

An Auditory Output Brain–Computer Interface for Speech Communication

Jonathan S. Brumberg, Frank H. Guenther and Philip R. Kennedy

Abstract Understanding the neural mechanisms underlying speech production can aid the design and implementation of brain–computer interfaces for speech communication. Specifically, the act of speech production is unequivocally a motor behavior; speech arises from the precise activation of all of the muscles of the respiratory and vocal mechanisms. Speech also preferentially relies on auditory output to communicate information between conversation partners. However, self-perception of one’s own speech is also important for maintaining error-free speech and proper production of intended utterances. This chapter discusses our efforts to use motor cortical neural output during attempted speech production for control of a communication BCI device by an individual with locked-in syndrome while taking advantage of neural circuits used for learning and maintaining speech. The end result is a BCI capable of producing instantaneously vocalized output within a framework of motor-based brain-computer interfacing that provides appropriate auditory feedback to the user.’

Introduction

One of the primary motivating factors in brain–computer interface (BCI) research is to provide alternative communication options for individuals who are otherwise unable to speak. Most often, BCIs are focused on individuals with locked-in

J. S. Brumberg (✉)

Department of Speech-Language-Hearing, University of Kansas, 1000 Sunnyside Ave.,
3001 Dole Human Development Center, Lawrence 66045 KS, USA
e-mail: brumberg@ku.edu

F. H. Guenther

Department of Speech, Language and Hearing Sciences, Department of Biomedical
Engineering, Boston University, Boston, MA, USA

P. R. Kennedy

Neural Signals, Inc, Duluth, GA, USA

syndrome (LIS) (Plum and Posner 1972), which is characterized by complete paralysis of the voluntary motor system while maintaining intact cognition, sensation and perception. One of the many reasons for this focus is that current assistive communication systems typically require some amount of movement of the limbs, face or eyes. The mere fact that many individuals with LIS cannot produce even the smallest amount of consistent motor behavior to control these systems is a testament to the severity of their paralysis. Despite such comprehensive motor and communication impairment, individuals with LIS are often fully conscious and alert, yet have limited or no means of self-expression.

A number of BCIs and other augmentative and alternative communication (AAC) systems provide computer-based message construction utilizing a typing or spelling framework. These interfaces often use visual feedback for manipulating the spelling devices, and in the case of BCIs, for eliciting neurological control signals. A common finding in patients with LIS is that visual perception is sometimes impaired, which may adversely affect subject performance when utilizing visually-based BCI devices. We address this issue through design of an intracortical auditory-output BCI for direct control of a speech synthesizer using a chronic microelectrode implant (Kennedy 1989). Part of our BCI approach benefits from prior findings for the feasibility of BCIs with dynamic auditory output (Nijboer et al. 2008). We extended the auditory output approach employing a motor-speech theoretical perspective, drawing from computational modeling of the speech motor system (Guenther 1994; Guenther et al. 2006; Hickok 2012; Houde and Nagarajan 2011), and our findings of motor-speech and phoneme relationships to neural activity in the recording site (Bartels et al. 2008; Brumberg et al. 2011), to design and implement a decoding algorithm to map extracellular neural activity into speech-based representations for immediate synthesis and audio output (Brumberg et al. 2010; Guenther et al. 2009).

Auditory Processing in Speech Production

Our speech synthesizer BCI decodes speech output using neural activity directly related to the neural representations underlying speech production. Computational modeling of the speech system in the human brain has revealed the presence of sensory feedback control mechanisms used to maintain error-free speech productions (Guenther et al. 2006; Houde and Nagarajan 2011). In particular, sensory feedback in the form of self-perception of auditory and somatosensory consequences of speech output is used to monitor errors and issue corrective feedback commands to the motor cortex. Our BCI design takes advantage of two key features: (1) auditory feedback in the form of corrective movement commands and (2) intact hearing and motor cortical activity typically observed in cases of LIS. These features are combined in our BCI to provide instantaneous auditory feedback driven through speech-motor control of the BCI. This auditory feedback is expected to engage existing neural mechanisms used to monitor and correct errors in typical speech production and send feedback commands to the motor cortex for updated control of the BCI.

