SPEECH AUDITORY BRAINSTEM RESPONSE SIGNAL PROCESSING: ESTIMATION, MODELING, DETECTION, AND ENHANCEMENT

ANWAR FALLATAH

Thesis submitted to the University of Ottawa
in partial Fulfillment of the requirements for the
Doctorate in Philosophy degree in Electrical and Computer Engineering

School of Electrical Engineering and Computer Science
Faculty of Engineering
University of Ottawa

Abstract

The speech auditory brainstem response (sABR) is a promising technique for assessing the function of the auditory system. This non-invasive technique has shown utility as a marker of central processing disorders, some types of learning difficulties in children, and potentially for fitting hearing aids. However, the sABR needs a long recording time to obtain a reliable signal due to the high background noise, which limits its clinical applicability. The objective of this work is to develop methods to detect the sABR in high background noise and enhance it based on a modeling approach and through experimental testing. First, sABR noise estimation based on LQ/QR decomposition is derived, and its mathematical proof is shown. Second, an autoregression model is used to estimate the single-trial sABR which is then used to test several sABR detection and enhancement methods. Third, a novel Artificial Neural Network (ANN) based detection approach is proposed and compared using modeled and recorded data to other detection methods in the literature: Optimal Linear Filter (LF), Online Estimator (OE), Mutual Information (MI) and Artificial Neural Network based on the Discrete Wavelet Transform and Approximate Entropy (ANN DA). Finally, comprehensive evaluation of several sABR enhancement methods is performed, based on the Wiener Filter (WF), Maximum-SNR Filter (Max-SNR), Adaptive Noise Cancellation (ANC) with Least-Mean-Square (LMS), Affine Projection (AP) and Recursive-Least-Square (RLS) adaptation algorithm. The results show that the developed LQ/QR decomposition estimated noise is similar to the actual noise, and the modeled data are statistically similar to the recorded data. Moreover, the proposed ANN-based detection method is more accurate and requires less processing time than other methods, and the comprehensive evaluation of enhancement methods shows that RLS has best overall performance in enhancing the sABR.

Therefore, the methods developed and evaluated in this work have the potential to reduce the required recording time for the sABR, and thus make it more practical as a clinical tool.

Table of Contents

List of Figures

Background and literature review:

Noise estimation paper:

Modeling paper:

Detection paper:

List of Tables

Modeling paper:

Detection Paper:

Enhancement paper:

List of Acronyms

ABR	Auditory Brainstem Response
sABR	Speech ABR
AE	Absolute Error
sAE	Slope Absolute Error
ANC	Adaptive Noise cancellation
ANN	Artificial Neural Network
ANN DA	ANN-based discrete wavelet transform and approximate entropy
ANN FF	ANN-based frequency features
ANOVA	Analysis of Variance
AP	Affine Projection
AR	Autoregressive model
ARMA	Autoregressive Moving-average
ASSR	Auditory Steady-State Response
CN	Cochlear Nucleus
Cz	Scalp location at the vertex
dB	Decibels
DWT	Discrete Wavelet Transform
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
LF	Optimal Linear Filter
LL	Lateral Lemniscus
LMS	Least-mean-square
MA	Moving-average
Max-SNR	Maximum SNR

MGB	Medial Geniculate Body
MI	Mutual Information
MSE	Mean Square Error
sMSE	Slope MSE
wMSE	Weighted MSE
swMSE	Slope Weighted MSE
OE	Online estimator
RLS	Recursive Least Square
SNR	Signal-to-Noise Ratio
VA	VA onset complex (transient response)
WF	Wiener Filter

List of Symbols

$.^T$	Matrix transposition
$.^H$	Matrix Hermitian
.*	Complex conjugate
.^	Estimated value
$.^-$	Mean value
u	Vector variables in boldfaced letter
E(x)	Statistical expectation operation
X	Matrix variables in capital
X(f)	Frequency domain
R	Autocorrelation matrix
p	Cross-correlation vector
σ_x	Variance of vector x
μ	Step-size parameter
R###	
M###	
N###	To describe the sABR signals, the first letter indicates the signal source: R: Human-recorded sABR. M: Modeled sABR that includes the signal and the noise part. N: Modeled noise of the sABR without the signal part. The following 2 or 3 digits indicate the signal quality which relates to the number of averaged trials: 36k: A clean sABR gathered from coherently averaging 36000 trials. 250: High quality sABR (i.e. with high SNR) coherently averaged over 250 trials. 20: Low quality sABR (i.e. with low SNR) averaged over 20 trials. 01: Estimated single-trial sABR. inf: Pure sABR without any noise content (a modeled signal that has infinite SNR).

1. Introduction

1.1 Motivation and Objectives

The Speech Auditory Brainstem Response (sABR) is a promising tool that can be used to objectively assess the auditory system. It has been shown to be useful for detecting certain types of central processing disorders in children (Johnson, Nicol, and Kraus, 2005) and it shows promise in evaluating other types of hearing impairment (Greenberg and Ainsworth, 2004). In addition, it is potentially useful for the objective fitting of hearing aids in patients such as infants who cannot communicate what they perceive (Dajani, Heffernan, and Giguère, 2013). In this measurement, electrodes placed on the scalp record brain activity in response to speech stimuli, such as vowels, syllables, or words, which are presented to the subject through earphones or loudspeakers. The main problem in obtaining a reliable sABR is the high background noise, and especially the Electroencephalogram (EEG) noise that is associated with general brain activity. Traditionally, synchronized presentation of multiple repetitions of the speech stimulus is performed to record and coherently average the responses over multiple trials to improve the signal-to-noise ratio (SNR). In practice, a very long recording (perhaps 15 minutes or more) may be needed to obtain a sufficiently high SNR through coherent averaging. This has limited the clinical applicability of this measurement.

As a consequence, the overall objective of this work is to improve the detection of the sABRs and enhance them in order to ultimately reduce the time needed to record them. To achieve this overall objective, the following are the specific objectives of the work:

1) Derive a noise estimation method that can estimate the background noise of sABR.
2) Model the responses across several "qualities" that correspond to different numbers of averaged responses, and in particular to model the single-trial response,
3) Develop a new method to detect the sABR that outperforms other reported methods,
4) When the sABR is detected, evaluate in detail the performance of several signal enhancement methods that can be applied to the recorded signal.

1.2 Contributions

This work consists of four main contributions:

1) Noise estimation based on LQ/QR decomposition for estimating the background noise of sABR. This work is summarized in a journal paper that has been submitted:

   Fallatah, A and Dajani, HR. LQ/QR decomposition noise estimation method for Speech Auditory Brainstem Response, *under review*.

2) A model of sABR signals is developed and the characteristics of the modeled responses are compared with those of recorded ones. The model is also used to estimate the single-trial sABR which is evaluated based on the theoretical SNR increase, the exponential

relation between the SNR and number of averaged trials, and power spectrum estimation. This work has been published in the following refereed conference paper:

Fallatah, A and Dajani HR. Estimating the properties of the single-trial speech auditory brainstem response using an accurate AR model, *Proceedings of the 14th International Conference on Signal Image Technology and Internet-based Systems (SITIS 2018)*, Las Palmas, Spain, Nov. 26-29, 2018.

3) A novel neural network-based method for detection of the sABR is developed and compared to four other detection methods, Optimal Linear Filter (LF), Online Estimator (OE), Mutual Information (MI), and Artificial Neural Network based on Discrete Wavelet transform and Approximate Entropy (ANNDA). The proposed method is shown to be more accurate and require less processing time than other methods. This work has been published in the following refereed journal paper:

   Fallatah, A and Dajani HR. Accurate detection of speech auditory brainstem responses using a spectral feature-based ANN method, *Biomedical Signal Processing and Control*, Vol. 44, pp.307-13, 2018.

4) Once the sABR is detected, it is then to be enhanced. In this work, the performance of five enhancement methods based on linear optimal and adaptive filters is evaluated in detail. The five methods are the Wiener Filter (WF), Maximum-SNR Filter (Max-SNR), Adaptive Noise Cancellation (ANC) with Least-Mean-Square (LMS) adaptation algorithm,

ANC with Affine Projection (AP), ANC with Recursive-Least-Square (RLS) adaptation. The results are summarized in a journal paper that is in preparation:

Fallatah, A and Dajani, HR. Assessment of linear optimal and adaptive filters for enhancing speech auditory brainstem responses, *in preparation*.

1.3 Outline of the Thesis

This dissertation is in the format of an article-based thesis. It is divided into eight chapters: Chapter 1 is an introduction which includes the motivation for the work, the main objectives, the thesis contributions, and this outline. Chapter 2 provides background information about the auditory nervous system and speech auditory brainstem responses, and a general literature review of sABR estimation, modeling, detection, and enhancement. Chapter 3 contains the first paper on LQ/QR decomposition noise estimation, Chapter 4 contains the second paper on sABR modeling, Chapter 5 contains the third paper on sABR detection, while Chapter 6 contains the fourth paper on sABR enhancement. Chapter 7 discusses the main results. Chapter 8 contains the conclusion and future work, and it is followed by a list of references. The Appendix contains the Ethics Approval certificate for this project and the approval of the thesis proposal committee for the use of the data collected under this project by the author of this thesis.

2. Background and literature review

2.1 The Auditory Nervous System

The auditory nervous system refers to the neural pathways from the inner ear through all different stages up to the auditory cortex in the brain. It has a vast number of interconnections between nerve cells that it is sometimes considered the most complex system of all sensory systems (Greenberg and Ainsworth, 2004). According to the pathway direction, the auditory nervous system can be divided into two parts:

Ascending auditory pathway

The ascending auditory pathway is the part that is responsible for sending signals from the inner ear to the auditory cortex. It can be in turn divided into two parts, the classical and non-classical pathways (Moller, 2003). The classical pathway involves hierarchal and parallel processing from the cochlea through three main layers: the cochlear nucleus (CN), the inferior colliculus nucleus (ICC) and the medial geniculate body (MGB) which is connected to the primary auditory cortex. The non-classical pathway is responsible for connecting this auditory system with other sensory systems. Its functionality is not fully understood, but it is believed to be responsible for lower perceptual activities, and it extends to the secondary auditory cortex or to other cortical areas.

Descending auditory pathway

The descending auditory pathway is responsible for the information transmission in the opposite direction of the ascending pathway. It starts from the auditory cortex all the way to the receptors in the cochlea. Its functionality is also not fully understood, but studies have shown that it is responsible for modulation of signal processing by the auditory system.

2.2 Speech Auditory Brainstem Responses

Auditory brainstem responses are evoked potentials that originate mainly in the brainstem in response to an acoustic stimulus. In speech Auditory Brainstem Responses (sABR), the acoustic stimulus is speech, which may be a vowel, syllable, word, or even a longer speech sample (Dajani, Purcell, Wong, Kunov, and Picton, 2005). In this work, the stimulus is a synthetic vowel with a fixed fundamental frequency and higher harmonics making the response similar to the widely studied Auditory Steady-State Response (ASSR) (Darwin, McKeown, and Kirby, 1989; Aiken and Picton, 2008). The stimulus is typically presented with insert earphones in alternating polarities between trials. The sABR can be recorded by placing electrodes on the surface of the subject's scalp, typically using a three electrodes configuration with one being the positive electrode on the vertex, another the negative electrode, and the third the ground. Figure 1 shows a schematic diagram of how the sABR is recorded, and the applied synthesized stimulus in the time and frequency domains. The three electrodes refer to the positive electrode, negative electrode, and the ground at left ear. The device delivers the stimulus to the subject's ears via insert earphones.

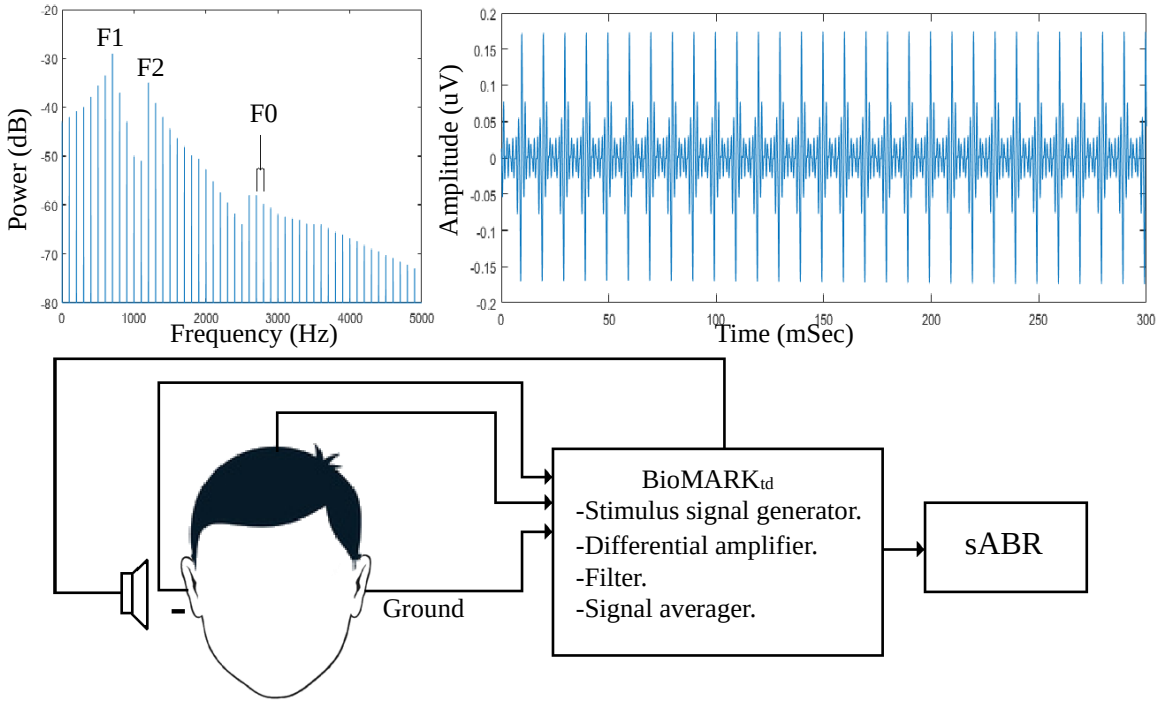


Figure 1 Schematic diagram of the sABR measurement system and the applied stimulus in the time and frequency domain. F0 refers to the fundamental frequency, and F1 and F2 are the 1st and 2nd formants.

A commonly used technique for assessing hearing functionality in the clinic is by ABR with sound stimuli such as clicks and chirps. However, it has been shown that using a speech stimulus (i.e. sABR) instead could extend these audiological tests. In (Russo, Nicol, Musacchia, and Kraus, 2004; Skoe and Kraus, 2010), for instance, the sABR is investigated to assess language-based learning problems in children by assessing the condition of central auditory neural pathways. In another study, the use of the sABR is proposed for evaluating hearing aids (Dajani, Heffernan, and Giguère, 2013) and for assessing impairments of auditory function in general (Johnson, Nicol, and Kraus, 2005; Johnson, Nicol, Zecker, Bradlow, Skoe, and Kraus, 2008).

Synthetic pure vowels and synthetic consonant-vowel syllables are the most commonly used stimuli in sABR (Skoe and Kraus, 2010). After obtaining the raw sABR, it can be processed further to give either the so-called Envelope Following Response (EFR) or the Frequency Following Response (FFR). The EFR follows the envelope of the stimulus speech signal and includes the fundamental frequency and its lower harmonics. It is obtained by averaging the responses to the stimulus in one polarity and those obtained from the inverted polarity stimulus. On the other hand, the FFR directly follows the stimulus harmonics, particularly in the region of the first formant and sometimes the second formant, if the frequency of the second formant is low enough to permit phase-locking of neural activity to the harmonics. It is obtained by averaging the responses to the stimulus in one polarity and the negative of the responses obtained from the inverted polarity stimulus (Aiken and Picton, 2008).

2.3 Literature review

This section provides a literature review on the four main topics investigated in this work, namely sABR noise estimation, modeling, detection, and enhancement. It discusses these four signal processing topics in biomedical applications and then focuses on EEG, evoked responses and other ABR-related signals since there is limited research on the processing of the sABR signal in the literature, and because other ABR-related signals share relevant characteristics with the sABR. Further targeted reviews of the literature specific to the topics of the four articles on which this thesis is based are found at the start of those articles.

ABR estimation

Signal estimation is a first step to gather information about the signal, and it is key to designing and optimizing any further processing technique such as detection and enhancement. Three aspects of the signal can be estimated: 1) The signal parameters such as its mean, variance and its autoregression coefficients, 2) The signal spectrum, and 3) The signal time-series. Optimal estimation tries to minimize the error based on a given criterion such as Mean Squared Error (MSE) and Least Mean Square (LMS). General discussion about estimation theory is found in (Alessio, 2015). Extensive information about signal estimation and how it is different from detection is found in (Poor, 1994). Principles of parameter estimation are discussed in (Levy, 2008). Estimating parameters for the purpose of signal processing is discussed in (Scharf and Demeure, 1991).

Early studies focused on estimating statistical parameters. A number of estimators have been proposed including Least Squares criterion (Johnson, 1940) and Maximum Likelihood estimator (Young, 1941). A comparison between Mean Square Error and Closeness criterion is presented in (Johnson, 1950) which found that each has its suitable application. Most of image noise-level estimations are based on Singular Value Decomposition (SVD) (Liu and Lin, 2012; Kumar, Sreenivasulu, and Rajan, 2016; Turajlić, 2018), while another image-noise estimation approach is the Laplace-convolution-based method (Turajlić, 2017). Regarding subspace systems, methods that are used include resonance-frequency estimation (Hirao and Adachi, 2009), estimating of parameters of linear relations (Wolfowitz, 1952; Kasami, 1961), linear regression model (Goldberger, 1962), linear and quadratic functionals (Collier, Comminges, and Tsybakov, 2017),

nonlinear generalized exponential autoregressive models (Chen, Gan, and Chen, 2018), and estimation of corrupted data of low-rank matrix (Gotze, Haardt, and Nossek, 1995).

Regarding estimation of ABR, early researchers focused on estimating its parameters. For instance, the quality of the averaged ABR is estimated since it is the key element to decide whether more recorded data are needed to enhance the averaged ABR. A basic quality estimation method is based on comparing the magnitude of the signal and noise (Elberling and Don, 1984). The importance of the signal stationarity for successful averaging and the estimation of weights to compensate for ABR non-stationarity are discussed in (Elberling and Wahlgreen, 1985). Estimation of fast wavelet coefficients that can enhance the averaged ABR is discussed in (Causevic, Morley, Wickerhauser, and Jacquin, 2005). Most proposed methods focus on estimating attributes of the targeted time-series signal such as its mean, power, and subspace coefficients. However, to the best of our knowledge, none of the mentioned studies propose speech ABR time-series estimation. The coherent averaging method can be considered as time-series estimation; however, it is a signal enhancement method since it usually produces an enhanced version of the targeted signal.

ABR modeling

Modeling is an important topic for data analysis, and its applications in stochastic behavior, prediction, control, feature extraction, knowledge discovery, validation, risk evaluation and decision making are well-known (Konishi and Kitagawa, 2008). Modeling biomedical signals helps in understanding the embedded signal-generating process at the physiological level which can contribute to developing better tools for biomedical diagnosis or treatment. Therefore, modeling biomedical signals has been of interest to many researchers in the field.

Early work focused on the frequency domain for studying signal statistical properties (Blackman and Tukey, 1958). The statistical properties of a signal have been often used to create the model, known as the statistical model. In (Akaike, 1974), a minimum information theoretic criterion (AIC) is used to as a means of model selection, and a detailed explanation of AIC is presented in (Konishi and Kitagawa, 2008). A theoretical description of linear modeling for EEG is presented in (Sato, 1963) which suggests that the stimulus harmonics affect EEG high frequencies. An autocorrelation-based model was used for pattern recognition and feature extraction for EEG signal classification (Hjorth 1973). Multichannel EEG was used for comparing the goodness of the model fit in (McEwen and Anderson 1975; Franaszczuk, Blinowska, and Kowalczyk, 1985). (McEwen and Anderson, 1975) also discusses the importance of wide-sense stationary analysis in modeling EEG signals. The importance of stationarity and Gaussian properties is also shown in (Sugimoto, Ishi, Iwata, and Suzumura, 1977). These statistical tools allow applying stochastic signals concepts in processing EEG signals. Autoregressive modeling and estimating the effective order of the model is presented in (Steinberg, Gasser, and Franke, 1985). An important finding of this research is that the EEG cannot be fully described by an AR model (in wider bandwidth) and some criteria have inconsistency in estimating the model order. Frequency domain modeling using Principal Component Analysis (PCA) is investigated for EEG analysis in (Valdés, Bosch, Grave, Hernandez, Riera, Pascual, and Biscay, 1992). This model considers the dominant frequency (i.e. with highest amplitude), and its bandwidth and amplitude only.

In recent years, researchers have focused on modeling ABRs. A non-linear ABR model is presented in (Harte, Rønne, and Dau, 2010) where different stages of auditory internal processing (such as in the basilar-membrane (BM), inner hair-cell (IHC), and auditory-nerve

synapse adaptation) are modeled mathematically. An exponential-based model is presented in (Jeng, Chung, Lin, Dickman, and Hu, 2011). A fuzzy-based model is presented in (Shirzhiyan, Jafarpisheh, Ahadi, and Jafari, 2013) which also reveals that fuzzy-logic has better performance in terms of error measures over AR in this application due to its nonlinearity. The auditory steady-state response (ASSR), which is a type of ABR that is composed of a single frequency component or multiple harmonics, is modeled in order to understand the origin of the background noise found in recordings with cochlear implant patients and to apply the right denoising technique in (Mina, Attina, Duroc, Veuillet, Truy, and Van, 2017). Four Independent Component Analysis (ICA) algorithms are investigated in this paper, and an important finding is that the detected response could be deformed if improper detection technique is used. To the best of our knowledge, there has been no previous attempt to model the speech auditory brainstem response (sABR), and in particular the single-trial response.

ABR detection

Signal detection is the essential step to determine the presence or absence of signals of interest. In biomedical applications, detection is a key element to reliable signal monitoring and diagnosis. ABR detection is important since the observed signal is buried in background noise which usually requires a long recording time to obtain an acceptable ABR quality. Principles of signal detection are discussed by (Levy, 2008), an extensive study of detection theory is found in (Song, Dagotto, Bae, and Kim, 2002), and detection of biomedical signals is presented in (Rangayyan, 2002b).

Earlier studies investigated the difference between the objective and the visual judgment approaches to detecting the ABR (Arnold 1985; Kidd Burkard and Mason, 1993), and they showed that objective approaches are more reliable than human visual-based detection. The importance

of ABR features in detection is presented first in (Sánchez, Riquenes, and Perez-Abalo, 1995), and an implementation of ABR features in an artificial neural-network is presented in (Tian, Grönfors, and Juhola, 1996). An overview of objective detection methods for ABR is presented in (Hyde, Sininger, and Don, 1998) where basic processing approaches (e.g. coherent averaging and correlation) and statistical tools (e.g. standard deviation and Hotelling T-test) in time and frequency domain are discussed. Wavelet-based feature extraction is used in (Delb, Strauss, and Plinkert, 2004), and Support Vector Machine (SVM) feature-based detection is used in (Acır, Özdamar, and Güzelis, 2006). Kernel-based ABR detection is used in (Corona-Strauss, Delb, Schick, and Strauss, 2010) where features such as phase, energy, and entropy in the time-frequency domain are used for detection. These methods were able to reduce the number of responses needed for detection to less than 200 responses. However, most of these methods are not suitable for sABR, since the most common forms of ABR are transient responses to short stimuli whereas sABR is usually dominated by a steady-state response to a sustained stimulus.

ABR enhancement

Signal enhancement is used to enhance the detected signal for better estimation of its characteristics, for improved display, and for decision-making. This is usually achieved by improving the SNR of the detected signal. The targeted signal, in EEG for instance, is usually present along with other unwanted signals (e.g. muscle artifact and 60Hz power-line interference), which makes it difficult to use further. However, by boosting relevant information and suppressing irrelevant background noise, the targeted signal can be better presented for clinical diagnosis and any decision-making process. The fundamentals of signal enhancement are

presented in (Benesty, Cohen, and Chen, 2017) while removing artifacts and filtering in biomedical signals are presented in (Rangayyan, 2002a).

Early studies focused on the use of coherent averaging for improving evoked response signals (Sato, Honda, Mimura, Teramoto, and Kitajima, 1962; Davis, 1964; Rappelsberger, 1985) and its effects on both steady-state and transient evoked responses since some of their content (such as at some frequencies) may be independent of the stimulus (Regan, 1966). Other averaging-based techniques such as weighted averaging have also been investigated. In (Davila and Mobin, 1992), it was shown that weighting the recorded signals before averaging can enhance the SNR up to 21%. The Wiener filter for enhancing evoked responses is presented in (Walter, 1968; Walter, 1970), and its better performance over coherent averaging was demonstrated. In (Albrecht, Lánský, Indra, and Radil-Weiss, 1977), different artificial signals were used to compare the Wiener filter against coherent averaging, and this study also concluded that the Wiener filter has better performance for enhancing evoked responses. The Wiener filter was also used for multi-channel enhancement in (Maki, Toda, Sakti, Neubig, and Nakamura, 2015). Other linear and optimal methods, including Kalman filter (Bartoli and Cerutti, 1983), Wavelets with Independent Component Analysis filters (Castellanos and Makarov, 2006), have also shown artifact suppression and signal enhancement. Knowledge-based enhancement, which classifies the recorded signals in real-time based on pre-stored characteristics, is presented in (Ifeachor, Hellyar, Mapps, and Allen, 1990). Fuzzy logic is also used for enhancement and clustering, so only signals with minimum distortion can pass to the next stage in processing (Riddington, Wu, Ifeachor, Allen, and Hudson, 1996). A neural-network-based technique is proposed in (Selvan and Srinivasan, 1999) as a solution for real-time EEG enhancement since their weights can be adapted

iteratively to minimize distortions. Transform domain methods including the Fourier Linear Combiner (Tahernezhadi, Karthik, and Kong, 1997) and Time-Frequency representation methods (Mohammadi, Pouyan, Abolghasemi, and Khan, 2017) have shown good performance in enhancing EEG signals.

Regarding ABR enhancement, early studies investigated the effect of the stimulus type on the quality of the observed response. In (Davis, 1964), it is reported that using a perceptually-discriminated auditory stimulus could enhance the evoked response. Weighted averaging has been implemented for ABR enhancement in (Davila, 1989). Adaptive filtering has been implemented in (Xue, Hecox, Tompkins, 1986; Madhavan, 1989). However, these studies focused on the transient response to very brief stimuli and so are not directly applicable to the sABR. Moreover, none of these studies has included a comprehensive evaluation of the performance or how the proposed methods can reduce the number of the recorded trials.

