Cortical encoding of phonemic context during word production

Emily M. Mugler-*IEEE Member*, Matthew Goldrick, and Marc W. Slutzky, *Senior Member, IEEE*

***Abstract*—Brain-computer interfaces that directly decode speech could restore communication to locked-in individuals. However, decoding speech from brain signals still faces many challenges. We investigated decoding of phonemes – the smallest separable parts of speech - from ECoG signals during word production. We expanded on previous efforts to identify specific phoneme by identifying phonemes by where in the word they were formed. We evaluated how the context of phonemes in words affects classification results using linear discriminant analysis. The decoding accuracy of our linear classifier indicated the degree to which the context of a phoneme can be determined from the cortical signal significantly greater than chance. Further, we identified the spectrotemporal features that contributed most to successful decoding of phonemic classes. Finally, we discuss how this can augment speech decoding for neural interfaces.**

I. Introduction

Neurally-controlled speech prostheses could aid people who are "locked-in" and unable to speak due to stroke or motor neuron disease. Current methods of augmentative communication, such as brain-computer interfaces (BCIs) or eye-tracking technologies, primarily involve a typing paradigm, in which communication or commands are selected letter by letter [1]. Alternatively, if BCIs could access and utilize cortical speech production signals, BCIs could substantially improve their efficiency and would be much more intuitive to use. However many technical challenges prevent realization of such speech decoding.

One paramount challenge to speech BCI development is that the change in motor cortical signal during speech production is not fully understood. Moreover, the organization of and how the motor cortex executes speech – from how words are first planned to how they are produced by the vocal tract – is currently debated. The organization of speech motor cortex has been theorized to be based on anatomical structures [2], [3], syllables [4], and phonemes [5], the smallest unique units of speech that can be combined into words. By investigating cortical dynamics during speech production, recent studies using electrocorticography (ECoG) have decoded specific words [6] and phonemes [7]–[9] using speech-related band power changes. If cortical representation were based on phonemes, there would be advantages for neural interface design, as phonemes can be combined in myriad ways and represent speech sounds of multiple languages [10]. Moreover, a better understanding of the actual encoding in the motor cortex would enable BCIs to be more biomimetic and hence more intuitive for the user.

Thus far, however, most successful studies that decoded phonemes from cortical signals have used small, closed sets of phonemes. One reason phoneme decoding has been limited in success when applied to larger sets may be that the vocal tract position can vary for a given phoneme. Moreover, each speech sound can be slightly different in execution of vocal tract articulators due to *contextual* placement within a word (e.g. the placement of the tongue during the /n/ phoneme of "tenth" and "ten"). These contextual differences may also change the cortical representation of a given phoneme. If this context is also encoded in the motor cortex, we could incorporate that information into an improved phoneme decoding algorithm. Information about position of phonemes could augment whole word decoding. Further, it would mean that phonemes can vary in context, and that the cortical plan does not only care about reaching acoustic targets (i.e. phonemes), but also the way phonemes are produced.

In this study, we investigate the high gamma power changes in motor cortex during production of consonants and their placement within words. Specifically, we investigate the difference in signal of isolated speech sounds, e.g. consonants, due to their contextual position within a word. By investigating and classifying single-trial differences in how a sound is produced at the beginning compared to the end of the word, we can determine how phonemes change in context of words. Moreover, identifying critical components of how the high gamma band power signal changes in relation to this context may aid further attempts at decoding speech from ECoG signal.

II. Methods

A. Signal Acquisition

We implanted an ECoG array (Integra Inc., 2.3 mm platinum electrodes, 1 cm inter-electrode spacing) in a subject undergoing surgery for intractable epilepsy. We recorded signals with a Nihon Kohden Neurofax system and 1 kHz sample rate. Electrical stimulation mapping determined that 8 electrodes were located in speech motor cortex. Electrode locations on the cortex were determined by co-registering the CT and MR images[11].

A USB large-diaphragm condenser microphone (MXL) was placed approximately 18 inches from the subject's mouth. Words were presented on a screen using BCI2000 software at a rate of 1 word every 4 seconds. Words were sampled from the Modified Rhyme Test and had a consonant-vowel-consonant structure [12]. The subject was

This research was supported by in part by the Doris Duke Charitable Foundation Clinical Scientist Development Award (Grant No. 2011039) and the National Science Foundation (Grant No. 0549489, 0718558).

E. M. Mugler is with the Department of Neurology, Northwestern University, Chicago, IL 60611 USA (312-503-6097; e-mail: emily.mugler@northwestern.edu).

