DOI:10.3217/978-3-85125-260-6-10

Auditory BCIs for Visually Impaired Users: Should Developers Worry About the Quality of Text-to-Speech Readers?

K. R. Laghari[1,2], R. Gupta[1,2], S. Arndt[3], J.-N. Antons[3], S. Möller[3], R. Schleicher[3], D. O'Shaughnessy[1], T. H. Falk[1,2]

[1]INRS-EMT, University of Quebec, Canada; [2]Centre for Research on Brain, Language, and Music (CRBLM), Canada; [3]Quality and Usability Lab, Berlin Institute of Technology, Germany

Correspondence: K. R. Laghari, INRS-EMT, University of Quebec, Canada. E-mail: khalil.laghari@gmail.com

Abstract. Auditory-based brain-computer interfaces (BCI) have been gaining significant grounds recently, particularly for visually impaired users. The majority of existing auditory BCIs are based on P300 auditory event related potentials. Previous studies have shown, however, that P300s may be affected by the quality of the presented speech stimuli. Since visually impaired users commonly rely on text-to-speech (TTS) readers to e.g., navigate websites and read documents, this study investigated if the quality of the TTS system had an effect on observed P300 amplitudes. An experiment with 14 healthy subjects showed that indeed TTS synthesizer quality plays a significant effect on P300 amplitude. Hence, it is recommended that auditory BCI developers pay careful attention to the quality of the TTS system commonly utilized by the user, as low-quality systems may negatively affect BCI performance.

Keywords: EEG, text-to-speech, P300, auditory BCI, quality

1. Introduction

Auditory BCIs have emerged as an attractive means of communication for blind and vision impaired individuals, such as those with amyotrophic lateral sclerosis (ALS). Auditory BCIs commonly involve the use of an oddball paradigm to elicit a P300 event related potential. More specifically, the user is required to pay attention to a specific ”hidden” sound amongst a collection of sounds. The simplest way of eliciting a P300 signal is by presenting subjects with tones of two different pitch frequencies (e.g., high and low), thus resulting in a binary BCI. Recent studies have also shown that spoken words [Guo et al., 2010] or environmental sounds [Klobassa et al., 2009] can be used to develop a multi-class BCI.

The ultimate goal in BCI is to allow users to interact with a computer and/or the environment around them. Within the human-computer interaction domain, there has also been an on-going effort to develop suitable screen reading interfaces for visually impaired individuals. For example, JAWS (Job Access With Speech) is a computer screen reader program for Microsoft Windows that utilizes text-to-speech (TTS) synthesis to read the material on the screen back to the user; VoiceOver provides similar capabilities for Mac OS products. As such, by combining an auditory BCI with a TTS-based screen reader, several applications could be enabled for visually impaired individuals, such as website navigation and document reading, to name a few.

However the quality of existing TTS systems are still poor and clearly distinguishable from naturally-produced speech. Moreover, our previous studies have shown that low-quality speech stimuli can negatively (and unconsciously) affect P300 potentials, both in terms of amplitude and latency. In this study, we investigated the effects of TTS quality on P300 potentials, with the ultimate goal of understanding if TTS reader quality can have a negative effect on P300-based auditory BCIs. An experiment with 14 healthy subjects showed that there is a significant (inverse) main effect of TTS quality on P300 amplitude (i.e., lower quality, higher amplitude). As such, care must be exercised when developing auditory BCIs for the visually impaired with the ultimate goal of controlling a TTS-based screen reader.

2. Material and Methods

Fourteen subjects with normal hearing were recruited to participate in the study (eight female; mean age = 21.57 years; SD = 3.55). Ethical approval was obtained from the affiliated institutes and participants freely consented to participate. Three different TTS systems were chosen from the well-known international TTS challenge [King and Karaiskos, 2009] to represent three different quality ranges, namely high (mean opinion quality score, MOS, of 3.7 out of 5), medium (MOS = 2.8) and low (MOS = 2.1). Four sentences were used to generate 12 synthesized speech samples of approximately 10-second duration each; this was performed for each of the three TTS systems.

