



HAL
open science

Electrocorticographic Representations of Segmental Features in Continuous Speech

Fabien Lotte, Jonathan Brumberg, Peter Brunner, Aysegul Gunduz, Anthony L.
Ritaccio, Cuntai Guan, Gerwin Schalk

► **To cite this version:**

Fabien Lotte, Jonathan Brumberg, Peter Brunner, Aysegul Gunduz, Anthony L. Ritaccio, et al.. Electrocorticographic Representations of Segmental Features in Continuous Speech. *Frontiers in Human Neuroscience*, 2015, 9 (97), pp.27. <10.3389/fnhum.2015.00097>. <hal-01159163>

HAL Id: hal-01159163

<https://inria.hal.science/hal-01159163v1>

Submitted on 2 Jun 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

Electrocorticographic Representations of Segmental Features in Continuous Speech

Fabien Lotte, Jonathan S Brumberg, Peter Brunner, Aysegül Gunduz, Anthony L Ritaccio, Cuntai Guan and Gerwin Schalk

Journal Name: Frontiers in Human Neuroscience
ISSN: 1662-5161
Article type: Original Research Article
Received on: 18 Sep 2014
Accepted on: 06 Feb 2015
Provisional PDF published on: 06 Feb 2015
Frontiers website link: www.frontiersin.org
Citation: Lotte F, Brumberg JS, Brunner P, Gunduz A, Ritaccio AL, Guan C and Schalk G(2015) Electrocorticographic Representations of Segmental Features in Continuous Speech. *Front. Hum. Neurosci.* 9:97. doi:10.3389/fnhum.2015.00097
Copyright statement: © 2015 Lotte, Brumberg, Brunner, Gunduz, Ritaccio, Guan and Schalk. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](http://creativecommons.org/licenses/by/2.0/). The use, distribution and reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

This Provisional PDF corresponds to the article as it appeared upon acceptance, after rigorous peer-review. Fully formatted PDF and full text (HTML) versions will be made available soon.



1

Electrocorticographic Representations of Segmental Features in Continuous Speech

Fabien Lotte^{1†} Jonathan S. Brumberg^{2*†} Peter Brunner^{3,4} Aysegul Gunduz⁵
Anthony L. Ritaccio⁴ Cuntai Guan⁶ and Gerwin Schalk^{3,4}

¹Inria Bordeaux Sud-Ouest/LaBRI, 2000 avenue de la vieille tour, 33405 Talence Cedex, France

²Department of Speech-Language-Hearing, University of Kansas, Lawrence, KS 66045, USA

³National Center for Adaptive Neurotechnologies, Wadsworth Center, Albany, NY 12201, USA

⁴Department of Neurology, Albany Medical College, Albany, NY 12208, USA

⁵J. Crayton Pruitt Family Dept. of Biomedical Engineering, University of Florida, Gainesville, FL 32611, USA

⁶Institute for Infocomm Research, A*STAR Agency for Science, Technology and Research, Singapore

Correspondence*:

Jonathan Brumberg

Department of Speech-Language-Hearing, University of Kansas, Lawrence, KS 66045, USA, brumberg@ku.edu

2 ABSTRACT

3 Acoustic speech output results from coordinated articulation of dozens of muscles, bones
4 and cartilages of the vocal mechanism. While we commonly take the fluency and speed of
5 our speech productions for granted, the neural mechanisms facilitating the requisite muscular
6 control are not completely understood. Previous neuroimaging and electrophysiology studies
7 of speech sensorimotor control has typically concentrated on speech sounds (i.e., phonemes,
8 syllables and words) in isolation; sentence-length investigations have largely been used to
9 inform coincident linguistic processing. In this study, we examined the neural representations
10 of segmental features (place and manner of articulation, and voicing status) in the context
11 of fluent, continuous speech production. We used recordings from the cortical surface
12 (electrocorticography (ECoG)) to simultaneously evaluate the spatial topography and temporal
13 dynamics of the neural correlates of speech articulation that may mediate the generation of
14 hypothesized gestural or articulatory scores. We found that the representation of place of
15 articulation involved broad networks of brain regions during all phases of speech production:
16 preparation, execution and monitoring. In contrast, manner of articulation and voicing status
17 were dominated by auditory cortical responses after speech had been initiated. These results
18 provide a new insight into the articulatory and auditory processes underlying speech production
19 in terms of their motor requirements and acoustic correlates.

