

Enhancement of Speech Auditory Brainstem Responses using Adaptive Filters

by

Anwar Fallatah

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

Master of Applied Sciences

The Ottawa-Carleton Institute for
Electric and Computer Engineering

School of Electrical Engineering and Computer Science

University of Ottawa

Ottawa, Ontario

2012

© Anwar Fallatah, Ottawa, Canada, 2012

Abstract

Several adaptive filters were investigated to enhance speech auditory brainstem responses (speech ABR). The objective was to shorten the long recording time currently needed by the standard coherent averaging method to obtain acceptable performance, which has limited the clinical adoption of speech ABR. Five algorithms were implemented: Wiener Filter (WF), Steepest Descent (SD), Adaptive Noise Cancellation (ANC) based on Least-Mean-Square error (LMS) and normalized LMS error (nLMS), and a multi-adaptive cascade combination of SD and LMS. The performance of the adaptive filters was assessed on speech ABR data gathered from several subjects and compared with coherent averaging using the overall Signal-to-Noise Ratio (SNR), the local SNR around the fundamental frequency and the first formant, and Mean-Square-Error (MSE) in the time and frequency domains. The adaptive filters could reduce the time needed, by at least one order of magnitude, for obtaining comparable signal quality as that obtained with coherent averaging.

Table of Contents

<i>Abstract</i>	ii
<i>List of Figures</i>	vi
<i>List of Tables</i>	x
<i>List of Acronyms</i>	xi
<i>List of Symbols</i>	xiii
CHAPTER 1: Introduction	1
1.1 Motivation	1
1.2 Thesis Outline	3
CHAPTER 2: Background.....	5
2.1 The Auditory Nervous System.....	5
2.1.1 Ascending auditory pathway.....	5
2.1.2 Descending auditory pathway.....	6
2.2 Speech Auditory Brainstem Responses.....	6
2.3 Time Series Analysis.....	10
2.3.1 Stationarity	10
2.3.2 Correlation function	10
2.4 Optimal Filters.....	11
2.4.1 The Wiener Filter	11
2.4.2 The Maximum SNR Filter	14
2.4.3 The Kalman Filter	14
2.5 Adaptive Filters	16
2.5.1 Method of Least-mean-square (LMS)	16
2.5.2 Method of Least Square (LS).....	17

2.6	ADAPTIVE FILTERS APPLICATIONS.....	17
2.6.1	Adaptive Line Enhancement (ALE)	18
2.6.2	Modelling.....	19
2.6.3	Adaptive Noise Cancellation (ANC)	20
CHAPTER 3: Methodology.....		22
3.1	Experiment setup.....	22
3.1.1	Audiometric Test	22
3.1.2	Construction of the Stimulus	23
3.1.3	ABR Recording.....	24
3.2	Signal Analysis and performance metrics	25
3.2.1	Signal-to-Noise Ratio.....	25
3.2.2	Mean Square Error	28
3.3	Adaptive Filter Design	28
3.3.1	Wiener Filter	29
3.3.2	Steepest-Descent algorithm	31
3.3.3	Adaptive Noise Cancellation	33
3.4	Validation with Noise Signals.....	34
CHAPTER 4: Results		43
4.1	The Wiener Filter and Steepest-Descent Algorithm	46
4.2	Adaptive Noise Cancellation.....	50
4.3	Multi-Adaptive technique	53
4.4	Comparison between the Algorithms.....	55
4.5	Performance of The Algorithms With Frequency Following Responses (FFR)	61
CHAPTER 5: Discussion.....		66
CHAPTER 6: Conclusion And Future Work.....		71

6.1	Conclusion.....	71
6.2	Thesis Contributions	72
6.3	Future Work	72
	Appendix A.....	74
	Appendix B	75
	Appendix C	76
	<i>References</i>	79

List of Figures

Figure 2-1 Schematic diagram of the speech ABR measurement system	7
Figure 2-2 Time domain waveforms of speech ABR of vowel /a/. The EFR is shown at the top and the FFR at the bottom. The transient onset response ends in less than 20 ms, and is then followed by the sustained response.....	9
Figure 2-3 Spectra of speech ABR of vowel /a/. The EFR is shown at the top and the FFR at the bottom. The EFR shows a strong component at F0 of the stimulus and its harmonics, whereas FFR emphasizes harmonics in the region of F1. .	9
Figure 2-4 ALE block diagram. AF refers to Adaptation Algorithm, u is the observation signal, y is the output, w are the coefficients, and z^{-m} is the time delay of m samples.....	18
Figure 2-5 System identification using adaptive filters. AF refers to Adaptation Algorithm, u is the observation signal, y is the output, w and g are the coefficients, and z^{-l} is the time delay.	20
Figure 2-6 Block diagram of Adaptive Noise Cancellation, AF refers to Adaptation Algorithm, u is the observation signal, v is the noise reference signal, w are the coefficients, and z^{-l} is the time delay.	21
Figure 3-1 Amplitude spectrum of the synthetic vowel /a/ stimulus	23
Figure 3-2 The noise interval used for calculating local SNR at F0 (ISNR100)	26
Figure 3-3 The noise interval used for calculating local SNR at F1 (ISNR700)	27
Figure 3-4 The estimated increase of overall SNR of the configured WF with different number of coefficients.	30
Figure 3-5 Convergence of the designed WF and SD wieghts. Three weights of WF (W_o) and SD (W_{sd}) are shown as they converge to the optimul values.	32
Figure 3-6 Effect of WF on random white noise	35
Figure 3-7 Effect WF on AR modeled noise	35
Figure 3-8 Effect of WF on one sample of EEG noise with a strong 60 Hz power-line noise component (EEGn1).....	36
Figure 3-9 Effect of WF on a second sample of EEG noise (EEGn2)	36

Figure 3-10 Block diagram of WF and SD algorithms, with a pre-processing threshold function that is used to control the reference signal switch (r). If speech ABR is detected, the desired signal will pass. Otherwise the noise signal (n) that is generated by LQ factorization will pass. 38

Figure 3-11 Block diagram of ANC algorithm, with a pre-processing threshold function is used to control the reference signal switch (r). If speech ABR is detected, the noise n (from LQ factorization) will pass. Otherwise the signal (s) will pass. 38

Figure 3-12 Illustration of time domain performance of the adaptive filters with the Electroencephalogram noise (EEGn1) that has a strong component at 60Hz, with pre-processing using the autocorrelation-based threshold function..... 41

Figure 3-13 Illustration of frequency domain performance of the adaptive filters with the Electroencephalogram noise (EEGn1) that has a strong component at 60Hz, with pre-processing using the autocorrelation-based threshold function..... 42

Figure 4-1 Illustration of time domain performance of all the adaptive filters with the Envelope Following Response (EFR). The sample response is a single block based on the coherent average of 250 repetitions (block 1 from subject 1) and is shown before and after filtering. The coherent average over twelve blocks and the desired speech ABR reference signal are also shown 44

Figure 4-2 Illustration of frequency domain performance of all the adaptive filters with the same single block Envelope Following Response (EFR) shown in the previous figure. The spectra of the coherent average over twelve blocks and the desired speech ABR reference signal are also shown. 45

Figure 4-3 Results of WF on speech EFR of the first subject. The solid line represents the unfiltered ABR, the circled line represents the results of the WF, the dotted line is WF of the response that contains no components related to the stimulus, and the dashed horizontal line represents the coherent average over the 12 blocks 46

Figure 4-4 Results of WF on speech EFR of the second subject..... 47

Figure4-5 Results of WF on speech EFR of the third subject 47

Figure 4-6 Results of SD on speech EFR of the first subject. The solid line represents the unfiltered ABR, the circled line represents the result of the SD algorithm, the dotted line represents the output of SD with a response that contains no components related to the stimulus, and the dashed horizontal line represents the coherent average over the 12 blocks 48

Figure 4-7 Results of SD on speech EFR of the second subject..... 49

Figure 4-8 Results of SD on speech EFR of the third subject 49

Figure 4-9 Results of ANC with LMS on speech EFR of the first subject. The solid line represents the unfiltered ABR, the circled line represents the result of the ANC with LMS algorithm, the dotted line represents the output of the LMS algorithm with a response that contains no components related to the stimulus, and the dashed horizontal line represents the coherent average over the 12 blocks 50

Figure 4-10 Results of ANC with LMS on speech EFR of the second subject 50

Figure 4-11 Results of ANC with LMS on speech EFR of the third subject..... 51

Figure 4-12 Results of ANC with nLMS on speech EFR of the first subject. The solid line represents the unfiltered ABR, the circled line represents the result of the ANC with nLMS algorithm, the dotted line represents the output of the nLMS algorithm with a response that contains no components related to the stimulus, and the dashed horizontal line represents the coherent average over the 12 blocks 51

Figure 4-13 Results of ANC with nLMS on speech EFR of the second subject. 52

Figure 4-14 Results of ANC with nLMS on speech EFR of the third subject..... 52

Figure 4-15 Results of SD+LMS on speech EFR of the first subject. The solid line represents the unfiltered ABR, the circled line represents the result of the SD + LMS, the dotted line represents the output of the SD + LMS algorithm with a response that contains no components related to the stimulus, and the dashed horizontal line represents the coherent average over the 12 blocks... 53

Figure 4-16 Results of SD+LMS on speech EFR of the second subject. 54

Figure 4-17 Results of SD+LMS on speech EFR of the third subject..... 54

Figure 4-18 The oSNR performance of all algorithms on the speech EFR in all blocks from 3 subjects. 55

Figure 4-19 Local SNR at 100 Hz (lSNR100) performance of all algorithms on the speech EFR in all blocks from 3 subjects.. 56

Figure 4-20 Time domain MSEq performance of all algorithms on the speech EFR in all blocks from 3 subjects. The bottom graph is a magnification of the bottom part of the top graph. 57

Figure 4-21 Time domain MSEd of all algorithms on the speech EFR in all blocks from 3 subjects. The bottom graph is a magnification of the bottom part of the top graph..... 58

Figure 4-22 Frequency domain MSEq (fMSEq) performance of all algorithms on the speech EFR in all blocks from 3 subjects. The bottom graph is a magnification of the bottom part of the top graph. 58

Figure 4-23 Frequency domain MSEd (fMSEd) performance of all algorithms on the speech EFR in all blocks from 3 subjects. The bottom graph is a magnification of the bottom part of the top graph. 59

Figure 4-24 Illustration of time domain performance of all the adaptive filters with the Frequency Following Response (FFR). The sample response based on the coherent average of 250 repetitions (block 1 from subject 1) is shown before and after filtering. The coherent average over twelve blocks is also shown. 62

Figure 4-25 Illustration of frequency domain performance of all the adaptive filters with the Frequency Following Response (FFR). The sample response based on the coherent average of 250 repetitions (block 1 from subject 1) is shown before and after filtering. The coherent average over twelve blocks is also shown. 63

Figure 4-26 The overall SNR (oSNR) performance of all algorithms on the speech FFR in all blocks from 3 subjects..... 64

Figure 4-27 The local SNR at 700 Hz (lSNR700) performance of all algorithms on the speech FFR in all blocks from 3 subjects 65

List of Tables

Table 3-1 Summary of the analysis performed on the 250-repetition speech ABR	29
Table 3-2 Performance of WF with noise signals.....	37
Table 3-3 oSNR performance of all algorithms with no-ABR where the it is averaged over the 36 responses that contains no components related to the stimulus, after removal of the components using LQ factorization, and with the four noise samples (AR, White Noise, EEGn1, and EEGn2). Pre-processing is performed first by passing the signal through the autocorrelation-based threshold function as described in the text.	40
Table 4-1 Performance of all algorithms with the EFR averaged over 36 individual blocks from the 3 subjects	59
Table 4-2 Performance of all algorithms averaged over 12 individual blocks from subject one	60
Table 4-3 Performance of all algorithms averaged over 12 individual blocks from subject two	60
Table 4-4 Performance of all algorithms averaged over 12 individual blocks from subject three	61
Table 4-5 Performance of all algorithms with the FFR averaged over 36 blocks from the 3 subjects	65
Table 5-1 Theoretical and measured performances with the EFR signals.....	68
Table 5-2 The theoretical and measured performances with the FFR signals	69

List of Acronyms

ABR	Auditory Brainstem Response
ALE	Adaptive Line Enhancement
ANC	Adaptive Noise cancellation
ANOVA	Analysis of Variance
ASSR	Auditory Steady-State Response
CN	Cochlear Nucleus
dB	Decibels
EEG	Electroencephalogram
EFR	Envelope Frequency Response
EP	Evoked Potential
F0	Fundamental Frequency
F1	First Formant
F2	Second Formant
FFR	Frequency Following Response
FIR	Finite Impulse Response
ICC	Inferior Colliculus
LL	Lateral Lemniscus
LMS	Least-mean-square
LS	Least Square method
MGB	Medial Geniculate Body
MSE	Mean Square Error
MSEd	Time domain MSE between observation and the desired signal
MSEq	Time domain MSE between observation and a reference signal generated by LQ factorization
fMSE	Frequency domain MSE
nLMS	Normalized Least-mean-square

oSNR	Overall SNR
lSNR	Local SNR
RLS	Recursive Least Square
SD	Steepest-decent method
SNR	Signal-to-Noise Ratio
SpEPs	Speech Evoked Potentials
VA	VA onset complex (transient response)
WF	Wiener Filter

List of Symbols

\cdot^T	Matrix transposition
\cdot^H	Matrix Hermitian
\cdot^*	Complex conjugate
$\hat{\cdot}$	Estimated value
$\bar{\cdot}$	Mean value
u	Vector variables in boldfaced letter
$E(x)$	Statistical expected operation
X	Matrix variables in capital letter
R	Autocorrelation matrix
p	Cross-correlation matrix
σ_x^2	Variance of vector x
μ	Step-size parameter

CHAPTER 1: Introduction

1.1 MOTIVATION

The Auditory Brainstem Response (ABR) is a reliable signal that can be used to objectively evaluate the function of the auditory system. In current clinical practice, the ABR is recorded when the auditory system is stimulated with an artificial sound such as a click, tone burst, or an amplitude modulated tone (Boston and Moller, 1985). More recently, the ABR generated by a speech stimulus (speech ABR) has been investigated because speech is of primary importance in human acoustic communication. In particular, speech ABR has been proposed as a marker of defects in central auditory processing in children with learning problems, and has also been proposed to study degradation in auditory processing in the aging auditory system (Johnson et al., 2005; Vander Werff and Burns, 2011). Since ABR is recorded non-invasively using electrodes placed on the surface of the scalp, the signal-to-noise (SNR) ratio of the ABR signal is generally very low. This low SNR has limited the application of speech ABR in the clinic because the conventional approach of coherently averaging the responses over multiple presentations of the relatively long duration speech stimulus requires an exceedingly long recording time that ranges from several minutes to tens of minutes with a single speech sample (Dajani et al., 2005; Laroche et al., in press; Prevost et al., in press).

Different approaches have been utilized to increase the SNR of the ABR. Some approaches are associated with using a certain type of stimulus, while others focus on implementing certain methodologies for obtaining ABR signals. Studies have shown that using a stimulus that has special features increases the chance of obtaining evoked responses with acceptable SNR. For instance, using synthetic vowel stimuli that have specified features is more effective than using natural vowel stimuli since the spectral content of the synthetic vowel is fixed and concentrated at harmonic frequencies of the fundamental. Another advantage of using synthetic vowels is that the fundamental and formant frequencies can be controlled and chosen as required (Krishnan, 2002). Other studies have shown that higher SNR can also be obtained if a subject, the listener, has

previous experience with the stimuli. For example, a number of experiments show that auditory-evoked responses to some complex acoustic stimuli are more robustly represented in musicians than in non-musicians or untrained listeners (Pitt, 1994; Pantev et al., 2001; Bidelman and Krishnan, 2010).

The methodology that is used to obtain the ABR also influences the measurement SNR. Methods can be divided into transient and steady-state techniques. This classification is made based on the periodicity of the stimulus. The transient auditory evoked potential is produced if the stimulus is a short tone burst or a click. On the other hand, the steady-state auditory evoked potential is produced if a stimulus is applied repeatedly with a steady periodicity, as for example in the so-called Auditory Steady-state Response (ASSR). Since ASSR is generally obtained over a longer period of time than transient evoked responses, it is believed to have a better SNR as previous studies have shown (Fridman et al., 1982; Woodworth et al., 1983; Jacquin et al., 2005).

All the methods that have been used for improving the quality of the ASSR can be classified into single and multiple channel methods. Most single channel methods are based on averaging or adaptive algorithms. Coherent averaging is the most popular method that has been used to increase the SNR of sensory evoked responses (Regan 1989). There are other related averaging methods that have been used, including weighted averaging in the frequency domain (John et al., 2001) and phase-locked averaging, based on the phase of a response in the frequency domain (Picton et al., 2001). Adaptive methods that have been used with ABR include nonlinear estimation (McNamara and Ziarani, 2004) and adaptive sinusoidal estimator (Cheah and Hou, 2010). Finally, multiple channel methods include independent component analysis and Wiener filter methods (Van Dun et al., 2007).

With speech ABR in particular, coherent averaging is the only method that has been used to improve response SNR. There is therefore a need to investigate other methods that can be applied to improve the SNR of the recorded speech ABR and so potentially reduce the required recording time. In this thesis, the use of adaptive filtering methods is investigated and is shown to perform significantly better than conventional coherent averaging. The success of these new approaches to the enhancement of speech ABR will

potentially facilitate the adoption of this technique in the clinical assessment of central auditory processing disorders and in monitoring the progress of auditory rehabilitation.

1.2 THESIS OUTLINE

This thesis is divided into six chapters including this chapter, which describes the motivation for the work, contains an up-to-date overview, and the thesis outline. The following briefly describes the content of the next five chapters:

Chapter 2: This chapter provides background information about several important topics relevant to the thesis. It contains six sections that deal with the auditory nervous system, auditory brainstem responses, time series analysis, optimal filters, adaptive filters, and adaptive filter applications.

Chapter 3: This chapter describes the research methodology used in this thesis. Discussion includes details of procedures used in the experimental collection of data and of the adaptive filters used to enhance the collected signals. In addition, a description of some signal quality and validity measures is provided.

Chapter 4: This chapter presents the results obtained in this work, including a comparison between different methods applied to the enhancement of speech ABR, and in particular the Envelope Following Response (EFR). These methods are the Wiener filter (WF), Steepest-Descent algorithm (SD), adaptive noise cancellation (ANC) with both Least-mean-square (LMS) and normalized LMS (nLMS) approaches, a multi-adaptive approach that combines two algorithms in cascade, and coherent averaging. Enhancement of noisier speech ABRs that are obtained using a different speech ABR configuration, the so-called Frequency Following Response (FFR), is also investigated to determine the limitations of these methods with more challenging signals.

Chapter 5: This chapter provides an in-depth discussion of the obtained results and the consequences for speech ABR measurement.

Chapter 6: This chapter concludes the thesis with a summary of the major findings and contributions of the work, and proposes some avenues for future work.

CHAPTER 2: BACKGROUND

2.1 THE AUDITORY NERVOUS SYSTEM

The auditory nervous system refers to neural systems that start at the inner ear and extend through several stages up to the auditory cortex of the brain. It is considered the most complex neural system when compared to the other sensory systems (Moller, 2006). The auditory nervous system is divided into two main pathways that process signals in opposite directions, namely, the ascending and descending pathways.

2.1.1 Ascending auditory pathway

The ascending auditory pathway is the sensory pathway that is responsible for signal transmission from receptors in the inner ear, also called the cochlea, to nuclei in the auditory cortex. The ascending pathway is classified into two subdivisions: the classical and non-classical ascending pathways (Moller, 2006).

The classical ascending pathway processes details of the auditory information in a hierarchical and parallel order. Prior to the auditory cortex, the pathway includes three main relays: the cochlear nucleus (CN), the inferior colliculus nucleus (ICC) in the upper brainstem, and the medial geniculate body (MGB), which connects to the primary auditory cortex. The acoustic energy is transformed into neural electrical pulses in the cochlea of both ears, which are then transmitted to the CN. After initial processing in the CN, the signals are transmitted to the ICC in the upper brainstem mainly via a group of fibres named the lateral lemniscus (LL). In fact, there is more than one connection that connects the cochlea, CN, ICC, MGB, and the primary cortex with each other and with other parts of the system that originates in the other ear.

The non-classical ascending pathway connects to other sensory systems, and terminates at the secondary auditory cortex or in other sensory cortices. The

function of the non-classical pathway is not well understood, but is thought to have lower perceptual acuity than the classical pathway.

2.1.2 Descending auditory pathway

The descending auditory pathway is responsible for information transmission in the opposite direction from that of the ascending pathway. It starts from the cortex and connects the different stages all the way to the receptors in the cochlea. It is believed that most stages of the ascending pathway have corresponding descending reciprocal connections. A number of studies have shown that descending pathways can modulate the auditory signal processing at each level of the auditory system, starting from the cochlea. For instance, there is evidence that signals from the auditory cortex influence frequency tuning in the ICC and MGB which may in turn enhance the evoked response to a stimulus buried in background noise (Zhang et al., 1997).

2.2 SPEECH AUDITORY BRAINSTEM RESPONSES

Auditory brainstem responses (ABR) refer to the auditory evoked potentials that originate in the brainstem in response to an acoustic stimulus. In clinical practice, ABR is measured by an electronic instrument that uses electrodes placed on the surface of the subject's scalp at specific locations. Figure 2-1 shows a block diagram of the ABR detection device that is used in this work. The three electrodes refer to the positive terminal, negative terminal, and the ground. The device also presents the stimulus to both ears through the earphones and can control the time synchronization between the presented stimulus and the recorded ABR signals.

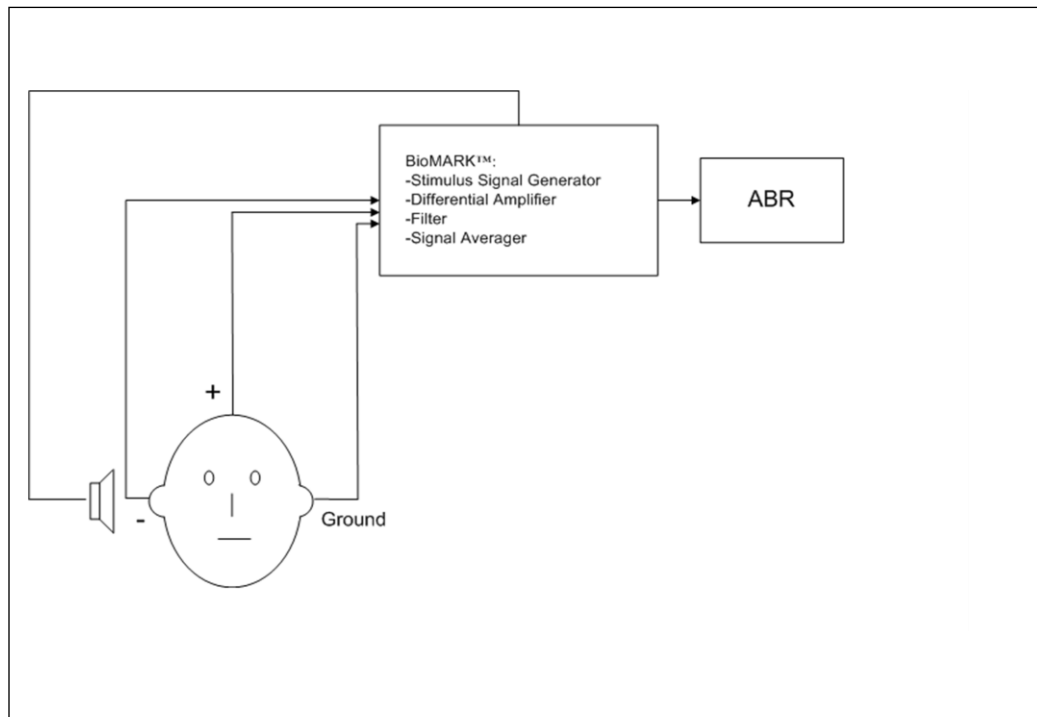


Figure 2-1 Schematic diagram of the speech ABR measurement system (adapted from Laroche, 2010)

In clinical applications, ABR is generated using pure tones, clicks, and other artificial sound stimuli to objectively assess the hearing function. There is evidence that ABR using a speech stimulus could extend these audiological tests to include:

1. Assessing the condition of central auditory neural pathways in children with language-based learning problems (Russo et al., 2004; Greenberg and Ainsworth, 2004; Skoe and Kraus, 2010).
2. Evaluating speech intelligibility with hearing aid usage (Aiken and Picton, 2006).
3. Evaluating the encoding of speech in the intact and impaired auditory system (Johnson et al., 2005; Johnson et al., 2008).

Most speech stimuli that have been used have been synthetic pure vowels or synthetic consonant-vowel syllables (Dajani et al., 2005; Skoe and Kraus, 2010; Prevost et al., in press). Speech ABRs are thought to originate from multiple sources in the brainstem, and

mainly from sources in the upper brainstem. They reflect the compound electrical activity of many neurons that is phase-locked to the different components of the stimulus.

Depending on how the speech ABR is analysed, two types of responses can be obtained, namely the Envelope Following Response (EFR) and the Frequency Following Response (FFR). The EFR represents neural activity that is phase-locked to the envelope of the speech stimulus, which is modulated at the fundamental frequency F0. Therefore the EFR occurs at F0 and some its harmonics, since the envelope is not sinusoidal. In practice, it is obtained using the so-called ‘+-’ configuration where responses to the stimulus presented in one polarity are averaged with an equal number of responses to the stimulus presented in inverted polarity (Aiken and Picton, 2008). On the other hand, the FFR represents neural activity that directly follows the harmonics of the stimulus, and in particular the strongest harmonics near the first formant F1 and possibly the second formant F2. As a result, the FFR occurs at the harmonics near F1 (and possibly F2). In practice, it is obtained using the ‘--’ configuration, where the responses to the stimulus in one polarity are averaged with the negative of the responses to the stimulus in inverted polarity. Figure 2-2 and 2-3 show the time waveform and frequency domain spectrum of the EFR and FFR that correspond to vowel /a/. The EFR signal is obtained by averaging coherently 36,000 repetitions over twelve different subjects while the FFR signal is obtained by averaging 9000 repetitions from three subjects only. The first figure also shows the transient onset response and the sustained response. The frequency domain spectrum clearly shows the F0 component of the stimulus (which is at 100Hz) and its harmonics in the EFR (Fig. 2-2) and the harmonics in the region of F1 in the FFR (Fig. 2-3).

