Objective Classification of Auditory Brainstem Responses to Consonant-Vowel Syllables Using Local Discriminant Bases

Zahra Shirzhiyan [a,b], Elham Shamsi [a,b], Amir Salar Jafarpisheh [c], Amir Homayoun Jafari [*a,b.]

[a] Biomedical Engineering and Medical Physics Department, Faculty of Medicine, Tehran University of Medical Sciences (TUMS), Tehran, Iran

[b] Research Center for Biomedical Technologies and Robotics (RCBTR), Tehran, Iran

[c] University of Social Welfare and Rehabilitation Sciences, Tehran, Iran

Corresponding author: Dr. Amir Homayoun Jafari at Biomedical Engineering and Medical Physics Department, Faculty of Medicine, Tehran University of Medical Sciences, Poursina Street, Tehran, Iran, E-mail: h_jafari@tums.ac.ir, Tel: +982166466383

Abstract

Objective: The main goal of this study is classifying the auditory brainstem responses to consonant-vowels /da/, /ba/, and /ga/ into three classes automatically.

Method: Auditory brainstem responses (ABRs) to consonant-vowel syllables /da/, /ba/, and /ga/ were recorded from twenty-seven normal subjects. Time domain features (i.e., energy, entropy, cosine distance and correlation coefficient) and frequency domain features (i.e., magnitude and phase of the frequency responses) were extracted. Also, local discriminant bases (LDB) method was employed to extract time-frequency features from wavelet packet coefficients. Then, random subset feature selection (RSFS) algorithm reduced the feature spaces. Afterward, discriminant analysis (DA), naive Bayes (NB), multiclass support vector machine (MSVM) and K-nearest neighbors (KNN) methods classified the selected features.

Results: The time-frequency domain features showed better results than other features (91.36%). The maximum accuracy was achieved for MSVM classifier when we used a combination of all the features (97.5%).

Conclusion:

This study shows the efficiency of frequency and time-frequency domains features. The results indicate that time-frequency features obtained by local discriminant bases were more successful in objective classifications of the responses. Besides, the selected features from the phase of the frequency responses were reliable in classifying the responses to consonant-vowels /da/, /ba/ and /ga/.

Significance: The importance of this study lies in the fact that it helps the objective classification of auditory brainstem responses underlying the different encoding of three consonant-vowels (/ba/, /da/ and /ga/) in frequency and time-frequency domains.

Keywords: Speech ABR; Speech encoding; Local discriminant bases

1. Introduction

For years, analysis of auditory brainstem responses (ABRs) to synthesized signals, such as clicks or tones, has been one of the major clinical tools for physicians and researchers to diagnose hearing impairments and auditory processing disorders [1]. However, there is no possibility for objective evaluation of the auditory performance for speech communication. Moreover, measurement of subcortical signals in response to speech stimuli makes it possible to understand how the brain processes the speech and helps the development of better diagnostic tools for objective evaluation of auditory performance for speech communication. Central auditory processing impairments sometimes lead to language and learning problems. It has been stated that speech ABR analysis could also help diagnose these problems [2, 3].

Some studies tried to demonstrate that speech-ABR may be very useful for hearing aid fitting [4], [5]. Additionally, several recent studies have pointed out that subcortical speech processing contributes to speech perception (both in silence and noise) [6]. Some settings of modern hearing aids could be adjusted by speech ABRs [7]. For instance, tracking the changes in the amplitudes of the response harmonics could be useful for the adjustment of the frequency-dependent gain and hearing aid compression levels. Also, the compression time constants of the hearing aid may influence both the amplitude and latency of the initial transient component of the speech ABRs to different initial consonants. These components of the response are prone to noise [2, 3]. It is not evident whether the resemblance between the stimulus and the response could be the best criterion for the assessment of hearing aid functionality. However, it has been shown that the correlation between the spectra of the speech-ABR and the stimulus would be maximized by tuning the hearing aid settings [4].

Subcortical neural processing of the speech is represented by the speech-ABR. The existence of speech-formants-following components in the speech-ABRs has been proved earlier [8]. Recently, evaluating the characteristics of the ABR to complex stimuli (e.g., speech) have been more popular. This response comprises two parts: the transient response and the sustained response. The former resembles the transient component of the click-ABR and lasts for less than 20 milliseconds after the stimulus onset and offset [9]. The ascending auditory pathway between the cochlear nerve and midbrain usually produces dominant peaks in the transient response. More specifically, its V-A complex reflects auditory processing in the upper part of the brainstem [10, 11]. The transient response can be considered as a response to the attack characteristics of the stimulus onset [9]. Consequently, the transient responses to different initial consonants are not the same, and they can be utilized to recognize the speech sounds [3, 9]. The sustained response of the speech-ABR appears after the periodic part of the speech presentation and follows it [9]. Previous studies on the speech-ABR or speech-evoked potentials have mainly focused on understanding the neurological auditory activities during speech processing, the origin of the speech-evoked potentials and new techniques for the diagnosis of hearing impairments and disorders [12, 13]. A large body of studies applied some synthesized consonant-vowels, such as /da/ [2, 14-24], a meaningful word namely /baba/ [25], and a single vowel as the stimulus to elicit and measure the speech-ABR [26, 27].

A potentially powerful tool for better understanding of human's auditory system is provided by speech recognition via speech-evoked potentials. Speech-evoked potentials of five English vowels were automatically recognized through a common classification method using transient and sustained response features. Thus, there is a strong relationship between different vowels and their corresponding speech-evoked potentials [28, 29]. The aforementioned research concluded

that speech-evoked potentials have some useful information for speech stimuli discrimination. In some studies, cortical responses were used for speech classification. In two cases, cortical neural activity patterns of the rat in response to English consonant-/a/-consonant stimuli were classified [30, 31]. Additionally, speech reconstruction is shown to be possible in human patients under neurosurgery through measurement and analysis of cortical surface potentials [32]. Furthermore, it has been found that the decoding of the cortical response to speech stimuli would help the researchers develop brain-computer interfaces (BCIs) for speech communication purposes [33, 34]. Besides, researchers have examined ABRs to three synthesized consonant-vowel stimuli (/ba/, /da/, and /ga/) in children without learning disorders. It has been revealed that the mentioned responses are distinctive according to the temporal and spectral features of the first 60 ms of the response [3]. It was also pointed out that among /da/, /ba/ and /ga/, the stimulus with higher onset frequency of F2 and F3 formants led to an earlier speech-ABR [3]. Also, they showed that the difference in the minor peak latencies of the speech-ABR to these consonant-vowels was more robust than that of the major peaks. Skoe and her colleagues tried to present an objective neural index of differentiation of speech-ABRs to /da/, /ba/ and /ga/ using the phase of power spectrum density in running windows. In a recent study, some nonlinear features for objective classification of the auditory brainstem response to diotic perceptually critical consonant-vowel syllables have been extracted [35]. However, there still exists a great need for further research on the speech-evoked potentials and how the speech stimuli are processed with distinct features in the subcortical regions.

The spectral distributions of the brainstem responses to /ba/, /da/, and /ga/ syllables are different [3]. Also, there are some discriminative features in the phase information of the power spectral density function [36]. Therefore, extracting features from the phase of the frequency responses

may be useful for the objective classification of the responses. Furthermore, because of the non-stationary nature of speech ABRs, time-frequency domain analysis might be helpful to discriminate these responses. According to the knowledge of authors, no study has performed the time-frequency classification of auditory brainstem responses to consonant-vowel combinations.

In this study, we used the time, frequency, and time-frequency domains features and combination of selected features for the objective classification of the auditory brainstem responses to /ba/, /da/, and /ga/ syllables and compared the results.

2. Materials and Method

2.1. Subjects

Twenty-seven volunteer students from Tehran University of Medical Sciences (13 female), 22 to 29 years old (mean = 24.34, SD = 1.95) participated in this study. None of the subjects had a history of auditory, learning, or neurological disorders. Based on self-report, all of them were Persian speakers and pure tone hearing thresholds for both of their ears were equal to or better than 20 dB HL for octave frequencies 250–8000 Hz. Subjects signed written informed consent form. The deputy of the research review board and ethics community of Tehran University of Medical Sciences approved all the procedures.

2.2. Stimuli

Three well-known synthesized voiced stop consonant-vowel syllables (/ba/, /da/ and /ga/) were provided by auditory neuroscience laboratory at Northwestern University and used as the stimuli. Figure 1 shows them in the time domain. The fundamental frequency (F0) of all the stimuli with 170 ms duration is 100 Hz. During the 50 ms formant transition period, F1 rises from 400 Hz to 720 Hz linearly. F2 reaches from 3000, 1700 and 900 to 1240 Hz, and F3 reaches from 3100, 1700, 900 to 2500 Hz for /ba/, /da/ and /ga/, respectively. Also, F4, F5, and F6 remain constant at

3300 Hz, 3750 Hz and 4900 Hz, respectively. For all the stimuli, the voicing onset occurs at 10 ms and its spectral contents are located around F4 and F5.

