe en un nombre fini et un très petit nombre de modes empiriques, appelés fonctions modale intrinsèque (IMFs), sans quitter le domaine temporel. Les IMFs sont appelés modes empiriques parce qu'ils ne sont ni prédéfinis ni dans un domaine de transformation particulier, comme c'est le cas avec les techniques d'analyse de signal traditionnelles, mais qu'ils sont dérivés empiriquement sur la base du signal d'entrée. En conséquence, les IMFs servent de fonctions adaptatives de base du EMD et chaque IMF représente une certaine tendance oscillatoire (rapide à lente) dans le signal original. Le signal complexe d'origine peut être complètement reconstruit en additionnant tous les IMF.

Cette section passe en revue EMD en quelques mots. Il existe une abondante littérature théorique et empirique sur l'EMD et son utilisation en sciences appliquées [18, 96]. Tous les détails concernant la mise en œuvre de l'algorithme EMD et les scripts Matlab correspondants sont entièrement disponibles dans [99].

Analyse EMD

La procédure d'analyse de données adaptative ad-hoc que EMD utilise pour extraire des IMFs du signal d'entrée original est appelée *le processus de tamisage (ou the Sifting process)*. En représentant et en analysant les signaux nonlinéaires et non-stationnaires, l'approche de base du processus de criblage a été de décomposer les signaux d'entrée en une combinaison linéaire de modes oscillatoires empiriques (IMF). Les modes empiriques ou les IMFs permettent de mieux comprendre la structure interne du signal et les différents composants impliqués. Pour qu'un signal élémentaire soit un IMF, il doit satisfaire aux deux propriétés importantes suivantes: [18]:

1. Le nombre d'extrema (maxima et minima) et le nombre de passages par zéro dans l'ensemble du signal d'entrée (durée totale du signal) doivent être égaux ou différer d'au plus un.
2. La valeur moyenne de l'enveloppe définie par les maxima locaux et les minima locaux est nulle en tout point.

L'algorithme EMD a été proposé à l'origine pour surmonter les limitations de la transformée de Hilbert, cette dernière sera présentée dans le chapitre suivant. Les deux contraintes ci-dessus admettent les transformations de Hilbert bien comportées. La première contrainte élimine les vagues d'équitation [1] en s'assurant que les maxima locaux d'un signal sont toujours positifs et les minima locaux sont négatifs, respectivement. La deuxième condition rend la forme d'onde symétrique par rapport à l'origine en supprimant toutes les fluctuations indésirables, ce qui simplifie l'analyse des données en extrayant l'information d'amplitude et de fréquence désirée sans résultats paradoxaux contradictoires [100]. Ces conditions garantissent que chaque IMF a un contenu de fréquence localisé en empêchant la propagation de fréquence due aux formes d'onde asymétriques [18].

La procédure complète du processus de tamisage pour décomposer une série temporelle en un ensemble de IMF est illustrée schématiquement à la Fig. A.13. Le lecteur est dirigé vers la discussion du processus de tamisage trouvé dans la section 6.2.

A titre d'exemple, la figure A.14 illustre un ensemble complet de IMFs résultant de l'application de EMD au signal parent $x(n)$. Les IMF sont dérivées itérativement à partir du mode de fréquence le plus élevé, IMF1, au mode de fréquence le plus bas, IMF6.

[1] Les vagues d'équitation sont les oscillations rapides sans aucun passage par zéro entre les extrema. Cela provoque des minima locaux positifs et des maxima locaux négatifs dans le signal. En général, les ondes circulantes sont définies comme des signaux transitoires qui interrompent le modèle prédominant de l'onde [97]

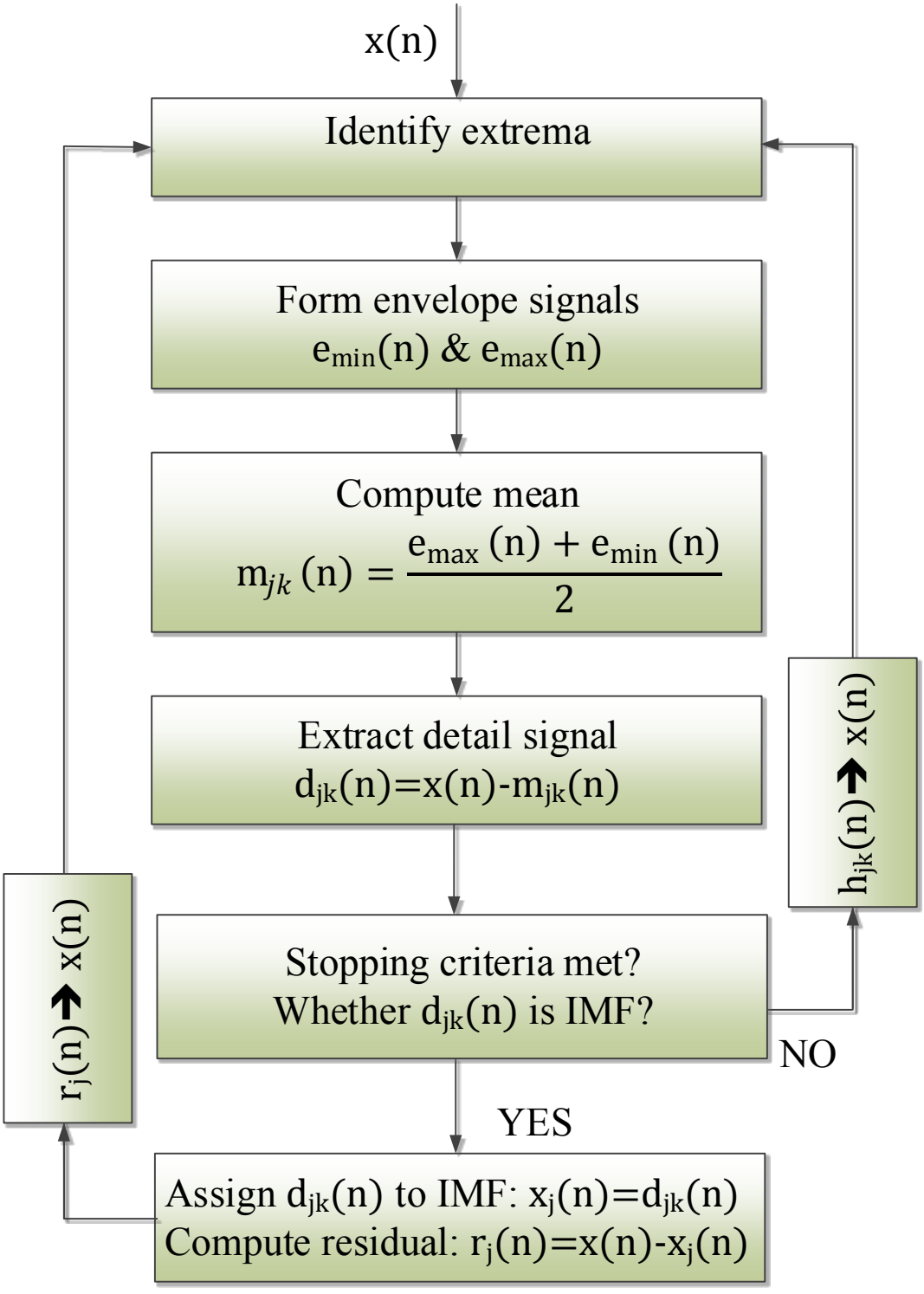


Figure A.13 – L'organigramme illustre la procédure du processus de criblage pour décomposer tout signal compliqué en un ensemble de IMFs.

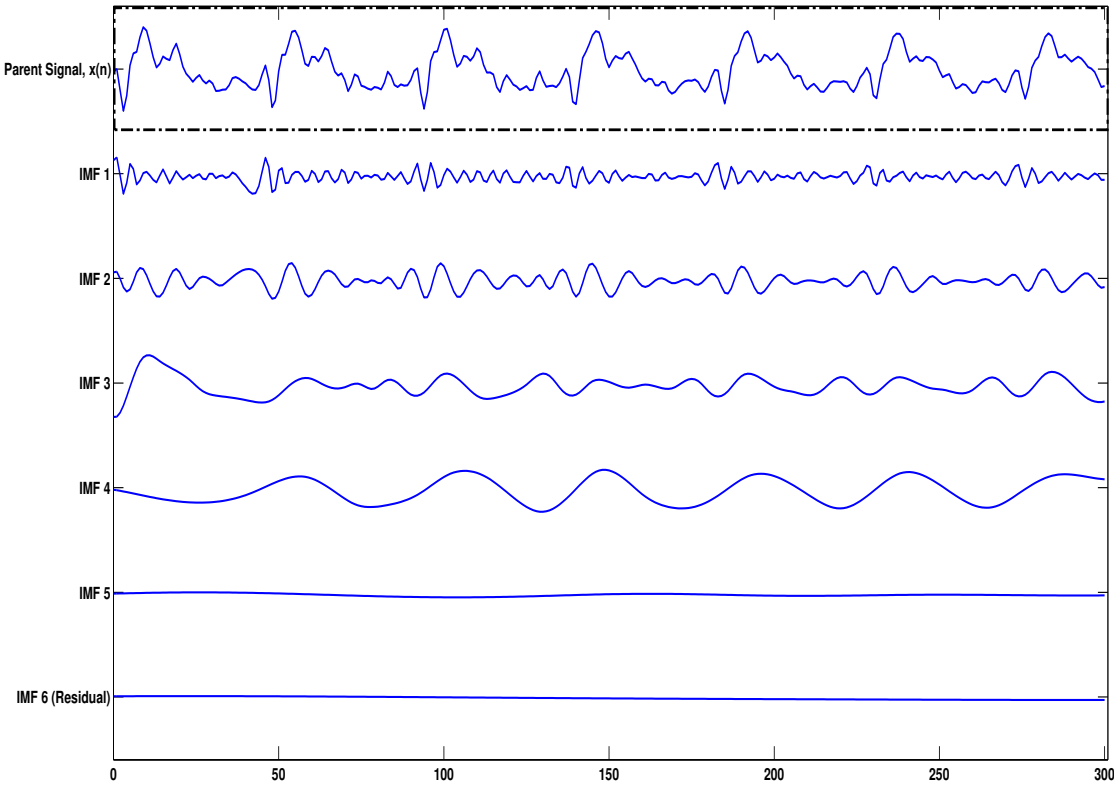


Figure A.14 – Une illustration d'EMD. Illustré est un signal parent donné (en haut) et les 6 IMFs résultants.

Cependant, les IMFs ne sont pas des composantes d'amplitude et de fréquence constantes comme dans l'analyse de Fourier, mais peuvent avoir une modulation d'amplitude et également des fréquences changeantes comme le montre la Fig. A.14. Les IMFs d'ordre supérieur sont ensuite plus lisses car nous enlevons les composants à haute fréquence avant leur extraction. Le signal résiduel (ou le dernier IMF) représente la tendance générale du signal d'entrée.