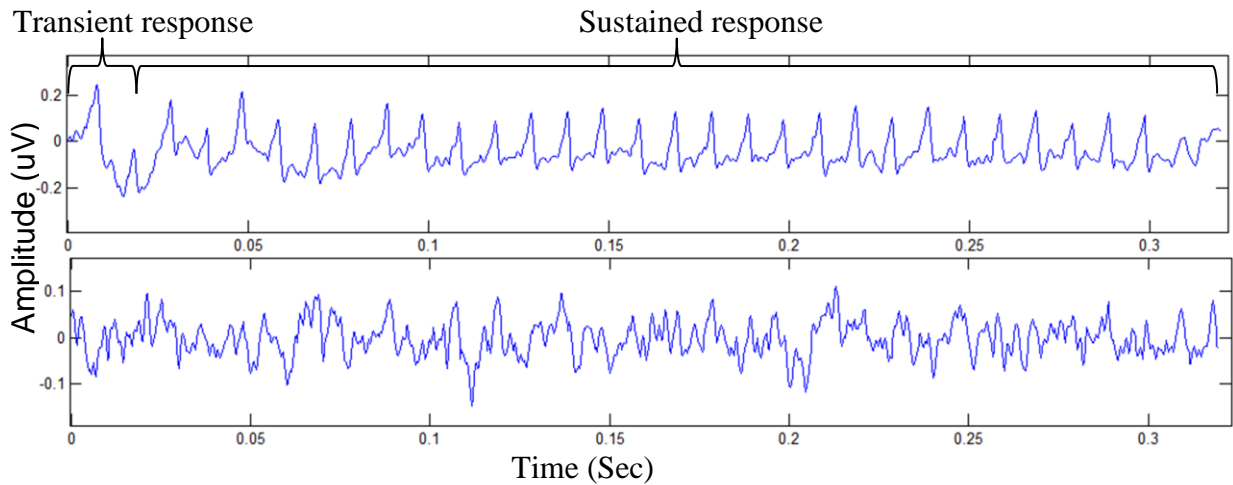


Figure 2-2 Time domain waveforms of speech ABR of vowel /a/. The EFR is shown at the top and the FFR at the bottom. The transient onset response ends in less than 20 ms, and is then followed by the sustained response.

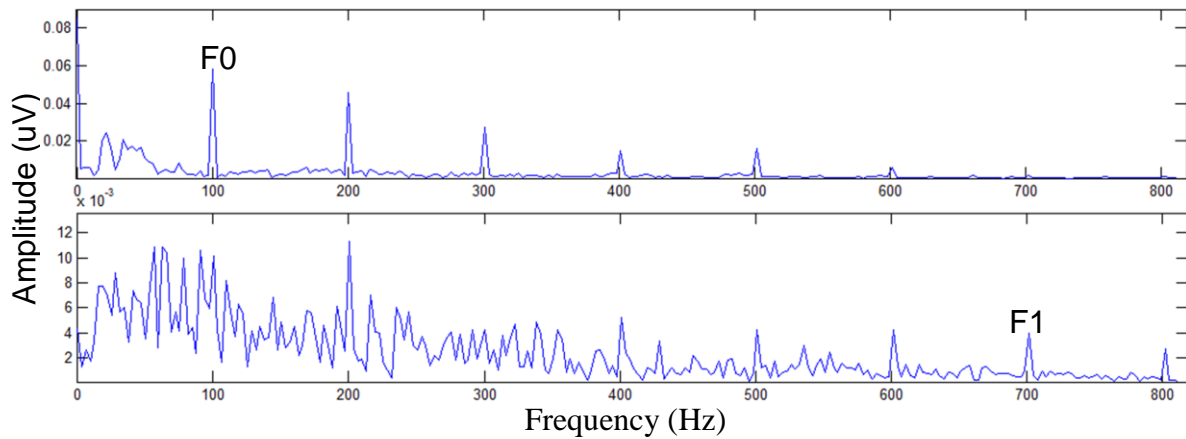


Figure 2-3 Spectra of speech ABR of vowel /a/. The EFR is shown at the top and the FFR at the bottom. The EFR shows a strong component at F0 of the stimulus and its harmonics, whereas FFR emphasizes harmonics in the region of F1.

2.3 TIME SERIES ANALYSIS

Since the obtained speech ABR is a time series, some important issues about time series analysis have to be addressed. Stationarity and correlation function are considered two of the most important topics when we deal with time series analysis using adaptive filters.

2.3.1 Stationarity

One of the most important signal characteristics in time series analysis is stationarity. A stationary process is that whose statistical parameters (such as mean and variance) do not vary with time. There are several types of stationarity based on the statistical moments of the process that do not change with time. The four most common types of stationary processes are first-order, second-order, wide-sense, and strictly stationary. The first-order stationary process is that which has invariant mean. The second-order is that which has invariant correlation function. The wide-sense stationary process is that which satisfies conditions of first and second-order stationary process. The strictly stationary process is that which all its statistical moments are time-invariant (Lessard, 2006).

The coherent averaging method depends on stationarity of the averaged signal. Its performance is ideal if the averaged signals are strongly stationary, and it is worse if there is no stationarity. This also applies to filters that do not adapt such as Wiener Filter (WF). In contrast, adaptive filters have the ability to track statistical change in the signal and so counter the effect of the nonstationarity (Sayed, 2011).

2.3.2 Correlation function

The correlation function is used to determine the relation between two sets of data (cross-correlation) or how a value of data is related to the other values in the same time series (autocorrelation function). Cross-correlation is a statistical tool that is

used to show how two time series are related, or if there is a linear relation between the two sets. On the other hand, autocorrelation highlights patterns, such as periodicities, in one time series (Lessard, 2006).

2.4 OPTIMAL FILTERS

In speech signal enhancement, it is common to use optimal filters to estimate the noise added to the desired signal and then suppress it (Benesty and Chen, 2011). These filters can be useful to enhance speech ABR signals since, as discussed above, these signals are “speech-like” with components at the fundamental frequency and its harmonics. In the following, three well-known linear optimal filters are introduced, namely the Wiener, Maximum SNR and Kalman filters. The Maximum SNR and Kalman Filter are only discussed in this section to provide wider background information about adaptive filters, but they have not been used in this work.

2.4.1 The Wiener Filter

The Wiener Filter is a digital filter that has coefficients that produce minimum mean-square-error (MSE) between the input and the desired signal (Benesty and Chen, 2011). It requires a stationary signal and prior knowledge of its autocorrelation and cross-correlation. Consider the following equation:

$$u(i) = s(i) + v(i) \tag{2.4.1}$$

where $u(i)$ is the observation signal that contains a signal $s(i)$ and an unwanted additive noise $v(i)$. The desired signal is estimated by:

$$\hat{\mathbf{d}} = \mathbf{w}_o^T \mathbf{u} \tag{2.4.2}$$

where \mathbf{w}_o is the optimal impulse response of the Finite Impulse Response (FIR) digital filter that minimizes the mean-square-error (MSE), and it is described as follows:

$$\mathbf{w}_o = \sigma_u^2 R^{-1} \mathbf{p} \quad 2.4.3$$

where σ_u^2 the variance of \mathbf{u} , R is the autocorrelation matrix of \mathbf{u} , and \mathbf{p} is the cross-correlation between \mathbf{u} and \mathbf{d} , the desired signal.

Before we check if our filter produces a false positive when the speech ABR is absent, we have to analyze the input signal to see if the speech ABR can be detected. Let us decompose the observation signal $u(i)$ into three terms as follows:

$$u(i) = p d(i) + d_i(i) + v(i) \quad 2.4.4$$

where the first term, $p d(i)$, is related to a proportional component of the desired signal, the second term, $d_i(i)$, is an interference accumulated to the desired signal, and the last term is the noise signal $v(i)$. By substituting equation (2.4.4) into equation (2.4.2):

$$\hat{d}(i) = w_o (p d(i) + d_i(i) + v(i)) \quad 2.4.5$$

This equation clearly shows the Wiener filter's output which contains the three terms: the filtered desired signal $d(i)$, the remaining residual of the interference, $d_i(i)$, and the noise $v(i)$ (note that $\hat{d}(i) = d(i)$ only in the ideal case where there is no noise).

Next, we will assume the same configuration for the Wiener filter (fixed w_o , $u(i)$, and $v(i)$), except for the absence of desired signal $d(i)$. In other words, let us assume that $d(i)$ equals zero (which leads to $\sigma_{pd}^2 = 0$) and substitute its value in equation (2.4.5). The new output will be:

$$y(i) = w_o (d_i(i) + v(i)) \quad 2.4.6$$

To contrast equation (2.4.5) and equation (2.4.6), the following definition of SNR that is based on variance of the signal is used. Let us define SNR of the input/observation signal by (Benesty and Chen, 2011):

$$iSNR \triangleq \frac{\sigma_s^2}{\sigma_v^2} \quad 2.4.7$$

And the output SNR of equation (2.4.5) and equation (2.4.6) will by:

$$oSNR_{\hat{a}} = \frac{\sigma_{pd}^2}{\sigma_{d_i}^2 + \sigma_v^2} \quad 2.4.8$$

$$oSNR_y = \frac{0}{\sigma_{d_i}^2 + \sigma_v^2} \quad 2.4.9$$

where

$$\sigma_{pd}^2 = var(w_o^T p d(i)) \quad 2.4.10$$

is the variance of the filtered desired signal, and

$$\sigma_{d_i}^2 = var(w_o^T d_i(i)) \quad 2.4.11$$

is the variance of the residual interference, and

$$\sigma_v^2 = var(w_o^T v(i)) \quad 2.4.12$$

is the variance of the residual noise.

Based on equation (2.4.8) and equation (2.4.9), we can draw the following two conclusions:

1. The Wiener Filter produces higher SNR if the input signal contains the desired signal (in our case the speech ABR).
2. Theoretically, in the ideal case, the output SNR is exactly equal to zero if the input signal does not contain the desired signal.

This is the behaviour that would be expected for a filter when introducing an input that does not contains a target signal (in our case, speech ABR) in background noise.

2.4.2 The Maximum SNR Filter

The Maximum SNR filter is an FIR digital filter that uses Rayleigh-Ritz characterization of eigenvalues (Sayed, 2011) to maximize the SNR of the input signal. It is often used as a signal enhancer and its weights are defined as follows:

$$\mathbf{w}_o = \zeta R_{iv}^{-1} \mathbf{p} \quad 2.4.13$$

where

$$R_{iv} = R_i + R_v \quad 2.4.14$$

is the sum of the autocorrelation of the interference and the noise, and ζ is a non-zero scalar. The Maximum SNR filter requires a prior knowledge of the correlation of the noise and interference.

2.4.3 The Kalman Filter

The Kalman filter is considered a powerful tool that can be used with both stationary and nonstationary data (Sayed, 2003). It can be derived based on numerous recursive algorithms; however, here we only discuss the least mean square (LMS) algorithm.

Consider the following equation:

$$\mathbf{d} = K \mathbf{u} \quad 2.4.15$$

where \mathbf{d} is a desired signal, and \mathbf{u} is the observation signal that is multiplied by a matrix K . The desired signal \mathbf{d} can be estimated by:

$$\hat{\mathbf{d}} = \mathbf{p}_{x,e} R_e^{-1} \mathbf{e} \quad 2.4.16$$

where \mathbf{e} is linearly generated from \mathbf{u} , such that \mathbf{e} and \mathbf{u} are uncorrelated. By using Gram-Schmidt linear algebra, \mathbf{e} is calculated as:

$$\mathbf{e} = A \mathbf{u} \quad 2.4.17$$

where A is a lower triangular matrix. As an illustration, for a (3x3) matrix containing the first three moments u_0 , u_1 , and u_2 , it is (Sayed, 2003):

$$A = \begin{bmatrix} I & & \\ -E(u_1 u_0^*) E(u_0 u_0^*)^{-1} & I & \\ X & X & I \end{bmatrix} \quad 2.4.18$$

and,

$$[X \ X] = E(u_2 [u_0^* u_1^*]) E \left(\begin{bmatrix} u_0 \\ u_1 \end{bmatrix} \begin{bmatrix} u_0 \\ u_1 \end{bmatrix}^* \right)^{-1} \quad 2.4.19$$

Now, let us decompose the vector e into three terms that refer to the desired signal e_d , the interference e_i , and the noise e_n . Rewriting equation (2.4.17) using these terms gives:

$$\hat{d} = p_{x,e} R_e^{-1} (e_d + e_i + e_n) \quad 2.4.20$$

The output SNR is defined by:

$$oSNR_{\hat{d}} = \frac{\sigma_{e_d}^2}{\sigma_{e_i}^2 + \sigma_{e_n}^2} \quad 2.4.21$$

To check that our filter does not produce a false positive when speech ABR is absent, assuming fixed correlation matrices (R , and p in equation (2.4.20), and e_d equal zero (which leads to $\sigma_{e_d}^2 = 0$), the output signal in that case will be:

$$y = p_{x,e} R_e^{-1} (0 + e_i + e_n) \quad 2.4.22$$

and its SNR will be:

$$oSNR_y = \frac{0}{\sigma_{e_i}^2 + \sigma_{e_n}^2} \quad 2.4.23$$

According to equation (2.4.23), we can draw the following two conclusions:

1. The Kalman filter produces higher SNR if the input signal contains the desired signal (the speech ABR).
2. Theoretically, in the ideal case, the output SNR of the Kalman filter is exactly zero if the input signal does not contain the desired signal.

This is the behaviour that would be expected for a filter when introducing an input that does not contains a target signal (in our case, speech ABR) in background noise.

2.5 ADAPTIVE FILTERS

Adaptive filters play a significant role in many applications where there is a need to track the changes in the statistical parameters of signals. They have the ability to compare the output signal to the desired signal and update the coefficients accordingly, which is something that static filters are unable to do. Adaptive filters are generally classified into two types based on the method of comparing the filter performance to the desired signal. These two types, namely Least-mean-square (LMS) and the least square (LS), are introduced briefly in this section, and then three of their most common applications are discussed (Farhang-Boroujeny, 1998).

2.5.1 Method of Least-mean-square (LMS)

The least-mean-square adaptive filters tend to minimize the MSE between the output and the desired signal. Their equations relate the desired signal d , the output signal y , the error e , and the weight updates:

$$y(i) = \mathbf{w}^T \mathbf{u} \quad 2.5.1$$

$$e(i) = d(i) - y(i) \quad 2.5.2$$

$$\mathbf{w}(i) = \mathbf{w}(i-1) + \mu u(i) e(i) \quad 2.5.3$$

where $\mathbf{w}(i-1)$ is the weight of the previous iteration, and μ is the step-size parameter, whose maximum value to ensure stability is calculated by:

$$\mu = \frac{2}{\lambda_{max}} \quad 2.5.4$$

where λ_{max} is the maximum eigenvalue of the autocorrelation function (R). It is recommended to have step-size that is below this maximum value to avoid instability.

2.5.2 Method of Least Square (LS)

The second type of adaptive filters is based on the least square method. These filters differ from LMS filters in that they track the deterministic changes of the output signal. Their equations are described as follows:

$$y(i) = \mathbf{w}^H(i-1) \mathbf{u} \quad 2.5.5$$

$$e(i) = d(i) - y(i) \quad 2.5.6$$

$$\mathbf{w}(i) = \mathbf{w}(i-1) + K(i) \mathbf{u}(i) e^*(i) \quad 2.5.7$$

where:

$$K = \lambda^{-1} \left[P_{i-1} - \frac{\lambda^{-1} P_{i-1} \mathbf{u}^H \mathbf{u} P_{i-1}}{1 + \lambda^{-1} \mathbf{u}^H P_{i-1} \mathbf{u}} \right] \quad 2.5.8$$

$$P = \Phi^{-1} \quad 2.5.9$$

$$\Phi = \lambda \Phi_{i-1} + \mathbf{u}^H \mathbf{u} \quad 2.5.10$$

where K and P are referred to as gain parameters, and the d , y , e , and w parameters are as described above.

2.6 ADAPTIVE FILTERS APPLICATIONS

Since an adaptive filter has the ability to adjust itself to approach the desired signal, a number of applications have been developed to apply adaptive filters to track either the statistical or deterministic parameters of signals. In this section, three of the most well-known adaptive filter applications are discussed; Adaptive Line Enhancement (ALE), Modeling, and Adaptive Noise Cancellation (ANC), which is the only filter of these three that has been used in this work. In addition, a discussion about how each application can serve to enhance the speech ABR is included.

2.6.1 Adaptive Line Enhancement (ALE)

One of the basic applications of adaptive filters is adaptive line enhancement (ALE). This application is suitable to extract a narrow-band signal buried in wide-band noise. ALE has two advantages over a fixed filter (non-adaptive filter) in that it can detect a signal of unknown frequency, and the time-varying amplitude of the signal does not affect the filter performance (Treichler, 1979).

Figure 2-4 shows the block diagram of the FIR-based ALE. The idea behind this filter is to increase the autocorrelation of the desired signal by introducing delays z^{-m} . The adaptive filter will minimize the difference between the observation signal u and a delayed version of u , which corresponds to a period of the desired frequency. Since there will be no correlation for the noise band, it will be suppressed, and that is why z^{-m} sometimes is called decorrelation delay. ALE also can be seen as a predictor that predicts samples at delay of m .

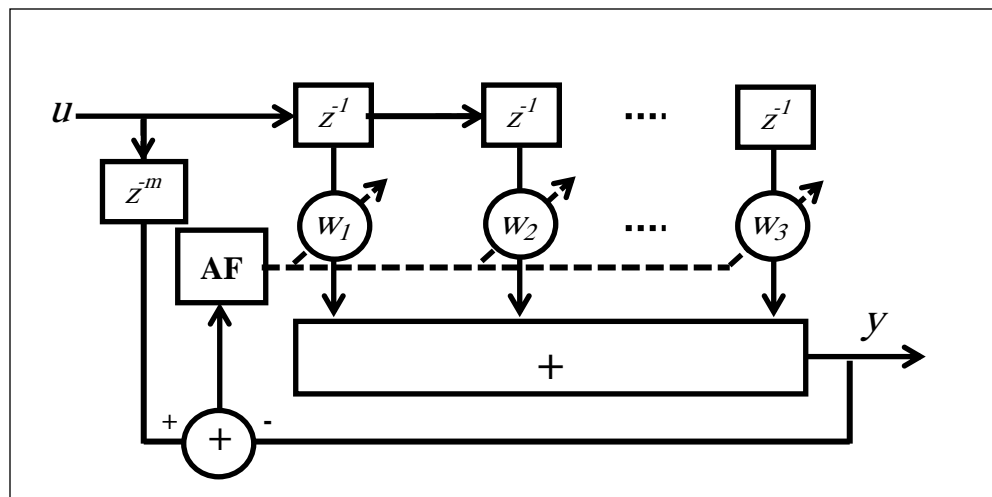


Figure 2-4 ALE block diagram. AF refers to Adaptation Algorithm, u is the observation signal, y is the output, w are the coefficients, and z^{-m} is the time delay of m samples.

ALE can also detect a certain frequency component in the observation signal that needs to be suppressed. For instance, the 60 Hz power line interference that sometimes contaminates the speech ABR signal can be detected and suppressed. This is done by subtracting the target frequency (at the filter's output) from the input signal.

An early study used a modified ALE (MALE) to extract the ABR to a click stimulus (Madhavan, 1989). In that study, the parameters of noise were calculated in a pre-stimulus interval, and then they were used to suppress the noise in the post-stimulus interval. Finally, the distorted output signals were averaged coherently over an optimal minimal number of repetitions. This method was successfully able to produce a performance (using only 40 repetitions) that is comparable to 2000 repetitions of coherent averaging (Madhavan, 1992).

2.6.2 Modelling

Modelling an unknown process or a plant is usually done to estimate the parameters of the process or the behaviour of the plant. A basic block diagram of adaptive filter modelling (also called system identification) is illustrated in Figure 2-5. In this figure, the unknown weights of \mathbf{g} are identified, and \mathbf{w} will approach \mathbf{g} in order to minimize MSE.

One of advantages of AF modelling is that it can estimate the unknown weights even in the presence of additive noise. If we consider the brainstem neural activity as the plant with unknown time-variant coefficients \mathbf{g} , we may be able to estimate these weights before presenting any stimulus. Then, when the system is identified, we can use these parameters to estimate the noise and then obtain a speech ABR with a higher SNR.

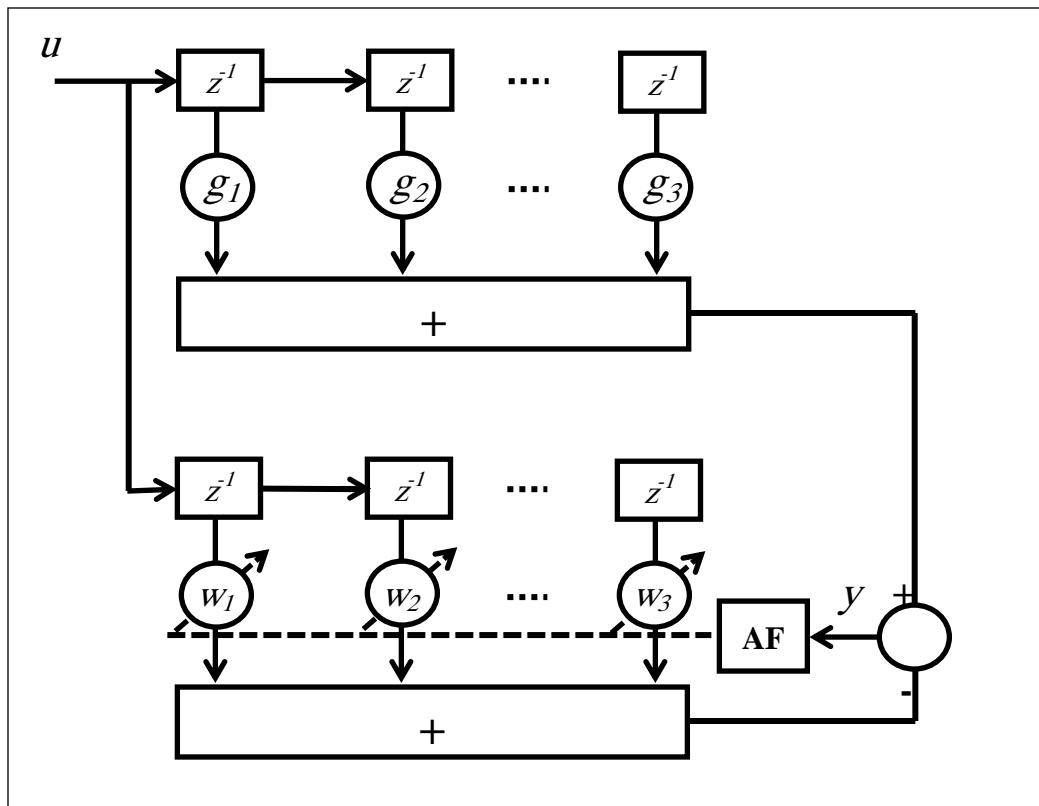


Figure 2-5 System identification using adaptive filters. AF refers to Adaptation Algorithm, u is the observation signal, y is the output, w and g are the coefficients, and z^{-1} is the time delay.

In the literature, a few studies successfully modeled the parameters or function of evoked potentials. Some of them consider only the linear behaviour of ABR (Cheah and Hou, 2010), while others consider the nonlinearity of ABR (Shi and Hecox, 1991). In addition, some of them focus on estimating the functional behaviour (Rauner et al., 1984), while others estimate the parameters (McNamara and Ziarani, 2004).

2.6.3 Adaptive Noise Cancellation (ANC)

Adaptive noise cancellation is usually used to suppress an additive interference, when a noise reference signal is available (Widrow et al., 1975). Figure 2-6 shows an FIR implementation of an ANC filter. One input to the FIR filter is the noise or the reference signal, and the output of the filter is subtracted from the observation signal to produce the final desired signal.

ANC has a variety of applications in the area of biomedical measurement. For example, ANC is suitable to remove any known equipment noise, 60Hz power line noise, EEG noise, or some of the body-activity related noise that can contaminate biomedical measurements (Widrow et al., 1975).

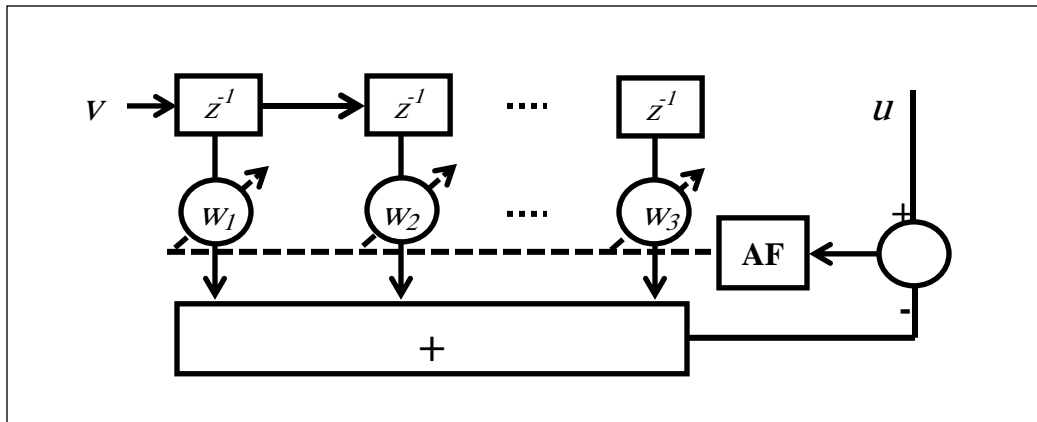


Figure 2-6 Block diagram of Adaptive Noise Cancellation, AF refers to Adaptation Algorithm, u is the observation signal, v is the noise reference signal, w are the coefficients, and z^{-1} is the time delay.

CHAPTER 3: METHODOLOGY

This chapter describes the methodology used in this thesis. We can divide the methods into four categories: 1) The experimental set-up which includes a description the experimental protocol, the speech stimulus, the ABR recording equipment, and the participating subjects, 2) The signal analysis and metrics used to assess performance of the filters, 3) The adaptive filters used to enhance the ABR response, and 4) pre-processing to reduce false positives when the inputs are pure noise signals.