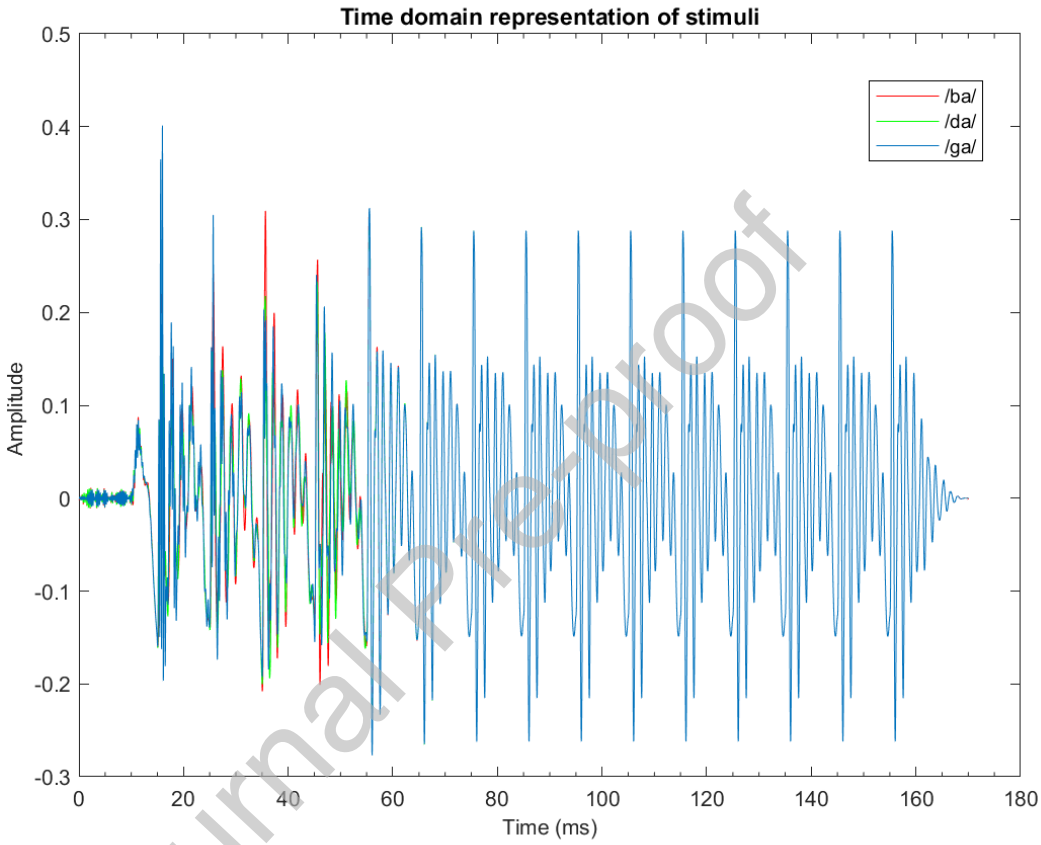

Figure 1: Time domain representation of the stimuli: The onset voicing of the stimulus is 10 ms. The first 50 ms is the formant transition period. These stimuli are different in their consonant and are identical in their vowel part.

2.3. Data acquisition

To ensure subjects' cooperation and promote stillness, all of them watched muted videotaped programs such as movies or cartoons of their choice with subtitles. They were instructed to attend to the video rather than to the stimuli. The stimuli were presented diotically with a high

precision synchronized stimuli-presenting system (National Instruments (NI DAQ- USB 6210)). The presentation rate was 4.65/s. Both stimuli polarities (i.e., condensation and rarefaction) were presented. The data were recorded in 4 blocks, each of which contained 1000 epochs (i.e., 500 rarefaction and 500 condensation polarities). The stimuli presentation paradigm is shown schematically in Figure 2. This paradigm was repeated for each consonant-vowel stimulus individually. The stimuli were presented to both ears through Etymotic ER-3 insert earphones (Etymotic Research, Elk Grove Village, IL) with the intensity of 83 dB SPL.

An active electrode was placed at Cz and two electrodes were placed on earlobes (i.e., averaged earlobes activity was used as the reference) with the forehead (Fpz) as ground. Electrode impedances were kept below 5 kΩ. Bio potentials recording system (g.USB Amp, g.tec) recorded the evoked potentials synchronized with auditory stimuli and digitized at 19200 Hz. Also, an online bandpass filter with cut off frequencies of 20 Hz and 2500 Hz was used.

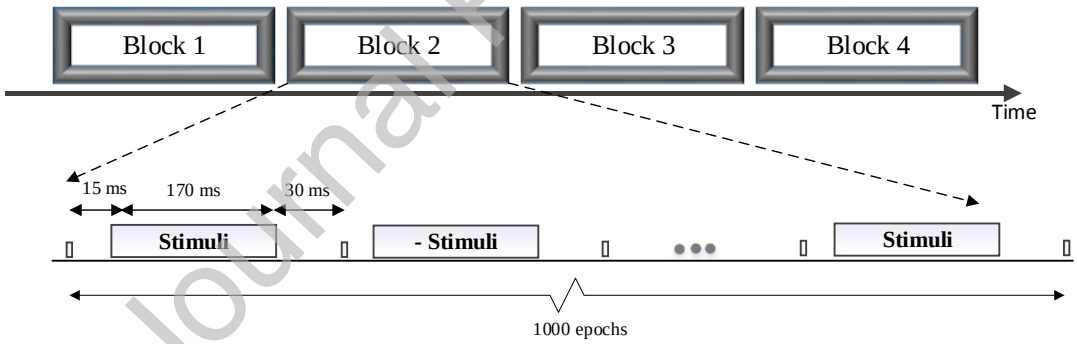

Figure 2: Schematic configuration of stimuli presentation paradigm. Arrows show the timing of the blocks. Each block contained 1000 epochs. The total duration of each epoch was 215 ms (i.e., 15 ms pre-stimulus + 170 ms stimulus + 30 ms post-stimulus).

2.4. Preprocessing

EEG recorded signals were filtered by fourth-order bandpass Butterworth filter with cutoff frequencies of 70 and 2000 Hz to isolate the brainstem responses. To avoid phase distortions, *filtfilt* function in MATLAB 2016b (MathWorks Inc., Natick, MA, USA) was used for zero-phase digital IIR filtering. Then, the EEG signals were divided into 215-ms segments corresponding to each epoch (15 ms pre-stimuli, 170 ms stimuli, and 30 ms post-stimuli). A threshold of $\pm$35 µV was applied to reject epochs containing myogenic artifacts. For each stimulus, the preprocessed segments were separately averaged according to their polarities and stored in channel A and B and then $(A+B)/2$ and $(A-B)/2$ were calculated. $(A+B)/2$ is considered mostly as speech-ABR responses [9]. ($A-B)/2$ signal mainly consists of high frequency spectral information[5] with contamination of noises and stimulus artifact [9].

Preprocessing method 1 (P1): LMS-based adaptive filtering

The LMS-based (Least mean squares-based) adaptive filter was used as the preprocessing method 1 **(P1)** for enhancing the speech-ABR responses. $(A+B)/2$ will highlight lower frequency components related to envelope information of signal with minimized stimulus artifact and cochlear microphonic (CM). $(A-B)/2$ biases the higher frequencies (spectral details) with maximum stimulus artifact contamination, cochlear microphonic and contains possible high frequency noise [9]. Therefore, we tried to remove the effect of (A-B)/2 components in (A+B)/2 signal by using LMS-based adaptive filter [37]. This process eliminated the undesirable components and preserves the envelope information while ignoring the spectral details in (A+B)/2.

According to Figure 3, plot A is obtained by subtracting channel B from channel A and contains spectral details. It may also consist of stimulus artifact and cochlear microphonic responses; Plot

B is the speech-ABR responses obtained from averaging of all the epochs in each block (i.e., (A+B)/2). These two signals are the inputs of the adaptive filter.

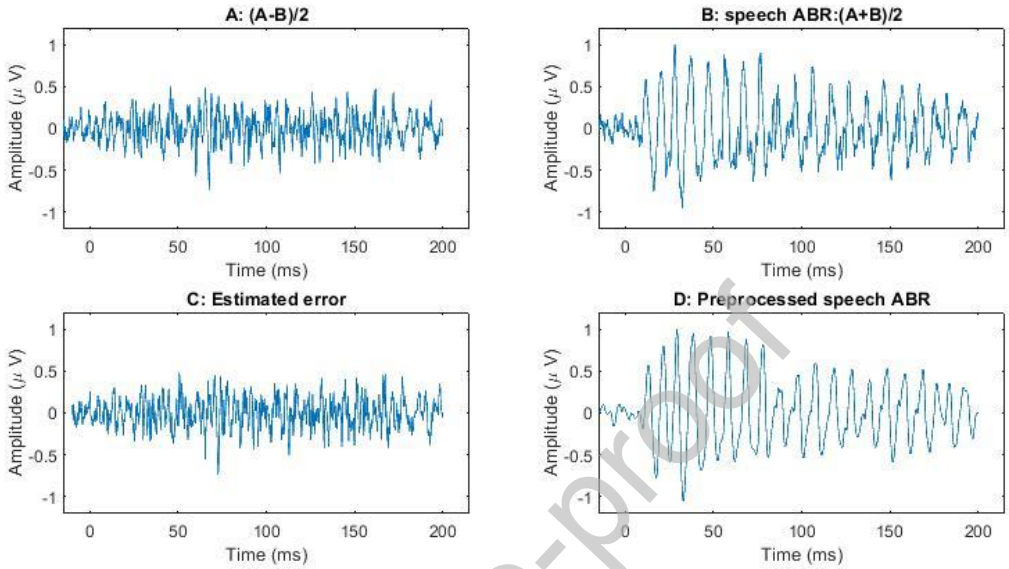

Figure 3: Speech-ABR preprocessing using LMS-based adaptive filter: A: (A-B) /2. B: (A+B)/2, C: error signal estimated by LMS-based adaptive filter, and D: Preprocessed speech-ABR.

The amplitude spectra of grand averages of $(A+B)/2$, $(A-B)/2$ and the preprocessed response by applying P1 method are shown in Figure 4. It can be inferred that $(A-B)/2$ has high frequency components, and also peaks at harmonics of fundamental frequency F0.

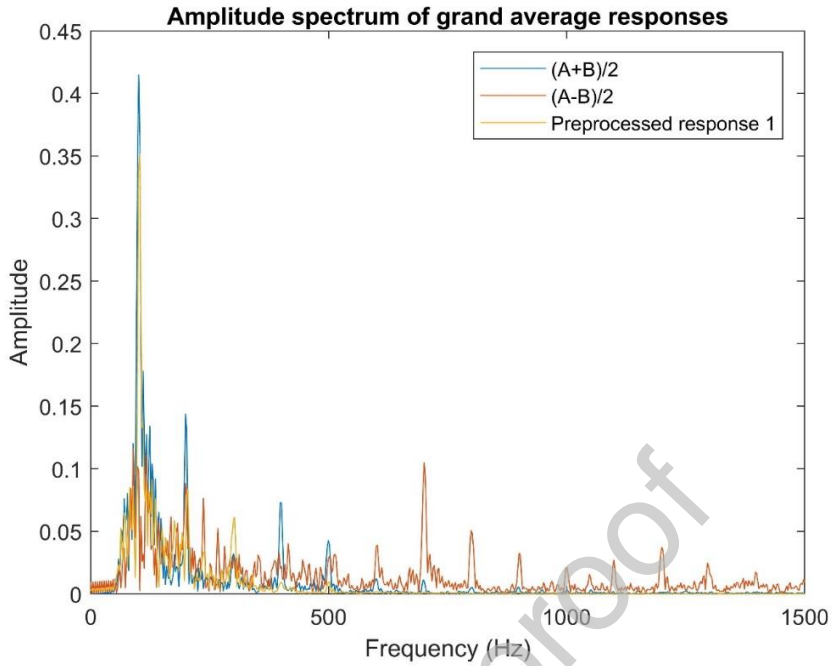

Figure 4: The amplitude spectra of (A+B)/2, (A-B)/2, and the preprocessed response by applying P1 method. The spectrum of (A-B)/2 contains more high frequency components.

