 DOI:10.3217/978-3-85125-260-6-108

The Use of Phonetic Similarity Cues in Auditory Spellers

Manson C-M. Fong[1], James W. Minett[1], Lin Zhou[1], Thierry Blu[1], William S-Y. Wang[1]
[1]*The Chinese University of Hong Kong, Hong Kong, China*

Correspondence: Manson C-M. Fong, Department of Electronic Engineering, The Chinese University of Hong Kong, Hong Kong, China.
E-mail: cmfong@ee.cuhk.edu.hk

Abstract. This study examines whether the ERPs elicited by a non-target stimulus in an oddball-based auditory speller are modulated by its phonetic similarity to the target. Both N200 and N400 amplitudes were significantly enhanced when a non-target contained the same initial consonant as the target, compared to when it did not. These additional modulations, which have not been exploited in previous studies, were used to train a classifier to provide partial information about target pronunciation, thereby complementing the standard target/non-target classifier. Offline classification results suggest that this approach is useful for enhancing speller performance.

Keywords: Auditory Speller, Phonetic Simiarlity, Speech, Electroencephalography, N2, Oddball Paradigm, Linear Discriminant Analysis

1. Introduction

Although both auditory spellers and visual spellers have generally been based on the oddball paradigm, the performance of auditory spellers using speech stimuli has been lower than that of visual spellers (see e.g., [Furdea et al., 2009]), especially when a “simple” task [Guo et al., 2010] is used. We posit that such a difference in performance arises in part because ERP waveforms are likely influenced by the phonetic similarity between target and non-target stimuli. In particular, the N2 component—one of the primary features for classification—is known to be enhanced in oddball tasks for both standards [Deguchi et al., 2010] and deviants [Nieuwenhuis et al., 2004] when the phonetic overlap among the stimuli is increased. The present study aims to characterize these modulations within the context of auditory spellers, and to assess whether they can be applied to enhance speller performance.

2. Material and Methods

Seventeen healthy native Hong Kong subjects (8 male and 9 female), aged 19–23, took part and were rewarded at about $7 per hour. Informed consent was obtained from each subject. The data for one subject were rejected since the subject reported that no sounds were heard in one ear during the first two blocks. Each subject completed two phases of training: constant-stimulus (CONST) and variable-stimulus training (VAR), each comprising 3 blocks of 18 runs. In each run, 6 Cantonese syllables were played 10 times each. The syllables were of 200–320 ms duration, and the SOA was 350 ms. Subjects were instructed to mentally repeat the target every time they perceived it.

In the CONST blocks, a single, constant set of 6 syllables was used as stimuli. For the VAR blocks, 18 distinct sets of 6 syllables served as stimuli to ensure that a representative range of phonetic information could be gathered during this training phase. In one of these blocks, syllables in each run were chosen to serve the following three roles: a target (T), a non-target which shared the same initial consonant (SI) as the target, and 4 non-targets which contained different initials (DI), e.g., T: geoi^2; SI: gung^6; and DI: bun^3, kwan^4, sang^1, and ming^4 (transcription in Jyutping, a standard romanization system for Cantonese). In the other two blocks, the roles of the syllables were adjusted such that the target was either a stimulus that had previously served as either SI (e.g., gung^6) or DI (e.g., bun^3). Two testing blocks, comprising runs in which the target shared the initial with 0, 1 or 2 non-target stimuli, were included to assess speller performance for stimuli sets with different degrees of phonetic overlap.

EEG data, acquired at 2048 Hz using a 32-channel BioSemi system, were down-sampled offline to 64 Hz, low-pass filtered at 6 Hz (which was found to provide optimal performance), and the linear trend within each epoch removed. ERP analyses were conducted to identify time-windows for selecting features. A standard target/non-target classifier was obtained for each type of training blocks using shrinkage LDA [Blankertz et al., 2011], with the feature vector associated with each epoch being constructed by concatenating all electrodes. A third classifier was trained using VAR blocks by combining the scores of two binary classifiers, the first discriminating T vs. DI and the second SI vs. T and DI. Since a high score for the second classifier indicated that the stimulus corresponded to a SI non-target, the stimuli with the same initial as the identified SI were more likely to be target—for these stimuli, the score associated with the identified SI was thus added to their scores for the first classifier.

