

2014

Direct Classification of All American English Phonemes Using Signals From Functional Speech Motor Cortex

Emily M. Mugler

James L. Patton

Robert D. Flint

Zachary A. Wright

Stephan U. Schuele

See next page for additional authors

Follow this and additional works at: https://digitalcommons.odu.edu/ece_fac_pubs



Part of the [Biomedical Commons](#), and the [Neurology Commons](#)

Repository Citation

Mugler, Emily M.; Patton, James L.; Flint, Robert D.; Wright, Zachary A.; Schuele, Stephan U.; Rosenow, Joshua; Shih, Jerry J.; Krusienski, Dean J.; and Slutzky, Marc W., "Direct Classification of All American English Phonemes Using Signals From Functional Speech Motor Cortex" (2014). *Electrical & Computer Engineering Faculty Publications*. 150.
https://digitalcommons.odu.edu/ece_fac_pubs/150

Original Publication Citation

Mugler, E. M., Patton, J. L., Flint, R. D., Wright, Z. A., Schuele, S. U., Rosenow, J., . . . Slutzky, M. W. (2014). Direct classification of all American English phonemes using signals from functional speech motor cortex. *Journal of Neural Engineering*, 11(3), 035015
doi:10.1088/1741-2560/11/3/035015

Authors

Emily M. Mugler, James L. Patton, Robert D. Flint, Zachary A. Wright, Stephan U. Schuele, Joshua Rosenow, Jerry J. Shih, Dean J. Krusienski, and Marc W. Slutzky



Published in final edited form as:

J Neural Eng. 2014 June ; 11(3): 035015. doi:10.1088/1741-2560/11/3/035015.

Direct classification of all American English phonemes using signals from functional speech motor cortex

Emily M. Mugler¹, James L. Patton¹, Robert D. Flint², Zachary A. Wright², Stephan U. Schuele², Joshua Rosenow², Jerry J. Shih⁵, Dean J. Krusienski⁶, and Marc W. Slutzky^{2,3,4}

¹Bioengineering, University of Illinois at Chicago, 851 S. Morgan Street, Chicago, IL, USA 606 07

²Neurology, Northwestern University, 303 E. Superior Street, Chicago, IL 60611

³Physiology, Northwestern University, 303 E. Superior Street, Chicago, IL 60611

⁴Physical Medicine and Rehabilitation, Northwestern University, 303 E. Superior Street, Chicago, IL 60611

⁵Neurology, Mayo Clinic; 4500 San Pablo Road, Jacksonville, FL 32224

⁶Electrical & Computer Engineering, Old Dominion University, Norfolk, VA 23529

Abstract

Although brain-computer interfaces (BCIs) can be used in several different ways to restore communication, communicative BCI has not approached the rate or success of natural human speech. Electrocorticography (ECoG) has precise spatiotemporal resolution that enables recording of brain activity that is distributed over a wide area of cortex, such as during speech production. In this study, we investigated words that span the entire set of phonemes in the General American accent using ECoG with 4 subjects. We classified phonemes with up to 36% accuracy when classifying all phonemes and up to 63% accuracy for a single phoneme. Further, misclassified phonemes follow articulation organization described in phonology literature, aiding classification of whole words. Precise temporal alignment to phoneme onset was crucial for classification success. We identified specific spatiotemporal features that aid classification, which could guide future applications. Word identification was equivalent to information transfer rates as high as 3.0 bits/s (33.6 words/min), supporting pursuit of speech articulation for BCI control.

Keywords

electrocorticography; speech production; phonemes; linear discriminant analysis

1. Introduction

Brain-computer interfaces (BCI) can be used to restore communication in several different ways (see [1], [2] for reviews). However, communicative BCIs using primarily evoked potentials have not approached the rate or success of natural human communication [3]. One

approach that could provide higher throughput is to classify and decode neural signals related to speech production [4], but research in this field has failed to approach the efficiency of speech. Advances in electrocorticography (ECoG), in which field potentials are recorded directly from the surface of the cortex, may be able to improve decoding efficiency by recording from cortical speech areas. ECoG has been used to decode movement kinematics and kinetics [5], [6] and classify rapid cognitive processes [7]. ECoG recordings have precise temporal and spatial resolution compared to noninvasive techniques [8] and enable recording of rapid electrophysiological processes over a wide area of cortex [9]. ECoG can therefore facilitate mapping of rapid neural changes related to speech production [10], which involves concurrent activation in a wide area of cortex.

Most studies of speech production using ECoG to date have been intentionally limited in scope. Studies that employ a *whole-word* approach, classifying cortical activation patterns primarily based upon the differences between full words, initially identified the cortical areas that are active during speech articulation [11]. Classification of articulated words with micro-ECoG electrodes over facial motor cortex successfully identified at best less than half of 10 words in one patient [12]. Another study classified pairings of initial and final consonants by comparing the ECoG activation relative to word onset, and achieved up to 45% classification of a single consonant pairing in one out of 8 subjects [13]. These whole-word studies demonstrate preliminary success in speech decoding, but ultimately such success rates cannot be extrapolated to more complex speech. Moreover, the current most efficient BCI for communication reports information rates of 2.1 bits/s [14], much lower than the average natural efficiency of human speech production at 25 bits/s [15]. Thus, perhaps the ultimate goal for a speech neuroprosthetic is an information transfer rate that approaches natural speech.

