Investigating Evoked EEG Responses to Targets Presented in Virtual Reality

Pawan Lapborisuth[1], Josef Faller[1,2], Jonathan Koss[1], Nicholas R. Waytowich[1,2], Jonathan Touryan[2], Paul Sajda[1]

Abstract— **Virtual reality (VR) offers the potential to study brain function in complex, ecologically realistic environments. However, the additional degrees of freedom make analysis more challenging, particularly with respect to evoked neural responses. In this paper we designed a target detection task in VR where we varied the visual angle of targets as subjects moved through a three dimensional maze. We investigated how the latency and shape of the classic P300 evoked response varied as a function of locking the electroencephalogram data to the target image onset, the target-saccade intersection, and the first fixation on the target. We found, as expected, a systematic shift in the timing of the evoked responses as a function of the type of response locking, as well as a difference in the shape of the waveforms. Interestingly, single-trial analysis showed that the peak discriminability of the evoked responses does not differ between image locked and saccade locked analysis, though it decreases significantly when fixation locked. These results suggest that there is a spread in the perception of visual information in VR environments across time and visual space. Our results point to the importance of considering how information may be perceived in naturalistic environments, specifically those that have more complexity and higher degrees of freedom than in traditional laboratory paradigms.**

I. Introduction

Virtual reality (VR) offers the potential to study brain function in complex, ecologically valid environments. Our understanding of brain function from highly controlled experimental setups that rely on simple abstractions of real-world objects or situations may not always directly translate to more natural behaviors and environments [1], [2]. Ecologically realistic VR environments may also prove essential for enabling real-world neurotechnology applications like brain-computer interfaces (BCIs).

One way of studying the brain is through event related potentials (ERPs), which are neuroelectrical responses that occur at a fixed time following an event. This work focuses specifically on the P300 evoked response which can be measured via the electroencephalogram (EEG) when human subjects perceive a "target" stimulus that is rare relative to other distracting stimuli [3]. Previous work has already shown that ERPs can be measured in a complex but non-VR target detection paradigm in which subjects are led through a maze environment [4]. However, this setup was neither immersive nor naturalistic with the subject's head being positioned on a chin rest in front of a monitor.

[1]Laboratory for Intelligent Imaging and Neural Computing, Department for Biomedical Engineering, Columbia University, New York City, NY, USA
Correspondence: pl2622@columbia.edu
[2]Human Research and Engineering Directorate, U.S. Army Research Laboratory, Aberdeen Proving Grounds, MD, USA.

Here, we focus on a similar target detection task in a more naturalistic VR environment using a head-mounted display (HMD). We look at two visual angles and the resulting changes in eye movement. We examine how different experimental conditions affect the latency and shape of the evoked response and the efficacy of different techniques for marking stimulus onset.

II. Materials and Methods

A. Subjects

Eleven neurologically healthy, unmedicated adults (9 males, 2 females, ages 20 to 40 years) participated in the study. One male subject was excluded from analysis since he showed no measurable P300 response. Written informed consent was obtained from all participants prior to the study and the experimental protocol was approved by the Institutional Review Board of Columbia University.

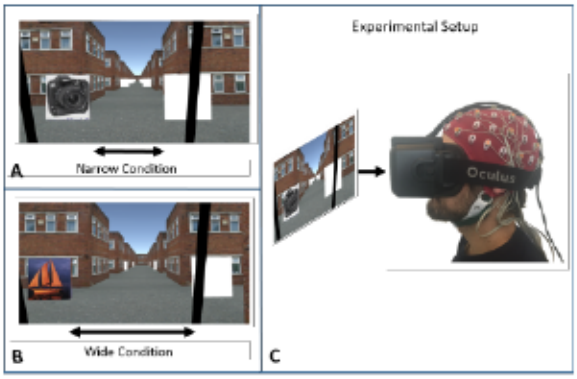

Fig. 1. Overview of the setup and the two experimental conditions. In the narrow condition (A), the billboards were placed closer to the middle of the alleyway whereas in the wide condition (B), the billboards were placed further apart. (C) A subject instrumented with an electroencephalography cap and head mounted display.

B. Experimental Paradigm

Subjects were passively driven through a simulated city-like VR environment. Initially, white billboards were placed on the left and right side of the street (see Fig. 1). The time to "drive" between billboards was 2.85 s. Whenever the subject was 0.75 s away from a set of billboards, an image randomly chosen from four categories — cameras, boats, laptops and grand pianos (50 images per category) would appear on one of the two billboards. These images were taken from a subset of images found in the CalTech101 database [5].