20 **Keywords:** electrocorticography (ECoG); speech processing; place of articulation; manner of articulation; voicing

1 INTRODUCTION

21 Speech and language are realized as acoustic outputs of an aeromechanical system that is coordinated
22 by a vast brain and muscular network. The interaction between neural structures, facial and vocal
23 tract musculature, and respiration provides humans with a dynamic speech production system capable
24 of forming simple sounds (e.g., mono-syllabic words) and complex sounds (e.g., fluent conversation).
25 These sounds are often represented by phonemes and syllables, which are fundamental linguistic bases
26 for constructing both simple and complex speech production (e.g., the ‘b’ in ‘bad’ is an example of a
27 phoneme while the ‘ba’ is an example of a consonant-vowel (CV) syllable), which in turn correspond
28 to stereotyped vocal-tract movements resulting in acoustic speech output. Examples of such vocal-
29 motor articulations range from the compression of the lungs for producing the air pressure needed for
30 vocalization, to movements of laryngeal muscles during phonation, to configurations of the upper vocal
31 tract for final shaping of speech output. These muscular actions are the behavioral consequences of the
32 speech neuromotor system, which is in turn driven by phonological constructs and lexical relationships
33 [1].

34 This type of communication relies on neural processes that construct messages and sensorimotor
35 commands to convey and receive communicative information. These processes have previously been
36 characterized in a theoretical neurolinguistic model, the Levelt-Roelofs-Meyer (LRM) model [2]. Using
37 this model as a framework, it is possible to investigate the behavioral, neurological, linguistic and motor
38 processes involved in vocal communication. The model consists of the following processing components:
39 conceptual preparation, lexical selection, morpho-phonological code retrieval, phonological encoding,
40 phonetic encoding and articulation [1, 2]. The first four processing levels in the LRM framework all
41 mediate perceptual processes underlying speech and language recognition in preparation for upcoming
42 vocal productions (e.g., reading, picture naming). These levels of processing have been well investigated
43 and were summarized in a meta-analysis of neuroimaging, electrophysiology and neuro-stimulation
44 studies of speech and language [1], and more recently by [3]. The final two stages, phonetic encoding
45 and articulation (of articulatory scores), describe the motor aspects of vocal communication and are the
46 focus of the present study. According to the LRM framework, the phonetic encoding stage translates a
47 phonological word (from the previous phonological encoding stage) into an articulatory score, which can
48 be processed and transmitted to the articulatory musculature for speech motor output.

49 The precise nature by which the brain realizes these phonetic encoding and articulation functions
50 are still unknown. One possible explanation for this lack of understanding stems from the difficulty in
51 measuring the neurological processes involved in the planning and production of speech. Indefrey and
52 Levelt estimate a total speech-language processing time of approximately only 600 ms (not including
53 articulation) from beginning to end, with individual durations of approximately 100-200 ms for each
54 processing component in their model [1]. Functional magnetic resonance imaging (fMRI), which is
55 the primary neuroimaging technique used in speech neuroscience, cannot resolve brain activity at that
56 temporal resolution. In contrast, electroencephalography (EEG) and magnetoencephalography (MEG) can
57 readily detect neurological signals at these temporal scales, but cannot precisely ascribe their source to a
58 particular location. In addition, EEG, MEG and fMRI are all susceptible to electrical and/or movement
59 artifacts created by speech articulation, and thus are typically used to investigate neurological activity
60 prior to articulation or speech perception.

61 Electrical signals recorded directly from the cortical surface (electrocorticography (ECoG)) have
62 recently begun to attract increasing attention for basic and translational neuroscience research, because
63 they allow for examination of the precise spatio-temporal evolution of neurological processes associated
64 with complex behaviors, including speech output. Specifically, ECoG has been used to investigate
65 neurological activity during a number of tasks including linguistic processing [4, 5], speech perception
66 and feedback processing [6, 7, 8, 9, 10], as well as articulation of phonemes, syllables, and words
67 [3, 11, 12, 13, 14, 15]. In the present study, we apply machine learning techniques to evaluate
68 the neurological activity during speech production based on segmental features (i.e., phonology, and
69 articulatory-acoustic descriptors) and the resulting ECoG signals. By analyzing these features, rather

70 than phonemes, syllables, or words, we are able to identify a low-dimensional and invariant basis by
71 which to interpret neural activity related to overt speech production that can be upscaled to more complex
72 vocalizations.