Synthèse EMD

Le résultat du processus de tamisage produit M IMFs et un signal résiduel constant $r(n)$ ($=r_M(n)$, l'indice M est ignorée par souci de simplicité). Le signal parent ($x(n)$) peut être entièrement reconstruit par le biais du processus de synthèse EMD, qui est simplement une somme directe des IMFs du domaine physique générées par l'EMD:

$$x(n) = \sum_{j=1}^{M} x_j(n) + r(n) \tag{A.9}$$

où $x_j(n)$ est le $j^{ème}$ IMF. Le procédé de tamisage a deux postulats: 1) le signal d'entrée pour le processus de tamisage ($x(n)$) peut être représenté comme une combinaison linéaire de ses IMF et 2) entrer un IMF au processus de criblage résultats en seulement l'entrée IMF avec facteur d'échelle 1. Par conséquent, les IMFs peuvent remarquer que les fonctions propres du processus de tamisage. De plus, les composantes du IMF forment une base complète et "presque" orthogonale pour le signal d'entrée [18, 96, 98]. Par conséquent, la capacité entièrement modulée par les données et l'adaptabilité de la méthode EMD expliquent qu'elle peut être considérée comme bien adaptée aux données nonlinéaires et non-stationnaires.

Applications de l'EMD

EMD effectue la décomposition du signal non supervisé sur la base de l'échelle de temps caractéristique locale des données. En outre, EMD est adaptatif, très efficace et ne laisse pas de domaine temporel. Ces propriétés de l'EMD ont été déclarées bien adaptées aux données nonlinéaires et non-stationnaires et ont incité de nombreux chercheurs à étudier la méthode EMD dans divers domaines de recherche. En conséquence, il y a eu des centaines d'articles dans la littérature au cours de la dernière décennie consacrée à l'application de la technique EMD à diverses applications d'ingénierie et non-ingénierie, par exemple, applications biomédicales [100, 105, 106, 107], traitement d'images et vision par ordinateur [101, 108, 109], météorologie et études climatiques [110, 111], études financières [112], des études sur les ondes océaniques et sismiques [113, 114], l'ingénierie mécanique [115, 116, 117] et de nombreux autres domaines de recherche.

Au cours des dernières années, l'application de l'EMD a également été étendue au traitement du signal vocal et audio. Comme de nombreux signaux du monde réel, les signaux de parole sont également fortement non-stationnaires, ce qui rend l'analyse de signal traditionnelle insatisfaisante en raison de la variation dynamique du contenu spectral à travers le temps. EMD est une meilleure alternative appropriée pour analyser des signaux hautement non-stationnaires comme la parole. Un exemple d'analyse de signal vocal utilisant EMD est illustré sur la figure A.15. Le signal d'entrée est un signal vocal propre échantillonné à 8kHz. EMD décompose le signal de la parole propre dans 18 IMFs; les 6 premiers IMFs sont représentés sur la figure A.15. Le premier IMF a une caractéristique passe-haut, mais contient également une énergie plus basse, un contenu à basse fréquence. Les IMFs d'ordre supérieur ont des caractéristiques de passe-bande superposées [118]. Il est nécessaire de souligner que la fréquence de coupure entre les IMFs consécutifs dépend du temps et du signal d'entrée. Plusieurs efforts ont été récemment déployés pour utiliser les IMF pour améliorer la parole. Les approches de l'amélioration de la parole et de l'élimination du bruit basées sur les EMD sont proposées dans la

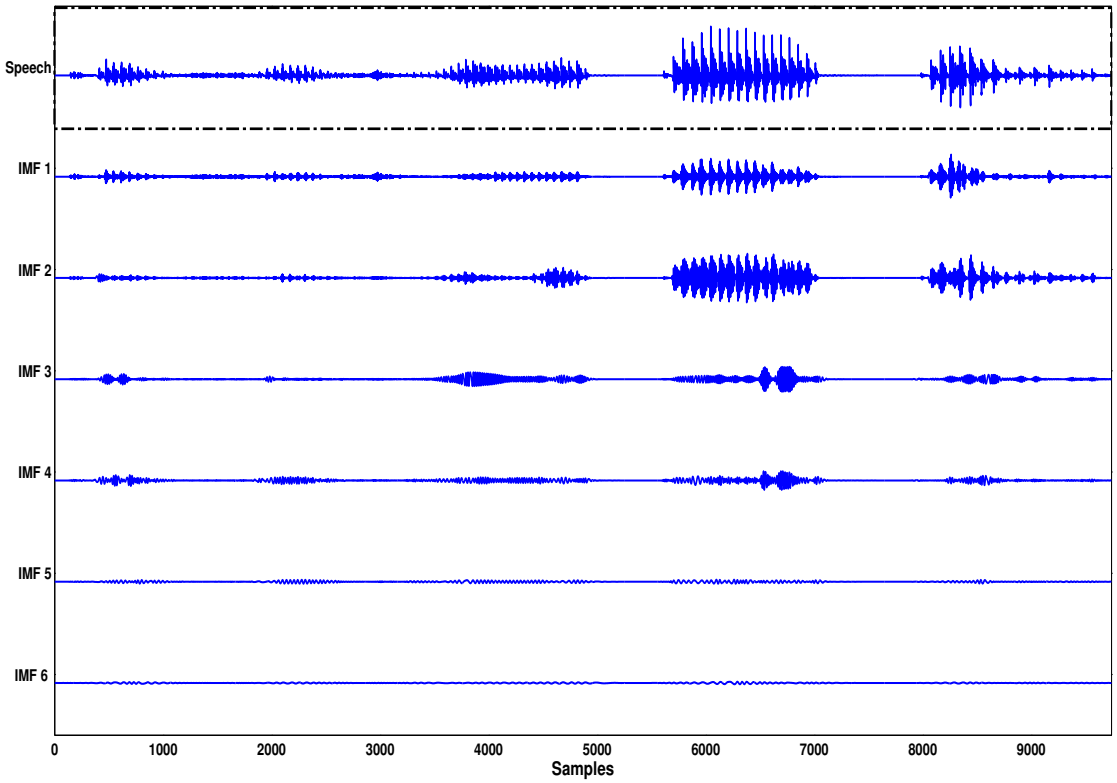


Figure A.15 – Une illustration d'EMD. Illustré est un signal de parole propre (en haut) et les 6 premiers IMFs.

littérature [118, 119, 120]. Récemment, une approche sous-bande basée sur EMD à AEC linéaire est rapportée dans [122]. D'autres applications de l'EMD dans le traitement de la parole et du son comprennent l'analyse de la parole [124], l'analyse de codage prédictif linéaire (LPC) [121], la séparation des sources [125], et l'estimation de hauteur tonale [127].

A.4.2 Application de l'EMD à NAEC

Cette section présente la première approche de NAEC basée sur les EMD. Le travail vise à démontrer l'application d'EMD dans le domaine temporel comme une solution potentielle à NAEC. La technique EMD adaptative aux données est plus appropriée pour décomposer un signal déformé nonlinéaire en un ensemble de IMF qui peuvent être caractérisés en tant que dominante nonlinéaire ou dominante linéaire. Cette classification du IMF en dominants nonlinéaires et linéaires est l'un des facteurs clés de l'annulation d'écho nonlinéaire. Il permet d'éliminer le traitement nonlinéaire pour les signaux à dominante écho linéaire (pour éviter une sur-modélisation) sans dégrader les performances AEC linéaires (dues au gradient-bruit). En profitant de ce principe, nous proposons un

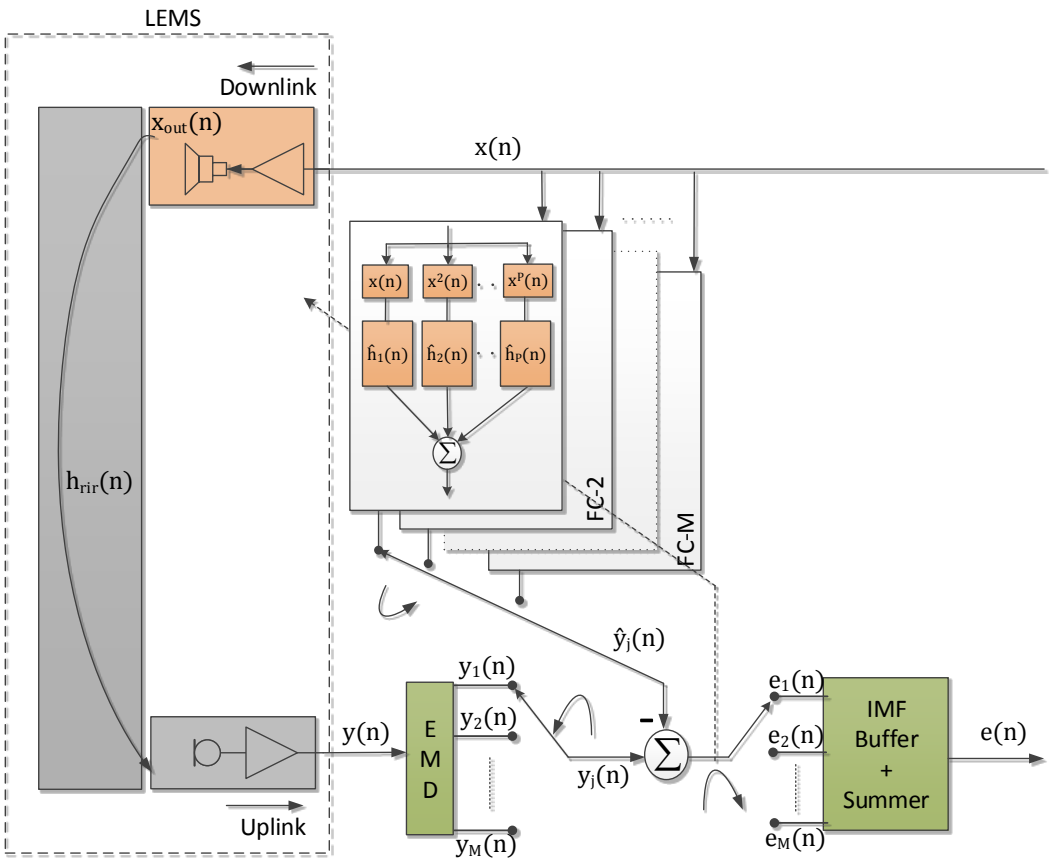


Figure A.16 – Structure de NAEC basée sur EMD.

nouveau schéma de NAEC. Ce travail a été publié dans [93].

L'approche est basée sur la décomposition d'un signal de microphone à bande complète en IMF utilisant EMD. Le NAEC est réalisé grâce à l'application d'un filtrage de puissance adaptatif conventionnel (approche parallèle) à chaque IMF en utilisant un signal de référence en bande complète $x(n)$. La structure du nouveau schéma NAEC basé sur EMD illustré sur la figure A.16 est essentiellement standard sauf pour la décomposition EMD, la re-synthèse et l'utilisation de plusieurs chambres de filtre (ou FC, pour Filter Chambers). Le signal de liaison descendante / référence est noté $x(n)$, le signal de sortie du haut-parleur par $x_{out}(n)$ et le signal de sortie uplink / microphone par $y(n)$. Dans cette première tentative d'employer EMD pour NAEC nous supposons pas de discours proche et pas de bruit de fond. Le signal de liaison montante contient donc uniquement un écho.