The published papers and papers in preparation:

3. The Noise Estimation Paper

LQ/QR decomposition noise estimation method for speech Auditory Brainstem Response

by Anwar Fallatah and Hilmi R. Dajani

Journal paper submitted for review

LQ/QR Decomposition Noise Estimation Method for Speech Auditory Brainstem Response

Anwar Fallatah[1] and Hilmi R. Dajani[1]
[1]School of Electrical Engineering and Computer Science, University of Ottawa, 161 Louis Pasteur, Ottawa, ON, Canada, K1N 6N5

Abstract

LQ (or QR) decomposition is one of the most important matrix factorizations. It has been used in various applications such as in solving linear least-square problems. In this paper, we introduce a novel application that applies LQ/QR decomposition to estimate additive background noise presented in an observation with a target signal that has a fundamental frequency and a number of harmonics. A mathematical proof, MATLAB script, and a practical implementation of estimating the background noise in speech Auditory Brainstem Response (sABR) are also included. The results show that proposed method is robust against interference and spike noise. Moreover, it produces less Mean Square Error (MSE) over iterative estimation methods such as Wiener Filter (WF) and Least Mean Square adaptive filter (LMS).

Keywords: LQ/QR decomposition, Fourier series, sine and cosine decomposition.

I. Introduction

LQ /QR decomposition is one of the most important matrix factorizations. It is often applied for solving linear least-square problems, dealing with square and non-square matrices (Golub, 1965), and for matrix rank revealing (Chan, 1987). It can also be used to avoid the operation of $Q^T Q$ product (for any given matrix Q) which cannot be implemented accurately if applied directly

(Golub and Van Loan, 2012). Recent studies show how this decomposition can identify subspace systems under deterministic type disturbance (Katayama, 2007; Katayama and Tanaka, 2007; Zhang, Liu, Hou, and Ni, 2017), where the deterministic disturbance occurs randomly with varying magnitude.

In this paper, we introduce a novel application that applies LQ/QR decomposition to estimate additive background noise present in an observation with a target signal that has a fundamental frequency and a number of harmonics. This method has the features of LQ decomposition, and in addition, it does not need prior knowledge about the noise or other transformations such as Fourier Transform of the observation which makes it ideal for time-series estimation when little is known about the nature of the noise. This method is different from other harmonic estimation methods such as Pisarenko (Ong and Kitney, 1986), MUSIC (Sun and Sun, 2008), and ESPRIT harmonic decomposition methods (Xiangwen, Zhe, and Wei, 2015) which estimate harmonic spectrum, and it is different from subspace methods such as (Katayama, 2007; Katayama and Tanaka, 2007; Zhang, Liu, Hou, and Ni, 2017) which implement the decomposition on the subspace parameters to estimate the system state. In contrast, the proposed method is used to decompose the time-series to estimate the noise part of the time-series.

This paper is divided into six sections. This section is the introduction. The second section discusses LQ decomposition and the proposed algorithm for noise estimation. The third section provides the mathematical proof. The fourth section demonstrates a practical implementation of how this algorithm can contribute in estimating the background noise present in the speech Auditory Brainstem Response (sABR). The fifth section shows Mean Square Error (MSE) performance with different noises compared to Wiener Filter (WF) and Least Mean Square

adaptive filter (LMS) estimation. The last section discusses the obtained results. The Appendix includes the MATLAB script for LQ/QR decomposition noise estimation.

II. Methodology

A. LQ decomposition:

Any invertible square matrix A can be factorized into $A = L\,Q$, where Q is an orthogonal matrix and a unitary positive definite matrix, and L is a Lower triangular matrix that has the following form:

$$L = \begin{bmatrix} l_{1,1} & 0 & 0 & 0 \\ l_{2,1} & l_{2,2} & 0 & 0 \\ \vdots & \vdots & \ddots & 0 \\ l_{m,1} & l_{m,2} & \dots & l_{m,m} \end{bmatrix} \qquad \text{Eq. 1}$$

which can be obtained from the following equation [1]:

$$L = A^T Q$$

This decomposition can be performed using numerous methods including the following:

1. Gram–Schmidt method.
2. Householder transformations.
3. Givens rotations.

B. Noise estimation:

For a given observation ***(b)*** which includes a periodic signal ***(x)*** buried in background noise ***(v)*** that is described as:

[1] In case of QR, the form $R=Q^T$ A is used instead to find R which is the right upper triangle matrix.

$$\boldsymbol{b} = \boldsymbol{x} + \boldsymbol{v}$$

where (***b***) is a vector in the form $b = [b_1, b_2, \dots, b_m]^T$,
the noise ($\hat{\boldsymbol{v}}$) can be estimated by LQ decomposition using following three steps:

1. **Create the combined matrix (*M*):** This matrix contains the observation signal *(**b**)* and other vectors that represent the fundamental frequency and harmonics of the target signal *(**x**)*. These vectors represent sine *(**s**)* and cosine *(**c**)* at each frequency. *M* is a (*j* . *m*) matrix where *j=2r + 1* and (*r*) is the number of frequency components present in the target signal ***x***, and it has the following form:

$$M = \begin{bmatrix} \boldsymbol{s}_1^T \\ \boldsymbol{c}_1^T \\ \vdots \\ \boldsymbol{s}_r^T \\ \boldsymbol{c}_r^T \\ \boldsymbol{b}^T \end{bmatrix} = \begin{bmatrix} s_{1,1} & s_{1,2} & \dots & s_{1,m} \\ c_{1,1} & c_{1,2} & \dots & c_{1,m} \\ \vdots & \vdots & \ddots & \vdots \\ s_{r,1} & s_{r,2} & \dots & s_{r,m} \\ c_{r,1} & c_{r,2} & \dots & c_{r,m} \\ b_1 & b_2 & \dots & b_m \end{bmatrix}$$

   where:

$$s_i(n) = sin(\omega_i\, n)$$

   and

$$c_i(n) = cos(\omega_i\, n)$$

   and where $\omega = 2\pi f$, *n=1,2, …, m* and *i = 1,2, …, r*.

2. **Obtain LQ decomposition:** By using any method of LQ decomposition, we can obtain the $\underline{L}$ and Q matrices of the combined matrix *M*. Notice that $\underline{L}$ is not a square matrix. Unlike *L* in Eq. 1, $\underline{L}$ is a (*j* . *m*) matrix since it is obtained from a non-square matrix (which is *M*), and it has the following form:

$$\underline{L} = [L\ 0]$$

where 0 is an all-zeros matrix with dimensions (j . $m - j$).

The second matrix (Q) is square, and it has the orthogonal property such that:

$$Q^T Q = I$$

where I is the identity matrix.

3. **Calculate the estimated noise:** The estimated noise ($\widehat{\boldsymbol{v}}$) is obtained by multiplying l' and q', where (· ') is a partition taken from the original matrix. To explain, l ' is a scalar which is the last element in the matrix L (which is equal to $l_{m,m}$ in Eq. 1), and q' is a vector equal to column number j in the matrix Q. In MATLAB script, they are represented as follows:

$$l' = L(end, end);$$

$$\boldsymbol{q}' = Q(j, :);$$

and finally, the estimated noise is equal to:

$$\widehat{\boldsymbol{v}} = [l'\boldsymbol{q}']^T$$

III. Mathematical proof

To prove that the proposed method estimates noise present in a given observation ($\boldsymbol{b}$), we use Fourier series to decompose the target periodic signal ($\boldsymbol{x}$), which is buried in a background noise ($\boldsymbol{v}$), into a sum of sine and cosine waves with different amplitudes (at the fundamental and its harmonics). Then, the proof will be satisfied if:

$$\boldsymbol{v} = \boldsymbol{b} - \boldsymbol{x} = l'\boldsymbol{q}' \qquad \text{Eq. 2}$$

The Fourier series decomposition states that any periodic signal is a sum of sine and cosine waves with different coefficients. Applying that into the target signal:

$$\boldsymbol{x} = \sum_{\boldsymbol{i=0}}^{\boldsymbol{r}} \boldsymbol{\alpha_i}\, \boldsymbol{cos}(\boldsymbol{\omega_i}\, \boldsymbol{n}) + \boldsymbol{\beta_i}\, \boldsymbol{sin}(\boldsymbol{\omega_i}\, \boldsymbol{n})$$

where the coefficients $\boldsymbol{\alpha_i}$ and $\boldsymbol{\beta_i}$ can be represented as follows:

$$\boldsymbol{\alpha_i} = \frac{\boldsymbol{c_i^T}\, \boldsymbol{x}}{\|\boldsymbol{c_i^T}\, \boldsymbol{c_i}\|}$$

and

$$\boldsymbol{\beta_i} = \frac{\boldsymbol{s_i^T}\, \boldsymbol{x}}{\|\boldsymbol{s_i^T}\, \boldsymbol{s_i}\|}$$

where $c_i = cos(\omega_i\, n)$ and $s_i = sin(\omega_i\, n)$.

Therefore, the left side of Eq. 2 (i.e. b - x) can be written as follow:

$$L.S. = \boldsymbol{b} - \sum_{\boldsymbol{i=0}}^{\boldsymbol{r}} \frac{\boldsymbol{c_i^T}\, \boldsymbol{x}}{\|\boldsymbol{c_i^T}\, \boldsymbol{c_i}\|}\, \boldsymbol{c_i} + \frac{\boldsymbol{s_i^T}\, \boldsymbol{x}}{\|\boldsymbol{s_i^T}\, \boldsymbol{s_i}\|}\, \boldsymbol{s_i}$$

Then, applying LQ decomposition on the combined matrix M using Gram–Schmidt method yields the following:

$$A = L\, Q$$

$$A = \begin{bmatrix} \boldsymbol{a_1} \\ \boldsymbol{a_2} \\ \vdots \\ \boldsymbol{a_m} \end{bmatrix}$$

$$\mathrm{L} = \begin{bmatrix} (\boldsymbol{e_1^T a_1}) & 0 & 0 & 0 \\ (\boldsymbol{e_1^T a_2}) & (\boldsymbol{e_2^T a_2}) & 0 & 0 \\ \vdots & \vdots & \ddots & 0 \\ (\boldsymbol{e_1^T a_m}) & (\boldsymbol{e_2^T a_m}) & \dots & (\boldsymbol{e_m^T a_m}) \end{bmatrix}$$

and

$$Q = [\boldsymbol{e_1} \quad \boldsymbol{e_2} \quad \cdots \quad \boldsymbol{e_m}]$$

where

$$\boldsymbol{e_i} = \frac{\boldsymbol{u_i}}{\|\boldsymbol{u_i}\|}$$

$$\boldsymbol{a_i} = \sum_{\boldsymbol{j=1}}^{\boldsymbol{m}} \left(\boldsymbol{e_j^T a_i}\right) \boldsymbol{e_j}$$

$$\boldsymbol{u_i} = \boldsymbol{a_i} - \sum_{\boldsymbol{j=1}}^{\boldsymbol{m-1}} \frac{\boldsymbol{e_j^T a_i}}{\left\|\boldsymbol{e_j^T e_j}\right\|} \boldsymbol{e_j}$$

Therefore, we can represent the matrix partitions *l*' and $\boldsymbol{q}$' as follows:

$$l' = \boldsymbol{e_m^T a_m}$$

$$\boldsymbol{q}' = \boldsymbol{e_m}$$

The inner product in *l*' produces a scalar number:

$$w_{\boldsymbol{m}} \triangleq \boldsymbol{e_m^T a_m}$$

and this is equal to $\|\boldsymbol{u_m}\|$. The $\boldsymbol{q}$' part is equal to:

$$\boldsymbol{q}' = \boldsymbol{e_m} = \frac{\boldsymbol{u_m}}{\|\boldsymbol{u_m}\|}$$

$$= \frac{\boldsymbol{u_m}}{w_{\boldsymbol{m}}}$$

$$= \frac{1}{w_m}\left[\boldsymbol{a_m} - \sum_{\boldsymbol{j=1}}^{\boldsymbol{m-1}} \frac{\boldsymbol{e_j^T a_m}}{\left\|\boldsymbol{e_j^T e_j}\right\|}\boldsymbol{e_j}\right]$$

Notice that e_j are normalized sine and cosine waveforms that alternate in the summation operation. Therefore, $\boldsymbol{q}'$ can be rewritten as the following equation:

$$\boldsymbol{q}' = \frac{1}{w_m}\left[\boldsymbol{a_m} - \sum_{\boldsymbol{j=1}}^{\boldsymbol{m-1}} \frac{\boldsymbol{c_j^T a_m}}{\left\|\boldsymbol{c_j^T c_j}\right\|}\boldsymbol{c_j} + \frac{\boldsymbol{s_j^T a_m}}{\left\|\boldsymbol{s_j^T s_j}\right\|}\boldsymbol{s_j}\right]$$

The estimated noise will be:

$$\hat{\boldsymbol{v}} = l'\boldsymbol{q}'$$

$$= w_m \frac{1}{w_m}\left[\boldsymbol{a_m} - \sum_{\boldsymbol{j=1}}^{\boldsymbol{m-1}} \frac{\boldsymbol{c_j^T a_m}}{\left\|\boldsymbol{c_j^T c_j}\right\|}\boldsymbol{c_j} + \frac{\boldsymbol{s_j^T a_m}}{\left\|\boldsymbol{s_j^T s_j}\right\|}\boldsymbol{s_j}\right]$$

The right side of the Eq. 2 (i.e. *l' q'*) is:

$$\text{R. S.} = \boldsymbol{a_m} - \sum_{\boldsymbol{j=1}}^{\boldsymbol{m-1}} \frac{\boldsymbol{c_j^T a_m}}{\left\|\boldsymbol{c_j^T c_j}\right\|}\boldsymbol{c_j} + \frac{\boldsymbol{s_j^T a_m}}{\left\|\boldsymbol{s_j^T s_j}\right\|}\boldsymbol{s_j}$$

Since $\boldsymbol{a_m}$ is the last column in the matrix *A*, it is equal to $\boldsymbol{b}$ the last column in *M* in *L.S.* Therefore, R.S. (right side) is equal to the L.S. (left side).

IV. sABR noise estimation

This section shows a practical implementation of LQ/QR noise estimation of sABR. Since sABR to a synthetic vowel contains a signal with a fundamental frequency and its harmonics (Prévost,

Laroche, Marcoux, and Dajani, 2013), its background noise can be estimated using LQ/QR method with a knowledge of the signal component only. A high-quality human recorded sABR (denoted by R36k[2]) is used to estimate the background noise that is present along with the evoked response which has components at the fundamental frequency (100 Hz) and 6 following harmonics (200, 300, 400, 500, 600, and 700 Hz). Figure 1 shows the observed sABR (i.e. R36k), the LQ-estimated noise (*v*), and the recovered signal (*x*) in time and frequency domains.

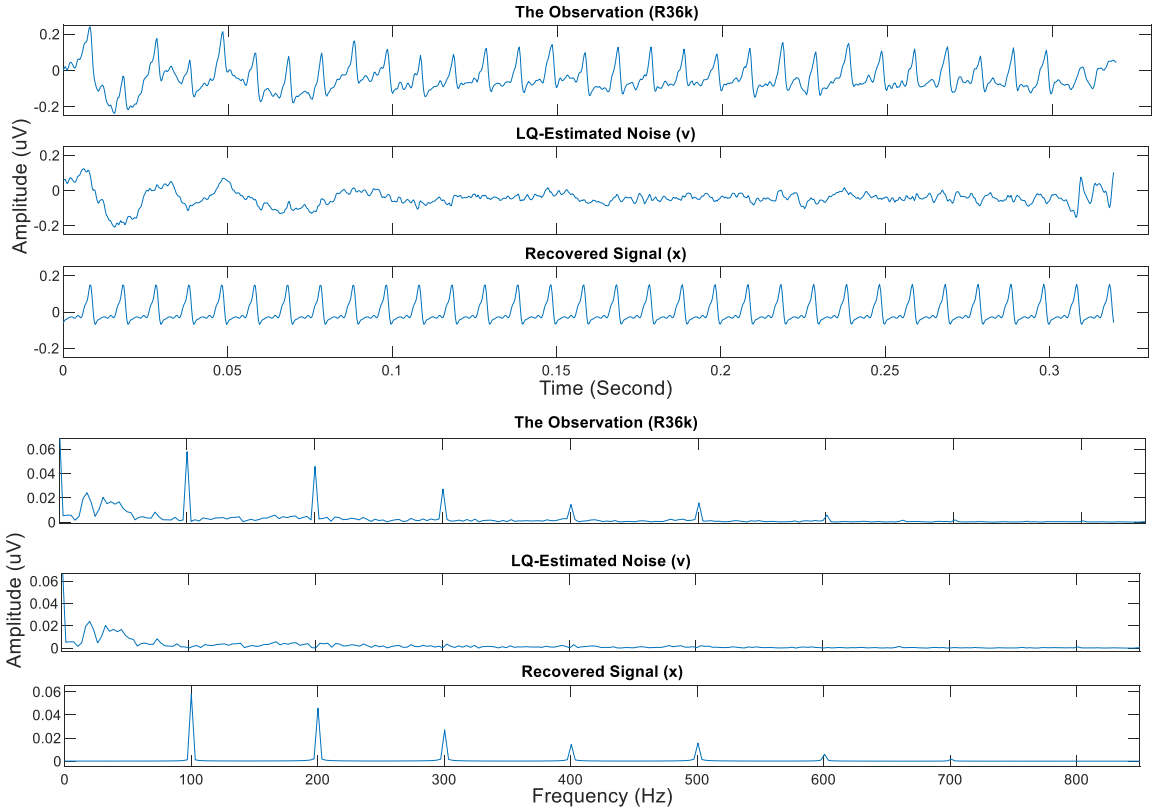


Figure 1 shows a practical implementation of LQ noise estimation of a background noise in sABR. The first three figures show the observation (R36k), the LQ-estimated noise (*v*), and the recovered signal (*x*) in the time domain and thefrequency domain in the following three figures.

As seen in Figure 1, the estimated noise captures the entire spectrum of the background noise. Therefore, when the estimated noise is subtracted from the observation, the recovered signal is

[2] Denoted by R36k since it is obtained from a human-recorded sABR with coherent averaging of 36,000 trials. Detailed explanation is found in (Fallatah, and Dajani, 2018).

obtained with components only at the fundamental frequency and its harmonics as seen in the last plot in Figure 1 confirming the effectiveness of the proposed method. The MSE between the actual noise and the LQ-estimated one is presented in the next section.

V. The Mean Square Error (MSE) performance

This section demonstrates the MSE performance of the proposed method to estimate noise in different conditions compared to performance of two other well-known filters. Moreover, the robustness of LQ/QR decomposition noise estimation against disturbance such as the power-supply interference and spike noise is shown. The proposed method is compared to two iterative-based[3] estimations: Wiener Filter (WF), which is well-known in optimal estimation, and Least-Mean-Square (LMS) adaptive filter. Both of these approaches are configured and optimized[4] to recover the targeted signal (i.e. x that is discussed in the previous section) from three different background noises:

1. From the actual recorded noise that present in R36k shown in Figure 1 (Denoted as R36k-noise).
2. From the R36k-noise with added 60Hz interference (R36k-noise + interference).
3. From the R36k-noise with 10 added spikes randomly placed in the time-domain. (R36k-noise + spike noise).

After obtaining the filter output (i.e. the estimated noise), MSE is used to measure the performance by comparing it with the actual noise which is used as a reference. This reference is obtained by retaining the noise spectrum only (i.e. the R36k spectrum except components at the fundamental

[3] Iterative-based methods are those that can produce an output even if they have not processed the entire length of the input. In FIR and IIR filters, for instance, there is an output after each iteration. On the other hand, block-based methods, such as LQ/QR decomposition, need the entire length of the input to obtain the output.
[4] 100-taps WF and 7-taps LMS are used. Detailed explanation about their configuration is presented in (Fallatah, 2012).

frequency and its harmonics). Figure 2 shows a summary diagram of the followed procedure for the performance comparison.

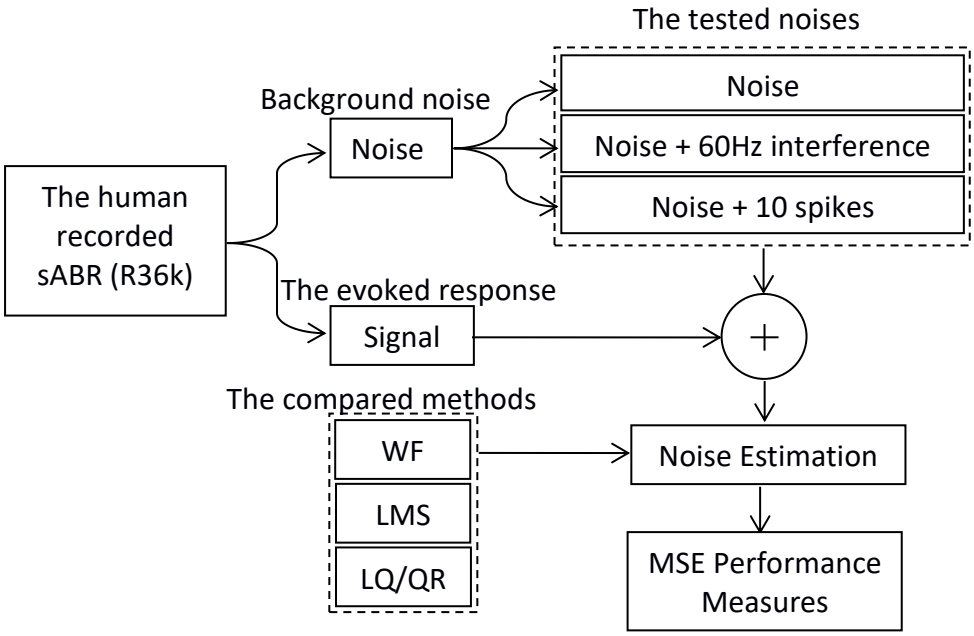


Figure 2 Summary diagram of the followed procedure for the performance comparison.

Figure 3 shows the signal in the time domain (the evoked response), the actual noise, and the estimated noise of the three estimation methods.

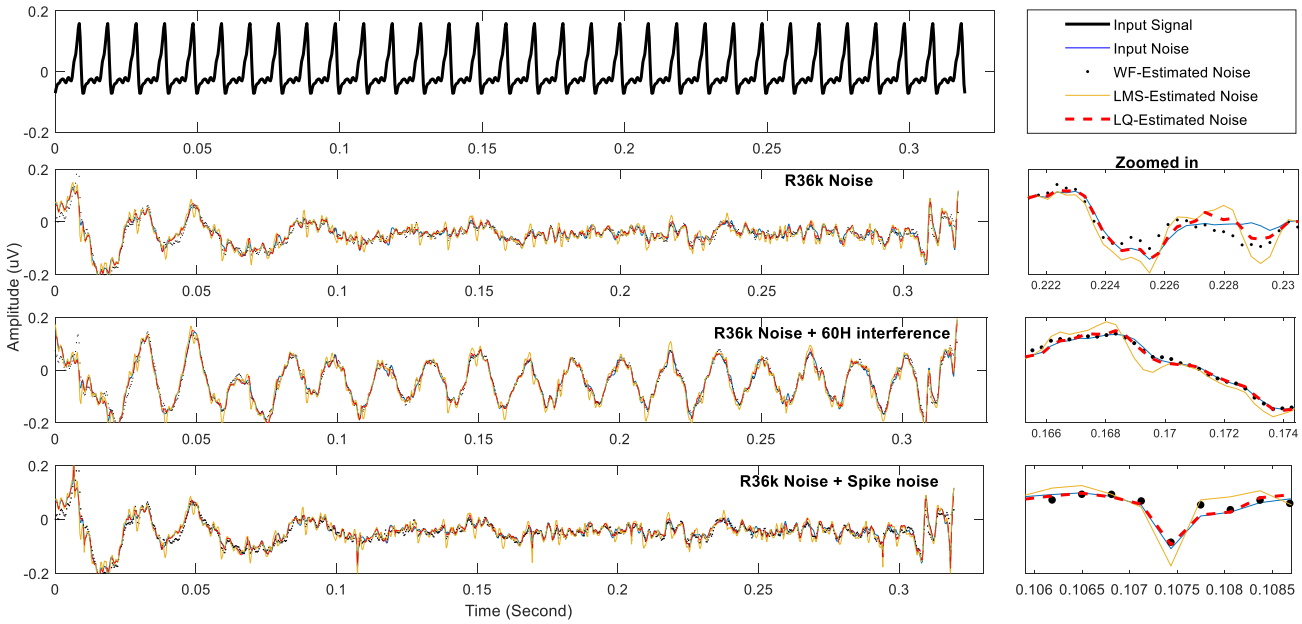


Figure 3 The target signal in time domain (top figure with bold black line), the noise (blue line), and the estimated noise of the three tested noises (R36k-noise, R36k-noise + interference, R36k-noise + spike noise) using the three estimation methods: WF (the black dots), LMS (orange line), and LQ (red dashed line).

Figure 4 shows the MSE performance of the three noise-estimation methods.

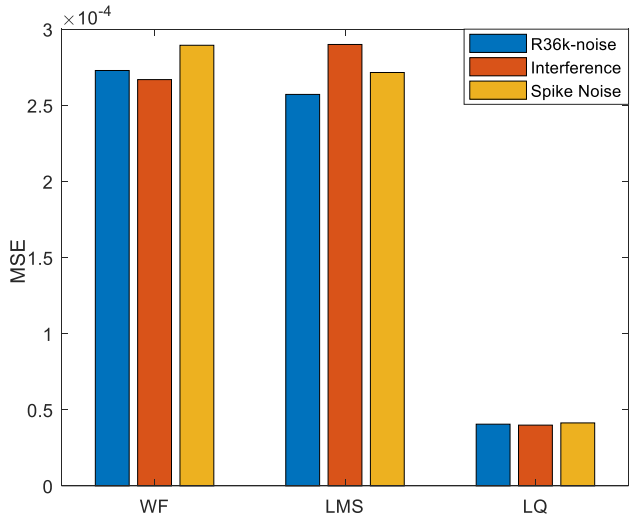


Figure 4 The MSE performance of the three noise-estimation methods (i.e. WF, LMS, and LQ) for the three noises: R36k-noise (the blue bars), R36k-noise + interference (the orange bars), R36k-noise + spike noise (the yellow bars).

The results in Figure 4 show that the LQ decomposition gives the best estimate of the noise compared to the two other methods. Moreover, the LQ method is robust against noise disturbance such as 60Hz interference and random spike noise.

VI. Discussion

The obtained results in the previous section confirm that the proposed method can affectively estimate the time-series of the noise. This is also what has been proven in Section **III**. The performance of iterative-based estimations such as WF and LMS is based on the signal correlation which does not consider short instances of noise, and they need time to track sudden changes in noise characteristics. In contrast, LQ\QR estimation decomposes the full spectrum of noise regardless of how it varies with time.

LQ/QR decomposition is suitable for estimating the noise time-series with knowledge of its bandwidth only, as seen in this study. In some cases where the signal and the noise spectra overlap, LQ/QR can be used to gather information about the noise characteristics to proceed to further processing such as adaptive noise cancellation, because LQ/QR decomposition recognizes all the bandwidth either as a noise or as a signal. This limitation needs further investigation which will be our objective for future research.

