 DOI: 10.3217/978-3-85125-467-9-23

# Development of a Real Time Speech Synthesizer Based Brain Computer Interface.

**J. S. Brumberg[1,2*], J. D. Burnison[2], K. Pitt[1]**

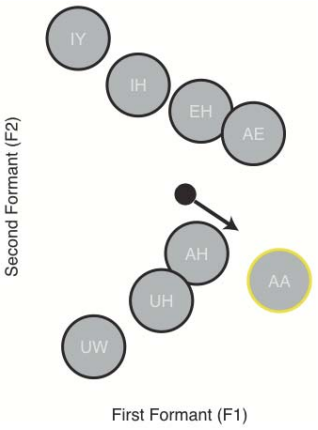
*[1]Speech-Langauge-Hearing, [2]Neuroscience Graduate Program, University of Kansas, Lawrence, KS, United States of America*
**Dole Human Developlment Center, Room 3001, 1000 Sunyside Ave, Lawrence, KS, USA. E-mail: brumberg@ku.edu*

*Introduction:* In the present study, we investigate an EEG-based BCI that directly controls a speech synthesizer for instantaneous acoustic speech output. Specifically, the BCI decodes modulations of the sensorimotor rhythm (SMR) in a motor imagery protocol for controlling two continuous parameters in a manner similar to 2D cursor control. However in this study, instead of defining graphical positions, the cursor defines the first two formant frequencies of speech, which can be used for instantaneous synthesis and auditory feedback [1] of as many as 8 vowel sounds in American English, though in the present study we only examine the production of 3 vowels.

*Material/Methods:* Three participants without neuromotor impairments have been recruited to participate in the study, with additional recruitment ongoing (21 participants are targeted). Participants are asked to complete a vowel imitation task that includes an initial training session followed by four sessions of online BCI control. During training, visual and/or auditory representations of vowel sounds are presented to the participant for four seconds depicting one of three target vowels (/aa/ [hot], /iy/ [heat] or /uw/ [hoot]). Our full participant pool will be separated into those receiving visual feedback only, auditory feedback only, or combined audio-visual feedback to test whether the auditory feedback leads to improved BCI performance. Visual stimuli are represented by the 2D formant position displayed as a cursor on the screen. Auditory stimuli are synthesized for both the training paradigm and online BCI feedback using a formant frequency synthesizer (Snack Sound Toolkit, KTH Royal Institute of Technology). Participants are instructed to imagine moving their left hand for /uw/ sounds, right hand for /aa/ sounds and both feet for /iy/ sounds. Model weights are then estimated for a Kalman filter decoder and used for online BCI control. In the control task, participants are presented with the audio and/or visual representation of a target vowel sound (20 trials per vowel) for 1.5 s and instructed to perform the appropriate kinesthetic limb motor imagery tasks to control the SMR-based BCI, which then outputs the continuously varying 2D formants from the center of the formant plane (the neutral vowel) to the target vowel (similar to the well-known center-out task). Instantaneous visual (cursor movements) and audio (synthesized vowel sounds) feedback is provided to the participants in the 6 s response period. All EEG data were recorded via 62-channel acquisition system (g.HIAmp, g.tec) at a sampling rate of 512 Hz. The sensorimotor rhythm was obtained using a fourth order low pass butterworth filter from 8-14 Hz, and the bandpower was calculated using the Hilbert transform.

*Results:* Trials were labeled correct when the predicted formants entered into the appropriate vowel region, and incorrect if they did not. Offline analysis of Kalman filter model weights show the sensorimotor regions contribute most to neural decoding, which is expected based on previous SMR studies of cursor control via limb motor imagery [2]. The mean vowel production accuracy from our preliminary study is approximately 70%.

*Discussion & Significance:* The advantage of the BCI described in this study its novel approach to decoding in formant frequencies of speech rather than attempting to use discrete classification of vowels. While motor imagery is not suitable for discrete classification of speech sounds (e.g., there are only 3-4 detectable motor imagery classes for 8 vowels and 30 consonants), it is suitable for low degree-of-freedom systems that use continuous control (e.g., 2D cursor control [2]). In the present study the trained sounds define the outer boundary of all vowel sounds in English, therefore, it is possible to produce all of the other vowels (e.g., /ih/ hid, /uh/ hood, /eh/ head, etc.) through combinations of imagined movements. For instance, the vowel /uh/ lies between /uw/ and /aa/ in the 2D formant plane, therefore, combined left and right hand movements will generate the appropriate formant frequencies to produce /uh/ without additional training. The successful completion of this project will provide the necessary data to proceed to test all 8 vowels, as well as to develop a new BCI in which users control a low degree of freedom articulatory synthesizer capable of producing all 38 phonemes in American English through use of just 3 or 4 continuous parameters (analogous to a cursor in 3D space).




*References*

[1] Guenther, F. H., & Brumberg, J. S. (2011). Brain-machine interfaces for real-time speech synthesis. In *Proceedings of the 33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC '11)*. Boston, MA.

[2] Wolpaw, J. R., & McFarland, D. J. (2004). Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(51), 17849.