One way of improving information rates may be to specifically decode the smallest isolated segments of speech, called *phonemes*. This approach would use phonemes, rather than words, as the “events” around which to analyze changes in brain signal. Speech BCIs using intracortically-recorded spikes to decode phonemes have achieved up to 21% classification success of all phonemes [16], and demonstrated up to 70% classification success of discrimination of 3 imagined vowels in an individual with locked-in syndrome [17]. Similar studies using ECoG succeeded in classifying small subsets of phonemes, isolated from the context of words (4 phonemes in [18], 2 in [19]). One ECoG-based BCI achieved an average of 84% discrimination of 2 vowels for 2 subjects in real-time [19]. One recent study detailed an approximate “phonemotopic” map of the areas to target within motor cortex using intermediate-density ECoG electrodes, updating traditional somatotopic maps for motor cortex [10]. These approaches demonstrate the potential to decode phonemes from cortical signals.

To our knowledge, no ECoG study has specifically investigated phonemes as independent events within words. ECoG approaches to date have not tried to classify phonemes with millisecond precision in phoneme onset detection. Further, no ECoG study has investigated classification of a *comprehensive* set of phonemes for a language.

In this study, we investigated production of words using the entire set of phonemes in the General American accent of English using ECoG. The rationale for this study was that once the smallest segments of speech articulation were related to corresponding cortical signals, the first critical step toward motor-based speech prosthetics would be established. We attempted to identify specific factors of decoding success or failure as a guide for future approaches. Furthermore, we hypothesized that precisely synchronizing analysis to each individual phoneme event is crucial for accurately discerning event-related cortical activity. This synchronization could reveal speech production dynamics in cortex, enabling decoding of individual phonemes within articulation of words.

2. Methods

2.1. Subjects

Four subjects (mean age 42, 2 female) who required extraoperative ECoG monitoring for treatment of their intractable seizures gave informed consent to participate in this study. The Institutional Review Boards of Northwestern University and the Mayo Clinic approved this study. Electrode coverage of cortex, determined by medical necessity, included some frontal and temporal areas in all subjects, although the degree of frontal coverage varied widely. Electrical stimulation mapping was performed for clinical purposes to determine areas corresponding to speech motor function, defined by movement of speech articulators in response to stimulation, and provided a gold standard for functional identification of brain regions (Figure 1). ECoG electrode placement was determined by co-registering pre-implant magnetic resonance images with post-implant computed tomography scans [20], [21].

2.2. Data Acquisition

We simultaneously collected speech audio signal (sampled at 44.1 kHz) from a USB microphone (MXL) using customized BCI2000 software [22] and a Tucker-Davis Bioamp system. We synchronized this signal with ECoG signals recorded on a clinical system (Nihon Kohden for NU subjects and Natus XLTEK for the MC subject). ECoG sampling frequencies, which varied due to clinical settings, were 500 Hz for Subject NU1, 1 kHz for Subjects NU2 and NU3, and 9.6 kHz for Subject MC1. ECoG was subsequently bandpass filtered from 0.5-300 Hz for NU2, NU3 and MC1 and 0.5-120 Hz for NU1 (Figure 2).

2.3. Experimental Protocol

Prior to the start of the experiment, subjects were screened for accent, mother tongue, and foreign accent exposure. Subjects read words from the Modified Rhyme Test (MRT), consisting of 300 monosyllabic words, predominantly with consonant-vowel-consonant structure [23]. The frequency of phonemes within the MRT set roughly approximates the phonemic frequency found in the English language [24], and it also has a high prevalence of rhyming structures. Because the MRT did not include all phonemes present in the General American accent of English, 20 additional words, which included 4 phonemes excluded from the MRT ($\text{\textbackslash}e\text{\textbackslash}$, $\text{\textbackslash}j\text{\textbackslash}$, $\text{\textbackslash}\partial\text{\textbackslash}$, and $\text{\textbackslash}ar\text{\textbackslash}$) were added to the stimulus set to create a comprehensive collection of General American phonemes in full words.

Using BCI2000, we presented one word on a screen for 3 s, followed by a blank screen for 1 s. Subjects were encouraged to read each word aloud as soon as it appeared. Total trials per subject varied from 320 words (Subjects NU1 and MC1) to 480 words (in which the first 160 words of the stimulus set were repeated; Subjects NU2 and NU3). This resulted in a minimum of 600 consonant and 320 vowel phonemes for each subject.

