Decoding Articulatory Properties of Overt Speech from Electrocorticography

E. M. Mugler[1], R. D. Flint[2], Z. A.Wright[2], S. U. Schuele[2], J. Rosenow[2], J. L. Patton[1], M. W. Slutzky[2,3,4]

[1]Bioengineering, University of Illinois at Chicago, Chicago, IL, USA; [2]Neurology, [3]Physiology, and [4]Physical Medicine and Rehabilitation, Northwestern University, Chicago, IL, USA

Correspondence: M. W. Slutzky, 303 E. Superior Street, Chicago, IL 60611. E-mail: mslutzky@northwestern.edu

Abstract. Brain-computer interface (BCI) applications for communication could greatly increase information transfer rates by directly decoding phonemes from cortical signals. We decoded individual phonemes and phonemic classes from electrocorticograms (ECoG) recorded while two subjects read words aloud from the Modified Rhyme Test. Information elicited from electrodes located over facial motor areas demonstrated distinct dynamic changes in high-γ and μ frequency power for different places of articulation, enabling us to correctly identify phoneme class in up to 64% of examples. We correctly classified 35% of all individual consonants by decoding the type and place of articulation. These results demonstrate successful decoding of phonemes from ECoG.

Keywords: Electrocorticography, Speech Production, Phonemes, Communication, Locked-in Syndrome

1. Introduction

A central goal of brain-computer interface (BCI) research is to enable communication for people with locked-in syndrome from such disorders as stroke or amyotrophic lateral sclerosis. BCIs have proven to enable such communication, predominantly via spelling paradigms. However, such BCIs have failed to achieve communication rates comparable to natural speech [Brumberg et al., 2010]. In contrast, recent advances in speech recognition technology have demonstrated that phonetic decoding can perform fast enough for natural speech production. Application of such a phonemic approach to decoding speech production from cortical signals could similarly improve communication rates for potential BCI users. In order to effectively decode intended speech, we first investigate what factors lead to successful decoding of overt speech using electrocorticography (ECoG).

Recent studies have shown promise in decoding speech from cortical signals but hold significant challenges to real BCI-to-speech application. [Brumberg et al., 2010] used action potentials from 31 spike clusters in facial motor cortex of a locked-in subject to classify correctly 21% of 38 American English phonemes, but only classified 24 phonemes above chance levels. ECoG may have greater longevity than spikes; one study used microwire ECoG from facial motor cortex to decode a set of 10 whole words with 48% accuracy in one subject [Kellis et al., 2010]. Studies employing standard-sized ECoG electrodes, with broader coverage area, have analyzed limited sets of phonemes, the best results of which had 91% classification of 2 isolated phonemes [Leuthardt et al., 2011; Blakely et al., 2008], and 41% decoding of phonemes from 4 vowels or 9 surrounding consonant pairs of 36 words [Pei et al., 2011]. Despite their successes, these studies were limited in comparison with the broad phonemic decoding that may be necessary for natural BCI-to-speech communication.

This is, to our knowledge, the first study to analyze all of the phonemes comprising American English as naturally spoken in words using ECoG, enabling several novel capabilities. This methodology enables us to isolate key neural aspects for successful decoding of speech sounds.

2. Material and Methods

This study was approved by the Institutional Review Board at Northwestern University, and subjects gave informed consent prior to experimental testing. Two subjects (female, age 30; male, age 50) who required extraoperative ECoG monitoring for treatment of their drug-resistant epilepsy participated in this study. Electrode placement was determined by medical necessity and included frontal and temporal areas in both subjects.

Subjects read aloud words presented on a monitor every 4 s using BCI2000. Words included the Modified Rhyme Test, a list of 300 monosyllabic words of consonant-vowel-consonant structure. Twenty additional words were included to create a comprehensive collection of General American phonemes in the set. Speech was recorded with a condenser microphone and digitized at 44.1 kHz. We recorded 42 and 48 channels of ECoG (1 cm spacing) at 500 Hz and 1 kHz and bandpass filtered from 0.5 to 250 or 300 Hz for subjects 1 and 2, respectively.