Appendix

This Appendix includes the MATLAB code for implementing LQ/QR decomposition noise estimation. The following code is for the LQ-based estimation function:

```
1.  %Noise Estimation by LQ decomposition
2.  %The code input is b(lx1) the observation and r the targeted
3.  %frequencies (Hz) in b, and it has two outputs x (signal) and
4.  %the estimated noise (v). An example of estimating a noise
5.  %contaminating a signal with components at 100, 200, …, 700Hz:
6.  %[x,v]=NEbyLQ(b,[100,200,300,400,500,600,700]);
7.  %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
8.
9.  function [x,v] =NEbyLQ (b,f)
10.   Fs = 3202; % The sampling rate
11.   l  = length(b);
12.   n  = (0:l-1)*(1/Fs);
13.   wn = 2*pi*n;
14.   r  =  length (f);
15.
16.   % Creating the combined matrix(M)
17.   M  = zeros(1,l);
18.   for i = 1:r
19.      s(i,:)=sin(wn*f(i));
20.      c(i,:)=cos(wn*f(i));
21.      M  = [M;s(i,:);c(i,:)];
22.   end
23.
24.   M  = [M(2:end,:);b'];
25.
26.   % Applying LQ decomposition
27.   [L,Q]=lq(M);
28.
29.   % Extracting L' and Q', and estimating the noise
30.   j = r*2+1;
31.   Ll = L(j,j);
32.   Ql = Q(j,:);
33.   v = (Ll*Ql)';
34.   x = b-v;
35. end
```

For the QR-based estimation, the last two steps of the LQ code can replaced by:

```
⋮                                ⋮
26.   % Applying QR decomposition
27.   [Q,R] = qr (M');
28.
```

```
    % Extracting Q' and R', and estimating the noise
    j = r*2+1;
    Rl = R(j,j);
    Ql = Q(:,j);
    v = (Rl*Ql);
    x = b-v;
end
```

References

Chan T. E, 1987. Rank revealing QR factorization. *Linear Algebra and its Applications*. **89**, pp. 67-82.

Fallatah, A., 2012. Enhancement of Speech Auditory Brainstem Responses using Adaptive Filters, *M.A.Sc. Thesis, School of Electrical Engineering and Computer Science*, University of Ottawa, Ottawa, Canada.

Fallatah, A., Dajani, H.R., 2018. Accurate detection of speech auditory brainstem responses using a spectral feature-based ANN method, *Biomedical Signal Processing and Control*, **44**(1), pp. 307-313.

Golub, G.H., 1965. Numerical methods for solving linear least squares problems. *Numerische Mathematik*. **7**(3), pp. 206-216.

Golub, G.H. and Van Loan, C.F., 2012. *Chapter 5: Orthogonalization and Least Squares. In Matrix Computations Vol.3*, JHU Press, Baltimore, United States, pp. 206-236.

Katayama, T., 2007. Role of LQ decomposition in subspace identification methods. *In Modeling, Estimation and Control*, Springer, Berlin, Heidelberg, pp. 207-220.

Katayama, T. and Tanaka, H., 2007. An approach to closed-loop subspace identification by orthogonal decomposition. *Automatica*, **43**(9), pp. 1623-1630.

Ong, H.G. and Kitney, R.I., 1986. Application of the Pisarenko harmonic decomposition method to physiological data. *Journal of Biomedical Engineering*, **8**(4), pp. 313-319.

Prévost, F., Laroche, M., Marcoux, A.M. and Dajani, H.R., 2013. Objective measurement of physiological signal-to-noise gain in the brainstem response to a synthetic vowel. *Clinical Neurophysiology*, **124**(1), pp. 52-60.

Sun, X., and Sun, L, 2008. MUSIC algorithm for harmonic frequency high resolution estimation in power system. *Electrical Measurement and Instrumentation*, **10**, pp. 28-31.

Xiangwen, S., Zhe, C. and Wei, T., 2015. Harmonic frequency estimation algorithm based on ESPRIT and MSWF in power system. *The Open Electrical & Electronic Engineering Journal*, **9**(1), pp. 518-523.

Zhang, S., Liu, T., Hou, J. and Ni, X., 2017. LQ decomposition based subspace identification under deterministic type disturbance. *Systems Science & Control Engineering*, **5**(1), pp. 243-251.

ADDENDUM

This section addresses some comments that were raised by the thesis committee:

In: *V. Mean Square Error (MSE) performance:*

1. The proposed method is compared to WF and LMS which are based on supervised training (they need a desired signal for tuning). A fair comparison could include "single-ended" methods, such as the mentioned harmonic estimations and classic PSD estimation methods, like spectrum subtraction, short-time spectral amplitude (STSA), and Modified Wiener Filtering.

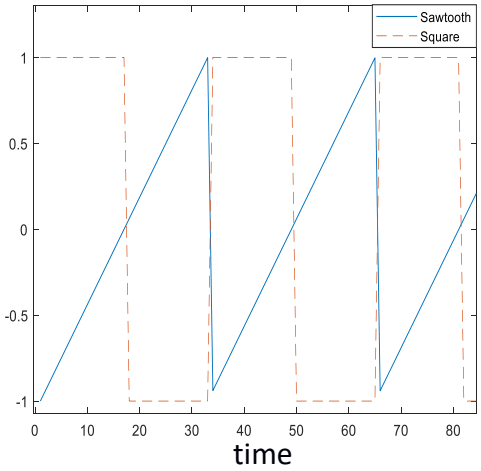
All mentioned harmonic estimation and PSD methods can be used for the comparison. However, these methods need the signal spectrum at some point. On the other hand, the applied methods in this paper do not need spectral analysis, and they can be implemented even on signals without harmonics.WF and LMS are based on supervised training which is considered in their favour, and the main reason for choosing them is that they generate the optimal solution theoretically.

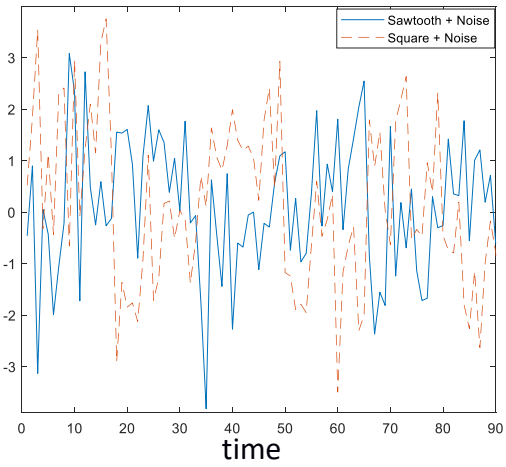
2. How is the Wiener filter an iterative method? Why did you use 100 taps for the Wiener filter and 7 taps for the LMS? What was the mu value for LMS and how was that chosen? Why not use NLMS instead of LMS? Some clarification is required here.

WF is an iterative method since it has a certain number of weights which is usually less than the input signal length. Therefore. the input signal must be iterated to process the entire data.
The reference (Fallatah, 2012) includes details regarding the selection of 100-tap WF and 7-tap LMS filter. Other filters could also have been chosen, but WF and LMS are chosen since they provide an optimal solution.

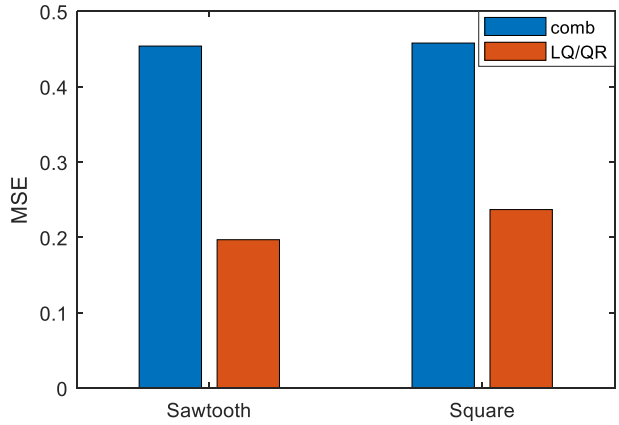
3. Why was this algorithm only tested for R36k? How does it work at other signal qualities? This is important because all other chapters investigated the performance of proposed techniques at different sABR signal qualities.

This method has been tested on all the sABR qualities, and the performance of estimating the noise (that is presented in the noise bandwidth) for the different qualities is similar to that of R36k. Only R36k is included in this paper because the signal and noise bandwidth can be considered as effectively separated. Several types of noise (R36k noise, 60 Hz interference, and spike noise) are tested to show the robustness of this method. The mathematical proof is included to generalize the performance of the method. In addition, the published code includes a demo with two other examples of estimating random background noise added to sawtooth and square signals (shown on the left below).





The figure on the right shows these signals with added random noise. Again, the MSE performance agrees with the reported results, and the figure below shows the MSE performance of LQ/QR decomposition compared to a comb filter centered on the fundamental frequency and its seven subsequent harmonics.



The published code is found in CodeOcean (https://codeocean.com/capsule/4854646/tree/v1)

In: *VI. Discussion:*
4. If the bandwidth of the noise needs to be separate from the bandwidth of the signal, then simple frequency selective filters (such as comb filter) could be used at lower cost.

Comb filter is an effective method that can be used instead of the proposed method in this particular example, since R36 has almost no noise at the fundamental frequency and harmonics. However, the proposed method can be used on signals without harmonics as mentioned. Moreover, the comb filter is considered an iterative method which would degrade its performance because it requires time to track sudden changes in the noise. Moreover, an experimental test has shown that the performance of WF and LMS (which are the methods used for comparison in the paper) is at least three times better than comb filter performance.

4. The Modeling Paper

Estimating the Properties of a Single-trial Speech Auditory Brainstem Response using an Accurate AR Model

by Anwar Fallatah and Hilmi R. Dajani

in Proceedings of the 14th International Conference on Signal Image Technology & Internet-based Systems (SITIS 2018), Las Palmas, Spain, Nov. 26-29, 2018.

2018 14th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)

Estimating the Properties of the Single-trial Speech Auditory Brainstem Response using an Accurate AR Model

Anwar Fallatah
School of Electrical Engineering and Computer Science
University of Ottawa
Ottawa, Canada

Hilmi R. Dajani
School of Electrical Engineering and Computer Science
University of Ottawa
Ottawa, Canada

***Abstract*—The human speech Auditory Brainstem Response (sABR) is an electrophysiological response with potentially important clinical and practical applications. However, because of the very low SNR of the signal, long recording times are usually needed over which the responses from a large number of trials are coherently averaged. Therefore, it is important to understand the properties of the single trial sABR, as this can help in developing methods to detect this response using a smaller number of trials. This paper presents a parametric model of averaged human sABR that is used to estimate the properties of a single-trial response. The Autoregression (AR) method is followed to model the sABR at four different signal qualities, based on recorded data coherently averaged over different numbers of trials. The properties of the modeled sABR are compared with the recorded ones in the time and frequency domains. This model is also used to estimate a single-trial sABR. The results show that the properties of the modeled responses (statistical distribution, SNR, noise power) are similar to the recorded sABRs. Moreover, coherent averaging based on the estimated single-trial sABR produces a comparable theoretical SNR increase, similar exponential relation between the SNR and number of averaged trials, and a similar power spectrum to the recorded sABR.**

***Index Terms*—Speech auditory brainstem response; Single-trial; Parametric modelling; Autoregression; Power spectrum estimation**

I. INTRODUCTION

The human speech auditory brainstem response (sABR) is an electrophysiological response that is obtained by presenting a speech stimulus to the ear and recording the responses evoked in the brain using surface electrodes placed on the scalp. The sABR has potentially important applications in assessing auditory function [1,2] and in developing a brain-computer interface for objective fitting of hearing aids [3] or automatic adjustment of other audio devices.

However, because of the typically very low SNR of the sABR, long recording times are usually needed over which the responses from a large number of trials (usually thousands) are coherently averaged. This long recording time has limited the practical use of this measured signal.

With coherent averaging over multiple trials, important amplitude and phase information in the individual trial responses may be lost [4]. Moreover, the non-coherent (noise) part of the single trial response is suppressed. This information could contribute to assessing the signal quality [5] and determining the location of the evoked response generators in the auditory pathway [6]. Denoising techniques other than coherent averaging have been applied to auditory evoked potentials, but not yet to the sABR, and these include: static [7] and adaptive filters [8], Fourier and Wavelet transform-based filters [9], and Independent Component Analysis [4]. Most of these methods aim to recover the coherent part of the stimulus response only, and the remainder of the spectrum is usually considered as pure background noise which is suppressed during the filtering process. The filtering often involves estimation of the power of the signal and the noise of the observed single-trial. Therefore, a better understanding of the properties of a single trial sABR, through mathematical modeling, may help in the development of methods for denoising this signal that outperform coherent averaging.

Basic ways to model Electroencephalogram (EEG) signals involve using white-noise and pink-noise [10, 11], while advanced modeling techniques are used for full spectrum estimation, linear prediction, and feature extraction [12]. Advanced modeling methods can be classified into parametric and nonparametric models [13]. The main difference between these two classes is that nonparametric models make no assumptions about the data probability distribution. Nonparametric methods include periodogram, the smoothed periodogram (known as Blackman-Tukey method), and multiple-averaged periodograms (Bartlett and Welch method). Parametric methods include autoregressive (AR), moving-average (MA), and autoregressive moving-average (ARMA) models.

In the literature, the EEG parametric model is discussed in [13, 14] for feature extraction, and a parametric model-based classification of different EEG signals is presented in [12]. The Auditory Steady State Response (ASSR), another auditory brainstem response, is modeled in [15]. To the best of our knowledge, none of these studies or others in the literature investigated in detail how the proposed modeled responses compare to the recorded responses.

978-1-5386-9385-8/18/$31.00 © 2018 IEEE
DOI 10.1109 /SITIS.2018.00020

In this study, the AR method is followed to model four different qualities of the noise part of the recorded sABR, based on recorded data coherently averaged over different numbers of trials. The properties of the modeled sABR (statistical distribution, SNR, noise power) are compared to those of the recorded data. The model is used to test several sABR detection methods in [16] and to estimate the single trial response in this paper. The estimated single-trial response properties are evaluated using: the theoretical SNR increase with coherent averaging, the exponential relation between SNR and number of averaged trials, and power spectral estimation of the noise.

II. Methodology

In this study, sABRs of three different qualities are recorded based on averaging of 20, 250, and 36000 trials. The term quality here relates to the noise level in the signal retained after coherent averaging over different numbers of trials, but in other contexts, could be due to a particular experimental condition (e.g. with background noise added to the stimulus [17]). All modeling and evaluation is performed in MATLAB (The MathWorks, Natick, MA).

A. Experimental Methodology and Recorded Responses

Sound stimuli were presented to subjects through an insert earphone in the right ear, while they sat in an electrically shielded sound proof booth. A recording electrode was placed on the scalp at the vertex (Cz), a reference electrode on the right earlobe, and a ground electrode on the left earlobe. The speech stimulus was the English vowel /a/ generated using a formant synthesizer with a fundamental frequency of 100 Hz and first three formants at 700, 1220 and 2600Hz. The stimulus was presented in alternating polarity at a calibrated level of 80 dB SPL. The responses to multiple trials were averaged coherently to provide what known as the envelope frequency responses. This type of sABR follows the envelope of the stimulus, and so it also has a fundamental frequency of 100 Hz and harmonics at 200 Hz, 300 Hz, etc. The data from human subjects were collected as part of previous studies [17, 18, 20] that were approved by the Research Ethics Board of the University of Ottawa. Further details on the experimental recording methods can be found in [20].

There are three sets of recorded data: R20, R250 and R36k. R20 were obtained from 5 subjects and include 75 blocks, with each block being the coherent average of responses from 20 individual trials. R250 were obtained from four subjects and include 36 blocks, with each block being the coherent average of 250 responses. R36k includes a single very high SNR block based on the coherent averaging of 36000 responses from 12 subjects (including the R250 and R20 responses). Each response sweep was recorded over 0.3198 sec and sampled at 3202 Hz, giving 1024 samples per sweep.

B. The Autoregression Model

1) Model of the noise part of the sABR: Autoregression (AR) is often preferred to model EEG signals [13], where it is used to obtain the optimal linear coefficients that minimize the least square criterion. Autoregression is described by:

$$AR_n = -\sum_{i=1}^{p} a_i v_{n-i} \quad (1)$$

where are a_i the AR parameters, v is zero-mean white noise, and AR is the obtained model signal. The selected order of the model (p) is six which is sufficient to describe signals with no, or few, distinct spectral components. The AR parameters are obtained using the autocorrelation method (also known as Yule-Walker) described by:

$$a_i = R_x^{-1} r_x \quad (2)$$

where is a_i the inverse of autocorrelation matrix of x and its autocorrelation vector is R_x. This model is denoted by N (since it describes the noise spectrum present in the sABR) followed by the number of averaged trials (e.g. N20 represents the noise after averaging 20 trials). Table I shows the obtained AR parameters of each sABR quality (i.e. N36k, N250, and N20). It should be noted that the noise spectrum in the sABR could reflect background EEG activity in the brain in addition to recording noise.

TABLE I
THE OBTAINED AUTOREGRESSIVE (AR) PARAMETERS THAT REPRESENT EACH NOISE QUALITY OF RECORDED SABR INDICATED IN THE TOP ROW

Coeffi-cients	*N36k*	*N250*	*N20*	*N01**
a_0	-1.66	-1.82	-1.71	-1.7
a_1	0.858	1.27	1.27	1.06
a_2	-0.0548	-0.473	-0.473	-0.374
a_3	-0.0935	0.135	0.174	0.188
a_4	-0.117	-0.0675	-0.0807	-0.0818
a_5	0.0823	0.00557	-0.00134	-0.00396

* The estimated single-trial parameters, discussed in Section "Modeling AR parameters at other signal qualities".

2) Model of the clean sABR: The clean response signal has a finite number of harmonic components since the sABR depends on neural phase-locking in the auditory system which decays at higher frequencies. Therefore, amplitudes and phases of these components are selected based on the clean recorded sABR (i.e. R36k). This model is denoted by MInf since it has no noise spectrum, and therefore its SNR is infinity.

3) Evaluation of the modeled sABR: To evaluate the model, 100 signals are obtained for each noise quality model (i.e. N36K, N250, and N20) and added to MInf. The obtained modeled sABRs are denoted by: M36k, M250, and M20. Fig. 1 shows averages of all samples of these signals in the time domain, while Fig. 2 shows them in the frequency domain. The Z-test is performed to compare the SNRs of the modeled data in the frequency domain to those of the recorded sABR. This test is not performed for M36k and R36k since more than 30 samples are needed [21]. Fig. 3 shows the distributions of the recorded and the modeled sABR time-series, and their boxplot representations are shown in Fig. 4.

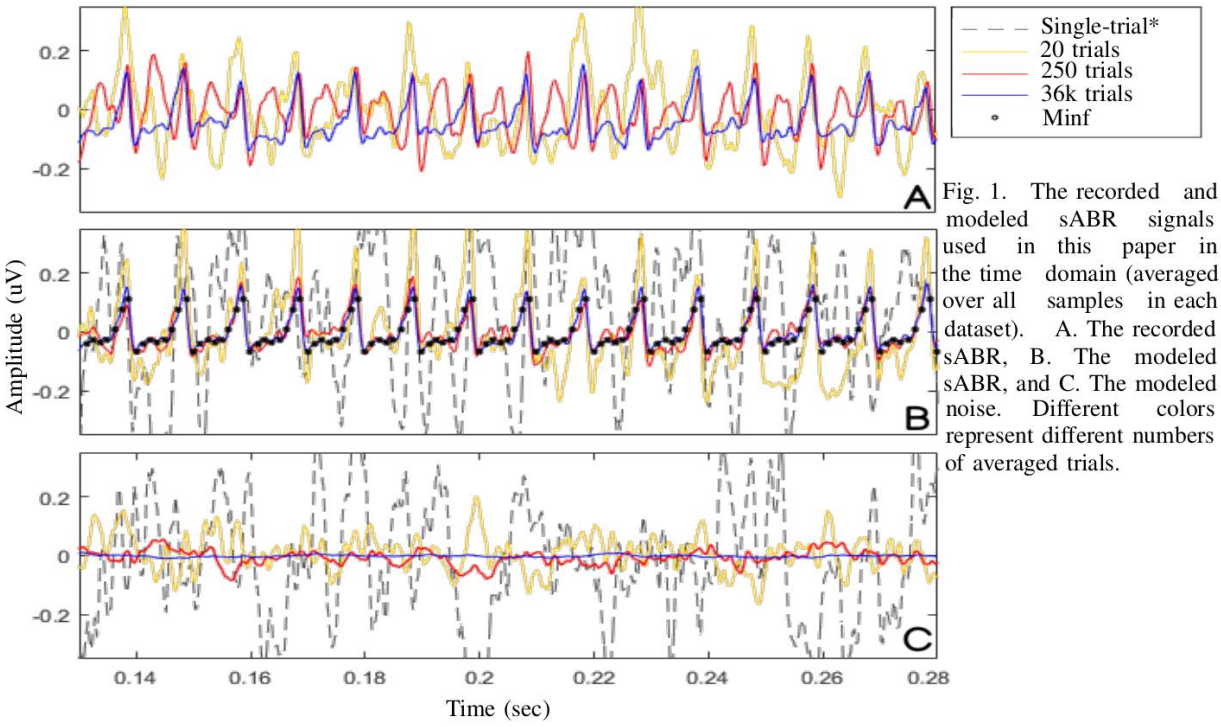


Fig. 1. The recorded and modeled sABR signals used in this paper in the time domain (averaged over all samples in each dataset). A. The recorded sABR, B. The modeled sABR, and C. The modeled noise. Different colors represent different numbers of averaged trials.

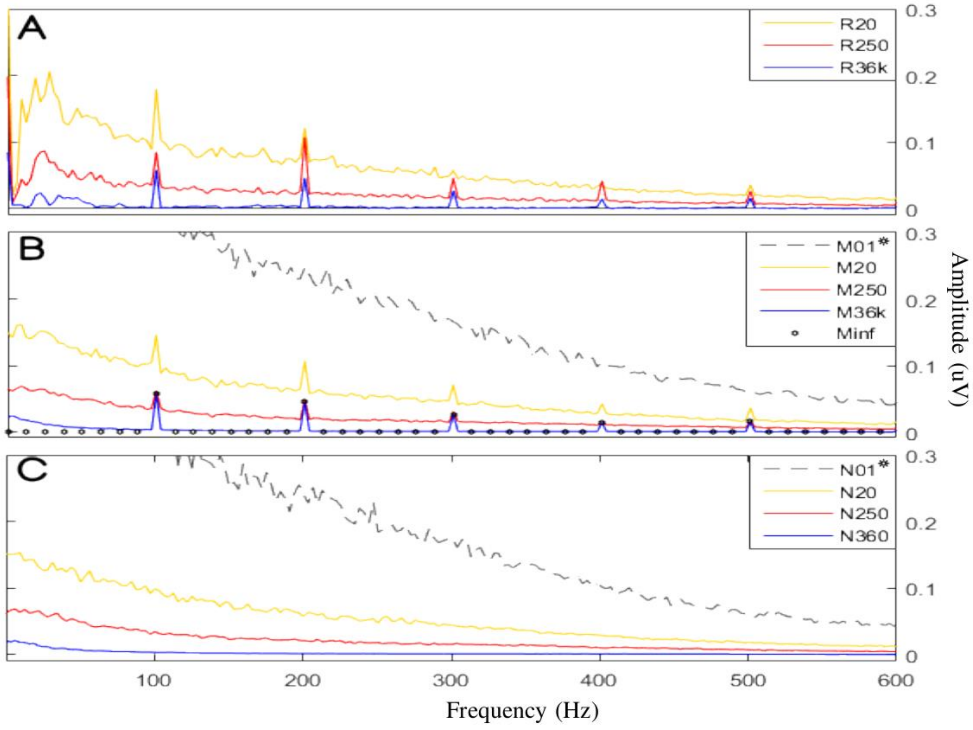


Fig. 2. The recorded and modeled sABR signals used in this paper in the frequency domain (averaged over all samples in each dataset). A. The recorded sABR, B. The modeled sABR, and C. The modeled noise. Different colors represent different numbers of averaged trials.

*The estimated single-trial.

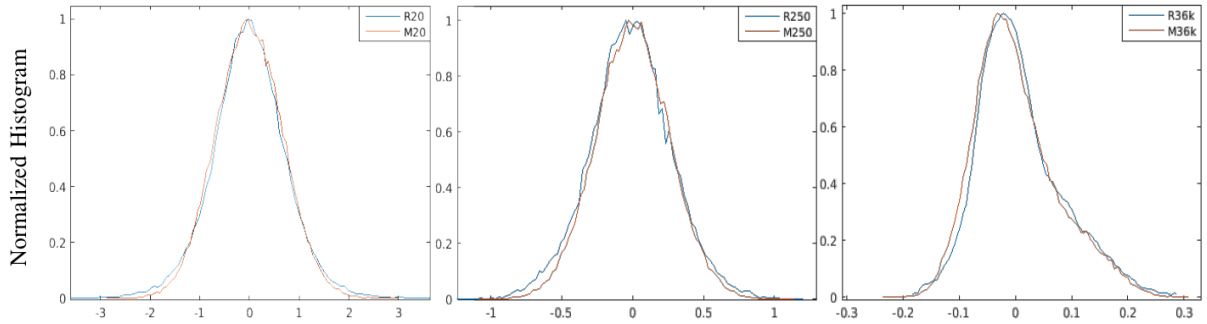


Fig. 3. Distributions of all recorded (in blue) and all modeled (in red) sABR time-series. The range of the x-axis in the three plots is different reflecting the decrease in width of the distribution when more trials are averaged.

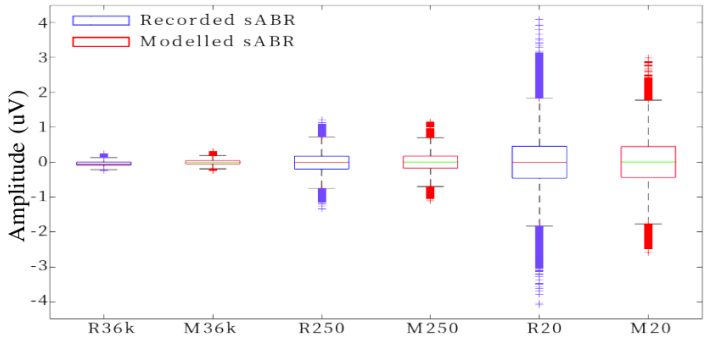


Fig. 4. Boxplots of all recorded (in blue) and all modeled (in red) sABR time-series.

4) Modeling AR parameters at other signal qualities: In order to model the sABR at other signal qualities, and in particular the single trial sABR, the AR parameters are modeled using linear regression for a0, a1, and a4 and exponential equations for the rest. These two types of models (i.e. linear regression and exponential equation) are selected according to their goodness of fit, which was less than 10^{-4} for the SumSquare-Error at a 95% confidence level. These models are used to obtain the parameters of the single-trial sABR shown in Table I under the N01 column.