Other groups have also investigated methods for directly decoding speech sounds from neural activity during speech production from a discrete classification approach using electroencephalography (DaSalla et al. 2009), electrocorticography (Blakely et al. 2008; Kellis et al. 2010; Leuthardt et al. 2011) and microelectrode recordings (Brumberg et al., 2011). These studies all illustrate that phoneme and word classification is possible using neurological activity related to speech production. The same LIS patient participated in both our microelectrode study of phoneme production and online speech synthesizer BCI control study. The results of our earlier study (Brumberg et al. 2011) confirmed the presence of sufficient information to correctly classify as many as 24 (of 38) phonemes above chance expectations (Brumberg et al. 2011). Each of these speech-decoding results could greatly impact the design of future BCIs for speech communication. In the following sections, we describe some of the advantages of using a low degree-of-freedom, continuous auditory output representation over discrete classification.

The BCI implementation (described below) employs a discrete-time, adaptive filter-based decoder which can dynamically track changes in the speech output signal in real-time. The decoding and neural control paradigms used for this BCI are analogous to those previously used for motor kinematic prediction (Hochberg et al. 2006; Wolpaw and McFarland 2004); specifically, the auditory consequences of imagined speech-motor movements used here are analogous to two-dimensional cursor movements in prior studies. Ideally, we would like to use motor kinematic parameters specifically related to the movements of the vocal tract as output features of the BCI device. Such a design is similar to predicting joint angles and kinematics for limb movement BCIs. However, there are dozens of muscles involved in speech production, and most motor-based BCIs can only accurately account for a fraction of the degrees of freedom observed in the vocal mechanism. We therefore chose a lower, two-dimensional acoustic mapping as a computational consideration for a real-time auditory output device.

The chosen auditory dimensions are directly related to the movements of the speech articulators. This dimension-reduction choice is similar to those made for decoding neurological activity related to the high degree of freedom movements of the arm and hand into two-dimensional cursor directions. Further, the auditory space, when described as a two-dimensional plane, is topographically organized with neutral vowels, like the ‘uh’ in ‘hut,’ in the center and vowels with extreme tongue movements along the inferior–superior and anterior–posterior dimensions around the perimeter (see Fig. 1 left, for an illustration of the 2D representation). In this way we can directly compare this BCI design to prior motor-based BCI utilizing 2D center-out and random-pursuit tasks.

Auditory Output BCI Design

The speech synthesis BCI consists of (1) an extracellular microelectrode (Bartels et al. 2008; Kennedy 1989) implanted in the speech motor cortex (2) a Kalman

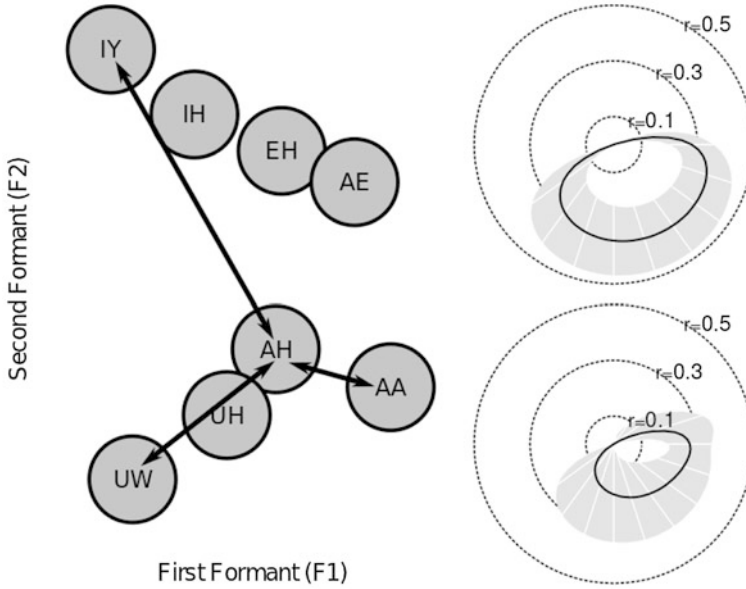


Fig. 1 *Left* 2D representation of formant frequencies. The *arrows* indicate formant trajectories used for training the neural decoder. *Right* examples of formant tuning preferences for two recorded units. The *black curve* indicates mean tuning preferences with 95 % confidence intervals in gray. The *top tuning curve* indicates a primarily 2nd formant preference while the *lower curve* indicates a mixed preference