978-1-5386-1311-5/19/$31.00 ©2019 IEEE

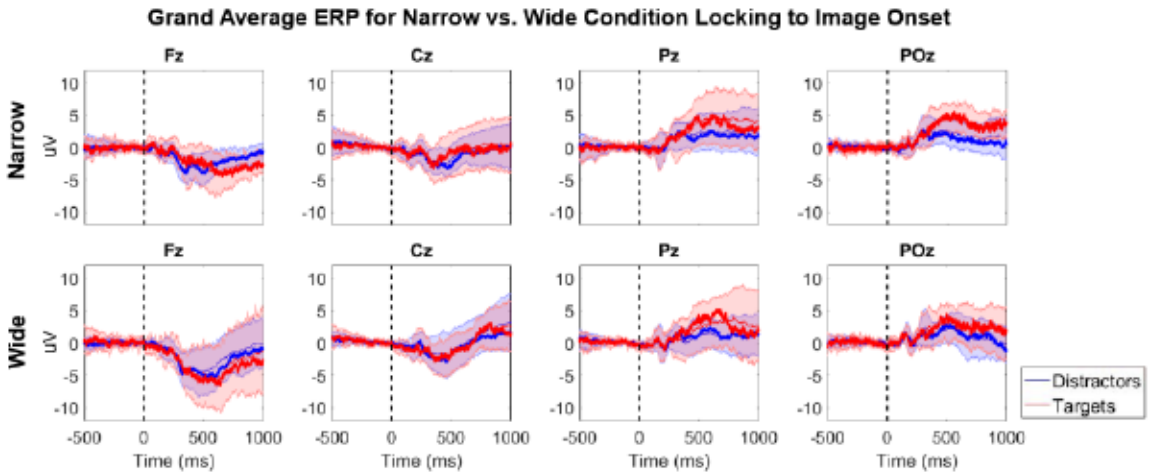

Fig. 2. Grand average event related potential (ERP) for the four mid-line electrodes Fz, Cz, Pz and POz, across 10 subjects for the narrow (top) and wide (bottom) conditions locking to image onset. Red indicates targets while blue indicates distractors. Thick lines represent the median ERP while thin lines represent the mean ERP across subjects. The shaded areas indicate standard deviation at each point in the time interval. Dotted vertical lines indicate time of image onset.

For each subject, one particular category was designated as the "target" stimulus. Subjects were instructed to count the number of targets that appeared every run and report to the experimenter. Forty images were presented to the subject during each run. Subjects performed two conditions of the task, "narrow" and "wide". In the narrow condition, the billboards were placed closer to the middle of the alley (see Fig. 1A). During this condition, subjects were instructed to keep their gaze fixed strictly between the two billboards throughout the run and to use their peripheral vision to observe the images presented on the billboards. In the wide condition, the billboards were placed further apart to the side of the alleyway (see Fig. 1B). Again, subjects were instructed to initially keep their gaze fixed in the middle between the two billboards. However, whenever an image appeared, the subjects were instructed to saccade to the image and then return their gaze back to the center after passing the billboard. Subjects performed a total of eight runs of each condition in four blocks of four runs per block in the following order—narrow, wide, narrow, wide. The paradigm was developed and implemented using NEDE, a Unity-based (Unity Technologies, San Francisco, CA, USA) scripting framework for virtual 3D environments [6]. The experiment was presented using a stereoscopic HMD (Oculus Rift DK2, Oculus VR, Menlo Park, CA, USA).

C. EEG-Eyetracking-VR Experimental Setup

EEG was measured from 64 Ag/AgCl scalp electrodes at a sampling rate of 2048 Hz (ActiveTwo, BioSemi B.V., Amsterdam, Netherlands). Gaze information was obtained using an eye tracker (SensoMotoric Instrument GmbH, Teltow, Germany) embedded within the HMD at a sampling rate of 60 Hz. All data streams were synchronized and recorded using the software framework Labstreaming Layer [7].