C. SNR and Noise Power Comparisons

Ideally, the obtained single-trial modeled signal would be compared to the recorded single-trial response. However, in this study, signal-trial sABRs were not recorded. Moreover, even if this signal is available, measuring its SNR directly is not possible due to the high power of noise. As a result, three estimation methods are used to assess the overall performance of the proposed modeling method as described in the following sub-sections.

1) The theoretical SNR increase: The theoretical increase of the SNR obtained after coherent averaging of a signal containing uncorrelated zero-mean background noise is [22]:

$$SNR_{inc} = \sqrt{n} \tag{3}$$

where n is the number of averaged trials, and the SNR of the signal before and after averaging are related by:

$$SNR_{inc} = \frac{SNR_{aft}}{SNR_{bef}} \tag{4}$$

Therefore, in decibels:

$$SNR_{bef} = SNR_{aft} - 10\ log_{10}(n) \tag{5}$$

Since both SNR_{aft} and n the number of averaged trials are available, the SNR of the recorded signal-trial sABR can be estimated using this equation. R20 is selected to estimate the single-trial response using this method since it has the lowest number of trials and therefore is the closest to the single-trial sABR.

2) Using the exponential relation between the number of trials and the SNR values: In practice, the SNR change with coherent averaging of sABR across trials may not follow the theoretical relation given above, particularly for higher number of trials. This could be due to a number of factors such as time jitter in the measurement hardware, non-stationarity of the background noise, and small changes in the target signal.

Consequently, the relation between the number of averaged trials and the SNR of the recorded sABR can be described by the following power function:

$$SNR = a(n)^b \tag{6}$$

where *a* and *b* are selected to minimize the Mean-Square-Error (MSE) between the modeled response SNR and the recorded response SNR. In this study, these constants were found to be 0.2412 and 0.4068 respectively, and the resulting MSE of the curve fit is 0.25 at a 95% confidence level.

3) By estimating the noise power: The noise part of the modeled sABR can be assessed independently, especially with the responses modeled based on low number of averaged trials where the noise power is larger than the target response power. A non-parametric approach is used to estimate the power spectrum of the noise part of the recorded sABR and of the modeled data (i.e. N01, N20, N250, and N36k). In the recorded signal, the noise is obtained by removing the target signal components in the frequency domain (i.e. at 100, 200, 700Hz) using the LQ factorization method [7].

Welchs estimation method is followed using Hamming window, sample size equal to 25% of the length of one sABR block, and 25% overlap between the samples [23]. The obtained levels of the noise power of the recorded data are also modeled to estimate the noise level of the single-trial response by following the method discussed in Section IIC but on the noise part only.

III. Results and discussion

The distributions of the modeled and recorded sABR are very similar as shown in Fig. 3. This is also confirmed in the time-series boxplots in Fig. 4. Moreover, the results of the Z-test (Table II) show that there is no statistically significant difference between the mean SNRs of the modeled and recorded data for the two signal qualities (M20 and M250).

Table III summarizes all estimated SNR and noise results. The upper part of the table (in ”1” and ”2”) shows SNR comparisons based on the two methods discussed in Sections IIC1 and IIC2 under ”SNR and noise power comparisons”, respectively.

TABLE II
Z-test of the mean SNR in the modeled and recorded data

	R250	**R20**
Population mean of SNR	2.2369	0.7321
The standard deviation	0.7545	0.3632
p-value (comparison with the modeled data)	0.8910	0.1186

The lower part of the table (in ”3”) contains noise power comparisons obtained using the third estimation method discussed in Section IIC3. The absolute error (in dB) between the estimates obtained using the proposed AR modeling approach and the three estimation methods described above is shown in the last column.

As can be seen, the single-trial response modeled using the proposed approach has an approximately 1dB SNR difference relative to that obtained using the theoretical SNR increase and the exponential SNR relation, and 0.6 dB difference in terms of the estimated spectral noise power.

Regarding the modeled responses corresponding to 20, 250, and 36k averaged trials, the proposed method gives SNR differences less than around 2 dB relative to those obtained using the exponential SNR relation, and also less than 2 dB difference in terms of the estimated spectral noise power. These results show that the proposed method can be used to model the time-series of the single-trial sABR and of other response qualities.

IV. Conclusion

In this paper, a new single-trial speech Auditory Brainstem Response estimation approach based on an Autoregression model is presented. The parameters of the single-trial are derived and its properties (statistical distribution, SNR, and noise power) are compared to those estimated for recorded single-trial responses. Results show that the proposed method can be used to estimate a single-trial sABR and can be generalized to model other sABR qualities. The measurement of sABR has important clinical applications but has been severely hindered by the long recording times that are required. Modeling the single-trial sABR will potentially help researchers to develop new signal processing approaches reduce the recording time as shown in [16].

TABLE III

SNR AND NOISE POWER OBTAINED USING THE PROPOSED AR MODELING APPROACH AND THE THREE ESTIMATION METHODS DESCRIBED IN THE TEXT

Results of:	A. The three estimation methods (dB) described in Section IIC under "SNR and noise power comparisons"			B. The proposed AR method (dB)	Absolute Error (dB)
SNR comparisons	1) Based on theoretical SNR increase	***R01***	-14.10	-13.38	0.72
	2) Based on exponential relation between SNR and trial number	***R01***	-12.35	-13.38	1.03
		R20	-1.77	-2.06	0.29
		R250	7.16	6.95	0.21
		R36k	24.72	27.04	2.32
Noise power comparisons	3) Based on noise power spectrum estimation	***R01***	12.38	11.82	0.56
		R20	3.52	5.27	1.75
		R250	-3.97	-4.71	0.74
		R36k	-18.56	-17.2	1.36

ACKNOWLEDGMENT

Partial support for this work was provided by the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

[1] H.R. Dajani, D. Purcell, W. Wong, H. Kunov , T.W. Picton " Recording human evoked potentials that follow the pitch contour of a natural vowel," IEEE Transactions on Biomedical Engineering, 52(9), pp. 16141618, 2005.

[2] S.J. Aiken, T.W. Picton. "Envelope and spectral frequency-following responses to vowel sounds," Hearing Research, 245(1-2), pp. 35-47, 2008.

[3] H.R. Dajani, B.P. Heffernan, C. Gigure. "Improving hearing aid fitting using the speech-evoked auditory brainstem response." IEEE Engineering in Medicine and Biology Society Conference (EMBC), pp. 28122815, 2013.

[4] D. Iyer, and G. Zouridakis, Single-trial evoked potential estimation: Comparison between independent component analysis and wavelet denoising, Clinical Neurophysiology, 118(3), 2007, pp. 495-504, 2007.

[5] C. Elberling and M. Don, Quality estimation of averaged auditory brainstem responses, Scandinavian Audiology, 13(3), pp. 187-197, 1984.

[6] E.D. Farahani, T. Goossens, J. Wouters, A. Wieringen, Spatiotemporal reconstruction of auditory steady-state responses to acoustic amplitude modulations: Potential sources beyond the auditory pathway, NeuroImage, 148, pp.240-253, 2017.

[7] B. Van Dun, J. Wouters, M. Moonen,, "Multi-Channel Wiener Filtering Based Auditory Steady-State Response Detection," IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 929-932, 2007.

[8] D.M. McNamara, A.K. Ziarani, "A new adaptive technique of estimation of steady state auditory evoked potentials," IEEE Engineering in Medicine and Biology Society Conference (IEMBS), pp. 4544-4547, 2004.

[9] A. Bradley, W. Wilson, On wavelet analysis of auditory evoked potentials, Clinical Neurophysiology, 115(5), pp 1114-1128, 2004.

[10] R.J. Morana, K.E. Stephanab, T. Seidenbecherc, H.C. Papec, R.J. Dolana, K.J. Fristona, " Dynamic causal models of steady-state responses," Neuroimage, 44(3), pp.796-811, 2009.

[11] M. Potter, N. Gadhok and W. Kinsner, "Separation performance of ICA on simulated EEG and ECG signals contaminated by noise," IEEE Canadian Conference on Electrical and Computer Engineering, 2, pp. 1099-1104, 2002.

[12] Y. Zhang, B. Liu, X. Ji, D. Huang, "Classification of EEG signals based on autoregressive model and wavelet packet decomposition," Neural Processing Letters, 45(2), pp. 1370-4621, 2017.

[13] J. Pardey, S. Roberts, L. Tarassenko, A review of parametric modelling techniques for EEG analysis, Medical Engineering and Physics, 8(1), pp. 2-11, 1996.

[14] R.D Pascual-marqui, P.A Valdes-sosa, A. Alvarez-amador, "A parametric model for multichannel EEG spectra," International Journal of Neuroscience, 40(1), pp. 89-99, 1988.

[15] F. Mina, V. Attina, Y.Duroc,E. Veuillet, E. Truy, H. Thai-Van, "Auditory steady state responses and cochlear implants: Modeling the artifactresponse mixture in the perspective of denoising," Public Library of Science (PloS), 12(3), pp. 1932-6203,2017.

[16] A. Fallatah, H.R. Dajani, Accurate detection of speech auditory brainstem responses using a spectral feature-based ANN method, Biomedical Signal Processing and Control, 44(1), pp. 307-313, 2018.

[17] F. Prvost, M. Laroche, A.M. Marcoux, H.R. Dajani, "Objective measurement of physiological signal-to-noise gain in the brainstem response to a synthetic vowel," Clinical Neurophysiology, 124(1), pp. 52-60, 2013.

[18] A. Sadeghian, H.R. Dajani, A.D.C. Chan, Classification of speechevoked brainstem responses to English vowels, Speech Communication, 68, pp. 69-84, 2015.

[19] S. Gholami-Boroujeny, A. Fallatah, B.P. Heffernan, H.R. Dajani, Neural network-Based adaptive noise cancellation for enhancement of speech auditory brainstem responses, Signal, Image, and Video Processing. 10(2) pp. 389-395, 2016.

[20] M. Laroche, H.R. Dajani, F. Prvost, A.M. Marcoux, Brainstem auditory responses to resolved and unresolved harmonics of a synthetic vowel in quiet and noise, Ear and Hearing 34 (1), pp. 6374, 2013.

[21] R.C. Sprinthall, The normal Curve and z Scores, in Basic Statistical Analysis. Boston: Pearson Allyn & Bacon, 2012.

[22] O. Ozdamar, R.E. Delgado, Measurement of signal and noise characteristics in ongoing auditory brainstem response averaging, Annals of Biomedical Engineering, 24(6), pp. 702-715, 1996.

[23] P.D. Welch, The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms, IEEE Transactions on audio and electroacoustics, pp. 7079, 1967.

ADDENDUM

This section addresses some comments that were raised by the thesis committee after publishing this paper:

In: II.B.*1) Model of the noise part of the sABR:*

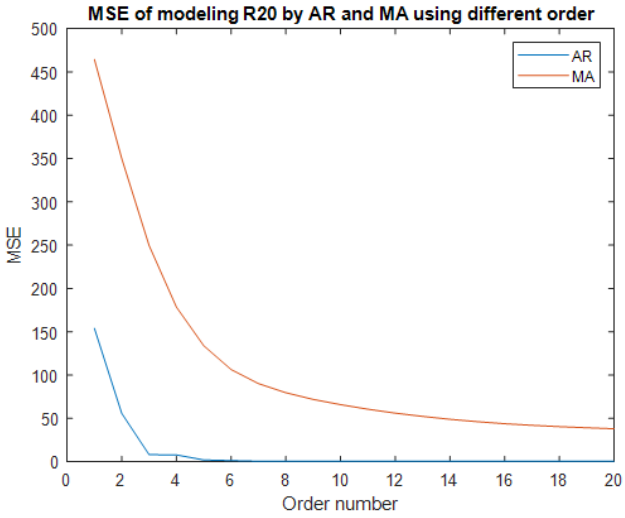
1. The Autoregression (AR) model is normally appropriate for modeling signals with peaks in its spectra whereas Moving Average (MA) is appropriate for signals with valleys in its spectra. Why is AR model used instead of MA?

While it is well-known that Autoregression (AR) modeling fits signals with peaks in their spectrum and Moving Average (MA) modeling fits signals with valleys in their spectrum, most researchers tend to use AR to model most natural behaviors such as the EEG.
The following paragraphs discuss:

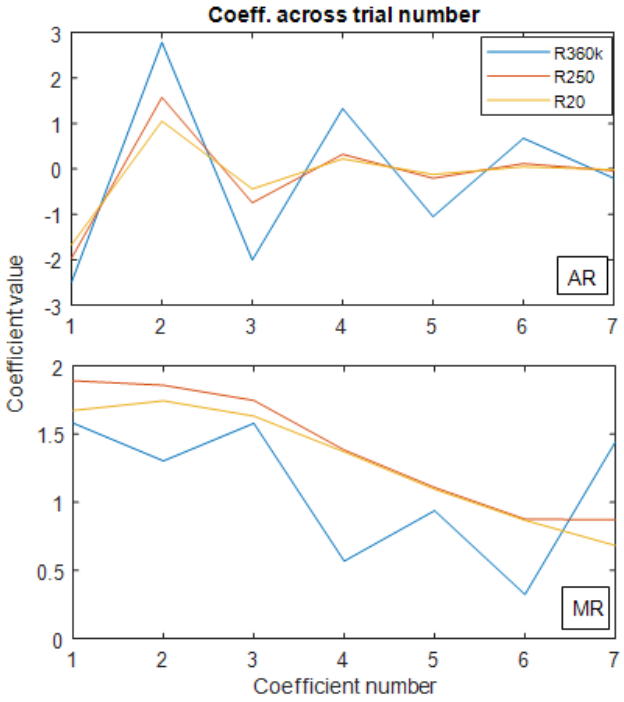
- whether it is better to model sABR using AR or MA modeling.
- our opinion as to why AR is preferred in the literature and in this research in particular.
- an experimental comparison between the two models of sABR.

Natural behaviors such as EEG and ABR are mostly modeled using AR [1,2] even if these signals do not have spectra with sharp peaks.
Possible reasons for this include:
1- MA is used to model signals with valleys, and these signals do not necessarily have valleys in their spectra.
2- AR modeling is less complex and easy to perform.
3- MA modeling has some conditions (e.g. invertibility) that make it more restricted than AR modeling.

In addition to the above-mentioned reason, our research shows that:
1- MA modeling of sABR yielded larger MSE compared to AR modeling with the same order. The following figure shows MSE between the human recorded sABR (i.e. R20) and the modeled data based on AR and MA:



2- MA modeling of several qualities of sABR is more consistent between coefficients compared to AR model. The following figure shows model coefficients of several qualities of sABR, and shows that AR coefficients are linearly dependent across trial number whereas MA coefficients are not. This property is important for estimating coefficients of other sABR qualities, as described in the paper regarding estimating the single-trial sABR.



In addition to the above discussion, it should be emphasized that in this work, AR is not used for modeling the whole signal (i.e. both response and noise) which has strong peaks in the spectrum, but it is only used for modeling the noise part of the spectrum which neither has strong peaks nor valleys.

References:

[1] J. Pardey, S. Roberts, L. Tarassenko, A review of parametric modelling techniques for EEG analysis, Medical Engineering and Physics, 8(1), pp. 2-11, 1996.

[2] M.I. Noorzi and I. Faye, A Review of EEG Signal Simulation Methods, Centre for Intelligent Signal and Imaging Research (CISIR), Springer, pp. 599–608, 2016.

2. In Eq.1, why are the past samples used for the regression written as white noise samples? Normally in the AR model synthesis, we have the sum of past model output samples each weighted by *a* coefficients, plus some white noise excitation input. But the past output samples are not white noise samples.

Eq.1 is adopted using whitening filter and AR general equation. However, after analysing this equation in more detail, we find that it is better to use the well-known AR general equation:

$$x_n = -\sum_{i=1}^{p} a_i x_{n-i} + v_i$$

where *x* is *AR* in Eq.1

3. In Table 1, normally *a*1 to *a*6 is used (and not *a*0 to *a*5) like in Eq. 1 for the AR coefficients, because there is an implicit *a*0=1.0 coefficient in some formulations of AR.

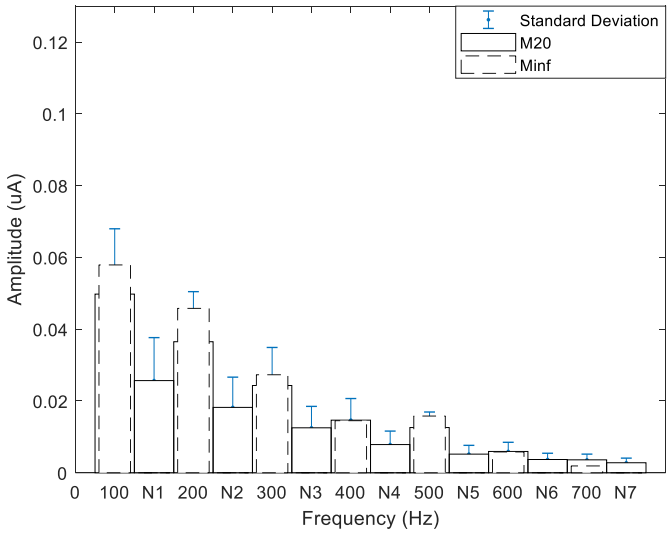
The coefficients listed in Table 1 are variable names and not the summation iterative variable "a_i" in Eq.1 (i.e. *an* not a_i, where *n* is from 0 to 5). However, it would have been to rename them according to Eq.1 in order to avoid confusion. This notation will be considered in our future research.

In: II.B.*2) Model of the clean sABR:*

4. The modeled data in this study are based on only one modeled evoked response, which is Minf, whereas in practice, each subject has his/her own evoked response. The modeled data must include different evoked responses to model the variability between different-subjects sABRs.

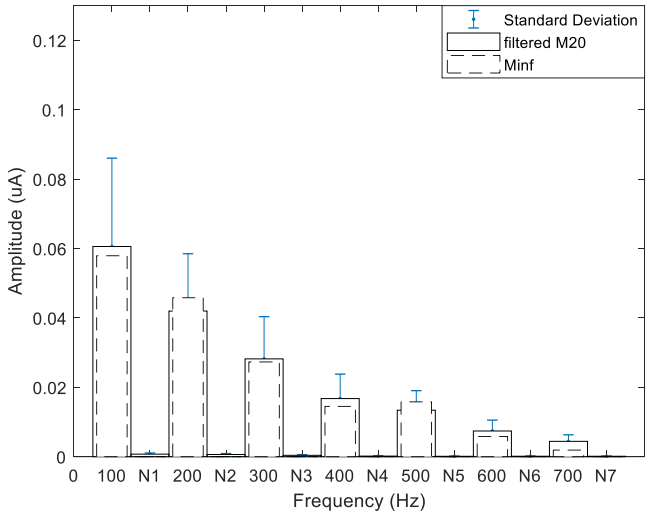
The main goal behind using a single evoked response in this study is to model the common evoked responses to the applied stimulus since this response (i.e. Minf) has been obtained from averaging 36,000 trials from 12 subjects. Therefore, the same evoked response must be obtained when averaging all modeled data. This study focuses on modeling the variability between different qualities of sABR since the ultimate goal is to estimate the single trial response. It may be of interest to model the variability between different subjects perhaps for other purposes, but this definitely needs a sufficient number of subjects that exceeds what is available in this study.

On the other hand, the generated model sABR captures the variability between different sABRs of the recorded data as seen in this paper. Moreover, even though only one Minf is applied to model the data, the resulting sABRs are unique. The following figures show the frequency domain content of M20.



where N# is the RMS of the noise amplitudes in between the harmonics of the evoked response, and the standard deviation is measured across M20.

This figure shows that the M20 mean are different than Minf, and there is variability between different modeled data. This is also confirmed in another study when these modeled data are used for assessing different enhancement methods (Fallatah and Dajani, in preparation). In that study, different filters are applied to recover the evoked response in the modeled data, and the obtained results always have different mean from Minf and different variance across the data. The following figure shows the frequency domain content of the M20 recovered evoked response.



Again, this figure shows amplitudes that are close to the original Minf, but these amplitudes still have different means and with variance across data which simulates the practical variance of the evoked response between sABR recorded in different individuals.

Reference:

A. Fallatah and H. Dajani. Assessment of Linear Optimal and Adaptive Filters for Enhancing Speech Auditory Brainstem Responses, paper in preparation.

In: *IV Conclusion:*
"the parameters of the single-trial" should be "the parameters of the modeled single-trial".

5. The Detection Paper

Accurate Detection of Speech Auditory Brainstem Responses using a Spectral Feature-based ANN Method

by Anwar Fallatah and Hilmi R. Dajani

in Biomedical Signal Processing and Control (BSPC), Vol. 44, pp.307-13, 2018.

Biomedical Signal Processing and Control 44 (2018 307-313)

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

Journal homepage: www.elsevier.com/locate/bspc

Accurate detection of speech auditory brainstem responses using a spectral feature-based ANN method

Anwar Fallatah*, Hilmi R. Dajani

School of Electrical Engineering and Computer Science, University of Ottawa, 800 King Edward Ave., Ottawa, Ontario, K1N 6N5, Canada

ARTICLE INFO

Article history:
Received 13 August 2017
Received in revised form 28 April 2018
Accepted 5 May 2018
Available online 21 May 2018

Keywords:
Speech auditory brainstem response
Detection
Artificial neural network
Mutual information
Discrete wavelet transform
Approximate entropy

ABSTRACT

The speech auditory brainstem response (sABR) is a promising tool that can be used for objectively assessing auditory function. The main problem in obtaining the sABR is the high background noise, especially noise associated with general brain activity. In practice, a very long recording is needed to detect the sABR. We therefore propose a new detection method of the sABR based on spectral feature extraction that will reduce the detection time without reducing the accuracy. This method involves a constructed feature-frequency vector fed to an artificial neural network. The performance of the proposed method is compared to four other methods reported in the literature: optimal linear filtering, online estimator, Mutual Information, and artificial neural network based on discrete wavelet transforms and approximate entropy. All the methods were evaluated with several datasets of recorded and simulated sABRs ranging from extremely noisy to relatively clean. The proposed method performed very well in terms of sensitivity, specificity, and overall accuracy in detecting the sABR, compared with the other methods The reduction in the required recording time promises to facilitate the application of this measurement technique in clinical settings.

1. Introduction

The speech auditory brainstem response (sABR) is a promising tool to objectively assess the function of the auditory system. In this approach, a speech stimulus is used to evoke the ABR, while the generated transient and steady-state components of the response can be used to study auditory function and diagnose different auditory disorders [1,2]. The sABR can also be potentially used for rehabilitation and hearing-aid fitting [3,4]. Often, the speech stimulus consists of a synthetic vowel with unchanging fundamental frequency and higher harmonics, and in that case, the steady-state components of the response are periodic, making them similar to the more widely studied auditory steady-state response (ASSR) [5]. Usually, the coherent averaging method is used to enhance the very weak targeted signals (measured in fractions of microvolts), and for detection, two statistical tools are mainly used in the literature [6], the F-test [5] and the Hotelling T-test [7]. These tests show whether a significant ABR frequency component is present in the obtained electroencephalogram (EEG) signal which includes background brain activity and environmental noise.

For general signals, several kinds of detection methods are proposed in the literature, and these can be categorized into linear and nonlinear methods. Linear optimal filters have been well investigated in several studies [8,9], and a number of nonlinear online detection methods have been published recently [10–12]. However, there has been little work on applying nonlinear methods with sABR signals. Our team has investigated the use of the artificial neural network (ANN) to adopt linear filters for enhancing sABR signals and showed that this approach outperforms conventional adaptive filters for this application [13]. Several studies have investigated ANNs in detecting other kinds of physiological information from the EEG. For instance, ANN-based methods have shown promising performance in detecting epileptic seizures [14,15].

Linear optimal filters have been used for enhancing [16] and detecting the sABR and the ASSR [8,9]. In Wiener filter detection, for instance, optimal coefficients are tuned using a clean response with very high signal-to-noise ratio (SNR). Then this filter is applied to the rest of the recorded low SNR signals. In [9] for example, optimal coefficients for each subject were created separately, and this method is reported to reduce the recording time and increase the sensitivity by 15% over unfiltered data. However, even with these improvements, the recording time required for each subject remained long. The Wavelet transform is another method that has been introduced for target signal detection in the EEG [17].

https://doi.org/10.1016/j.bspc.2018.05.007
1746-8094/

However, use of the Wavelet transform to improve the performance increases the time and complexity of the detection process.

Two recent papers have also applied a well-known tool in information theory, Mutual Information (MI), in detecting and classifying ABRs [18,19]. This powerful tool is typically used to measure the shared information between the spectrum of the stimulus and the observed response signal to obtain MI values that can be used both as a detector or as a classifier with a great reduction in processing time and complexity compared with other methods. However, this fast processing comes at the cost of the detection performance.

In this paper, the recorded time-domain data is transformed into the frequency domain, a smaller dimension space that maintains important features of the original signal. The next logical step for the case of sABR would then be to narrow the frequency domain representation to include the main targeted set of frequencies and remove any redundancies. We take a further step and investigate ANN performance in detecting sABRs instead of using it only to adapt the linear filter as done in our previous work [13]. We propose a constructed frequency-domain feature vector as an input to the feed-forward ANN. This achieves a reduction in processing time and higher accuracy by taking advantage of the parallel processing and the nonlinearity inherent in the ANN.

The performance of the proposed non-linear filter (denoted by ANN FF) is evaluated by comparing it with four other approaches, linear filtering (LF), online estimator (OE), Mutual Information (MI), and the ANN based on Discrete Wavelet Transforms and Approximate Entropy (ANN DA). In addition, the results are compared with those obtained from the original coherently averaged unfiltered data (UF). A Receiver Operating Characteristic (ROC) curve is constructed for all the applied algorithms and we evaluate performance by comparing sensitivity, specificity, overall accuracy, and the processing time needed to obtain the final detection decision.

2. Methods

In sABR measurements, recordings are synchronized to repeated stimulus presentations and then the collected responses are coherently averaged to increase the SNR. We used recorded signals based on averages of 20, 250, and 36000 trials. Also, as described below, we created modeled signals, corresponding to different SNR levels, to test the algorithms. Implementations were performed in MATLAB (The MathWorks, Natick, MA).

2.1. Experimental methodology and recorded responses

Subjects were asked to sit and wear insert earphones which deliver the acoustic stimulus. Three electrodes were placed on the scalp, a recording one at the vertex (Cz), a reference on the right earlobe, and a ground on the left earlobe. The speech stimulus was a synthetic vowel/a/with a fundamental frequency (F0) at 100 Hz and first three formants at 700, 1220 and 2600 Hz (shown in Fig. 1). It was presented in sequentially alternating polarity at a level of 80 dB SPL, and the responses to multiple trials were coherently averaged to give the sABR known as the envelope frequency response. This response follows the periodicity of the envelope of the signal, and so it also has a fundamental frequency of 100 Hz and some higher harmonics at multiples of F0. The data from human subjects were collected in the course of previous studies [13,20,21] that were approved by the Research Ethics Board of the University of Ottawa.