La sortie du microphone $y(n)$ est décomposée par EMD en M IMFs selon l'approche décrite dans la Section 6.2. Chaque IMF est alors estimé de manière adaptative à partir du signal de liaison descendante à bande-complète/référence $x(n)$ par l'une des

M chambres de filtration(FCs). Chaque FC contient le $P^{ème}$ order modèle de filtre de puissance classique [48] illustré dans la figure 5.10. Le modèle de filtre de puissance est une approche relativement efficace pour l'identification des chemins d'écho acoustiques nonlinéaires comme discuté au Chapitre 5. Les sous-filtres évaluent de manière adaptative la réponse impulsionnelle du canal acoustique et du haut-parleur, collectivement appelée LEMS, comme illustré sur la Fig. A.16.

La décomposition du signal du microphone $y(n)$ produit des signaux M IMF $y_j; j = 1, \cdots, M$ où chaque IMF représente une gamme de fréquences distincte. En conséquence, chaque FC correspondante nécessite moins de prises de filtre que ce qui serait autrement nécessaire dans le cas d'un signal à bande-complète. Avec un contrôle dépendant de la fréquence sur les IMF, l'ordre des filtres de puissance, P, peut être ajusté individuellement dans chaque FC en fonction des propriétés spectrales du IMF correspondant. Cette structure offre également un degré de liberté supplémentaire pour choisir les paramètres système du modèle de filtre de puissance (tels que l'ordre de nonlinéarité P, la longueur des sous-filtres L_p, les paramètres de filtres adaptatifs, etc.) chaque FC en fonction de la gamme spectrale des IMF. La sortie de chaque FC, $\hat{y}_j(n)$, est soustraite du IMF correspondant, $y_j(n)$, générant ainsi des signaux d'erreur individuels $e_j(n)$. Chaque signal d'erreur est utilisé de manière classique pour mettre à jour les coefficients du sous-filtre FC $\hat{h}_{p,j}(n); p = 1, \cdots, P$ et $j = 1, \cdots, M$. Le paramètre j dans $\hat{h}_{p,j}(n)$ est ignoré pour le reste du chapitre pour des raisons de simplicité et sera appelé $\hat{h}_p(n)$. Enfin, les signaux d'erreur individuels sont sommés ensemble pour reconstruire le signal d'erreur de bande complète:

$$e(n) = \sum_{j=1}^{M} e_j(n) \tag{A.10}$$

Plus de détails sur la structure NAEC et la configuration expérimentale sont discutés dans la Section 6.3.

Ce qui suit rapporte une comparaison des performances de la nouvelle approche basée sur EMD à NAEC à une approche de filtrage de puissance de base. Toutes les expériences ont été menées avec des signaux vocaux et le signal d'écho nonlinéaire est généré de manière empirique (en utilisant le modèle GPHM en identifiant un vrai haut-parleur de téléphone mobile) comme discuté dans les chapitres précédents (voir section 4.1). La performance est évaluée en termes d'ERLE.

Les résultats ERLE pour l'EMD et les approches de filtre de puissance de base pour

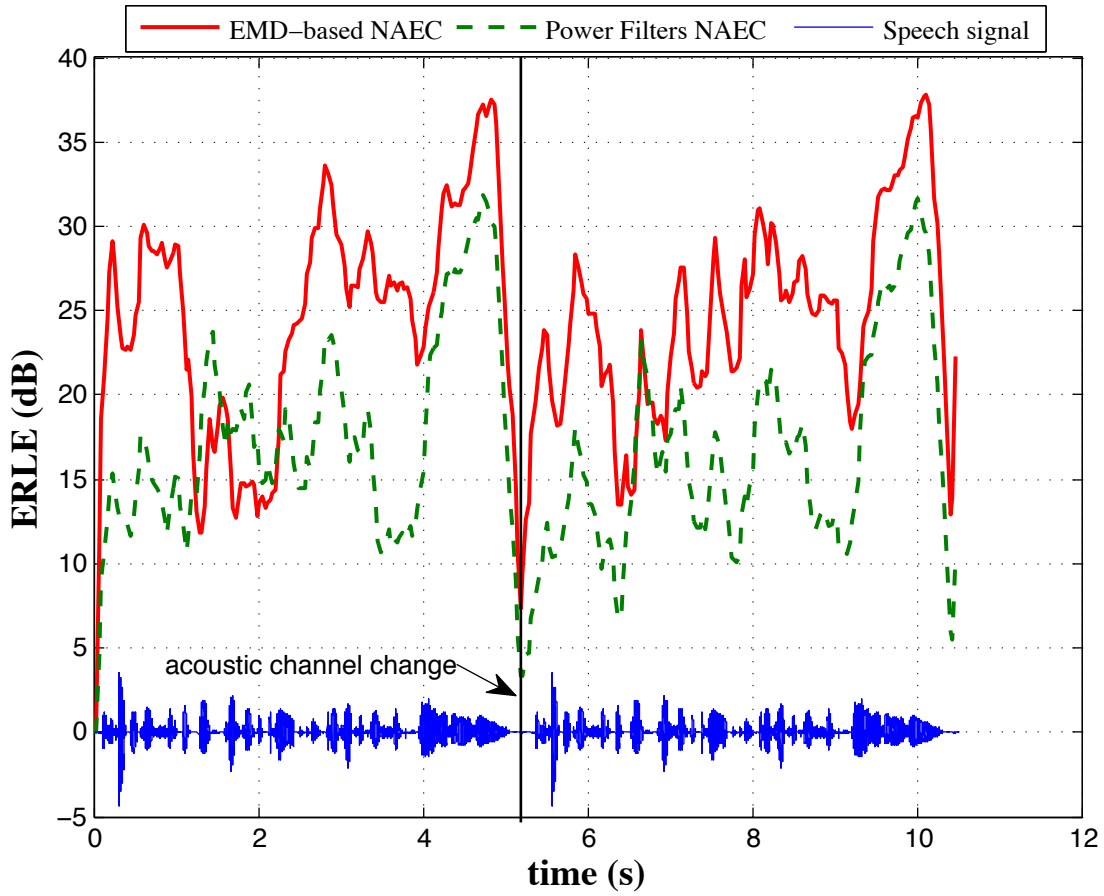


Figure A.17 – Une comparaison des performances en termes d'ERLE pour la nouvelle approche EMD basée sur NAEC et une approche de filtre de puissance de base.

NAEC sont illustrés sur la Fig. A.17 pour une excitation commune. L'approche EMD est montré pour surpasser le système de base; il atteint un niveau plus élevé d'ERLE, environ 8-10 dB de plus que la ligne de base. L'utilisation de différents ordres de filtres de puissance fournit un moyen pratique d'améliorer les performances NAEC, minimisant ainsi le bruit de gradient dû à la sur-modélisation.

Bien que le NAEC basé sur EMD proposé offre non seulement une atténuation d'écho moyenne supérieure, une convergence plus rapide et donc de meilleures performances dans le cas d'un trajet acoustique à changement dynamique, il n'est pas sans coût. Cela entraîne une complexité de calcul accrue, principalement due à la décomposition EMD et à l'utilisation de plusieurs FC. Bien qu'il soit possible de réduire la charge de calcul par une optimisation supplémentaire, le système actuel est environ 1,8 fois plus exigeant en termes de calcul. Bien qu'il existe un algorithme EMD en ligne [96], le travail rapporté ici a été réalisé avec une implémentation "hors ligne", c'est-à-dire par application d'EMD à des signaux entiers. Cela a été délibéré afin de démontrer l'application de l'EMD à

l'annulation d'écho nonlinéaire tout en évitant les problèmes supplémentaires inhérents au traitement en ligne [96].

Bien que la solution NAEC proposée n'utilise pas d'analyse de Fourier (ou d'outils d'analyse de signaux linéaires), l'interprétation sous-jacente de la distorsion nonlinéaire (en tant que distorsion harmonique) dépend toujours de l'analyse temps-fréquence de Fourier et de la série Volterra. Dans le prolongement de l'algorithme EMD décrit ci-dessus, la section suivante introduit une nouvelle méthode d'analyse temps-fréquence appelée transformation de Hilbert-Huang (HHT). Comme méthode alternative à l'analyse temps-fréquence basée sur Fourier, nous allons appliquer cette méthodologie HHT pour analyser les haut-parleurs nonlinéaires dans la section suivante.

A.5 Une interprétation alternative des nonlinéarités de haut-parleur

Cette section présente une nouvelle approche de la caractérisation nonlinéaire des haut-parleurs utilisant la transformée de Hilbert-Huang (HHT). Basé sur l'EMD et la transformée de Hilbert, le HHT décompose les signaux nonlinéaires en bases adaptatives qui révèlent des effets nonlinéaires dans des détails plus grands et plus fiables que les approches actuelles. Les techniques conventionnelles de décomposition du signal, telles que les techniques de Fourier et d'ondelette, analysent la distorsion nonlinéaire à l'aide de la théorie des transformées linéaires. Cela limite la distorsion nonlinéaire à la distorsion harmonique. Ce travail montre que la vraie distorsion de haut-parleur nonlinéaire est plus complexe. HHT offre une autre vue à travers les effets *cumulatifs* des harmoniques et de la modulation d'amplitude et de fréquence intra-onde. Le travail remet en question l'interprétation de la distorsion nonlinéaire par les harmoniques et les points vers un lien entre les sources physiques de nonlinéarité et la modulation d'amplitude et en fréquence. Ce travail, publié dans [128], remet en outre en question la pertinence des approches traditionnelles d'analyse des signaux tout en donnant du poids à l'utilisation de l'analyse HHT dans les travaux futurs.

A.5.1 Fréquence instantanée et transformée de Hilbert

Pour comprendre le comportement complexe des systèmes nonlinéaires, une analyse temps-fréquence approfondie est nécessaire au niveau de précision de la fréquence instantanée (IF) et de l'amplitude instantanée (IA). La fréquence instantanée devrait révéler des détails plus précis du phénomène sous-jacent de la distorsion nonlinéaire.

