Optimizing Stimulus Presentation for a Spatial Auditory P300 Speller

Nicholas Boldt **Shadi Sartipi**

University of Rochester, department of ECE
nboldt@u.rochester.edu, ssartipi@ur.rochester.edu

ABSTRACT

Recently, a brain-computer interface (BCI) using virtual sound sources has been proposed for estimating user intention via electroencephalogram (EEG) in an oddball task. However, its performance is still insufficient for practical use. In this work, we propose using multiple unique sound images coming from six directions in the form of spoken syllables. We aim to optimize the presentation sequence codebook to make a practical auditory P300 speller. To this aim, we have shown the differences of P300 brain response between playing from one direction versus two different directions using support vector machine classification. The results show that the brain generates stronger P300 response to the stimuli presentation from two different direction, allowing us to design more efficient auditory BCIs

1. INTRODUCTION

Brain-computer interfaces provide additional means for humans to interact with their environment. Signals recorded from the brain such as EEG provide meaningful information streams that can be classified into commands that the user intends to execute. This is especially useful for people who have lost their traditional means of communication - allowing paralyzed people to operate robotic arms [1], wheelchairs [2], or to spell out words [3].

There are several EEG brain activity patterns that can be utilized for a brain-computer interfaces including P300, SSVEP, or sensorimotor rhythms. The P300 event-related potential is the electrophysiological response to infrequent stimuli occurring during an oddball paradigm. The P300 waveform is characterized by a peak 300 ms after stimulus onset, and is on the scale of millivolts [4]. Hence, it is vital to use proper stimuli to elicit strong P300 responses.

For a P300-based BCI speller visual stimuli are most commonly used, where letters are arranged in a grid. Rows and columns are flashed while the user attends a single letter, and a binary classifier determines whether the flashed produced a P300 response. The probability of all letters are adjusted until they pass a confidence threshold and a letter is chosen [4].

However, brain-computer interface are designed for populations that include people with neurodegenerative disorders such as ALS or Parkinson's disease. For these people that have lost their ability to move their limbs, it is difficult to attend their gaze to different visual stimuli, and an alternative form of stimulus presentation is necessary [5]. One approach is to use audition, however these P300-spellers typically present individual stimuli sequentially which results in a slow, inefficient system.

There is recent work using auditory stimuli coming from different directions to produce P300 responses. Schreuder et al. 2010 first utilized spatialized audio in a multi-class brain-computer interface using eight speakers arranged around listeners, obtaining about 70% classification performance [6]. In 2013, Nambu et al. 2013 utilized virtual spatialized audio presented through binaural headphones to achieve similar performance for six virtual speakers, with single-trial performance of 70% to 89% when averaging over multiple trials [7]. Carabez et al. 2017 used convolutional neural networks to achieve 90% single-trial performance for spatialized audio from six directions [8].

There exists work that makes use of this stimulation technique. Sugi et al. tested the impact of using different stimulus onset asynchronies (SOA), which are the durations between presented sounds [9]. The reasoning was that while decreasing SOA speeds up the task, sound perception make become too difficult, with fewer and smaller P300 responses. They had their participants count the number of times a white noise stimulus came from a specific direction, where all directions used the same sound image. They found that SOAs should be chosen based on the user - most users could optimally perform around 400 ms, but some could perform optimally at even shorter SOAs around 200 ms.

We aim to further expand upon this research. When using multiple unique sound images coming from each direction in the form of spoken syllables, additional complications arise. One concern is the choice of stimulus presentation order. The P300 response is influenced by intertarget times (ITT), where more time between presentations of the attended stimuli enhance the P300 response. Conversely, playing the same stimulus too soon can have negative consequences on the P300. Hence our goal is to design and test different presentation codebooks that optimize the relationship between SOA and ITTs for an auditory P300 utilizing many unique auditory stimuli.

2. METHODS

2.1 Experimental Design

2.1.1 EEG Recording

In this study the data corresponds to the evoked p300 waves on an auditory BCI paradigm has been used. A digital electroencephalogram system (Active Two, BioSemi, Amsterdam, Netherlands [10]) at 2048 Hz was used for recording the data. The recording device consists of 64 electrodes distributed over the head of the subject based on standard 10-20 system. The distribution of the electrodes was shown in Fig. 1.