Synthesized speech stimuli was presented to the participants in a quasi-randomized fashion, with the only constraint that a given sentence s_i synthesized by the low-quality system had to be presented to the participants prior to

Published by Graz University of Technology Publishing House, sponsored by g.tec medical engineering GmbH

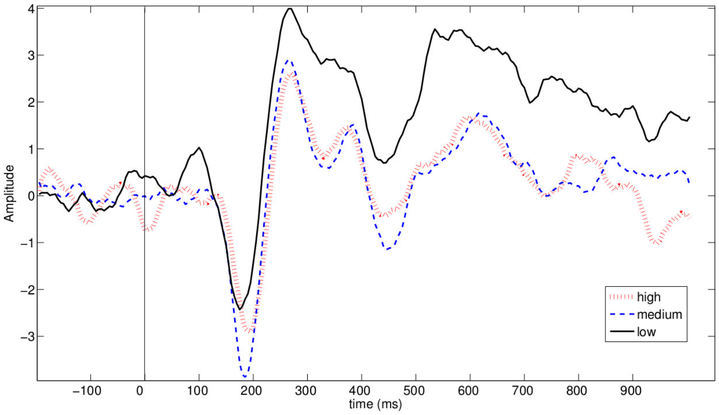

Figure 1: *P300 potentials elicited by low-, medium-, and high-quality TTS stimuli.*

presentation of the same sentence s_i synthesized by the high-quality TTS system, thus to avoid any memory biases. Moreover, an inter-stimulus-interval of 20 seconds was also used to decrease any remaining memory biases of the spoken content. Stimuli were presented binaurally at the individual's preferred listening level through in-ear headphones. And subjects were asked to judge the quality of TTS stimuli by pressing button with options ”pleasant” and ”unpleasant” as soon as possible at the start of the sentence. A 128-channel BioSemi ActiveII EEG system was used but only the following subset was recorded: 64 EEG-electrodes, 4 EOG-electrodes, and two mastoid-electrodes (right and left). Data was recorded at 512 Hz but down-sampled to 200 Hz and band-pass filtered between 1 and 40 Hz for offline analysis. To quantify the deviance-related effects of P300, we measured the peak amplitude in a fixed time window relative to the pre-stimulus baseline at electrode CPz. The time window for P300 quantification was set from 200 to 600 ms after stimulus onset.

3. Experimental Results and Discussion

P300 signal measured from electrode CPz is depicted in Fig. 1. Here it can be observed that P300 amplitude increases with a decrease in TTS quality. A repeated measures ANOVA was used to verify the difference between P300 peak amplitudes across three different TTS quality conditions, which showed a significant (inverse) main effect for TTS quality on P300 peak amplitude ($F(2,22) = 4.42$, $p < .05$). It is believed that the increase observed in the P300 amplitude with a decrease in TTS quality is due to increased cognitive functioning needed to fully comprehend the low-quality sentences being played back. Further studies are still needed to investigate if these negative effects persist even once the user has become habituated to the poor quality synthesized stimuli. In summary, it is recommended that auditory BCI developers aimed at developing computer interaction applications should pay careful attention to the possible effects of TTS screen reader quality on BCI performance.

Acknowledgments

The authors are grateful to the Ministere du Developpement Economique, Innovation et Exportation du Quebec, the National Science and Engineering Research Council of Canada, the German Federal Ministry of Education and Research (Grant FKZ 01GQ0850), the German Research Foundation (DFG-1013 ‘Prospective Design of Human-Technology Interaction’) and to the German Academic Exchange Service (DAAD) for funding this work.

References

Guo, J., Gao, S., and Hong, B. (2010). An auditory brain-computer interface using active mental response. *IEEE Trans Neural Syst Rehabil Eng*, 18(3):230–235.

King, S. and Karaiskos, V. (2009). The blizzard challenge 2009. In *Proceedings of the International Blizzard Challenge TTS Workshop*.

Klobassa, D. S., Vaughan, T. M., Brunner, P., Schwartz, N. E., Wolpaw, J. R., Neuper, C., and Sellers, E. W. (2009). Toward a high-throughput auditory p300-based brain-computer interface. *Clin Neurophysiol*, 120(7):1252–1261.