La transformée de Hilbert (HT) est une technique bien connue dans le traitement du

signal pour calculer la fréquence et l'amplitude instantanées. Le HT peut être interprété comme un déphaseur 90^o. En revenant temporairement à la notation continue, pour toute série temporelle arbitraire, $x(t)$, la transformation de Hilbert (HT), $y(t)$, est obtenue comme suit [130]:

$$y(t) = \frac{1}{\pi} P \int_{-\infty}^{+\infty} \frac{x(\tau)}{t-\tau} d\tau \tag{A.11}$$

où P représente la valeur principale de Cauchy. Avec cette définition, $x(t)$ et $y(t)$ forment une paire conjuguée complexe conduisant à un signal analytique:

$$z(t) = x(t) + jy(t) = a(t)e^{j\theta(t)} \tag{A.12}$$

dans lequel

$$\begin{aligned} a(t) &= \sqrt{(x^2(t) + y^2(t))} \\ \theta(t) &= \arctan \frac{y(t)}{x(t)} \end{aligned} \tag{A.13}$$

Ici, $a(t)$ est l'IA et $\theta(t)$ est la phase instantanée. L'IF peut être calculé comme suit:

$$\omega(t) = \frac{d\theta(t)}{dt} \tag{A.14}$$

Cette définition de IF de la théorie des ondes classique est calculée par différentiation plutôt que par intégration. Le FI est donc local, et non global, et reflète la modulation de fréquence intra-onde [18]. La modulation de fréquence intra-onde représente le changement de FI dans un cycle d'oscillation (ou dans une période d'une onde). Cependant, cette façon de calculer IF et parfois le concept même d'IF a été sujet à controverses et limitations [124, 129, 131, 132]. Cohen [129] a montré que le HT produit un IF significatif pour les signaux *monocomposant* tandis que les théorèmes de Bedrosian et Nuttall [131, 132] imposent

d'autres contraintes, e.g. spectres d'amplitudes non chevauchants ($a(t)$) et les spectres du cosinus ($cos(\theta(t))$). Si pour une fonction donnée, les spectres de $a(t)$ et $cos(\theta(t))$ sont superposés, cette fonction ne peut pas être exprimée sous la forme de signal analytique donnée dans l'équation A.12 . De même, tous les signaux réels avec des minima locaux positifs et des maxima locaux négatifs (les signaux dits multicomposants) ne peuvent pas non plus être exprimés sous la forme d'un signal analytique, ce qui signifie que HT n'existe pas. Malheureusement, ces conditions sont trop restrictives et la plupart des données pratiques ne répondent pas à ces exigences. En conséquence, le plein potentiel du HT a dû attendre le développement de la décomposition modale empirique (EMD).

A.5.2 La transformation de Hilbert-Huang (HHT)

Comme discuté dans le section précédent, le EMD a été introduit pour résoudre les limites de la HT. Rappelant ce qui a été dit dans le section précédent, EMD décompose tout signal en un ensemble fini de signaux élémentaires à travers le processus de tamisage et pour qu'un signal élémentaire soit un IMF, il doit satisfaire les deux propriétés importantes suivantes [18]:

1. Le nombre d'extrema (maxima et minima) et le nombre de passages par zéro dans l'ensemble du signal d'entrée (durée totale du signal) doivent être égaux ou différer d'au plus un.

2. La valeur moyenne de l'enveloppe définie par les maxima locaux et les minima locaux est nulle en tout point.

Les IMF générées par l'EMD satisfont aux contraintes et/ou aux limitations d'admettre des HT bien agencées et des fréquences locales instantanées significatives en fonction du temps. EMD avec le HT est ce que Huang et al. Ont appelé la transformation de Hilbert-Huang (HHT) [18]. La transformation de Hilbert-Huang (HHT) est une approche d'analyse de signal qui est bien adaptée aux signaux nonlinéaires et non-stationnaires [18,96]. L'application de HHT implique deux étapes. Le premier décompose un signal discret du domaine temporel $y(n)$ en un ensemble de M fonctions modale intrinsèque (IMFs), $y_j\,(n)\,;\, j = 1, \cdots, M$, en utilisant la décomposition modale empirique (EMD) de telle sorte que:

$$y\,(n) = \sum_{j=1}^{M} y_j\,(n) + r\,(n) \tag{A.15}$$

où $r(n)$ est le résidu. La deuxième étape détermine la fréquence instantanée (IF) et l'amplitude instantanée (IA) de chaque IMF y_j en utilisant la transformation de Hilbert. A partir de ceux-ci, on peut construire directement la distribution temps-fréquence-énergie appelée le spectre de Hilbert [18, 96].

Hilbert-Huang Spectrum

Le HT est facilement appliqué à chaque IMF afin de déterminer l'IA ($a_j(n)$; $j = 1, \cdots, M$) et IF ($\omega_j(n)$; $j = 1, \cdots, M$) selon Eqs. A.13 and A.14 respectivement. La représentation analytique du signal d'entrée peut alors être exprimée comme:

$$y'(n) = \sum_{j=1}^{M} a_j(n) e^{i \int \omega_j(n) dn} \tag{A.16}$$

où, puisqu'il est constant, le résidu $r(n)$ est omis. Le signal d'entrée original, $y(n)$, est la partie réelle du signal analytique. Les IAs ($a_j(n)$; $j = 1, \cdots, M$) et IFs ($\omega_j(n)$; $j = 1, \cdots, M$) alors donne une représentation temps-fréquence-amplitude du signal, appelée spectre de Hilbert-Huang [18, 96]. Une représentation graphique de la distribution temps-fréquence de IA^2 (carré l'amplitude) illustre la densité d'énergie de la même manière qu'un spectrogramme conventionnel.

Relation aux techniques de Fourier

Exprimé en tant que somme de sinusoïdes, le signal d'entrée est donné par:

$$y'(n) = \sum_{j=1}^{\infty} a_j e^{i \omega_j n} \tag{A.17}$$

où a_j et ω_j sont respectivement des termes d'amplitude et de fréquence constants. Puisque la fréquence de chaque fonction sinusoïdale est indépendante du temps, l'analyse de Fourier est capable de construire des données stationnaires seulement. De plus, comme les ondes sinusoïdales utilisées pour décrire un signal ont une étendue infinie, l'analyse de Fourier est considérée comme un outil d'analyse globale. La précision dépend donc de manière critique de la longueur et de la stationnarité des données, mais les données pratiques sont généralement de courte durée et de durée arbitraire.

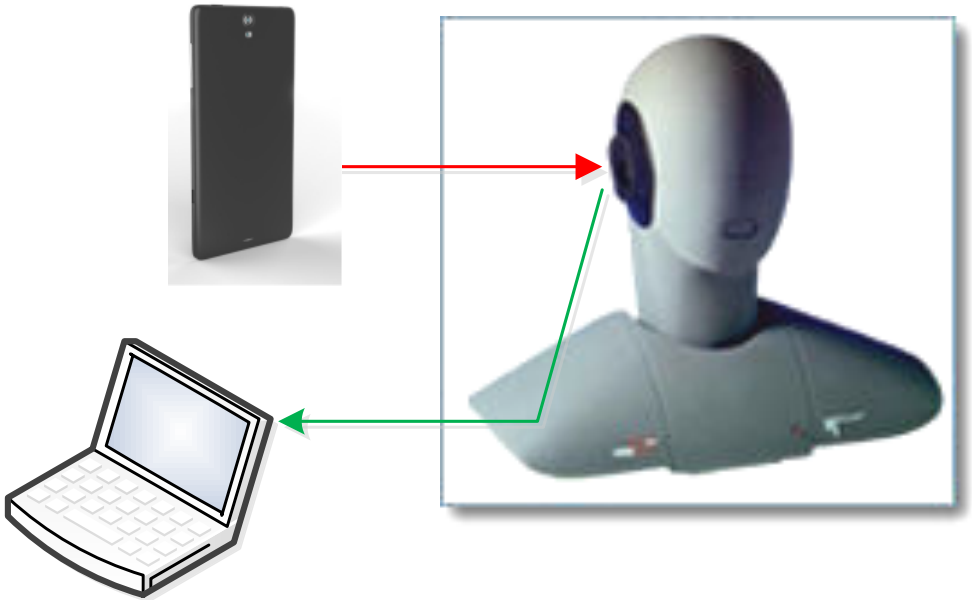


Figure A.18 – Montage expérimental dans une chambre anéchoïque pour mesurer les sorties des haut-parleurs.

La comparaison des Eqs. A.16 et A.17 montre que le HHT est une expansion de Fourier généralisée mais avec une amplitude et une fréquence variables dans le temps qui accommodent des données nonlinéaires et non-stationnaires. La représentation de Fourier implique une énergie constante à une fréquence donnée, c'est-à-dire une onde harmonique régulière qui reste inchangée tout au long de l'enregistrement complet. L'analyse HHT, en revanche, reflète la probabilité *local* d'énergie à une fréquence donnée. Une brève description de l'analyse HHT est présentée dans cette section et pour des présentations plus détaillées, les lecteurs sont référés à [18, 96, 98].

A.5.3 Analyse de distorsion de haut-parleur

Cette section présente notre première tentative d'application de HHT à l'analyse de la distorsion nonlinéaire produite par des haut-parleurs miniatures. Le travail est notre première étape pour aligner l'analyse de la distorsion nonlinéaire sur ses origines physiques. Ce travail a été réalisé à l'aide de véritables enregistrements de haut-parleurs de téléphones mobiles.

La réponse nonlinéaire d'un haut-parleur est observée depuis sa sortie vers un seul signal d'excitation sinusoïdal. Cette approche a été utilisée pour caractériser un véritable haut-parleur de téléphone portable placé devant un mannequin de tête et de torse à une distance de 30 cm dans une chambre anéchoïque. La montage expérimentale utilisée est illustrée sur la Fig. A.18. Le dispositif est configuré pour fonctionner en mode mains

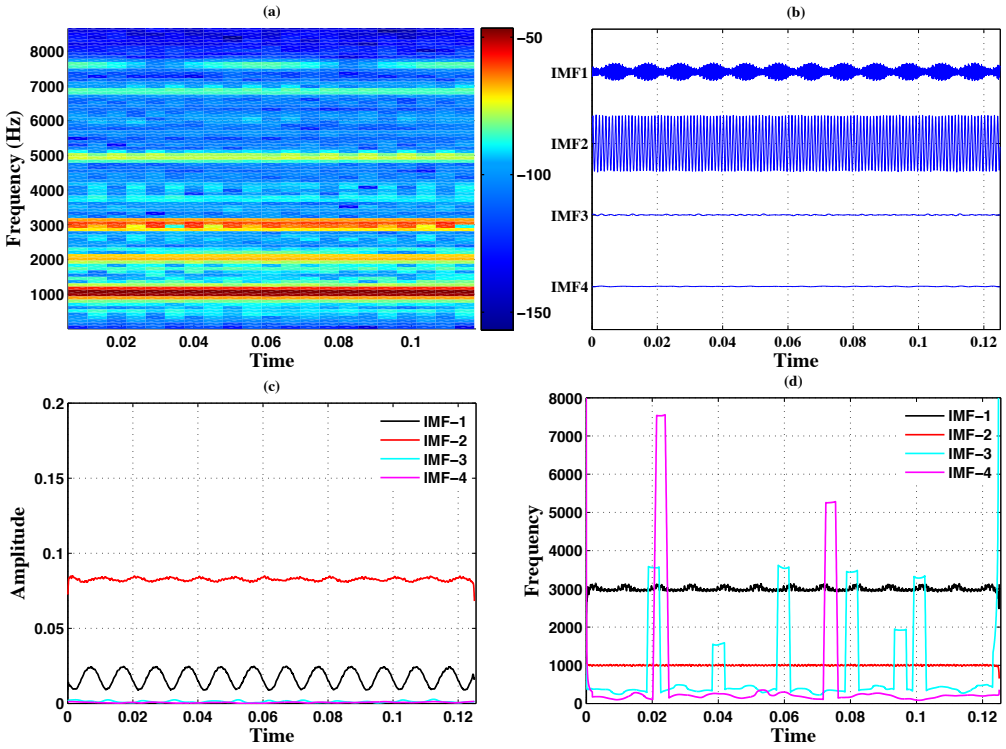


Figure A.19 – (a) Spectrogramme STFT de la réponse d'un haut-parleur de téléphone mobile à une entrée sinusoïdale pure à 1 kHz, échantillonnée à 48 kHz; (b) La réponse du haut-parleur au tonus sinusoïdal de 1 kHz est décomposée par le EMD, ce qui donne 8 IMFs, les 4 premiers IMF sont énumérés ci-dessus et les autres ne sont pas affichés puisqu'ils sont presque nuls; (c) Profils IA des IMF obtenus par HHT; (d) Profils IF des IMF obtenus par HHT