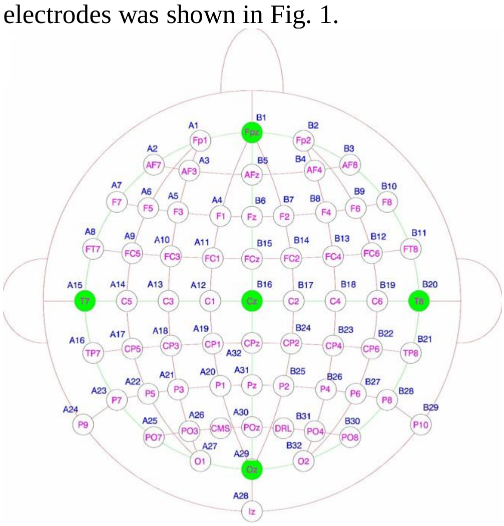

Figure 1: 64-channel electroencephalogram (EEG) layout used in the experiment.

2.1.2 Proposed Experiment Design

In this work, we propose a setup where we will test the subjects P300 brain responses in paying attention to the target stimuli among real source separated loudspeakers. The subjects are presented with stimuli from six direction semi-simultaneously. On every trial, listeners are cued by a 400 ms audiovisual cue to attend one of the six directions.

The stimuli consist of 26 English letters, which are split across the six external speakers. Thus, a participant knows to attend to a specific speaker. The set of possible stimuli at each speaker remains constant, but the order of the stimuli varies within-speaker. While the order across-speakers depends on our codebook strategy, all speakers will play a stimulus in sets until all stimuli are depleted (single trial). Timing is one factor we are investigating, where we present stimuli at different presentation rates, with stimulus onset asynchronies (duration between target presentations) ranging from 100 ms to 300 ms. Once a trial is complete, we repeat additional trials with the same attended stimulus until we reach 12 trials, forming a single block. The placement of the speakers is shown in Fig. 2, with the flow of the experiment shown in Fig. 3.

The designing of the experiment was done by using expyfun [16].

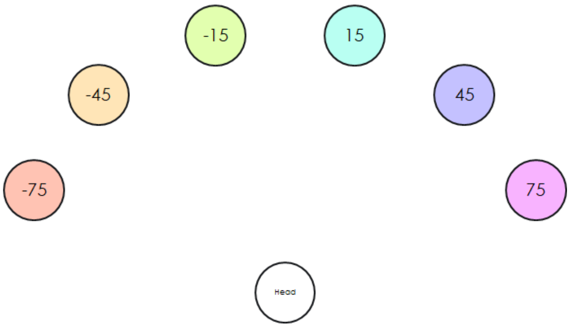

Figure 2: Real disposition of the six sound directions relative to the user.

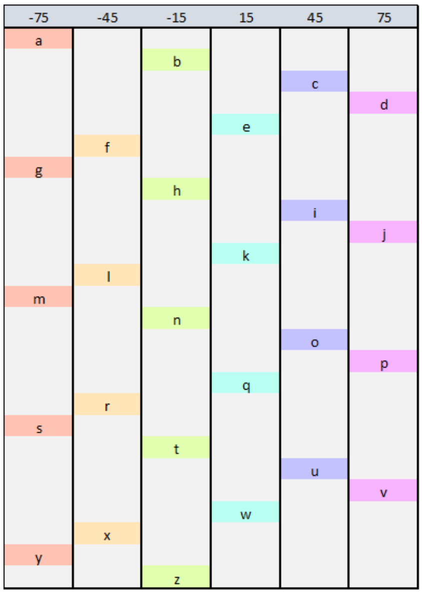

Figure 3: Presentation codebook (task constitution).

2.1.3 Tested Experimental Design

The subject participated in two different experiments. First, they were resented with stimuli from two direction simultaneously. On every trial, listeners were cued by a 500 ms audiovisual cue to attend one of the two directions. Second, the stimuli were presented at one of the two directions.

Each trial consisted of playing 26 English letters. For multiple speakers, the stimuli were evenly split, and a participant could attend to their stimulus at a predetermined direction (Fig 4.). The set of possible stimuli at each speaker remains constant, but the order of the stimuli varies within-speaker. Each trial was repeated 10 times for every letter.