2.4. Data Preprocessing

Data was reduced to time-frequency features for each trial [25] and further separated by phoneme (Figure 2 depicts a summary of data preprocessing). We inspected visual and auditory spectral changes in the microphone signal to manually label the onset of each phoneme within each word. Phoneme assignment was determined using the CMU Pronouncing Dictionary, which assumes General American pronunciation. ECoG signals were common-average referenced in the time domain. Signals were split into 4-s trials centered on word onset, defined as the start of the first phoneme of a word. We computed short-time Fourier Transforms (FFTs) on moving 150 ms windows of each ECoG electrode. A 2 Hz frequency step size was used. Power was computed relative to power in baseline activity during the first second of each trial (Matlab).

To create spectrotemporal features for each phoneme, we combined FFT coefficients within each frequency band to denote overall power changes for that band for each electrode. We used the following 5 frequency bands: high-gamma, separated into 3 segments that avoided the harmonics of 60 Hz noise (65-115 Hz, 125-175 Hz, and 185-250 Hz) [26], mu (7-13 Hz), and beta (15-30 Hz). We also used the local motor potential [27], the time-domain signal smoothed over 150 ms. We computed features by shifting the 150-ms window in 50-ms increments, and used the features from 300 ms prior to 300 ms after phoneme onset. These 12 time bins included the entire phoneme (mean phoneme duration was 176 ms) as well as any preparatory movements prior to phoneme production. While the period surrounding phoneme onset could contain some information about neighboring phonemes, this information quickly diminished with increasing number of samples of a given phoneme since neighboring phonemes differed in each sample.

The time-frequency power features were then sorted by phoneme. For the full 320 word stimulus set, 981 phonemes were analyzed; for subjects who completed 480 words, 1470 phonemes were analyzed (67.3% of which were consonants). To reduce feature dimensionality and ensure phoneme class separability, features were ranked according to p-values from one-way ANOVAs across phonemes during training.

2.5. Classification

The 140 features with the lowest p-values were selected to classify phonemes using linear discriminant analysis (LDA) [25], [28]. The optimal number of features was determined by systematically increasing the value until decoding performance reached a maximum. LDA was selected primarily because it allowed us to identify features that led to successful categorization. We used 10-fold cross-validation with randomly-selected test sets, which were fully independent of each other and of the training sets, to compute success rates.

We determined the most informative frequency bands by independently decoding with each individual frequency band using all electrodes. Similarly, we determined the most informative time features by investigating performance of individual time bins. Further, we evaluated performance with increasing numbers of time bins, adding bins to the feature set until performance no longer improved. Time bins were added in order of highest to lowest significance, and included all electrodes and frequency bands.

To determine peak performance for each subject, we excluded electrodes that did not contain significant features (p -values > 0.1) or that had no significant activity changes prior to or during phoneme onset. This prevented electrodes associated with auditory areas from confounding motor speech decoding. Features were selected from these electrodes for each subject to determine their best possible decoding performance (i.e., 140 features were selected from at least 288 possible features – at least 4 electrodes, 12 time bins, and 5 frequency bands plus LMP – from each subject Figure 1). To ensure decoding results were not influenced by phoneme distributions within folds, decoding percentages were determined by averaging 5 separate 10-fold cross-validation results (i.e., average of 50 folds). Chance classification percentages were determined by randomly shuffling phoneme labels and re-classifying; this process was repeated 100 times.

2.6. Estimation of Information Transfer Rate

The goal for this speech BCI approach is to decode phonemic information during speech production, but phonemes exist in combinations within words. We therefore analyzed how phonemic decoding of combinations of consonants could be applied to identify words of the data set. We further investigated this performance when constraining predictions of phoneme combinations to those existing in the stimulus set in the order of posterior probability. To calculate gross information transfer rate (ITR) from these results, we first calculated average word duration (520 ms) and phoneme duration (176 ms). ITR was then determined by multiplying the information capacity (in bits/phoneme) by classification success and rate of speech production [29]. This procedure was generalized to other results reported in the literature using speech duration times from our results. Conversion to words per minute from bits per second was estimated using bit rates for syllable production of speech [15].

3. Results

3.1. Classification performance

Vowels and consonants were analyzed separately [30]. Decoding results were significant from chance levels despite varying widely over subjects, largely due to the wide variation in coverage of face motor areas (Figure 3). Subject NU2 had the highest overall performance, in which 36.1% of consonant phonemes were correctly classified. The maximum performance for classifying any one phoneme was 63% ($\backslash k \backslash$ for Subject NU2). Averaged across all subjects, $20.4 \pm 9.8\%$ of all phonemes were classified correctly, significantly greater than chance decoding (7.4%, $p < 0.001$, t -test). Average classification performance for vowels across all subjects was $19.2 \pm 3.7\%$, also significantly greater than chance (12.9%, $p < 0.01$, t -test), with the best performance in NU2 of 23.9%. The most successfully classified

vowels were produced with similar articulatory positions to the most successfully classified consonants for a given subject. For example, Subject NU2 had the most accurate classifications of $\backslash i \backslash$, which involves a high back tongue, and $\backslash k \backslash$, which is also articulated with the back of the tongue close to the palate.