D. Time-Locking EEG Analysis

EEG data was band-pass filtered from 0.5 to 50 Hz. Independent component analysis (ICA) was used to remove eye blinks and horizontal eye movement artifacts. The gaze information from the eyetracker and the paradigm information from Unity were used to epoch the EEG data into three different categories - image onset (IO) locked, saccade intersection (SI) locked and first fixation (FF) locked. In this study, IO is defined as the time in which the image appeared on the billboard closest to the subject. SI is defined to be the first moment in which the subject's gaze intersected the bounding box of the image. Lastly, FF is the first moment in which the subjects fixated their gaze within the bounding box of the image. A pre-determined threshold for gaze velocity was used to determine the point of first fixation following the subjects' saccade to the billboard. Automatic channel and epoch rejections were performed using EEGLAB [8]. All epochs were baseline corrected from -200 ms to 0 ms. The evoked response was used to train a logistic regression classifier in order to determine the time and levels of peak discriminability between the targets and distractors for the different types of locking. Separate classifiers were trained for each 50 ms window of the evoked response. Leave-one-out cross validation and the area under the receiver operating characteristic curve (Az) were used to evaluate the performance of the trained classifier.

III. Results

A. Evoked Response

We found classic P300 evoked responses, in terms of shape and latency, for both the narrow and wide conditions when locking to IO (Fig. 2). This result suggests that our paradigm reliably generated P300 responses in both conditions (narrow and wide) which provides a good basis for investigating the effects of different time locking functions on P300 shape and

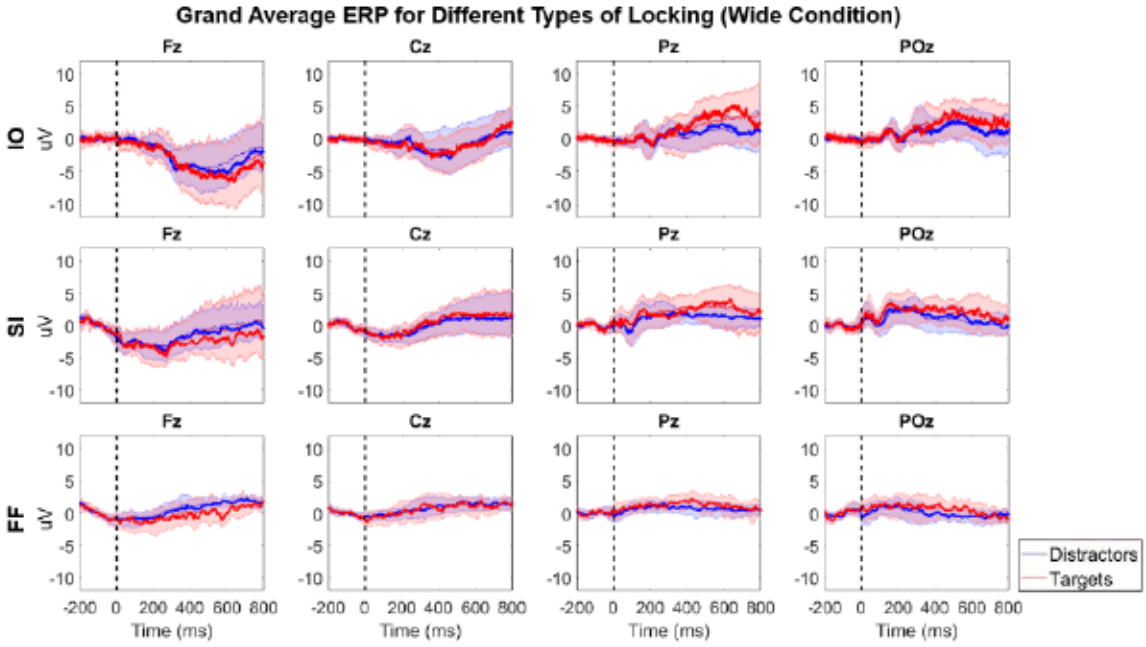

Fig. 3. Grand average event related potential (ERP) for the four mid-line electrodes Fz, Cz, Pz and POz, across 10 subjects for the wide condition locking to image onset (IO—top), saccade intersection (SI—middle) and first fixation (FF—bottom). Red indicates targets while blue indicates distractors. Thick lines represent the median ERP while thin lines represent the mean ERP across subjects. The shaded areas indicate standard deviation at each point in the time interval. Dotted vertical lines indicate time of locking.

latency. In the wide condition, we found classic P300 evoked responses both in shape and latency when locking to IO and locking to SI but not when locking to FF, where the peaks were flattened out and displayed less separability between targets and distractors (Fig. 3).