libres et au volume maximal auquel la distorsion nonlinéaire est la plus grande. Les signaux d'entrée échantillonnés à 48 kHz sont des sinusoïdes pures avec des fréquences comprises entre 100 Hz et 3800 Hz à des intervalles de 100 Hz. Ils sont stockés dans la mémoire du téléphone portable et lus à l'aide d'un lecteur VLC pré-installé. Les sorties des haut-parleurs sont enregistrées avec un microphone (linéaire) de haute qualité monté dans l'oreille du mannequin. Les signaux enregistrés sont stockés sur un PC à la même fréquence d'échantillonnage de 48 kHz.

A titre d'exemple, nous considérons un véritable haut-parleur de téléphone portable soumis à une seule excitation sinusoïdale de fréquence 1kHz. Fig. A.19(a) montre les résultats de l'analyse STFT. Plusieurs harmoniques d'ordre élevé sont visibles, représentant la vision traditionnelle de la distorsion nonlinéaire.

La Fig. A.19(b) illustre les quatre (sur huit) IMFs qui résultent de la décomposition du signal de haut-parleur en utilisant EMD et les routines disponibles dans [99]. Puisque

EMD extrait le IMF le plus élevé en premier, IMF-1 est l'harmonique déformée causée par les nonlinéarités du haut-parleur. IMF-2 est la fréquence naturelle déformée à 1kHz alors que les autres IMF ont une énergie négligeable.

La Fig. A.19(c) illustre les profils IA des quatre composantes du IMF qui présentent une modulation d'amplitude intra-onde, à savoir une variation de l'amplitude dans le temps. La Fig. A.19(d) illustre les profils IF correspondants qui présentent une modulation de fréquence intra-onde. Ceci est dû au déplacement du diaphragme du haut-parleur qui n'est plus une fonction purement sinusoïdale en raison d'une distorsion nonlinéaire. Une harmonique de troisième ordre relativement forte est également générée en raison de nonlinéarités de haut-parleurs asymétriques.

La déformation du profil d'onde provoquée par la distorsion nonlinéaire est le résultat du contenu harmonique accumulé et de la modulation d'amplitude et de fréquence intra-onde. Cet effet *cumulatif* est observé dans la réponse dans le domaine temporel du haut-parleur représenté sur la figure A.20. La déformation de la forme d'onde n'est pas constante, mais varie de haut en bas et vice versa, conformément au profil IA de la figure A.19(c). L'ampleur de la déformation dépend de l'amplitude des harmoniques supplémentaires et de la force de la modulation d'amplitude et de fréquence intra-onde. Une observation attentive des profils IA et IF sur les figures A.19(c) & (d) montre respectivement que la variation de fréquence des composantes du IMF augmente lorsque leur amplitude diminue et vice versa. Ceci est indicatif de *adoucissement* nonlinéarité [133].

Les effets décrits ci-dessus ne sont pas reflétés dans le spectrogramme STFT traditionnel qui montre à la place des harmoniques fallacieux. Les estimations dérivées du HHT peuvent donc refléter un comportement nonlinéaire plus fiable que les estimations dérivées de STFT. Huang et al. dans [18] a déclaré que la modulation de fréquence intra-onde est la marque de la distorsion nonlinéaire, où la fréquence du système change avec la position même en une période d'oscillation. Par ailleurs, les auteurs soutiennent que les bases a priori définies dans les techniques traditionnelles d'analyse du signal imposent de nombreuses harmoniques et qu'elles ne sont rien d'autre qu'un artefact mathématique, sans lien avec une source physique [18, 96]. Contrairement aux approches traditionnelles, EMD adapte les bases au signal lui-même et peut donc produire des résultats plus pertinents physiquement. L'analyse HHT conduit à une nouvelle interprétation physique de la distorsion nonlinéaire. Au lieu de la distorsion harmonique est le concept de l'effet *cumulatif* du contenu harmonique et de la modulation d'amplitude et de fréquence intra-onde. La vraie question, cependant, est de savoir si la modulation de fréquence et d'amplitude illustrée sur la figure A.19 a une source physique réelle ou s'il s'agit simplement d'un artefact du HHT.

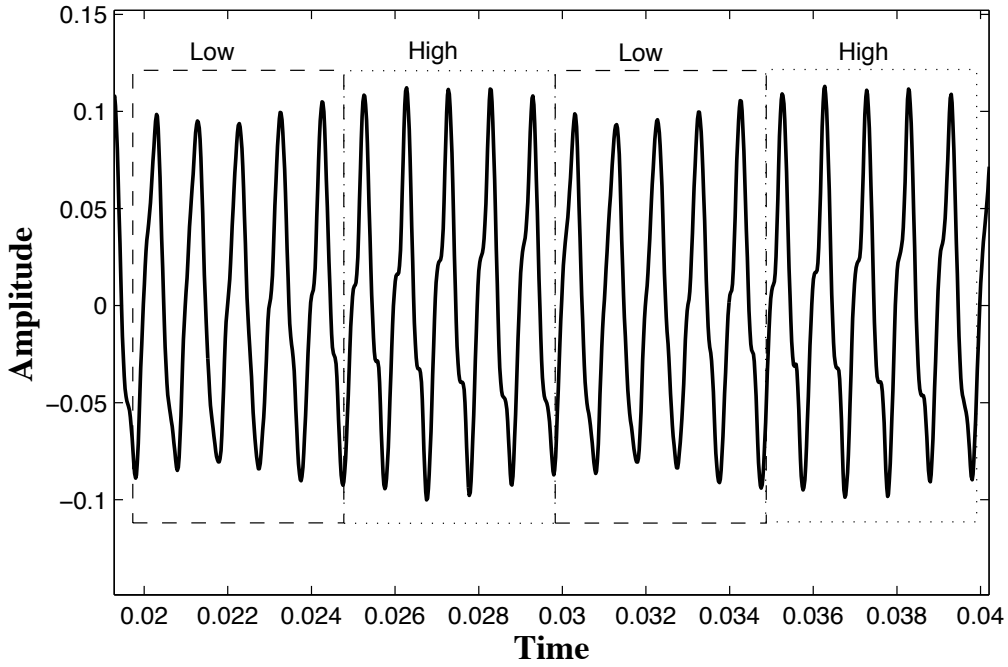


Figure A.20 – Une vraie réponse de haut-parleur de téléphone mobile à ton pur sinus 1kHz. La déformation du profil d'onde causée par la distorsion nonlinéaire n'est pas constante tout au long du temps.

A.5.4 Validation de HHT

Le HHT est entièrement validé dans [18] avec des exemples analytiques. Cette section vise à valider la technique HHT comme moyen de caractériser le comportement des haut-parleurs nonlinéaires.

Les figures A.21(a) and A.21(b) illustrent les profils spectrogramme et IF d'une réponse de haut-parleur de téléphone portable à une entrée sinusoïdale pure à 500Hz. Les profils IF affichent *cumulative* la distorsion harmonique et la modulation nonlinéaire. Il n'y a qu'une faible harmonique de troisième ordre et une modulation significative de l'amplitude et de la fréquence intra-onde, alors que le spectrogramme montre une harmonique de troisième ordre forte et plusieurs harmoniques faibles. Fig. A.21(c) montre les profils IF correspondants lorsque le haut-parleur du téléphone mobile est remplacé par un haut-parleur (linéaire) de haute qualité et montre une absence totale de modulation amplitude-fréquence. Fig. A.21(d) montre les profils IF de la sortie du haut-parleur (haute qualité) lorsqu'une distorsion harmonique de troisième ordre simulée est ajoutée à l'entrée. Encore une fois, il n'y a pas de modulation d'amplitude et de fréquence indiquant que la distorsion observée en (b) a des origines physiques et n'est pas simplement un artefact du traitement HHT.

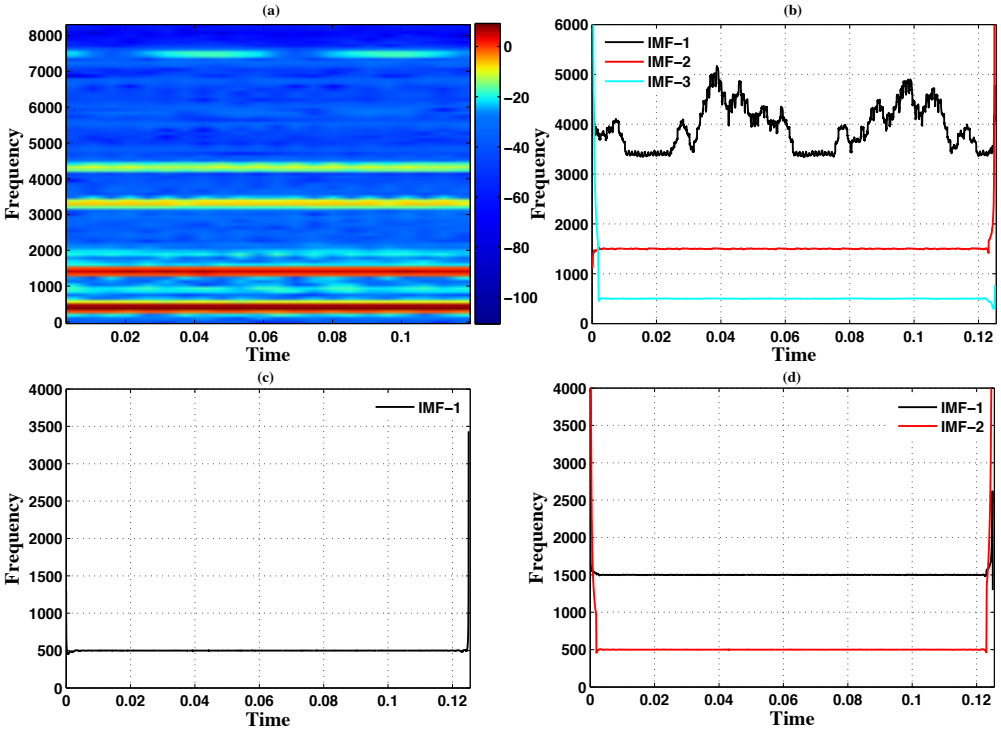


Figure A.21 – Distributions temps-fréquence-énergie: a) spectrogramme STFT de la réponse d'un haut-parleur de téléphone mobile à une entrée sinusoïdale pure à 500 Hz, échantillonnée à 48 kHz; (b) les profils IF obtenus par HHT; (c) profils IF pour une réponse de haut-parleur de haute qualité à la même entrée; (d) les profils IF du même haut-parleur soumis à une excitation d'entrée composée de sinusoïdale pure à 500 Hz et de son troisième harmonique.