3.1 EXPERIMENT SETUP

Fifty-one blocks of speech ABR signals were gathered from four subjects. Thirty-six of them correspond to the coherent average of individual responses to 250 repetitions of the synthetic vowel /a/. Fifteen other blocks correspond to Silence¹ stimulus (see Appendix A for more detail). Two other signals were gathered: one is used as ABR reference signal (referred to as the “desired signal”) and it was obtained by averaging coherently 3000 repetitions over twelve other subjects for a total of 36,000 repetitions (shown in Figure 2-2 in the case of the EFR), and the other signal is a sample of EEG noise that contains a strong component at 60 Hz. Some of the responses were collected as part of other studies conducted in the lab.

3.1.1 Audiometric Test

The four subjects (all male, age range: 25-45 years) were asked to sit in an acoustical booth located in the University of Ottawa Health campus. All subjects provided their informed consent in accordance with the guidelines of the University of Ottawa Research Ethics Board. An initial audiometric test was made to ensure that they have normal hearing with a threshold below 15 dB HL in both ears for pure tones at 500, 1000, 2000, and 4000 Hz. This test was done using an Interacoustics® AC40 Clinical Audiometer. The subjects were then asked to

¹ i.e. without presentation of any stimulus

watch a muted subtitled movie to help them stay awake for the rest of the experiment.

3.1.2 Construction of the Stimulus

A formant synthesizer was used to generate a 300 msec male speech stimulus vowel /a/. This stimulus has a fundamental frequency (F0) of 100 Hz, first formant (F1) at 700 Hz, second formant (F2) at 1220 Hz, and third formant (F3) at 2600 Hz (Klatt, 1980). Figure 3-1 shows the spectrum of the generated synthetic vowel /a/.

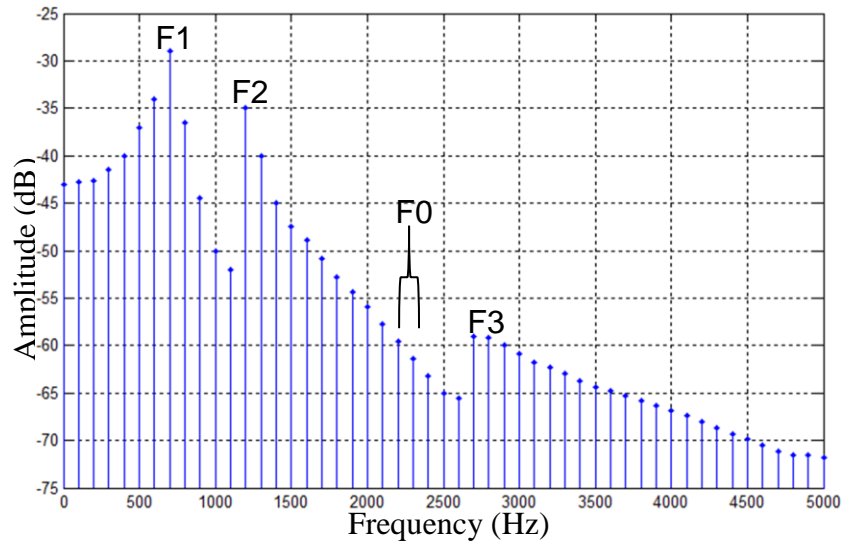


Figure 3-1 Amplitude spectrum of the synthetic vowel /a/ stimulus
(adapted from Laroche, 2010)

3.1.3 ABR Recording

The speech ABR signals were collected using the BioMARK™ system (Biological Marker of Auditory Processing, Bio-logic Systems Corporation). This device delivers the stimulus through small foam earphones that are inserted in both ears. The presentation of the stimulus over multiple repetitions was done in alternate (positive and negative) polarity.

The ABR recording system uses three electrodes that are placed on the surface of the scalp. The recording electrode is placed at the vertex (Cz), the reference electrode is placed on right earlobe, and the ground electrode is placed on the left earlobe. The obtained ABR signal passes through an in-line band-pass filter (30 – 1000Hz), which suppresses low frequency noise and emphasizes the targeted stimulus components of F0 and F1 (at 100 and 700Hz). Each ABR was recorded synchronously with the presented stimulus, over an epoch 319.8 msec long. In a recorded epoch, 1024 points are obtained giving a sampling frequency of approximately 3202 Hz.

Two types of speech ABRs are obtained by coherently averaging the obtained responses to the multiple repetitions, the Envelope Following Response (EFR) and the Frequency Following Response (FFR). The EFR is obtained by averaging the responses to the two stimulus polarities and it emphasizes brain activity that follows the envelope of the stimulus (Aiken and Picton, 2008). It is therefore suitable for analysing the response at the fundamental frequency (F0) and its harmonics. The FFR, on the other hand, is obtained by averaging the differences of the responses to the two stimulus polarities and it emphasizes brain activity that directly follows the stimulus harmonics. It is therefore suitable for analysing the response at the first formant (F1).

3.2 SIGNAL ANALYSIS AND PERFORMANCE METRICS

The speech ABR signal consists of an initial transient response of around 20 ms followed by a quasi-periodic steady-state response, which lasts to the end of stimulus. The interest in this work is in enhancing the steady-state portion of the response, especially at F0 and F1, and so the initial transient of 19.8 ms is removed prior to further analysis.

A number of metrics were considered to assess the performance of the proposed algorithms including some which have been used in assessing speech enhancement algorithms such as Average Segmental SNR (ASSNR), Log-Area Ratio (LAR), and Perceptual Evaluation of Speech Quality (PESQ) (Mustiere et al., 2007). However, we decided to use the signal-to-noise ratio (SNR) and mean square error (MSE) because they are not dependent of perceptual considerations like some of the other metrics mentioned above and so are most straightforward to interpret. This section discusses the performance metrics that were used to assess algorithm performance. It was also important to test if the algorithms produced “false positives”, that is components related to the stimulus when they should not be present. This forms part of the algorithm validation and is discussed in a later section.

3.2.1 Signal-to-Noise Ratio

The Signal-to-Noise ratio (SNR) in dB is calculated after obtaining the Discrete Fourier Transfer (DFT) of the speech ABR time series using the following equation:

$$SNR_{dB} = 10 \log_{10} \left(\frac{P_{signal}}{P_{noise}} \right) \quad 3.2.1$$

where P is the power of signal and noise components.

Two different types of SNR are calculated to assess performance, the overall and the local SNR. In the overall SNR, referred to as oSNR, the power at the fundamental frequency and at its harmonics (100, 200,... up to and including

700Hz) is combined by adding the square of the root mean square amplitudes (RMS) at these frequencies, while the power in the remaining signal components from 1 to 790 Hz is combined to give noise power. They are calculated as follow:

$$P_{signal} = \sum_{i=1}^n a_i^2 \quad 3.2.2$$

$$P_{noise} = \sum_{i=1}^m b_i^2 \quad 3.2.3$$

where n is number of the signal components (which are seven at 100, 200, ... 700 Hz) in the frequency domain, and m is number of noise components (which go from 1 to 790 Hz excluding the seven signal components).

On the other hand, the local SNR is calculated separately at F0 (100Hz) in the EFR and at F1 (700 Hz) in the FFR. The local SNR at F0 is referred to as ISNR100 and is calculated in equation (3.2.1) using the signal power at 100 Hz and the noise power over the interval between 100 – 200 Hz, excluding the components at 100 and 200 Hz as illustrated in Figure 3-2. So n equals one in equation (3.2.2) and m in equation (3.2.3) is the number of noise components between 100 and 200 Hz.

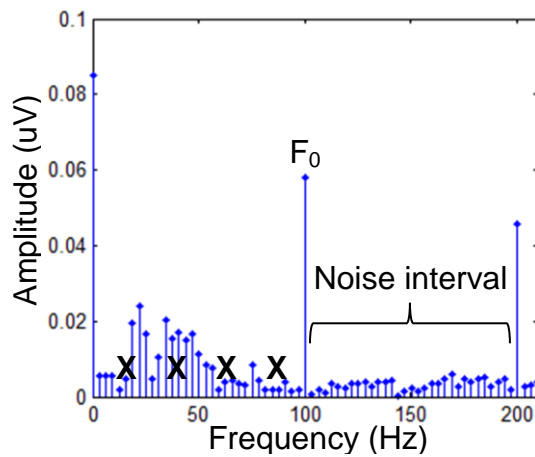


Figure 3-2 The noise interval used for calculating local SNR at F0 (ISNR100)

Note that the interval before F0 is ignored in this performance measure since this interval includes any 60Hz electromagnetic noise and the high noise power belonging to EEG activity that is unrelated to auditory processing.

The local SNR at F1 is referred to as ISNR700 and is calculated in equation (3.2.1) using the signal power at 700 Hz and the noise power over the interval between 600 – 800 Hz, excluding the components at 600, 700, and 800 Hz as illustrated in Figure 3-3. So a is calculated only at 700 Hz in equation (3.2.2), and m in equation (3.2.3) is the number of noise components between 600 and 800 Hz excluding 700.

It should be noted that when the SNR is averaged, for example over subjects, it is first converted to its linear form as the square-root of the ratio of signal to noise power, averaged, and then converted back to its logarithmic form shown in equation (3.2.1).

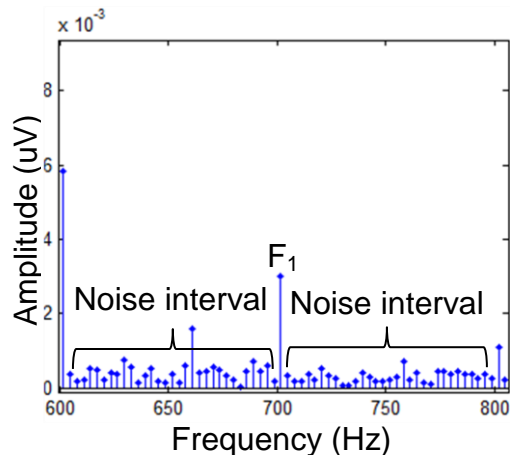


Figure 3-3 The noise interval used for calculating local SNR at F1 (ISNR700)

3.2.2 Mean Square Error

The Mean Square Error (MSE) is a fundamental measure that can be used to assess the difference between two signals. In our case, we calculated the MSE between the recorded ABR signals and the reference signal (MSE_q) (estimated by LQ factorization) and the desired signal (MSE_d) (the ground ABR signal with 36,000 repetitions) in both time and frequency domains. MSE can be expressed using the following equation:

$$MSE = E[(\mathbf{u} - \mathbf{d})^2] \quad 3.2.4$$

3.3 ADAPTIVE FILTER DESIGN

Four different algorithms were designed to enhance the speech ABR: The Wiener filter (WF), Steepest-descent (SD) adaptive filter, and Adaptive Noise Cancellation (ANC) based on LMS and normalized LMS (nLMS) algorithms. WF and SD were configured according to two signals, and then fixed and applied to the 250 repetition speech ABR signals. Since ANC needs prior knowledge of the noise, an LQ factorization method (see Appendix B) was used to estimate the noise in the LMS and nLMS algorithms. In addition, a multi-adaptive approach consisting of the cascade combination of ANC-LMS and SD was applied. Table 3-1 summarizes the algorithms that are performed on the speech ABRs obtained by coherently averaging responses over 250 repetitions.

Table 3-1 Summary of the analysis performed on the 250-repetition speech ABR

250-repetition speech ABR	
With stimulus components	Stimulus components removed
Wiener Filter	
Steepest-descent	
ANC-LMS	
ANC-nLMS	
SD + ANC-LMS	

The configurations of the four different algorithms that were implemented for enhancing ABR signals are described as follows:

3.3.1 Wiener Filter

The Wiener filter (WF) is the only optimal filter that has been used in this work. It has an un-adapted configuration and was configured based on the speech ABR response obtained with 250 repetitions. Since the number of repetitions is the significant factor in affecting the important statistical parameters of the response, a configured WF based on a certain number of repetitions can be applied to enhance another ABR from another subject obtained with the same number of repetitions.

To calculate the WF optimal weights (w_o), equation (2.4.3) was used where the cross-correlation matrix (p) is obtained between the speech ABR signal of block one from the first subject, since its autocorrelation is the closest one to the mean autocorrelation based on MSE criterion (Appendix C), and the desired signal, which is the ABR reference signal based on the average over 12 subjects of 3000 repetitions obtained from each subject. The autocorrelation vector (R) is obtained from the speech ABR of block one from the first subject.

The number of coefficients of the FIR filter affects SNR performance since it affects the length of the autocorrelation matrix R (as will be seen in equation 3.3.1). This factor has to be considered when we want to compare the performance of the filter with that of coherent averaging method. Comparing these two approaches while selecting improper number of coefficients would result in an unfair comparison. For instance, if a small number of coefficients has been selected, the output of the filter would be expected to have a small SNR, from which we could conclude that coherent averaging method provides a better SNR. An appropriate number of coefficients then would be at a point when the effect of increasing the coefficients this number would not be significant. The following figure illustrates the overall SNR (oSNR) performance of the designed WF with different number of coefficients.

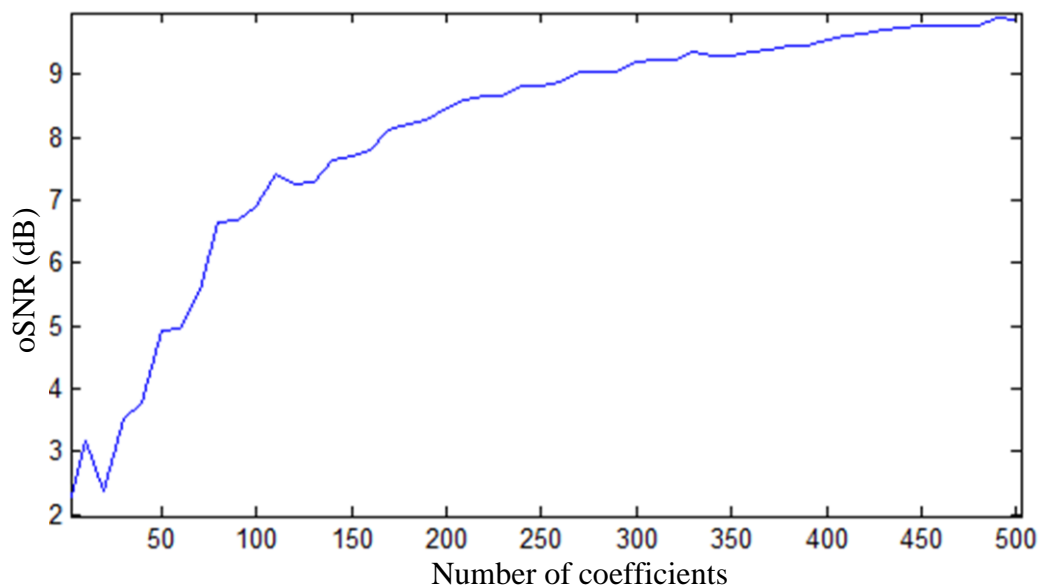


Figure 3-4 The estimated increase of overall SNR of the configured WF with different number of coefficients.

The number of coefficients is chosen to be 100, which is considered acceptable for this WF since its performance for suppressing noise beyond 100-taps was found to not significantly improve. This number of coefficients is also used for

SD algorithm which is discussed in the next section. Theoretically, the overall SNR performance of WF can be calculated by the following equation:

$$oSNR = \frac{\mathbf{w}^T R_y \mathbf{w}}{\mathbf{w}^T R_n \mathbf{w}} \quad 3.3.1$$

where R_y is the autocorrelation of the filter's output and R_n is the autocorrelation of the noise (Benesty and Chen, 2011). Based on this equation, the WF is expected to increase the input ABR's SNR by around 7.45 dB. Its MSE is calculated by (Benesty and Chen, 2011):

$$MSE = \sigma_d^2 + \mathbf{w}^T R_y \mathbf{w} - 2\sigma_d^2 \mathbf{w}^T \mathbf{p} \quad 3.3.2$$

and it is equal 0.0056. These values are expected in the present ABR signal that corresponds to the coherent average of 250 repetitions of the stimulus vowel /a/. On the other hand, when applying WF on the ABR with stimulus components removed using LQ factorization, a linear SNR of zero is expected in the ideal case (Section 2.4.1). Also, higher MSE is expected.

3.3.2 Steepest-Descent algorithm

The Steepest-Descent (SD) algorithm is considered a fundamental adaptive filter that uses the gradient iteration for searching for the optimal solution. It starts with initial arbitrary values for its coefficients, usually set at zero, and then iterates to reach to the optimal values that result in the minimal cost, MSE in our case, relative to a desired signal. The SD update equation is described as follows (Sayed, 2011):

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu(\mathbf{p} - R\mathbf{w}(n)) \quad 3.3.3$$

Where p and R are the same value as used in the WF and the step-size (μ) is chosen to be $1/\lambda_{max}$ which is less than the maximum value shown in equation (2.5.4), in order to avoid instability. If we compare this equation with the equation of the LMS algorithm (equation 2.5.3), we note that the SD algorithm's update is

based on the correlation function, while LMS is based on the previous input and the error signal. This feature allows SD to resist producing a false positive when the target signal is not present, in addition to functioning as an enhancer when the target signal is present. The LMS, on the other hand, tends to reduce the error between the input and the desired signal whatever the input is. This does not apply to ANC with LMS (section 3.3.3) in which the noise signal is the input signal. SD's performance is expected to be similar to that obtained with WF because they use the same correlation function described in the previous discussion about WF, and the same cost function. However, SD adapts itself from the initial value until it reaches the optimal values gradually, and it requires less complexity, to process one block, than WF which requires matrix inversion. Figure 3-5 shows three weights of WF (W_o) and SD (W_{sd}) as they converge to the optimal values. SD converges approximately after the 250th iteration, which will result in some difference in the performance relative to that obtained by WF as will be seen later. Removing the first 19.8 ms before calculating the SD algorithm performance will minimize the effect of the initial transient convergence.

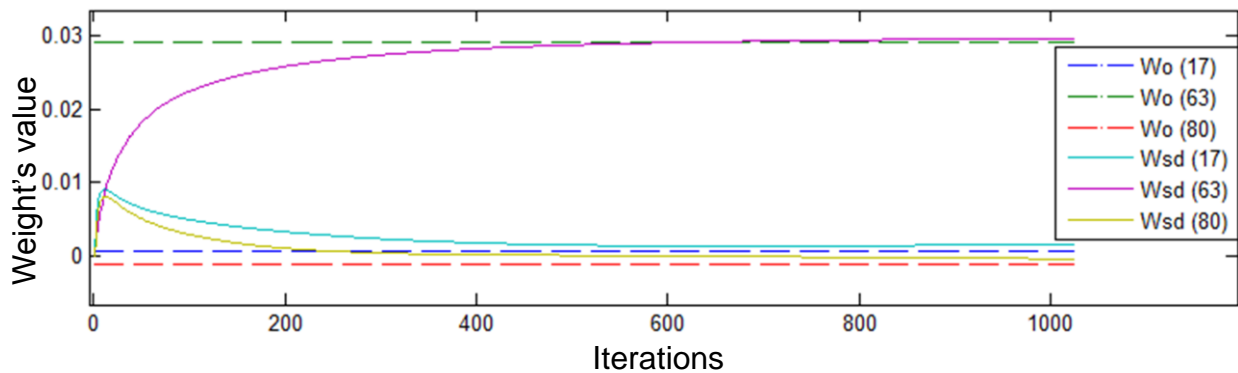


Figure 3-5 Convergence of the designed WF and SD weights. Three weights of WF (W_o) and SD (W_{sd}) are shown as they converge to the optimal values.

3.3.3 Adaptive Noise Cancellation

Adaptive noise cancellation (ANC) is used to suppress the background noise in the speech ABR signals. It is structured as seen in Figure 2-4. In this figure, the input of the FIR filter is the noise (v), and the adaptive filter (AF) depends on the difference between the FIR filter's output and the ABR signal. In online applications, the noise terminal can be taken from the source of the targeted noise. For instance, if the goal to suppress is the 60 Hz noise, the signal from the wall outlet can be taken as noise reference.

Since we have pre-knowledge of the expected ABR which will corresponds to vowel /a/ stimulus, a factorization method can be applied onto the obtained ABR to estimate the noise reference. For this reason, LQ factorization (see appendix B) has been used to estimate all components that are not related to the stimulus and used as a noise reference in the ANC algorithm.

In the ANC algorithm, a seven tap FIR filter has been used with two different adaptive filters from the LMS family, namely LMS and normalized LMS (nLMS). The update equation of LMS is shown in equation 2.5.3, while the update equation for nLMS is as follows (Sayed, 2011):

$$w(i) = w(i - 1) + \mu \frac{u(i) e(i)}{\|u(i)\|^2} \quad 3.3.4$$

where $w(i)$ is the coefficient vector, $e(i)$ is the error, and $u(i)$ is the observation at (i). μ was selected to be 0.04 for LMS and 0.1 for nLMS.

3.4 VALIDATION WITH NOISE SIGNALS

Quality is a critical factor in applications where important decisions have to be made based on detected signals. In the biomedical field, for example, measurements have to be of acceptable quality and reflect well the statistical aspects of the underlying measured signals. Failing to do so could lead to errors in the assessment of critical medical status or prevent the detection of early stage disease. It is also critical that the measuring instrument would not produce, or at least minimize, “false positives”. In the current application, the estimated output signal should not contain components related to the speech stimulus when they should not be present, for example when testing a patient with profound hearing impairment. In the ideal case, the used algorithm would exactly estimate the underlying components of the speech ABR and nothing else, and it would do so if and only if these components were present in the measured signal.

In this section, a discussion of how the FIR filters, used in this study perform when four different types of signals that do not contain the target speech ABR (i.e. noise-only signals) are used as input. These four signals are: EEG noise with strong 60Hz power-line noise, another EEG noise signal recorded with no stimulus present, white noise, and an autoregressive (AR) model signal generated using an autoregressor of 5th order. Since all proposed algorithms in this work are FIR-based, the performance of the Wiener Filter (WF) with noise inputs is used to guide the pre-processing performed prior to applying the enhancement algorithms. The performance of all the five algorithms with noise signals is tested experimentally and reported at the end of this section.

As mentioned earlier, the WF is expected to detect ABR signals if there is a response to the stimulus, and it produces a lower SNR if there is no response related to the stimulus. However, the WF is also expected to enhance the components of any noise-only signals that correspond in frequency to the stimulus components of interest. This enhancement depends on statistical properties of the noise. For example, if highly correlated noise is used as input, there is a higher chance that the output signal will have similar characteristics to a valid ABR signal.

Now let us consider the four different types of noise, two of which were generated computationally (the random Gaussian white noise and AR model noise), while the others were recorded in the lab from two participating subjects when no stimulus was present (EEG noise). Figures 3-6, 3-7, 3-8 and 3-9 show these noise samples (in time and frequency domain) before and after processing with WF.