There are three sets of recorded data: 1) R20: 75 blocks based on the coherent average of 20 responses from five subjects, 2) R250: 36 blocks based on the coherent average of 250 responses gathered from four subjects, and 3) R36k: one very high SNR block based on the coherent average of 36000 responses from twelve subjects (including the R250 and R20 responses). Each response sweep was recorded over 0.3198 s and was sampled at 3202 Hz, giving 1024 samples per sweep. Data was also recorded without any stimulus presentation giving an EEG signal containing no response (denoted as EEG). Fig. 2 shows the spectrum of this EEG signal that contains no response. Table 1 summarizes all gathered data and their details and Fig. 3 shows R36k and the average of R250 and R20 in time domain (top left) and frequency domain (top right).

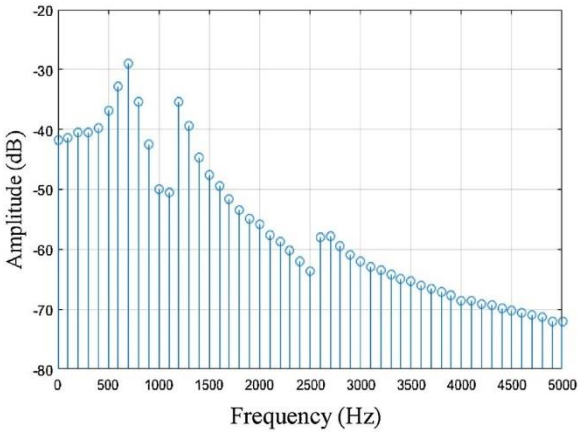


Fig. 1. The spectrum of the applied synthetic vowel/a/stimulus with a fundamental frequency at 100 Hz, and first three formants at 700, 1220 and 2600 Hz. Note that the amplitudes shown in the figure are not calibrated to sound level.

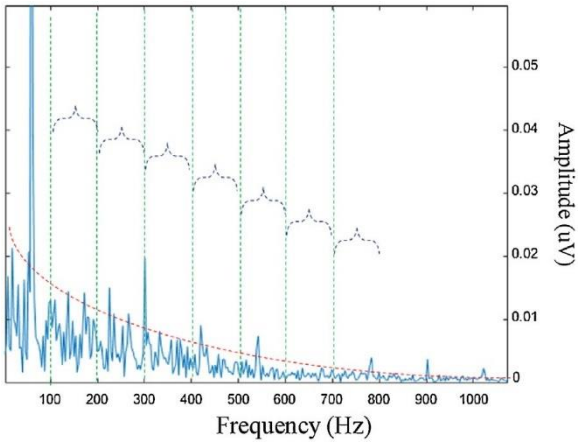


Fig. 2. The spectrum of a sample of EEG noise with the selected frequency features: the fundamental and its first 6 harmonics (dashed green), noise bands (dashed brackets), and the peaks above the threshold curve for the significant amplitudes (dashed red line). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2.2. Modeled signals

The main propose of creating modeled signals is to expand the tests to include different signal "qualities" in performance comparisons. Different qualities can represent different SNR levels in the response, different stimulus qualities (such as SNR changes in the stimulus), or different trial numbers as mentioned above in the description of the datasets. The modeled sABR signal is created by obtaining the signal response spectrum from R36k (the very high SNR sABR obtained by averaging 36000 trials), and only retaining components at the fundamental frequency and its harmonics. This signal is denoted by MInf (Fig. 3, middle). Its corresponding modeled noise (denoted as N36k) is created by passing white noise through an Autoregressive (AR) filter of order six (Table 2). These

Table 1
Datasets used in the study.

Symbol	description	Details	SNR (dB)
R36k	sABR with 36000 trials	1 block from 12 subjects	24.8
R250	sABR with 250 trials	36 blocks from 4 subjects	7 (3–12)
R20	sABR with 20 trials	75 blocks from 5 subjects	−1 (−8 to 5)
EEG	EEG with no response	1 block with 60 Hz harmonics	N/A
MInf	Modeled clean sABR	1 block, has sABR signal spectrum only	Infinity
N36k	Noise present in 36000 sABR	100 blocks, modeled by AR of order 6	N/A
M36k	Modeled 36000 trials	100 blocks, MInf + N36k	27 (25.8–28.5)
N250	Noise of 250 trials	100 blocks, modeled by AR of order 6	N/A
M250	Modeled 250 trials	100 blocks, MInf + N250	7 (6–8)
N20	Noise of 20 trials	100 blocks, modeled by AR of order 6	N/A
M20	Modeled 20 trials	100 blocks, MInf + N20	−2 (−4 to −1)
N01	Noise of 1 trial (estimated)	100 blocks, modeled by AR of order 6	N/A
M01	Modeled 1 trial	100 blocks, MInf + N01	−13 (−15 to −12)

Table 2

Coefficients	N36k	N250	N20	N01
$\mathbf{a}_0$	1.66	1.82	1.71	1.7
$\mathbf{a}_1$	0.858	1.27	1.27	1.06
$\mathbf{a}_2$	−0.0548	−0.473	−0.473	−0.374
$\mathbf{a}_3$	−0.0935	0.135	0.174	0.188
$\mathbf{a}_4$	−0.117	−0.0675	−0.0807	−0.0818
$\mathbf{a}_5$	0.0823	0.00557	−0.00134	−0.00396

Autoregressive (AR) Modeling Coefficients Used to Generate Each Type of noise in the Top Row.

two created signals (MInf and N36k) are added together to form the signal denoted by M36k.

The second modeled signal (denoted as M250) represents the 250-trial sABR and is created by combining the M36k with noise also modeled by an AR filter of order 6 but with different coefficients (Table 2). The 20-trial modeled signals (denoted as M20) is created by combining M36k with noise also modeled by another AR filter. An extra modeled signal (denoted as M01) is created to represent a very low SNR signal that is used as a very challenging task. Table 1 summarizes all modeled data and their characteristics. Fig. 3 (middle and bottom) shows all modeled signals in time and frequency domains.

2.3. The proposed ANN frequency features (ANN FF) algorithm

There are three main steps in the proposed ANN FF detector:

1. Obtain the Discrete Fourier transform (DFT)
2. Extract the main features in the frequency domain
3. Train the Feed-Forward ANN using backpropagation algorithm

In the first step, the 1024 samples of the observed signal are converted into 512-bins in the frequency domain. Then this number is reduced to 28 frequency features.

This can be achieved by several statistical tools used to measure irregularity (such as approximate entropy, or ApEn). However, we propose our own method that is simpler and only focuses on locating three types of frequency features (shown for the EEG noise signal in Fig. 2): 1. The 14 most statistically significant amplitudes. To achieve this, we obtain a detection threshold (dashed red line in Fig. 2) from a 100-point moving average of the spectrum shifted upward to retain the 14 most significant amplitudes (which correspond to the highest 2.7% components in the spectrum).

2. The components that represent amplitudes at the stimulus fundamental and 6 harmonic frequencies (7 components at the dashed vertical green lines in Fig. 2). The highest frequency harmonic considered is chosen to be 700 Hz because the envelope

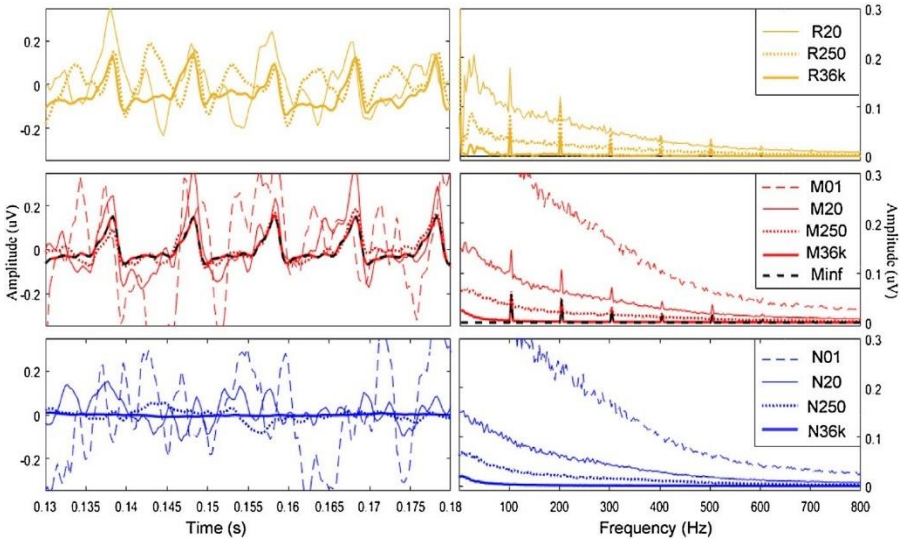


Fig. 3. The time-series (shown over 50 ms) and frequency domain content of the different data types used in this study. Top: The recorded speech auditory brainstem responses (sABR) R36k, and average of R250 and R20. Middle: The modeled sABR signals Minf, M36k, M250, M20 and M01 (note that Minf and M36k overlap significantly). Bottom, the modeled noise signals N36k, N250, N20 and N01.

following response is typically very small above this frequency. Also, a component may be recognized twice, here and in (1).

3. Features that represent the root-mean-square of amplitudes in different noise bands (7 bands − the dashed brackets).

These three feature sets are combined and used as input to a feed-forward multilayer perceptron neural network with one input layer, a single hidden layer with 5 neurons, and one output layer. The neural network has a single binary output to represent 0 if only noise is detected or 1 if sABR signal is detected

The multilayer perceptron neural network is described by:

$$y = f_2\left[\sum_{j=0}^{H} w_2(j).f_1\left[\sum_{i=0}^{M} w_1(j,i).x(i)\right]\right]$$

where x is the obtained features vector, w_2 is the hidden weights matrix, w_2 is the output weights vector, and f is the applied activation function (sigmod function based on the hyperbolic tangent for f_1 and log function for f_2).

The neural network is trained using back-propagation. K-fold cross-validation method (discussed below) is employed so that the entire dataset is tested for presence of sABR signal, and the possibility for overfitting is minimized.

2.4. Other applied algorithms

2.4.1. Optimal linear filtering (LF)

This linear detection method has been presented in the literature in several ways, including the optimal Wiener Filter [8] and maximum SNR Filter [9]. These detection methods use optimal filters to increase the SNR in noisy signals. To obtain the optimal filter coefficients, a high SNR sABR is needed for tuning the linear filter, which is then applied to any further measurements with similar noise characteristics (e.g. from same subjects). We have applied the generalized eigenvector SNR optimal algorithm that is described in [9] to maximize the SNR of any given sABR. For robust detection, the obtained filtered signal and the original signal have to pass the threshold (obtained by F-value as discussed in Section 2.5). In this application, we obtain the optimal filter coefficients for each block individually to increase the detection ability.

2.4.2. Online estimator (OE)

This method is based on amplitude and frequency estimation [11] which is suitable for online detection of a signal with unknown parameters, since it starts from an initial value (usually zero) and iterates to find any periodic signal. This approach is achieved using a number of derivative equations of the targeted signal that are used to obtain the frequency and amplitude of any detected periodic signal [10,11]. Since we are not interested in obtaining the original signal amplitudes, we reduce these equations to only obtain the frequencies and then use the F-test to measure the significance level. An online simulation method is needed since our recorded signals have limited duration (which is 0.3198 s). We therefore applied this filter only on the modeled signals since we can extend them to 1.28 s, a duration which allows reasonable results. Fig. 4 shows how this method increases the F-value of the challenging task M01(but not with N01), as a function of time, and so how it performs well even with this challenging task.

2.4.3. Mutual information (MI)

This algorithm is applied to the frequency representation to measure the probability of the shared information between the clean reference and the targeted signal [18,19]. We use a scaled copy of MInf as a reference to measure both the joint and marginal probability in 7 frequency bins at the stimulus fundamental and its harmonic frequencies (100, 200, . . ., 700 Hz) and in 7 noise bands

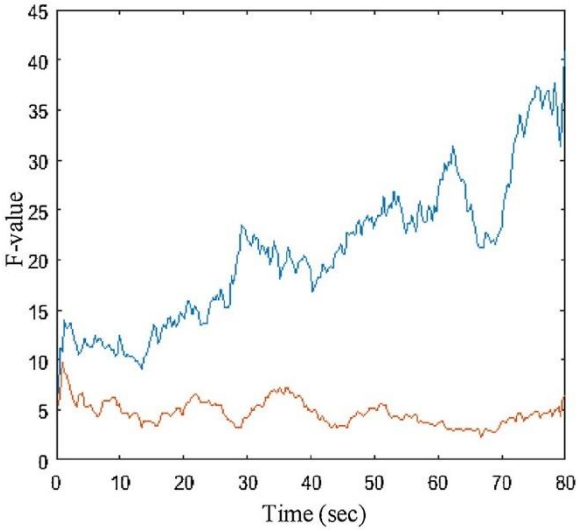


Fig. 4. F-value obtained with the challenging task M01 (the blue line) and the noise signal N01 (the red line) using the online estimator. The x-axis indicates the recording time needed to obtain a given F-value. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(based on the root-mean-square amplitude from 100–200 Hz, 200–300 Hz. . .up to 700–800 Hz, and excluding the harmonic frequencies themselves). The resulting MI values are different depending on the sABR quality (at 36k trials, 250 trials, 20 trials, and 1 trial) as expected since they have different shared information with the stimulus. A normalization step (between 0 and 1) and thresholding at 0.5 are also performed for detection.

2.4.4. ANN based on DWT and ApEn (ANN DA)

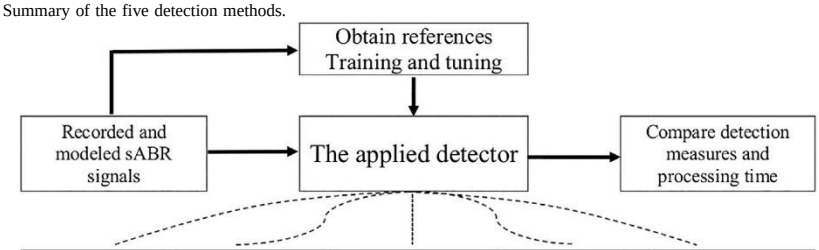
The last method used for comparison is an ANN based on Discrete Wavelet Transforms (DWT) and Approximate Entropy (ApEn) feature extraction method.

A recent study using this method shows high sensitivity and specificity for detecting and classifying several types of epileptic seizures from EEG signals [15]. We have introduced this method to compare it to our proposed method since it also employs an ANN and differs only in the feature extraction approach. In this method, we use DWT to decompose sABRs into several smaller blocks created using different window sizes that aim to obtain more detailed information in low and high frequencies. From each sABR signal, 6 other blocks are obtained with different number of samples (by dividing by 2 with each level down). ApEn then is used to obtain a number from each of the DWT blocks which represents the irregularity in this windowed signal. This is done by setting the ApEn window to 0.95 times the DWT size and the threshold to 1. To increase ApEn performance, all its inputs are rescaled to have a variance of 5 and then all long DWT windows (greater than 45 samples) are shortened by synchronized averaging (by dividing the window into smaller synchronized parts with equal length) to maintain the regularity in the signal and reduce the processing time. The obtained 6 numbers are then fed to an ANN for training and testing. The same network structure and k-fold crossvalidation training approach used for ANN FF method are also used here. Further details about this approach are found in the original paper [15]. Table 3 summarizes the five applied methods.

2.5. Cross-validation for ANN training

The main reason to apply this validation method in ANN training is to include the entire dataset in the testing phase, while at the same time avoiding overfitting by keeping the training, validation,

Table 3
Summary of the five detection methods.



Linear Filter (LF)	Online Estimator (OE)	Mutual Information (MI)	ANN DWT and ApEn (ANN DA)	ANN Frequency Features (ANN FF)
Obtain the optimal filter coefficients that maximize SNR. (100 tap FIR filter)	Longer modeled ABRs as input with (1024x4 samples each)	Calculate MI value between MInf and the rest of sABR data	Apply DWT (6 blocks with length of 40 to 512 samples)	Obtain DFT (512 bins)
Apply the filter	Initialize all variables and start iterations	Normalize the results	Obtain ApEn features (6 values)	Extract frequency features (28 values)
Measure F-value	Measure F-value	Apply the detection threshold at 0.5	Train the ANN and apply the detection threshold	Train the ANN and apply the detection threshold

and testing sets separate. K-fold cross-validation method is applied as follows:

1. Select k to represent 33% of the total number of the data.
2. Divide the entire data into three subgroups (training, validation, and testing sets) selected randomly each with the size of k.
3. Start the training procedure using the training and validation sets, and obtain performance using testing set.
4. Alternate among groups, so the data used for training will be used for validation or testing set and so on.
5. Repeat the training process (in step 3) with each different group of sets (of step 4) giving a total of 6 different training instances.
6. Average the final results over the number of the testing instances for each signal.

2.6. Detection measurements

One effective way to measure the significance level in sABR is the F-test [6]. The F-value can be calculated by:

$$\text{F-value} = J\frac{a_s^2}{\sum a_n^2}$$

where a_s is the signal amplitude, a_n represents the J bins of noise amplitudes around a_s, and the degrees of freedom = (2,2J-1). The amplitude at the fundamental frequency of 100 Hz is selected for a_s, and 31 bins after the fundamental frequency of the stimulus, up to but not including 200 Hz, are selected for a_n. Frequencies lower than 100 Hz are not included to represent the noise in order to avoid large amplitude non-auditory related EEG background activity. The F-test is not applied in the case where the output was normalized between 0 and 1 (in the MI and ANN methods). In those cases, the detection threshold is specified at 0.5.

2.7. Comparison parameters

The five detectors are compared using the following parameters:

1. Correctly identified responses.
2. Detection sensitivity.
3. Detection specificity.
4. Detection accuracy.
5. Processing time.

Correctly identified responses are the percentage of samples that a detection method is able to identify containing the stimulus response in case of sABR (i.e. the true positives) or not in case of only noise (i.e. the true negatives). The percentage is calculated by dividing by the total number of samples found in this dataset. The detection sensitivity is the number of correctly detected positive samples (true positives) divided by the total number of positive samples, and the detection specificity is the number of correctly detected negative samples (true negatives) divided by the total number of negative samples, where positive and negative samples refer to the presence and absence of a response, respectively. The accuracy is the total number of the true positives and true negatives divided by the total number of the samples. The performance of each method in the identification of stimulus response or noise was compared to that of the other methods using McNemar's test with Bonferroni corrections for multiple comparisons. The results of the comparisons between each method and the method that is immediately lower in terms of accuracy are reported as p-values in brackets.

The processing time represents how long it takes to process all sABR signals and obtain the final detection decision. This parameter was obtained using the MATLAB function "timeit" on a Windows PC with a 2.16 GHz 64-bit processor. The time to train the ANN is excluded from the total processing time since this is performed one time only at the beginning of the filter design.

3. Results

Table 4 summarizes the overall percentage of correct identification obtained using the five different methods, and that obtained with unfiltered coherently averaged signals.

The detection parameters (sensitivity, specificity, and accuracy) are shown in Table 5, and the ROC curves obtained with the unfiltered data and the five methods tested in this paper are shown in Fig. 5. Based on these results the proposed method has the best performance ($p < 0.001$) followed by ANN DWT and ApEn, Mutual Information, Online Estimator, Linear filter, and finally the unfiltered coherently averaged signals. On the other hand, MI requires the least processing time whereas ANN DWT and ApEn requires the longest time. Overall, the proposed method has the best performance with low processing time.

Table4
Correct Identification (%) of the Tested Detection Methods.

	CorrectlyIdentified(%)					
	Unfiltered	LinearFilter	OnlineEstimator	MutualInformation	ANNDA	ANNFF
R36k	100	100	N/A	100	100	100
R250	52.8	72	N/A	100	81	92
R20	60	73	N/A	61.3	80	87
EEG	100	100	N/A	100	100	100
N36k	82	83	74	99	100	100
M36k	100	100	100	100	100	100
N250	83	83	80	92	93	99
M250	67	85	100	94	97	91
N20	83	84	81	75	70	91
M20	49	68	98	85	86	96
N01	84	84	79	59	45	65
M01	11	29	26	59	68	57

Table5
Performance of the tested methods over all datasets.

Method	Sensitivity	Specificity	Accuracy	Processing time (sec)
Unfiltered	0.57	0.83	**0.68**	0.047
Linear filter	0.71	0.84	**0.77**	4.7
Online Estimator	0.81	0.78	**0.79**	2.8
Mutual Information	0.82	0.81	**0.81**	0.46
ANNDA	0.86	0.77	**0.82**	45.2
ANNFF	0.87	0.89	**0.88**	0.85

4. Discussion

4.1. Optimal linear filtering (LF) [9]

The results obtained with LF were slightly better than the unfiltered signals in terms of sensitivity (as also shown in [9]). However, after combining it with the F-value of the unfiltered sABR (i.e. where at least one of them has to pass the threshold for detection), the overall algorithm behavior improves, as shown in Table 5. In particular, the performance of LF is better than that obtained with unfiltered signals ($p < 0.001$). Linear filtering requires prior knowledge of the optimum filter coefficients for each subject (or each sABR dataset, such as R250, R20 etc.). This can be obtained from a clean sABR, but collecting this for a given subject or class of subjects is time consuming and defeats the purpose of reducing recording time.

4.2. Online estimator (OE) [11]

The performance of OE is better than that obtained with LF ($p < 0.01$). Online estimator is one of the best methods to detect periodic signals with unknown parameters, and for robust processing of a very low SNR signal. Moreover, it has less complexity than the optimal linear filter. However, it needs longer signal duration time (in our case, at least 1.2 s) to have acceptable performance. Table 4 shows the performance of this method with modeled sABR only and not with recorded sABRs because of the short duration of the recorded sweeps (around 0.3 s). With the modeled signals, this method performs well with low SNR signals (such as M20 and M01).

4.3. Mutual information (MI) [19]

The performance of MI is better than that obtained with OE ($p < 0.001$). The Mutual Information method has lower complexity than the other evaluated methods since it only processes 14 spectral bins of the tested signal (half of what the proposed method is using), which leads to a large reduction in the processing time. MI exhibits a trade-off between the processing time and detection accuracy, and as more information is available for matching, both the accuracy and processing time would increase. It should be noted that the obtained results are based on applying one reference signal and a global threshold assigned for all sABR datasets (as described in Methods section). Changing any of these factors to fit each dataset separately could enhance the results.

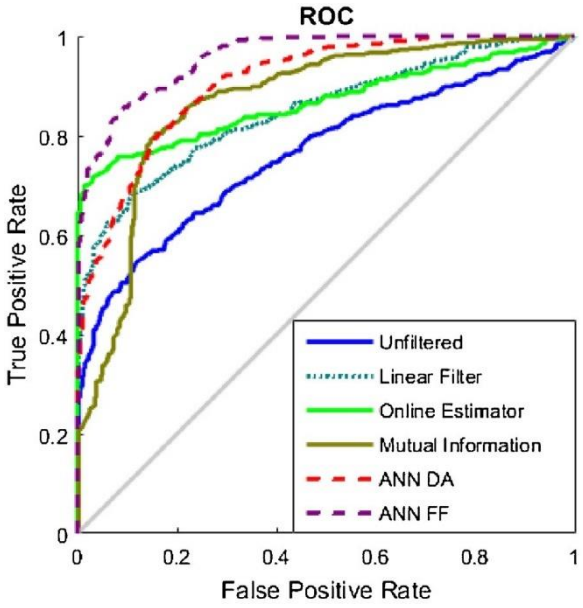


Fig. 5. ROC curves obtained with the unfiltered data and the five methods tested in this paper.

4.4. ANN based on DWT and ApEn (ANN DA) [15]

There was no statistically significant difference between the performance of ANN DA and MI. The feature extraction method by DWT and ApEn provides sABR features which can be considered comparable to those obtained with the proposed feature extraction method. Therefore, its performance is expected to be close to that of the proposed method, and this is what the results show (Tables 4 and 5). The difference in the results obtained with the two methods could be due to the limitation of ApEn in focusing on the signal regularity which is easily disturbed by noise. In addition, the tested signals have a wide range of variance across different sABR qualities, and across different DWT decompositions that would not be suitable for a fixed setting of ApEn. Moreover, it should be noted that a major disadvantage of the method is the very long processing time.

4.5. ANN based on frequency features (ANN FF)

The performance of the proposed method (ANN FF) is better than that obtained with ANN DA ($p < 0.001$). The proposed method has the best detection performance, as also seen by the purple dashed ROC curve in Fig. 5, with a fast and less complex algorithm compared to the other evaluated approaches.

5. Conclusion and future work

We have proposed a spectral extraction method for detection of speech ABRs that selects frequency features and processes them with an ANN. This method performs better, with high sensitivity, specificity, and accuracy, compared to the traditional coherent averaging method and others available in literature. It also requires low processing time.

There are many types of features that can be obtained from a time-series or from the frequency domain, but they can be divided into two categories: numerical and symbolic features. Our proposed method is an example of the use of numerical features, where every frequency feature is represented by a number (i.e. the amplitude in uV). However, it is also possible to provide a non-numerical description of physical data, which is often preferred in clinical diagnosis. For example, a recent study achieved 99% accuracy in identifying four different types of hearing conditions from transient ABRs using a symbolic representation of the time-series [22]. Given this reported performance, including meaningful symbolic representation will be our next target for detecting or classifying speech ABRs.

Conflict of interest statement

The authors declare that there is no conflict of interest regarding the publication of this article.

Acknowledgement

Partial support for this work was provided by the Natural Sciences and Engineering Research Council of Canada.

References

[1] K.L. Johnson, T.G. Nicol, N. Kraus, Brain Stem Response to Speech: a biological marker of auditory processing, Ear Hear. 26 (5) (2005) 424–434.

[2] H.R. Dajani, D. Purcell, W. Wong, H. Kunov, T.W. Picton, 'Recording human evoked potentials that follow the pitch contour of a natural vowel,', IEEE Trans. Biomed. Eng. vol. 52 (no.9) (2005) 1614–1618.

[3] S. Anderson, N. Kraus, The potential role of the cABR in the assessment and management of hearing impairment, Int. J. Otolaryngol. 1–10 (2013) 2013 (Article ID 604729).

[4] H.R. Dajani, B.P. Heffernan, C. Giguère, Improving hearing aid fitting using the speech-evoked auditory brainstem response, 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (2013) 2812–2815.

[5] M.S. John, T.W. Picton, MASTER: a Windows program for recording multiple auditory steady-state responses, Comput. Methods Programs Biomed. 61 (2000) 125–150.

[6] R.A. Dobie, M.J. Wilson, A comparison of t-test, F test, and coherence methods of detecting steady-state auditory evoked potentials, distortion-product otoacoustic emissions, or other sinusoids, J. Acoustic Soc. Am. 100 (1996) 2236–2246.

[7] T.W. Picton, J. Vajsar, R. Rodriguez, K.B. Campbell, Reliability estimates for steady-state evoked potentials, Electroencephalogr. Clin. Neurophysiol. 68 (2) (1987) 119–131.

[8] B. Van Dun, J. Wouters, M. Moonen, Multi-channel wiener filtering based auditory steady-state response detection, IEEE International Conference on Acoustics Speech and Signal Processing – ICASSP 2 (2007) 929–932.

