Feature Extraction and Enhancement of the Auditory N1-P2 Complex—Prediction of Perceptual Distances between Synthetic Vowels

Manson Cheuk-Man Fong [1,2]

[1] Joint Research Centre for Language and Human Complexity

[2] Department of Electronic Engineering

The Chinese University of Hong Kong, Hong Kong

Abstract

Purpose: An important aspect in the study of perception is to determine the distance relationship among perceptual objects. In this study, we tested whether or not the N1–P2 auditory evoked potentials, measured using electroencephalography (EEG), contained information for predicting the perceptual distances between vowels.

Method: We adopted a machine-learning framework to classify the auditory evoked responses to four synthetic mid-vowels differing only in second formant frequency (F2 = 840, 1200, 1680, and 2280 Hz). 6 subjects attended 4 EEG sessions each on separate days. Classifiers were trained using time-domain, single-trial features extracted according to two methods. In the first method, peak amplitudes and latencies were extracted for the N1-P2 complex. In the second method, a general linear model (GLM) was used to fit the single-trial EEG waveforms, based upon a design matrix comprising (i) Gaussian functions centered at the peaks and (ii) rectangular functions centered midway between the peaks. The resulting coefficients were selected as features.

Results: Based on the mean binary classification accuracy obtained by 20 trials, it was found that the GLM coefficient representations outperformed the traditional peak amplitude and latency features by about 10 %, and lead to comparable accuracy (> 70 %) to the case when the feature vector was formed by stacking all the time-samples across all electrodes. To assess how well perceptual distances among the vowels were reflected in our results, discriminability indices (d') were computed using both the behavioral results in a screening test and the classification results. It was found that although the behavioral discriminability indices were significantly correlated to their EEG counterparts obtained for both sets of features, the GLM coefficient representations gave rise to a stronger correlation.

Conclusion: Our results demonstrated that (i) the N1-P2 complex contains important features for predicting the perceptual distances between vowels; and (ii) the traditional peak features (amplitudes and latencies) could be enhanced using GLM to yield a representation that captures more information about the perceptual vowel space.