filter decoding mechanism and (3) a formant-based speech synthesizer. The Kalman filter decoder was trained to predict speech formant frequencies (or formants) from neural firing rates. Formants are acoustic measures directly related to vocal tract motor execution used in speech production, and just the first two formants are needed to represent all the vowels in English. According to our speech-motor approach, we hypothesized that formants were represented by the firing rates of recorded neural units. This hypothesis was verified from offline analyses of the recorded signals (Guenther et al. 2009).

BCI Evaluation

To evaluate our speech synthesizer BCI, a single human subject with LIS participated in an experimental paradigm in which he listened to and repeated sequences of vowel sounds, which were decoded and fed back as instantaneously synthesized auditory signals (Brumberg et al. 2010; Guenther et al. 2009). We minimized the effect of regional dialects by using vowel formant frequencies that were obtained from vocalizations of a healthy speaker from the subject's family. The total delay from neural firing to associated sound output was 50 ms. The

subject performed 25 sessions of vowel–vowel repetition trials, divided into approximately four blocks of 6–10 trials per session. At the beginning of each session, the decoder was trained using the neural activity obtained while the subject attempted to speak along with a vowel sequence stimulus consisting of repetitions of three vowels (AA [hot], IY [heed], and UW [who’d]) interleaved with a central vowel (AH [hut]). The vowel training stimuli are illustrated graphically in Fig. 1. These four vowels allowed us to sample from a wide range of vocal tract configurations and determine effective preferred formant frequencies, examples of which are shown on the right in Fig. 1.

Following training, the decoder parameters were fixed and the subject participated in a vowel-repetition BCI control paradigm. The first vowel was always AH (*hut*) and the second vowel was chosen randomly between IY (*heed*), UW (*who’d*) or AA (*hot*). By the end of each session, the participant achieved 70 % mean accuracy (with 89 % maximum accuracy on the 25th session) and significantly ($p < 0.05$, t test of zero-slope) improved his performance as a function of block number for both target hit rate and endpoint error (see Fig. 2). The average time to target was approximately 4 s.

These results represent classical measures of BCI performance. However, the true advantage of a system that can synthesize speech in real-time is the ability to create novel combinations of sounds on-the-fly. Using a two-dimensional formant representation, steady monophthong vowels can be synthesized using a single 2D position while more complex sounds can be made according to various trajectories through the formant plane. Figure 3 illustrates an example in which the 2D formant space can be used to produce the words “I” (left) and “you” (middle), and the phrase “I owe you a yo-yo.” These words and phrases do not require any additions

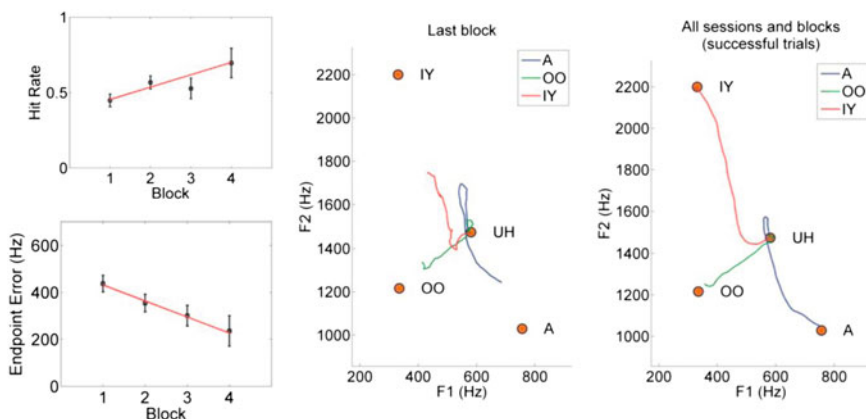


Fig. 2 Results from the speech synthesizer BCI control study. *Left* classical measures of performance, vowel target accuracy (*top*) and distance from target (*bottom*). *Middle* average formant trajectories taken for each of the three vowel–vowel sequences over all trials. *Right* average formant trajectories for each vowel–vowel sequence for successful trials only

