TOWARD A BRAIN INTERFACE FOR TRACKING ATTENDED AUDITORY SOURCES

Marzieh Haghighi[1], *Mohammad Moghadamfalahi*[1], *Hooman Nezamfar*[1],
Murat Akcakaya[2], *Deniz Erdogmus*[1]

[1] Northeastern University
Boston, MA 02115 USA
{haghighi, erdogmus}@ ece.neu.edu

[2] University of Pittsburgh
Pittsburgh, PA 15261 USA
akcakaya@pitt.edu

ABSTRACT

Auditory-evoked noninvasive electroencephalography (EEG) based brain-computer interfaces (BCIs) could be useful for improved hearing aids in the future. This manuscript investigates the role of frequency and spatial features of audio signal in EEG activities in an auditory BCI system with the purpose of detecting the attended auditory source in a cocktail party setting. A cross correlation based feature between EEG and speech envelope is shown to be useful to discriminate attention in the case of two different speakers. Results indicate that, on average, for speaker and direction (of arrival) of audio signals classification, the presented approach yields 91% and 86% accuracy, respectively.

Index Terms— Auditory BCI, auditory attention

1. INTRODUCTION

Brain-computer interfaces (BCIs) are proving to be feasible communication channel for people with severe physical disabilities, such as amyotrophic lateral sclerosis (ALS) or spinal cord injury. Auditory BCIs have been successful in this domain recently. Another emerging application area for auditory BCIs is attended speaker identification; in this paper, using electroencephalography (EEG), we show successful classification of spatial and frequency features of attended acoustic source in a cocktail party setting. EEG has been extensively used in BCI designs due to its high temporal resolution, noninvasiveness, and portability.

Auditory-evoked P300 BCI spelling system for locked-in patients is widely studied [1], [2], [3], [4], [5], [6]. Fundamental frequency, amplitude, pitch and direction of audio stimuli are distinctive features, which can be processed and distinguished by brain. Also, recent studies have shown cortical entrainment to the temporal envelope of speech using EEG measurements [7], [8], [9]. A study on the quality of cortical entrainment to audio stimulus envelope by topdown cognitive attention has shown enhancement of obligatory auditory processing activity in top-down attention responses when competing auditory stimuli differ in space direction [10] and frequency [11]. An early effect of coherence -115 to 185 ms - during passive listening and larger and longer effect of coherence - up to 265 ms - during passive listening of stochastic figureground (SFG) stimulus developed by [12] have been illustrated [13]. Recently, works with successful classification of attended versus unattended speaker using 60 second [14] and 20 seconds [15] of data has been published. Even though these results are still far from requirements to be incorporated in an online setting for a hearing aid application, they are motivating for further investigation on this area.

In this paper, we investigate the role of frequency and spatial features of audio sources in selective auditory attention activation of the brain. Diotic (binaural) stimulus presentation with different story tellers is an attempt to examine frequency based attention sensory processing. Dichotic (stimulation of each ear using different sounds simultaneously) stimulus presentation tries to explore the effect of spatial direction of stimulus on attention processing of the brain.

2. DATA COLLECTION

2.1. EEG Neurophysiological Data

Four individuals (2 male, 2 female) between ages 25 to 30 years old with no history of serious hearing impairment or neurological problems participated in this study, which followed an approved protocol. EEG signals were recorded using a g.USBamp biosignal amplifier using active g.Butterfly electrodes with cap application from g.Tec (Graz, Austria) at 256 Hz. Sixteen EEG channels (F3, F4, T7, T8, C3, C4, CZ, CPZ, PZ, P1, P2, P3, P4, PZ, O1, O2 and POZ according to International 10/20 system) were selected to capture auditory related brain activities over the scalp. Signals were filtered by g.Tecs built-in analog bandpass ($[0.5, 60]$Hz) and

This work is supported by NSF IIS-1149570, IIS-1118061, CNS-1136027, SMA-0835976; by NIH 5R01DC009834; and by NIDR-H133E140026. A paper package containing code and data can be found here: https://repository.library.northeastern.edu/collections/neu:rx914948d

978-1-5090-0746-2/16/$31.00 ©2016 IEEE

notch (60Hz) filters.