B. *Single-trial Analysis*

We performed a single-trial analysis to determine the quality of information in the EEG evoked response for different locking conditions in classifying the stimulus type (target vs distractor). Classification performance showed similar peak Az values for the narrow (M=0.70, SD=0.08) and wide (M=0.72, SD=0.06) conditions when locked to image onset (see Fig. 4 top). However, the mean time of peak Az relative to image onset was significantly longer for the wide condition (M=573 ms; SD=176 ms) compared to the narrow condition (M=450 ms; SD=112 ms; p=0.03; d=0.8; see Fig. 5 top). For the wide condition, we found a statistically significant effect of within-subject factor locking-type (three levels IO, SI and FF) on peak Az (repeated-measures ANOVA; $F(2,18)$=12.2, p=0.0005, η^2 = 0.58) (see Fig. 4 bottom). Post hoc t-tests where we adjusted alpha for three comparisons (Bonferroni, α_{corr} = 0.017) showed significantly lower peak Az when locking to FF (M=0.64; SD=0.04) relative to when locking to IO (M=0.72, SD=0.06; p=0.0002, d=1.6). No statistically significant difference was found when comparing peak Az for locking to IO relative to locking to SI (M=0.71; SD=0.07). We also found a statistically significant effect of within-subject factor locking-type (three levels IO, SI and FF) on peak Az time (repeated-measures ANOVA; $F(2,18)$=8.8, p=0.002, η^2 = 0.49; see Fig. 5 bottom). Post hoc t-tests where we adjusted alpha for three comparisons (Bonferroni, α_{corr} = 0.017) showed significantly lower peak Az time when locking to FF (M=263 ms, SD=189 ms) relative to when locking to IO (M=573 ms; SD=176 ms; p=0.0015, d=1.7). No statistically significant difference was found when comparing peak Az time for locking to IO relative to locking to SI (M=408 ms; SD=194 ms).

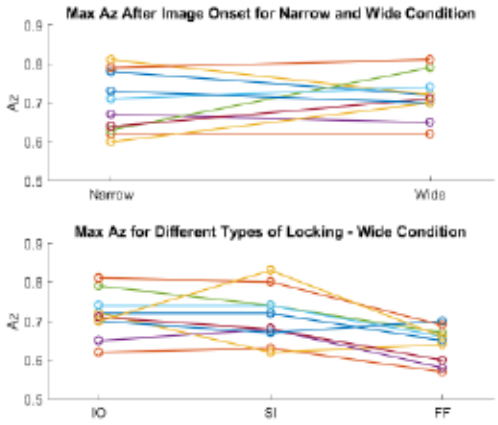

Fig. 4. Top panel displays maximum area under the receiver operating curve (Az) for ten subjects when locked to image onset in two conditions - narrow and wide. Bottom panel displays maximum Az for ten subjects in the wide condition for three different locking functions - image onset (IO), saccade intersection (SI) and first fixation (FF).

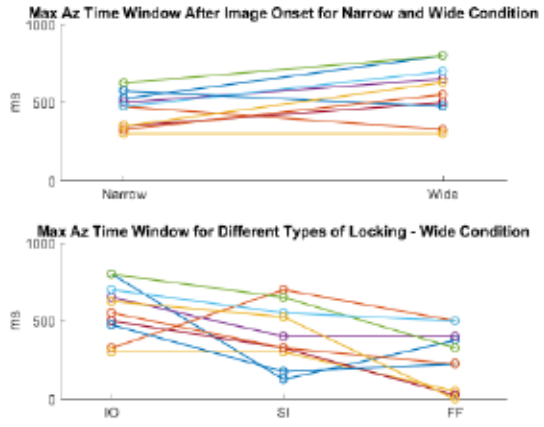

Fig. 5. Top panel displays time window following image onset which results in maximum area under the receiver operating curve (Az) for ten subjects in two conditions - narrow and wide. Bottom panel displays time window following three different locking functions - image onset (IO), saccade intersection (SI) and first fixation (FF) which results in maximum Az for ten subjects in the wide condition.