73 A recent ECoG study [14] employed such an articulation-based approach in which subjects were
74 required to produce isolated CV syllables. The authors observed both a topographic and temporal
75 organization of ECoG signals over the speech-motor cortex related to speech articulation. Specifically,
76 their results showed that the production of isolated syllables resulted in differential neurological activity
77 clustered by articulatory feature (e.g., lip and tongue movements). These findings greatly contributed
78 to our understanding of the motor cortical representations of isolated syllable production; however, in
79 typical speech, syllable production is rarely performed in isolation. Here, we generalize and improve
80 upon these results by investigating articulation as it occurs during continuous, fluent speech. One major
81 difference between isolated production of speech sounds and continuous speech is the presence and degree
82 of coarticulation, or the influence of past and future speech requirements over current productions [16].
83 The two varieties of coarticulation include: 1) carry-over, in which upcoming speech productions are
84 based on the vocal tract configurations of past utterances; and 2) anticipatory, in which the production
85 of current speech sounds is altered based on expected requirements of future sounds. The extent to
86 which segmental and phonological boundaries influence the degree of coarticulation [17] is currently
87 subject of debate (e.g., whether a boundary facilitates or inhibits coarticulation). In our study, we assume
88 coarticulation is occurring as participants produce speech, and our results are based solely on the amount
89 of speech information present in the ECoG signal.

90 In our experiments, we asked subjects to perform an out-loud speech production task. We recorded the
91 subjects' acoustic output with a microphone and ECoG from widespread perisylvian areas that included
92 locations with known involvement in the planning, execution and perception of speech. For each subject,
93 we then converted the subject's acoustic output into speech feature categories at the phonetic level (given
94 in Table 2) and applied machine learning techniques to identify differential brain activity resulting from the
95 production of specific speech features. The features used in our work were: place of articulation, manner
96 of articulation, voicing status and phonological category of consonant or vowel. These techniques allowed
97 us to investigate the topographical as well as temporal distributions of brain activity that differentiates each
98 type of speech feature amongst other features, which may temporally overlap in continuous speech. The
99 analysis techniques used in our study can also be used to predict the occurrence of a speech feature from
100 the ECoG signals. Therefore, our study provides important insights into the coordination of individual
101 articulatory neuromotor processes as they are sequenced together for production of fluent speech output,
102 and should provide an important basis for future development of a brain-to-text brain-computer interface
103 (BCI).

104 The results of our analyses revealed a broad network involving fronto-motor and temporal cortices
105 that were active during the preparation, execution and feedback monitoring of place of articulation. In
106 contrast, ECoG responses labelled by manner of articulation involved a widespread auditory cortical
107 network that was active near the start of speech onset and persisting throughout the feedback monitoring
108 process. Analysis of voicing status largely mirrored the manner of articulation results suggesting that the
109 production of different manners of articulation and voicing involve large auditory cortical networks for
110 processing for proper speech motor control, while place of articulation more equally weights processing
111 at all three stages of production. Interestingly, our analysis of both the manner and voicing conditions
112 included a focal motor response that likely reflects specific differences in the motor control of voicing
113 (e.g., voiced vs voiceless production). We elaborate on these results and their interpretation in the sections
114 that follow.

Table 1. Clinical profiles of participants

Subject	Age	Sex	Handedness	Performance IQ	Verbal IQ	Seizure Focus	# electrodes	# words
A	29	F	R	136	118	Left temporal	96	278
B	30	M	R	90	64	Left temporal	83	109
C	29	F	R	90	91	Left temporal	101	283
D	19	M	R	85	87	Left frontal	84	411
E	26	F	R	117	106	Left temporal	109	411
F	56	M	R	87	82	Left temporal	97	411
G	29	F	R	95	111	Left temporal	112	411

2 MATERIAL AND METHODS

2.1 HUMAN SUBJECTS AND DATA COLLECTION

115 The seven subjects who participated in this study were patients with intractable epilepsy at Albany
 116 Medical Center. Subjects underwent temporary placement of subdural electrode arrays to localize seizure
 117 foci prior to surgical resection of epileptic tissue. All gave informed consent to participate in the study,
 118 which was approved by the Institutional Review Boards of the hospital, had performance IQs of at least
 119 85, and were mentally, visually and physically capable of performing the task. Table 1 summarizes the
 120 subjects' clinical profiles.