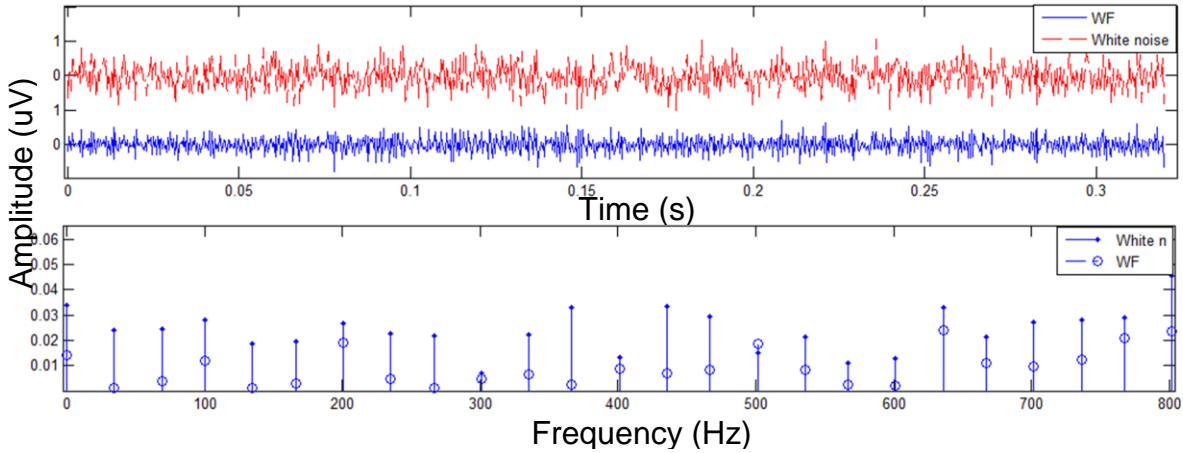


Figure 3-6 Effect of WF on random white noise

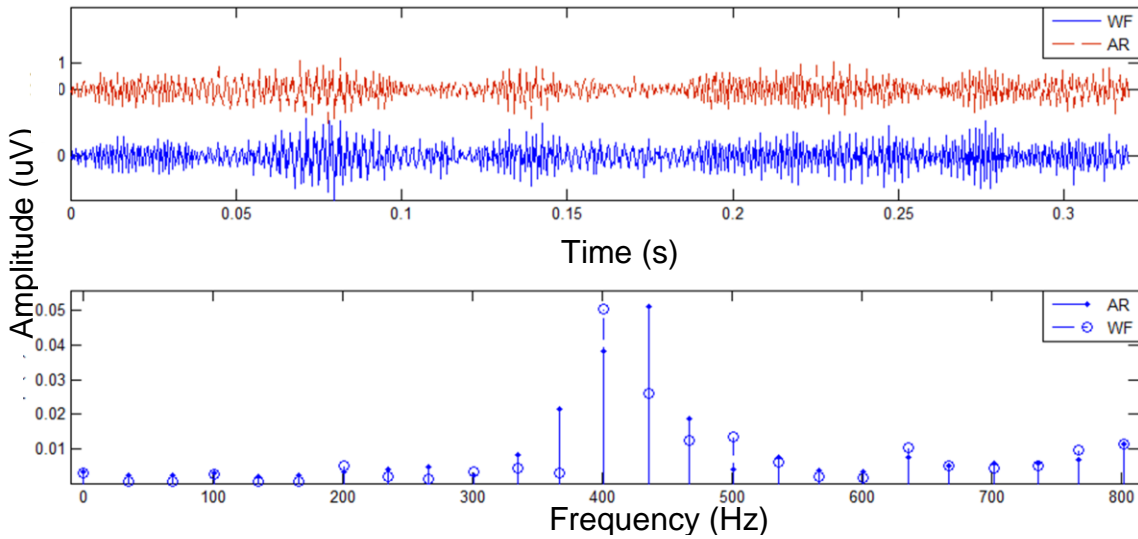


Figure 3-7 Effect WF on AR modeled noise

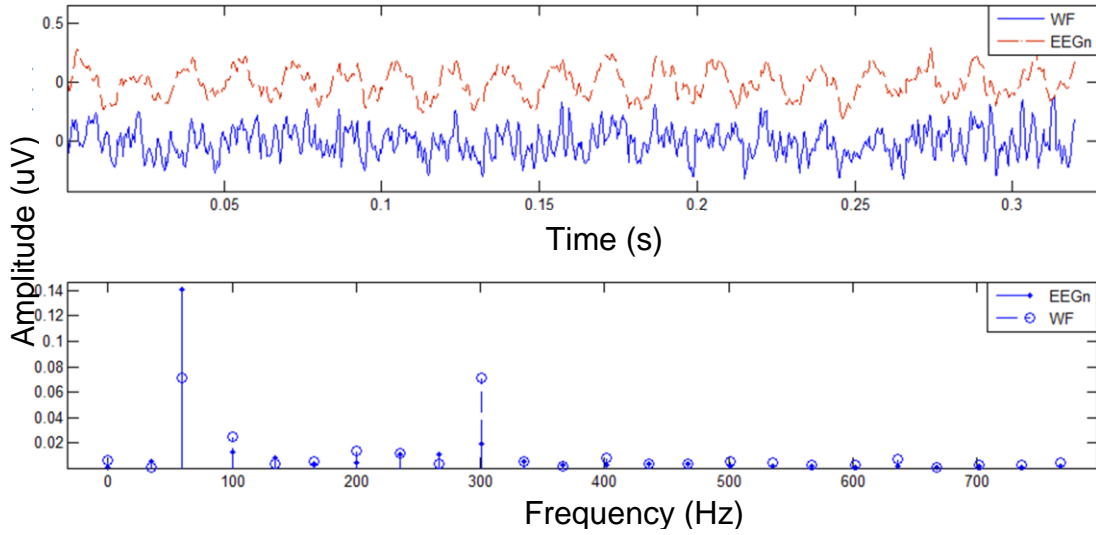


Figure 3-8 Effect of WF on one sample of EEG noise with a strong 60 Hz power-line noise component (EEGn1)

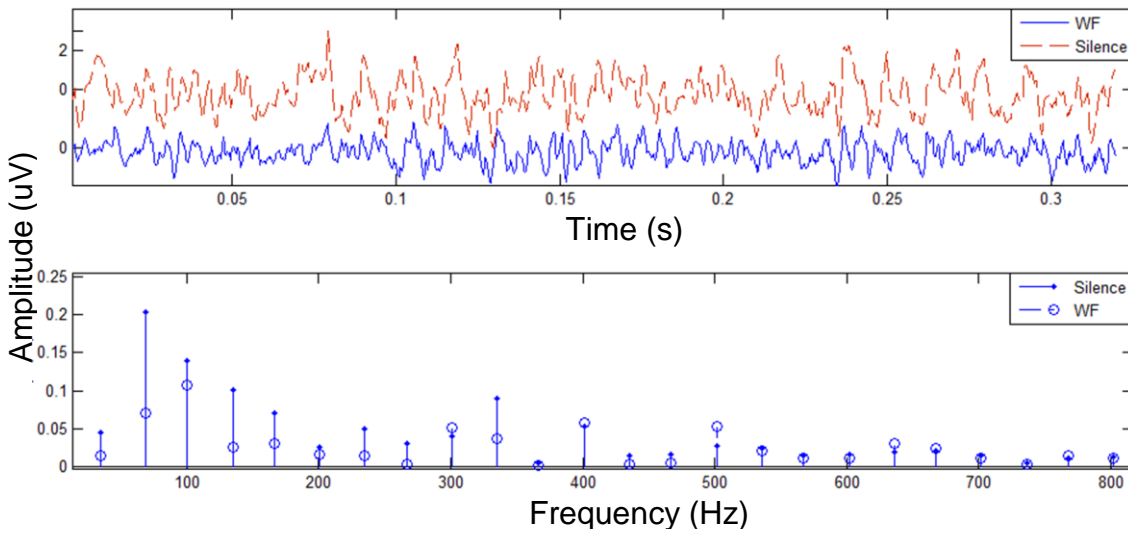


Figure 3-9 Effect of WF on a second sample of EEG noise (EEGn2)

Table 3-2 shows that there is some SNR enhancement that varies with the type of noise. Such unwanted enhancement, however, can be prevented if a threshold detector is implemented prior to the WF.

Table 3-2 Performance of WF with noise signals

	Unfiltered Variance	Unfiltered oSNR (dB)	WF oSNR(dB)	oSNR increase (dB)	ISNR100 increase (dB)	ISNR700 increase (dB)
White noise	0.13	-0.98	2.00	2.99	0.97	-3.77
AR	0.13	-0.59	1.00	1.59	-0.63	-2.61
EEGn1	0.02	-2.31	9.18	11.49 ²	1.39	-7.80
EEGn2	0.95	-4.03	0.97	4.99	2.50	-3.58

Most of methods that have been used to detect periodicity (such as the vowel's pitch) in speech could be applied to the speech ABR signals since these periodicities are usually represented in the response. A well-known simple method that has been used to detect any pattern in time series is the autocorrelation function. The simple autocorrelation function of the input signals is used here to see if they lie within the range of autocorrelation functions obtained from signals known to contain responses, and if they follow a similar peak pattern (Appendix C). In future work, more advanced techniques for applying autocorrelation function in this detection task could be used, including segmented autocorrelation (Atkinson et al., 1995), wavelet correlation (Kader, 2000) and the modified autocorrelation method by Yang et al. (2010).

The following figures show in block diagram the pre-processing of the signal with the autocorrelation threshold function prior to WF and SD in Figure 3-10 and prior to ANC in Figure 3-11. Using this threshold function prior to enhancement ensures that only input signals that contain response components are enhanced and so suppresses the

² This high oSNR resulted mainly from a component at 300Hz in the unfiltered EEGn1 spectrum as seen in Figure 3-7.

possibility of enhancing input signals that are pure noise. The major difference between these two diagrams is that in the WF and SD diagram the desired signal will be selected as reference if the threshold detects a speech ABR whereas in the ANC diagram the noise signal will be selected as a reference if the threshold detects a speech ABR signal. It is common to switch the adaptive filter ON and OFF according to a threshold. However, switching between different reference signals instead allows the adaptive filter to suppress the interference (the noise components at the frequency of interest).

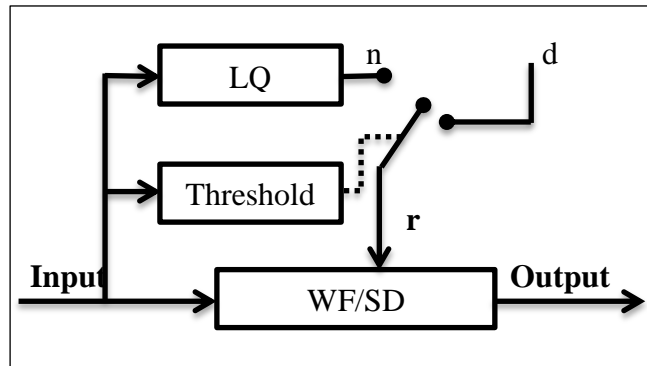


Figure 3-10 Block diagram of WF and SD algorithms, with a pre-processing threshold function that is used to control the reference signal switch (r). If speech ABR is detected, the desired signal will pass. Otherwise the noise signal (n) that is generated by LQ factorization will pass.

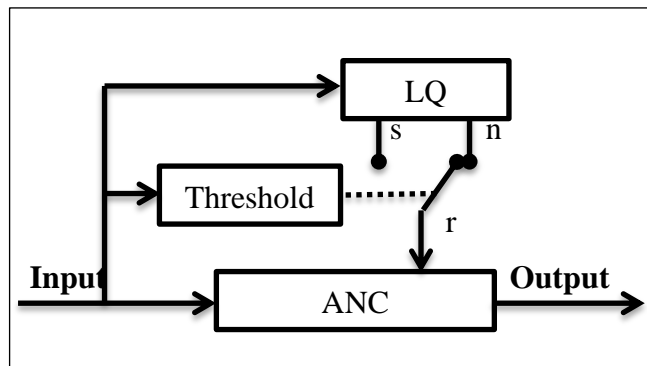


Figure 3-11 Block diagram of ANC algorithm, with a pre-processing threshold function is used to control the reference signal switch (r). If speech ABR is detected, the noise n (from LQ factorization) will pass. Otherwise the signal (s) will pass.

All algorithms use the following steps:

1. The observation signal (u) enters the Input terminal.
2. The observation signal enters the threshold block to be evaluated, and also enters the LQ Factorization block to estimate the noise (n) and signal (s).
3. a) In WS and SD, if the threshold detects a speech ABR signal, the desired signal (d) will be selected as reference signal. In ANC, if the threshold detects a speech ABR signal, the noise (n) will be selected as a reference signal.
b) In WF and SD, if the threshold detects a noise signal, the noise (n) will be selected as reference signal, In ANC, if the threshold detects a noise signal, the signal (s) will be selected as reference signal.
6. The adaptive filter (WF, SD or ANC) uses the selected reference and the input observation signal (u) to produce the output signal.

Table 3-3 shows the oSNR performance of all algorithms with pre-processing using the threshold function based on autocorrelation as described above. The performance is shown with “no-ABR” where the input signals are the 36 responses that contain no components related to the stimulus after removal of these components using LQ factorization and averaged over 36. The performance is also shown with the four noise samples (AR, White Noise, EEGn1, and EEGn2). As can be seen, at least in the case of the EEG noise samples, the filters now do not result in an increase of the oSNR relative to the unfiltered signal (compare with Table 3-2).

Table 3-3 oSNR performance of all algorithms with no-ABR where it is averaged over the 36 responses that contains no components related to the stimulus, after removal of the components using LQ factorization, and with the four noise samples (AR, White Noise, EEGn1, and EEGn2). Pre-processing is performed first by passing the signal through the autocorrelation-based threshold function as described in the text.

oSNR (dB)	Unfiltered	WF	SD	LMS	nLMS	SD+LMS
no-ABR	-12.83	-7.56	-7.85	-6.80	-2.37	-0.91
AR	-0.59	-0.90	-0.03	-1.02	-7.56	-1.15
White Noise	-0.98	-2.31	-1.60	-1.31	-8.46	-2.16
EEGn1	-2.31	-7.85	-8.12	-2.46	-17.58	-8.22
EEGn2	-4.03	-5.05	-4.97	-7.81	-4.86	-6.06

The following two figures illustrate the time and frequency domain performance of all the adaptive filters with the Electroencephalogram noise sample (EEGn1) that has a strong component at 60Hz, with pre-processing using the autocorrelation-based threshold function. As can be seen, with this EEG noise signal, there is no enhancement of the components at the harmonics of F0 which correspond to the speech ABR.

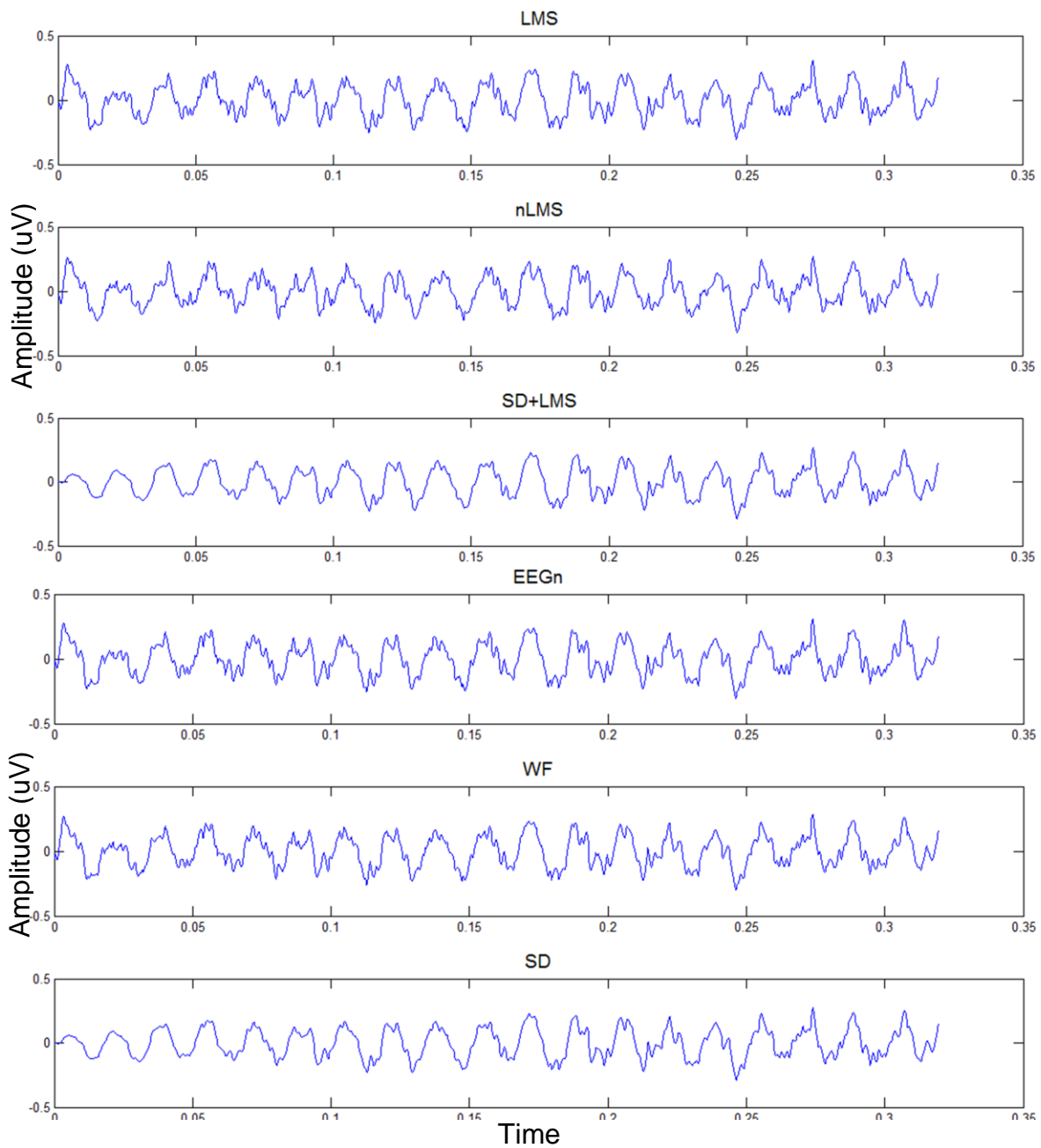


Figure 3-12 Illustration of time domain performance of the adaptive filters with the Electroencephalogram noise (EEGn1) that has a strong component at 60Hz, with pre-processing using the autocorrelation-based threshold function.

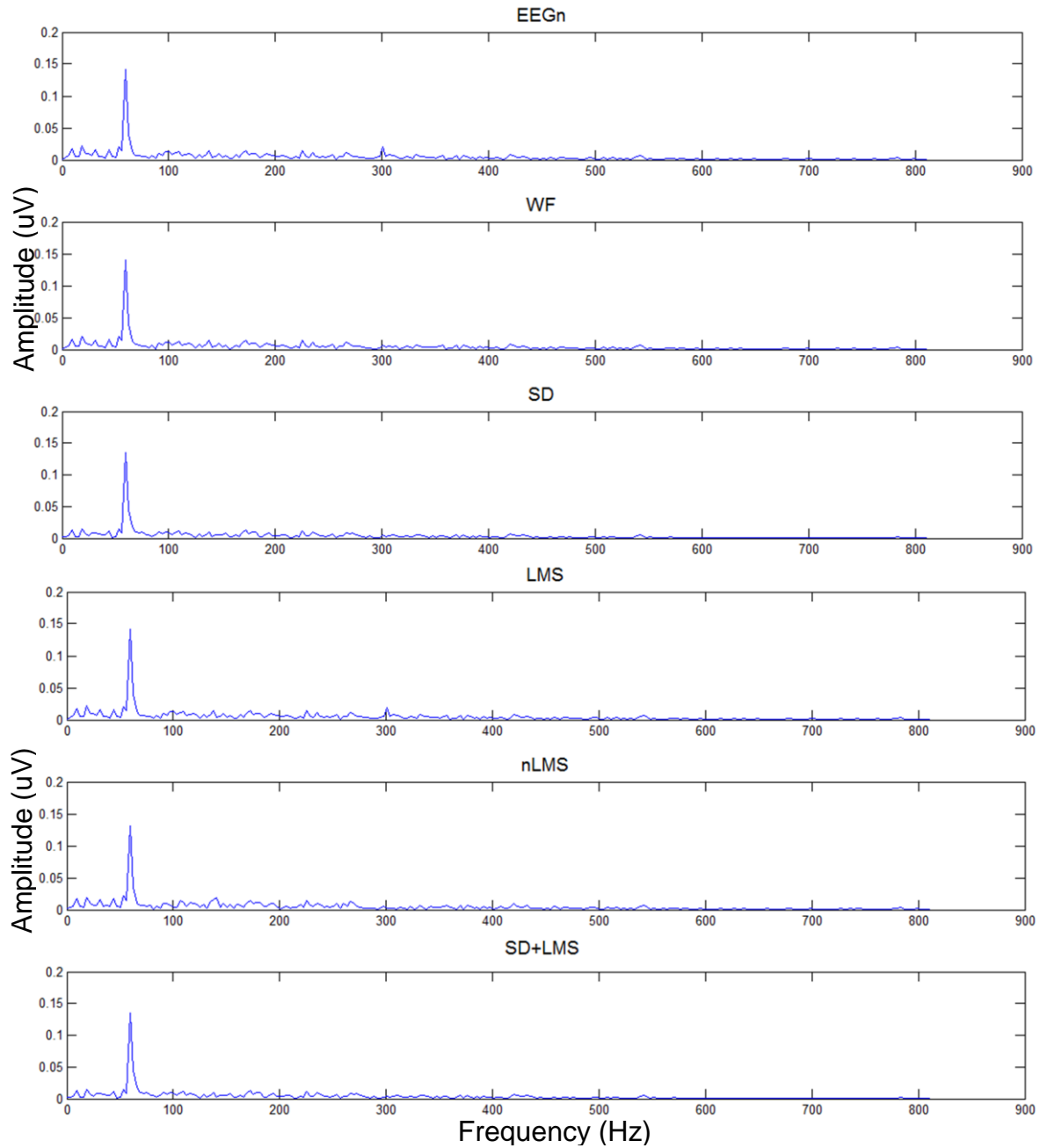


Figure 3-13 Illustration of frequency domain performance of the adaptive filters with the Electroencephalogram noise (EEGn1) that has a strong component at 60Hz, with pre-processing using the autocorrelation-based threshold function.

CHAPTER 4: RESULTS

This chapter presents the results of applying five algorithms on the recorded speech Envelope Following Response (EFR), which is based on the coherent average of the responses to 250 stimulus repetitions. The algorithms are the Wiener Filter (WF), Steepest-Descent (SD), Adaptive Noise Cancellation (ANC) based on Least-Mean-Square error (LMS) and normalized LMS (nLMS), and a multi-adaptive algorithm based on the cascade combination of SD and LMS (SD+LMS). The results of the WF and SD algorithms are shown in the first section. The results of ANC with both LMS and nLMS are discussed in the second section, entitled Adaptive Noise Cancellation. The third section, entitled Multi-adaptive Technique, shows the results of applying two algorithms (SD and LMS) from each of the first two sections connected in cascade. The fourth section compares all obtained results. The last section shows the results with the Frequency Following Response (FFR).

However, before presenting the detailed results, and as an illustration of the performance of the different algorithms, the following two Figures (4-1 and 4-2) show one block of speech ABR (EFR) from one subject in the time and frequency domains before and after filtering with all the adaptive filters. The sample is referred to as a single block because it is based on the coherent average over a block of 250 individual responses. Also included are the coherent average over twelve blocks from the same subject, and the desired signal is the speech ABR reference signal that was obtained by averaging coherently 3000 repetitions over twelve other subjects (Section 3.1).

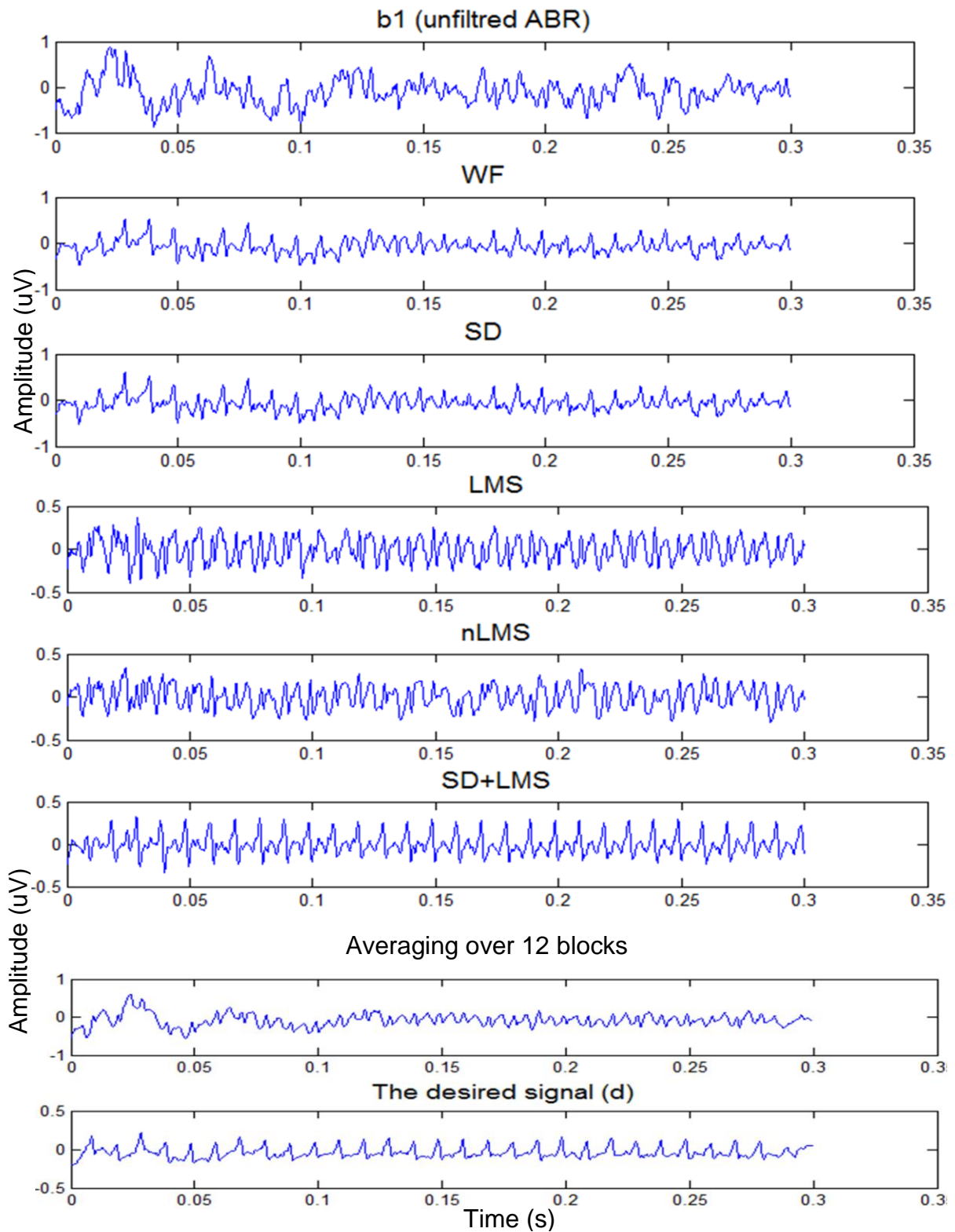


Figure 4-1 Illustration of time domain performance of all the adaptive filters with the Envelope Following Response (EFR). The sample response is a single block based on the coherent average of 250 repetitions (block 1 from subject 1) and is shown before and after filtering. The coherent average over twelve blocks and the desired speech ABR reference signal are also shown

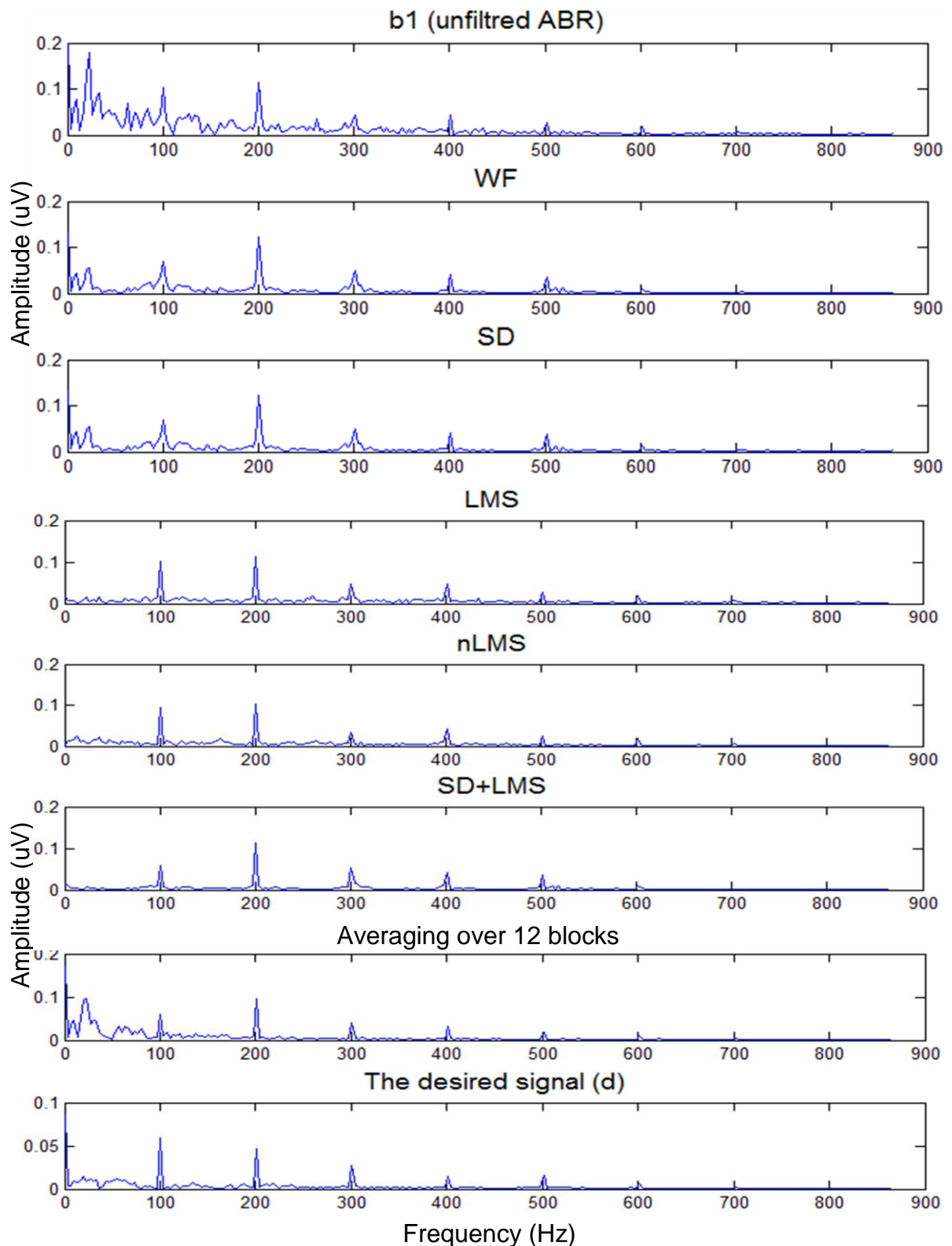


Figure 4-2 Illustration of frequency domain performance of all the adaptive filters with the same single block Envelope Following Response (EFR) shown in the previous figure. The spectra of the coherent average over twelve blocks and the desired speech ABR reference signal are also shown.