Interestingly, decoding results supported phonetic categorizations (Figure 4). When a phoneme was misclassified, it was typically classified as its nearest neighbor within the International Phonetic Alphabet (IPA) chart of pulmonic consonants [31], the standard for phonology and linguistics. The IPA chart organizes sounds across languages by their articulation location along the vocal tract and the degree or manner of that articulation [31]. Our results indicate that consonants similar in articulation location and manner are more often confused during classification. When misclassification incidence is compared via ratios of IPA to non-IPA neighbors, a phoneme is 15.5% more likely to be mislabeled as a direct neighbor than a phoneme distant on the IPA chart (24.2% for Subject NU2). Similarly, phonemes are 34.3% more likely to be mislabeled as a phoneme with a similar vocal tract constriction location (49.8% for Subject NU2). Further, notable exceptions to this rule reveal other properties of IPA organization. Nonpulmonic consonants, which include affricates ($\backslash tʃ \backslash$ in **chip**, $\backslash dʒ \backslash$ in **jump**) as well as approximants ($\backslash w \backslash$ in **win**), are most often confused with their closest pulmonic relatives. Thus $\backslash w \backslash$, a labialized velar approximant, is confused with both labial phoneme $\backslash b \backslash$ and velar phoneme $\backslash g \backslash$. Finally, poorest classification of phonemes occurred in those with few numbers of trials (< 10 repetitions).

We analyzed how performance varied with the number of phonemes classified, the number of phoneme samples, and the accuracy of phoneme onset (Figure 5). To investigate the degree to which the number of phonemes affects results, classification was restricted to subsets of phonemes included in descending order of their frequency of occurrence in the data set. This process yields a maximum performance of 72.3% using 4 phonemes for Subject NU2 (Figure 5a). For all subjects, performance decreased until approximately 15 phonemes were included (at which point additional, less common phonemes had fewer than 10 samples). To investigate how performance varied with quantity of phoneme samples, we classified all 24 consonant phonemes using increasing number of samples (Figure 5b). Performance increased with the number of samples for all subjects, which suggests that results could improve with more data. To investigate the effect of the precision of alignment to phoneme onset, we added variability to the onset time in the form of Gaussian noise (Figure 5c). Performance sharply decreased when the standard deviation of onset time variability increased to 100 ms, which is notably less than the 176 ms average length of a phoneme in time (ranging from 75 ms for $\backslash b \backslash$ to 282 ms for $\backslash s \backslash$). This result demonstrates the critical need for temporal precision in phonemic analysis, as performance sharply decreases as timing offset noise increases.

3.2. Analysis of Feature Contribution

To determine the ECoG factors that influenced performance, we decoded phonemes using each frequency band and time bin independently. The three segments of high-gamma band (65-250 Hz) produced the highest decoding performance of all frequency bands (mean of 79.8% of maximum decoding values across subjects for the 65-115 Hz band). The mu band

also showed significantly better than chance decoding (50.7% of maximum decoding values, $p < 0.05$). Local motor potential and beta bands could not decode phonemes significantly better than chance performance ($p > 0.1$). The most informative time bin occurred right at phoneme onset (0-50ms) across subjects. When combined, features spanning 200 ms before to 200 ms after phoneme onset accounted for 88.1% of peak performance. Most of the significant ECoG activity therefore occurs in immediate preparation for and during onset of phoneme production (i.e., causal activity). This minimized the influence of neighboring phonemes. All subjects had at least 4 electrodes with causal information about speech for decoding speech articulation.

Performance was predominantly best when only incorporating data from the electrodes located over traditional primary motor cortex. Although we discovered atypical functional organization for decoding phonemes for subject NU3, these subtemporal areas also responded to facial motor activity and facial sensation during electrical stimulation. Overall, in all subjects, the most information about phonemes was obtained from functional facial motor areas.

3.3. Information Transfer Rate

We calculated the classification success of all consonant combinations in our stimulus set with our best performing subject (NU2). We successfully identified 14.8% of these combinations without having ever trained our algorithm to decode words (chance = 0.83%, $p < 0.0001$, t-test). This investigation of phonemes within words outperforms simple joint probability of phoneme classification. When we constrained the predictions of phonemes for a whole word to only words used in the stimulus set, results improved to 18.8%. We calculated an information transfer rate of 3.0 bits/sec (equivalent to 33.6 words per minute) for a hypothetical BCI (Figure 6). This is higher than what would be expected for a similar motor control behavioral paradigm due to the high information transfer of speech.

4. Discussion

This study is the first to decode the entire set of phonemes from American English using ECoG. It is also the first to successfully analyze and classify individual phonemes within word production. We found that that an event-related methodology enabled us to decode phonemes within words by aligning to the onset of each speech sound. Although other ECoG studies have classified overt phonemes by comparing words with similar phonemes, by analyzing phonemes directly in context, we reveal properties of speech production that corroborate decades of phonetics research. Misclassification of cortical activity follows the similarities in designations of phonemes by the International Phonetic Alphabet. Using the guidelines of the IPA to reduce words to their phonemic components, and their corresponding identifiable patterns of cortical activity, we can then decode speech information efficiently.