Preprocessing method 2 (P2): Wavelet Denoising

Wavelet transform is a multi-resolution method, which decomposes signals by translation and dilation of a function (i.e., the so-called Mother Wavelet: ($\psi_{a,b}(t)$)). The continuous wavelet transform (CWT) of the signal x is defined as Equation 1, where a and b are the scale and translation parameters, respectively [38]:

$$W_\Psi X(a,b) = \langle x, \Psi_{a,b} \rangle\,,\ \Psi_{a,b} = |a|^{-\frac{1}{2}} \Psi(\tfrac{t-b}{a}) \quad (1)$$

Continuous wavelet transform provides redundant information. The practical usage of wavelet transform is the dyadic wavelet transform defined in discrete scales j and samples of k in which the scales and time are respectively $a_j = 2^j$ and $b_{jk} = 2^j k$. The DWT (Digital Wavelet

Transform) algorithm divides the signal into approximation and details coefficients at scales of j [38].

Auditory evoked potentials elicited in response to speech syllables are mixed with ongoing EEG activities. Therefore, the responses were decomposed using DWT into 5 levels with Mother Wavelet db6 for response enhancement and denoising. The signal recorded during the pre-stimulus period (15 ms before stimulus presentation time) was considered as the estimation of noise.

Donoho's denoising method:

Donoho and colleagues proposed a method for denoising, which tries to denoise responses by thresholding the wavelet coefficients at level j of signal x denoted by $X(j,k)$ [39].

For each scale of j, a threshold of T_j is obtained by Equation 2:

$$T_j = \sigma_j\sqrt{2lnN} \tag{2}$$

where N is the number of wavelet coefficients and σ_j is calculated by Equation 3:

$$\sigma_j = Median\{|X_{j,1} - \tilde{X}_j|, |X_{j,2} - \tilde{X}_j|, |X_{j,2} - \tilde{X}_j|, \dots\dots\dots, |X_{j,k} - \tilde{X}_j|\}/0.6745 \tag{3}$$

Denoising is done by hard thresholding of the coefficients $X(j,k)$ as presented in Equation 4:

$$X_{den}(j,k) = \begin{cases} X(j,k) & if\ |X_{j,k}| > T_j \\ 0 & if\ |X_{j,k}| < T_j \end{cases} \tag{4}$$

By applying the threshold (T_j) on wavelet coefficients, the denoised coefficients (X_{den}) were obtained. The denoised signals were reconstructed from these coefficients. A sample of the denoised signals by preprocessing method 2 (P_2) is shown in Figure 5.

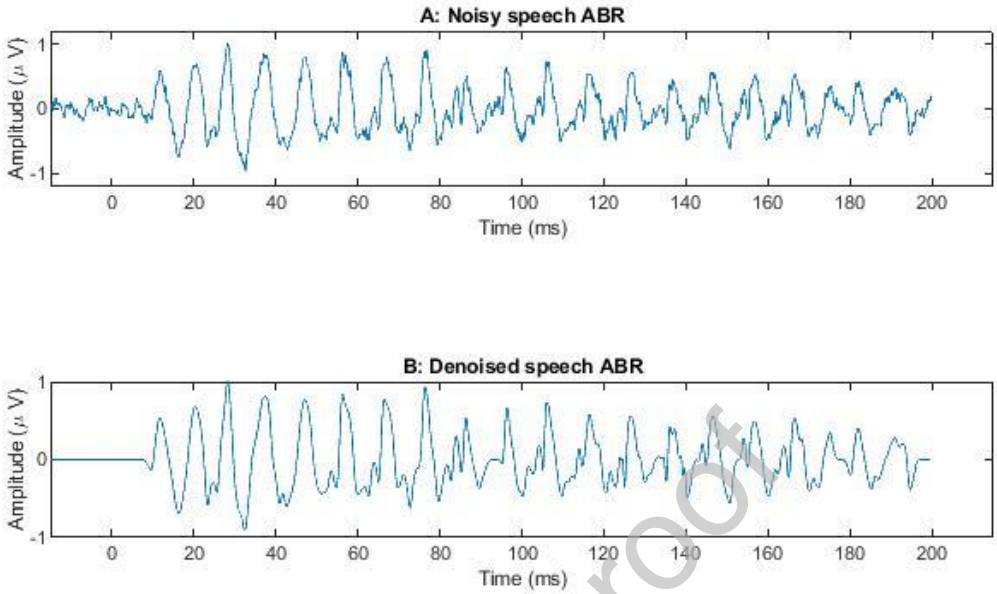

Figure 5: The results of preprocessing method 2: A: noisy speech ABR, B: denoised speech-ABR using Donoho method.

The overall procedures of preprocessing are demonstrated in Figure 6

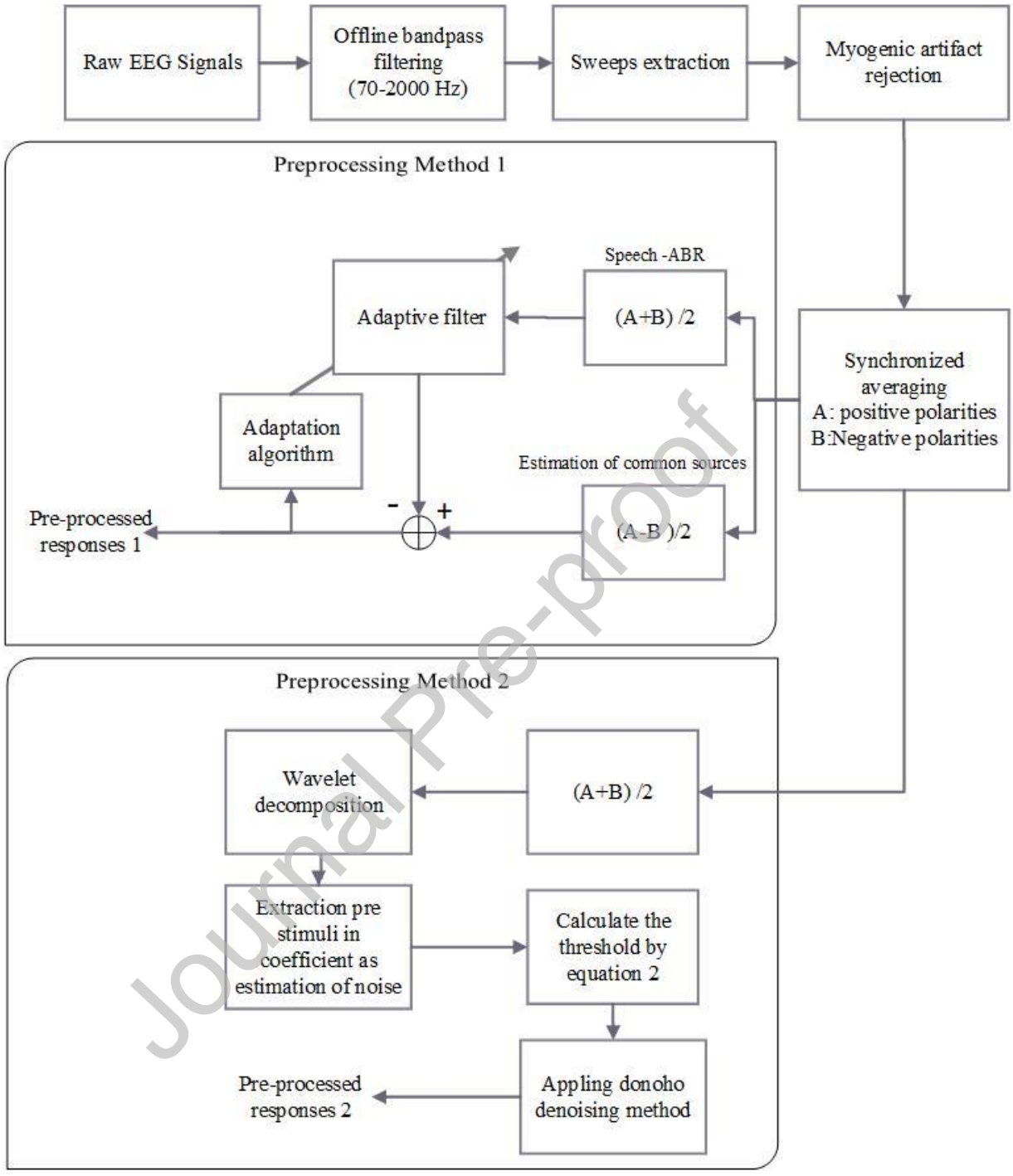

Figure 6: The block diagram of speech ABR extraction from raw EEG and enhancing the responses.

2.5. Feature extraction

The objective features for discriminating the responses to consonant-vowels /da/, /ba/ and /ga/ are extracted in the time, frequency and time-frequency domains. Because the vowel in these syllables is the same and they only differ in their initial part, which is related to the formant transition, we extracted the features in the first 50 ms of the responses.

All of the features were extracted from the responses in three steps: the first step (P0) included the responses obtained from synchronized averages of negative and positive polarities ($(A+B)/2$); the second and third steps including the preprocessed response 1 and preprocessed response 2, respectively (Figure 6).

Time domain features

The peaks of the transient part of the responses can represent the differences between the responses to consonant vowels /da/,/ba/ and /ga/ in the time domain[3]. However, marking these peaks requires expertise and it is subjective, time-consuming, and error-prone. Here, the time domain features were considered, which could be objectively represented. We have selected energy, correlation coefficient, entropy, and cosine distance as the time domain features (Table 1)[40]. The p is the speech ABR response and the grand average of the responses is p_{ref}.