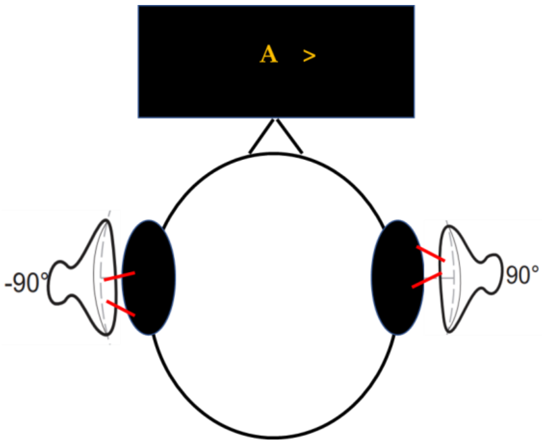

Figure 4: Flow for reduced experiment.

2.2 Data Preprocessing

EEG data preprocessing is conducted as follows: The EEG data were high-pass and low-pass filtered. The low-pass cutoff frequency was 0.1 Hz and the high-pass cutoff frequency was 15 Hz. The artifacts were removed by applying independent component analysis (ICA) using MNE package [11]. For each channel, we averaged the last 100 ms of the signal before stimulus onset as a baseline correction, and subtracted it from the measured data. Then we down-sampled the data to 128 Hz.

2.2 Feature Extraction and Feature Selection

2.2.1 Event related potentials (ERP)

ERP is the feature that commonly considered in analyzing the EEG data which are related to the specific event. Audio P300 response most of the time occurs in the time range between 200 ms and 350 ms after the stimuli onset. Thus, the ERPs in this time range considered as the features for the rest of the evaluations.

2.2.2 Wilcoxon Rank Sum Test

In statistics, the Wilcoxon test []is a nonparametric test of the null hypothesis that it is equally likely that a randomly selected value from one population will be less than or greater than a randomly selected value from a second population.

This test can be used to investigate whether two independent samples were selected from populations having the same distribution. A similar nonparametric test used on dependent samples is the Wilcoxon signed-rank test. With applying this test we were able to find which features from which channels could be the discriminant features between target and non-target groups.

2.3 Classification

Support vector machines so-called as SVM is a supervised learning algorithm that can be used for classification and regression problems as support vector classification (SVC) and support vector regression (SVR). Gaussian RBF(Radial Basis Function) is one of the popular Kernel methods used in SVM models for more. RBF kernel is a function whose value depends on the distance from the origin or from some point. Gaussian Kernel is of the following format:

$$K(X_1, X_2) = \exp(-\gamma \|X_1 - X_2\|^2)$$

Behavior: As the value of γ increases the model gets overfits. As the value of γ decreases the model underfits. Here we set the values to auto to select the best parameters for classification.

3. RESULTS

In this section the results for the two different experiment are presented. The set of channels relate to the selected features are "P5, Oz, POz, Cz, CP4, CP2, P2, P4, P6", and " P5, P7, O1, Pz, CPz, P4, P8" for one speaker and two speakers respectively. Figure 5 shows the ERPs for the experiment related to the one speaker and two speakers. The results show that the P300 response to the two speaker case is much more clear and could make discrimination between two groups around 250 ms.

Since the data is imbalance, the AUC is the good metric to show the effectiveness of the method. The receiver operating characteristic curve (ROC) for the two experiment considering the whole targets together is presented in Fig. 6. By considering each letter stimulus individually the ROC increase up to 0.8 for two speaker condition and 0.73 to one speaker condition.