M. Goldrick is with the Department of Linguistics, Northwestern University, Evanston, IL 60208 USA. (e-mail: matt-goldrick@northwestern.edu).

M. W. Slutzky is with the Departments of Neurology, Physiology, and Physical Medicine and Rehabilitation, Northwestern (e-mail: mslutzky@northwestern.edu).

978-1-4244-7929-0/14/$26.00 ©2014 IEEE

instructed to read the word aloud upon display (Fig. 1). Neural and audio data were synchronized offline using a syncing pulse signal produced by a Tucker-Davis Technologies Bioamp system.

The study was approved by the Institutional Review Board of Northwestern University and the patient provided informed consent for his participation.

B. *Signal Processing*

We bandpass filtered the ECoG signals from 0.53 Hz to 300 Hz and applied a common-average reference spatial filter. We extracted the power in the high-gamma band (70-300 Hz) by computing the amplitude of a Hilbert transform on 8 separate 20-Hz bands within the high gamma range, excluding 60-Hz harmonics, and calculating the mean of these 8 sub-bands. High gamma band power was then smoothed using a 5-ms, 3rd order Savitzky–Golay filter for each electrode. This signal was then normalized to mean high gamma power for that electrode over the course of the entire recording session.

We labeled onset of each phoneme within each word of the data set manually, using visual and auditory inspection of microphone voltage and audio spectrogram. This time label of the onset of each phoneme occurred at the acoustic release of the consonant, or the time at which the vocal tract's acoustic energy is greatest during phoneme utterance.

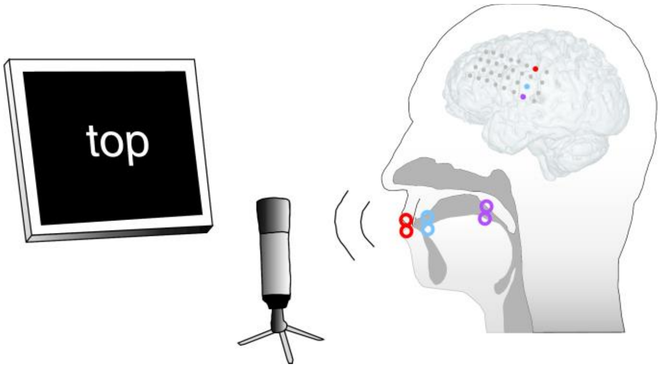

Figure 1. Recording set-up with schematic for ECoG subject and the co-registered electrode array for that subject. The highlighted electrodes demonstrate locations exhibited peak high gamma activity for /p/ (red), /s/ and /t/ (blue) and /k/ (purple). Open circles along the subject's vocal tract demonstrate the location of the primary constriction during consonant production in corresponding colors.

C. *Experimental Protocol*

To determine the extent to which a consonant's position within a word was encoded in motor cortex high gamma activity, we trained a linear classifier on the high gamma power in two different contexts: beginning and end of the word.

For 3 consonant phonemes with at least 90 examples in our data set (/k/, /p/, and /t/), we isolated the words which included that phoneme. We organized these instances into two classes: words with the phoneme as the initial consonant and those with phoneme as the final consonant. We aligned the ECoG data for each trial to the marked onset of the phoneme release. The high number of repetitions instances helped ensure that signals based upon surrounding phonemes would get averaged out. The mean duration of phonemes, measured from until acoustic signal ended or until the following phoneme started was 184.3 ms, 144.5 ms, 323.9 ms and 209.2 ms for /k/, /p/, /s/ and /t/, respectively.

For each consonant, we analyzed the ECoG electrode that varied the most in the entire high gamma band range during its production. This electrode was determined by performing a one-way ANOVA on all ECoG electrodes for that consonant. Restricting our analysis to a single electrode of ECoG activity ensured that neighboring electrodes potentially containing information of unrelated speech features could not confound decoding results. This criterion for electrode selection was used to ensure we were using only electrodes related to speech. The representative electrode for each phoneme roughly correlated with the somatotopic organization of the motor cortex for speech articulators [2], [3]: lip, tongue tip, and tongue body representation in dorsal to ventral order.

We used 10-fold, cross-validated linear discriminant analysis (LDA) to decode the location of the consonant within the word (initial or final position) [13], [14]. The features that served as inputs to the decoder were high gamma power in 25 ms bins from 100 ms prior to onset of acoustic release of the consonant through 50ms after release. Importantly, this time window contained no information outside of the preparation for and utterance of the phoneme of interest.