IV. Discussion

Our results indicate that the evoked neural response following the perception of visual information in VR is spread out temporally across different time locking functions. In the narrow and wide conditions, when locked to image onset, the shape of the P300 and the peak Az values remained similar to each other. This result suggests that the P300 is reliably generated in both conditions of our experimental paradigm. However, for the wide condition, the time from image onset to peak Az was significantly longer than for the narrow condition. This suggests that when the visual angle of the targets increases, the latency of the evoked response also increases. In order to further investigate when the evoked response is generated, we performed three different types of time locking for the wide condition. The ERP results showed qualitatively similar P300 shape and latency when locking to IO and SI but not when locking to FF where the responses was less sharp. Similarly, the single-trial analysis peak Az for locking to IO and locking to SI were similar whereas peak Az for locking to FF was significantly lower. Furthermore, the time of peak Az for locking to IO, SI and FF showed an overall decreasing trend, indicating the expected systematic shift in the timing of the evoked response as a function of the type of time locking. This suggests that the generation of the evoked P300 response takes place between the onset of the visual information and the end of the saccade prior to the fixation onto the targets.

Being able to identify when the evoked response is generated in VR can serve as useful information in developing BCIs for use in realistic environments. While the field of non-invasive BCIs has been growing rapidly, most systems are still limited to highly controlled experimental setups in which the subjects have to rely on information provided on a 2-D monitor/screen [9]. However, for BCIs to be applicable to real world situations, they must also function when subjects are exposed to realistic environments where the visual angle is much wider than that of a screen. The results of this study suggest that evoked responses can be decoded during saccade intersection which could be useful for developing BCIs for more complex environments. There are several noteworthy limitations to our current results and interpretations. One limitation is the design and presentation of the simulated driving game. A number of subjects reported feeling nauseated by the end of the experiment due to the movements of the environment in the VR headset. Another limitation is how we determined the point of first fixation during each trial, which was influenced to some degree by subjects' task-compliance and quality of the eye-tracking data in the constantly moving environment.

V. Conclusion

The results of this study advance our understanding of brain function in complex, ecologically valid environments. Our work highlights some of the challenges in studying the brain's responses to visual information in an environment that is no longer limited by a computer screen. Our finding that evoked responses occur between the appearance of the target and the visual saccade prior to the subject's fixation on the target object indicates that brain responses to visual stimuli in natural environments can be distributed across time. Future work is needed to identify the precise temporal and visual features involved in generating the observed evoked responses in such complex environments.

Acknowledgment

The work was funded by the Army Research Laboratory under Cooperative agreement number W911NF-10-2-0022, the Army Research Office (ARO) under grant number W911NF-11-1- 0219, and the National Science Foundation under grants IIS-1527747 and CHS-1816363. We would like to thank Kaylo Littlejohn for his contributions.

References

[1] C. J. Bohil, B. Alicea, and F. A. Biocca, "Virtual reality in neuroscience research and therapy," *Nat. Rev. Neurosci.*, vol. 12, no. 12, pp. 752–762, 2011.

[2] T. D. Parsons, "Virtual Reality for Enhanced Ecological Validity and Experimental Control in the Clinical, Affective and Social Neurosciences," *Front. Hum. Neurosci.*, vol. 9, 2015.

[3] J. Polich, "Updating P300: an integrative theory of P3a and P3b," *Clin. Neurophysiol.*, vol. 118, no. 10, pp. 2128-2148, 2007.

[4] D. C. Jangraw, J. Wang, B. J. Lance, S. F. Chang, and P. Sajda, "Neurally and ocularly informed graph-based models for searching 3D environments," *J. Neural Eng.*, vol. 12, no. 4, 2014.

[5] L. Fei-Fei, R. Fergus, and P. Perona, "Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories," *Comput. Vis. Image Underst.*, vol. 106, pp. 59–70, 2007.

[6] D. C. Jangraw, A. Johri, M. Gribetz, and P. Sajda, "NEDE: An open-source scripting suite for developing experiments in 3D virtual environments," *J. of Neurosci. Methods*, vol. 235, pp. 245–251, 2014.

[7] C. A. Kothe, "Labstreaming layer (LSL): A system for unified collection of measurement time series in research experiments," 2013.

[8] A. Delorme, and S. Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *J. Neurosci. Methods*, vol. 134, no. 1, pp. 9-21, 2004.

[9] R. Fezal-Rezai, "P300 brain computer interface: current challenges and emerging trends," *Front. Neuroeng.*, vol. 5, 2012.