Our results suggest specific spatiotemporal guidelines for future endeavors into speech decoding using ECoG, advancing the science behind speech BCI development. Recording with higher electrode density over or neighboring sensorimotor cortices likely would improve decoding performance substantially. Kellis and colleagues demonstrated that 5

electrodes over facial motor cortex with 1 mm spacing yielded best results for their 10-word stimulus set [12]. Our results expand on that result and suggest that there may be additional information necessary to decode other details of articulation across a 4-cm mediolateral span of speech sensorimotor cortex [10]. Thus, a high-density (1-2 mm) electrode array over an area of at least 4 cm of speech motor cortex may be optimal for decoding speech. Investigation of frequency content showed the high-gamma band provided the most information about speech motor activity. This is consistent with prior studies on hand and arm movements [25], [26], [32], [33]. Finally, although speech production, speech reception, and visual stimuli are strongly related (e.g. McGurk effect [34]), the current results are unlikely to be related to speech reception or visual confounds. Phonemes were classified using causal features made in preparation for or during movements of articulation. Further, misclassification of phonemes adhered to phonological designations for articulation and not potentially confounding visual or auditory factors. For example, /s/ and /f/, often aurally confused because of their similarities in frequency, are not confused within our data.

Comparisons with other speech ECoG studies are difficult due to large variability in the set of phonemes, number of trials, types of decoding, and electrode coverage. However, we can approximate comparisons by computing the efficiency (ITR) of our system (Error! Reference source not found.). Our results compare favorably with those of Kellis et al. [12], which identified 10 words with 48% success using their 5 best micro-ECoG electrodes, and Pei et al [13], which identified 4 vowels and 9 consonant pairs at 40.7% (Error! Reference source not found.). Although we did not directly train our decoders on whole words, we successfully identified phonemes in 14.8% of our 320 word set on a first attempt. A best-performance volitional control of a single /k/ channel for Subject NU2, computed similarly, could yield a theoretical ITR of 32 words per minute, higher than many current BCI communication systems. Finally, speech recognition algorithms could be applied to phonemic results to leverage the frequency of phonemes within words in the English language to exclude impossible scenarios (e.g. words starting with /ŋ/).

Although word identification was not sufficient for communication purposes at a mere 18.8%, it is notable that words can be identified from phonemic analysis alone. This result outperforms the joint probability of isolated phoneme prediction and indicates that classification across some words is better than others. It is important to note that we are not strictly decoding words, as we have prior information as to where phonemic onset occurs. However, the correct identification of phonemes within a word is a concrete step toward whole word decoding from a broader set of words.

Finally, to analyze neural activity during phoneme production, phonemes need to be precisely identified as events. The high temporal precision required to accurately decode phonemes using this method suggests that there will be challenges for translating these algorithms to a real-time BCI for locked-in patients. Other methods may be necessary to detect onset of attempted speech production to enable speech decoding.

Despite these limitations, this study establishes that decoding phonemic information from spoken words is not only possible, but follows findings of phonetics research. Our results confirm that ECoG is capable of decoding rapid motor components of speech production.

While current performance is good, we anticipate substantial improvement in classification with more phoneme repetitions, which were limited here due to clinical constraints. Phonemic decoding using an ECoG-based BMI may provide efficient and intuitive communication. Similar to the ways in which a keyboard can provide a higher information transfer rate than a mouse, such interfaces may be useful to individuals with locked-in syndrome and other communication disorders.

Acknowledgments

Grants: This work was supported in part by the Doris Duke Charitable Foundation (Grant No. 2011039), the National Science Foundation (Grant No. 0549489, 0718558, and 1064912), the National Institutes of Health (NIBIB/NINDS EB00856), and Mayo Clinic Foundation CR20 Grant.