For each fold of the cross-validation procedure, we randomly selected the instances of each consonant. We repeated the 10-fold cross-validation 5 times and used the mean performance as the overall accuracy of decoding for each consonant. This random selection and repetition was performed to avoid potential confounds from nonstationarity over the course of the recording session. Because the frequency of consonant position (initial *vs.* final) was non-uniform, chance decoding performance was determined empirically by shuffling the *labels* of the data and recalculating LDA accuracy percentages 100 times. Significance was determined by calculating the 95% confidence interval using the decoding accuracies of these 100 runs. To further compare between the disproportionate position information values, we calculated the sensitivity index d', a statistic commonly used in signal theory, which compares the signal and noise means to the distribution of the noise [15]. Therefore the higher the sensitivity index, the more separable the signal is from the noise.

III. Results

We decoded the context of the consonant in the word with a mean accuracy of 76.94% when high gamma band power changes for words were aligned to the onset of the consonant release (Table 1). This far outperforms chance performance for each phoneme compared to the frequency of occurrence of position for that phoneme. This frequency of occurrence value is represented by the calculated ratio of initial to total consonants for each phoneme.

Generally, the consonants at the beginning of the word exhibited an earlier peak in high gamma band power increase (Fig. 2). This difference averaged 208 ms ± 74 ms between peak onsets when aligned to acoustic release onset. Peak high gamma power values were consistent (43.32 ± 2.5

for initial consonant, 42.14 ± 2.7 for final consonant, normalized power units) despite contextual position, but the 100 ms prior to and the 50 ms following the peak activity differed significantly (Fig 2). This enabled our classifier to decode the position better than chance.

TABLE I. DECODING INITIAL VERSUS FINAL PHONEME POSITION FROM ONSET

Phoneme	**Decoding Results**				
	n	***initia total***	*Accuracy (%)*	***95% CI for shuffled trials (%)***	***d'***
/k/	94	0.46	78.25	59.70 - 68.78	22.11
/p/	96	0.63	68.67	49.60 - 62.11	9.66
/s/	108	0.64	74.53	47.58 - 60.32	13.33
/t/	124	0.33	78.22	29.78 - 44.67	16.29

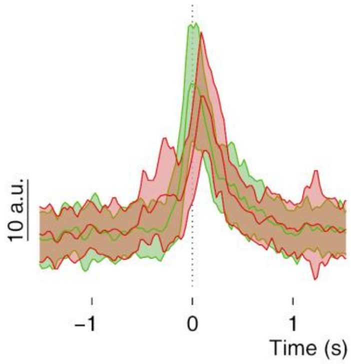

Figure 2. High gamma band values over time for a single ECoG electrode for the 43 words that start with /k/ (green) and the 42 words that end with /k/ (red). Values are in arbitrary units (a.u.) of normalized power. Mean (red and green lines) and confidence bounds (shaded areas) are aligned to the onset of the acoustic release.

IV. DISCUSSION

We used motor cortical signals to distinguish the position of consonants within spoken words. Further, this contextual position was distinguishable with the use of only a single electrode of ECoG activity. This implies that the primary motor cortex contains substantial information about the context of phonemes.

Our results further highlight that the cortical dynamics in motor cortex mimic the actual motor activity, given that our algorithm can distinguish the context of phonemes and how the vocal tract changes from cortical signal. Moreover, this result suggests that the acoustic target – the phoneme – may not be the only, or even the primary, speech parameter encoded in speech motor cortex. Rather, the primary motor cortex may also encode the speech motor plans of each articulator.

Understanding how a phoneme's position affects its production may explain other recent results in the reported speech-ECoG literature. Practically, in speech, no phoneme exists alone – phonemes must be combined to have any meaning in words. We have found that we can decode phonemes within words from cortex [16], but context and position could further improve these decoding results. In order to utilize the combinatory power of phoneme decoders, we should steer towards studying phonemes in the context of words [7]. Moreover, our results aligned with the reported somatotopic map of the speech motor cortex during overt vocal production [3], suggesting that motor targets matter more than acoustic targets (phonemes) in motor cortical activity.

For engineering brain-computer interfaces, this result builds on the ability to determine what phoneme is produced by revealing likelihood of where such a phoneme is within the word. When combined with prior knowledge of phonemic context within a language, these statistics can be used to supplement decoding of speech. A better understanding of how phonemes differ in context should enable improvement of phoneme-based BCIs for decoding of speech.