[9] W. Biesmans, A. Bertrand, J. Wouters, M. Moonen, Optimal spatial filtering for auditory steady-state response detection using high-density EEG, IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP) 8 (2015) 857–861.

[10] D.M. McNamara, A.K. Ziarani, A new adaptive technique of estimation of steady state auditory evoked potentials, Eng, Medicine and Biology Society (IEMBS) International Conference of the IEEE (2004) 4544–4547.

[11] F.L.A. Cheah, M. Hou, Real-time detection of auditory steady-state responses, Annual International Conference of the IEEE Engineering in Medicine and Biology (2010) 1382–1385.

[12] M. Hou, Amplitude and frequency estimator of a sinusoid, IEEE Trans. Autom. Control 50 (6) (2005) 855–858.

[13] S. Gholami-Boroujeny, A. Fallatah, B.P. Heffernan, H.R. Dajani, Neural network-Based adaptive noise cancellation for enhancement of speech auditory brainstem responses, Signal, Image Video Process. 10 (2) (2016) 389–395.

[14] V. Srinivasan, C. Eswaran, N. Sriraam, Approximate entropy-Based epileptic EEG detection using artificial neural networks, IEEE Trans. Inf. Technol. Biomed. 11 (3) (2007) 288–295.

[15] Y. Kumar, M.L. Dewal, R.S. Anand, Epileptic seizures detection in EEG using DWT-Based ApEn and artificial neural network, signal, Image Video Process. 8 (7) (2014) 1323–1334.

[16] A. Fallatah, Enhancement of Speech Auditory Brainstem Responses Using Adaptive Filters, M.A.Sc. Thesis, Ottawa, Canada, 2012.

[17] L.M. Patnaik, O.K. Manyam, Epileptic EEG detection using neural networks and post-Classification, Comput. Methods Progr. Bomed. 91 (2) (2008) 100–109.

[18] G.M. Bidelman, Objective information-Theoretic algorithm for detecting brainstem-Evoked responses to complex stimuli, J. Am. Acad. Audiol. 25 (8) (2014) 715–726.

[19] G.M. Bidelman, S.P. Bhagat, Objective detection of auditory steady-state based on mutual information, Int. J. Audiol. 55 (1) (2016) 313–319.

[20] M. Laroche, H.R. Dajani, F. Prévost, A.M. Marcoux, Brainstem auditory responses to resolved and unresolved harmonics of a synthetic vowel in quiet and noise, Ear Hear. 34 (1) (2013) 63–74.

[21] A. Sadeghian, H.R. Dajani, A.D.C. Chan, Classification of speech-evoked brainstem responses to English vowels, Speech Commun. 68 (2015) 69–84.

[22] M.E. Molina, A. Perez, J.P. Valente, Classification of auditory brainstem responses through symbolic pattern discovery, Artif. Intell. Med. 70 (2016) 12–30.

ADDENDUM

This section addresses some comments that were raised by the thesis committee after publishing this paper:

In: *2.1. Experimental methodology and recorded responses:*

1. There is a strong 60Hz component in Fig.2. Does this component exist in the recorded and the modeled sABR?

The 60Hz component shown in Fig. 2 occurred during the recording of EEG noise when implementing the experimental set-up early on. By adjusting the wiring and the position of the different devices used in the experimental set-up, the 60 Hz component was largely suppressed. As a result, this signal is the only one that has a strong 60Hz component in this study. The modeled sABR is meant to simulate the recorded data, and therefore, it does not contain this component either.

In: *2.2. Modeled signals:*

2. The modeled data are based on coefficients listed in Table 2 which are used to generate both the training and the testing dataset. Coefficients that have been used to generate the training dataset should not be used to generate the testing dataset.

The modeled data is meant to simulate the recorded data. Therefore, the coefficients listed in Table 2 are generated using the recorded data. The variability that exists in the recorded data is found in the modeled data (we have published a paper of how this model represents the recorded data well, Fallatah, and Dajani, 2018), and since the recorded data are fed to the ANN without any separation based on the data characteristics, the modeled data are used for training and testing the ANN without further consideration of their coefficients. The following response addresses the separation of data based on their subjects which includes performance of training the ANN based on all modeled data and testing based on all recorded data.

Reference:

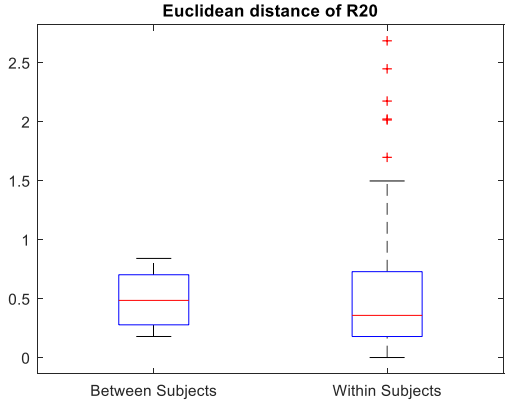
Fallatah, A and Dajani HR. Estimating the properties of the single-trial speech auditory brainstem response using an accurate AR model, Proceedings of the 14th International Conference on Signal Image Technology and Internet-based Systems (SITIS 2018), Las Palmas, Spain, Nov. 26-29, 2018.

In: *2.3. The proposed ANN frequency features (ANN FF) algorithm:*

3. The reported results are based on ANN that has been trained and tested using same subjects. Is it possible that the high performance of the proposed method is due to training and testing the ANN using data from same subjects?

The reported results of the ANN are obtained using a training dataset that is randomly separated from the testing set. The separation between the training and testing datasets was not based on subjects (not subject-based division). This is because the number of subjects in the study was limited (total of 9 subjects), and there is no recorded data for EEG noise in each subject for testing the true negative condition.

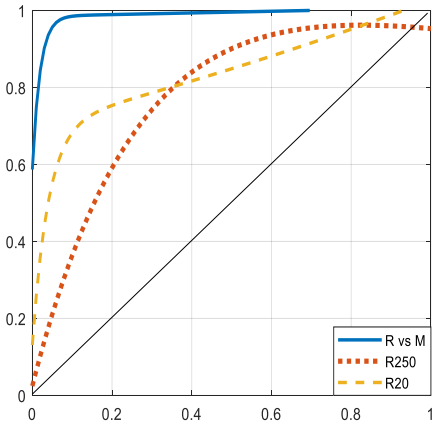
The Euclidean distance is used to measure the variance of between-subjects and within-subjects of the proposed spectral-feature vectors. The following figure shows the boxplots of Euclidean distance of spectral features of R20 between and within subjects which have means with relatively small difference. This indicates that the ANN will not suffer much from not having subject-based separation of the training and testing dataset.



To confirm the last conclusion, three diffreent training configrations of subject-based separation (i.e. some subjects are selected for training while the rest are used for testing) were investigated:

1. Recorded vs Modeled (RvsM) configration: The ANN is trained using only the modeled date (i.e. M01,M20,M250, and M36k), and then it is tested using the recorded data (R20,R250, and R36k).
2. R250-subjects configuration: The ANN is trained using 75% of R250 subjects (i.e. 3 subjects), and then it is tested on the rest of R250 subjects.
3. R20-subjects configration: The ANN is trained using 80% of R20 subjects (i.e. 4 subjects), and then it is tested on the rest of R20 subjects.

The following figure shows the detection performance of the three training configurations:



As seen in the performance plot, there is a relation between the accuracy and the number of subjects that are included in the training set. The high performance of R vs M is due to including all qualities in the training set (with total of 400 blocks). R20 is better than R250 performance since it has a greater number of subjects (5 subjects for R20 and 4 for R250). This confirms that the ANN FF results plotted in Figure 5 in the paper will not suffer from having subject-based separation in training the ANN.

In: *2.5. Cross-validation for ANN training:*

4. The percentages mentioned for cross-validation methods should also be presented in terms of number of subjects.

As mentioned in the previous response, there is no subject-based separation in the published paper (but this question is addressed in the response above). Percentages mentioned in cross-validation section are from the full dataset. The total number of data is 914 blocks, and 33% of them is equal to 301 blocks.

In: *3. Results-Table 5:*

5. The processing time in of some methods such as linear filter and ANN DA is longer than expected.

The applied linear filter algorithm is the one described in [8] which is more complex than basic well-known linear filters such as Wiener filter (WF). This also applies to ANN DA that is described in [15]. Implementing the conventional algorithm of WF or DWT would result in faster processing but at the cost of degrading the performance. However, note that in practice accuracy is a more important measure than processing time, since the processing of individual blocks, which are 319.8 ms long, is possible in real-time with all the tested methods.

6. The Enhancement Paper

Assessment of Linear Optimal and Adaptive Filters for Enhancing Speech Auditory Brainstem Responses

by Anwar Fallatah and Hilmi R. Dajani

Journal paper in preparation

Assessment of Linear Optimal and Adaptive Filters for Enhancing Speech Auditory Brainstem Responses

Anwar Fallatah, Hilmi Dajani

School of Electrical Engineering and Computer Science, University of Ottawa, 800 King Edward Ave., Ottawa, Ontario, Canada, K1N 6N5

ABSTRACT

The speech Auditory Brainstem Response (sABR) is a promising tool that can be used to objectively assess the auditory system. Due to the high background noise which includes general brain activity, the sABR is usually obtained by coherent averaging over many stimulus repetitions. No alternative enhancement method has been established to allow reduction of the number of trial repetitions. In this paper, five linear optimal and adaptive filters are proposed for enhancing the sABR. Two of the applied filters are static optimal filters, namely the Wiener Filter (WF) and Maximum-SNR Filter, and the other three filters are types of Adaptive Noise Cancellation (ANC) that employ different adaptation algorithms, namely the Least-Mean-Square (LMS), Affine Projection (AP), and Recursive-Least-Square (RLS) filters. The signal-to-noise ratio (SNR) and six error measures, evaluated in the frequency domain, are used to assess the performance of the applied methods and traditional coherent averaging by comparing the enhanced response with a reference response. These error measures focus on quantifying differences between amplitudes of frequency components and slopes between components, with a low value of the error measure indicating that the spectral pattern is preserved after enhancement. The results show that the applied linear filters would substantially reduce the required number of the recorded trials while maintaining the signal quality, and among the five compared filters, RLS has the best overall performance.

Keywords:
Speech auditory brainstem response; Enhancement; Optimal filter; Adaptive filter;

1. Introduction

The speech auditory brainstem response (sABR) is a promising tool for assessing the functionality of the auditory system [1-3] and possibly for objective fitting of hearing aids [4]. This measurement is achieved by presenting a speech stimulus and recording the evoked responses in the brain electroencephalogram (EEG). In practice, coherent averaging over a large number of trials is used to increase the signal-to-noise ratio (SNR) of the obtained response. This requires a long recording time and so limits the use of this technique in clinical applications. In the literature, various methods have been proposed for enhancing some other related ABRs. In [5,6], static and adaptive filters are used for estimating the Auditory Steady State Response (ASSR). In [7,8], the Wavelet Transform and Independent Component Analysis are used for enhancing the transient ABR and a response composed of a single frequency component. In contrast, the sABR is composed of multiple frequency components that follow the fundamental frequency (F0) of the speech stimulus and its harmonics.

There is currently no established method, apart from coherent averaging, for enhancing the sABR. If a method was developed that can outperform coherent averaging, then it would allow conducting a smaller number of trials, and so make the sABR technique more practical in clinical applications.

In this paper, five filters are applied to enhance the sABR: Wiener Filter (WF), Maximum-SNR Filter,

and three types of Adaptive Noise Cancellation (ANC) that employ different adaptation algorithms, namely Least-Mean-Square (LMS), Affine Projection (AP), and Recursive-Least-Square (RLS). This work follows on our previous research for accurate sABR detection [9]. The concept is that the sABR is first determined as present or not in the recorded signal, and then it is enhanced if present.

Signal-to-noise ratio (SNR) is used to compare the coherent-averaging method to the applied filters, and six error measures are used to compare the filtered signal to a reference evoked response obtained by averaging over a very large number of trials. These error measures evaluate the ability of the filters to maintain the spectral content and pattern of the response after enhancement. As such, the differences in the amplitudes of spectral components are measured using: 1. The absolute error (AE), 2. The mean square error (MSE), and 3. The weighted MSE (wMSE), and the differences in slopes between spectral components are measured using: 4. The slope absolute error (sAE), 5. The slope MSE (sMSE), and 6. The slope weighted MSE (swMSE). The slope error measures are utilized to evaluate how the spectral pattern of the response is affected by the filter. Maintaining relative differences in spectral components after filtering would also be important when assessing responses across different experimental conditions (e.g. [10]).

2. Methodology

The data used in this study is a set of recorded sABR from human subjects in response to a speech stimulus consisting of a synthetic vowel /a/ with a fundamental frequency (F0) at 100 Hz and first three formants at 700, 1220 and 2600 Hz, and a set of modeled responses based on the recorded sABR. The stimuli were presented alternately in opposite polarities and the responses were averaged giving what is known as Envelope Frequency Response (EFR). Each response is 0.3198 seconds long and contains 1024 time-samples, resulting in a sampling frequency of approximately 3202 Hz. Since the vowel was generated using a formant synthesizer, the modeled response contains signal components only at the fundamental frequency of the stimulus and its harmonics (i.e. 100 Hz, 200 Hz etc.).

The recorded data consists of sets of three different qualities: R20, R250 and R36k. R20 were obtained from 5 subjects and include 75 blocks, with each block being the coherent average of responses from 20 individual trials. R250 were obtained from four subjects and include 36 blocks, with each block being the coherent average of 250 responses. R36k includes a single very high SNR block based on the coherent averaging of 36000 responses from 12 subjects (including the R250 and R20 responses). The average SNRs corresponding to R20, R250, R36k, are -1, 7, 24.8 dB, respectively.

The modeled data include 400 sABRs at four different signal qualities (100 sABRs each). These signal qualities correspond to single trials (M01), coherent averages over 20 trials (M20), coherent averages over 250 trials (M250), and coherent averages over 36000 trials (M36k). Each modeled response is also 0.3198 seconds long and contains 1024 time-samples. The average SNRs corresponding to M01, M20, M250, M36k, are -13, -2, 7, 27 dB, respectively. In addition, a reference modeled response (Minf) was created that contains no noise and is based on the amplitudes and phases of the response components in R36k at F0 and 10 of its harmonics. Details of the modeling approach and the accuracy the modeled signals are discussed in [11]. Evaluation of several detection methods using the modeled data is presented in [9].

The main reason for including the modeled data is to extend testing of the enhancement methods to comprise the single-trial sABR, and to have a fair comparison across different sABR qualities since the available recorded data has different number of blocks across different qualities. Therefore, the analysis of the performance of the enhancement methods is conducted in detail with the modeled data. The recorded data is used at the end in an overall validation of the results.

2.1 The enhancement methods:

This section discusses the five applied enhancement methods. For the used static optimal filters, the Wiener Filter (WF) and Maximum-SNR (Max-SNR) algorithms are applied to estimate the 100-tap FIR-filter parameters (w) that minimize the error (e) between the filtered signal (s) and the reference (d):

$$s_i = w^T u \quad \text{Eq.1}$$

$$e_i = d_i - s_i \quad \text{Eq.2}$$

where (u) is the observed signal and the reference d is Minf.

For the used Adaptive Noise Cancelation (ANC) methods, three adaptive algorithms are applied to update the 7-tap FIR parameters to minimize the error (e) between the estimated noise (n) and the observation (u):

$$n_i = w^T \boldsymbol{v} \quad \text{Eq.3}$$

$$e_i = u_i - n_i \quad \text{Eq.4}$$

where ($\boldsymbol{v}$) is the noise reference obtained by LQ factorization. LQ factorization decomposes the observation (u) into a signal that contains most of the noise spectrum [12]. The resulting enhanced signal (in Eq.2 and Eq.4) is e_i. The following subsections provide the equations that are used to obtain and update the weights:

2.1.1 Wiener Filter (WF) [13]: This is the simplest optimal filter to derive. For two given signals, an observed signal (u) with variance (σ_u^2), a desired signal (d), an autocorrelation matrix (R_u) and crosscorrelation vector (p_{ud}), Wiener filter optimal coefficients are obtained by:

$$w = \sigma_{\boldsymbol{u}}^2 \boldsymbol{R_u^{-1}}\, \boldsymbol{p_{ud}} \quad \text{Eq.5}$$

The Wiener filter produces an estimated desired signal (d) that minimizes the MSE.

2.1.2 Maximum-SNR Filter [14]:

Maximum SNR filter tends to maximize SNR, and it is defined by:

$$w = \zeta \boldsymbol{R_v^{-1}}\, \boldsymbol{p_{ud}} \quad \text{Eq.6}$$

where (R_v) is autocorrelation of the noise and ζ is a scaling factor [14].

2.1.3 Least-Mean-Square (LMS) [15]: LMS is considered an adaptive filter with low processing complexity, and it minimizes the MSE by:

$$w_i = w_{i-1} + \mu\, u_i\, e_i \quad \text{Eq.7}$$

where μ is the step-size parameter.

2.1.4 Affine Projection (AP) [16]: AP has the ability to balance between processing complexity and the convergence time. Its weights are updated by:

$$w_i = w_{i-1} + \mu\, U_i\, (U_i^H\, U_i + R)^{-1}\, e_i \quad \text{Eq.8}$$

where U is the observed input in matrix form and R is a scaled diagonal matrix.

2.1.5 Recursive-Least-Square (RLS) [17]: RLS involves more complex processing than LMS but has faster convergence time. Its weights are updated by:

$$w_i = w_{i-1} + K_i\, u_i\, e_i \quad \text{Eq.9}$$

where K_i is the gain factor, and it can be calculated by:

$$K_i = \lambda^{-1}\left[K_{i-1} - \frac{\lambda^{-1}\, K_{i-1} u_i^* u_i K_{i-1}}{1 + \lambda^{-1} u_i K_{i-1} u_i^*}\right]$$

where λ is a forgetting factor.

2.2 Performance measures:

This section explains the measures that were used to assess the performance of the five enhancement methods evaluated in this work. The SNR and the error measures described below are all obtained based on the frequency spectrum of the sABR.

2.2.1 The signal-to-noise ratio (SNR):

Overall SNR is calculated by combining the signal component RMS values (i.e. at 100 Hz, 200 Hz, …up to 700 Hz) and dividing by the combination of noise-band RMS values between the signal components (i.e. 100-200 Hz, 200-300 Hz, … and 600-700 Hz, excluding the signal components). The upper frequency limit is chosen at 700 Hz because the EFR is usually very weak above this frequency.

2.2.2 The error measures:

This section discusses the six error measures that are used to quantify the difference between the spectral components in the filtered signals and in the reference signal (i.e. "Minf"). The error measures are based on differences in amplitudes and differences in slopes between signal components in the filtered signals and the reference signal.

The signal components at the fundamental frequency of the stimulus and its first 6 harmonics (i.e. 100 Hz, 200 Hz etc. up to 700 Hz) are considered and for the slope error measures, 6 slopes are obtained for each filtered signal. These slopes are obtained between the signal components at 100 and 200 Hz, between 200 and 300 Hz etc. up to the slope between 600 and 700Hz).

The following are the six amplitude and slope error measures:

1) The absolute error (AE): AE is obtained by measuring the error between the filtered signal "***s***" and the reference "***d***" using the following equation:

$$AE = mean(| \boldsymbol{d} - \boldsymbol{s} |) \qquad \text{Eq.10}$$

2) The mean-square-error (MSE): MSE is used as a criterion in many optimization problems, and is obtained by:

$$MSE = mean((\boldsymbol{d} - \boldsymbol{s})^2) \qquad \text{Eq.11}$$

3) The weighted mean-square-error (wMSE): Weights are applied since most of the amplitude means across all filtered sABRs are not equal to amplitudes in the reference (i.e. Minf). The method described in [18] is followed to obtain seven weights which are at the stimulus fundamental frequency and six harmonics (i.e. 100 Hz, 200 Hz, … 700 Hz). Then wMSE is obtained using Equation 11 on the weighted values.

4) The slope absolute error (sAE): sAE is obtained using the calculation discussed for AE (i.e. Equation 10) but for the 6 obtained slopes.

5) The slope mean-square-error (sMSE): sMSE is obtained using Equation 11 but for the 6 obtained slopes.

6) The slope weighted mean-square-error (swMSE): 6 weights are obtained to scale the filtered sABR slope means to the slopes of the reference signal. swMSE is then obtained using Equation 11 on the weighted values.

The six error measures are applied across the four sABR signal qualities and the five applied filters.

2.2.3 Overall measurements of performance:

In order to compare the applied filters using fewer metrics, in addition to the SNR increase for the single trial response after filtering, two other types of measurements are obtained:

1) Normalized error measures (NEMs) obtained by normalizing (assigning 0 to the minimum value of the error measure and 1 to the maximum value) and averaging each error measure across all signal qualities, or for each signal quality across the six error measures. An overall normalized error measure (ONEM) is obtained by normalizing and averaging across the six error measures and then normalizing and averaging again across all signal qualities. Normalization prior to averaging is needed because the ranges of the error measures may differ greatly.

2) Equivalent coherent-averaging number of trials calculated by averaging the unfiltered responses to obtain the same SNR of the filtered single-trial response. Figure 1 summarizes the procedure that is followed for assessing the sABR enhancement methods.

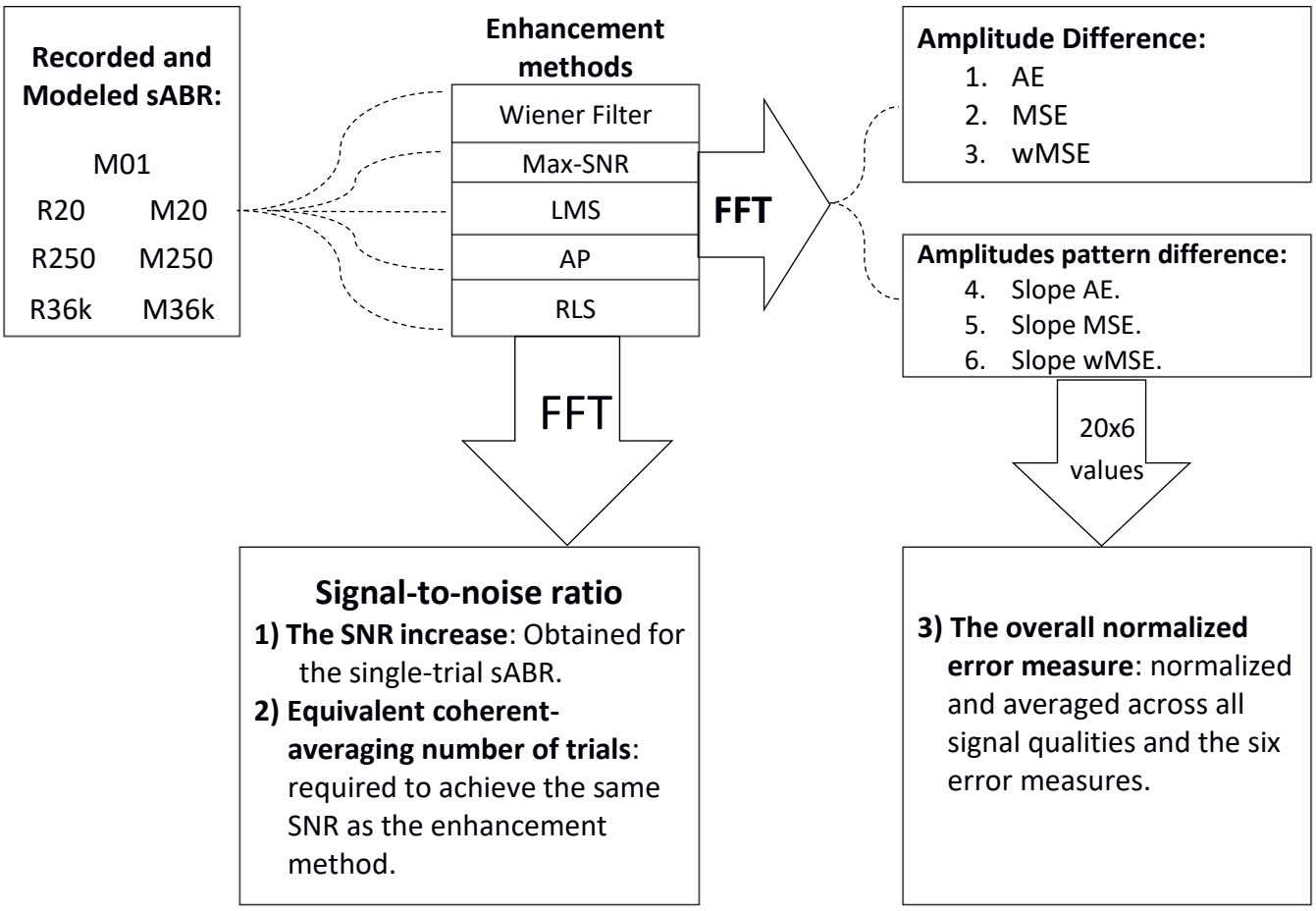


Figure 1 Summary of the procedure followed to assess the sABR enhancement methods (see text for explanation of all acronyms).

3. Results

Figure 2 shows the frequency domain content of the modeled single-trial sABR before and after applying the five filters. Figure 3 shows amplitudes at the fundamental frequency (i.e. 100Hz) of the unfiltered and the five filtered responses compared to the reference (i.e. Minf) for all modeled response signal qualities (i.e. M01, M20, M250, M36k). The amplitude at F0 is of particular interest in sABR studies (e.g. [8]).

Table 1 shows the overall performance of the five applied filters with the modeled responses. It includes the SNR increase for the single trial response, the overall normalized error measure (ONEM), and the number of trials needed to obtain the same performance using coherent averaging. Figure 4 shows boxplots of the SNR of the unfiltered modeled sABRs and the five filtered modeled ones, for all the response signal qualities.

Figure 5 shows the normalized error measures (NEMs) for the different modeled sABR qualities averaged across the six error measures (top), and for the six error measures averaged across the four signal qualities (bottom).