References

- [1]. Birbaumer N, Cohen LG. Brain-computer interfaces: communication and restoration of movement in paralysis. *J. Physiol.* 2007; 579(Pt 3):621–36. [PubMed: 17234696]
- [2]. Wolpaw JR, Birbaumer N, McFarland DJ, Pfurtscheller G, Vaughan TM. Brain-computer interfaces for communication and control. *Clin. Neurophysiol.* Jun.2002 113(6):767–91. [PubMed: 12048038]
- [3]. Leuthardt, EC.; Cunningham, J.; Barbour, D. Towards a Speech BCI Using ECoG. Guger, C.; Allison, BZ.; Edlinger, G., editors. Springer Berlin Heidelberg; Berlin, Heidelberg: 2013. p. 93-110.
- [4]. Kennedy PR, Bakay R. a. Restoration of neural output from a paralyzed patient by a direct brain connection. *Neuroreport.* Jun.1998 9(8):1707–11. [PubMed: 9665587]
- [5]. Schalk G, Leuthardt EC, Brunner P, Ojemann JG, Gerhardt L. a, Wolpaw JR. Real-time detection of event-related brain activity. *Neuroimage.* Nov.2008 43(2):245–9. [PubMed: 18718544]
- [6]. Wang W, Collinger JL, Degenhart AD, Tyler-Kabara EC, Schwartz AB, Moran DW, Weber DJ, Wodlinger B, Vinjamuri RK, Ashmore RC, Kelly JW, Boninger ML. An electrocorticographic brain interface in an individual with tetraplegia. *PLoS One.* Jan.2013 8(2):e55344. [PubMed: 23405137]
- [7]. Manning JR, Sperling MR, Sharan A, Rosenberg E. a, Kahana MJ. Spontaneously reactivated patterns in frontal and temporal lobe predict semantic clustering during memory search. *J. †.* 2012; 32(26):8871–8878.
- [8]. Slutzky MW, Jordan LR, Krieg T, Chen M, Mogul DJ, Miller LE. Optimal spacing of surface electrode arrays for brain-machine interface applications. *J. Neural Eng.* Apr.2010 7(2):26004. [PubMed: 20197598]
- [9]. Schalk G. Can Electrographic (ECoG) Support Robust and Powerful Brain-Computer Interfaces? *Front. Neuroeng.* Jan.2010 3(June):9. [PubMed: 20631853]
- [10]. Bouchard KE, Mesgarani N, Johnson K, Chang EF. Functional organization of human sensorimotor cortex for speech articulation. *Nature.* Feb.2013
- [11]. Pei X, Leuthardt EC, Gaona CM, Brunner P, Wolpaw JR, Schalk G. Spatiotemporal dynamics of electrocorticographic high gamma activity during overt and covert word repetition. *Neuroimage.* Oct.2010
- [12]. 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.* Oct.2010 7(5):056007. [PubMed: 20811093]
- [13]. 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.* Jul.2011 8(4): 046028. [PubMed: 21750369]
- [14]. Bin G, Gao X, Wang Y, Li Y, Hong B, Gao S. A high-speed BCI based on code modulation VEP. *J. Neural Eng.* Apr.2011 8(2):025015. [PubMed: 21436527]

- [15]. Reed CM, Durlach NI. Note on Information Transfer Rates in Human Communication. *Presence Teleoperators Virtual Environ.* Oct.1998 7(5):509–518.
- [16]. Brumberg JS, Wright EJ, Andreasen DS, Guenther FH, Kennedy PR. Classification of intended phoneme production from chronic intracortical microelectrode recordings in speech-motor cortex. *Front. Neurosci.* May; 2011 5(May):1–12. [PubMed: 21390287]
- [17]. Guenther FH, Brumberg JS, Wright EJ, Nieto-Castanon A, Tourville J. a, Panko M, Law R, Siebert S. a, Bartels JL, Andreasen DS, Ehirim P, Mao H, Kennedy PR. A wireless brain-machine interface for real-time speech synthesis. *PLoS One.* Jan.2009 4(12):e8218. [PubMed: 20011034]
- [18]. Blakely TM, 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.* Jan.2008 2008:4964–7. [PubMed: 19163831]
- [19]. Leuthardt EC, 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.* Jun.2011 8(3):036004. [PubMed: 21471638]
- [20]. Miller KJ, Hebb AO, Hermes D, den Nijs M, Ojemann JG, Rao RNP. Brain surface electrode co-registration using MRI and X-ray. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* Jan.2010 2010:6015–8. [PubMed: 21097113]
- [21]. Hermes D, Miller KJ, Noordmans HJ, Vansteensel MJ, Ramsey NF. Automated electrocorticographic electrode localization on individually rendered brain surfaces. *J. Neurosci. Methods.* Jan.2010 185(2):293–8. [PubMed: 19836416]
- [22]. Schalk G, McFarland DJ, Hinterberger T, Birbaumer N, Wolpaw JR. BCI2000: a general-purpose brain-computer interface (BCI) system. *IEEE Trans. Biomed. Eng.* Jun.2004 51(6):1034–43. [PubMed: 15188875]
- [23]. House AS, Williams C, Hecker MHL, Kryter KD. Psychoacoustic speech tests: A modified rhyme test. *J. Acoust. Soc. Am.* 1963; 35:1899.
- [24]. Mines MA, Hanson BF, Shoup JE. Frequency of occurrence of phonemes in conversational English. *Lang. Speech.* 1978; 21(3):221–41. [PubMed: 732398]
- [25]. 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.* Jun.2012 9(4):046006. [PubMed: 22733013]
- [26]. Flint RD, Ethier C, Oby ER, Miller LE, Slutzky MW. Local field potentials allow accurate decoding of muscle activity. *J. Neurophysiol.* Jul.2012 108(1):18–24. [PubMed: 22496527]
- [27]. Schalk G, Kubánek J, Miller KJ, Anderson NR, Leuthardt EC, Ojemann JG, Limbrick D, Moran D, Gerhardt L. a, Wolpaw JR. Decoding two-dimensional movement trajectories using electrocorticographic signals in humans. *J. Neural Eng.* Sep.2007 4(3):264–75. [PubMed: 17873429]
- [28]. Slutzky MW, Jordan LR, Lindberg EW, Lindsay KE, Miller LE. Decoding the rat forelimb movement direction from epidural and intracortical field potentials. *J. Neural Eng.* Jun.2011 8(3):036013. [PubMed: 21508491]
- [29]. Wolpaw JR, Birbaumer N, McFarland DJ, Pfurtscheller G, Vaughan TM. Brain-computer interfaces for communication and control. *Clin. Neurophysiol.* Jun.2002 113(6):767–91. [PubMed: 12048038]
- [30]. 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.* Aug.2011 8(4):046028. [PubMed: 21750369]
- [31]. Brown A. International Phonetic Alphabet. *Encycl. Appl. Linguist.* 2013
- [32]. Stark E, Abeles M. Predicting movement from multiunit activity. *J. Neurosci.* Aug.2007 27(31):8387–94. [PubMed: 17670985]
- [33]. Ball T, Demandt E, Mutschler I, Neitzel E, Mehring C, Vogt K, Aertsen A, Schulze-Bonhage A. Movement related activity in the high gamma range of the human EEG. *Neuroimage.* Jun.2008 41(2):302–10. [PubMed: 18424182]
- [34]. McGurk H, MacDonald J. Hearing lips and seeing voices. *Nature.* 1976; 264