Après avoir analysé les trois différentes sorties de micro haut-parleurs pour des excitations sinusoïdales pures à différentes fréquences, nous avons pu déterminer que la distorsion nonlinéaire est causée par l'effet cumulatif du contenu harmonique et la modulation de fréquence et d'amplitude intra-onde. En outre, le contenu des harmoniques dépend du niveau du signal d'entrée. Pour un micro haut-parleur sur-piloté, le contenu harmonique est plus fort mais est limité à la distorsion du troisième ordre. À un niveau d'excitation modéré, la distorsion de modulation est plus préjudiciable que le contenu harmonique. D'autre part, la force de la modulation de fréquence et d'amplitude intra-onde dépend nonlinéairement du niveau et de la fréquence du signal d'entrée.

La modulation de fréquence intra-onde est plus forte même à des niveaux d'excitation modérés si la fréquence du signal d'entrée est proche de la fréquence de résonance naturelle du micro haut-parleur par rapport aux autres fréquences. La déformation de la forme d'onde se développe dès que l'indice de modulation de fréquence intra-onde (ou le pourcentage de modulation) dépasse un certain seuil, ce qui est différent pour différents

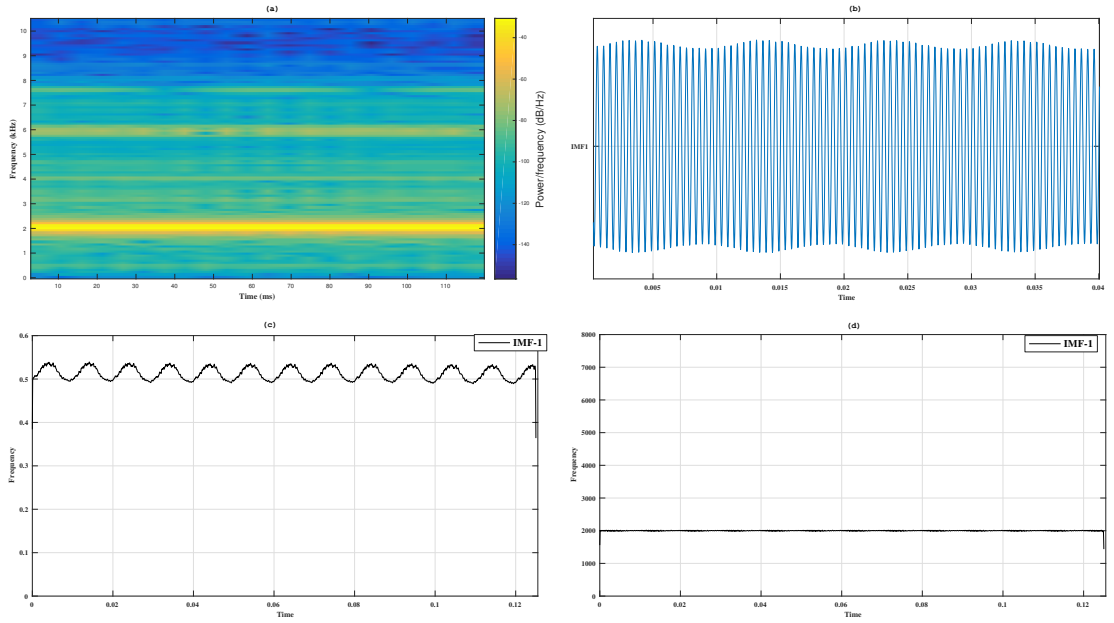


Figure A.22 – a) spectrogramme STFT de la réponse d'un haut-parleur de téléphone mobile à une entrée sinusoïdale pure à 2 kHz, échantillonnée à 48 kHz; (b) La réponse du haut-parleur au signal sinusoïdal de 2 kHz (zoom avant) est décomposée par le EMD, ce qui donne un seul IMF, ce qui signifie que la réponse du haut-parleur satisfait elle-même aux propriétés EMD; (c) profil IA du IMF obtenu par HHT; (d) Profil IF du IMF obtenu par HHT, indiquant un très faible pourcentage de modulation

micro haut-parleurs. Par exemple, la figure A.22 représente l'analyse STFT et HHT d'une réponse de micro haut-parleur à un signal sinusoïdal de 2 kHz. Dans la Fig. A.22(a), le spectrogramme traditionnel montre une troisième harmonique relativement plus forte et une seconde harmonique faible. La réponse réelle du haut-parleur est représentée par un IMF1 sur la Fig. A.22(b) n'indique aucune distorsion visuelle de la forme d'onde sauf la modulation d'amplitude. La même chose peut être observée dans l'analyse HHT en termes de IA et IF sur les Fig. A.22(c) et A.22(d) respectivement. Comme il n'y a pas de déformation de la forme d'onde dans le signal, l'indice de modulation de fréquence intra-onde est très faible et aucun signe de contenu harmonique. Par conséquent, il est sûr de considérer que le degré global de nonlinéarité produit par un micro haut-parleur est largement dû à la distorsion de modulation avec un impact limité dû au contenu harmonique.

En partant de ces résultats empiriques, nous pensons que cette nouvelle interprétation alternative des nonlinéarités des haut-parleurs peut être potentiellement appliquée pour résoudre le problème NAEC. Une solution possible pour estimer l'écho nonlinéaire consiste à incorporer l'effet de modulation amplitude-fréquence intra-onde comme pré-processeur dans le système NAEC pour modéliser la distorsion du haut-parleur et un filtre adaptatif

linéaire traditionnel en cascade pour modéliser l'echopath acoustique linéaire. . Le principal avantage par rapport aux solutions traditionnelles est que le système NAEC ne nécessite pas de nombreux ordres d'harmoniques pour modéliser les nonlinéarités de liaison descendante car les modulations intra-ondes intègrent les nonlinéarités inhérentes dans l'echopath acoustique. Cependant, le développement d'un modèle probabiliste qui émule explicitement l'effet cumulatif du contenu harmonique et de l'effet de modulation de fréquence et d'amplitude intra-onde dans les micro haut-parleurs est une partie très difficile de la conception du NAEC. Cette interprétation alternative peut également être étendue à d'autres domaines de recherche similaires comme la modélisation de haut-parleurs et la linéarisation de haut-parleurs.

En dépit de nombreux avantages caractéristiques de l'utilisation de la technique EMD/HHT pour l'étude des signaux nonlinéaires et non-stationnaires, la technique de décomposition actuelle souffre encore de certaines limitations. Plus de détails sur les limitations EMD et les avancées récentes de la norme EMD sont discutés dans les sections 7.6 et 7.7 respectivement.

En conclusion, cette thèse est l'occasion de faire un premier pas substantiel vers la recherche de réponses à des questions comme ce que la distorsion nonlinéaire réelle provoque aux signaux. La thèse commence par une interprétation traditionnelle de la distorsion nonlinéaire dans les haut-parleurs et se termine par une interprétation nouvelle et précise de la distorsion nonlinéaire, qui marque un nouveau début de la recherche NAEC.

Bibliography

[1] M. Jeub *et al.*, "Aachen impulse response (air) database," 2009. [Online].

[2] E. Hansler and G. Schmidt, *Acoustic Echo and Noise Control: A Practical Approach*. Wiley-Interscience, 2004.

[3] M. Mossi, *Non-linear acoustic echo cancellation with loudspeaker modelling and pre-processing*. PhD thesis, University of Nice - Sophia Antipolis, Oct. 2012.

[4] A. N. Birkett and R. A. Goubran, "Limitations of handsfree acoustic echo cancellers due to nonlinear loudspeaker distortion and enclosure vibration effects," in *Proc. ASPAA*, Oct. 1995.

[5] M. I. Mossi, N. W. D. Evans, and C. Beaugeant, "An assessment of linear adaptive filter performance with nonlinear distortion," in *Proc. ICASSP*, Mar. 2010.

[6] M. Mossi, N. W. D. Yemdji, C.and Evans, and C. Beaugeant, "A comparative assessment of noise and non-linear echo effects in acoustic echo cancellation," in *Proc. ICSP*, Oct. 2010.

[7] A. Farina, A. Bellini, and E. Armelloni, "Non-linear convolution: A new approach for the auralization of distorting systems," in *Audio Engineering Society Convention 110*, May 2001.

[8] A. Novak, L. Simon, F. Kadlec, and P. Lotton, "Nonlinear system identification using exponential swept-sine signal," *IEEE Transactions on Instrumentation and Measurement*, vol. 59, Aug. 2010.

[9] W. Klippel, "Loudspeaker nonlinearities - causes, parameters, symptoms," in *Audio Engineering Society Convention 119*, Oct. 2005.

[10] M. I. Mossi, C. Yemdji, N. W. D. Evans, C. Hergoltz, C. Beaugeant, and P. Degry, "New models for characterizing mobile terminal loudspeaker distortions," in *Proc. IWAENC*, Aug. 2010.

[11] M. Soria-Rodriguez, M. Gabbouj, N. Zacharov, M. S. Hamalainen, and K. Koivuniemi, "Modeling and real-time auralization of electrodynamic loudspeaker nonlinearities," in *Proc. ICASSP*, May 2004.

[12] A. Stenger, L. Trautmann, and R. Rabenstein, "Nonlinear acoustic echo cancellation with 2nd order adaptive volterra filters," in *Proc. ICASSP*, Mar. 1999.

[13] K. Lashkari, "A novel volterra-wiener model for equalization of loudspeaker distortions," in *Proc. ICASSP*, May 2006.

[14] D. Franken, K. Meerkotter, and J. Wassmuth, "Passive parametric modeling of dynamic loudspeakers," *IEEE Transactions on Speech and Audio Processing*, vol. 9, Nov. 2001.

[15] H. K. Jang and K. J. Kim, "Identification of loudspeaker nonlinearities using the NARMAX modeling technique," *The Journal of the Audio Engineering Society*, vol. 42, Feb. 1994.

[16] J. F. Barrett, *Lectures on Nonlinear Systems.* Technical University Eindhoven, 1976.

[17] A. Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *Audio Engineering Society Convention 108*, Feb. 2000.