Table 1: The definitions of time domain features.

Time domain features	
Dissimilarity type	Definition
Energy	$\sum_{i=1}^{n} p_i^2$
Entropy	$\sum_{i=1}^{n} p_i \log p_i$
Cosine distance	$\frac{\sum_{i=1}^{N}(p_i\, p_i{}^{ref})}{\sqrt{(\sum_{i=1}^{N} p_i\ \sum_{i=1}^{N} p_i{}^{ref})}}$
Correlation coefficient	$\rho_{p_i,p_i{}^{ref}} = \frac{cov(p_i,p_i{}^{ref})}{\sigma_p\ \sigma_{p^{ref}}}$

Frequency domain features

Frequency representation of speech ABRs could encompass a measure of phase-locking to the presented stimuli. The amplitude and phase of speech ABR spectrum were calculated in the frequency domain of responses that include the fundamental frequency ($F_0 = 100$ Hz), its harmonics (H_2 to H_{10}), and first formant interval ($F_1 = [400\ 720]$ Hz). These frequencies are in the frequency range that the phase-locking property of the brainstem exists. After feature extraction, the random subset feature selection (RSFS) method was used for feature selection.

Time-Frequency domain features

Speech-ABRs are nonstationary signals. Therefore, conventional analysis methods such as time and frequency domain analysis are not completely suitable for processing of these responses.

Therefore, there is a crucial need for an alternative domain being capable of concurrent representation of both the time and frequency characteristics of the speech ABRs.

Among various time-frequency analysis techniques, the wavelet transform uses variable-length windows employing translation and dilation of a function, so it provides a multi-resolution analysis approach in time and frequency domain. This transform has been widely used in the biomedical signal analysis [41] because the most interesting features of those signals are simultaneously localized in time and frequency. Figure 7 illustrates the spectrogram of stimuli and responses of one representative subject. According to Figure 1, the stimuli do not reveal significant differences in the time domain, while we can see there are obvious differences between the stimulus spectrograms in the formant transition period and the frequencies above the phase-locking property of the auditory system (shown with yellow ovals). Therefore, there may be some differences in time-frequency domain features of the responses exist, that could be identified by time-frequency analysis of responses.

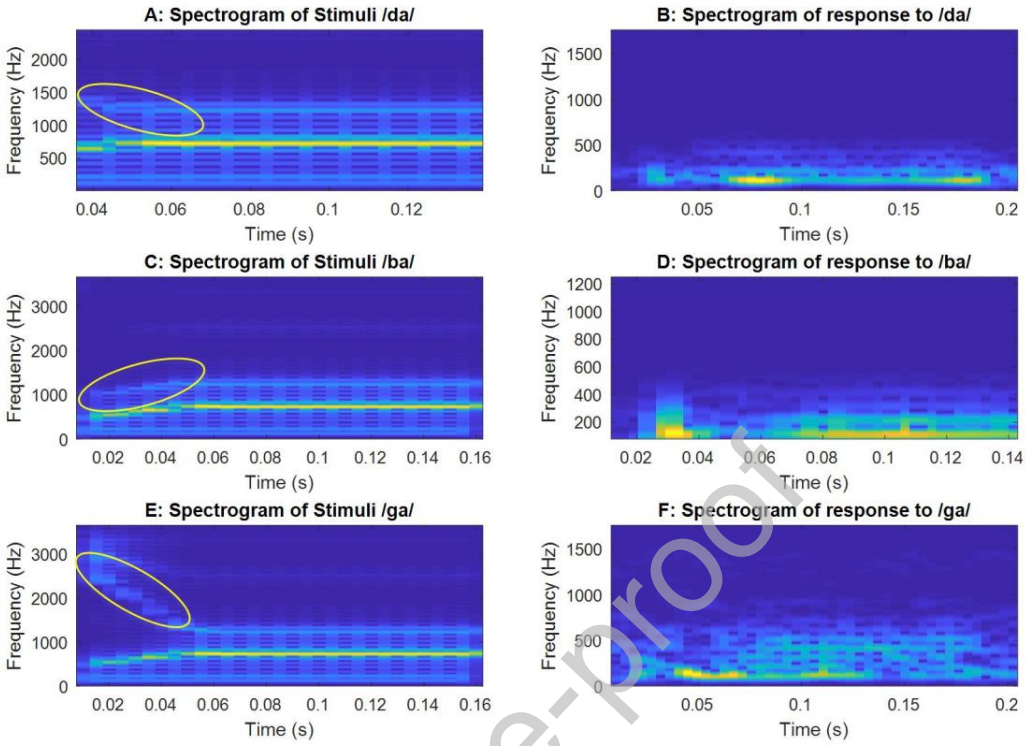

Figure 7: The spectrogram of stimuli and their corresponding speech-ABRs of one representative subject. The yellow ovals show the differences between the presented stimuli.

A promising wavelet packet-based feature extraction method is the local discriminant bases (LDB) algorithm [42]. In this algorithm, a basis function is selected from a dictionary of computationally efficient orthonormal bases that emphasizes the dissimilarities among classes. Subsequently, all the coefficients corresponding to the selected bases are considered as a feature space and will be given to the classifier.

LDB algorithm tries to find the bases, which reveal the maximum dissimilarities between two signals, that are similar in time or frequency domain. Brainstem responses to /ba/, /da/ and /ga/ are very similar in time and frequency domains. Thus, the LDB algorithm can find dissimilarities

in time-frequency domain and can be a promising choice for the classification of the brainstem responses to /ba/, /da/ and /ga/.

Wavelet Packet

The wavelet packet transform (WPT) is a time-frequency domain analysis approach that provides a great deal of freedom in analyzing non-stationary signals. WPT leads to a binary tree of orthonormal bases (the so-called nodes), which is shown in Figure 8. Unlike wavelet transform, WPT divides not only the low but also the high-frequency sub-bands; however, bandwidths of the nodes in each level of WPT decomposition are equal. Traditional wavelet transforms split the input signal into approximation and detail sub-bands. Then, the second level approximation sub-band is divided into approximation and detail and this process continues until reaching the last level. In contrast, WPT splits not only approximation sub-bands but also detail sub-bands into two sub-bands. This capacity of WPT to divide detail sub-bands provides more than one basis function at a given scale. If all sub-bands in each level are decomposed, the result is a complete basis tree. The previous level nodes (i.e., parents) lead to new nodes that are called children.

Let us assume that Ω_{00} denotes the first parent node from WPT decomposition (i.e., the original signal). According to the representation of Figure 8, node Ω_{jk} is the *k*th node in the jth level of decomposition. The relation between the child ($\Omega_{j+1\ 2k}$ and $\Omega_{j+1\ 2k+1}$) and parent nodes (Ω_{jk}) is defined in Equation 5. This process continues until reaching the last level of decompositions.

$$\Omega_{jk} = \Omega_{j+1\ 2k} \oplus \Omega_{j+1\ 2k+1} \quad (5)$$

The symbol $\oplus$ shows the direct sum of two subspaces $\Omega_{j+1\ 2k}$ and $\Omega_{j+1\ 2k+1}$.

The aim is to find the best subspaces that provide maximum dissimilarities between the three classes of the responses. This can be done by pruning the WPT tree using LDB.

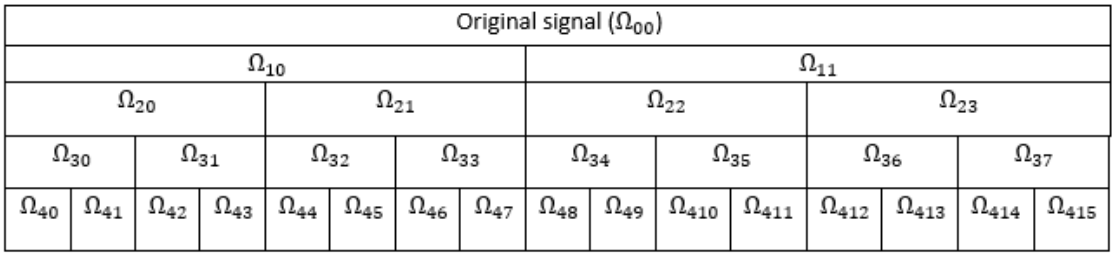

Figure 8: The schematic representation of the wavelet packet decomposition tree in four levels, where Ω_{00} is the original signal and Ω_{ij} represents the jth node of the ith level of decomposition.

Time-Frequency Feature Selection

The wavelet packet decomposition was performed in seven levels decomposition. The number of decomposition levels was determined by MATLAB R2016b built-in function *wmaxlev*. The db6 was chosen as Mother Wavelet because it is symmetric and suitable for feature detection. This kind of Gaussian wavelet function is better for extraction of uncommon features between classes [43]. Figure 9 denotes the waveforms of scaling and wavelet functions of db6. Then, the following process was carried out for extracting the time-frequency features from the observations.

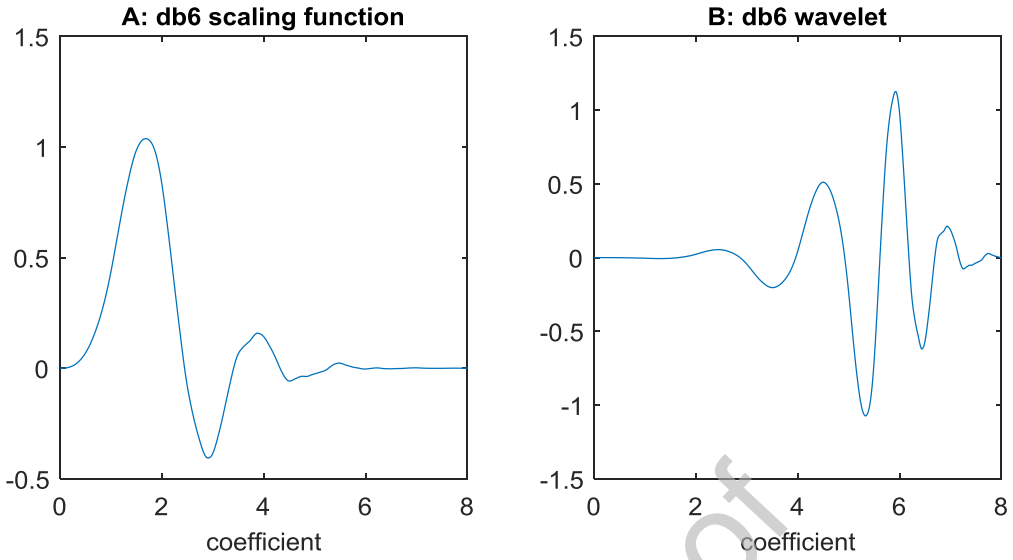

Figure 9: Mother Wavelet db6, plot A is the waveform of scaling and plot B is a wavelet function of db6.