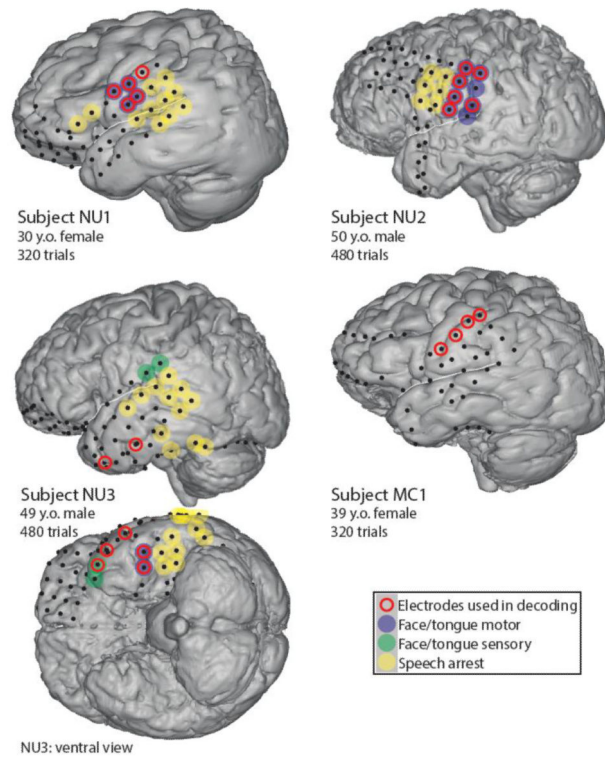


Figure 1.

Subject information and ECoG electrode locations (1cm spacing). Electrode coverage varied due to each patient's clinical needs. Red rings denote electrodes that contributed to best classification performance, which predominantly occurred in areas identified as facial motor cortex during stimulation. Blue, green, and yellow circles denote type of response to electrical stimulation mapping. White lines denote Sylvian fissure locations.

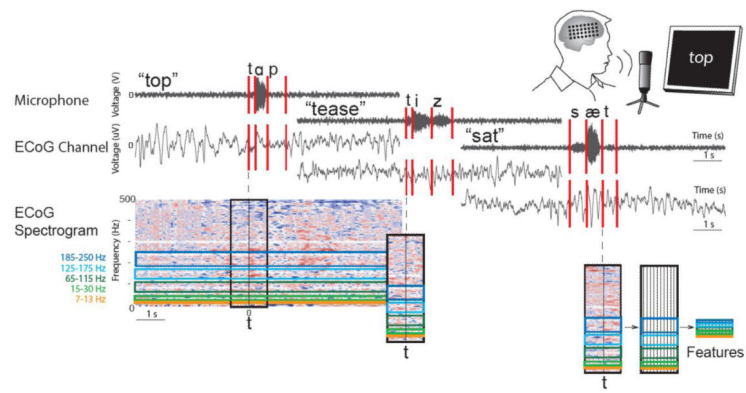


Figure 2.

Overview of data preprocessing. Speech signal is recorded simultaneously with ECoG signal (apparatus inset). These signals are marked according to onset of phoneme time and aligned with to the ECoG signal. An FFT is performed on the ECoG signal, and converted into features by combining FFT coefficients to form each frequency band of interest and 50 ms time windows relative to phoneme onset time.