[18] N. E. Huang *et al.*, "The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, vol. 454, Mar. 1998.

[19] T. Ogunfunmi, *Adaptive Nonlinear System Identification: The Volterra and Wiener Model Approaches.* Springer, 2007.

[20] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions, With Formulas, Graphs, and Mathematical Tables.* Dover Publications, Incorporated, 1974.

[21] W. J. Rugh, *Nonlinear System Theory: The Volterra / Wiener Approach.* The Johns Hopkins University Press, Aug. 1981.

[22] V. Mathews, "Adaptive polynomial filters," *Signal Processing Magazine, IEEE*, vol. 8, July 1991.

[23] N. Wiener, *Nonlinear Problems in Random Theory.* M.I.T. Press, 1958.

[24] T. Burton, R. Goubran, and F. Beaucoup, "Nonlinear system identification using a subband adaptive volterra filter," *IEEE Transactions on Instrumentation and Measurement*, vol. 58, May 2009.

[25] J. Nemeth, I. Kollar, and J. Schoukens, "Identification of volterra kernels using interpolation," *IEEE Transactions on Instrumentation and Measurement*, vol. 51, Aug. 2002.

[26] G. M. Raz and B. Van Veen, "Baseband volterra filters for implementing carrier based nonlinearities," *IEEE Transactions on Signal Processing*, vol. 46, Jan. 1998.

[27] M. Tsujikawa, T. Shiozaki, Y. Kajikawa, and Y. Nomura, "Identification and elimination of second-order nonlinear distortion of loudspeaker systems using volterra filter," in *Proc. ISCAS*, Oct. 2000.

[28] Y. Kajikawa, "Subband parallel cascade volterra filter for linearization of loudspeaker systems," in *Proc. EUSIPCO*, Aug. 2008.

[29] 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. ICASSP*, vol. 5, May 2006.

[30] A. Guerin, G. Faucon, and R. Le Bouquin-Jeannes, "Nonlinear acoustic echo cancellation based on volterra filters," *IEEE Transactions on Speech and Audio Processing*, vol. 11, Nov. 2003.

[31] W. Liu, J. C. Principe, and S. Haykin, *Kernel Adaptive Filtering: A Comprehensive Introduction*. John Wiley, 2010.

[32] R. Raich, H. Qian, and G. T. Zhou, "Digital baseband predistortion of nonlinear power amplifiers using orthogonal polynomials," in *Proc. ICASSP*, Apr. 2003.

[33] D. Lei, *Digital Predistortion of Power Amplifiers for Wireless Applications*. PhD thesis, Georgia Institute of Technology, Mar. 2003.

[34] F. Alonge, F. D'Ippolito, and T. Cangemi, "Identification and robust control of DC/DC converter hammerstein model," *IEEE Transactions on Power Electronics*, vol. 23, Nov. 2008.

[35] P. Brunet, *Nonlinear System Modeling and Identification of Loudspeakers*. PhD thesis, Northeastern University, Apr. 2014.

[36] A. Voishvillo, A. Terekhov, E. Czerwinski, and S. Alexandrov, "Graphing, interpretation, and comparison of results of loudspeaker nonlinear distortion measurements," *The Journal of the Audio Engineering Society*, vol. 52, Apr. 2004.

[37] A. Voishvillo, "Assessment of nonlinearity in transducers and sound systems - from THD to perceptual models," in *Audio Engineering Society Convention 121*, Oct. 2006.

[38] A. Dobrucki, "Nonlinear distortions in electroacoustic devices," *Archives of Acoustics*, vol. 36, Jan. 2011.

[39] D. Konstantinos, R. Arthur, and L. Domine, *Wide-bandwidth high dynamic range D/A converters*. Springer, 2006.

[40] A. Markus, "Nonlinear distortion studies in wide-band analog-to-digital converters," Master's thesis, Tampere University of Technology, Mar. 2009.

[41] K. L. Chan, J. Zhu, and I. Galton, "Dynamic element matching to prevent nonlinear distortion from pulse-shape mismatches in high-resolution DACs," *IEEE Journal of Solid-State Circuits*, vol. 43, Sept. 2008.

[42] C. Hoisington, *Android Boot Camp for Developers Using Java: A Guide to Creating Your First Android Apps*. Cengage Learning, 2014.

[43] A. Dobrucki, B. Merit, V. Lemarquand, and G. Lemarquand, "Modeling of the nonlinear distortion in electrodynamic loudspeakers caused by the voice-coil inductance," *10éme Congrés Français d'Acoustique*, vol. 10, Apr. 2010.

[44] R. Ravaud, G. Lemarquand, V. Lemarquand, and T. Roussel, "Ranking of the nonlinearities of electrodynamic loudspeakers," *Archives of Acoustics*, vol. 35, Feb. 2010.

[45] A. Novák, *Identification of Nonlinear Systems in Acoustics*. PhD thesis, Université du Maine, Apr. 2009.

[46] K. Shi, X. Ma, and G. Zhou, "Acoustic echo cancellation using a pseudocoherence function in the presence of memoryless nonlinearity," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 55, Oct. 2008.

[47] A. Stenger and W. Kellermann, "Adaptation of a memoryless preprocessor for nonlinear acoustic echo cancelling," *Elsevier Signal Processing*, Sept. 2000.

[48] F. Kuech, A. Mitnacht, and W. Kellermann, "Nonlinear acoustic echo cancellation using adaptive orthogonalized power filters," in *Proc. ICASSP*, Mar. 2005.

[49] M. I. Mossi, C. Yemdji, N. W. D. Evans, C. Beaugeant, and P. Degry, "Robust and low-cost cascaded non-linear acoustic echo cancellation," in *Proc. ICASSP*, May 2011.

[50] R. H. Flake, "Volterra series representation of nonlinear systems," *Transactions of the American Institute of Electrical Engineers, Part II: Applications and Industry*, vol. 81, Jan. 1963.

[51] A. J. M. Kaizer, "Modeling of the nonlinear response of an electrodynamic loudspeaker by a volterra series expansion," *J. Audio Eng. Soc*, vol. 35, June 1987.

[52] S. P. Boyd, *Volterra series: Engineering Fundamentals*. PhD thesis, Univeristy of California, Berkeley, May 1985.

[53] M. Zeller, L. Azpicueta-Ruiz, J. Arenas-Garcia, and W. Kellermann, "Adaptive volterra filters with evolutionary quadratic kernels using a combination scheme for memory control," *IEEE Transactions on Signal Processing*, vol. 59, Apr. 2011.

[54] G. B. Stan, J. J. Embrechts, and D. Archambeau, "Comparison of different impulse response measurement techniques," *J. Audio Eng. Soc*, vol. 50, Apr. 2002.

[55] M. Rebillat, R. Hennequin, E. Corteel, and B. F. G. Katz, "Prediction of harmonic distortion generated by electro-dynamic loudspeakers using cascade of hammerstein models," in *Audio Engineering Society Convention 128*, May 2010.

[56] L. K. Gudupudi, C. Beaugeant, N. W. D. Evans, M. I. Mossi, and L. Lepauloux, "A comparison of different loudspeaker models to empirically estimated non-linerities," in *Proc. HSCMA*, May 2014.

[57] L. K. Gudupudi, C. Beaugeant, and N. W. D. Evans, "Characterization and modelling of non-linear loudspeakers," in *Proc. IWAENC*, Sept. 2014.

[58] ITU, "Objective measuring apparatus, head and torso simulator for telephonometry," in *ITU-T P.58: Terminals and Subjective and Objective Assessment Methods*, Aug. 1996.

[59] H. Acoustics, "ACQUA - information material." [Online].

[60] S. T. Neely and J. B. Allen, "Invertibility of a room impulse response," *The Journal of the Acoustical Society of America*, vol. 66, July 1979.

[61] F. Everest and K. Pohlmann, *Master Handbook of Acoustics*. McGraw-Hill Education, 2009.

[62] A. W. Rix, J. G. Beerends, M. P. Hollier, and A. P. Hekstra, "Perceptual evaluation of speech quality (pesq)-a new method for speech quality assessment of telephone networks and codecs," in *Proc. ICASSP*, May 2001.

[63] J. Gil-Cacho, T. Van Waterschoot, M. Moonen, and S. Jensen, "Study and characterization of the odd and even nonlinearities in electrodynamic loudspeakers by periodic random-phase multisines"," in *Audio Engineering Society Convention 127*, Oct. 2009.

[64] L. K. Gudupudi, M. I. Mossi, C. Beaugeant, and N. W. D. Evans, "Comprehensive performance and stability analysis of NAEC algorithms," tech. rep., EURECOM, Sophia Antipolis, France, 2015.

[65] T. Gupta, S. Suppappola, and A. Spanias, "Nonlinear acoustic echo control using an accelerometer," in *Proc. ICASSP*, Apr. 2009.

[66] P. Shah, I. Lewis, S. Grant, and S. Angrignon, "Nonlinear acoustic echo cancellation using feedback," in *Proc. ICASSP*, May. 2013.

[67] P. Shah, I. Lewis, S. Grant, and S. Angrignon, "Nonlinear acoustic echo cancellation using voltage and current feedback," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, Oct. 2015.

[68] W. A. Frank, "An efficient approximation to the quadratic volterra filter and its application in real-time loudspeaker linearization," *Signal Processing*, vol. 45, Jul. 1995.

[69] W. Frank, R. Reger, and U. Appel, "Realtime loudspeaker linearization," in *Proc. NDSP*, Oct. 1993.

[70] M. Mossi, C. Yemdji, N. Evans, and C. Beaugeant, "Non-linear acoustic echo cancellation using online loudspeaker linearization," in *Proc. WASPAA*, Oct. 2011.

[71] S. Gustafsson, R. Martin, and P. Vary, "Combined acoustic echo control and noise reduction for hands-free telephony," *Signal Processing*, vol. 64, Jan. 1998.

[72] O. Hoshuyama and A. Sugiyama, "Nonlinear echo cancellation based on spectral shaping," in *Speech and Audio Processing in Adverse Environments*, pp. 267–283, Springer Berlin Heidelberg, 2008.

[73] F. Kuech and W. Kellermann, "Nonlinear residual echo suppression using a power filter model of the acoustic echo path," in *Proc. ICASSP*, Apr. 2007.

[74] K. Shi, X. Ma, and G. T. Zhou, "A residual echo suppression technique for systems with nonlinear acoustic echo paths," in *Proc. ICASSP*, Apr. 2008.

[75] D. Bendersky, J. Stokes, and H. S. Malvar, "Nonlinear residual acoustic echo suppression for high levels of harmonic distortion," in *Proc. ICASSP*, Apr. 2008.

[76] A. Schwarz, C. Hofmann, and W. Kellermann, "Spectral feature-based nonlinear residual echo suppression," in *Proc. WASPAA*, Oct. 2013.

[77] O. Hoshuyama and A. Sugiyama, "An acoustic echo suppressor based on a frequency-domain model of highly nonlinear residual echo," in *Proc. ICASSP*, May. 2006.