Local Discriminant Bases (LDB)

Assume that A_{jk} and Δ_{jk} are the local discriminant basis of Ω_{jk} and the discriminant criterion, respectively. LDB algorithm works as follows [44]:

Step 1: Decompose the responses using WPT.

Step 2: Choose a discriminant criterion to be applied to each node in the WP tree.

Step 3: For j = max level, assume $A_{jk} = \Omega_{jk}$ and calculate the discriminant criterion Δ_{jk} for each node (k =0 ,…,$2^j - 1$)

Step 4: find the best bases with the following rules:

$$\begin{cases} A_{jk} = \Omega_{jk} & \text{if } \Delta_{j,k} \geq \Delta_{j+1,2k} + \Delta_{j+1,2k+1} \\ A_{jk} = A_{j+1\ 2k} \cup A_{j+1\ 2k+1} & \text{otherwise} \end{cases}$$

Step 5: Sort the selected nodes according to their discriminant power.

Step 6: Choose the nodes whose power is more than the mean of discriminant powers for all nodes to classify the responses.

Dissimilarity Measures:

Here, similar to time domain features, relative entropy, energy difference, correlation index, Jenson difference and cosine distance of each node in the wavelet packet decomposition were used as dissimilarity measures.

Relative Entropy:

In the best bases selection algorithm of Coifman et al [45], each subspace utility in the WPT tree is measured by Shannon entropy. This measurement focuses on the distribution of energy in each subspace. The relative entropy is calculated according to the following Equations 6 and 7 (the discriminant criterion is $\delta_{j,k}$).

$$\delta_{j,k}(p_{j,k}^1, p_{j,k}^2) = \sum_{i=1}^{n} p_i^1 \log \frac{p_i^1}{p_i^2} \tag{6}$$

$$p_{j,k}^1 = psd(\Omega_{jk}^1) \tag{7}$$

Where $\sum_{i=1}^{n} p_{jk}^1(i) = 1$ and $\sum_{i=1}^{n} p_{jk}^2(i) = 1$ (p_{jk}^1 and p_{jk}^2 are nonnegative sequences. The superscripts 1 and 2 denote two different sequences). Also, *psd* denotes the power spectral density function. This criterion measures how much p_{jk}^1 and p_{jk}^2 are differently distributed. The probability density function of each node can be considered as nonnegative sequences p_{jk}^1 and p_{jk}^2 representing energy distribution). Then, the relative entropy of the two sequences of probability density function can be used as the discriminant criterion in the LDB algorithm. This

criterion will be zero if p_{jk}^1 and $p_{jk}^2$are the same. High values of the relative entropy indicate the large power of the discrimination of two nodes.

Energy Difference

Normalized energy difference was calculated in each node according to Equations 8 and 9:

$$\delta_{j,k} = E\left(\Omega_{jk}^1\right) - E\left(\Omega_{jk}^2\right). \quad (8)$$

$$E = \sum_{i=1}^{N} \left(\Omega_{\,jk}\ (i)\right)^2 / \sum_{i=1}^{N} \left(\Omega_{\,00}\ \right)^2. \quad (9)$$

Jenson Difference

The Jenson difference is used as a similarity measure for two nodes:

$$Jenson\left(\Omega'^1_{jk}, \Omega'^2_{jk}\right) = \quad (10)$$

$$\frac{1}{2}\sum_{i=1}^{k}\{\Omega'^1_{jk}\ log_2\Omega'^2_{jk}\ +\ \Omega'^2_{jk} log_2\Omega'^1_{jk} - \left(\Omega'^1_{jk} + \Omega'^2_{jk}\right) log_2\left(\frac{\Omega'^1_{jk} + \Omega'^2_{jk}}{2}\right)$$

Where

$$\Omega'^1_{jk} = \Omega^1_{jk} / \sum \Omega^1_{jk},\ \Omega'^2_{jk} = \Omega^2_{jk} / \sum \Omega^2_{jk}$$

The Jenson difference was used to define a dissimilarity measure as follows:

$$\delta_{j,k} = 1 - Jenson(\Omega^1_{jk}, \Omega^2_{jk}) \quad (11)$$

Cosine Distance

Cosine distance, which is defined as an angle-based similarity measure reveals the similarity of two normalized sequences (Table 1) and here it is used as a dissimilarity measure of nodes as Equation 12:

$$\delta_{j,k} = 1 - cosine(\Omega^1_{jk}, \Omega^2_{jk}). \quad (12)$$

The total dissimilarity measures are defined as Equation 13:

$$\Delta_{j,k} = \sum_{a=1}^{L-1} \sum_{b=a+1}^{L-1} \delta_{jk} \left(\Omega_{jk}^{a}, \Omega_{jk}^{b}\right) . \quad (13)$$

Where L is the number of classes (in this study, it is three). The detailed method proposed by this study is shown in Figure 10.

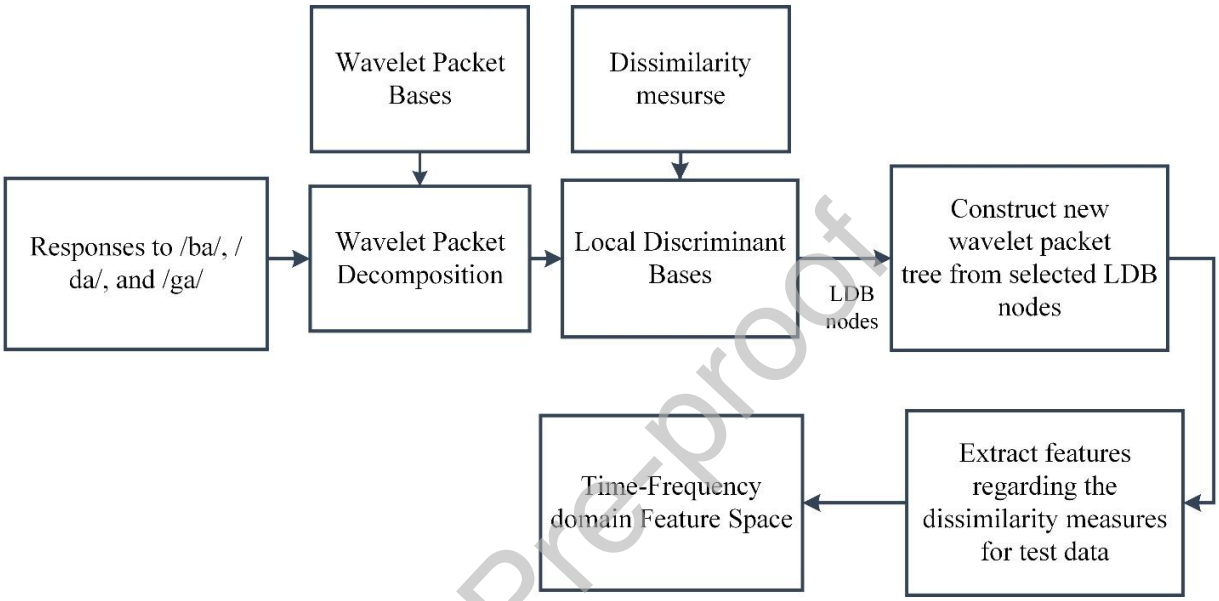

Figure 10: The schematic representation of extracting time-frequency domain features using LDB algorithm and classification.

Feature selection process:

In this study, we used random subset feature selection (RSFS) method as a feature selection algorithm [46]. This method tries to find a subset of features that are advantageous in the intended classification problem. The feature space is randomly divided into all possible subspaces. Then, the features selected by classification performances of these subspaces.

RSFS algorithm [46]:

1) Select randomly by uniformly distributing a subset of features f_x from the total feature space (F).
2) Apply the KNN classification for the selected subset and measure the value of the true positive rate as a criterion function.
3) Calculate and update the relevance r_x values for all features f_x from Equation 14.

$$r_x \leftarrow r_x + c_i - E\{c\} \tag{14}$$

Where c_i is the positive rate in iteration i and $E\{c\}$ is the mean value of all positive rates of previous iterations.

4) Repeat the above steps for new feature subspace.
5) Select randomly of dummy features subspace.
6) Update the relevance values of the dummy feature.
7) Get the probability of normal cumulative distribution function of true features which is higher than the mean of relevance values of dummy features.
8) Select the features with the probability above 0.99.

This algorithm was used for feature selection in time, frequency and time-frequency domain separately. We also applied it to all features obtained to select the combined features.

2.6. Classification:

The selected features from 324 observation (27subjects × 4 blocks × 3 syllables) were fed into four different classifiers: the k-nearest neighbor (KNN), naive Bayes (NB), multiclass support vector machine (MSVM) and discriminant analysis (DA). Figure 11 shows the feature classification process. The leave-one-out approach was used for cross-validation. According to

Table 2, the general performance of the classifiers is indicated by the accuracy and validity (i.e., the weighted average of the sensitivity of each class), which are defined in Equations 15 and 16, respectively. They were obtained from the confusion matrices.