121 The implanted electrode grids (Ad-Tech Medical Corp., Racine, WI) consisted of platinum-iridium
 122 electrodes (4 mm in diameter, 2.3 mm exposed) that were embedded in silicon and spaced at an inter-
 123 electrode distance of 1 cm. Subject G had implanted electrodes with 6 mm grid spacing (PMT Corp,
 124 Chanhassen, MN). All subjects received electrode grid implantations over the left hemisphere, though the
 125 total number of electrodes implanted was different for each subject. Grid placement and duration of ECoG
 126 monitoring were based solely on the requirements of the clinical evaluation without any consideration of
 127 this study.

128 Grid locations were verified in each subject using a co-registration method that included pre-operative
 129 structural magnetic resonance (MR) imaging and post-operative computed tomography (CT) imaging
 130 [18]. We then used Curry software (Compumedics, Charlotte, NC) to extract three-dimensional cortical
 131 models of individual subjects, to co-register the MR and CT images, and to extract electrode locations.
 132 Electrode locations are shown for each subject in Figure 1. Electrode locations were further assigned to
 133 cortical lobe using the Talairach Daemon (<http://www.talairach.org>, [19]).

134 ECoG signals were recorded at the bedside using eight 16-channel g.USBamp biosignal acquisition
 135 devices (g.tec, Graz, Austria) at a sampling rate of 1200 Hz, and stored for further analyses. Electrode
 136 contacts distant from epileptic foci and areas of interest were used for reference and ground and any
 137 channels with obvious electrical or mechanical artifacts removed. The total number of electrodes used per
 138 subject is listed in Table 1.

2.2 EXPERIMENTAL PARADIGM

139 In this study, subjects were asked to perform an overt speech production task in which stimuli consisted
 140 of well-known political speeches or nursery rhymes ranging between 109 and 411 words in length. The
 141 stimulus text was presented visually and scrolled across a computer screen from the right to the left at a
 142 constant rate and subjects repeated each word as it appeared on the screen. The rate was set for each subject
 143 to be appropriate for the subject's level of attentiveness, cognitive, and comprehension abilities (see Table
 144 1). The computer screen was placed approximately 1 m from the subjects. A single experimental run
 145 consisted of reading an entire stimulus passage, and subjects completed between 2–4 runs. All subjects
 146 completed the experiment in a single session except for Subject D, who required two sessions. Data
 147 collection from the g.USBamp acquisition devices, as well as control of the experimental paradigm were

148 accomplished simultaneously using BCI2000 software [20, 21]. A schematic illustrating the experimental
149 setup is shown in Figure 2).

2.3 SIGNAL PROCESSING AND ANALYSIS

150 The goal of our study was to identify those locations or times in which differential ECoG activity
151 was found between overtly produced speech utterances based on articulatory-acoustic and phonological
152 features (e.g., segmental features) of phonemes¹. In this work, we used a vowel versus consonant contrast
153 as the primary phonological discriminatory dimension. In addition, we examined the articulatory-acoustic
154 dimension by testing the manner (e.g., voicing quality: obstruent vs. sonorant) and the place (e.g., location
155 of articulatory closure or constriction) of speech articulation, and voicing (e.g., quasiperiodic oscillations
156 of the vocal folds: voiced vs. voiceless). The place features are primarily used to characterize consonant
157 sounds, while the manner and voicing features can be used in both consonant and vowel descriptions. We
158 conducted an analysis of a feature representing the tongue configurations involved in the production of
159 vowel sounds (e.g., height & frontness within the oral cavity); however, it did not reveal different patterns
160 of spatiotemporal activations and will not be discussed in subsequent sections.

