 DOI:10.3217/978-3-85125-260-6-35

# Pairwise Classification of Phoneme Imagery Using CSP

**X. F. Yang[1], M. H. Jiang[1]**

[1]*Center for Psychology and Cognitive Science, Tsinghua University, Beijing, China*

Correspondence: X. F. Yang, Room 313, Building No. 33, Tsinghua University, Haidian District, Beijing, China.
E-mail: xfyang.thu@gmail.com

***Abstract.*** We recorded EEG signals in three healthy subjects while imagining both the vocalization and places of articulation of four vowels (/a/, /i/, /u/, and /y/) and four consonants (/m/, /f/, /n/, and /ŋ/) in Mandarin Chinese, and a no action state as control. We examined the time-frequency features of phonemes as precise motor imagery and found the significant effects in spectral power lying in 2-12 Hz. Common spatial patterns were applied to EEG signals for spatial filtering, and the performance of pairwise classifications indicates potentiality for a speech neural prosthesis using phoneme imagery with more than 75% phoneme vs. control accuracy averaged across subjects.

*Keywords:* EEG, Phoneme Imagery, Communivative BCI, Common Spatial Patterns, Phonetic Features

## 1. Introduction

Brain-computer interfaces (BCIs) have been developed to restore communication and functionality to individuals with severe neuromuscular disorders. In this article, we propose a control scheme towards a true speech neural prosthesis using phoneme speech imagery. Electroencephalography (EEG) was recorded in three healthy subjects while imagining both the vocalization and places of articulation of four vowels (/a/, /i/, /u/, and /y/) and four consonants (/m/, /f/, /n/, and /ŋ/) in Mandarin Chinese as well as a no imagination state as control.

## 2. Material and Methods

### 2.1. Recording

EEG data were recorded from 32 Ag/AgCl electrodes (sampled at 500 Hz, bandpass filtering .01-100 Hz, notch filtering 50 Hz) positioned in a NeuroScan headcap on the scalp, according to the international 10/20 system.

### 2.2. Participants

Three undergraduates, recruited from Tsinghua University, participated in this study (two males, mean age 19.3 years, SD 2). All had normal or corrected to normal vision and none of them had history of neurological disease.

### 2.3. Stimuli

Nine kinds of stimuli were presented at the center of the screen in font SimSun-ExtB and size 30. Four of the stimuli are vowels; each represents a category of rhymes in the phonological system of Mandarin Chinese. Another four of the stimuli are consonants with distinguishable places of articulation. The phonetic features of the stimuli, together with a non-phonetic symbol # for control, with regard to the articulators are depicted in the Table 1.

### 2.4. Procedure

Coached beforehand and rehearsed with real movements to ensure correct articulation, subjects were instructed to imagine the movement of articulators as well as the vocalization of the phoneme as soon as they saw the visual target appeared on the screen. Each trial began with a fixation cross against a gray background for a randomized time interval between 1 to 2 s followed by a randomly chosen visual stimulus for 2 s, during which participants were instructed to perform the task of phoneme imagery and maintain the static imagination until the visual cue disappeared. After that, a blank gray screen was displayed for 3 s to serve as a rest interval. 50 trials were performed for each visual stimulus.

***Table 1.*** *Phonetic features of the stimuli concerning articulators.*

| Articulator | Status | /a/ | /i/ | /u/ | /y/ | /m/ | /f/ | /n/ | /ŋ/ | # |
|---|---|---|---|---|---|---|---|---|---|---|
| lips | unrounded | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 |
| | rounded | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| teeth | | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| tongue | high | 0 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 0 |
| | low | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | front | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |
| | back | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 |
| jaw | | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| alveolar ridge | | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| hard palate | | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| larynx | | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 |

### 2.5. Data Processing

All data processing was performed offline using MATLAB (version 7.14.0, MathWorks, Inc., Natick, MA) and EEGLAB [Delorme and Makeig, 2004]. The recorded EEG signal was re-referenced to the mean of the left and right mastoids and bandpass filtered between 1 and 45 Hz for preprocessing. We performed time-frequency analysis of phonemes vs. control and further filtered the data with the significant frequency band of 2-12 Hz. For all epochs, we selected a time window from 0 to 800 ms and randomly selected 40 of the 50 epochs 20 times per task to compose training data, upon which we applied the CSP method for spatial filtering. EEG training and testing data were decomposed using the spatial filters and then fed into an SVM model with a linear kernel for classification.

## 3. Results

In Fig. 1, we demonstrate the results of time-frequency analysis, common spatial patterns, and binary classification where we restrict the analysis to the pairwise problem condition vs. condition.

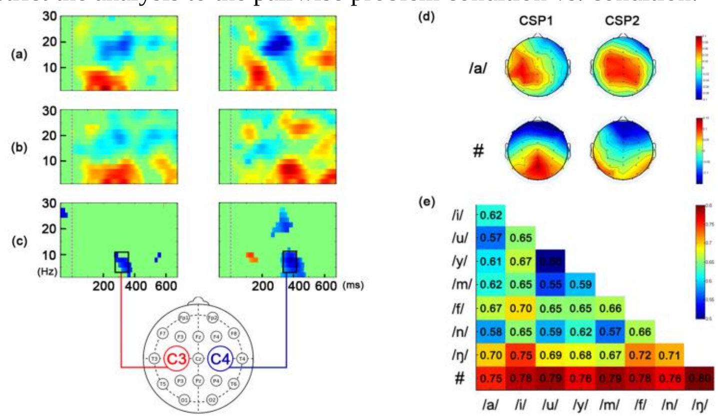


***Figure 1.*** ***(a)*** *ERSP to imagery of /a/,* ***(b)*** *ERSP to control condition, and* ***(c)*** *difference ERSP (/a/ minus control) on electrodes C3 & C4 for subject 3, in which significant phoneme imagery effect in spectral power was assessed with a bootstrap method and coloured in EEGLab ($p < 0.025$).* ***(d)*** *The two most important common spatial patterns of /a/ vs. control classification for subject 3.* ***(e)*** *Pairwise classification accuracies averaged across three subjects.*

## 4. Discussion

The significance of making BCI speech production more natural and fluent has been elaborated in [Brumberg and Guenther, 2010], and [DaSalla et al., 2009] has performed direct vowel prediction using EEG signals based on speech motor imagery with articulators corresponding to regions of the cerebral cortex [Guenther et al., 2006]. In our research, we have not only expanded imaginary speech movements to other phonemes, but also examined the frequency band of speech processing for EEG signals as in [Wang et al., 2012] and verified that speech articulation can be detected both in the scalp and cortex [Bouchard et al., 2013].

### Acknowledgements

This work was supported by NNSF in China (61171114) and Tsinghua self-determination Project.

### References

Bouchard KE, Mesgarani N, Johnson K, Chang EF. Functional organization of human sensorimotor cortex for speech articulation. *Nature*, available online, 2013.

Brumberg JS, Guenther FH. Development of speech prostheses: current status and recent advances. *Exp Rev Med Devices*, 7(5):667-679, 2010.

DaSalla CS, Kambara H, Sato M, Koike, Y. Single-trial classification of vowel speech imagery using common spatial patterns. *Neural Networks*, 22:1334-1339, 2004.

Delorme A, Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J Neurosci Meth*, 134:9-21, 2004.

Guenther FH, Ghosh SS, Tourville JA. Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain Lang*, 96:280-301, 2006.

Wang R, Perreau-Guimaraes M, Carvalhaes C, Suppes P. Using phase to recognize English phonemes and their distinctive features in the brain. *PNAS*, 109(50):20685–20690, 2012.