4.1 THE WIENER FILTER AND STEEPEST-DESCENT ALGORITHM

The results of applying the Wiener Filter (WF) on the ABR signal were measured and compared to the expected values. Figures 4-3, 4-4, and 4-5 show the overall SNR (oSNR) with the 12 blocks of EFR responses, each based on the average of 250 repetitions, from subject one, two and three respectively. The solid line represents the oSNR of each block before any processing. The circled line represents the output of the WF, and it shows that WF increases the ABR's oSNR by 7.34 dB when averaged over all the blocks. The dotted line represents the output of the WF on the EEG noise obtained from LQ factorization of the ABR (see Appendix B). Since this signal suppressed components that are related to the stimulus, its SNR is always less than 0 dB. The dashed horizontal line represents the oSNR of the coherent average of the 12 blocks for each subject. As can be seen, the WF obtains an oSNR with most individual blocks that is higher than the oSNR of the coherent average of 12 blocks.

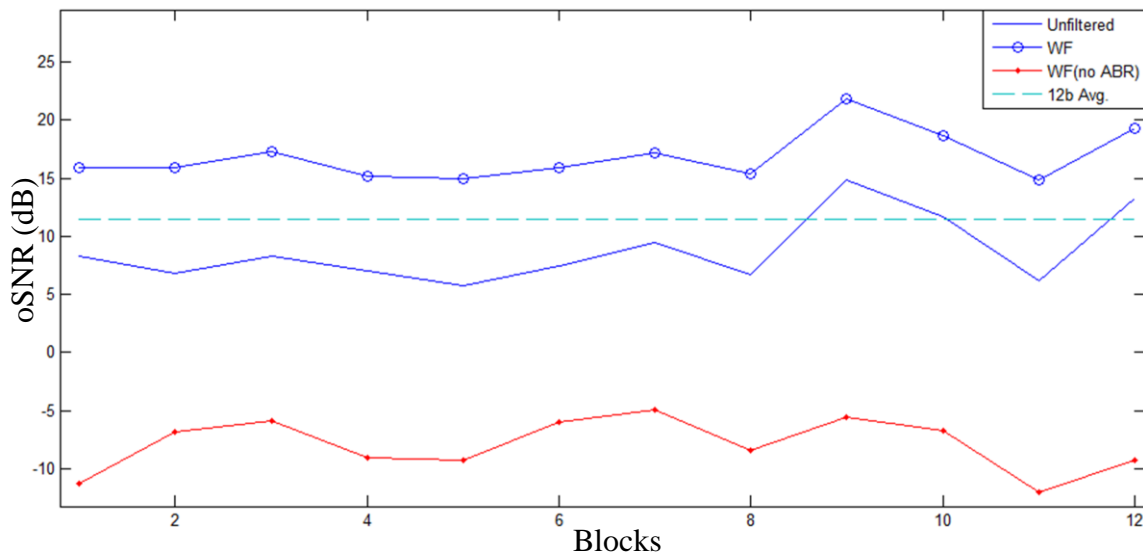


Figure 4-3 Results of WF on speech EFR of the first subject. The solid line represents the unfiltered ABR, the circled line represents the results of the WF, the dotted line is WF of the response that contains no components related to the stimulus, and the dashed horizontal line represents the coherent average over the 12 blocks

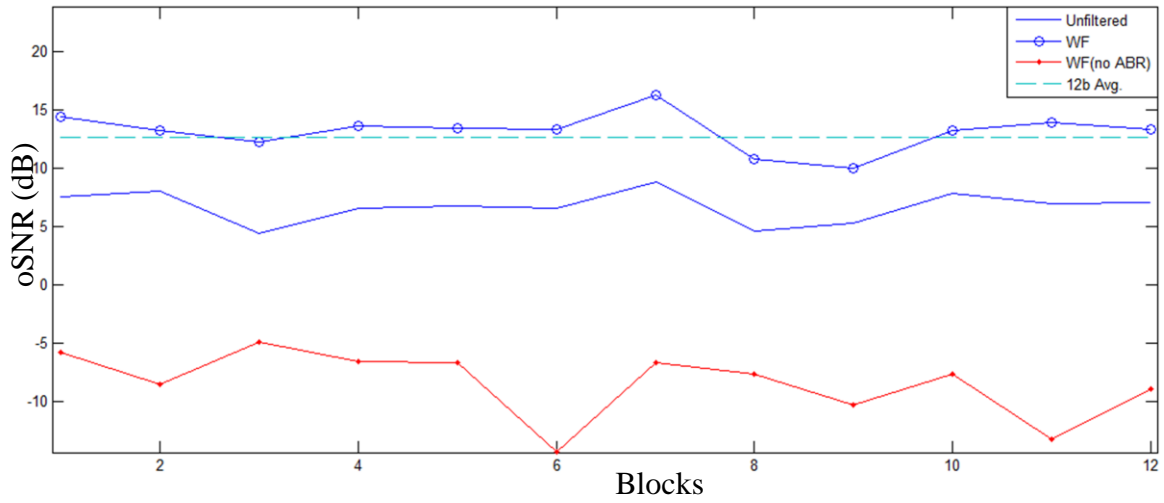


Figure 4-4 Results of WF on speech EFR of the second subject

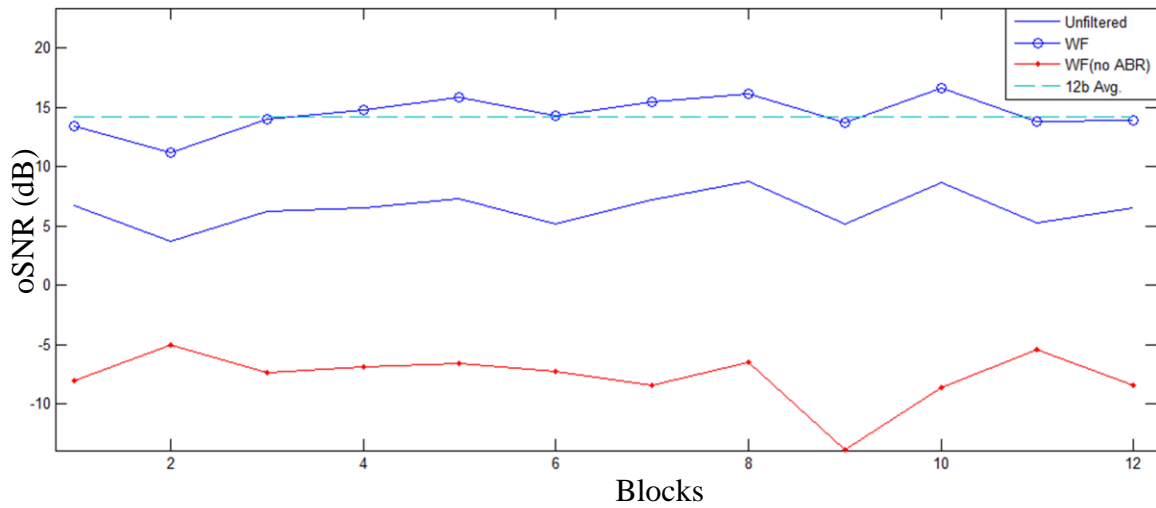


Figure4-5 Results of WF on speech EFR of the third subject

Figures 4-6, 4-7, and 4-8 show the oSNR with the application of the Steepest-Descent (SD) algorithm on the speech EFR from the three subjects. These results are close to WF performance.

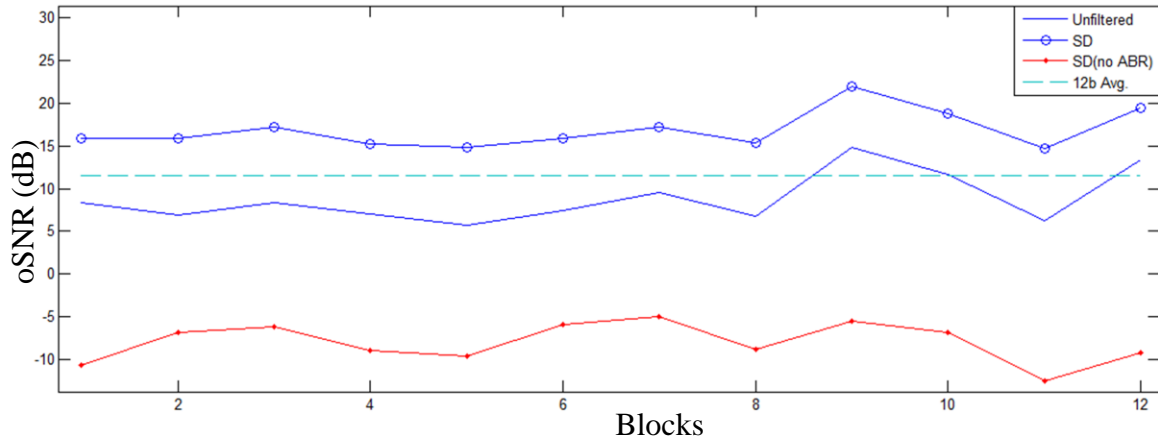


Figure 4-6 Results of SD on speech EFR of the first subject. The solid line represents the unfiltered ABR, the circled line represents the result of the SD algorithm, the dotted line represents the output of SD with a response that contains no components related to the stimulus, and the dashed horizontal line represents the coherent average over the 12 blocks

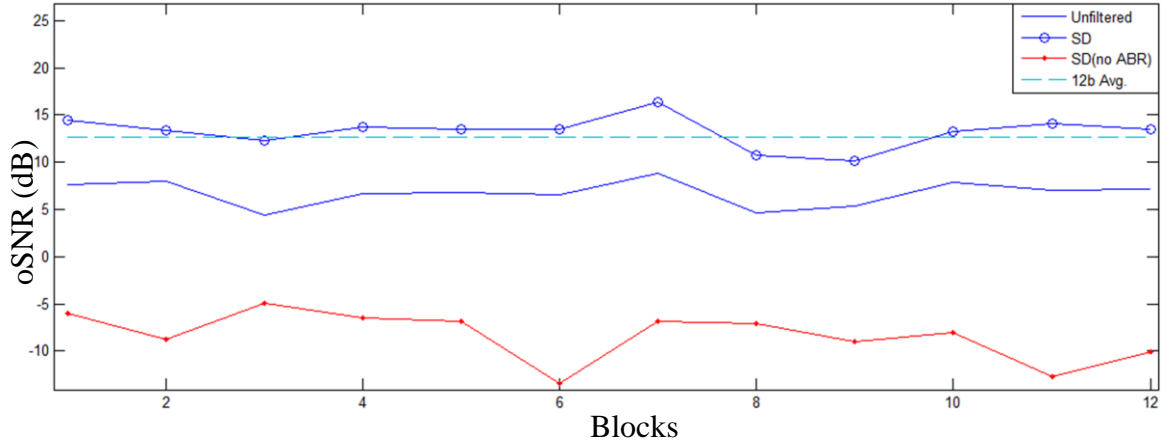


Figure 4-7 Results of SD on speech EFR of the second subject

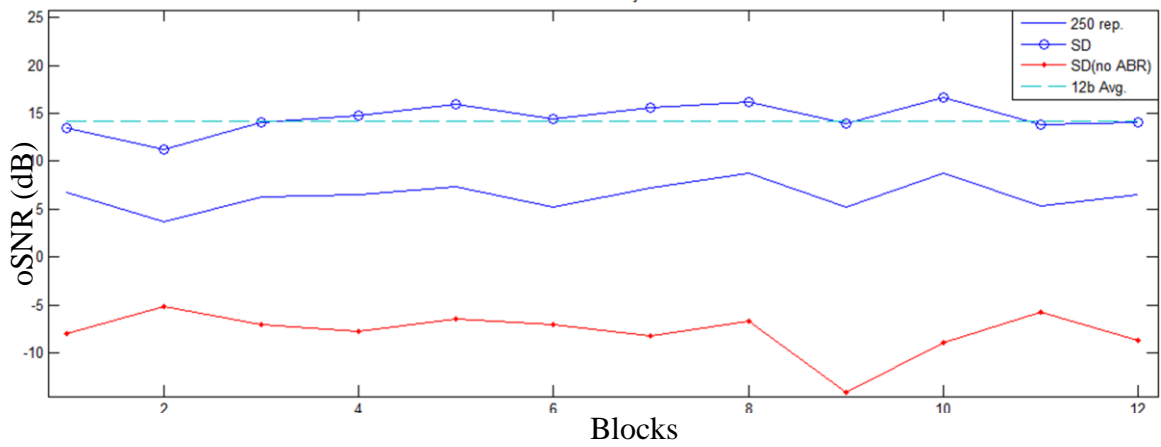


Figure 4-8 Results of SD on speech EFR of the third subject

4.2 ADAPTIVE NOISE CANCELLATION

Figures 4-9, 4-10, and 4-11 show the performance of Adaptive Noise Cancellation with LMS, and figures 4-12, 4-13, and 4-14 show the performance of ANC with nLMS on the blocks of speech EFR responses that are each based on the coherent average of 250 individual responses from the three subjects. As can be seen, ANC with LMS enhances the oSNR by around 10.34 dB, on average, whereas nLMS enhances the oSNR by around 11.52 dB

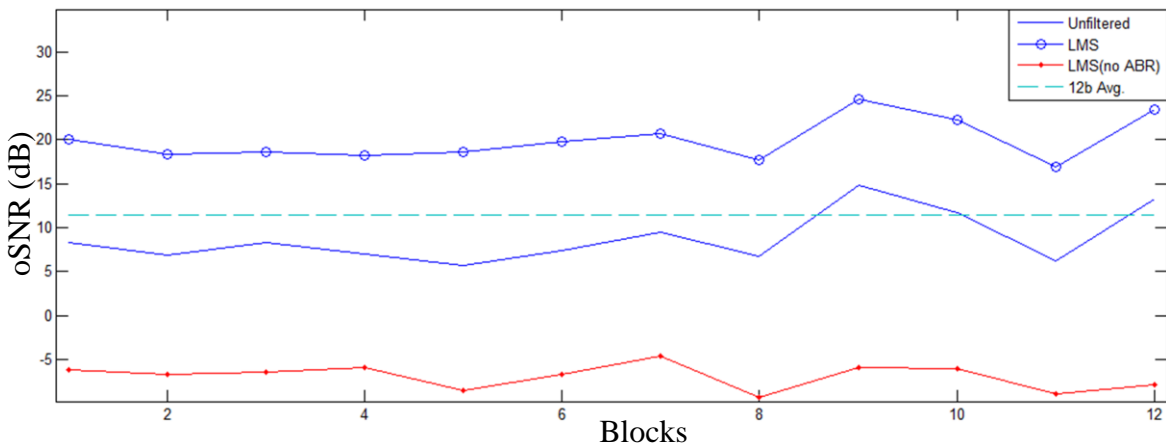


Figure 4-9 Results of ANC with LMS on speech EFR of the first subject. The solid line represents the unfiltered ABR, the circled line represents the result of the ANC with LMS algorithm, the dotted line represents the output of the LMS algorithm with a response that contains no components related to the stimulus, and the dashed horizontal line represents the coherent average over the 12 blocks

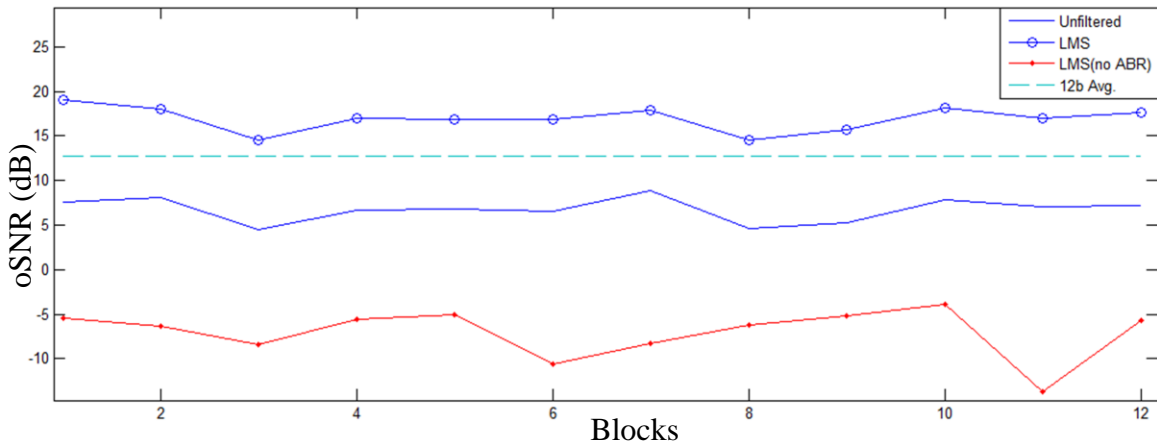


Figure 4-10 Results of ANC with LMS on speech EFR of the second subject

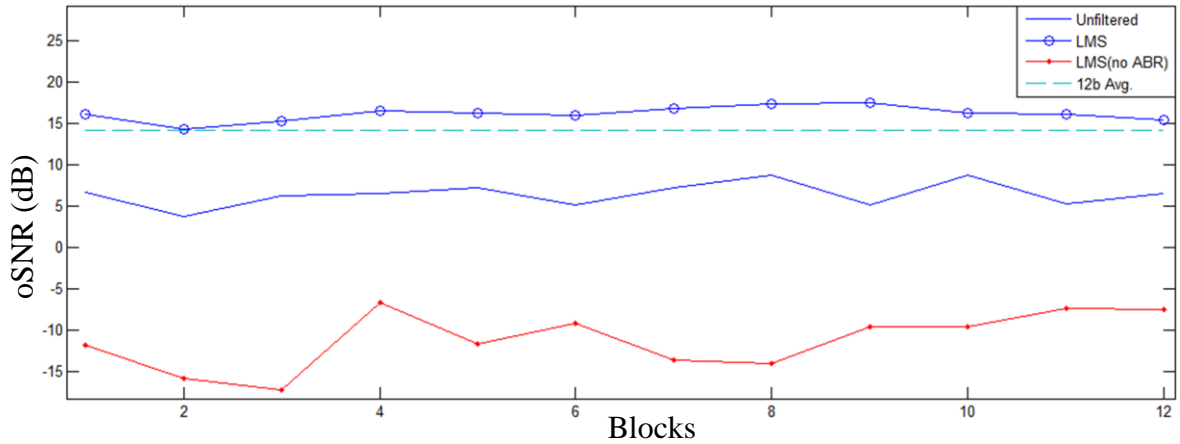


Figure 4-11 Results of ANC with LMS on speech EFR of the third subject

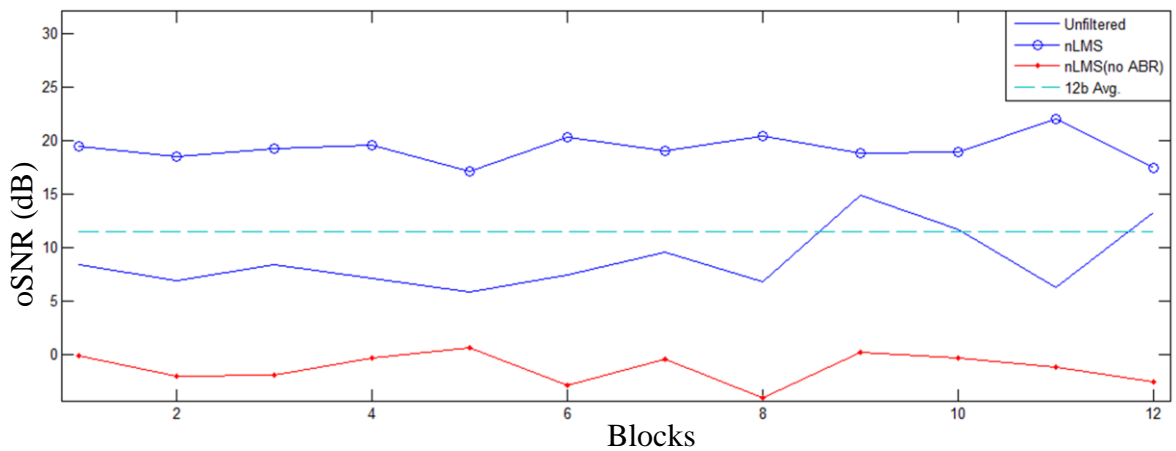


Figure 4-12 Results of ANC with nLMS on speech EFR of the first subject. The solid line represents the unfiltered ABR, the circled line represents the result of the ANC with nLMS algorithm, the dotted line represents the output of the nLMS algorithm with a response that contains no components related to the stimulus, and the dashed horizontal line represents the coherent average over the 12 blocks

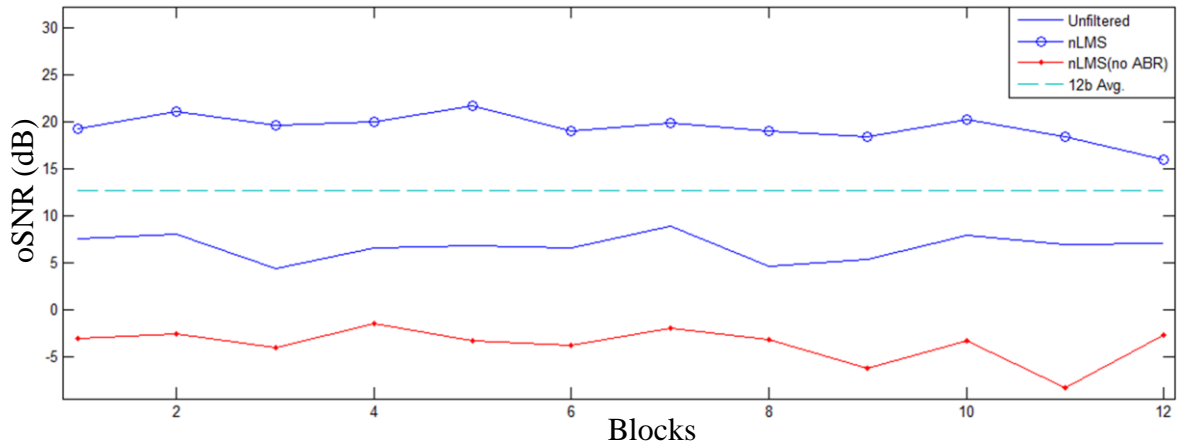


Figure 4-13 Results of ANC with nLMS on speech EFR of the second subject.

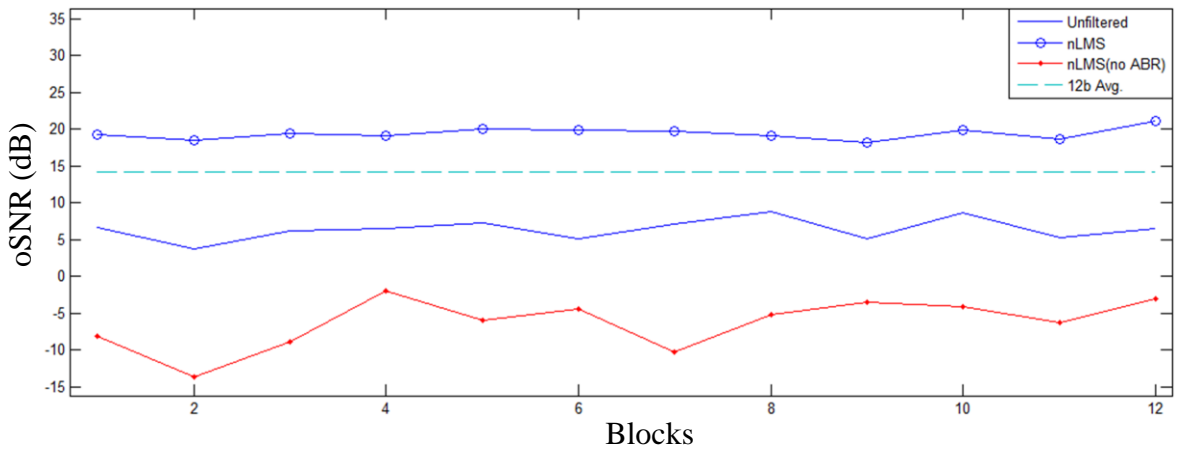


Figure 4-14 Results of ANC with nLMS on speech EFR of the third subject.