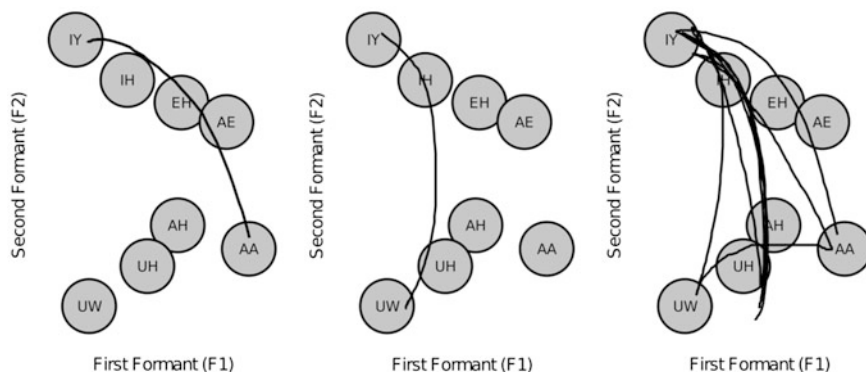


Fig. 3 An example of possible trajectories using manual 2D formant plane control. From *left to right*: selection of single formant pairs yields monophthong vowels; Trajectory from AA to IY yields the word “I”; Trajectory from IY to UW yields the word “you”; Complex trajectory shown yields the voiced sentence “I owe you a yo-yo”

to a decoding dictionary, as would be needed by a discrete classification BCI. Instead, the novel productions arise from new trajectories in the formant space.

Conclusion

These results are the first step toward developing a BCI for direct control over a speech synthesizer for the purpose of speech communication. Classification-based methods and our filter-based implementation for decoding speech from neurological recordings have the potential to reduce the cognitive load needed by a user to communicate using BCI devices by interfacing with intact neurological correlates of speech. Direct control of a speech synthesizer with auditory output yields further advantages by eliminating the need for a typing processes, freeing the visual system for other aspects of communication (e.g., eye contact) or for additional control in BCI operation. Future speech BCIs may utilize hybrid approaches in which discrete classification, similar to what is used for automatic speech recognition, are used in parallel to continuous decoding methods. The combination of both types of decoders has the potential to improve decoding rates while making the BCI a general communication device, capable of both speaking and transcribing intended utterances. Further, we believe that speech-sound feedback is better suited to tap into existing speech communication neural mechanisms, making it a promising and intuitive modality for supporting real-time communication.

The system as currently implemented is not capable of representing a complete speech framework, which includes both vowels and consonants. However, the results of our vowel-synthesizer BCI have led to a new line of research for

development of a low-dimensional (2-3D) articulatory-phonetic synthesizer for dynamic production of vowels and consonants. In addition, we are currently conducting studies using a non-invasive EEG-based sensorimotor (SMR) rhythm BCI for control of the vowel synthesizer as an alternative to invasive implantation. Early results from the non-invasive study with a healthy pilot subject have shown promising performance levels (~ 71 % accuracy) within a single 2-hour recording session. We expect users of the non-invasive system to improve performance after multiple training sessions, similar to other SMR approaches.

Acknowledgments Supported in part by CELEST, a National Science Foundation Science of Learning Center (NSF SMA-0835976) and the National Institute of Health (R03 DC011304, R44 DC007050-02).