2.2. Experimental Design

Participants were asked to passively listen to four speech stimuli sessions through earphones. Each session contained 20 different 60 second trials of two competing speakers with 4 sec rest between each two consecutive trials. Speech stimuli were selected from audio books of well known novels from the literature. One male and one female speaker narrated their stories simultaneously in each trial for all sessions. As summarized in figure 1, speech stimuli were presented diotically in the first two sessions such that both speakers narrated their stories simulataneously to both ears. In the last two sessions, speech stimuli were presented dichotically such that different speakers narrated their stories to different ears. In the first and third sessions participants were asked to attend to male voice whereas in the second and fourth sessions subjects were asked to attend the female voice. In the dichotic sessions target stimulus was randomly played in one of the ears but subjects were asked to focus on target (male/female) voice independent of its direction and direction of target was also shown on the screen using the following direction symbols "$>$","$<$" to reduce confusion. Amplitude of speech stimulus signal in each trial scaled to have equal energy for target and distractors.

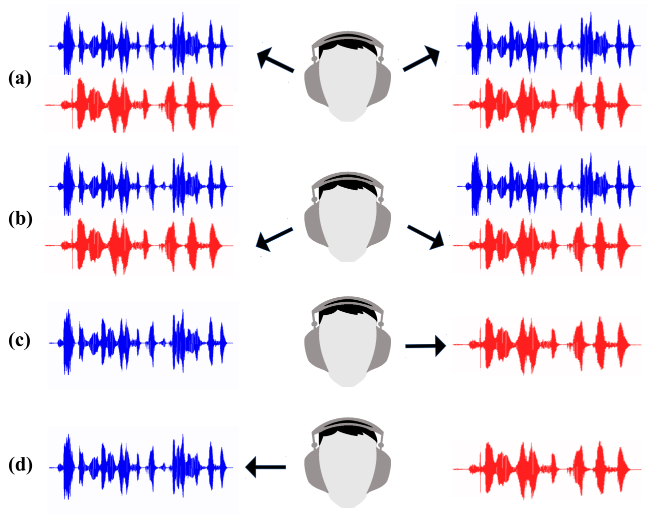

Fig. 1. Experimental paradigm visualization. (a) Diotic audio presentation and male is target. (b) Diotic audio presentation and female is target. (c) Dichotic audio presentation and male is target. (d) Dichotic audio presentaion and female is target.

3. METHODS

3.1. Preprocessing

Acquired EEG signals were digitally filtered by a linear-phase bandpass filter ($[1.5, 42]$Hz). For each trial, τ sec of EEG signal following each stimulus and time locked to the onset of each stimulus was extracted.

The acoustic envelope of speech stimulus signals were calculated using the Hilbert transform and filtered by a low pass filter (with 20Hz cut-off frequency). Then, τ seconds of acoustic envelope signals following every stimulus and time locked to the stimulus onset were extracted.

Optimizing τ to get maximum classification accuracy with minimum time window is an important factor in the design of online auditory BCI systems. In this paper, we performed grid search to coarsely optimize τ as reported in Section 4.

3.2. Feature Extraction

Since top down attention differentially modulates envelope tracking neural activity at different time lags [7], [8], [9], using different time lag values, $\boldsymbol{t}_0$, we calculate the cross correlation (CC) between the extracted EEG signals and target and distractor acoustic envelopes. Every $\boldsymbol{t}_0 = [t_{0_1}, \cdots, t_{0_i}, \cdots, t_{0_N}]^T$ is a vector of time lag values. In our analysis we choose $t_{0_i} \in [t_1, t_2]$ seconds. We also investigate the effect of choosing different time windows as $\boldsymbol{t}_0$ values on the BCI performance in Section 4. For each channel, we calculate the cross correlations between the EEG and the male and female speakers' acoustic envelopes for the time lag values defined in $\boldsymbol{t}_0$. Assuming that $\boldsymbol{t}_0$ is an $N \times 1$ vector, we concatenate the cross correlation values from male and female speakers into a single vector and hence each feature vector is $2N \times 1$ dimensional.