4.3 MULTI-ADAPTIVE TECHNIQUE

A higher performance may be achieved if two or more algorithms are applied in cascade, when the two use different reference signals. In this section, two algorithms are applied, the first being SD that has the desired signal as a reference, and the second being ANC with LMS that uses the contaminating noise as a reference. They are applied in this order, so that the filter output can approach the targeted signal, and then the remaining noise is suppressed by the second filter. This multi-adaptive technique is applied on the speech EFR based on the coherent average of 250 individual responses from the three subjects and the results are shown in Figures 4-15, 4-16, and 4-17. As can be seen, this multi-adaptive approach enhances the oSNR by 14.8 dB, on average, which is higher than the enhancement due to each of the two algorithms separately.

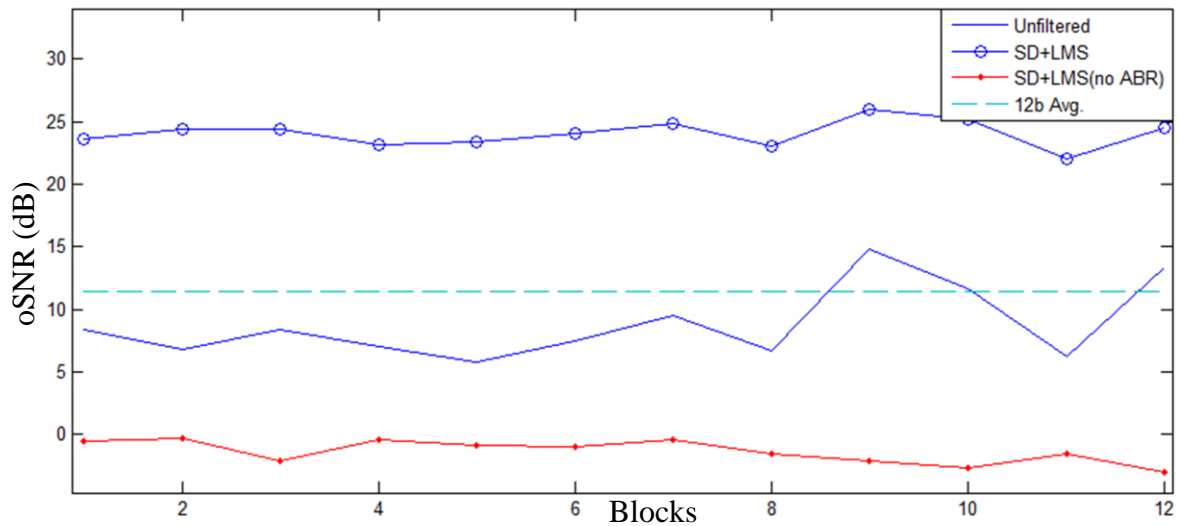


Figure 4-15 Results of SD+LMS on speech EFR of the first subject. The solid line represents the unfiltered ABR, the circled line represents the result of the SD + LMS, the dotted line represents the output of the SD + LMS algorithm with a response that contains no components related to the stimulus, and the dashed horizontal line represents the coherent average over the 12 blocks

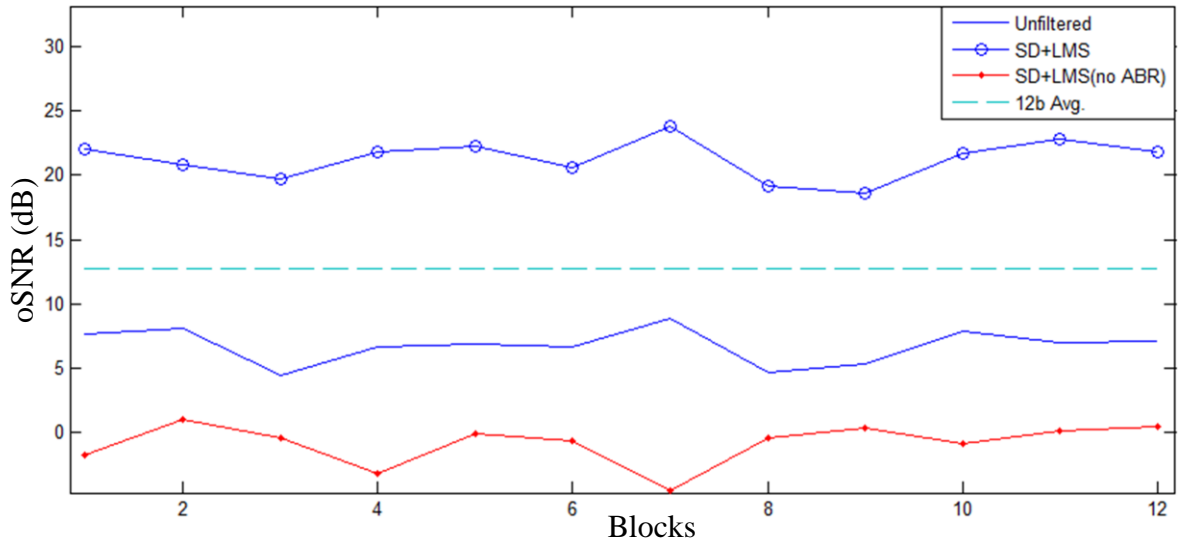


Figure 4-16 Results of SD+LMS on speech EFR of the second subject.

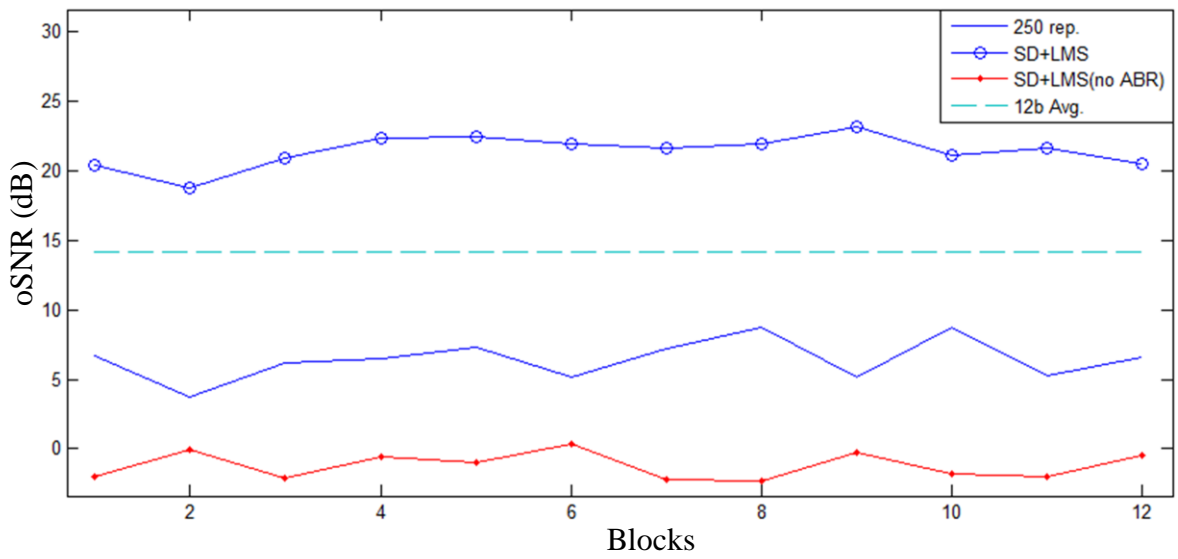


Figure 4-17 Results of SD+LMS on speech EFR of the third subject.

4.4 COMPARISON BETWEEN THE ALGORITHMS

In this section, the performances of the four adaptive filter algorithms and the multi-adaptive technique in enhancing the speech EFR are compared based on several metrics: overall SNR (oSNR), local SNR at 100Hz (lSNR100), MSE between the responses and reference signal estimated by LQ factorization (MSEq), and MSE between the responses and the desired signal obtained by averaging coherently 3000 repetitions over twelve other subjects (MSEd) in both time and frequency domains.

Figure 4-18 shows oSNR performance of all algorithms across the 12 blocks from three subjects (36 blocks in total). As can be seen, the multi-adaptive combination gives the highest oSNR in most blocks, and the oSNR of WF and SD are nearly identical.

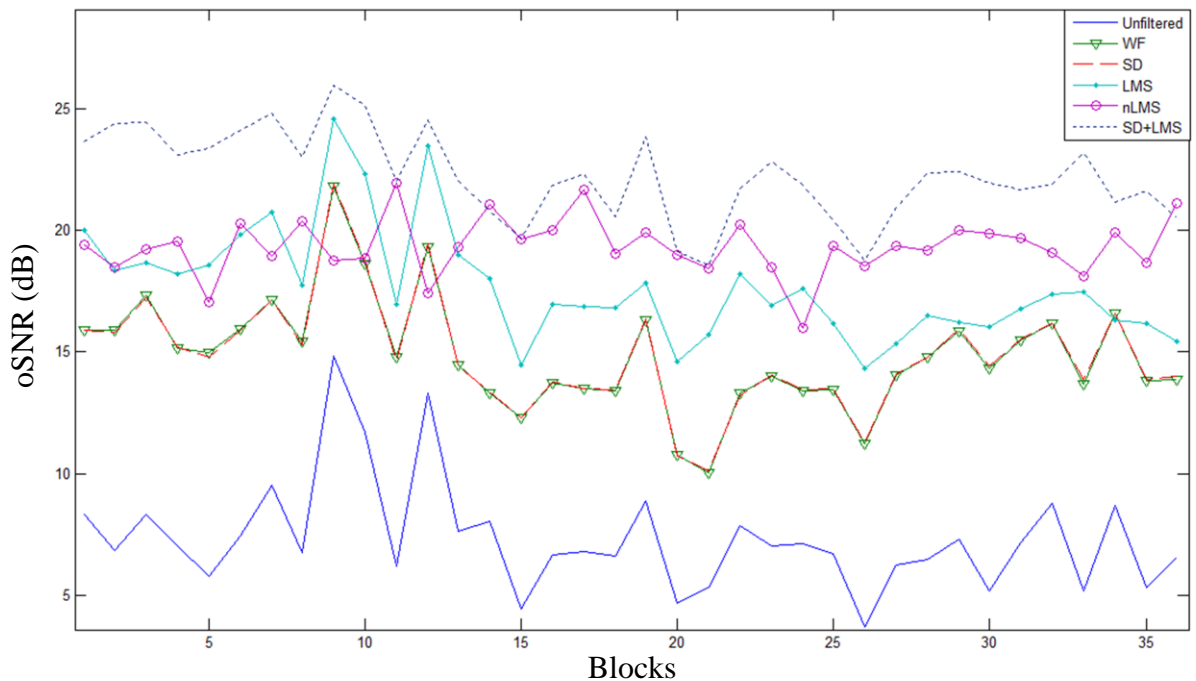


Figure 4-18 The oSNR performance of all algorithms on the speech EFR in all blocks from 3 subjects.

Figure 4-19 shows the local SNR at the fundamental frequency of 100Hz (ISNR100) from the 3 subjects across all 36 blocks. WF enhances ISNR100 by about 1.56 dB, on average, while SD enhances it by about 2.05 dB. ANC with LMS and nLMS enhance ISNR100 by 9.60 and 9.56 dB respectively, SD+LMS obtains the best performance with an average enhancement of 10.51 dB.

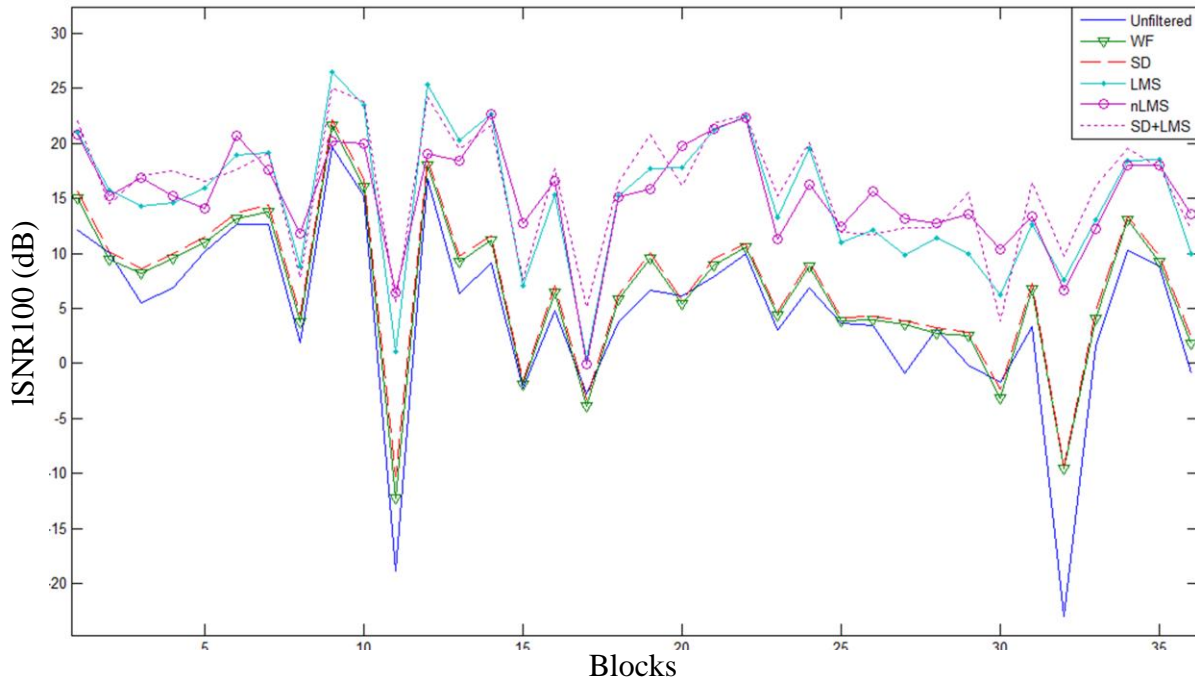


Figure 4-19 Local SNR at 100 Hz (ISNR100) performance of all algorithms on the speech EFR in all blocks from 3 subjects..

Figure 4-20 shows the MSE that is calculated in the time domain using the reference signal that is generated using LQ factorization (MSE_q). In general, WF and SD show better (i.e. lower) MSE_q than LMS and nLMS, and SD+LMS has the lowest MSE_q. Figure 4-21 shows the MSE in the time domain between the responses and the desired signal which is obtained by coherently averaging responses based on 3000 repetitions from 12 other subjects (MSE_d). The MSE_d obtained with SD+LMS is close to that obtained with WF and SD, which indicates that LMS component of SD+LMS only has a small effect on MSE_d. Figures 4-22 and 4-23 show MSE_q and MSE_d obtained by comparing the frequency domain spectra of the signals.

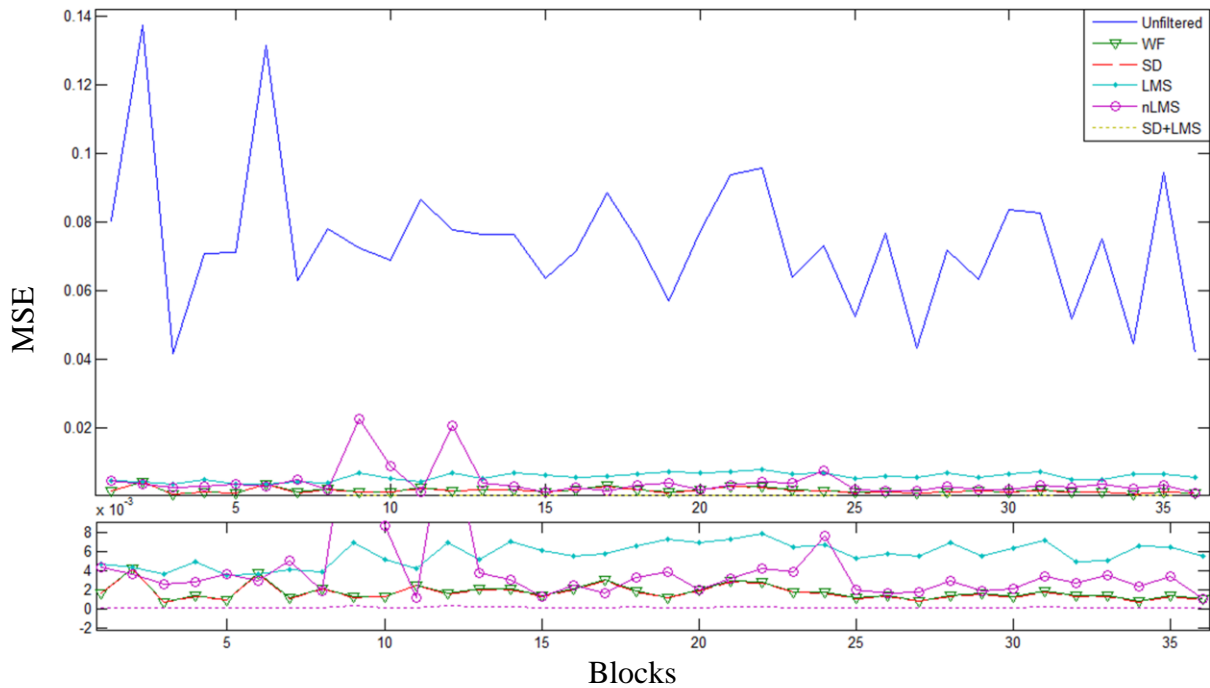


Figure 4-20 Time domain MSE_q performance of all algorithms on the speech EFR in all blocks from 3 subjects. The bottom graph is a magnification of the bottom part of the top graph.

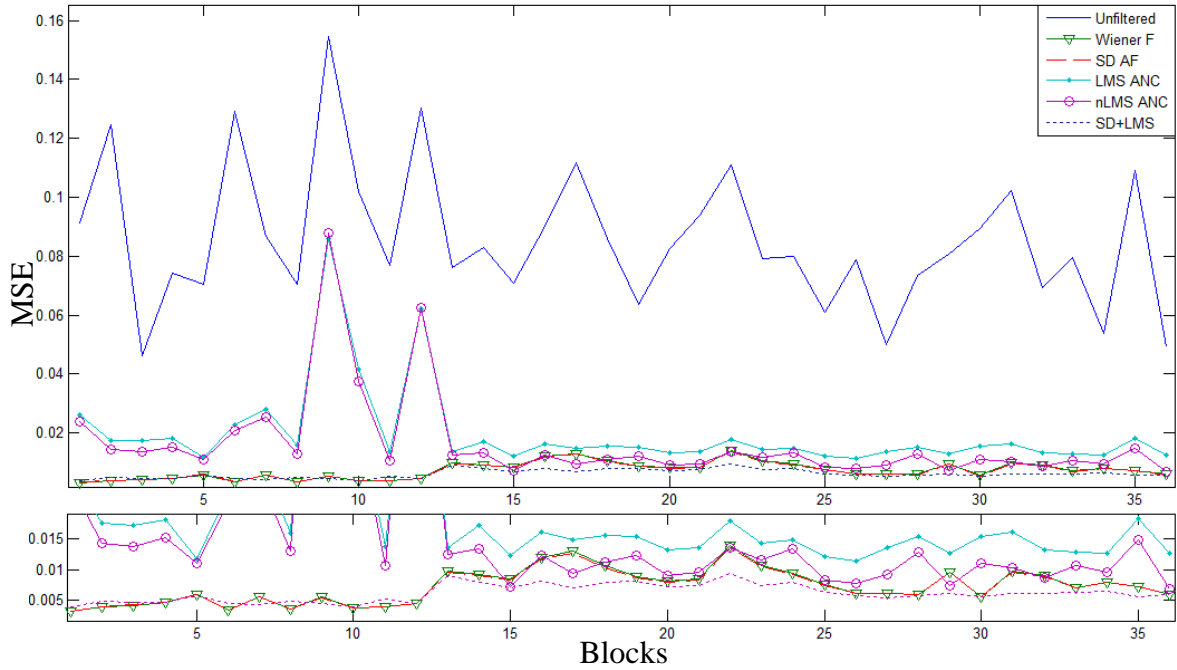


Figure 4-21 Time domain MSE of all algorithms on the speech EFR in all blocks from 3 subjects. The bottom graph is a magnification of the bottom part of the top graph

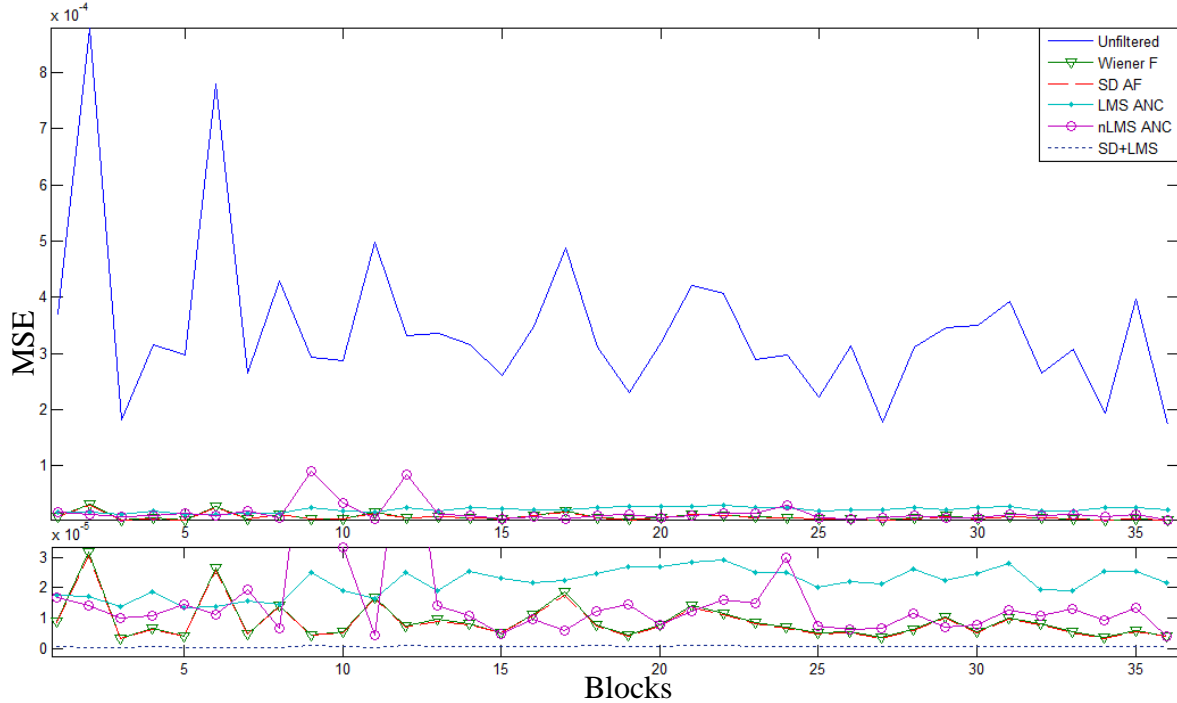


Figure 4-22 Frequency domain MSE (fMSE) performance of all algorithms on the speech EFR in all blocks from 3 subjects. The bottom graph is a magnification of the bottom part of the top graph.

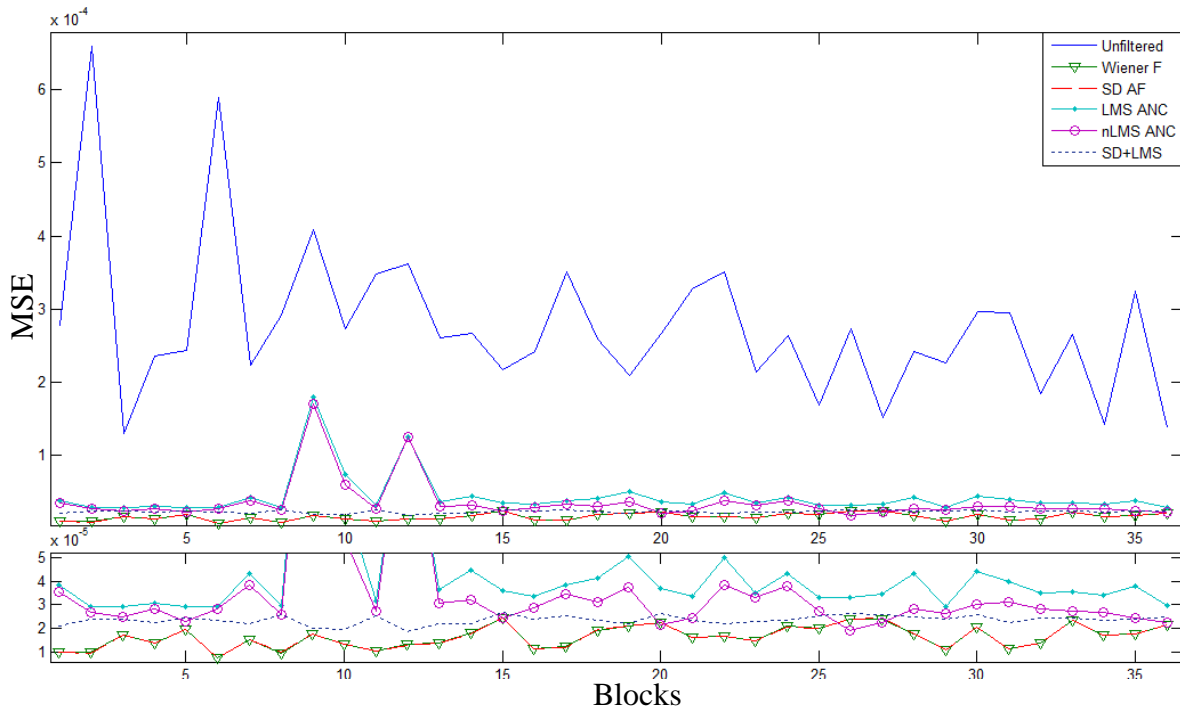


Figure 4-23 Frequency domain MSEd (fMSEd) performance of all algorithms on the speech EFR in all blocks from 3 subjects. The bottom graph is a magnification of the bottom part of the top graph.