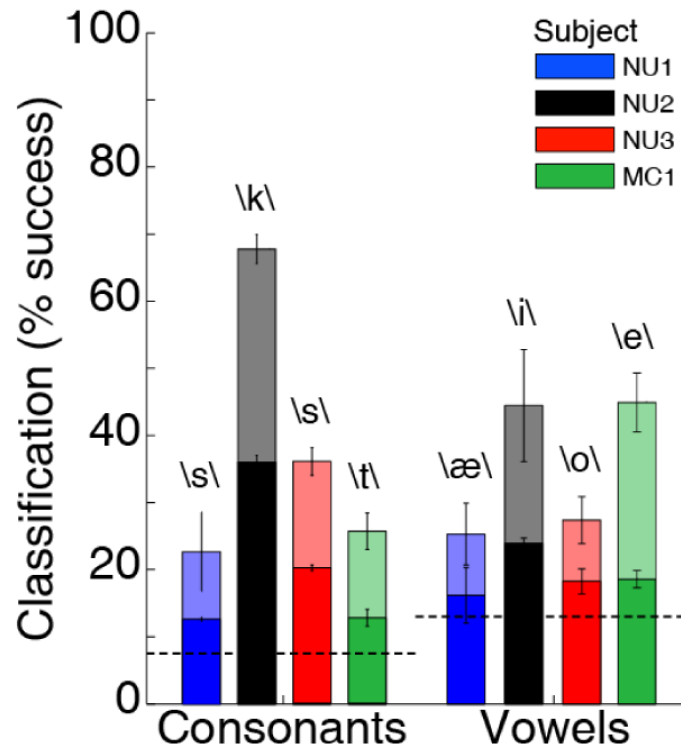


Figure 3. Classification results for phonemes for each subject. Chance percentage for each category is highlighted by a dotted line. Dark bars indicate performance across all phonemes for each subject; shaded bars indicate best performance of a single phoneme for each subject.

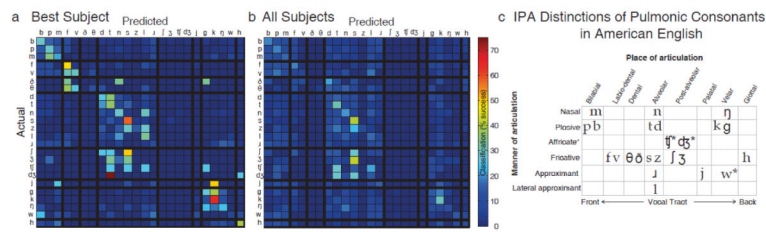


Figure 4.

Confusion matrices of decoding results of consonant phonemes for the (a) best performing subject (NU2) and (b) an average across all subjects, with respect to probability of that phoneme occurring within the data set. Phonemes are grouped by place of articulation, one axis of the IPA pulmonic consonants chart [31], with thick black lines dividing each articulation location. Decoding errors most often misclassified phonemes as neighboring phonemes according to IPA designation. (c) A condensed chart of the pulmonic consonants of the International Phonetic Alphabet used in American English. Asterisks denote non-pulmonic consonants.

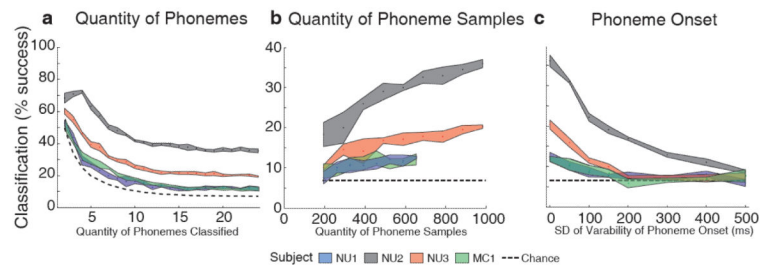


Figure 5. Dependence of classification performance on (a) quantity of phonemes classified, (b) total quantity of phoneme samples, and (c) accuracy of phoneme onset determination. Traces correspond to classification results for each subject, plotting the mean and standard deviation for 5 repetitions of 10-fold cross-validated classification.

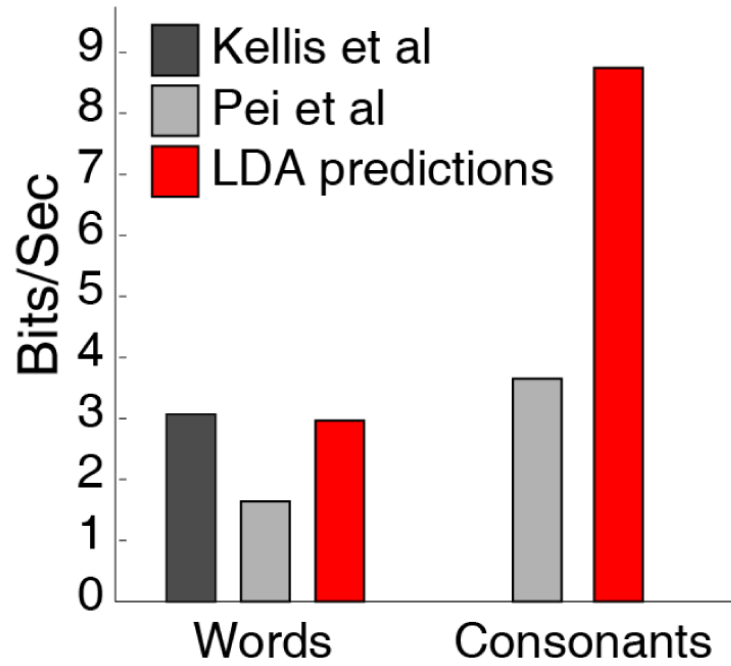


Figure 6. Information transfer rate for whole words and for consonant phonemes, with corresponding estimates from other studies of overt speech decoding using ECoG.