3.3. Classification of Speaker/direction

As explained in Section 2, the participants are asked to direct their auditory attention to a target speaker during data collection. The other speaker is the distractor. The labeled data collected in this manner is used in the analysis of discrimination between two speakers or discrimination between directions in a binary auditory attention classification problem. Using the $2N \times 1$ dimensional cross-correlation values as the feature vectors, we use Regularized Discriminant Analysis (RDA) as the classifier in our analysis. RDA is a modification of Quadratic Discriminant Analysis (QDA). QDA assumes that data is generated by two Gaussian distributions with unknown mean and covariances and requires the estimation of these means and covariances of the target and non-target classes before the calculation of the likelihood ratio. However, since, N, the length of $\boldsymbol{t}_0$, as defined in Section 3.2, is usually high and the calibration sessions are short, the covariance estimates are rank deficient.

RDA eliminates the singularity of covariance matrices by introducing shrinkage and regularization steps. Assume each $\mathbf{x}_i$ is a $2N \times 1$-dimensional feature vector and y_i is its binary label showing if the feature belongs to speaker 1 or 2, that is $y_i \in \{1, 2\}$. Then the maximum likelihood estimates of the class conditional mean and the covariance matrices are computed as follows:

$$\begin{aligned} \boldsymbol{\mu}_k &= \frac{1}{N_k} \sum_{i=1}^{N} \mathbf{x}_i \delta(y_i, k) \\ \boldsymbol{\Sigma}_k &= \frac{1}{N_k} \sum_{i=1}^{N} (\mathbf{x}_i - \boldsymbol{\mu}_k)(\mathbf{x}_i - \boldsymbol{\mu}_k)^T \delta(y_i, k) \end{aligned} \tag{1}$$

where $\delta(\cdot, \cdot)$ is the Kronecker-δ function, k represents a possible class label (here $k \in \{1, 2\}$, and N_k is the number of realizations in class k. Accordingly, the shrinkage and regularization of RDA is applied respectively as follows:

$$\begin{aligned} \widehat{\boldsymbol{\Sigma}}_k(\lambda) &= \frac{(1-\lambda) N_k \boldsymbol{\Sigma}_k + (\lambda) \sum_{k=0}^{1} N_k \boldsymbol{\Sigma}_k}{(1-\lambda) N_k + (\lambda) \sum_{k=0}^{1} N_k} \\ \widehat{\boldsymbol{\Sigma}}_k(\lambda, \gamma) &= (1-\gamma) \widehat{\boldsymbol{\Sigma}}_k(\lambda) + (\gamma) \frac{1}{p} tr[\widehat{\boldsymbol{\Sigma}}_k(\lambda)] \mathbf{I}_{2N} \end{aligned} \tag{2}$$

Here, $\lambda, \gamma \in [0, 1]$ are the shrinkage and regularization parameters, $tr[\cdot]$ is the trace operator and $\mathbf{I}_{2N}$ is an identity matrix of size $2N \times 2N$. In our system we optimize the values of λ and γ to obtain the maximum area under the receiver operating characteristics (ROC) curve (AUC) in a 8-fold cross validation framework. Finally, the RDA score for a trial with the observation vector $\mathbf{x}_i$, which is defined as:

$$\delta = \log \left(\frac{f_{\mathcal{N}}(\mathbf{x}_i; \boldsymbol{\mu}_1, \widehat{\boldsymbol{\Sigma}}_1(\lambda, \gamma))}{f_{\mathcal{N}}(\mathbf{x}_i; \boldsymbol{\mu}_0, \widehat{\boldsymbol{\Sigma}}_0(\lambda, \gamma))} \right) \tag{3}$$

where $f_{\mathcal{N}}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the Gaussian probability density function with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. Here δ values are used to plot the ROC curves and to compute the AUC values.

4. ANALYSIS AND RESULTS

In the first two analysis below, we chose $\tau = 58$ sec.

Target / distractor correlation with EEG signal: In Figures 2 & 3, for one representative participant, we illustrate the cross correlation values averaged over trials and channels for diotic and dichotic presentations, respectively. This pattern is consistent accross different participants.

For diotic presentation, the range with the highest absolute correlation can be identified in the range $[50, 350]$ms ($\boldsymbol{t}_0$ is extracted from this range). In this range, we observe a negative correlation for both target and distractor speakers followed by an early positive correlation for target stimulus and delayed and suppressed version of that positive correlation for the distractor stimulus. Table 1 reports the temporal latency and the magnitude of the peak in cross correlation responses for every participant.