Tables 4-1, 4-2, 4-3 and 4-4 summarize the performance of all four algorithms and the multi-adaptive algorithm averaged over the three subjects in Table 4-1 (i.e. 36 blocks) and over the subjects individually in the other tables (i.e. over 12 blocks). The results show that the multi-adaptive technique (SD+LMS) gives the best performance based on all the metrics, except fMSEd.

Table 4-1 Performance of all algorithms with the EFR averaged over 36 individual blocks from the 3 subjects

	Unfiltered	WF	SD	LMS	nLMS	SD+LMS
oSNR (dB)	7.66	14.98	15.00	18.00	19.18	22.43
ISNR₁₀₀ (dB)	7.56	9.12	9.60	17.16	17.11	18.07
MSE_q(e-2)	7.40	0.18	0.17	0.58	0.42	0.02
MSE_d(e-2)	8.56	0.73	0.72	1.96	1.63	0.61
fMSE_q(e-5)	34.43	0.89	0.85	2.17	1.61	0.06
fMSE_d(e-5)	27.75	1.63	1.63	4.38	3.62	2.36

Table 4-2 Performance of all algorithms averaged over 12 individual blocks from subject one

	Unfiltered	12block Avg.	WF	SD	LMS	nLMS	SD+LMS
oSNR (dB)	9.44	11.48	17.28	17.22	20.52	18.94	24.25
ISNR₁₀₀ (dB)	12.49	17.87	14.05	14.59	19.84	18.30	20.22
MSE_q(e-2)	8.16	4.14	0.19	0.19	0.47	0.67	0.02
MSE_d(e-2)	9.64	6.01	0.43	0.43	3.01	2.80	0.46
fmSE_q(e-5)	41.06	20.98	1.12	1.08	1.75	2.64	0.06
fmSE_d(e-5)	33.70	13.91	1.30	1.30	5.59	5.12	2.26

Table 4-3 Performance of all algorithms averaged over 12 individual blocks from subject two

	Unfiltered	12block Avg.	WF	SD	LMS	nLMS	SD+LMS
oSNR (dB)	6.83	12.70	13.24	13.26	17.01	19.33	21 .34
ISNR₁₀₀ (dB)	5.81	13.47	7.25	7.70	17.52	17.68	18.38
MSE_q(e-2)	7.61	1.48	0.21	0.20	0.66	0.34	0.02
MSE_d(e-2)	8.57	2.24	1.02	1.00	1.49	1.13	0.79
fmSE_q(e-5)	33.51	6.05	0.95	0.90	2.47	1 .26	0.07
fmSE_d(e-5)	26.93	4.75	1.75	1.75	3.99	3.12	2.36

Table 4-4 Performance of all algorithms averaged over 12 individual blocks from subject three

	Unfiltered	12block Avg.	WF	SD	LMS	nLMS	SD+LMS
oSNR (dB)	6.40	14.19	14.43	14.50	16.22	19.37	21.47
ISNR₁₀₀ (dB)	2.87	1.38	4.47	4.93	12.34	13.82	13.99
MSE_q(e-2)	6.52	0.89	0.14	0.13	0.60	0.24	0.02
MSE_d(e-2)	7.47	1.68	0.73	0.72	1.39	0.98	0.59
fMSE_q(e-5)	28.72	3.90	0.618	0.58	2.29	0.91	0.00
fMSE_d(e-5)	22.61	2.87	1.84	1.85	3.57	2.61	2.46

4.5 PERFORMANCE OF THE ALGORITHMS WITH FREQUENCY FOLLOWING RESPONSES (FFR)

As mentioned before, there are two possible configurations of the obtained speech ABR signals, the Envelope Following Response (EFR) and the Frequency Following Response (FFR). The EFR emphasizes components of the response that follow the fundamental frequency of the envelope (100 Hz) and its harmonics, while the FFR emphasizes components of the response that directly follow the stimulus harmonics and in particular the components related to the first formant at 700 Hz. The performance of the algorithms with the EFR was described in the previous sections. This section describes the performance with the FFR.

As an illustration of the performance of the different algorithms, the following two figures (Figures 4-24 and 4-25) show one block of speech ABR (FFR) from one subject which is based on the coherent average of 250 individual responses. The figures show the FFR in the time and frequency domains before and after filtering with all the adaptive filters. The coherent average over twelve blocks is also included.

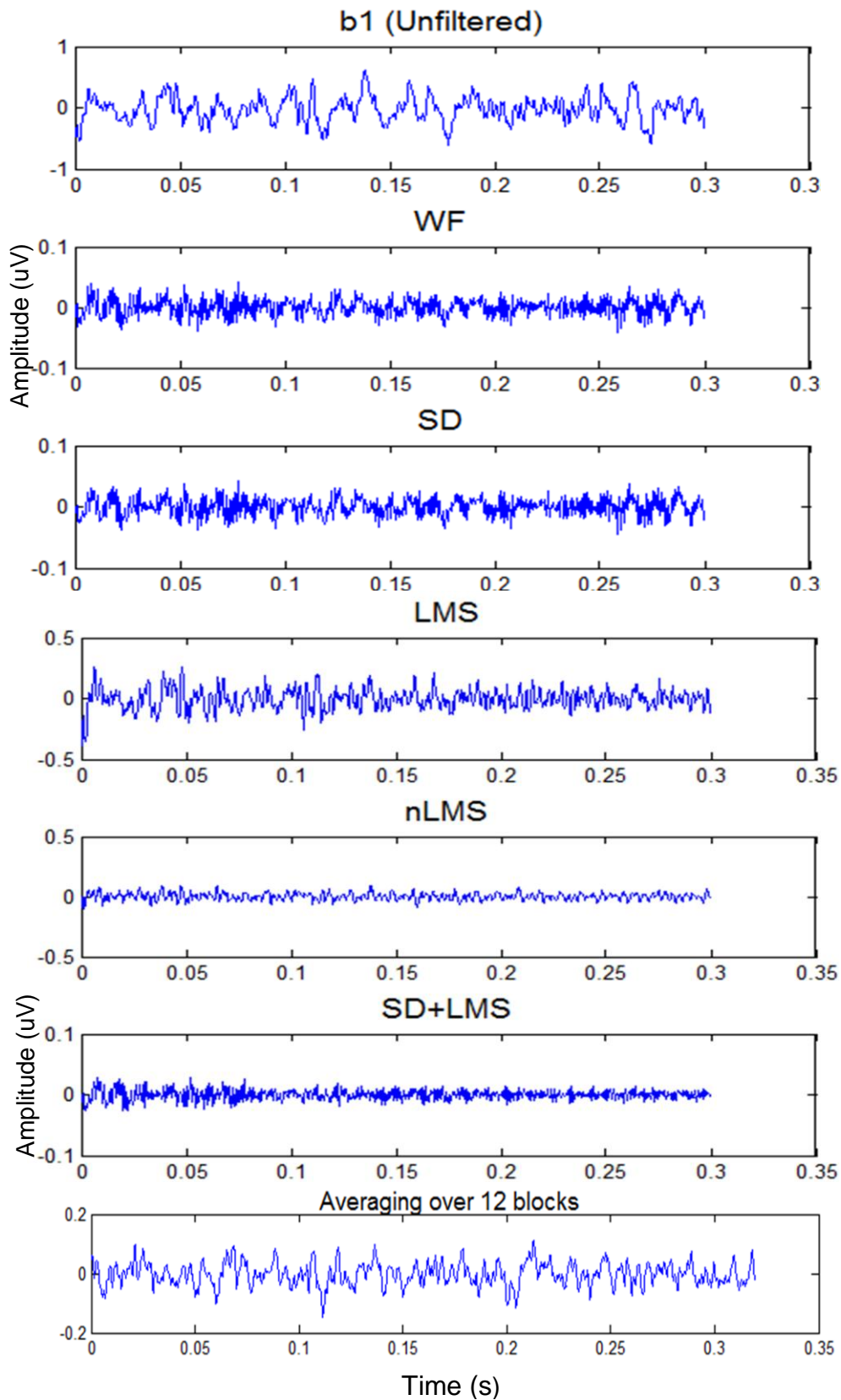


Figure 4-24 Illustration of time domain performance of all the adaptive filters with the Frequency Following Response (FFR). The sample response based on the coherent average of 250 repetitions (block 1 from subject 1) is shown before and after filtering. The coherent average over twelve blocks is also shown.

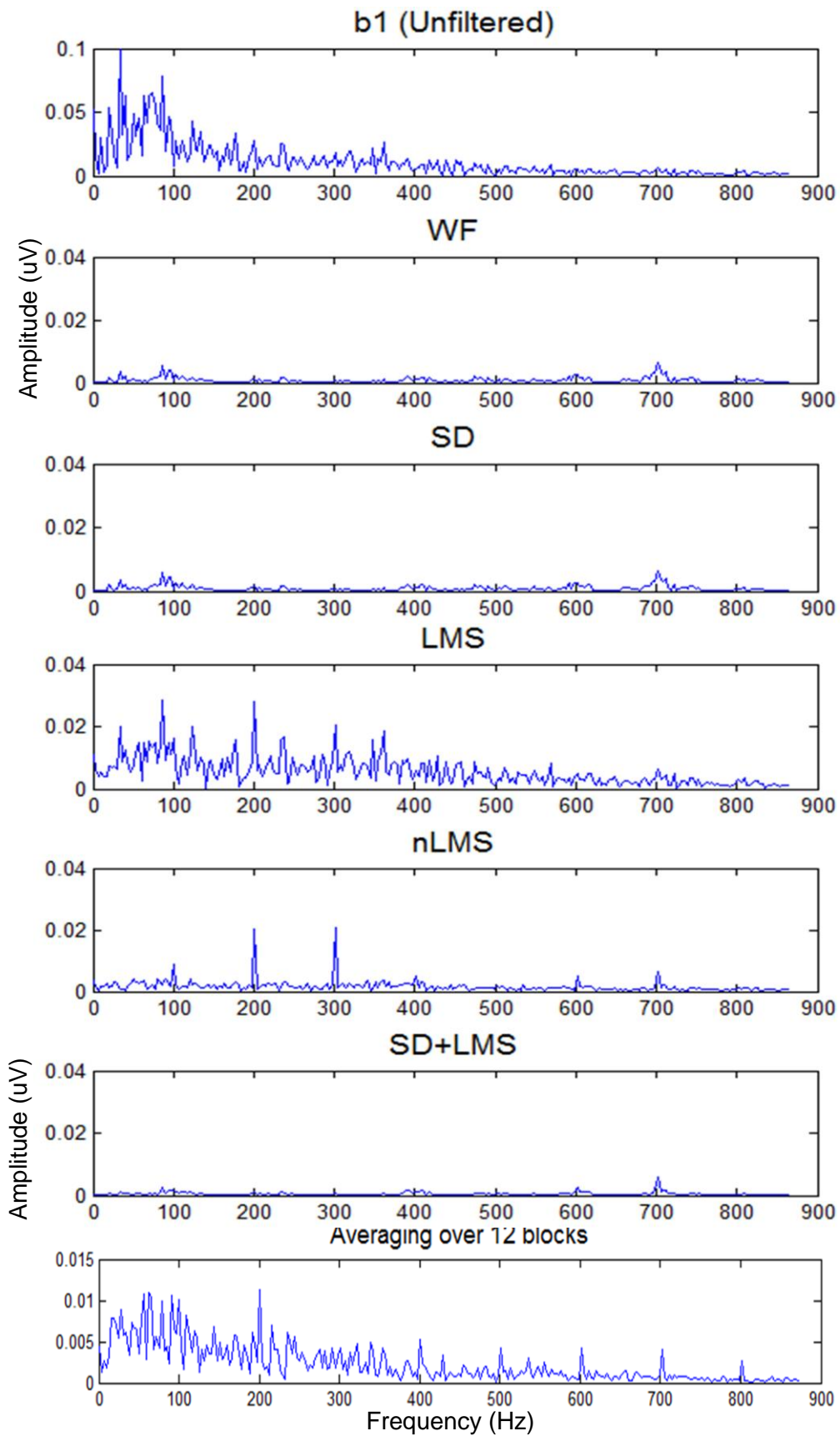


Figure 4-25 Illustration of frequency domain performance of all the adaptive filters with the Frequency Following Response (FFR). The sample response based on the coherent average of 250 repetitions (block 1 from subject 1) is shown before and after filtering. The coherent average over twelve blocks is also shown.

Figure 4-26 shows the oSNR of the FFR of the 36 blocks from the 3 subjects before and after filtering. Before filtering, the oSNR of the FFR is lower than that for the EFR (shown in Figure 4-18). Among the used algorithms, the performance of the nLMS is the highest with an average enhancement of 18.4 dB relative to the unfiltered signal. Figure 4-27 shows the local SNR at the first formant of 700Hz (ISNR700) that is obtained with all the algorithms. Table 4-5 summarizes the performance of all algorithms averaged over the 36 blocks of FFR from the three subjects. The results show that nLMS provides the highest oSNR, while SD+LMS gives the highest ISNR700. In terms of MSE, WF gives the best MSE_q and fMSE_q, while SD+LMS gives the best MSE_d.

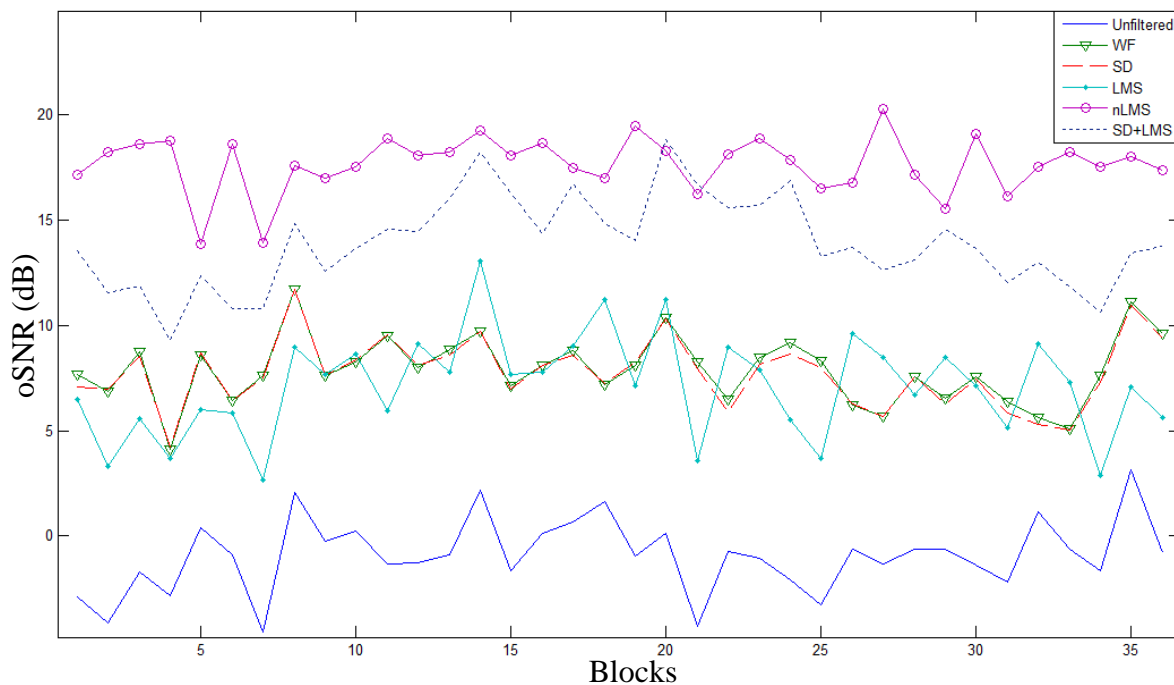


Figure 4-26 The overall SNR (oSNR) performance of all algorithms on the speech FFR in all blocks from 3 subjects

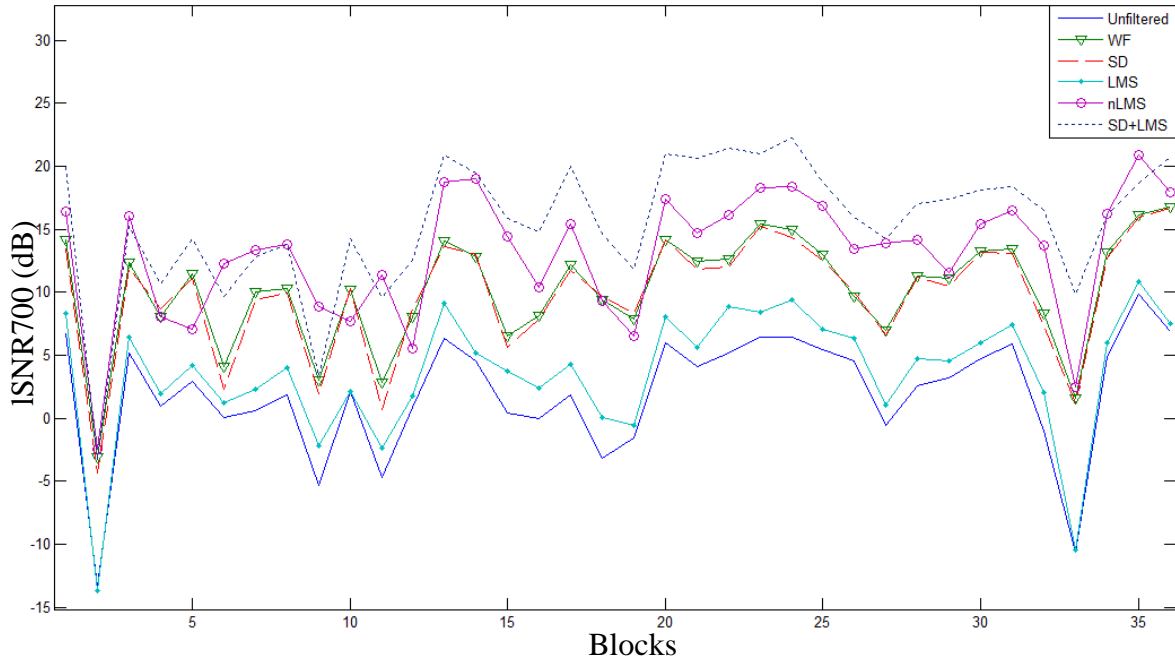


Figure 4-27 The local SNR at 700 Hz (ISNR700) performance of all algorithms on the speech FFR in all blocks from 3 subjects

Table 4-5 Performance of all algorithms with the FFR averaged over 36 blocks from the 3 subjects

	Unfiltered	WF	SD	LMS	nLMS	SD+LMS
oSNR (dB)	-0.71	7.94	7.79	7.53	17.69	14.36
ISNR₇₀₀ (dB)	3.11	11.31	11.00	4.98	14.19	17.31
MSE_q	0.05	2.44e-12	3.42e-6	0.01	5.78e-4	6.37e-7
MSE_d	0.05	1.54e-4	1.54e-4	0.01	1.4e-3	1.51e-4
fMSE_q	2.08e-4	9.21e-15	1.30e-8	2.61e-5	2.18e-6	2.25e-9
fMSE_d	2.08e-4	6.43e-7	5.67e-7	2.75e-5	3.85e-6	5.79e-7

CHAPTER 5: DISCUSSION

This chapter discusses the results presented in the previous chapter along with the comparison of the performance of the applied adaptive algorithms and the traditional method of coherent averaging. Theoretically, coherent averaging should reduce the noise power (or variance, σ_v^2) according to the following equation (Ozdamar and Delgado, 1996):

$$\text{decrease in noise power} = \frac{\sigma_v^2}{N}$$

where N is the number of averages.

As a consequence, coherent averaging will theoretically augment the SNR (in dB) by $10\log_{10}(N)$. However, the obtained results show that there is difference between obtained and theoretical SNR improvement. The expected increase of oSNR due to coherent averaging of 12 blocks is 10.8 dB, whereas the obtained improvement in oSNR is 5.5 dB (the average over 3 subjects of the increase of oSNR due to coherent averaging of 12 blocks in Tables 4-2, 4-3, and 4-4). Such a difference between the experimental and theoretical values could be due to one of three reasons:

1. The stimulus-related component is not identical in the averaged response blocks.
2. The noise signal is correlated in the averaged blocks.
3. The noise signal is not stationary.

All the above reasons can contribute to the difference between experimental and theoretical gain in SNR, and in the literature there are some studies that attempt to reduce the effect of these factors. For instance, an approach that has been proposed is the ongoing-computational-averaging technique which verifies that the ABR signals meet the required conditions for coherent averaging before averaging them (Ozdamar and Delgado, 1996).

The overall SNR (oSNR) obtained with the adaptive algorithms applied to the Envelope Following Response (EFR) are close to what is expected. WF and SD are expected to increase the SNR by 7.45 dB, by using equation 3.3.1 on the signal and noise configuration discussed in same section, and they increase it by 7.34 dB (the difference between Unfiltered and WF in Table 4-1). Adaptive Noise Cancellation with LMS and nLMS are expected to increase oSNR by 9.6 dB for both (using equation 3.3.1 as described above), and they increase it by 10.3 and 11.5 dB respectively (the difference between Unfiltered and LMS and nLMS in Table 4-1). The multi-adaptive technique (SD+LMS) is expected to increase oSNR by 14.2 dB, using equation 3.3.1, and it increases it by 14.8 dB (the difference between Unfiltered and SD+LMS).

Comparing the experimental results against coherent averaging, the performance of WF and SD is equivalent to averaging 5 blocks, based on a theoretical increase in SNR with coherent averaging equal to $10\log_{10}(N)$. LMS and nLMS are equivalent to coherently averaging 10 and 14 blocks, respectively, and SD+LMS is equivalent to coherently averaging 30 blocks. However, as noted earlier, in practice coherent averaging did not improve the SNR as much as is expected theoretically. Therefore, in reality, the improvement of oSNR due to the adaptive filters would be expected to be equivalent to an even greater number of coherent averages than indicated here.

The theoretical MSEd is calculated using equation 3.3.2, and it is expected to be 0.0056 for WF and SD, while the obtained result is 0.0073. The MSEd for ANC with LMS and nLMS are expected to be 0.025 and 0.016 respectively (since in this case MSE is equal to $\sigma_d^2 + \mathbf{w}^T R_y \mathbf{w} - 2 \mathbf{w}^T \mathbf{p}$), and the obtained values are 0.020 and 0.016. The multi-adaptive technique with SD+LMS is expected to give MSEd of 0.0033, while the obtained value is 0.0061. Table 5-1 summarizes the theoretical and measured performances with the EFR for all the algorithms.

Table 5-1 Theoretical and measured performances with the EFR signals

	Theoretical increase of oSNR (dB)	Measured increase of oSNR (dB)	Theoretical MSEd (e-2)	Measured MSEd (e-2)	Equivalent averaged blocks (Theoretical)
WF and SD	7.45	7.34	0.56	0.73	5
LMS	9.6	10.3	2.5	2.0	10
nLMS	9.6	11.5	1.6	1.6	14
SD+LMS	14.2	14.8	0.33	0.61	30

The local SNR at F0 (ISNR100) and the other MSE metrics show similar improvements using the adaptive filters, with the multi-adaptive technique usually giving the best performance. The ISNR100 gives useful information about processing of the fundamental frequency of speech, which perceptually corresponds to the pitch (Laroche et al., in press; Prevost et al., in press) while MSE metrics are useful for assessing adaptive filters performance, especially if the filters have been designed based on MSE criterion.

The oSNR and MSEd performance of the adaptive algorithms applied to the Frequency Following Response (FFR) are also close to what is expected, although with some of the metrics there is more discrepancy between the theoretical and obtained responses than with the EFR. This difference is probably due to the higher variance in the autocorrelation of the ABR signals that are obtained by FFR which only have one component of interest (at 700 Hz) and have a lower amplitude than EFR. WF and SD are expected to enhance the oSNR by 10.59 dB, and the enhancement is measured to be 8.67 dB. This enhancement is equivalent to averaging 7 blocks coherently. Their MSEd is expected to be 0.0017 and it is measured at 0.0015. ANC with LMS and nLMS are expected to enhance the oSNR 9.06 dB and 16.41 dB, respectively, and the enhancement is measured to be 8.24 and 18.39 dB. The enhancements they achieved are equivalent to coherently averaging 6 and 69 blocks, respectively. Their MSEd's are expected to be 0.0073 and 0.0020, respectively, whereas they were measured at 0.074 and 0.0014. The multi-adaptive technique with SD+LMS is expected to give an oSNR enhancement of

10.49 dB, while it is measured to be 15.1 dB. This enhancement is equivalent of averaging 32 blocks coherently. Its MSEd is expected to be 0.0017, and it is measured at 0.0015. Table 5-2 summarizes the theoretical and measured performances with the FFR for all the algorithms.

Table 5-2 The theoretical and measured performances with the FFR signals

	Theoretical increase of oSNR (dB)	Measured increase of oSNR (dB)	Calculated MSEd (e-2)	Measured MSEd (e-2)	Equivalent averaged blocks (Theoretical)
WF and SD	10.59	8.6	0.17	0.15	7
LMS	9.06	8.2	0.73	0.74	6
nLMS	16.41	18.4	0.20	0.14	69
SD+LMS	10.49	15.1	0.17	0.15	32

As the results show, adaptive filters are a potentially useful tool that can be used to enhance speech ABR signals by using a reference signal, which can be either a desired ground signal or a signal that is correlated with the background noise. Two algorithms are used (WF and SD) to enhance the speech ABR signals by using a desired ground signal. Two other ANC algorithms use a noise reference generated using the LQ factorization method. A combination of these two types of filters (SD+LMS) proved capable of enhancing the individual speech ABR response blocks more than can be achieved by each of the individual algorithms alone.

The results of applying the adaptive algorithms on the speech ABR show significant enhancement of the signal. This would potentially correspond to significant reduction in recording time compared to the traditional method of coherently averaging acquired responses. For example, as seen in Table 5-1, the recording time may be reduced by a factor of up to 30 times if the multi-adaptive technique is used and overall SNR is used as the performance metric. As mentioned above, the actual reduction in recording time can probably be higher since the practical performance of coherent averaging does not meet

the theoretical expectation. This advantage could be very helpful particularly in applications where recording time is critical, such as in the case of assessment of the auditory function in infants or children who would find it difficult to stay still for the duration of the recording.