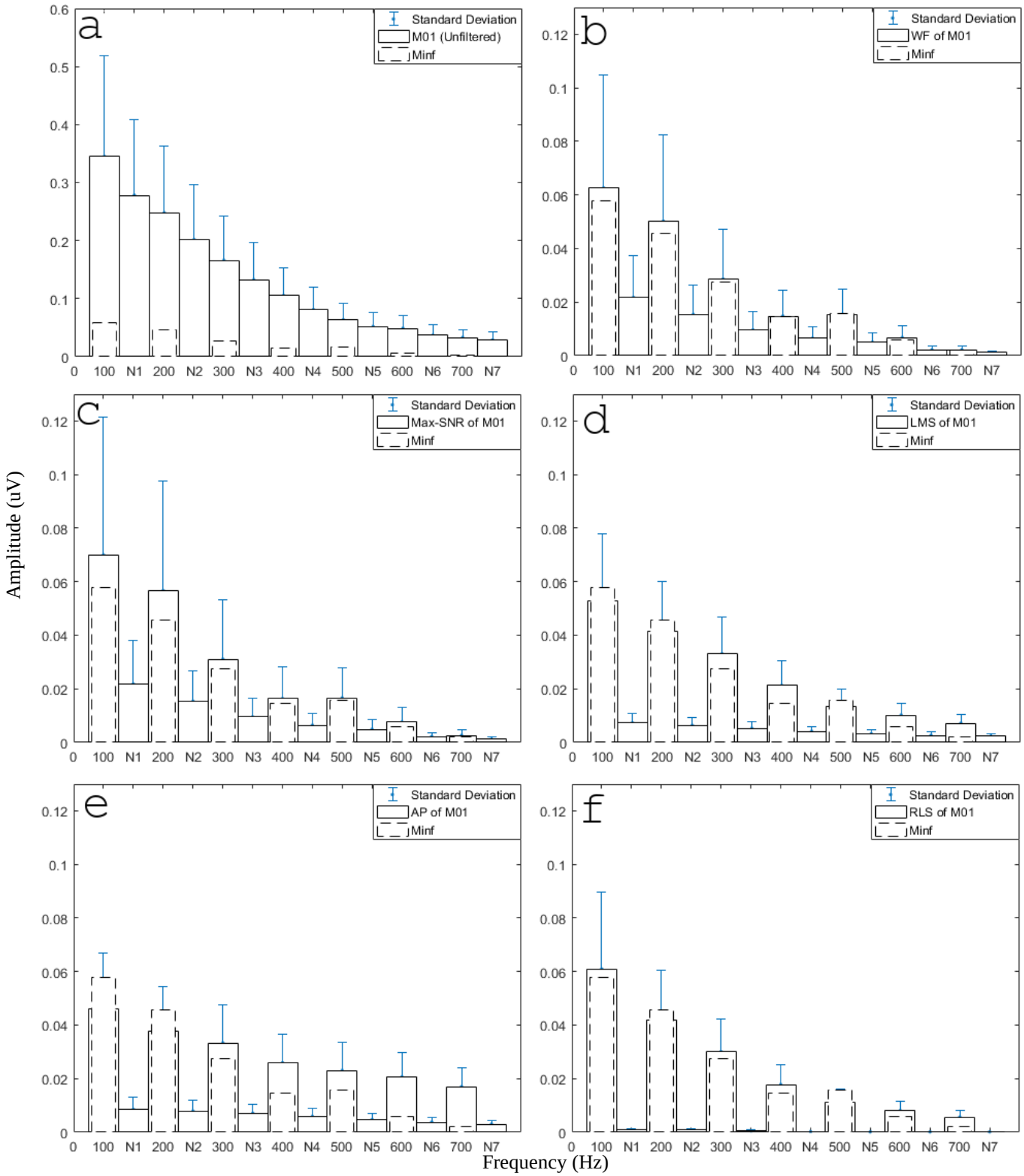


Figure 2 Frequency domain content of the modeled single-trial sABR before and after applying the five filters. Fourteen bins are shown corresponding to the response components (i.e. 100, 200, ..., and 700Hz) and noise-band RMS values between them (i.e. N1, N2, ..., and N7). The dashed bars show the reference signal (i.e. Minf). All plots have the same amplitude scale except plot "**a**".

Table 1 Overall performance of the five applied filters with the modeled responses

	WF	Max-SNR	LMS	AP	RLS
SNR increase for a single trial response					
dB	22.77	23.10	31.90	44.34	48.63
Overall Normalized Error measure					
Normalized Error	0.43	1.00	0.45	0.05	0.01
Equivalent coherent averaging					
# of trials	442	456	2462	63740	265950

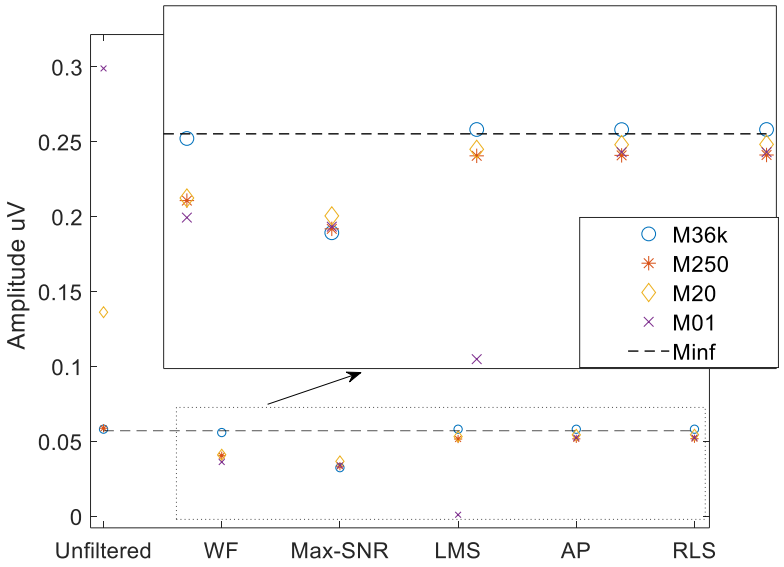


Figure 3 Amplitudes at the fundamental frequency (100Hz) of the unfiltered and the five filtered modeled sABRs compared to the reference (Minf) which is shown as a dotted line.

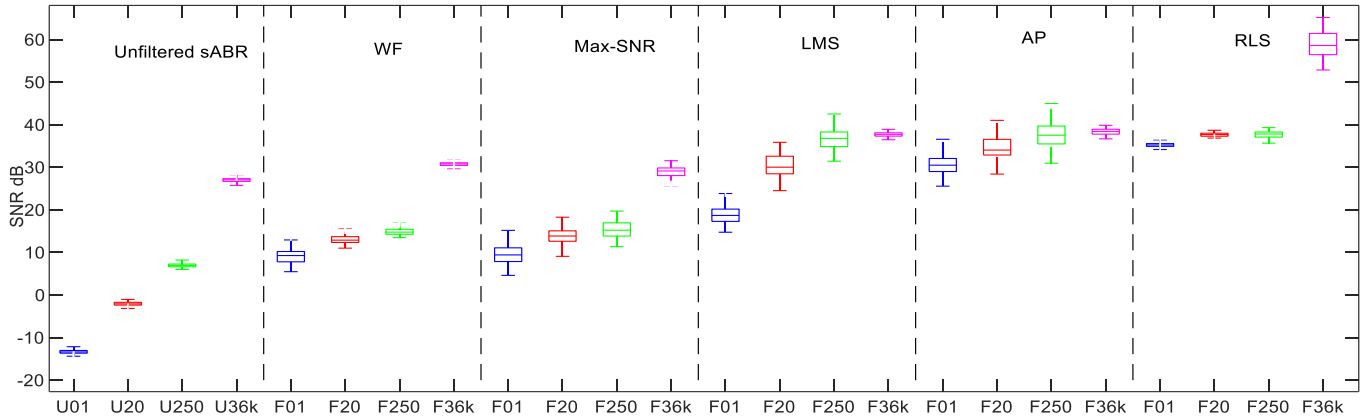


Figure 4 Boxplots of the SNR in dB of the unfiltered modeled sABR (M01, M20, M250, and M36k) and the outputs (F01, F20, F250, and F36k) after applying the five filters (WF, Max-SNR, LMS, AP, RLS).

4. Discussion

The overall measurements of performance with the modeled sABR in Table 1 reflect the frequency domain plots in Figure 2, the amplitudes at the fundamental frequency in Figure 3, the SNR boxplots in Figure 4, and the normalized error plots in Figure 5. For instance, the performance of the Max-SNR filter was overall the worst among the five filters, while RLS was the best.

The equivalent coherent averaging numbers of trials indicated in Table 1 confirm that all the denoising techniques would significantly reduce the required recording time for sABR. In particular, RLS increases the SNR of the single-trial sABR by 48.63 dB which is equivalent to coherently averaging 265,950 trials. Therefore, if the primary goal is to increase the response SNR, then RLS would allow reduction of the recording time by this factor. Moreover, in practice, the number of trials needed to achieve the same SNR may be even higher because of non-ideal recording and signal conditions.

The normalized error plots in Figure 5 confirm that the performance of the filters varies with the different sABR signal qualities and with the different error measures, a point which must be considered when selecting the denoising technique. For instance, when the response is from a single-trial, the best overall performance is obtained when it is filtered by RLS (the red dotted line with filled circles in the top figure), while with a higher quality sABR, such as M36k, the best performance is obtained when it is filtered by WF (the blue dashed line with diamonds in the top figure). This also applies to the six error measures. Based on MSE, for instance, WF performs better than Max-SNR (the blue solid line with crosses in the bottom figure), whereas based on wMSE Max-SNR performs better (the blue solid line with diamonds in the bottom figure).

On the other hand, if the goal is to estimate the response amplitude at the fundamental frequency accurately with a single trial, LMS is suggested since it combines a low mean error and relatively low

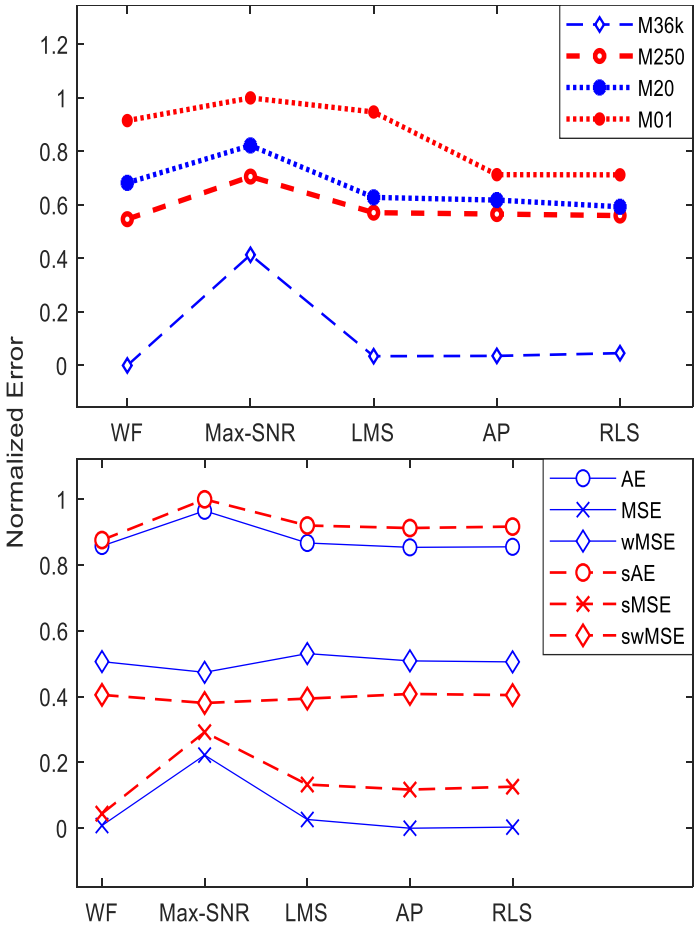


Figure 5 The normalized error measures for the different modeled sABR qualities averaged across the six error measures (top), and for the six error measures averaged across the four signal qualities (bottom).

standard deviation at the fundamental frequency as seen in Figure 2. These considerations demonstrate the importance of selecting the right filter and error measure that fit the application. Moreover, it should be noted that all reported results are based on the configurations reported in the Methodology section which try to optimize the settings for the different filters.

All previously reported results are based on the modeled data since, as mentioned earlier, they have the same number of blocks for each sABR quality which allows for a fair comparison among filters. The recorded data are used only to validate the overall trends in the obtained results. Figure 6 shows all obtained performance results with both recorded and modeled sABR. In general, all obtained recorded-based results are similar to those obtained by the

modeled data. The relatively small differences between the two results could be because of the insufficient number of the recorded data samples which could result in biased results. In the case of R36k, there is only one recorded data sample which can be an outlier if compared to 100 blocks of M36k (the R36k outlier is shown in the top plot of Figure 6 in the last position on the right which represents the SNR of the RLS-filtered response). This study evaluated different linear and optimal adaptive filters for enhancing the sABR. The proposed performance measures allow for extensive assessment of the filters across different stimulus and response conditions. This would help researchers investigating the use of the sABR in assessing hearing function. For instance, the slope error measures that are here applied across different spectral components of the response can be used across different responses under different experimental conditions (e.g. with different stimuli presented in quiet and in noise at various SNR levels as done in [8]). This would reveal how these experimental conditions affect the sABR and help elucidate the neural mechanisms involved in processing of speech by the auditory system.

5. Conclusions

This paper presents an assessment of several speech auditory brainstem response (sABR) enhancement methods. Two optimal linear filters, Wiener Filter, Maximum-SNR Filter, and three Adaptive Noise Cancellation algorithms including Least-Mean-Square, Affine Projection, and Recursive-Least-Square are applied. To the best of our knowledge, this is the only study to comprehensively evaluate sABR enhancement methods and several error measures beyond the conventional method of coherent averaging.

Modeled sABRs of several qualities are enhanced and evaluated by analyzing the filtered signal in the frequency domain, and recorded ABR are used to validate the results. The performance results confirm that the enhancement methods would significantly reduce the number of trials needed to obtain a given response SNR compared to the standard method of coherent averaging. This has important implications to the use of the sABR in clinical settings, where the long recording times have limited the usability of this technique.

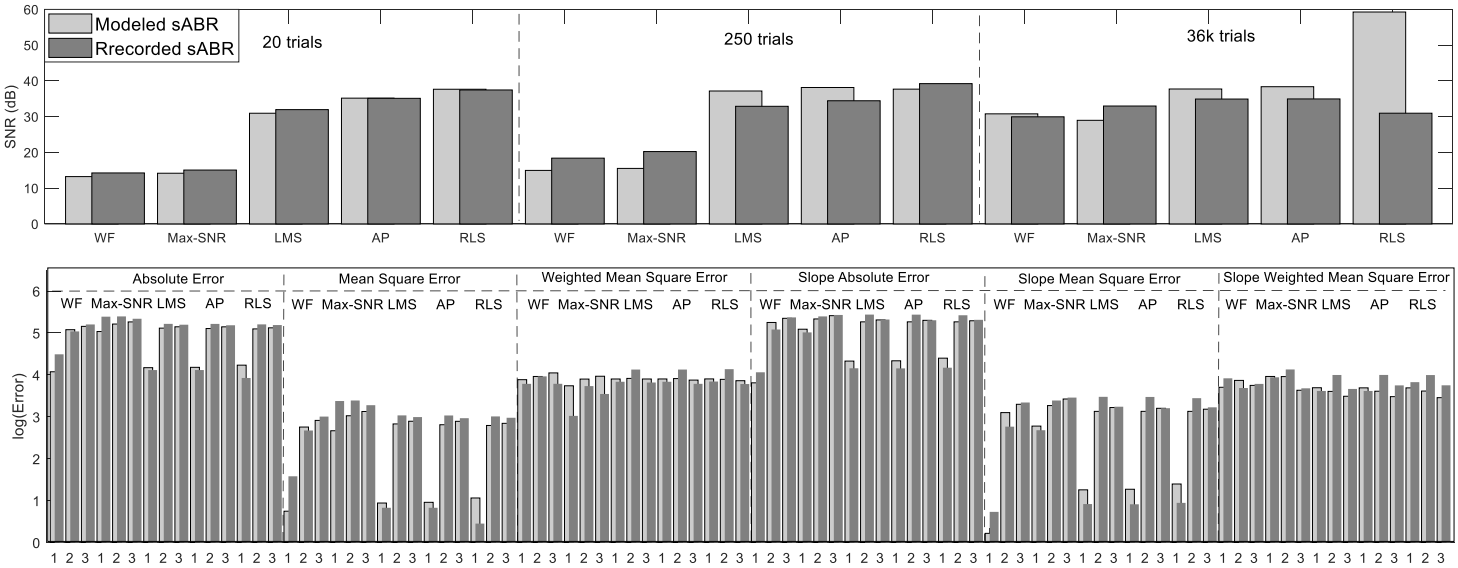


Figure 6 SNR (top) and error measure (bottom) performance of the recorded (dark grey) and modeled (grey) sABR across different qualities, filters, and error measures. In x-axis of the bottom plot, '1'represents 20-trials, '2' represents 250 trials, and '3' represents 36k-trials.

While all the applied filters are able to reduce the recording time and increase the SNR, RLS has the best performance overall when all error measures and signal qualities are considered. On the other hand, other filters sometimes outperform the RLS on an individual error measure or with a particular signal quality, suggesting that the selection of the filter should depend on the application.

This work also shows that different error measures quantify different aspects of the compared signals, a subject which needs further investigation. Moreover, evaluation of non-linear enhancement methods should be further investigated in future work since the brain neural activity can be influenced in non-linear ways by sensory stimuli [19, 20, 21].

References

[1] H. R. Dajani, D. Purcell, W. Wong, H. Kunov, T. W. Picton," Recording human evoked potentials that follow the pitch contour of a natural vowel," IEEE Transactions on Biomedical Engineering, 52(9), pp.1614-1618, 2005.

[2] S. J. Aiken, T. W. Picton, "Envelope and spectral frequency-following responses to vowel sounds," Hearing Research, 245(1-2), pp. 35-47, 2008.

[3] E. C. Thompson, K. W. Carr, T. White-Schwoch, S. Otto-Meyer, N. Kraus, "Individual differences in speech-in-noise perception parallel neural speech processing and attention in preschoolers," Hearing Research, 344, pp. 148-157, 2017.

[4] H. R. Dajani, B. Heffernan, C. Giguère. "Improving hearing aid fitting using the speech-evoked auditory brainstem response." IEEE Engineering in Medicine and Biology Society Conference (EMBC), 2013.

[5] B. V. Dun, J. Wouters, M. Moonen, "Multi-Channel Wiener Filtering Based Auditory Steady-State Response Detection," IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP, 2, pp. 929-932, 2007.

[6] D. M. McNamara, A. K. Ziarani, "A new adaptive technique of estimation of steady state auditory evoked potentials," IEEE Engineering in Medicine and Biology Society Conference EMBC), pp. 4544-4547, 2004.

[7] A. Bradley, W. Wilson, "On wavelet analysis of auditory evoked potentials," Clinical Neurophysiology, 115(5), pp. 1114-1128, 2004.

[8] D. Iyer, G. Zouridakis, "Single-trial evoked potential estimation: Comparison between independent component analysis and wavelet denoising," Clinical Neurophysiology, 118(3), pp. 495-504, 2007.

[9] A. Fallatah, H. R. Dajani, "Accurate detection of speech auditory brainstem responses using a spectral feature-based ANN method," *Biomedical Signal Processing and Control*, 44(1), pp. 307-313, 2018.

[10] F. Prévost, M. Laroche, A. Marcoux, H. R. Dajani, "Objective measurement of physiological signal-to-noise gain in the brainstem response to a synthetic vowel," *Clinical Neurophysiology*, 124(1), pp. 52-60, 2013.

[11] A. Fallatah, H. R. Dajani, "Estimating the properties of the single-trial speech auditory brainstem response using an accurate AR model," research article in preparation.

[12] A. Fallatah, "Enhancement of Speech Auditory Brainstem Responses using Adaptive Filters," M.A.Sc. Thesis, School of Electrical Engineering and Computer Science, University of Ottawa, 2012.

[13] D. J. Doyle, M. L. Hyde, "Wiener Filtering of Auditory Brainstem Responses," *Journal of Clinical Engineering*, 10(4), pp. 331-338, 1985.

[14] J. Benesty, J. Chen, "Single-Channel Noise Reduction with a Filtering Vector," in *Optimal Time-Domain Reduction Filters: A*

Theoretical Study, Heidelberg: Springer, 2011, pp 3-22.

[15] A. Sayed, “Linear Estimation,” in *Fundamentals of Adaptive Filtering*, Hoboken, NJ: Wiley, 2003, pp. 47-77.

[16] J. Satyanarayana, G. Amjad Khan, "Affine projection algorithm applied to adaptive noise cancellation," *International Journal of Electronics & Communication Technology*, 3(1), 2012.

[17] Y. Tang, A. Norcia, "An adaptive filter for steady-state evoked responses," *Clinical Neurophysiology*, 96(3), pp. 268–277, 1995.

[18] M. S. John, A. Dimitrijevic, T. W. Picton, "Weighted averaging of steady-state responses," *Clinical Neurophysiology*, 112(3), pp. 555-562, 2001.

[19] O. David, J. M. Kilner, K. J. Friston, "Mechanisms of evoked and induced responses in MEG/EEG," Neuroimage, 31(4), pp. 1580-91, 2006.

[20] M. Laroche, H. R. Dajani, F. Prévost, A. M. Marcoux, “Brainstem auditory responses to resolved and unresolved harmonics of a synthetic vowel in quiet and noise,” Ear and Hearing 34 (1), pp. 63–74, 2013.

[21] S. Gholami-Boroujeny, A. Fallatah, B. P. Heffernan, H. R. Dajani, “Neural network-Based adaptive noise cancellation for enhancement of speech auditory brainstem responses,” Signal Image Video Processing 10(2) pp. 389–395, 2016.

ADDENDUM

This section addresses some comments that were raised by the thesis committee:

In: *2.1 The enhancement methods:*
1. The static and adaptive solutions are compared for different setups, this creates confusion. Wiener could also be used for Eq. 3 and Eq. 4, with the same number of taps as LMS, RLS, AP (7). This would allow new comparisons (Wiener in 2 different setups, and Wiener vs. adaptive ones in 2nd setup)

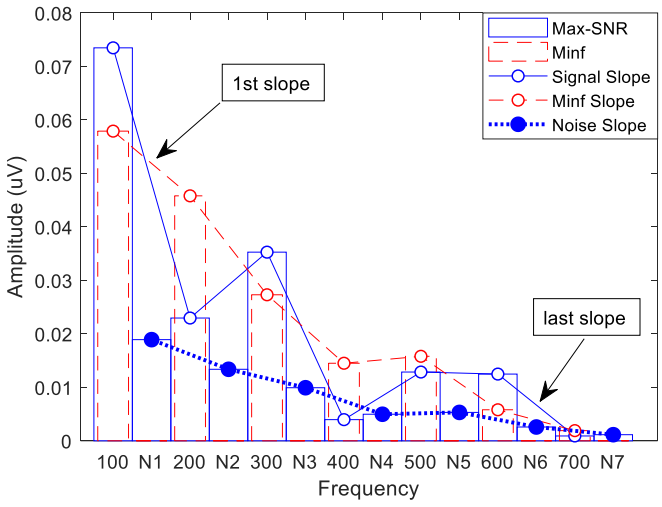
Wiener filter can be set up according to adaptive filter's configuration using Eq.3 and Eq.4. However, this will reduce WF performance. For example, the reported WF performance is based on 100 coefficients. Reducing the number of coefficients to 7 will reduce the performance dramatically. In (Fallatah, 2012), Figure 3-4 shows SNR of WF with different coefficient number. Also, configuring WF to estimate noise instead of the signal (i.e. Eq.3 and Eq.4) will reduces the performance since WF depends on the autocorrelation, and the correlation of the signal is higher than the noise correlation. A test was implemented on the high quality sABR (i.e.R36k) which shows 7dB SNR difference between these two configurations.

2. The Minf is used as a reference to tune optimal filters. Why is the applied stimulus not used as a reference instead?

The signal Minf is used as a reference since it contains the spectrum of the evoked response. The applied stimulus cannot be used as a reference since it is different from the targeted evoked response which is the product of several neural transformation in the nuclei of the auditory system. For example, the amplitude at the fundamental frequency of the applied stimulus is less prominent than that of the evoked response.

In: *2.2.2 The error measures:*
3. Slope Errors is an uncommon measure, so it is better if it is demonstrated by a graph.

The main reason behind using Slope Error measures is to evaluate the slope difference of the enhanced signal relative to the reference signal, and so evaluate how well the filtering approach recovers the overall spectral pattern of the signal (at the fundamental and its harmonics). The following figure shows the frequency domain content of a Max-SNR filtered sABR. As seen in the figure, there is difference between the targeted slope (the red dashed line of Minf) and the filtered signal slope (the blue solid line). This implies that the applied filter alters the signal slope while maintaining a stable Noise slope (the blue dotted line).



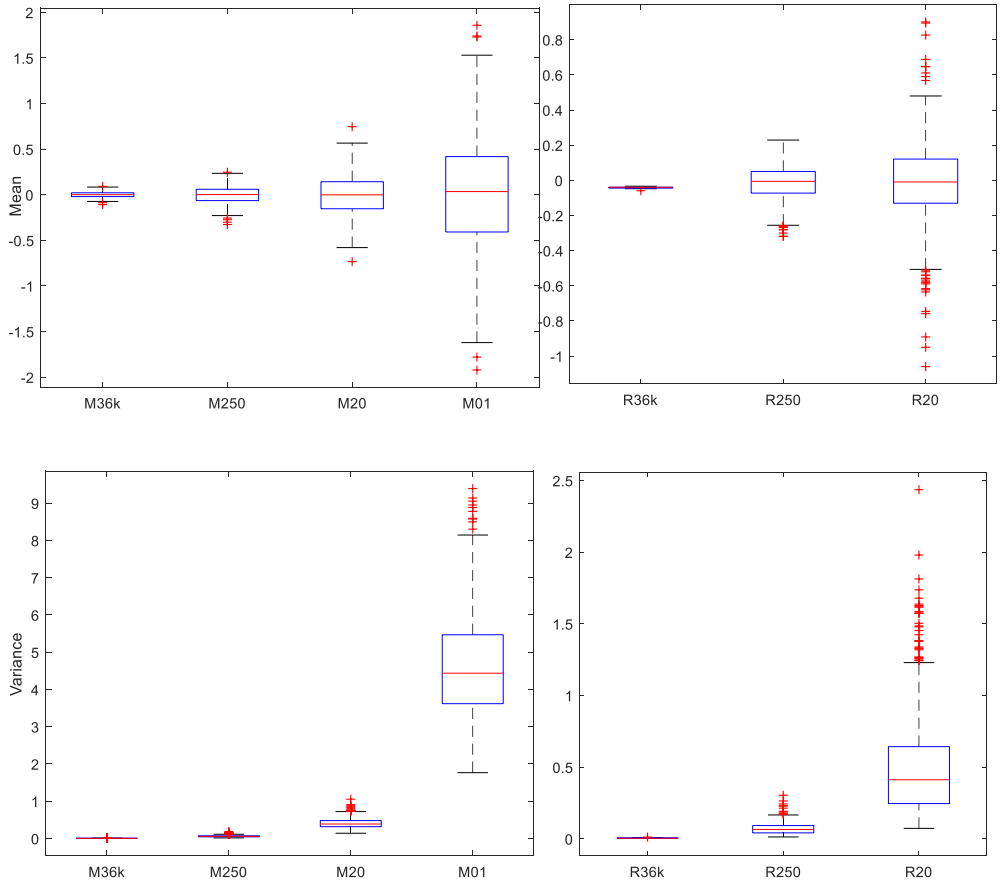
Reference:

A. Fallatah, "Enhancement of Speech Auditory Brainstem Responses using Adaptive Filters," M.A.Sc. Thesis, School of Electrical Engineering and Computer Science, University of Ottawa, 2012.