Correlation Features	**Positive Peak Magnitude**		**Time Lag of Peak (ms)**	
Stimulus	Target	Distractor	Target	Distractor
Participant 1	0.02	0.015	180	215
Participant 2	0.015	0.011	183	207
Participant 3	0.008	0.005	367	191
Participant 4	0.014	0.006	164	320

Table 1. Time latency and magnitude of peak in cross correlation responses for each participant

For dichotic presentation, a more complicated pattern emerges. A pattern similar to the diotic case is observed for the correlation of brain responses and target stimuli in both ears. However, the response to distractor stimulus behaves differently in right and left ear in general. For two participants, in the right ear a delayed and suppressed, and in the left ear earlier and suppressed version of the target cross correlation is observed (see figure 3).

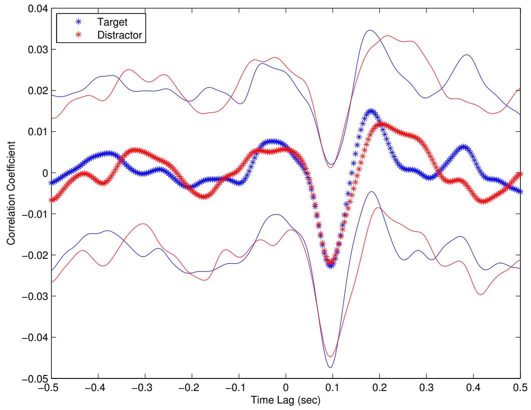

Fig. 2. Correlation coefficients of target/distractor with envelope of speech signal at different time lags which is averaged across trials and channels. Surrounding lines show one standard deviation above and below the mean.(Diotic stimulus presentation)

Single channel classification analysis: Following the plots in Figures 2 and 3, we chose the window $[50, 350]$ms as the most informative window for classification of target versus distractor responses ($\boldsymbol{t}_0$ vector is formed). We applied the classifier described in Section 3.3 on the extracted features for each channel independently to localize the selective attention responses. Figure 4 shows topographical maps of classification performance (AUC) for both diotic and dichotic auditory presentations over the scalp, for all participants. For the diotic presentation, the maximum AUC for a single channel classification, are as follows: 100% for participant 1, 97% for

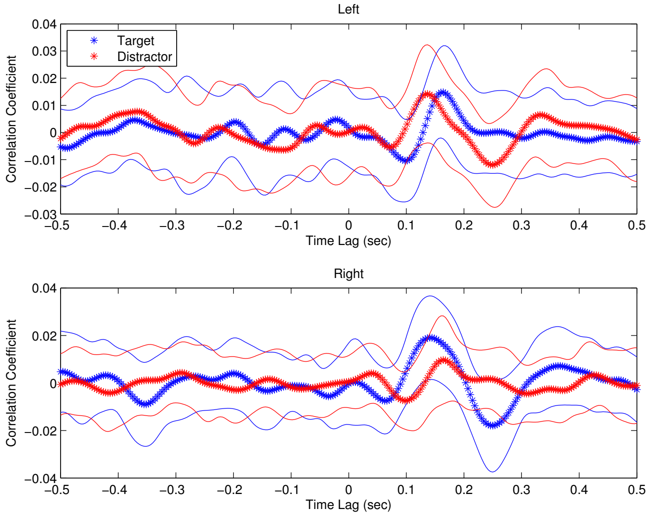

Fig. 3. Correlation coefficients of target/distractor with envelope of speech signal at different time lags which is averaged across trials and channels. Dashed thin lines show one standard deviation above and below the mean.(Dichotic stimulus presentation)

participant 2, 88% for participant 3, 96% for participant 4. Similarly, the maximum classification accuracies in a single channel classification scheme for dichotic auditory presentation are: 91% accuracy for participant 1, 90% for participant 2, 93% for participant 3, 71% for participant 4 for dichotic stimulus presentation.

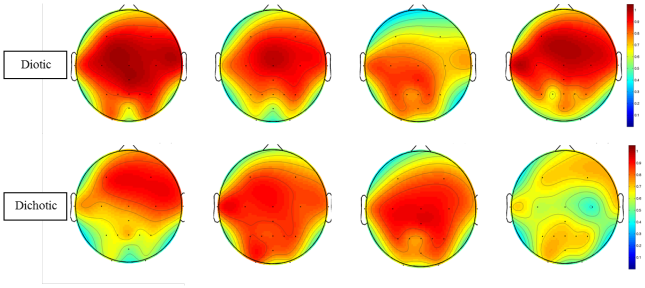

Fig. 4. Topographic map of classification performance over the scalp for classifying attended versus unattended speakers.