161 *2.3.1 Articulatory-acoustic feature descriptions* The articulatory features used in the present study
162 generally characterize the vocal tract movements and configurations required for speech production. The
163 place of articulation defines a location where speech articulators either close or constrict the vocal tract. In
164 our analysis, high-level descriptions of place of articulation broadly describe the closure of the lips (labial)
165 and the location in the oral cavity where the tongue contacts or approaches the hard and soft palates
166 (coronal and dorsal) [23]. The manner of articulation describes the relative closure of the vocal tract and
167 resultant airflow path during phonation; it can be coarsely grouped into obstruents (those articulations
168 that impede airflow in the vocal tract) and sonorants (those which maintain an open vocal tract) [23]. The
169 voicing feature indicates whether the vocal folds are active and oscillating during production of speech
170 sounds. Speech sounds are classified as “voiced” if the vocal folds are oscillating and “voiceless” if they
171 are not. All sonorant sounds, including all vowels, in English are considered voiced (with only a few
172 exceptions) while obstruents have voiced and voiceless pairs (e.g., the bilabial pair ‘b’ [voiced] and ‘p’
173 [voiceless]).

174 Both the place and manner of articulation can be specified at increasingly refined levels. For place,
175 some examples of the labial feature includes bilabials (‘b’) and labiodentals (‘v’), an example of a coronal
176 includes alveolars (‘d’ in “dog”) and palatals, and the dorsal group includes consonants with contact on
177 the velum or soft palate (‘g’ in “good”). Additionally, the dorsal group can be used to describe the relative
178 movements of all the vowels, though not their specific configurations. These additional place descriptors
179 can further refine the locations of the hard and soft palates contacted by the tongue and vice versa as well
180 (e.g., they describe the portions of the tongue used to contact the palate). The manner of articulation can
181 also be described with finer levels of detail, with examples of the obstruent category including features for
182 stops (‘b’ in “boy”), fricatives (‘v’ in “vast”), and affricates (‘ch’ in “chest”) while the sonorant category
183 contains the features for approximants (‘l’ in “less”) and nasals (‘n’ in “nine”). These additional levels
184 of description characterize specific differences in airflow resulting from speech production. To simplify
185 the analysis and provide sufficient data for estimation of our machine learning models, we concentrated
186 on the high-level categorical groupings: obstruent versus sonorant for manner of articulation, and labial
187 versus coronal versus dorsal for place of articulation. A summary of the phonetic feature descriptions used
188 in this study can be found in Table 2.

189 *2.3.2 Speech segmentation into phonemes* We first segmented the acoustic speech signals into
190 individual phonemes. This segmentation served to (1) separate each individual spoken word and (2)

¹ As defined by the International Phonetic Association, a phoneme is “the smallest segmental unit of sound employed to form meaningful contrasts between utterances” [22]

Table 2. Features and frequencies observed in the speech stimuli.

Place of articulation			Manner of articulation		
Feature	Frequency	phonemes	Feature	Frequency	phonemes
Labial	22.9%	/b p f v m w/	Obstruent	59.1%	/b p g k d t f v/ /tʃ dʒ ð θ s z ʃ ʒ/
Coronal	78.0%	/t d θ ð s z ʃ ʒ n/ /tʃ dʒ r l j/	Sonorant	37.5%	/i ɪ ɛ æ ɑ ə u ʊ/ /ɜ˞ ɑɪ eɪ aʊ oʊ ɔɪ/ /w j r l m n ŋ/
Dorsal	12.4%	/k g w/			
Phonological			Voicing		
Feature	Frequency	#	Feature	Frequency	phonemes
Consonant	60.8%	24	Voiced	78.0%	/i ɪ ɛ æ ɑ ə u ʊ/ /ɜ˞ ɑɪ eɪ aʊ oʊ ɔɪ/ /w j r l m n ŋ/ /b d g v ʃ ʒ dʒ/
Vowel	39.2%	15	Voiceless	22.0%	/p t k f s z tʃ/

191 identify and temporally locate phonemes within each word. Our segmentation procedure obtained
 192 phonetic transcriptions using a semi-automated algorithm that first isolated the spoken words from silence
 193 followed by identification of constituent phonemes. The onset and termination of spoken words were
 194 manually located in the audio signal waveforms. Initial manual segmentation of word boundaries was
 195 necessary for accurate speech analysis, and was often completed with minimal effort. Following word
 196 segmentation, phonemes were automatically labeled and aligned to the audio signal, using a Hidden
 197 Markov Model (HMM) classifier with Mel-Frequency Cepstrum Coefficients (MFCC) and their first and
 198 second derivatives as features [24]. The phonetic transcription and alignment was performed using the
 199 HMM ToolKit (HTK) [25]. Our rationale for automated phonetic transcription was to minimize human
 200 errors and provide an objective solution for a fair comparison between participants. Each phoneme was
 201 then classified as (1) a consonant or a vowel [phonological], (2) an obstruent or sonorant [manner], (3)
 202 according to vocal tract contacts or constrictions [place] and (4) voicing status [voicing].