We first determined phoneme onset from spectral audio data. We computed power using short-time Fourier transforms on common-average-referenced ECoG signal for each channel (Hanning window, length 50 ms, 2 Hz bands), normalized by the average spectrum from a 1-s baseline. We extracted features by averaging band-power both over a 50 ms window and over δ (0-4 Hz), μ (8-13 Hz), β (15-30 Hz), and high-γ (70-200 Hz) frequency bands. We decoded the phonemic class (e.g. "bilabial" or "plosive" according to International Phonetic Association distinction) of consonants using linear discriminant analysis (LDA) on the features, selecting features with the lowest p-values (one-way ANOVA), with 10-fold cross-validation [Flint et al., 2012]. We performed simultaneous LDA classifications of place and manner of articulation as well as vocalization properties. We then multiplied posterior probabilities of each classifier to decode individual consonants (Bayesian decoding).

3. Results

Successful decoding of phonemic class (place of articulation) correlated strongly with electrode location. We classified 61% of consonant phonemes in the correct phonemic class (chance = 27%, $p < 10^{-43}$, *t*-test) in subject S2, who had 8 electrodes identified in both tongue and throat motor areas by electrical stimulation mapping. In contrast, subject S1 had only 4 electrodes covering tongue motor cortex; we classified 36% of her phonemes correctly ($p < 10^{-14}$), primarily those involving tongue. Interestingly, phonemes that were misclassified were most often confused for neighboring places of articulation. We correctly classified individual consonants in 35% of cases for S2. This rate was far above chance decoding (chance = 7%, $p < 10^{-131}$ for consonants).

The most discriminative features were high-γ band power increases and mu and beta band power decreases in laryngeal, tongue, and lip motor electrodes. We found a correspondence between p-value and articulatory location. Informative feature times varied by phonemic class: causal features (-250-0 ms) were most significant for bilabial phonemes, while acausal features (0-200 ms) dominated for post-alveolar and dental consonants. This suggests that movement *from* consonant articulation position mattered more for decoding than the movement *to* articulation position for the latter two classes. Features corresponding to lip articulation preceded features for vocalization.

4. Discussion

We attribute decoding success to articulation-based analysis aligned precisely to phoneme articulation onset. Our incorporation of the entire spectrum of articulation enables decoding of consonants using established articulatory principles of phonetics. Importantly, in S2, we found distinct high-γ band power patterns for lip, tongue, and laryngeal activity – a motor somatotopy. Since features for some articulatory classes (e.g. labiodental, palatal) overlapped multiple electrodes, we hypothesize that higher density ECoG grids may better isolate features from neighboring classes. Coverage of laryngeal, tongue, jaw, and lip motor areas best enables phoneme decoding. If phonemic decoding can be further improved, this technology has the potential to result in faster communicative BCI.

Acknowledgements

This work was supported in part by the Doris Duke Charitable Foundation and the NSF under Grant No. 0549489.

References

Blakely T, Miller KJ, Rao RPN, Holmes MD, Ojemann JG. Localization and classification of phonemes using high spatial resolution electrocorticography (ECoG) grids. *Conf Proc IEEE Eng Med Biol Soc*, 4964–4967, 2008.

Brumberg JS, Nieto-Castanon A, Kennedy PR, Guenther FH. Brain-Computer Interfaces for Speech Communication. *Speech Comm*, 52:367–379, 2010.

Brumberg JS, Wright JE, Andreasen DS, Guenther FH, Kennedy PR. Classification of intended phoneme production from chronic intracortical microelectrode recordings in speech-motor cortex. *Front Neurosci*, 5:1–12, 2011.

Flint RD, Lindberg EW, Jordan LR, Miller LE, Slutzky MW. Accurate decoding of reaching movements from field potentials in the absence of spikes. *J Neural Eng*, 9:046006, 2012.

Kellis S, Miller K, Thomson K, Brown R, House P, Greger B. Decoding spoken words using local field potentials recorded from the cortical surface. *J Neural Eng*, 7:056007, 2010.

Leuthardt E, Gaona C, Sharma M, Szrama N, Roland J, Freudenberg Z, Solis J, Breshears J, Schalk G. Using the electrocorticographic speech network to control a brain-computer interface in humans. *J Neural Eng*, 8:036004, 2011.

Pei X, Barbour DL, Leuthardt EC, Schalk G. Decoding vowels and consonants in spoken and imagined words using electrocorticographic signals in humans. *J Neural Eng*, 8:046028, 2011.