For direction classification, we applied the RDA classifier on samples from only male (or female) speaker and classified the direction of the incoming voice. Topographical maps of classification preformance results are summarized in Figure 5. Considering the channels which provide the best classification accuracy, we observed 80% for participant 1, 85% for participant 2, 90% for participant 3, 100% for participant 4 for male speaker and 80% accuracy for participant 1, 75% for participant 2, 85% for participant 3, 90% for participant 4 for female speaker.

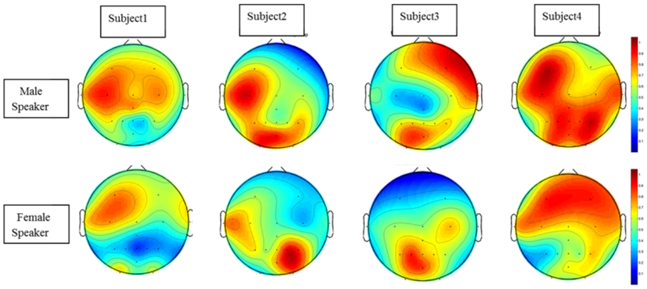

Fig. 5. Topographic map of classification performance of target speaker direction.

Evaluation of trial length for classification performance: The goal of this analysis is to choose the best τ, the length of EEG data to be extracted for use in cross correlation. We report the classification accuracies for speaker sound frequency (male/female) discrimination in diotic and speaker sound direction in dichotic sessions in Figures 6 and 7, respectively. In these figures, there is one curve for each participant and these plots show the classification performance at channel Cz as a function of τ. In this analysis, we proposed to use Cz since on average it was producing the best performance among users. To generate these figures we increased τ by two seconds at every step. In these two figures, we observe an average incremental pattern in the classification accuracies for both dichotic and diotic sessions as τ is increasing. The inconsistency for longer durations might be due to silent periods or user attention drifts.

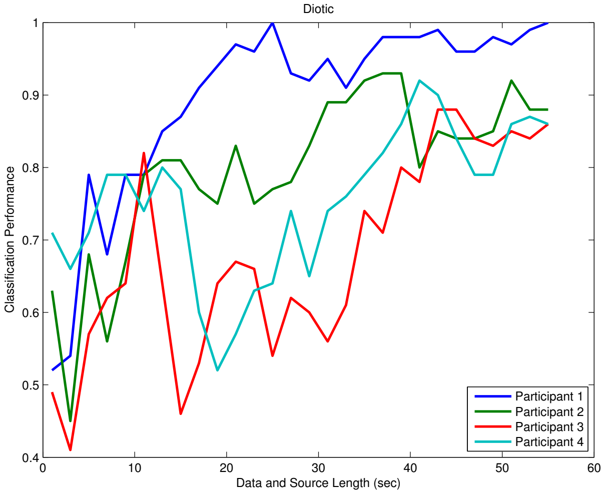

Fig. 6. Speaker frequency based classification performance using different τ values for diotic stimulus presentation at channel Cz.

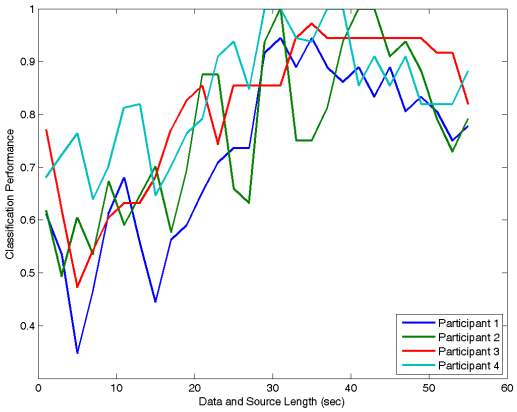

Fig. 7. Speaker direction based classification performance using different τ values for dichotic stimulus presentation at channel Cz.

5. CONCLUSION

In this paper, we presented results from a preliminary attempt to investigate the feasibility of online classification of auditory attention using a noninvasive EEG-based brain interface. Analyzing experimental data from four participants offline to evaluate the effect of spatial and frequency characteristic of audio stimuli, we identified informative data length and cross-correlation time lags for feature extraction.

