Available online at www.sciencedirect.com

ScienceDirect

Speech Communication 68 (2015) 69–84

SPEECH COMMUNICATION

www.elsevier.com/locate/specom

# Classification of speech-evoked brainstem responses to English vowels ☆

Amir Sadeghian [a], Hilmi R. Dajani [b,*], Adrian D.C. Chan [a]

[a] *Department of Systems and Computer Engineering, Carleton University, Ottawa, Canada*
[b] *School of Electrical Engineering and Computer Science, University of Ottawa, Ottawa, Canada*

Received 10 November 2013; received in revised form 24 November 2014; accepted 12 January 2015
Available online 29 January 2015

## Abstract

This study investigated whether speech-evoked auditory brainstem responses (speech ABRs) can be automatically separated into distinct classes. With five English synthetic vowels, the speech ABRs were classified using linear discriminant analysis based on features contained in the transient onset response, the sustained envelope following response (EFR), and the sustained frequency following response (FFR). EFR contains components mainly at frequencies well below the first formant, while the FFR has more energy around the first formant. Accuracies of 83.33% were obtained for combined EFR and FFR features and 38.33% were obtained for transient response features. The EFR features performed relatively well with a classification accuracy of 70.83% despite the belief that vowel discrimination is primarily dependent on the formants. The FFR features obtained a lower accuracy of 59.58% possibly because the second formant is not well represented in all the responses. Moreover, the classification accuracy based on the transient features exceeded chance level which indicates that the initial response transients contain vowel specific information. The results of this study will be useful in a proposed application of speech ABR to objective hearing aid fitting, if the separation of the brain's responses to different vowels is found to be correlated with perceptual discrimination.

*Keywords:* Speech-evoked auditory brainstem response; Envelope following response; Frequency following response; Classification of evoked responses; Auditory processing of speech; Fitting hearing aids

## 1. Introduction

In recent years, there has been increasing interest in measuring brain signals in response to speech stimuli. The ultimate goal of this research is to understand brain processing of speech in order to develop better clinical tools for both the diagnosis and treatment of sensory and cognitive impairments. Currently, hearing assessment is limited by diagnostic tests which usually employ artificial signals like tones or clicks that do not allow a clear assessment of auditory function for speech communication. While there are tests of speech perception that rely on subjective responses, these are of no value for assessing the hearing of infants and uncooperative individuals. Speech evoked responses (speech ABRs) could thus fill the need to objectively assess auditory performance in these cases (Anderson and Kraus, 2013). Recent studies have also demonstrated that the speech ABR may help identify children with language and learning problems that derive from central auditory processing impairments (Russo et al., 2004; Johnson et al., 2008).

On the treatment side, speech ABRs may prove to be very useful for the objective fitting of hearing aids (Aiken and Picton, 2008; Anderson and Kraus, 2013). Currently,

---

☆ A preliminary version of this research was published in the Proceedings of the International Conference of the IEEE Engineering in Medicine and Biology Society, Boston, USA, 2011 (Sadeghian et al., 2011).

* Corresponding author at: School of Electrical Engineering and Computer Science, 161 Louis Pasteur, Ottawa, Ontario K1N 6N5, Canada. Tel.: +1 613 562 5800x6217; fax: +1 613 562 5175.
*E-mail address:* hdajani@eecs.uottawa.ca (H.R. Dajani).

http://dx.doi.org/10.1016/j.specom.2015.01.003
0167-6393/© 2015 Elsevier B.V. All rights reserved.

hearing aid fitting is often based on diagnostic tests that use simple stimuli, which do not allow for selective acoustic treatments (Johnson et al., 2005). Since the speech ABR is believed to mostly originate in the upper brainstem (inferior colliculus, lateral lemniscus), it provides a window into subcortical processing of speech (Banai et al., 2007; Chandrasekaran and Kraus, 2010). A number of studies have presented evidence that subcortical processing of speech at this level provides the substrate for speech perception in quiet and noise (e.g. Krishnan et al., 2005; Hornickel et al., 2009; Anderson et al., 2013a,b).