Table 2: The confusion matrix of classification

		Actual response		
		/da/ Class 1	/ba/ Class2	/ga/ Class3
Predicted response	/da/ Class1	C_{11}	C_{12}	C_{13}
	/ba/ Class2	C_{21}	C_{22}	C_{23}
	/ga/ Class3	C_{31}	C_{32}	C_{33}

$$Total\ Accuracy = P_1 \frac{C_{11}}{C_{11} + C_{12} + C_{13}} + P_2 \frac{C_{22}}{C_{22} + C_{21} + C_{23}} + P_3 \frac{C_{33}}{C_{33} + C_{32} + C_{31}} \quad (15)$$

$$Total\ Validity = P_1 \frac{C_{11}}{C_{11} + C_{21} + C_{31}} + P_2 \frac{C_{22}}{C_{22} + C_{12} + C_{32}} + P_3 \frac{C_{33}}{C_{33} + C_{23} + C_{13}} \quad (16)$$

$$P_l = \frac{n_l}{\sum_{l=1}^{3} n_l}$$

where n_l $(l = 1, 2, 3)$ is the number of data in class l. C_{ij} is the number of data, which belong to the jth class and were predicted as an ith class.

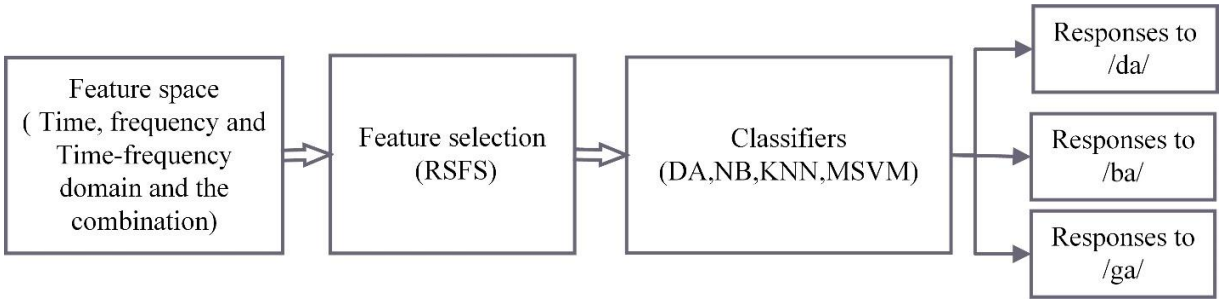

Figure 11: The schematic representation of the classification process.

Discriminant analysis:

Discriminant analysis is a classification method introduced by R.A Fisher in 1936, and it considers the data in different classes that have various Gaussian distributions [47].

Naïve Bayes:

The Naïve Bayes classifier is a Bayesian classifier, which classifies the feature space based on the conditional probability of different classes given the features (i.e., posterior probabilities for each class) and tries to assign each observation to the class that has the highest posterior probability [48].

K-Nearest Neighbor (KNN):

KNN is a common and well-known classifier for ordering complex problems with high performance [49]. This method assumes a distance type and tries to classify the feature spaces. A class is constructed by features, in which their distance is close to each other and far between other classes. For having a good classification performance, distance type and the number of neighbors are two important parameters in KNN classifier that need to be optimized.

Multiclass SVM:

Support vector machines determine the optimal hyperplane, which maximizes the margin and are used for binary classification, so many studies have been done to develop SVM for multiclass classification solutions [50-53]. Here, we employed one versus one approach for our three-class classification problem.

MATLAB built-in functions, namely *fitcdiscr*, *fitcnb*, *fitcknn* and *fitcecoc* were put to use for optimizing the classifiers DA, NB, KNN, and MSVM, respectively.

Statistical Analysis:

Statistical analysis was performed with SPSS 16 (SPSS Inc. IBM Corp, Chicago) to compare the results of the preprocessing methods of P_0, P_1, and P_2. The classification results in time, frequency and time-frequency domain features for three preprocessing methods, namely P_0, P_1, and P_2, were compared.

Repeated measures ANOVA with a Greenhouse-Geisser correction with a significance level $\alpha = 0.05$ was employed to analyze the differences between the accuracies of preprocessing methods P_0, P_1, and P_2 for time domain features.

Regarding frequency and time-frequency domain accuracies, non-parametric Friedman test with a significance level $\alpha = 0.05$ was employed. Then, the posthoc analysis with Bonferroni correction was applied to compare the results of different preprocessing methods while α set 0.016.

3. Results

Classification performance:

The evaluation of extracted features in three domains of time, frequency and time-frequency were done by calculating the accuracy and validity of classification of the three classes with a chance level of 33.33 %.

The classification results are reported for the different preprocessing approach of P_0, P_1, and P_2.

Time domain features classification results:

The time domain extracted features provided the lowest classification accuracy.The maximum classification accuracy was 50.31 % obtained from correlation features that were selected by RSFS classified by MSVM classifier and it was related to the preprocessing method 2. The detailed results of classification according to the time domain features are shown in Table 3.

Table 3: The results of the time domain features classification.

		Time domain features classification results		
		P_0	P_1	P_2
KNN	**Total Accuracy (%)**	47.8	44.0	45.0
	Total Validity (%)	44.6	44.0	43.4
NB	**Total Accuracy (%)**	42.3	45.0	44.7
	Total Validity (%)	42.1	39.9	44.8
DA	**Total Accuracy (%)**	46.9	44.0	48.4
	Total Validity (%)	44.8	45.1	43.8
MSVM	**Total Accuracy (%)**	40.3	43.1	50.3
	Total Validity (%)	37.8	43.6	48.4

Frequency domain features classification results:

The frequency domain features were extracted with a resolution of 1 Hz up to 1500 Hz. Therefore, the order of frequency features was put to order incrementally from low frequencies to high. Then, RFSF algorithm selected the most powerful features. The relevance value of features extracted from RSFS is related to positive rate values. Therefore, from the relevance values, the relative goodness of features could be inferred. The relevance values of all the extracted and selected features of the frequency domain are shown in Figure 12. It can be observed that the relevance value of selected phase features in the frequency domain are higher than that of amplitude features. The results of classification derived from these features are reported in Table 4.

In terms of classification performance, the frequency domain features outperformed time domain features. Moreover, the phase features results were better compared to the amplitude of frequency domain features. Here, the best classification accuracy was 80.25% obtained by KNN classifier and the preprocessing method 2 (Figure 14).

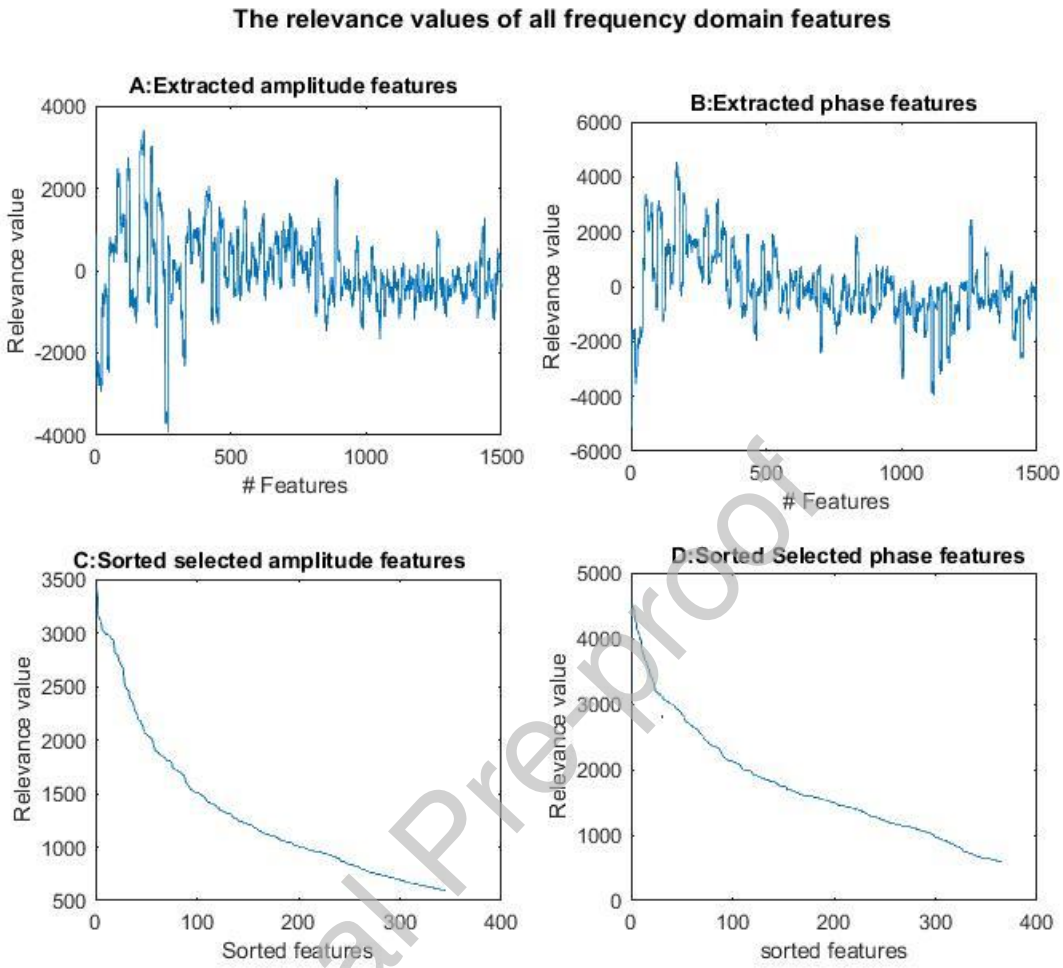

Figure 12: The relevance values of extracted and selected frequency domain features. A and B represent the relevance values of extracted amplitude and phase features of the frequency domain, while C and D show the relevance values of selected amplitude and phase features, respectively. It is noteworthy that the feature number in A and B match with frequency order.

Table 4: The results of frequency domain features classification.

		Frequency domain features classification results					
		Amplitude features			**Phase features**		
		P_0	P_1	P_2	P_0	P_1	P_2
KNN	**Total Accuracy (%)**	53.8	56.3	58.0	70.9	69.1	80.2
	Total Validity (%)	53.6	57.9	56.7	72.3	69.1	80.2
NB	**Total Accuracy (%)**	49.3	44.3	58.3	66.9	69.3	70.9
	Total Validity (%)	49.3	45.9	56.9	67.4	68.0	70.8
DA	**Total Accuracy (%)**	47.2	52.3	58.9	63.8	65.2	74.6
	Total Validity (%)	47.4	52.1	57.7	64.5	65.4	74.8
MSVM	**Total Accuracy (%)**	63.5	60.3	61.1	67.9	69.0	77.4
	Total Validity (%)	63.5	60.4	59.6	68.5	68.5	77.4

Time-Frequency domain features classification results:

The discriminative power of selected nodes of the wavelet packet obtained from step 3 of the LDB algorithm is shown in Figure 13. The classification results derived from these features are

reported in Table 5. The best classification accuracy of these features was 91.36 % and its corresponding validity was 93.14 % obtained via MSVM classifier and the preprocessing method2.