[78] O. Hoshuyama and A. Sugiyama, "Evaluations of an echo suppressor based on a frequency-domain model of highly nonlinear residual echo," in *Proc. IWAENC*, Sept. 2006.

[79] S. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 27, Apr. 1979.

[80] J. G. Proakis and D. G. Manolakis, *Digital Signal Processing (3rd Ed.): Principles, Algorithms, and Applications*. Prentice-Hall, Inc., 1996.

[81] H. Malvar, "A modulated complex lapped transform and its applications to audio processing," in *Proc. ICASSP*, Mar. 1999.

[82] A. Stenger and R. Rabenstein, "Adaptive volterra filters for nonlinear acoustic echo cancellation," in *Proc. NSIP*, 1999.

[83] M. Z. Ikram, "Non-linear acoustic echo cancellation using cascaded Kalman filtering," in *Proc. ICASSP*, May 2014.

[84] F. Kuech and W. Kellermann, "Partitioned block frequency-domain adaptive second-order volterra filter," *IEEE Transactions on Signal Processing*, vol. 53, Feb. 2005.

[85] S. Malik and G. Enzner, "Fourier expansion of hammerstein models for nonlinear acoustic system identification," in *Proc. ICASSP*, May 2011.

[86] A. Fermo, A. Carini, and G. L. Sicuranza, "Simplified volterra filters for acoustic echo cancellation in gsm receivers," in *Proc. EUSIPCO*, Sept. 2000.

[87] D. Mansour and A. H. G. Jr., "Frequency domain non-linear adaptive filter," in *Proc. ICASSP*, Apr. 1981.

[88] L. A. Azpicueta-Ruiz, M. Zeller, J. Arenas-Garcia, and W. Kellermann, "Novel schemes for nonlinear acoustic echo cancellation based on filter combinations," in *Proc. ICASSP*, Apr. 2009.

[89] S. Malik and G. Enzner, "State-space frequency-domain adaptive filtering for nonlinear acoustic echo cancellation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, Sept. 2012.

[90] A. N. Birkett and R. A. Goubran, "Acoustic echo cancellation using nlms-neural network structures," in *Proc. ICASSP*, May. 1995.

[91] A. Birkett and R. Goubran, "Nonlinear echo cancellation using a partial adaptive time delay neural network," in *Proc. WNNSP*, Aug. 1995.

[92] C. Danilo, *Adaptive Algorithms for Intelligent Acoustic Interfaces*. PhD thesis, Sapienza University of Rome, Dec. 2011.

[93] L. K. Gudupudi, N. Chatlani, C. Beaugeant, and N. W. D. Evans, "Non-linear acoustic echo cancellation using empirical mode decomposition," in *Proc. ICASSP*, Apr. 2015.

[94] C. Chui, *An Introduction to Wavelets*. Wavelet analysis and its applications, Academic Press, 1992.

[95] R. Gray and J. Goodman, *Fourier Transforms: An Introduction for Engineers*. The Springer International Series in Engineering and Computer Science, Springer US, 2012.

[96] P. Flandrin, P. Goncalves, and G. Rilling, *Hilbert-Huang Transform and its Applications*, ch. 4. World Scientific, Sept. 2005.

[97] S. Kizhner, T. P. Flatley, N. E. Huang, K. Blank, and E. Conwell, "On the hilbert-huang transform data processing system development," in *Proc. IEEE Aerospace Conference*, vol. 3, Mar. 2004.

[98] G. Rilling, P. Flandrin, and P. Gonçalves, "On empirical mode decomposition and its algorithms," in *Proc. NSIP*, June 2003.

[99] P. Flandrin *et al.*, "Matlab/c codes for EMD and EEMD with examples," 2007. [Online].

[100] A. Zeiler, *Weighted Sliding Empirical Mode Decomposition and its Applications to Neuromonitoring Data*. PhD thesis, Universität Regensburg, 2012.

[101] A. Linderhed, *Adaptive Image Compression with Wavelet Packets and Empirical Mode Decomposition*. PhD thesis, Linköping University, Sweden, 2004.

[102] C. R. Sharpley and V. Vatchev, "Analysis of the intrinsic mode functions," *Constructive Approximation*, vol. 24, 2006.

[103] N. E. Huang, M. C. Wu, S. R. Long, S. S. P. Shen, W. Qu, P. Gloersen, and K. L. Fan, "A confidence limit for the empirical mode decomposition and hilbert spectral analysis," *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 459, 2003.

[104] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: A noise-assisted data analysis method," *Advances in Adaptive Data Analysis*, vol. 01, 2009.

[105] C. Gaochao, Y. Yunchao, T. Tanaka, and C. Jianting, "EEG energy analysis for evaluating consciousness level using dynamic MEMD," in *Proc. IJCNN*, July 2014.

[106] X. Navarro, F. Poree, and G. Carrault, "ECG removal in preterm eeg combining empirical mode decomposition and adaptive filtering," in *Proc. ICASSP*, Mar. 2012.

[107] C. Park, D. Looney, P. Kidmose, M. Ungstrup, and D. P. Mandic, "Time-frequency analysis of EEG asymmetry using bivariate empirical mode decomposition," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 19, Aug. 2011.

[108] N. Rehman, D. Looney, T. M. Rutkowski, and D. P. Mandic, "Bivariate EMD-based image fusion," in *IEEE/SP 15th Workshop on Statistical Signal Processing*, Aug. 2009.

[109] J. C. Nunes and E. Deléchelle, "Empirical mode decomposition: Applications on signal and image processing," *Advances in Adaptive Data Analysis*, 2009.

[110] H. En-Ching, *Hilbert -Huang transform analysis of hydrological and climatic time series*. PhD thesis, Purdue University, 2006.

[111] J. I. Salisbury and M. Wimbush, "Using modern time series analysis techniques to predict enso events from the soi time series," *Nonlinear Processes in Geophysics*, vol. 9, 2002.

[112] N. E. Huang, M. Wu, W. Qu, S. R. Long, and S. S. P. Shen, "Applications of hilbert huang transform to nonstationary financial time series analysis," *Applied Stochastic Models in Business and Industry*, vol. 19, 2003.

[113] P. A. Hwang, N. E. Huang, and D. W. Wang, "A note on analyzing nonlinear and nonstationary ocean wave data," *Applied Ocean Research*, vol. 25, Aug. 2003.

[114] J. Han and B. Mirko, "Empirical mode decomposition for seismic time-frequency analysis," *GEOPHYSICS*, vol. 78, 2013.

[115] L. Da-Chao, L. Zhang, P. Feng, and L. Fan, "Elimination of end effects in empirical mode decomposition by mirror image coupled with support vector regression," *Mechanical Systems and Signal Processing*, vol. 31, 2012.

[116] Z. K. Peng, W. T. Peter, and F. L. Chu, "An improved hilbert-huang transform and its application in vibration signal analysis," *Journal of Sound and Vibration*, vol. 286, Aug. 2005.

[117] P. P. Frank and N. P. Anthony, "HHT-based nonlinear signal processing method for parametric and non-parametric identification of dynamical systems," *International Journal of Mechanical Sciences*, vol. 50, Dec. 2008.

[118] N. Chatlani and J. J. Soraghan, "EMD-based filtering (emdf) of low-frequency noise for speech enhancement," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, May 2012.

[119] N. Chatlani and J. J. Soraghan, "Speech enhancement using adaptive empirical mode decomposition," in *Proc. DSP*, July 2009.

[120] L. Zao, R. Coelho, and P. Flandrin, "Speech enhancement with EMD and hurst-based mode selection," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, May 2014.

[121] A. Bouzid and N. Ellouze, "Empirical mode decomposition of voiced speech signal," in *First International Symposium on Control, Communications and Signal Processing*, 2004.

[122] X. He, R. Goubran, and P. Liu, "A novel sub-band adaptive filtering for acoustic echo cancellation based on empirical mode decomposition algorithm," *International Journal of Speech Technology*, vol. 17, Mar. 2014.

[123] K. A. Lee, W. S. Gan, and S. M. Kuo, *Subband Adaptive Filtering: Theory and Implementation*. Wiley Publishing, 2009.

[124] S. Sandoval and P. L. De Leon, "Theory of the hilbert spectrum," *arXiv preprint arXiv:1504.07554*, 2015.

[125] B. Mijovic, M. De Vos, I. Gligorijevic, J. Taelman, and S. Van Huffel, "Source separation from single-channel recordings by combining empirical-mode decomposition and independent component analysis," *IEEE Transactions on Biomedical Engineering*, vol. 57, Sept. 2010.

[126] M. K. I. Molla, K. Hirose, S. K. Roy, and S. Ahmad, "Adaptive thresholding approach for robust voiced/unvoiced classification," in *Proc. ISCAS*, May 2011.

[127] S. K. Roy and W. P. Zhu, "Pitch estimation of noisy speech using ensemble empirical mode decomposition and dominant harmonic modification," in *Proc. CCECE*, May 2014.

[128] L. K. Gudupudi, N. Chatlani, C. Beaugeant, and N. Evans, "An alternative view of loudspeaker nonlinearities using the Hilbert-huang transform," in *Proc. WASPAA*, Oct. 2015.

[129] L. Cohen, *Time-frequency analysis*. Prentice-Hall, 1995.

[130] F. W. King, *Hilbert Transforms*. Cambridge University Press, 2009.

[131] E. Bedrosian, "A product theorem for hilbert transforms," *Proc. IEEE*, vol. 51, May 1963.

[132] A. H. Nuttall, "On the quadrature approximation to the hilbert transform of modulated signals," *Proc. IEEE*, vol. 54, Oct. 1966.

[133] A. Nayfeh, *Introduction to perturbation techniques*. Wiley, 1981.

[134] D. Looney, A. Hemakom, and D. P. Mandic, "Intrinsic multi-scale analysis: a multi-variate empirical mode decomposition framework," *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 471, no. 2173, 2014.

[135] Y. S. Lee, S. Tsakirtzis, A. F. Vakakis, L. A. Bergman, and D. M. McFarland, "Physics-based foundation for empirical mode decomposition," *AIAA Journal*, vol. 47, 2009.

[136] R. Deering and J. F. Kaiser, "The use of a masking signal to improve empirical mode decomposition," in *Proc. ICASSP*, Mar. 2005.

[137] N. Ur Rehman, C. Park, N. E. Huang, and D. P. Mandic, "Emd via MEMD: Multivariate noise-aided computation of standard EMD," *Advances in Adaptive Data Analysis*, vol. 05, no. 02, 2013.

[138] Z. Wu and N. E. Huang, "A study of the characteristics of white noise using the empirical mode decomposition method," *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 460, 2004.

[139] P. Flandrin, G. Rilling, and P. Goncalves, "Empirical mode decomposition as a filter bank," *IEEE Signal Processing Letters*, vol. 11, Feb. 2004.

[140] S. Sandoval and P. L. De Leon, "http://www.hilbertspectrum.com." Accessed: 2017-04-05.