References

- J. Bartels, D. Andreasen, P. Ehirim, H. Mao, S. Seibert, E.J. Wright, P. Kennedy, Neurotrophic electrode: method of assembly and implantation into human motor speech cortex. *J. Neurosci. Methods* **174**(2), 168–176 (2008)
- J.S. Brumberg, A. Nieto-Castanon, P.R. Kennedy, F.H. Guenther, Brain–computer interfaces for speech communication. *Speech Commun.* **52**(4), 367–379 (2010)
- C.S. DaSalla, H. Kambara, M. Sato, Y. Koike, Single-trial classification of vowel speech imagery using common spatial patterns. *Neural Netw.* **22**(9), 1334–1339 (2009)
- F.H. Guenther, A neural network model of speech acquisition and motor equivalent speech production. *Biol. Cybern.* **72**(1), 43–53 (1994)
- F.H. Guenther, S.S. Ghosh, J.A. Tourville, Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain Lang.* **96**(3), 280–301 (2006)
- F.H. Guenther, J.S. Brumberg, E.J. Wright, A. Nieto-Castanon, J.A. Tourville, M. Panko, R. Law, S.A. Siebert, J.L. Bartels, D.S. Andreasen, P. Ehirim, H. Mao, P.R. Kennedy, A wireless brain-machine interface for real-time speech synthesis. *PLoS ONE* **4**(12), e8218 (2009)
- G. Hickok, Computational neuroanatomy of speech production. *Nat. Rev. Neurosci.* **13**(2), 135–145 (2012)
- L.R. Hochberg, M.D. Serruya, G.M. Friebs, J.A. Mukand, M. Saleh, A.H. Caplan, A. Branner, D. Chen, R.D. Penn, J.P. Donoghue, Neuronal ensemble control of prosthetic devices by a human with tetraplegia. *Nature* **442**(7099), 164–171 (2006)
- J.F. Houde, S.S. Nagarajan, Speech production as state feedback control. *Frontiers Human Neurosci.* **5**, 82 (2011)
- J.S. Brumberg, E.J. Wright, D.S. Andreasen, F.H. Guenther, P.R. Kennedy, Classification of intended phoneme production from chronic intracortical microelectrode recordings in speech-motor cortex. *Frontiers Neurosci.* **5**(65), 1–14 (2011)
- S. Kellis, K. Miller, K. Thomson, R. Brown, P. House, B. Greger, Decoding spoken words using local field potentials recorded from the cortical surface. *J. Neural Eng.* **7**(5), 056007 (2010)
- P.R. Kennedy, The cone electrode: a long-term electrode that records from neurites grown onto its recording surface. *J. Neurosci. Methods* **29**(3), 181–193 (1989)
- E.C. Leuthardt, C. Gaona, M. Sharma, N. Szrama, J. Roland, Z. Freudenberg, J. Solis, J. Breshears, G. Schalk, Using the electrocorticographic speech network to control a brain–computer interface in humans. *J. Neural Eng.* **8**(3), 036004 (2011)
- F. Nijboer, A. Furdea, I. Gunst, J. Mellinger, D.J. McFarland, N. Birbaumer, A. Kübler, An auditory brain-computer interface (BCI). *J. Neurosci. Methods* **167**(1), 43–50 (2008)

- F. Plum, J.B. Posner, The diagnosis of stupor and coma. *Contemp. Neurol. Series* **10**, 1–286 (1972)
- T. Blakely, K.J. Miller, R.P.N. Rao, M.D. Holmes, J.G. Ojemann, Localization and classification of phonemes using high spatial resolution electrocorticography (ECoG) grids, in *IEEE Engineering in Medicine and Biology Society*, vol. 2008, pp. 4964–4967, 2008
- J.R. Wolpaw, D.J. McFarland, Control of a two-dimensional movement signal by a noninvasive brain–computer interface in humans. *Proc. Natl. Acad. Sci. U. S. A.* **101**(51), 17849 (2004)

Brain-Computer Interface Research

A State-of-the-Art Summary

Guger, C.; Allison, B.; Edlinger, G. (Eds.)

2013, VI, 123 p. 40 illus., Softcover

ISBN: 978-3-642-36082-4