3. Results

Fig. 1(a) shows the grand-averaged ERP waveforms recorded at the vertex (Cz) during the VAR training blocks. Two separate repeated-measures ANOVAs were run with factors: *Stimulus Type* [T, SI, DI] and *Electrode* [Fz, Cz, Pz] to examine the differences in potential within the N200 (188–250 ms) and N400 (375–438 ms) windows. The interaction effects were significant for both N200 [$p < 0.025$; Fig. 1(b)] and N400 [$p < 0.001$; Fig. 1(c)] windows. Post-hoc pairwise *t*-tests with Bonferroni correction were run to assess the differences across conditions. Of special note is that the average amplitude across the three electrodes was significantly larger for SI than DI for both windows, suggesting that these two windows contain cues that can be used to discriminate SI from DI.

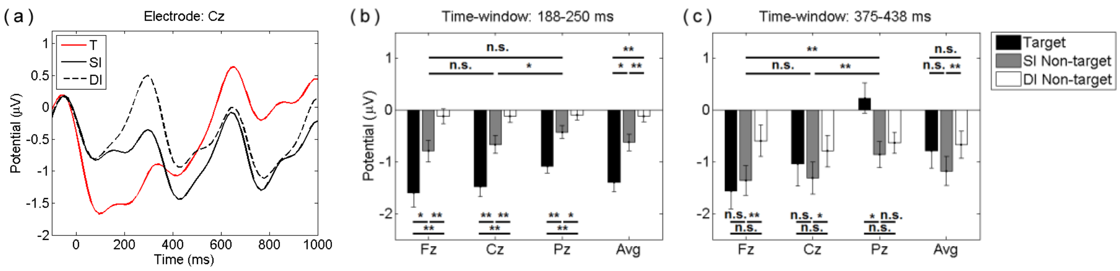

Figure 1. *(a) ERP measured at Cz during the VAR training blocks for the three conditions (T: target; SI: same-initial non-target; DI: different-initial non-targets). (b) Bar chart showing the average amplitudes measured at three electrodes (Fz, Cz, Pz) and the average of these electrodes within the time window 188–250 ms for the three conditions. (c) Bar chart showing the average amplitudes measured within the time window 375–438 ms for the three conditions.* ** $p < 0.005$; * $p < 0.05$; *n.s.: non-significant.*

Target classification accuracy was assessed using all 10 repetitions using either the CONST or VAR blocks to train the classifier. For runs comprising stimuli that did not overlap in initial, the accuracies obtained using a standard binary classifier were 67.2% (CONST) and 63.5% (VAR), while the corresponding accuracies for runs comprising stimuli that overlapped in initial were 69.8% (CONST) and 67.7% (VAR). The time-window for selecting features in all four cases was 0–1 s. For the latter type of runs, however, by training a classifier that discriminated SI from the other two types of stimuli using features from 0.3–1 s to complement the T vs. DI classifier, an enhanced performance of 73.2% was obtained.

4. Discussion

This study confirms that ERP amplitudes are modulated according to whether a syllable stimulus has the same initial consonant as the target. These modulations can be used to derive phonetic similarity cues for target identification. We have shown that partial information about target pronunciation can be obtained within the time-window 0.3–1 s, resulting in enhanced accuracy of about 5.5% in runs comprising stimuli that overlap phonetically with the target.

Acknowledgements

This work was supported by a research grant awarded to William S-Y. Wang by the Patent Committee of The Chinese University of Hong Kong (Grant No.: TBF/12/ENG/019).

References

Blankertz B, Lemm S, Treder M, Haufe S, Müller K-R. Single-trial analysis and classification of ERP components—A tutorial. *NeuroImage*, 56(2):814–825, 2011.

Deguchi C, Chobert J, Brunellière A, Nguyen N, Colombo L, Besson M. Pre-attentive and attentive processing of French vowels. *Brain Res*, 1366:149–161, 2010.

Furdea A, Halder S, Krusienski DJ, Bross D, Nijboer F, Birbaumer N., Kübler A. An auditory oddball (P300) spelling system for brain–computer interfaces. *Psychophysiol*, 46(3):617–625, 2009.

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

Nieuwenhuis S, Yeung N, Cohen JD. Stimulus modality, perceptual overlap, and the go/no-go N2. *Psychophysiol*, 41(1):157–160, 2004.