ACKNOWLEDGMENT

We thank the subject for his participation in our study, as well as Drs. Stephan Schuele and Joshua Rosenow for assistance with patient recruitment. We also thank Zachary Wright, Robert Flint, and our EEG technicians for assistance with recordings.

REFERENCES

[1] J. S. Brumberg, A. Nieto-Castanon, P. R. Kennedy, and F. H. Guenther, "Brain-Computer Interfaces for Speech Communication," *Speech Commun.*, vol. 52, no. 4, pp. 367–379, Apr. 2010.

[2] W. Penfield and E. Boldrey, "Somatic motor and sensory representation in the cerebral cortex of man as studied by electrical stimulation.," *Brain A J. Neurol.*, vol. 60, pp. 389–443, 1937.

[3] K. E. Bouchard, N. Mesgarani, K. Johnson, and E. F. Chang, "Functional organization of human sensorimotor cortex for speech articulation," *Nature*, Feb. 2013.

[4] G. Hickok, "Computational neuroanatomy of speech production," *Nat. Rev. Neurosci.*, vol. 13, pp. 135–145, 2012.

[5] F. H. Guenther, "Cortical interactions underlying the production of speech sounds," *J. Commun. Disord.*, vol. 39, no. 5, pp. 350–65, 2006.

[6] S. Kellis, K. Miller, K. Thomson, R. Brown, P. House, and B. Greger, "Decoding spoken words using local field potentials recorded from the cortical surface," *J. Neural Eng.*, vol. 7, no. 5, p. 056007, Oct. 2010.

[7] X. Pei, D. L. Barbour, E. C. Leuthardt, and G. Schalk, "Decoding vowels and consonants in spoken and imagined words using electrocorticographic signals in humans," *J. Neural Eng.*, vol. 8, no. 4, p. 046028, Jul. 2011.

[8] E. C. Leuthardt, C. Gaona, M. Sharma, N. Szrama, J. Roland, Z. Freudenberg, J. Solis, J. Breshears, and G. Schalk, "Using the electrocorticographic speech network to control a brain-computer interface in humans," *J. Neural Eng.*, vol. 8, no. 3, p. 036004, Jun. 2011.

[9] E. M. Mugler, R. D. Flint, Z. A. Wright, S. U. Schuele, J. Rosneow, and J. L. Patton, “Decoding Articulatory Properties of Overt Speech from Electrocorticography,” *Proc. Fifth Int. Brain-Computer Interface Meet. 2013, Pacific Grove, CA, June 3-7, 2013.*, pp. 4–5, 2013.

[10] A. Brown, “International Phonetic Alphabet,” *Encycl. Appl. Linguist.*, 2013.

[11] D. Hermes, K. J. Miller, H. J. Noordmans, M. J. Vansteensel, and N. F. Ramsey, “Automated electrocorticographic electrode localization on individually rendered brain surfaces,” *J. Neurosci. Methods*, vol. 185, no. 2, pp. 293–8, Jan. 2010.

[12] A. S. House, C. Williams, M. H. L. Hecker, and K. D. Kryter, “Psychoacoustic speech tests: A modified rhyme test,” *J. Acoust. Soc. Am.*, vol. 35, p. 1899, 1963.

[13] R. D. Flint, E. W. Lindberg, L. R. Jordan, L. E. Miller, and M. W. Slutzky, “Accurate decoding of reaching movements from field potentials in the absence of spikes,” *J. Neural Eng.*, vol. 9, no. 4, p. 046006, Jun. 2012.

[14] E. M. Mugler, R. D. Flint, Z. A. Wright, S. U. Schuele, J. Rosenow, and J. L. Patton, “Decoding Articulatory Properties of Overt Speech from Electrocorticography,” pp. 5–6, 2013.

[15] P. Afshar, D. Moran, A. Rouse, X. Wei, and T. Denison, “Validation of chronic implantable neural sensing technology using electrocorticographic (ECoG) based brain machine interfaces,” *2011 5th Int. IEEE/EMBS Conf. Neural Eng.*, pp. 704–707, Apr. 2011.

[16] E. M. Mugler, J. L. Patton, R. D. Flint, J. Wright, Zachary A. Schuele, Stephan U. Rosenow, J. J. Shih, D. J. Krusienski, and M. W. Slutzky, “Direct classification of all American English phonemes using signals from functional speech motor cortex,” *J. Neural Eng.*