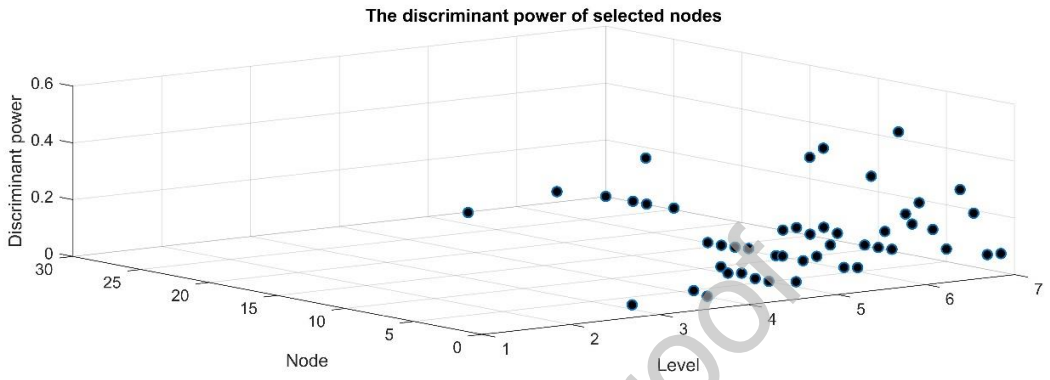

Figure 13: The discriminant power of selected nodes in time-frequency domain subspaces

Table 5: The time-frequency domain features classification results

		The Time-Frequency domain features classification results		
		P_0	P_1	P_2
KNN	**Total Accuracy (%)**	73.5	75.4	91.3
	Total Validity (%)	73.7	75.4	93.1
NB	**Total Accuracy (%)**	61.1	70.3	71.6
	Total Validity (%)	61.4	69.7	73.7
DA	**Total Accuracy (%)**	63.5	67.9	72.5
	Total Validity (%)	64.6	67.8	73.7
MSVM	**Total Accuracy (%)**	79.6	80.0	89.5
	Total Validity (%)	86.2	83.1	89.6

Combined features classification results:

The results of classification accuracies derived from combined features are reported in table 6. The best result achieved for combined features was 97.5 % classification accuracy and 97.7 %

classification validity obtained by MSVM classifier and the preprocessing method 2. Figure 14 illustrates the confusion matrices of the best classification results of all the feature categories. It can be inferred that the combined features provided the highest classification accuracy and validity.

Table 6: The classification results of combined features

		Combined features classification results		
		P_0	P_1	P_2
KNN	**Total Accuracy (%)**	84.2	85.4	96.3
	Total Validity (%)	84.3	85.7	96.3
NB	**Total Accuracy (%)**	76.5	76.2	84.8
	Total Validity (%)	76.6	76.8	85.0
DA	**Total Accuracy (%)**	80.8	80.5	88.8
	Total Validity (%)	80.9	80.9	88.8
MSVM	**Total Accuracy (%)**	88.5	87.0	97.5
	Total Validity (%)	88.7	87.2	97.7

Statistical Results:

There were significant differences between accuracies of the preprocessing methods for time domain features (F (1.742,54.011) =11.632, P=0.001, Repeated measures ANOVA)**.** Post-hoc analysis using Bonferroni correction with α= 0.016 showed significant differences between the preprocessing method P_2 compared to P_1 and P_0. Nonetheless, no significant difference was found between (P_2, P_0) and (P_1, P_0) pairs.

Concerning frequency domain features, significant differences between the results of preprocessing methods were observed ($\chi^2(2) = 13.625$, P=0.001, Friedman). Post hoc analysis with Wilcoxon signed-rank tests with Bonferroni correction resulted in a significance level of P < 0.016. There were no significant differences between P_1 and P_0 (Z=-0.207, P=0.836), In contrast, there were significant differences between the results of P_2 and P_1 (Z=-3.39, P=0.001). Also, the results of P_2 was significantly higher than P_0 method (Z=-3.258, P=0.001).

For the time-frequency features, there were significant differences between the conditions ($\chi^2(2) = 12.25$, P=0.002, Friedman). More specifically, there was no significant difference between the accuracies of P_2 and P_1 (Z=-0.7, P=0.484, Wilcoxon signed-rank), but the accuracy results derived from P_2 was significantly higher than that of P_1 (Z=-2.521, P=0.012, Wilcoxon signed-rank) and P_0 (Z=-2.521, P=0.012, Wilcoxon signed-rank).

Figure 14: The confusion matrix of best classification results. The combined features lead to a higher classification accuracy of 97.5%.

4. Discussion

The purpose of this study was to objectively discover the dissimilarities of speech-ABRs to three different consonant-vowel syllables (i.e., /ba/, /da/, and /ga/). These syllables are very similar to each other and differ only in their F2 and F3 formants in the formant transition period. Achieving this purpose may help us understand how the brainstem processes different consonant-vowels. Johnson and colleagues [3] showed that the responses to these consonant-vowels are different in terms of their latencies in the formant transition period. They showed that the latency variability was due to the differences in second and third formants that are above the phase-locking frequency range of the brainstem.

The consonant vowels with higher formant frequencies (especially F2 and F3) shift the minor peaks of responses to the lower latencies [3]. Latency differences also affect the phase information of the responses. Our objective classification results based on the phase of frequency domain features confirm this effect. The higher classification accuracy based on phase compared to amplitude information indicates that the phase information conveys more valuable information. This is consistent with Skoe and colleagues finding that there is good information in the cross-pahseogram of the responses [36]. Besides, Sadeghian et al. objectively classified different vowels with the highest accuracy of 83.33% [29]. They confirmed that the auditory responses to different speech stimuli could be automatically classified into distinct classes by using the combined transient and sustained features. Therefore, these kinds of studies might help to find useful features as biomarkers for clinical applications.

Also, the results confirm that the encoded information of the consonant-vowel stimuli in the subcortical regions can be identified by using the time-frequency analysis on the brainstem responses. In other words, the brainstem responses to speech stimuli contain important information for discriminating the different speech stimuli in the time-frequency domain.

This study was designed to extract objective features in the different domains of time, frequency and time-frequency. Besides, we used three preprocessing methods and the statistical results indicated significant differences between the preprocessing methods. Specifically, the preprocessing method 2 (P_2: wavelet denoising) had better performance. Also, no significant effect was observed in the results concerning the preprocessing method 1 (P_1: LMS-based adaptive filtering) for enhancing the responses. The results indicate that the wavelet denoising method that is removing noise in both time and frequency domain simultaneously had better performance. However, using an adaptive filter for preprocessing did not lead to significantly better classification results. This may show that wavelet denoising by using pre-stimulus data as an estimation of noise is more successful compared to adaptive filter in our study.

The extracted frequency features were richer than the time domain ones (see Table 3 and Table 4). The frequency information in the F2 and F3 formants of consonants vowels /ba/, /da/ and /ga/ are different. These formants are mostly above the phase-locking range properties of the brainstem [8, 54, 55]. Therefore, it is not expected these formants to appear in the spectrum of the responses with considerable power. However, the effect of these formant differences in stimuli leads to lag differences between the responses [8] and in amplitude and phase of the spectrum of responses in lower frequencies. This is maybe due to the nonlinear processing of the brainstem in lower frequencies [56]. This point could be seen in the relevance values of RSFS (reflecting the power of discrimination among those features) for phase and amplitude features of the frequency domain (Figure 12).

The time-frequency features results were better compared to the frequency and time domain features. This emphasizes that the time-frequency domain information is richer than others, and they contain valuable information in the time-frequency subspaces for speech discrimination in the subcortical region. Besides, consonant-vowels are distinguishable because of their spectral and temporal characteristics at the same time. Therefore, both temporal and spectral information of the responses are discriminant at the same time. This could replicate the acoustic possibilities of distinguishing the consonant-vowel sequences which means that the starting point of spectral components (formants) is equally important than the progression of formants in time (formant transitions). Then, we combined features to take advantage of each category to increase classification accuracy. This led to higher accuracy of 97.5 % in the objective classification of responses to the similar consonant-vowels.

We used four different classifiers to evaluate the features. All of them were relatively successful in classifying the responses. Among these classifiers, MSVM and KNN were more successful. The superior performance of MSVM is may because of its capability to solving the classification question of high dimensional feature space using their various kernel function [57, 58] .

Future studies may be helpful to classify the speech-ABRs to consonant-vowels using advanced nonlinear methods. Moreover, using advanced methods for speech-ABR denoising and preprocessing may improve the classification accuracy and reduce the number of needed epochs

for averaging. Therefore, it may become possible to automatically separate speech-ABR to words and other complex speech stimuli.

Conflict of Interests

None of the authors have potential conflicts of interest to be disclosed.

Acknowledgment

This work was supported by a grant no.30946 from the deputy of the research review board and ethics community of Tehran University of Medical Sciences.

Declaration of competing interests

We wish to confirm that there are no known conflicts of interest associated with this Publication and there has been no significant financial support for this work that could have influenced its outcome.

We confirm that the manuscript has been read and approved by all named authors and that there are no other persons who satisfied the criteria for authorship but are not listed. We further confirm that the order of authors listed in the manuscript has been approved by all of us.

We confirm that we have given due consideration to the protection of intellectual property associated with this work and that there are no impediments to publication, including the timing of publication, with respect to intellectual property. In so doing we confirm that we have followed the regulations of our institutions concerning intellectual property.

We further confirm that any aspect of the work covered in this manuscript that has involved either experimental animals or human patients has been conducted with the ethical approval of all relevant bodies and that such approvals are acknowledged within the manuscript.