203 A summary of all phoneme transcriptions and data features used in this study is provided in Table 2.
 204 Each speech feature was assigned in a binary fashion in which ‘+’ indicated the presence of a feature,
 205 and ‘-’ the absence. Importantly, while the features were coded as binary, any one phoneme may code for
 206 multiple combinations of features (e.g., consonant+, obstruent+, labial+ and voicing+ for the ‘b’ sound).
 207 In other words, a particular phonemic feature was assigned a value (‘+’ or ‘-’) for each phoneme. Overall,
 208 we identified 33 different phonemes with 1226 – 4872 combined occurrences per subject. Each phoneme
 209 was defined by a particular onset and offset time that was used for subsequent neurophysiological analyses.
 210 An example of audio signal transcription and feature labeling (for the feature: vowel) is given in Figure 3
 211 along with synchronized ECoG recordings (gamma band power) at two electrode sites.

212 The automatic speech recognition system described above was adapted from the original implementation
 213 to achieve robust and accurate speaker-dependent classification for use with all of our study participants.
 214 The classifier was first trained on an “ideal” source based on a triphone acoustic model to establish a
 215 baseline. Then, the classifier was adapted to account for each participant’s individual speech acoustic
 216 characteristics using the speech recorded from each subject, creating a speaker-dependent recognition
 217 and phonetic transcription system. The speaker-dependent model outperformed the speaker-independent

218 model in terms of producing more accurate phoneme boundaries. All automatic phoneme alignments were
219 visually checked by a speech recognition expert who confirmed their quality.

220 *2.3.3 ECoG segments extraction and labeling* We analyzed event-related changes in 700 ms ECoG
221 epochs aligned to phoneme acoustic onset. To do this, we first high-pass filtered the continuous ECoG
222 recordings using a cutoff frequency of 0.5 Hz and a forward-backward Butterworth filter of order 4 to
223 remove DC signal components (Matlab functions `filtfilt` and `butter`). The data were then notch-
224 filtered at 120 Hz using a forward-backward infinite impulse response (IIR) notch filter with a Q-factor of
225 35 ($q = \omega_0/bw$, where $\omega_0 = 120$ and $q = 35$) [26] to remove the power line harmonics (first harmonic)
226 interference. Note that we did not filter the signals at the fundamental frequency of the power line (60
227 Hz) nor its other harmonics (180 Hz, 240 Hz, etc.) since our analysis only involved the gamma band (70–
228 170 Hz) of the ECoG signals. Following filtering, the ECoG signals were re-referenced to the common
229 average reference (CAR), separately for each grid of implanted electrodes.² Finally, the ECoG gamma
230 band power was obtained by applying a bandpass filter in the range of 70–170 Hz using a fourth order
231 forward-backward Butterworth filter, squaring the result and log-transforming the signal.

232 After preprocessing the recorded ECoG signals, we extracted a 700 ms window of data from the
233 continuous recording. This window was aligned to the onset of each phoneme identified by the semi-
234 automated phoneme transcription procedure described above. Each window was centered on the phoneme
235 onset, and thus consisted of a 350 ms pre-phoneme interval and a 350 ms post-phoneme interval, which
236 provides sufficient opportunity to examine the neurological processing per phoneme. Each window was
237 tagged with the phoneme's feature vector (i.e., '+' or '-' definition for each phonemic feature) for
238 subsequent classification / discrimination analysis.