Classification accuracies for attended speaker, obtained using regularized discriminant analysis of extracted EEG features, ranged around 95% and 86% for speaker discrimination in diotic and dichotic cases, respectively. For the direction identification, in average we observed 89% and 82% accuracies for identification of male and female voice directions, respectively. In our future work, using data from a larger group of participants, we will pursue a real-time implementation of a brain interface that can track attended speaker and direction.

6. REFERENCES

[1] M. Schreuder, T. Rost, and M. Tangermann, “Listen, you are writing! speeding up online spelling with a dynamic auditory bci,” *Frontiers in neuroscience*, vol. 5, 2011.

[2] J. Höhne, M. Schreuder, B. Blankertz, and M. Tangermann, “A novel 9-class auditory erp paradigm driving a predictive text entry system,” *Frontiers in neuroscience*, vol. 5, 2011.

[3] A. Kübler, A. Furdea, S. Halder, E. M. Hammer, F. Nijboer, and B. Kotchoubey, “A brain–computer interface controlled auditory event-related potential (p300) spelling system for locked-in patients,” *Annals of the New York Academy of Sciences*, vol. 1157, no. 1, pp. 90–100, 2009.

[4] S. Halder, M. Rea, R. Andreoni, F. Nijboer, E. Hammer, S. Kleih, N. Birbaumer, and A. Kübler, “An auditory oddball brain–computer interface for binary choices,” *Clinical Neurophysiology*, vol. 121, no. 4, pp. 516–523, 2010.

[5] A. Furdea, S. Halder, D. Krusienski, D. Bross, F. Nijboer, N. Birbaumer, and A. Kübler, “An auditory oddball (p300) spelling system for brain-computer interfaces,” *Psychophysiology*, vol. 46, no. 3, pp. 617–625, 2009.

[6] S. Kanoh, K.-i. Miyamoto, and T. Yoshinobu, “A brain-computer interface (bci) system based on auditory stream segregation,” in *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*. IEEE, 2008, pp. 642–645.

[7] N. Ding and J. Z. Simon, “Cortical entrainment to continuous speech: functional roles and interpretations,” *Frontiers in human neuroscience*, vol. 8, 2014.

[8] S. J. Aiken and T. W. Picton, “Human cortical responses to the speech envelope,” *Ear and hearing*, vol. 29, no. 2, pp. 139–157, 2008.

[9] Y.-Y. Kong, A. Mullangi, and N. Ding, “Differential modulation of auditory responses to attended and unattended speech in different listening conditions,” *Hearing research*, vol. 316, pp. 73–81, 2014.

[10] A. J. Power, E. C. Lalor, and R. B. Reilly, “Endogenous auditory spatial attention modulates obligatory sensory activity in auditory cortex,” *Cerebral Cortex*, vol. 21, no. 6, pp. 1223–1230, 2011.

[11] C. M. Sheedy, A. J. Power, R. B. Reilly, M. J. Crosse, G. M. Loughnane, and E. C. Lalor, “Endogenous auditory frequency-based attention modulates electroencephalogram-based measures of obligatory sensory activity in humans,” *NeuroReport*, vol. 25, no. 4, pp. 219–225, 2014.

[12] D. M. Groppe, T. P. Urbach, and M. Kutas, “Mass univariate analysis of event-related brain potentials/fields i: A critical tutorial review,” *Psychophysiology*, vol. 48, no. 12, pp. 1711–1725, 2011.

[13] J. A. O’Sullivan, S. A. Shamma, and E. C. Lalor, “Evidence for neural computations of temporal coherence in an auditory scene and their enhancement during active listening,” *The Journal of Neuroscience*, vol. 35, no. 18, pp. 7256–7263, 2015.

[14] J. A. O’Sullivan, A. J. Power, N. Mesgarani, S. Rajaram, J. J. Foxe, B. G. Shinn-Cunningham, M. Slaney, S. A. Shamma, and E. C. Lalor, “Attentional selection in a cocktail party environment can be decoded from single-trial eeg,” *Cerebral Cortex*, p. bht355, 2014.

[15] C. Horton, R. Srinivasan, and M. DZmura, “Envelope responses in single-trial eeg indicate attended speaker in a cocktail party,” *Journal of neural engineering*, vol. 11, no. 4, p. 046015, 2014.