References

1. Burkard, R.F., J.J. Eggermont, and M. Don, *Auditory evoked potentials: basic principles and clinical application*. 2007: Lippincott Williams & Wilkins.
2. Russo, N., et al., *Brainstem responses to speech syllables.* Clinical Neurophysiology, 2004. **115**(9): p. 2021-2030.
3. Johnson, K.L., et al., *Brainstem encoding of voiced consonant–vowel stop syllables.* Clinical Neurophysiology, 2008. **119**(11): p. 2623-2635.
4. Anderson, S. and N. Kraus, *The potential role of the cABR in assessment and management of hearing impairment.* International journal of otolaryngology, 2013. **2013**.
5. Aiken, S.J. and T.W. Picton, *Envelope and spectral frequency-following responses to vowel sounds.* Hearing research, 2008. **245**(1-2): p. 35-47.
6. !!! INVALID CITATION !!! [6-9].
7. Dajani, H.R., B.P. Heffernan, and C. Giguere. *Improving hearing aid fitting using the speech-evoked auditory brainstem response*. in *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*. 2013. IEEE.
8. Greenberg, S., *Wpp, no. 52: Temporal neural coding of pitch and vowel quality.* 1980.
9. Skoe, E. and N. Kraus, *Auditory brainstem response to complex sounds: a tutorial.* Ear and hearing, 2010. **31**(3): p. 302.
10. Banai, K., D. Abrams, and N. Kraus, *Sensory-based learning disability: Insights from brainstem processing of speech sounds.* International Journal of Audiology, 2007. **46**(9): p. 524-532.
11. Chandrasekaran, B. and N. Kraus, *The scalp-recorded brainstem response to speech: Neural origins and plasticity.* Psychophysiology, 2010. **47**(2): p. 236-246.
12. Martin, B.A., K.L. Tremblay, and P. Korczak, *Speech evoked potentials: from the laboratory to the clinic.* Ear and hearing, 2008. **29**(3): p. 285-313.
13. Kraus, N. and T. Nicol, *Brainstem origins for cortical 'what'and 'where'pathways in the auditory system.* Trends in neurosciences, 2005. **28**(4): p. 176-181.
14. Johnson, K.L., T.G. Nicol, and N. Kraus, *Brain stem response to speech: a biological marker of auditory processing.* Ear and hearing, 2005. **26**(5): p. 424-434.
15. Cunningham, J., et al., *Neurobiologic responses to speech in noise in children with learning problems: deficits and strategies for improvement.* Clinical Neurophysiology, 2001. **112**(5): p. 758-767.
16. Plyler, P.N. and A. Ananthanarayan, *Human frequency-following responses: Representation of second formant transitions in normal-hearing and hearing-impaired listeners.* Journal-American Academy of Audiology, 2001. **12**(10): p. 523-533.
17. King, C., et al., *Deficits in auditory brainstem pathway encoding of speech sounds in children with learning problems.* Neuroscience letters, 2002. **319**(2): p. 111-115.
18. Wible, B., T. Nicol, and N. Kraus, *Atypical brainstem representation of onset and formant structure of speech sounds in children with language-based learning problems.* Biological psychology, 2004. **67**(3): p. 299-317.
19. Russo, N.M., et al., *Auditory training improves neural timing in the human brainstem.* Behavioural brain research, 2005. **156**(1): p. 95-103.
20. Johnson, K.L., et al., *Auditory brainstem correlates of perceptual timing deficits.* Journal of Cognitive Neuroscience, 2007. **19**(3): p. 376-385.

21. Banai, K., et al., *Brainstem timing: implications for cortical processing and literacy.* Journal of Neuroscience, 2005. **25**(43): p. 9850-9857.
22. Banai, K., et al., *Reading and subcortical auditory function.* Cerebral cortex, 2009. **19**(11): p. 2699-2707.
23. Burns, K., A. Soll, and K. Vander Werff. *Brainstem responses to speech in younger and older adults*. in *American Auditory Society Annual Meeting Scottsdale, AZ*. 2009.
24. Chandrasekaran, B., et al., *Context-dependent encoding in the human auditory brainstem relates to hearing speech in noise: implications for developmental dyslexia.* Neuron, 2009. **64**(3): p. 311-319.
25. Asgharzadeh-Alavr, A., et al., *A Study of the Effect of Two Meaningful Syllables Stimuli in Auditory Brainstem Responses Using Correlation and Coherence Analyses.* Frontiers in Biomedical Technologies, 2015. **2**(2): p. 80-86.
26. Prévost, F., et al., *Objective measurement of physiological signal-to-noise gain in the brainstem response to a synthetic vowel.* Clinical Neurophysiology, 2013. **124**(1): p. 52-60.
27. Laroche, M., et al., *Brainstem auditory responses to resolved and unresolved harmonics of a synthetic vowel in quiet and noise.* Ear and hearing, 2013. **34**(1): p. 63-74.
28. Sadeghian, A., H.R. Dajani, and A.D.C. Chan. *Classification of English vowels using speech evoked potentials*. in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 2011.
29. Sadeghian, A., H.R. Dajani, and A.D. Chan, *Classification of speech-evoked brainstem responses to English vowels.* Speech Communication, 2015. **68**: p. 69-84.
30. Engineer, C.T., et al., *Cortical activity patterns predict speech discrimination ability.* Nature neuroscience, 2008. **11**(5): p. 603.
31. Centanni, T., et al., *Detection and identification of speech sounds using cortical activity patterns.* Neuroscience, 2014. **258**: p. 292-306.
32. Pasley, B.N., et al., *Reconstructing speech from human auditory cortex.* PLoS Biol, 2012. **10**(1): p. e1001251.
33. Brumberg, J.S., et al., *Brain–computer interfaces for speech communication.* Speech communication, 2010. **52**(4): p. 367-379.
34. Denby, B., et al., *Silent speech interfaces.* Speech Communication, 2010. **52**(4): p. 270-287.
35. Jafarpisheh, A.S., et al., *Nonlinear feature extraction for objective classification of complex auditory brainstem responses to diotic perceptually critical consonant-vowel syllables.* Auris Nasus Larynx, 2016. **43**(1): p. 37-44.
36. Skoe, E., T. Nicol, and N. Kraus, *Cross-phaseogram: objective neural index of speech sound differentiation.* Journal of neuroscience methods, 2011. **196**(2): p. 308-317.
37. Shirzhiyan, Z., et al. *Enhancement of complex auditory brainstem response to a voiced stop consonant-vowel syllable, by using LMS-based Adaptive Filter*. in *Biomedical Engineering and 2016 1st International Iranian Conference on Biomedical Engineering (ICBME), 2016 23rd Iranian Conference on*. 2016. IEEE.
38. Mallat, S., *A wavelet tour of signal processing*. 1999: Elsevier.
39. Donoho, D.L. and I.M. Johnstone. *Threshold selection for wavelet shrinkage of noisy data*. in *Engineering in Medicine and Biology Society, 1994. Engineering Advances: New Opportunities for Biomedical Engineers. Proceedings of the 16th Annual International Conference of the IEEE*. 1994. IEEE.
40. Cha, S.-H., *Comprehensive survey on distance/similarity measures between probability density functions.* City, 2007. **1**(2): p. 1.
41. Akay, M., *Wavelet applications in medicine.* IEEE spectrum, 1997. **34**(5): p. 50-56.

42. Saito, N. and R.R. Coifman. *Local discriminant bases*. in *SPIE's 1994 International Symposium on Optics, Imaging, and Instrumentation*. 1994. International Society for Optics and Photonics.
43. Couderc, J., et al. *Detection of abnormal time-frequency components of the QT interval using a wavelet transformation technique*. in *Computers in Cardiology 1997*. 1997. IEEE.
44. Saito, N. and R.R. Coifman. *Local discriminant bases*. in *Wavelet Applications in Signal and Image Processing II*. 1994. International Society for Optics and Photonics.
45. Saito, N. and R.R. Coifman, *Local discriminant bases and their applications.* Journal of Mathematical Imaging and Vision, 1995. **5**(4): p. 337-358.
46. Räsänen, O. and J. Pohjalainen. *Random subset feature selection in automatic recognition of developmental disorders, affective states, and level of conflict from speech*. in *INTERSPEECH*. 2013.
47. Fisher, R.A., *The use of multiple measurements in taxonomic problems.* Annals of human genetics, 1936. **7**(2): p. 179-188.
48. Harris, B., *The Estimation of Probabilities: An Essay on Modern Bayesian Methods (IJ Good).* SIAM Review, 1966. **8**(1): p. 118.
49. Polikar, R. and M. Akay, *Pattern recognition, in wiley encyclopedia of biomedical engineering.* Ed. Akay, M., New York, NY: Wiley, 2006.
50. Hsu, C.-W. and C.-J. Lin, *A comparison of methods for multiclass support vector machines.* IEEE transactions on Neural Networks, 2002. **13**(2): p. 415-425.
51. Cristianini, N. and J. Shawe-Taylor, *An introduction to support vector machines and other kernel-based learning methods*. 2000: Cambridge university press.
52. Platt, J.C., N. Cristianini, and J. Shawe-Taylor. *Large margin DAGs for multiclass classification*. in *Advances in neural information processing systems*. 2000.
53. Ramanan, A., S. Suppharangsan, and M. Niranjan. *Unbalanced decision trees for multi-class classification*. in *Industrial and Information Systems, 2007. ICIIS 2007. International Conference on*. 2007. IEEE.
54. Kraus, N., et al., *Auditory neurophysiologic responses and discrimination deficits in children with learning problems.* Science, 1996. **273**(5277): p. 971-973.
55. Bidelman, G.M. and A. Krishnan, *Neural correlates of consonance, dissonance, and the hierarchy of musical pitch in the human brainstem.* Journal of Neuroscience, 2009. **29**(42): p. 13165-13171.
56. Musacchia, G., et al., *Musicians have enhanced subcortical auditory and audiovisual processing of speech and music.* Proceedings of the National Academy of Sciences, 2007. **104**(40): p. 15894-15898.
57. Gavrishchaka, V.V. and S.B. Ganguli, *Support vector machine as an efficient tool for high-dimensional data processing: Application to substorm forecasting.* Journal of Geophysical Research: Space Physics, 2001. **106**(A12): p. 29911-29914.
58. Ghaddar, B. and J. Naoum-Sawaya, *High dimensional data classification and feature selection using support vector machines.* European Journal of Operational Research, 2018. **265**(3): p. 993-1004.