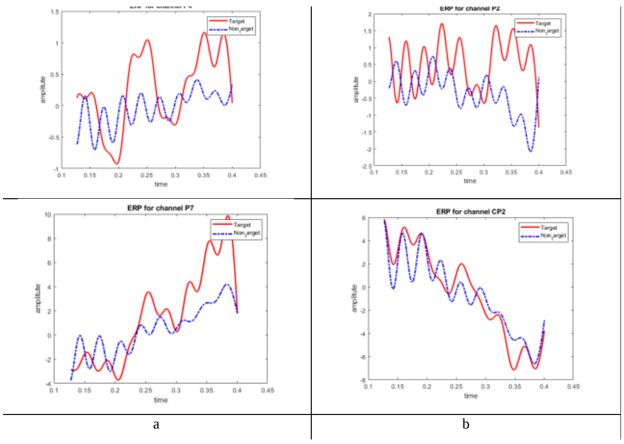

Figure 5: ERPs for (a) two speaker experiment , and (b) one speaker experiment

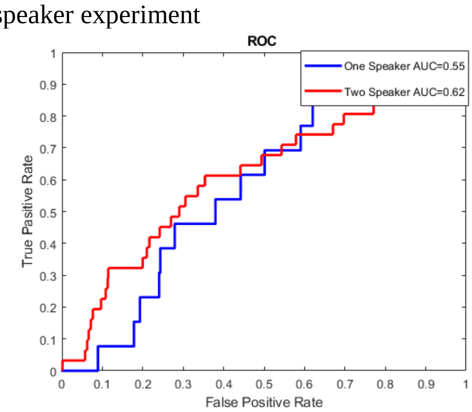

Figure 6: AUC Curves for One vs Two Speakers

4. CONCLUSIONS

Here we demonstrate the usefulness of using multiple speakers for an auditory P300 BCI setup. We compared the classification performance between one and two speakers with AUC analysis, which showed better performance with the latter condition.

In the near future we plan to test our six speaker approach, which we proposed in this paper. Furthermore, we would utilize deep-learning with CNN.

5. REFERENCES

[1] B. Edelman, J. Meng, D. Suma, C. Zurn, E. Nagarajan, B. Baxter, C. Cline, and B. He: “Noninvasive Neuroimaging Enhances Continuous Neural Tracking for Robotic Device Control.” *Science Robotics*, Vol. 4, No. 31, 2019.

[2] T. Bastos, S. Muller, A. Benevides, and M. Sarcinelli-Filho: “Robotic Wheelchair Commanded by SSVEP, Motor Imagery and Word Generation.” *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2011.

[3] A. Rezeika, M. Benda, P. Stawicki, F. Gembler, A. Saboor, and I. Volosyak: “Brain–Computer Interface Spellers: A Review.” *Brain Sciences*, Vol. 8, No. 4, 2018.

[4] R. Fazel-Rezai, B. Allison, C. Guger, E. Sellers, S. Kleih, and A. Kübler: “P300 Brain Computer Interface: Current Challenges and Emerging Trends.” *Frontiers in Neuroengineering* Vol. 5, 2012.

[5] L. Carelli, F. Solca, A. Faini, P. Meriggi, D. Sangalli, P. Cipresso, G. Riva, N. Ticozzi, A. Ciammola, V. Silani, and B. Poletti: “Brain-Computer Interface for Clinical Purposes: Cognitive Assessment and Rehabilitation.” *BioMed Research International 2017*, 2017.

[6] M. Schreuder, B. Blankertz, and M. Tangermann: “A New Auditory Multi-Class Brain-Computer Interface Paradigm: Spatial Hearing as an Informative Cue.” *PLoS ONE*, Vol. 5, No. 4, 2010.

[7] I. Nambu, M. Ebisawa, M. Kogure, S. Yano, H. Hokari, and Y. Wada: “Estimating the Intended Sound Direction of the User: Toward an Auditory Brain-Computer Interface Using Out-of-Head Sound Localization.” *PLoS ONE*, Vol. 8, No. 2, 2013.

[8] E. Carabez, M. Sugi, I. Nambu, and Y. Wada: “Identifying Single Trial Event-Related Potentials in an Earphone-Based Auditory Brain-Computer Interface.” *Applied Sciences*, Vol. 7, No. 11, 2017.

[9] M. Sugi, Y. Hagimoto, I. Nambu, A. Gonzalez, Y. Takei, S. Yano, H. Hokari, and Y. Wada: “Improving the Performance of an Auditory Brain-Computer Interface Using Virtual Sound Sources by Shortening Stimulus Onset Asynchrony.” *Frontiers in Neuroscience*, Vol. 12, 2018.

[10] Biosemi. https://www.biosemi.com/index.htm

[11] A. Gramfort. "MNE software for processing MEG and EEG data." *Neuroimage* 86 (2014): 446-460.

[12] B. Allison and J. Pineda: “Effects of SOA and Flash Pattern Manipulations on ERPs, Performance, and Preference: Implications for a BCI System.” *International Journal of Psychophysiology*, Vol. 59, No. 2, 2006.

[13] G. Ruxton: “The Unequal Variance t-Test Is an Underused Alternative to Students t-Test and the Mann–Whitney U Test.” *Behavioral Ecology* Vol. 17, No. 4, 2006.

[14] S. Vishwanathan., and M. Murty: “SSVM: a Simple SVM Algorithm.” *Proceedings of the 2002 International Joint Conference on Neural Networks*, 2002.

[15] Y. Kim: “Convolutional Neural Networks for Sentence Classification.” *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014.

[16] M. Cappelloni. "AUDIOVISUAL PARSING OF POLYPHONIC MUSIC."