The speech ABR could therefore objectively measure the effect of adjusting any of the multiple settings of modern hearing aids on the auditory system's response. Dajani et al. (2013) have proposed several ways in which the measurement of speech ABR could be used to improve this process. For example, it may be possible to use changes in the amplitudes of the response harmonics to tune the frequency dependent gain and compression levels of the hearing aid. The internal SNR of speech ABR, which was estimated in Prévost et al. (2013), may also be useful as an indicator of the quality of the internal neural representation of the speech sound after processing by the hearing aid. Moreover, since the amplitude and latency of the initial transient complex of the speech ABR depend on the initial consonant in a speech sound and are affected by noise (Russo et al., 2004; Johnson et al., 2008), they will likely be dependent on the compression time constants of the hearing aid. Anderson and Kraus (2013) have also suggested that hearing aid settings could be adjusted to maximize correlation between the spectra of the speech ABR and the stimulus. However, given that the speech ABR reflects signal transformations from the auditory periphery to the upper brainstem, it is currently unclear whether the similarity between the stimulus and the response would be the best measure of hearing aid performance. Much more systematic study of the speech ABR is needed to determine the best use of these responses in hearing aid fitting (Clinard and Tremblay, 2013).

Although the auditory brainstem response (ABR) to simple artificial stimuli is widely used in the clinic, it provides little understanding about the auditory processing of complex stimuli such as speech sounds. The speech ABR, on the other hand, reflects neural processing of the different components of speech. In an early study, Greenberg showed that components that follow speech formants are present in the evoked response (Greenberg, 1980). More recent work has led to a more detailed characterization of the auditory brainstem response to vowel stimuli. This response consists of two parts: (1) the transient response and (2) the sustained response. The transient response is short ($\lesssim 20$ ms) and is similar to the transient response to click stimuli (Skoe and Kraus, 2010). It usually contains prominent peaks which originate from the ascending auditory pathway between the cochlear nerve and midbrain, and the VA complex of the transient response in particular signifies auditory processing in the upper brainstem (Banai et al., 2007; Chandrasekaran and Kraus, 2010). As such the transient response may be thought to be a response to the "attack" characteristics of the stimulus onset (Skoe and Kraus, 2010). The transient response can thus differ depending on the initial consonant, and may contribute to the identification of specific speech sounds (Johnson et al., 2008; Skoe and Kraus, 2010). However, there has been no previous work on whether it is able to convey phonetic information when the stimulus is a pure vowel.

The sustained response ($\gtrsim 20$ ms) of speech ABR follows the periodic components of the speech stimulus. Depending on how the response signals are analyzed, it can correspond to the Envelope Following Response (EFR) or Frequency Following Response (FFR) (Aiken and Picton, 2006). The EFR is calculated by taking the average between the responses to the stimulus in one polarity and an equal number of responses to the stimulus in inverted polarity (Avg. speech ABRs + Avg. Inv polarity speech ABRs/2), while the FFR is calculated by taking the average between the response to the stimulus in one polarity and an equal number of the negative of the responses to the stimulus in inverted polarity (Avg. speech ABRs – Avg. Inv polarity speech ABRs/2) (Aiken and Picton, 2008). In the notation of some recent studies, the EFR corresponds to the response to the temporal envelope in the stimulus (or ENV), and the FFR corresponds to the response to the temporal fine structure (or TFS) (e.g. Anderson et al., 2013a,b; Zhong et al., 2014).

The EFR mainly reflects auditory neural activity that is phase-locked to the envelope of the speech stimuli, and so it has a fundamental frequency equal to the stimulus F0 (Aiken and Picton, 2008; Dajani et al., 2005). The response at F0 is probably introduced primarily through the rectification of the envelope during inner hair cell transduction (Cebulla et al., 2006; Aiken, 2008; Aiken and Picton, 2008). Energy would also appear at the early harmonics of F0 because the envelope is non-sinusoidal. Other nonlinearities within the cochlea and in higher neural pathways produce multiple intermodulation distortion products in response to pairs of stimulus harmonics which also appear in the brainstem response and contribute to the EFR. Since the synthetic vowel used in this study only contains energy at F0 and integer harmonics, these distortion products would also occur only at integer harmonics of F0 (Aiken and Picton, 2008).

On the other hand, the FFR is formed as a result of auditory neural phase-locking that directly follows the harmonics of the speech stimulus, and in particular near the first formant F1 since these harmonics are typically the most intense in the stimulus and are usually well within the phase-locking frequency limit of neurons in the auditory brainstem (Krishnan, 2002; Skoe and Kraus, 2010). Intermodulation distortion products, however, may also contribute to the FFR (Aiken and Picton, 2008).

In addition to the contribution of activity in the ascending auditory system to the speech ABR, top-down influences