The used adaptive filters can prevent false positives by using a threshold condition such as an autocorrelation-based criterion. Table 3-3 shows how the used algorithms suppress both the noise in the signal with no component related to stimulus (no-ABR) and false positives with four different signals (white noise, AR, EEGn1, and EEGn2). The WF and SD are the best algorithms that can prevent a false positive with no-ABR. The LMS is the best algorithm that can prevent a false positive with white noise and EEGn2. The nLMS is the best algorithm that can prevent false positives with white noise, AR, and EEGn1. This analysis was performed in a limited number of samples, so additional samples of EEG noise in particular are needed to confirm that the adaptive filters with pre-processing using a threshold condition do not produce a large number of false positives. As indicated earlier, a number of pitch detection algorithms that are available in the literature could be used in the pre-processing signal detection stage.

Other potential limitations of the study reported in this thesis should be mentioned. Collecting data from a larger pool of subjects would be desirable to confirm the advantages of using adaptive filters. Another limitation is that testing was done on normal hearing subjects. More testing on subjects with various types of hearing impairment is needed to determine if the advantage of the adaptive filter is retained. This should be done before these techniques can be applied to a clinical environment.

CHAPTER 6: CONCLUSION AND FUTURE WORK

6.1 CONCLUSION

This thesis uses four different adaptive filter algorithms, namely the Wiener Filter (WF), Steepest Descent (SD), Adaptive Noise Cancellation (ANC) based on Least-Mean-Square error (LMS) and normalized LMS error (nLMS), and a multi-adaptive algorithm based on the cascade combination of SD and LMS (SD+LMS) to enhance the recorded speech ABR signals. Thirty-six blocks of speech ABR were recorded from three subjects using a synthetic vowel stimulus /a/ (each response is based on the coherent average of 250 repetitions), and fifteen other blocks were recorded with no stimulus from a fourth subject (each based on 20 repetitions). Thirty-six other blocks with no components related to stimulus (no-ABR) were created by applying LQ factorization on the recorded speech ABRs. The recorded speech ABR was analysed in two configurations, the Envelope Following Response (EFR) and the Frequency Following Response (FFR). Finally, a ground signal of 3000 repetitions of speech ABR averaged coherently from twelve other subjects (for a total of 36,000 repetitions) was used as a desired reference signal. WF and SD algorithms were configured and fixed according to the response signal that had an autocorrelation that was closest to the mean autocorrelation (based on MSE criterion) of the 36 blocks and the desired signal. The ANC was configured to suppress only the background noise components estimated using LQ factorization.

Pre-processing using a threshold condition based on autocorrelation was used to detect responses. Four different noise samples (autoregressive noise (AR), white noise, EEG noise with a strong component at 60 Hz, and EEG with Silence stimulus) were used to check if the configured algorithms produced a false positive when there was no stimulus.

All algorithms successfully enhanced both EFR and FFR speech ABR with significantly less time needed than the coherent averaging method based on four metrics used to assess the performance (σ SNR, ISNR, MSR_d, and MSE_q). In addition, the algorithms appear to have the ability to suppress false positives when noise samples are used as inputs.

6.2 THESIS CONTRIBUTIONS

The contributions of the thesis are summarized as follows:

- This thesis investigated the implementation of several types of adaptive filters to enhance speech ABR signals, using two types of reference signals, a desired ground signal and a noise sample correlated with the background noise. The thesis also proposed and tested a multi-adaptive technique that uses the two types of reference signals in cascade.
- This thesis demonstrates that these adaptive filters can significantly enhance the speech ABR signal, with the performance assessed using several metrics. The multi-adaptive technique gave the best performance on most of the metrics.
- This thesis demonstrates that the gains in SNR using the adaptive filters can potentially result in a reduction in response recording time of at least one order of magnitude, and possibly more, compared to the currently used method of coherent averaging.
- Portions of the research have been disseminated in: Fallatah A and Dajani HR, 2012. "Adaptive filters for enhancing auditory brainstem responses to synthetic vowel stimuli." poster presented at the *International Conference on Adult Hearing Screening*, Cernobbio, Italy, June 2012.

6.3 FUTURE WORK

Since this thesis only focused on subjects with normal hearing and used a single synthetic vowel stimulus, testing these algorithms on various types of hearing impairments and other speech samples including natural speech stimuli is needed to demonstrate the performance of adaptive filters in these different conditions. Moreover, the used

algorithms are based on basic principles of adaptive filters, such as the use of an FIR structure. Future work can include investigating more advanced algorithms with IIR, and lattice structures. For example, linear prediction theory provides several algorithms that have the ability to estimate low power components of interest in background noise using IIR and lattice filters (Madhavan, 1989). In addition, multi-channel adaptive filters are a promising technique that can be used to process more than one block of ABR at the same time. Therefore, an extension of the adaptive filter algorithms can include the multi-channel algorithms that have been described in the literature (Westerkamp and Slifka, 1989; Van Dun et al., 2007). In addition, with the ANC algorithms, EEG noise samples will be collected from each subject interleaved with the collection of the speech ABR responses. Because these noise samples should be highly correlated with the noise contaminating the measured response signal in each subject, the ANC algorithms may achieve better performance.

APPENDIX A

This appendix discusses two topics related to recording the speech ABR from the three subjects, namely the calibration of sound stimulus level, and the used software. All data collected was performed by Amir Sadeghian³, except for the desired signal which was collected by Marilyn Laroche (Laroche, 2010).

Calibration of sound stimulus:

The stimulus levels were calibrated at 80.5 dB SPL with the earphone connected to a 2cc coupler attached to a Brüel&Kjaer Artificial Ear type 4152, and a Sound Level Meter (SLM) Type 2230.

The Used software:

Synthesizer software based on Klatt (1980) and written by M. Laroche (2010):

- Used for generating the vowel stimulus.

Matlab® version 7.6.0 (The MathWorks, Inc.) for:

- Plotting all time domain waveforms and frequency spectra.
- Designing all used adaptive filters (except ANC which is based on a built-in function) and implementing all performance measurements.
- Applying LQ factorization.

³ A. Sadeghian is an M.A.Sc student with the Department of Systems and Computer Engineering, Carleton University, Ottawa, ON K1S 5B6, Canada

APPENDIX B

LQ-factorization

LQ stands for Lower triangular (L) and orthogonal (Q) matrix. This method of matrix factorization is able to decompose a given vector (\mathbf{v}) into signal-plus-noise form ($\mathbf{s}+\mathbf{n}$); (Van Dun et al., 2007; Sayed, 2011).

A vector $\mathbf{v}=[v_0, v_1, \dots, v_k]$ which contains signal components at certain frequencies $\mathbf{S}=[\mathbf{s}_0, \mathbf{s}_1, \dots, \mathbf{s}_{m \times 2}]$ and noise components can be decomposed into an L matrix and \mathbf{q} vector as follows:

$$\begin{bmatrix} \mathbf{S} \\ \mathbf{v} \end{bmatrix} = L \mathbf{q}$$

where L is a lower triangular $l \times l$ matrix in the following form:

$$L = \begin{bmatrix} d_{0,0} & 0 & 0 & 0 \\ d_{1,0} & d_{1,1} & 0 & 0 \\ \vdots & \vdots & \ddots & 0 \\ d_{l,0} & d_{l,1} & \dots & d_{l,l} \end{bmatrix}$$

and \mathbf{q} is orthogonal $l \times 1$ vector, so that:

$$\mathbf{q}^T \mathbf{q} = I = 1$$

The noise vector \mathbf{n} is:

$$\mathbf{n} = L' \mathbf{q}'$$

where L' is a matrix taken from the lower triangular L , and it is equal to:

$$L' = \begin{bmatrix} d_{m+1,m+1} & 0 & 0 & 0 \\ d_{m+2,m+1} & d_{m+2,m+2} & 0 & 0 \\ \vdots & \vdots & \ddots & 0 \\ d_{l,m+1} & d_{l,1} & \dots & d_{l,l} \end{bmatrix}$$

where m is the number of signal components multiplied by two, and \mathbf{q}' is equal to:

$$\mathbf{q}' = \mathbf{q}(m+1:l)$$

It should be noted that the LQ factorization method assumes that the matrix $\begin{bmatrix} \mathbf{S} \\ \mathbf{v} \end{bmatrix}$ is invertible (Sayed, 2011).

APPENDIX C

Calculating Autocorrelation Matrix:

Autocorrelation is a statistical tool that measures the similarity between observations as a function of the time lag between them. It is used to find any pattern of periodicity in a signal. Usually, it is written as a toeplitz matrix (Vaidyanathan, 2007).

For a given observation $x(i)$, we can calculate its autocorrelation matrix (R) as follows:

$$R(\tau) = \frac{E[(x_t - \mu)(x_{t+\tau} - \mu)]}{\sigma^2}$$

where μ is the mean of the observation and τ is the time lag. R can also be estimated using a matrix method. For example, if we want to estimate the (3x3) autocorrelation (R) of a (4x1) observation vector (x), we first have to create the matrix X as follows:

$$X = \begin{bmatrix} x_0 & 0 & 0 \\ x_1 & x_0 & 0 \\ x_2 & x_1 & x_0 \\ x_3 & x_2 & x_1 \\ 0 & x_3 & x_2 \\ 0 & 0 & x_3 \end{bmatrix};$$

Then we can easily calculate R as follows:

$$R = (X^T X)$$

Selecting R for WF and SD:

Figure C-1 shows the autocorrelation of block one (b1) from subject 1 that is used in designing the WF and SD algorithms for the speech ABR blocks that are based on the coherent averaging of 250 individual responses. It was selected since it has the minimum MSE to the average of the 36 autocorrelations (C Avg.) seen in this figure.

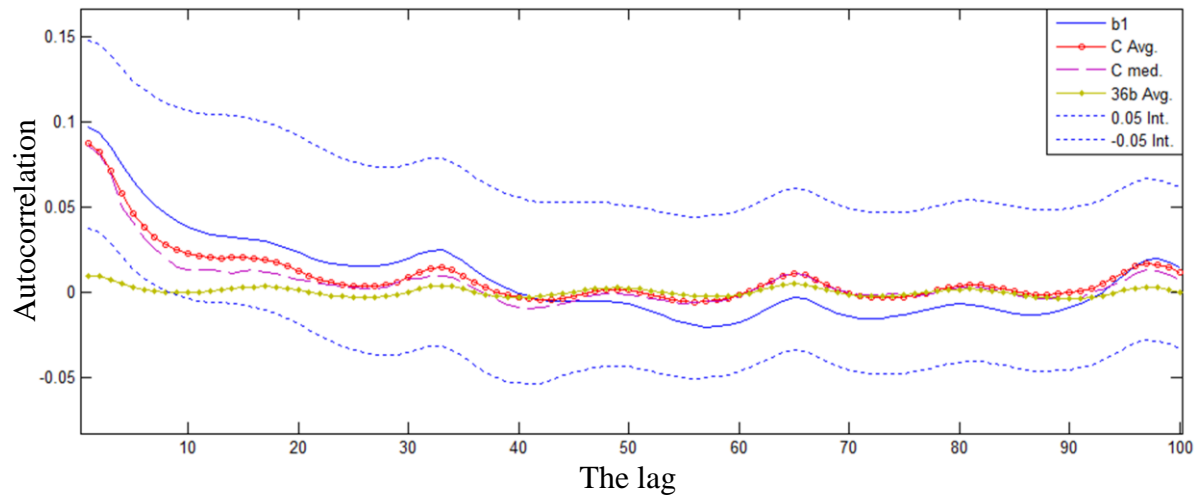


Figure C-1 .This figure shows the autocorrelation for block 1 from the first subject (solid line), the ± 0.05 interval (dots), the 36 blocks coherently averaged (dotted line), the average of all autocorrelations (red circled line), and the median of all autocorrelations (striped line)

The selected autocorrelation threshold:

The gathered 36 blocks from the three subjects share the following features about their autocorrelation vector (extracted from any column from the matrix R):

- 1) They lie in a narrow interval (mostly within ± 0.053 as seen in Figure C-2).
- 2) They have six peaks located at a period of 16, which represents the stimulus fundamental frequency, $F_0 \times \text{signal duration}/2$ ($100 \times 0.32/2 = 16$).

These two features have been used as a threshold. If the autocorrelation of the input signal lies in the interval and follows the peak pattern (at least three peaks at a period of 16), the signal will be considered as ABR signal. Otherwise the signal will be considered as noise.

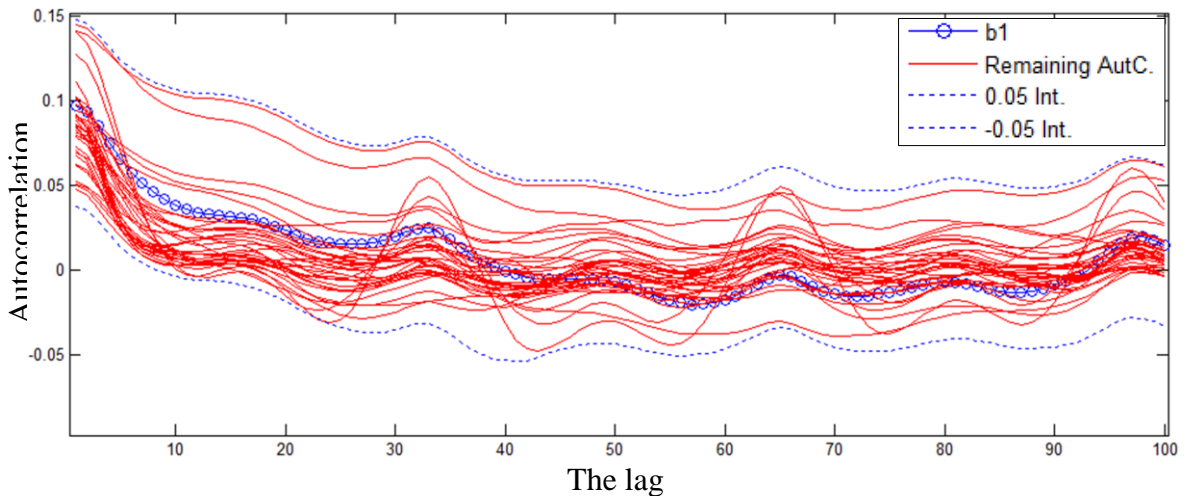


Figure C-2 Autocorrelations of the 36 speech ABR blocks lie in ± 0.05 interval. In addition, all have the same peak pattern that repeats with a period of 16.

Figure C-3 shows the autocorrelations of the four noise signals that have been used in section 3.4. These autocorrelations of the noise signals lie beyond the ± 0.05 interval, and they have different peak patterns than the speech ABR signals.

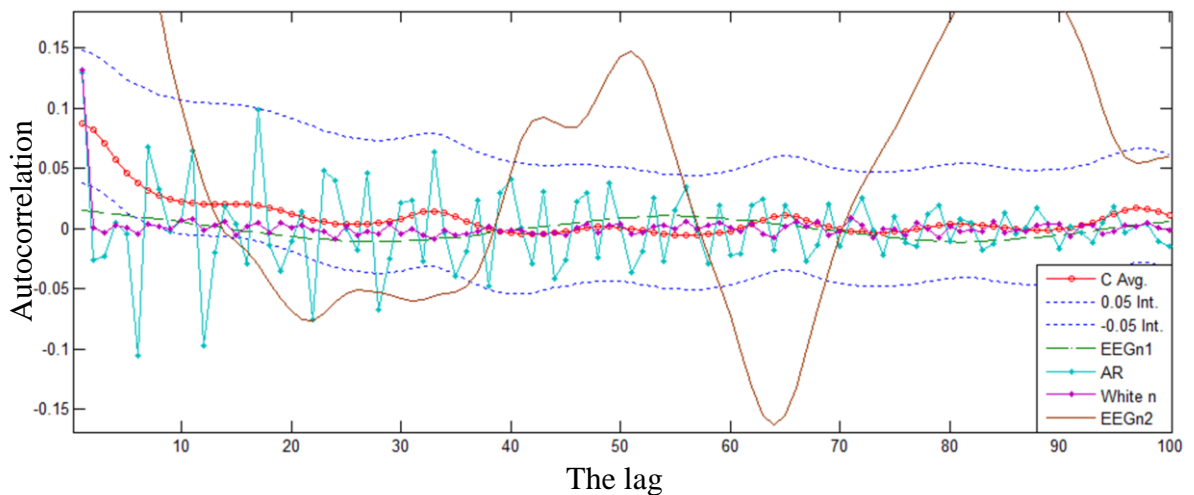


Figure C-3 Autocorrelations of the four noise signals AR (dotted green line), white noise (dotted purple line), EEGn1 (long striped line) and EEGn2 (solid line). The ± 0.05 interval and the average of the 36 autocorrelations are shown.

References

Aiken, S.J. and T.W. Picton. 2006. "Envelope following responses to natural vowels." *Audiol. Neurootol.* 11(4):213-232.

Aiken, S. J. and T. W. Picton. 2008. "Envelope and spectral frequency-following responses to vowel sounds." *Hear. Res.* 245(1-2):35-47.

Atkinson, I.A., A.M. Kondo, and B.G. Evans. 1995. "Pitch detection of speech signals using segmented autocorrelation." *Electronics Letters* 31(7):533-535.

Benesty, J. and J. Chen 2011. "Signal-Channel Noise Reduction with a Filtering Vector." pp. 3-22 in "*Optimal Time-Domain Reduction Filters: A Theoretical Study*". University of Quebec: Springer.

Bidelman, G. M. and A. Krishnan. 2010. "Effects of reverberation on brainstem representation of speech in musicians and non-musicians." *Brain Res.* 1355:112-125.

Boston, J. R. and A. R. Moller. 1985. "Brainstem auditory-evoked potentials." *Crit. Rev. Biomed. Eng.* 13(2):97-123.

Cheah, L. A. and M. Hou. 2010. "Real-time detection of auditory steady-state responses." *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2010:1382-1385.

Dajani, H. R., D. Purcell, W. Wong, H. Kunov, and T. W. Picton. 2005. "Recording human evoked potentials that follow the pitch contour of a natural vowel." *IEEE Trans. Biomed. Eng.* 52(9):1614-1618.

Fallatah A and Dajani HR, 2012. "Adaptive filters for enhancing auditory brainstem responses to synthetic vowel stimuli." *poster presented at the International Conference on Adult Hearing Screening*, Cernobbio, Italy, June 2012.

Fan, Y., L. Ming-hui, X. Sun-hua, and P. Guo-feng. 2010. "Research on a new method of preprocessing and speech synthesis pitch detection." *Computer Design and Applications (ICCCA), IEEE International Conference* 1:399-401.

- Farhang-Boroujeny, B. 1998. "Linear Filters." pp. 1-27 in *Adaptive Filters: Theory & Applications*. Chichester, England: John Wiley & Sons.
- Fridman, J., E. R. John, M. Bergelson, J. B. Kaiser, and H. W. Baird. 1982. "Application of digital filtering and automatic peak detection to brain stem auditory evoked potential." *Electroencephalogr. Clin. Neurophysiol.* 53(4):405-416.
- Greenberg, S. and W. A. Ainsworth 2004. "Speech processing in the auditory system." pp. 1-63 in *Chapter 1:Speech Processing in The Auditory System: An Overview*, edited by Popper A.N. and Fay R.R. 1st ed. New York: Springer.
- Jacquin, A., E. Causevic, R. John, and J. Kovacevic. 2005. "Adaptive complex wavelet-based filtering of EEG for extraction of evoked potential responses." *Acoustics, Speech, and Signal Processing, IEEE International Conference* 5:393-396.
- John, M. S., A. Dimitrijevic, and T. W. Picton. 2001. "Weighted averaging of steady-state responses." *Clin. Neurophysiol.* 112(3):555-562.
- Johnson, K. L., T. G. Nicol, and N. Kraus. 2005. "Brain stem response to speech a biological marker of auditory processing." *Ear Hear* 26(5):424-434.
- Johnson, K. L., T. Nicol, S. G. Zecker, A. R. Bradlow, E. Skoe, and N. Kraus. 2008. "Brainstem encoding of voiced consonant-vowel stop syllables." *Clin. Neurophysiol.* 119(11):2623-2635.
- Kader, N.A. 2000. "Pitch detection algorithm using a wavelet correlation model." *Radio Science IEEE Seventeenth National Conference* C33/1--C33/8.
- Klatt, H. D. 1980. "Software for a cascade/parallel formant synthesizer." *J. Acoust. Soc. Am.* 67(3):971-995.
- Krishnan, A. 2002. "Human frequency-following responses: representation of steady-state synthetic vowels." *Hear. Res.* 166(1-2):192-201.

Laroche, M. 2010. "A Study of Auditory Speech Processing using Brainstem Evoked Responses under Quiet and Noisy Conditions." *M.A.Sc. Thesis, Ottawa-Carleton Institute for Biomedical Engineering*, University of Ottawa.

Laroche, M., H. R. Dajani, F. Prevost, and A. M. Marcoux. . "Brainstem Auditory Responses to Resolved and Unresolved Harmonics of a Synthetic Vowel in Quiet and Noise." *Ear Hear.* in press.

Lessard, C. S. 2006. "Stationary and Ergodic Random Processes." pp. 49-57 in "*Signal Processing of Random Physiological Signals*". Texas A&M University: Morgan & Claypool Publishers.

Madhavan, G. 1989. "Modified adaptive line enhancement for evoked potential extraction." *Engineering in Medicine and Biology Society, IEEE Annual International Conference* 2:708--709.

Madhavan, P. G. 1992. "Minimal repetition evoked potentials by modified adaptive line enhancement." *IEEE Trans. Biomed. Eng.* 39(7):760-764.

McNamara, D. M. and A. K. Ziarani. 2004. "A new adaptive technique of estimation of steady state auditory evoked potentials." *Engineering in Medicine and Biology Society, IEEE Annual International Conference* 2:4544-4547.

Moller, A. R. 2006. "The Auditory Nerveous System." pp. 75-192 in "*Hearing: Anatomy, Physiology, and Disorders of The Auditory System*". Dallas: Elsevier Science.

Mustiere, F., M. Bouchard, and M. Bolic. 2007. "Quality Assessment of Speech Enhanced using Particle Filters." *Acoustics, Speech and Signal Processing, IEEE International Conference* 3:1197-1200.

Ozdamar, O. and R. E. Delgado. 1996. "Measurement of signal and noise characteristics in ongoing auditory brainstem response averaging." *Ann. Biomed. Eng.* 24(6):702-715.

Pantev, C., L. E. Roberts, M. Schulz, A. Engelien, and B. Ross. 2001. "Timbre-specific enhancement of auditory cortical representations in musicians." *Neuroreport* 12(1):169-174.

- Picton, T.W., A. Dimitrijevic, M. Sasha John, and P. Van Roon. 2001. "The use of phase in the detection of auditory steady-state responses." *Clinical Neurophysiology* 112(9):1698-1711.
- Pitt, M.A. 1994. "Perception of pitch and timbre by musically trained and untrained listeners." *J. Exp. Psychol. Hum. Percept. Perform* 20(5):976-986.
- Prevost, F., M. Laroche, A. M. Marcoux, and H. R. Dajani. "Objective measurement of physiological signal-to-noise gain in the brainstem response to a synthetic vowel." *Clin. Neurophysiol.* in press.
- Rauner, H., W. Wolf, and U. Appel. 1984. "System identification techniques for noise reduction in evoked potential processing." *Acoustics, Speech, and Signal Processing, IEEE International Conference* 9:426-429.
- Regan, D. 1989. *"Human Brain Electrophysiology: Evoked Potentials and Evoked Magnetic Fields in Science and Medicine"*. The University of Michigan: Elsevier.
- Russo, N., T. Nicol, G. Musacchia, and N. Kraus. 2004. "Brainstem responses to speech syllables." *Clin. Neurophysiol.* 115(9):2021-2030.
- Sayed, A. H. 2003. "Linear estimation." pp. 47-77 in *"Fundamentals of Adaptive Filtering"*. Los Angeles: John Wiley & Sons.
- Sayed, A. H. 2011. "Linear Algebra." pp. 12-24 in *"Adaptive Filters"*. Los Angeles: John Wiley & Sons.
- Shi, Y. and K. E. Hecox. 1991. "Nonlinear system identification by m-pulse sequences: application to brainstem auditory evoked responses." *IEEE Trans. Biomed. Eng.* 38(9):834-845.
- Skoe, E. and N. Kraus. 2010. "Auditory brain stem response to complex sounds: a tutorial." *Ear Hear.* 31(3):302-324.
- Treichler, J. R. 1979. "Transient and Convergent Behavior of Adaptive Line Enhancer." *Transactions on acoustics, speech, and signal processing* 27(1):pp. 53-62.

- Vaidyanathan, P.P. 2007. "The Optimal Linear Prediction Problem." pp. 5-18 in *"The Theory of Linear Prediction"*. Carnegie Mellon University: Morgan & Claypoor Pub.
- Van Dun, B., J. Wouters, and M. Moonen. 2007. "Improving Auditory Steady-State Response Detection Using Independent Component Analysis on Multichannel EEG Data." *Biomedical Engineering, IEEE Transactions on* 54(7):1220-1230.
- Vander Werff, K. R. and K. S. Burns. 2011. "Brain stem responses to speech in younger and older adults." *Ear Hear.* 32(2):168-180.
- Weiner, N. 1949. "Extrapolation, interpolation, and smoothing of stationary time series: with engineering applications." *Journal of the American Statistical Association* 47:258-319.
- Westerkamp, J. J. and J. L. Slifka. 1989. "Adaptive estimation and detection of evoked potentials." *Engineering in Medicine and Biology Society, Images of the Twenty-First Century., Proceedings of the Annual International Conference of IEEE* 2:702-703.
- Widrow, B., JR J. Glover, John McCool, John Kaunitz, Charles Williams, Robert Hearn, James Zeidler, Eugene Dong, and Robert Goodlin. 1975. "Adaptive Noise Canceling: Principles and Applications" *Proc IEEE* 63(12):1692-1716.
- Woodworth, Walker, Stanley Reisman, and A. B. Fontaine. 1983. "The Detection of Auditory Evoked Responses Using a Matched Filter." *Biomedical Engineering, IEEE Transactions on BME*-30(7):369-376.
- Young, E. D. and M. B. Sachs. 1979. "Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory-nerve fibers." *J. Acoust. Soc. Am.* 66(5):1381-1403.
- Zhang, Y., N. Suga, and J. Yan. 1997. "Corticofugal modulation of frequency processing in bat auditory system." *Nature* 387(6636):900-903.