In: *3. Results:*

4. The results of LMS/AP/RLS outperform the results of WF, whereas we would expect WF to provide the best MSE results since it obtains the solution that minimizes MSE if the signal is stationary.

The WF result is not obtaining the optimized solution likely since sABR is not completely stationary. It is known that a (wide sense) stationary signal has invariant mean and variance. The figures below show boxplots of mean and variance of 128-sample windows of the recorded and modeled sABR which clearly shows variability in the mean and variance across time for the different sABR qualities.



5. To have the RLS outperform the AP by such a margin is an anomaly which brings some concerns.

The reported results are based on setting and optimizing the filters to obtain the highest SNR for the available sABR. Changing the data characteristics or analyzing another type of data could lead to different results. In another experiment, AP outperforms RLS when the length of the input signal is extended to three times the current duration of 0.3198 sec.

6. Parameters need to be specified (forgetting factor and initial value of inverse correlation matrix in RLS, step-size in LMS and AP, and regularization factor AP). They determine the performance for adaptive algorithms (there is always a tradeoff between convergence speed and MMSE performance achieved).

Below are the parameters that were used for the adaptive filters:

LMS parameters:

Forgetting factor: 1

Initial coefficients: [0.51,0.33,0.15,0.03,-0.01,-0.0069,0.0071];

Step size: 0.010

AP parameters:

Projection order: 2

Initial coefficients: [0.51,0.33,0.15,0.03,-0.01,-0.0069,0.0071];

Step size: 0.010

Regularization factor: 1

RLS parameters:

Forgetting factor: 1

Initial coefficients: [0.51,0.33,0.15,0.03,-0.01,-0.0069,0.0071];

Initial inverse covariance: 1000

Three factors are considered in setting these parameters: To have a unified setting as much as possible, to avoid negative impact on overall performance, and to ensure that a filter is not in the tracking mode when obtaining the performance measurements. The usual null weights have been shown to negatively affect the performance of LMS in favor of the other filters, so using the initialization weights given above results in a more fair comparison considering the 3 mentioned factors.

Regarding the mu value of LMS and AP, using the max-step of mu will always result in instability on some of the data, and so having a small step-size will ensure overall stable performance and high SNR.

7. Discussion

In this thesis, the main idea is to detect the signal first and then to enhance it if it is detected. The estimation and the modeling parts are needed to study the signal characteristics. The estimation part is used to estimate the background noise which is then used to create an accurate AR model in the modeling part. This model is also used to estimate the characteristics of a single-trial response, which is only accessible through modeling because in experimental recordings it is completely submerged in noise. An accurate ANN-based detector is developed and compared to several other detection methods. Afterwards, several linear static and adaptive algorithms are investigated in detail for enhancing the detected sABR. The following paragraphs discuss the main results of the four parts in more detail.

The experimental results in the estimation section and the provided mathematical proof confirm that the proposed method can estimate the time-series of the noise. The results illustrate how the proposed method outperforms WF and LMS in term of MSE and robustness against the noise type. The proposed method is suitable for estimating the noise time-series for sABR with knowledge of the stimulus fundamental frequency.

The distribution curves of the recorded and modeled sABRs are very similar as shown in the modeling section. This is also confirmed in their time-series boxplots and through the Z-test. The estimated characteristics of the single-trial is similar to those estimated using other methods with difference less than 3 dB. These results show that the proposed method can be used to model the sABR and to estimate its single-trial time-series.

The performance of the proposed detection method is better than several other methods available in the literature as indicated by the ROC and the detection performance table. This shows that this method can be used for high accuracy detection of the sABR.

In the enhancement section, the detailed study of several linear filters shows that RLS outperforms other filters based on SNR and overall error measures. The equivalent coherent averaging number of trials is shown for each method. This number indicates how a method can reduce the required recording time to obtain better sABR quality. In addition, the six reported error measures confirm that the performance of a filter varies with the different sABR signal qualities and with the different error measures, a point which must be considered when selecting the denoising technique. For instance, when the goal is to enhance a single-trial sABR, RLS should be used for best overall performance. On the other hand, if the goal is to enhance a higher quality sABR, WF has shown to perform very well. This also applies to each one of the reported error measures.

8. Conclusion and Future Work

8.1 Conclusion

In this work, four processing stages, namely estimation, modeling, detection, and enhancement are comprehensively studied in order to improve the functionality of the speech auditory brainstem response as a tool to better understand how the auditory system processes speech in normal hearing and hearing impaired individuals. Currently, the most commonly-used method for detecting and enhancing sABR, the coherent averaging method, is very time-consuming which has limited wider use of this technique in the clinic.

The results show that the designed model can estimate the single-trial sABR, and the statistical and other characteristics of the modeled sABR are similar to those obtained from the recorded sABR. The proposed ANN-based detection has higher accuracy than the other evaluated methods and fast processing time, and among the compared enhancement methods, RLS has the best overall performance in term of SNR and several error measures.

8.2 Future Work

Several research avenues are open in each of the modeling, detection, and enhancement components of the work. These include:

In the estimation component, the proposed estimation method can estimate noise if its bandwidth is separated from the signal bandwidth, but suffers when there is a conjunction of their bandwidths. Therefore, further research is required to address this issue.

In the modeling component, the proposed model can estimate sABR of any quality, which in this work was strictly related to the number of averaged trials. However, this approach can be

generalized to include response qualities that are related to different experimental conditions such as the type of stimulus and the presence or absence of background acoustic noise in the stimulus as has been studied in for example (Prévost, Laroche, Marcoux, and Dajani, 2013).

In the detection component, a different approach to selecting features should be explored. In this work, the proposed detection approach is based on spectral features; however, clinicians prefer features that help in diagnosis and that are relevant to clinical outcomes. For instance, the most commonly-used marker for hearing problems is the shape and latency of ABR waves in response to brief stimuli such as clicks. In recent research (Molina, Perez, and Valente, 2016), the shape and the latency of the time-series are used as features for detection. Integrating symbolic features-that are based on markers for hearing ability may be a fruitful avenue to explore for improving sABR detection.

As for the enhancement component, in this work linear filters are investigated for enhancement of sABR; however, non-linear filters have been shown to have good performance since the sABR originates through non-linear neural processes (Gholami-Boroujeny, Fallatah, Heffernan, and Dajani, 2016). Therefore, further investigation of non-linear filters for this application may yield useful results.

References

Acır, N., Özdamar, Ö. and Güzeliş, C., 2006. Automatic classification of auditory brainstem responses using SVM-based feature selection algorithm for threshold detection. *In Engineering Applications of Artificial Intelligence*, **19**(2), pp. 209-218.

Albrecht, V., Lánský, P., Indra, M. and Radil-Weiss, T., 1977. Wiener filtration versus averaging of evoked responses. *In Biological Cybernetics,* **27**(3), pp. 147-154.

Alessio, S.M., 2015. Chapter 9: General concepts in estimation theory, *In Digital Signal Processing and Spectral Analysis for Scientists: Concepts and Applications*. Springer. New York, United States, pp. 371-401.

Aiken, S.J. and Picton, T.W., 2008. Envelope and spectral frequency-following responses to vowel sounds. *In Hearing Research*, **245**(1-2), pp. 35-47.

Akaike, H., 1974. A new look at the statistical model identification. *In IEEE Transactions on Automatic Control*, **19**(6), pp. 716-723.

Arnold, S.A., 1985. Objective versus visual detection of the auditory brain stem response. *In Ear and Hearing*, **6**(3), pp. 144-150.

Bartoli, F. and Cerutti, S., 1983. An optimal linear filter for the reduction of noise superimposed to the EEG signal. *Journal of Biomedical Engineering*, **5**(4), pp. 274-280.

Benesty, J., Cohen, I. and Chen, J., 2017. *Chapter 2: Signal Enhancement. In Fundamentals of Signal Enhancement and Array Signal Processing*, John Wiley & Sons, New York, United States. pp. 19-61.

Blackman, R., and Tukey, 1958. *The Measurement of Power Spectra*, Dover publications, New York, United States.

Castellanos, N.P. and Makarov, V.A., 2006. Recovering EEG brain signals: artifact suppression with wavelet enhanced independent component analysis. *Journal of Neuroscience Methods,* **158**(2), pp. 300-312.

Causevic, E., Morley, R.E., Wickerhauser, M.V. and Jacquin, A.E., 2005. Fast wavelet estimation of weak biosignals. *IEEE Transactions on Biomedical Engineering*, **52**(6), pp. 1021-1032.

Chen, G.Y., Gan, M. and Chen, G.L., 2018. Generalized exponential autoregressive models for nonlinear time series: Stationarity, estimation and applications. *Information Sciences*, **438**, pp. 46-57.

Collier, O, Comminges, L, and Tsybakov, A.B., 2017. Minimax estimation of linear and quadratic functionals on sparsity classes. *The Annals of Statistics*, **45**(3), pp. 923-958.

Corona-Strauss, F.I., Delb, W., Schick, B. and Strauss, D.J., 2010. A Kernel-based novelty detection scheme for the ultra-fast detection of chirp evoked auditory brainstem responses. In *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE,* Buenos Aires, Argentina, pp. 6833-6836.

Cox, D.R. and Hinkley, D.V., 1979. Theoretical Statistics. *Chapman and Hall/CRC*. Taylor &. Francis Group. Florida, United States.

Dajani, H.R., Purcell, D., Wong, W., Kunov, H., and Picton, T.W., 2005. Recording human evoked potentials that follow the pitch contour of a natural vowel. *IEEE Transactions on Biomedical Engineering*, **52**(9), pp. 1614–1618.

Dajani, H.R., Heffernan, B.P., and Giguère, C., 2013. Improving hearing aid fitting using the speech-evoked auditory brainstem response. *IEEE Engineering in Medicine and Biology Society Conference (EMBC)*, Osaka, Japan, pp. 2812-2815.

Darwin, C.J., McKeown, J.D. and Kirby, D., 1989. Perceptual compensation for transmission channel and speaker effects on vowel quality. *Speech Communication*, **8**(3), pp. 221-234.

Davila, C.E., 1989. BAEP enhancement by weighted ensemble averaging. *In Engineering in Medicine and Biology Society. Images of the Twenty-First Century, 1989. Proceedings of the 11th Annual International Conference of the IEEE Engineering in Medicine and Biology Society,* Seattle, United States, pp. 713-714.

Davila, C.E. and Mobin, M.S., 1992. Weighted averaging of evoked potentials. *IEEE Transactions on Biomedical Engineering*, **39**(4), pp. 338-345.

Davis, H., 1964. Enhancement of evoked cortical potentials in humans related to a task requiring a decision. *Science*, **145**(3628), pp. 182-183.

Debruyne, F., 1984. Estimation of the slope of the audiogram with the auditory brain stem response. *Clinical Otolaryngology & Allied Sciences*, **9**(6), pp. 341-345.

Delb, W., Strauss, D.J. and Plinkert, P.K., 2004. A time-frequency feature extraction scheme for the automated detection of binaural interaction in auditory brainstem responses. *International Journal of Audiology*, **43**(2), pp. 69-78.

Elberling, C. and Don. M., 1984. Quality estimation of averaged auditory brainstem responses. *Scandinavian Audiology*, **13**(3), pp. 187-197.

Elberling, C. and Wahlgreen, O., 1985. Estimation of auditory brainstem response, ABR, by means of Bayesian inference. *Scandinavian Audiology*, **14**(2), pp. 89-96.

Franaszczuk, P.J., Blinowska, K.J. and Kowalczyk, M., 1985. The application of parametric multichannel spectral estimates in the study of electrical brain activity. *Biological Cybernetics,* **51**(4), pp. 239-247.

Gholami-Boroujeny, S., Fallatah, A., Heffernan, B.P. and Dajani, H.R., 2016. Neural network-based adaptive noise cancellation for enhancement of speech auditory brainstem responses. *Signal, Image and Video Processing*, **10**(2), pp. 389-395.

Goldberger, A.S., 1962. Best linear unbiased prediction in the generalized linear regression model. *Journal of the American Statistical Association*, **57**(298), pp. 369-375.

Gotze, J., Haardt, M. and Nossek, J.A., 1995. Subspace estimation using unitary Schur-type methods. *In 1995 International Conference on Acoustics, Speech, and Signal Processing*, pp. 1153-1156.

Greenberg, S. and Ainsworth, W.A., 2004. *Chapter 1: Speech Processing in the Auditory System: An Overview. In Speech Processing in the Auditory System*, Vol. 18, 1 ed., Springer, New York, United States, pp. 1-6.

Harte, J., Rønne, F.M. and Dau, T., 2010. Modeling human auditory evoked brainstem responses based on nonlinear cochlear processing*, In 20th International Congress on Acoustics 2010,* Sydney, Australia, pp. 422-431.

Hirao, T. and Adachi, S., 2009. Resonance frequency estimation of time-series data by subspace method. *In Proceedings of 2009 ICCAS-SICE*, pp. 4913-4916.

Hjorth, B., 1973. The physical significance of time domain descriptors in EEG analysis. *Electroencephalography and Clinical Neurophysiology*, **34**(3), pp. 321-325.

Hyde, M., Sininger, Y.S. and Don, M., 1998, Objective detection and analysis of auditory brainstem response: an historical perspective. *In Seminars in Hearing,* **19**(1), pp. 97-113.

Ifeachor, E.C., Hellyar, M.T., Mapps, D.J. and Allen, E.M., 1990. October. Knowledge-based enhancement of human EEG signals. *In IEE Proceedings F-Radar and Signal Processing*, **137**(5), pp. 302-310.

Jeng, F.C., Chung, H.K., Lin, C.D., Dickman, B. and Hu, J., 2011. Exponential modeling of human frequency-following responses to voice pitch, *International Journal of Audiology*, **50**(9), pp. 582-593.

Johnson, E., 1940. Estimates of parameters by means of least squares. *The Annals of Mathematical Statistics*, **11**(4), pp.453-456.

Johnson, K.L., Nicol, T.G. and Kraus, N., 2005. Brain stem response to speech: a biological marker of auditory processing. *Ear and hearing*, **26**(5), pp.424-434.

Johnson, K.L., Nicol, T.G., Zecker, S.G., Bradlow, A.R., Skoe, E. and Kraus, N., 2008. Brainstem encoding of voiced consonant–vowel stop syllables. *Clinical Neurophysiology*, **119**(11), pp. 2623-2635.

Johnson, N.L., 1950. On the comparison of estimators. *Biometrika*, **37**(3/4), pp. 281-287.

Kasami, T., 1961. Optimum linear finite-dimensional estimator of signal waveforms. *IRE Transactions on Information Theory*, **7**(4), pp. 206-215.

Kidd, G., Burkard, R.F. and Mason, C.R., 1993. Auditory detection of the human brainstem auditory evoked response. *Journal of Speech, Language, and Hearing Research*, **36**(2), pp.442-447.

Konishi, S. and Kitagawa, G., 2008. Chapter 2: Statistical models. *In Information Criteria and Statistical Modeling*, Springer, New York, United States, pp.29-74.

Kumar, K.S., Sreenivasulu, G. and Rajan, S.V., 2016, Block based SVD approach for Additive White Gaussian Noise level estimation in satellite images. *In 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom),* pp. 1464-1468.

Levy, B.C., 2008. Principles of Signal Detection and Parameter Estimation. *Springer US*, United States.

Liu, W. and Lin, W., 2012, Gaussian noise level estimation in SVD domain for images. *In 2012 IEEE International Conference on Multimedia and Expo*, pp. 830-835.

Lutman, M.E. and Sheppard, S., 1990. Quality estimation of click-evoked oto-acoustic emissions. *Scandinavian Audiology*, **19**(1), pp. 3-7.

Madhavan, G., 1989. Modified adaptive line enhancement for evoked potential extraction. *In Engineering in Medicine and Biology Society, 1989. Images of the Twenty-First Century., Proceedings of the Annual International Conference of the IEEE Engineering, Seattle, United States*, pp. 708-709.

Maki, H., Toda, T., Sakti, S., Neubig, G. and Nakamura, S., 2015. An evaluation of EEG ocular artifact removal with a multi-channel Wiener filter based on probabilistic generative model*. In Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE,* Milan, Italy, pp. 2775-2778.

McEwen, J.A. and Anderson, G.B., 1975. Modeling the stationarity and gaussianity of spontaneous electroencephalographic activity. *IEEE Transactions on Biomedical Engineering*, **7**(5), pp. 361-369.

Mina, F., Attina, V., Duroc, Y., Veuillet, E., Truy, E. and Thai-Van, H., 2017. Auditory steady state responses and cochlear implants: Modeling the artifact-response mixture in the perspective of denoising. *Public Library of Science (PLoS) one*, **12**(3), pp. 17-44.

Mohammadi, M., Pouyan, A.A., Abolghasemi, V. and Khan, N.A., 2017. Enhancement of the spikes attributes in the time-frequency representations of real EEG signals. *In Knowledge-Based Engineering and Innovation (KBEI), 2017 IEEE 4th International Conference on,* Tehran, Iran, pp. 768-772.

Molina, M.E., Perez, A. and Valente, J.P., 2016. Classification of auditory brainstem responses through symbolic pattern discovery. *Artificial Intelligence in Medicine*, **70**, pp. 12-30.

Moller, A.R., 2003. Sensory Systems: Anatomy, Physiology and Pathophysiology, *Academic Press*, San Diego, CA, United States.

Prévost, F., Laroche, M., Marcoux, A.M. and Dajani, H.R., 2013. Objective measurement of physiological signal-to-noise gain in the brainstem response to a synthetic vowel. *Clinical Neurophysiology*, **124**(1), pp. 52-60.

Poor, H.V., 1994. An Introduction to Signal Detection and Estimation. Springer-Verlag, New York, United States.

Rangayyan, R.M., 2002a, *Chapter 3: Filtering for Removal of Artifacts. In Biomedical Signal Analysis: A Case-Study Approach*, IEEE Press, Calgary, Canada, pp. 73-175.

Rangayyan, R.M., 2002b, *Chapter 4: Event Detection. In Biomedical Signal Analysis: A Case-Study Approach*, IEEE Press, Calgary, Canada, pp. 262-263.

Rappelsberger, P, 1985. Averaging of evoked potentials, *EEG-Labor,* **7**(2), pp. 53-62.

Regan, D., 1966. Some characteristics of average steady-state and transient responses evoked by modulated light, *Electroencephalography and Clinical Neurophysiology*, **20**(3), pp. 238-248.

Riddington, E.P., Wu, J., Ifeachor, E.C., Allen, E.M. and Hudson, N.R., 1996. Intelligent enhancement and interpretation of EEG signals, *IEE Colloquium on Artificial Intelligence Methods for Biomedical Data Processing,* London, United Kingdom, pp. 1-7.

Russo, N., Nicol, T., Musacchia, G. and Kraus, N., 2004. Brainstem responses to speech syllables. *Clinical Neurophysiology*, **115**(9), pp. 2021-2030.

Sanchez, R., Riquenes, A. and Perez-Abalo, M., 1995. Automatic detection of auditory brainstem responses using feature vectors. *International Journal of Bio-medical Computing*, **39**(3), pp. 287-297.

Sato, K., Honda, N., Mimura, K., Ozaki, T., Teramoto, S. and Kitajima, K., 1962. A simplified method for determining the average response time-pattern of the evoked potential in electroencephalography. *Clinical Neurophysiology,* 14(5), pp. 764-766.

Sato, K., 1963. On the linear model of the brain activity in electroencephalographic potential, *Psychiatry and Clinical Neurosciences*, *17*(2), pp. 156-166.

Selvan, S. and Srinivasan, R., 1999. Removal of ocular artifacts from EEG using an efficient neural network based adaptive filtering technique. *IEEE Signal Processing Letters*, **6**(12), pp. 330-332.

Scharf, L.L. and Demeure, C., 1991. Statistical Signal Processing: Detection, Estimation, and Time Series Analysis. Addison-Wesley. Reading, MA, United States.

Shirzhiyan, Z., Jafarpisheh, A.S., Ahadi, M. and Jafari, A.H., 2013. The representation of fuzzy model for auditory brainstem response to one syllable speech stimuli /da/. *In Biomedical Engineering (ICBME), 2013 20th Iranian Conference*, pp. 31-36.

Skoe, E. and Kraus, N., 2010. Auditory brainstem response to complex sounds: a tutorial. *Ear and Hearing*, **31**(3), pp. 302-324.

Song, I., Dagotto, E.Y., Bae, J. and y Kim, S., 2002. *Advanced Theory of Signal Detection*. Springer, Berlin, Germany.

Stapells, D.R. and Oates, P., 1997. Estimation of the pure-tone audiogram by the auditory brainstem response: a review. *Audiology and Neurotology*, **2**(5), pp. 257-280.

Steinberg, H.W., Gasser, T. and Franke, J., 1985. Fitting autoregressive models to EEG time series: An empirical comparison of estimates of the order. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, **33**(1), pp. 143-150.

Strauss, D.J., Delb, W. and Plinkert, P.K., 2004. Analysis and detection of binaural interaction in auditory evoked brainstem responses by time-scale representations. *Computers in Biology and Medicine*, **34**(6), pp. 461-477.

Sugimoto, H., Ishii, N., Iwata, A. and Suzumura, N., 1977. Stationarity and normality test for biomedical data. *Computer Programs in Biomedicine*, **7**(4), pp. 293-304.

Tahernezhadi, M., Karthik, R. and Kong, X., 1997. A transform domain approach to enhancement of evoked potential EEG signals. *In Engineering in Medicine and Biology Society, 1997. Proceedings of the 19th Annual International Conference of the IEEE,* Chicago, United States, pp. 1413-1416.

Tian, J., Grönfors, T. and Juhola, M., 1996. Research using Neural Networks for Auditory Brainstem Response Detection. *Studies in Health Technology and Informatics*, pp. 791-798.

Turajlić, E., 2017. A fast noise level estimation algorithm based on adaptive image segmentation and Laplacian convolution. In 2017 40th International Convention on Information and Communication Technology, *Electronics and Microelectronics (MIPRO),* pp. 486-491

Turajlić, E., 2018. Adaptive SVD domain-based White Gaussian Noise level estimation in images. *IEEE Access*, **6**, pp. 72735-72747.

Valdés, P., Bosch, J., Grave, R., Hernandez, J., Riera, J., Pascual, R. and Biscay, R., 1992. Frequency domain models of the EEG. *Brain Topography*, **4**(4), pp. 309-319.

Walter, D.O., 1968. A posteriori "Wiener filtering" of average evoked responses. *Electroencephalography and Clinical Neurophysiology,* **27**, p. 61.

Walter, D.O., 1970. Enhancing evoked responses by optimum a posteriori filtering. *Electroencephalography and Clinical Neurophysiology,* **28**(1), p. 93.

Wolfowitz, J., 1952. Consistent estimators of the parameters of a linear structural relation. *Scandinavian Actuarial Journal*, **1952**(3-4), pp. 132-151.

Young, G., 1941. Maximum likelihood estimation and factor analysis. *Psychometrika*, **6**(1), pp. 49-53.

Xue, Q.Z., Hecox, K. and Tompkins, W.J., 1986. Adaptive filtering of auditory evoked potentials. *In proceedings of 8th annual conference of the IEEE Engineering in Medicine and Biology Society, New York, United States*, pp. 402-405.

Appendix

Feb. 20, 2019

To whom it may it concern,

As the members of Mr. Anwar Fallatah's thesis proposal defense committee, we agree that he be allowed to use the data that has been collected under the project entitled "Measurement and Applications of Speech-evoked Auditory Brainstem Response (sABRs)" for his doctoral thesis project.

Sincerely,

Adrian Chan, Professor, Carleton University

Martin Bouchard, Professor, University of Ottawa

Miodrag Bolic, Professor, University of Ottawa

Hilmi Dajani, Associate Professor, University of Ottawa

Université d'Ottawa University of Ottawa

Bureau d'éthique et d'intégrité de la recherche Office of Research Ethics and Integrity

Ethics Approval Notice

Health Sciences and Science REB

Principal Investigator / Supervisor / Co-investigator(s) / Student(s)

First Name	Last Name	Affiliation	Role
Hilmi	Dajani	Engineering / Computer Science	Principal Investigator
Rida	Al-Osman	Engineering / Electrical Engineering	Co-investigator
Christian	Giguère	Health Sciences / Others	Co-investigator
Brian	Heffernan	Health Sciences / Occupational Therapy	Co-investigator
Anwar	Fallatah	Engineering / Electrical Engineering	Research Assistant

File Number: H05-14-22B **Type of Project:** Professor

Title: Measurement and Applications of Speech-evoked Auditory Brainstem Response (sABRs)

Renewal Date (mm/dd/yyyy)	Expiry Date (mm/dd/yyyy)	Approval Type
12/02/2018	12/01/2019	Renewal

Special Conditions / Comments:

N/A

550, rue Cumberland, pièce 154 550 Cumberland Street, room 154
Ottawa (Ontario) K1N 6N5 Canada Ottawa, Ontario K1N 6N5 Canada
(613) 562-5387 • Téléc./Fax (613) 562-5338
www.recherche.uottawa.ca/deontologie/ www.research.uottawa.ca/ethics/

Université d'Ottawa University of Ottawa

Bureau d'éthique et d'intégrité de la recherche Office of Research Ethics and Integrity

This is to confirm that the University of Ottawa Research Ethics Board identified above, which operates in accordance with the Tri-Council Policy Statement (2010) and other applicable laws and regulations in Ontario, has examined and approved the ethics application for the above named research project. Ethics approval is valid for the period indicated above and subject to the conditions listed in the section entitled "Special Conditions / Comments".

During the course of the project, the protocol may not be modified without prior written approval from the REB except when necessary to remove participants from immediate endangerment or when the modification(s) pertain to only administrative or logistical components of the project (e.g., change of telephone number). Investigators must also promptly alert the REB of any changes which increase the risk to participant(s), any changes which considerably affect the conduct of the project, all unanticipated and harmful events that occur, and new information that may negatively affect the conduct of the project and safety of the participant(s). Modifications to the project, including consent and recruitment documentation, should be submitted to the Ethics Office for approval using the "Modification to research project" form available at: https://research.uottawa.ca/ethics/forms.

Please submit an annual report to the Ethics Office four weeks before the above-referenced expiry date to request a renewal of this ethics approval. To close the file, a final report must be submitted. These documents can be found at: https://research.uottawa.ca/ethics/forms.

If you have any questions, please do not hesitate to contact the Ethics Office at extension 5387 or by e-mail at: ethics@uOttawa.ca.

Signature:

Marc Alain Bonenfant
Research Ethics Coordinator
For Catherine Paquet, Director of the Office of Research Ethics and Integrity

550, rue Cumberland, pièce 154 550 Cumberland Street, room 154
Ottawa (Ontario) K1N 6N5 Canada Ottawa, Ontario K1N 6N5 Canada
(613) 562-5387 • Téléc./Fax (613) 562-5338
www.recherche.uottawa.ca/deontologie/ www.research.uottawa.ca/ethics/