239 *2.3.4 Classification analysis technique* In the following sections, we describe the method used
240 to evaluate the spatial and temporal patterns of neurological activity involved in speech production.
241 Specifically, we employed a classification analysis to determine which brain regions differ in their patterns
242 of activity during the production of speech that varies by place of articulation, manner of articulation,
243 voicing and phonological category of consonant or vowel (Section 2.3.6). We include also a classification
244 analysis of brain activity during active speaking versus silence (Section 2.3.5). The same procedure was
245 used for all classification analyses, and is summarized as follows:

- 246 1. Process and segment speech signal for features of interest (e.g., speech vs silence, place, manner and
247 voicing features, phonological features)
- 248 2. Preprocess ECoG gamma band power (as in Section 2.3.3)
- 249 3. Choose analysis features based on the number of ECoG electrodes, and reduce feature dimensionality
250 according to the minimal Redundancy Maximal Relevance (mRMR) feature selection procedure [27].
- 251 4. Train and apply a regularized linear discriminant analysis (LDA) classifier [28] for distinguishing
252 selected features using 5 fold cross-validation for each subject and run. Note that feature selection
253 was performed, for each fold of the cross-validation, on the training data only.
- 254 5. Evaluate classifier using receiver operating characteristics (ROC) curves, and obtain the area under
255 the curve (AUC) as the primary performance measure.

256 LDA regularization was achieved using covariance matrix shrinkage according to the Ledoit and Wolf
257 method for automatically estimating large dimensional covariance matrices from small data observations
258 [29]. Regularized LDA using this technique has been previously used in brain-machine interfacing
259 experiments where data and feature dimensionality are consistently problematic [30, 31]. According to
260 our cross-validation procedure, the data were split into five non-overlapping subsets, four of which were

² Most subjects had more than one ECoG grid implanted; therefore, the electrodes from each grid were re-referenced to the grid average.

261 used for LDA training and feature selection and the remaining, mutually-exclusive data set, used for
262 testing. The training and testing procedures were repeated five times, once for each mutually exclusive
263 validation set, and the performance was averaged over all test-set results. Note, classifier training and
264 feature selection were performed only on the training part of each cross-validation fold.

265 Additionally, we chose area under the ROC curve as the measure of performance since it is specifically
266 designed for unbalanced binary classification problems [32]. In our study, the number of phonemes labeled
267 '+' for a speech feature was not necessarily the same as the number of phonemes labeled '-', therefore
268 the classification problem was unbalanced. The '+' class was used as the positive class for ROC curves
269 computation. Statistical significance of the obtained AUC values was determined using the Hanley and
270 McNeil formula for estimating standard error [33]. The resulting p -value was then corrected for multiple
271 comparisons (number of subjects \times number of ECoG electrodes per subject) using the false discovery
272 rate (FDR) approach [34].

273 *2.3.5 Subject screening and inclusion* As a screening measure, we first determined which of the
274 subjects produced ECoG signals that were different between spoken words and silence. Subjects whose
275 classification results exceeded our threshold (see below for details) were analyzed further for the speech
276 feature analysis. According to the classification procedure described in Section 2.3.4, we first manually
277 obtained the boundaries of all words from the acoustic signal and extracted ECoG gamma band power
278 from a 700 ms window centered on each word. We then obtained an equal number of ECoG segments
279 taken from 700 ms windows of silence and labeled the segments as "speech" or "silence." For each
280 electrode, the pre-processed 700ms ECoG signal was segmented in time using 50 ms long windows with
281 25 ms overlap based on the parameters from prior studies [8, 9]. This procedure resulted in an initial set of
282 27 gamma-band features per electrode (between 83–112 electrodes per subject), which were taken from
283 cortical areas covering the perisylvian and Rolandic cortices (e.g., primary motor, premotor, auditory and
284 somatosensory cortices; Broca's and Wernicke's areas). We then used the mRMR procedure to reduce the
285 feature dimension by selecting 50 features from the larger data set. Last, we obtained the ROC curve and
286 set a threshold of $AUC > 0.8$ for inclusion in the remainder of the speech feature analysis. An AUC
287 of 0.5 represents chance performance, we therefore utilized a higher threshold for use as a screening
288 criterion.

289 *2.3.6 Classification of articulatory features* Determination of the differential neurological activity
290 used in the production of each articulatory-acoustic and phonological features (described in Section 2.3.1)
291 was split into separate analyses of spatial topography and temporal dynamics. In the spatial topography
292 analysis, we projected the results onto the cortical surface and plotted the results over time for the temporal
293 dynamics analysis. In these two procedures, the spatial analysis considered ECoG activity at each location
294 throughout each windowed epoch; the temporal analysis considered ECoG activity at a particular time but
295 across all locations.

296 **Spatial topography analysis.** Using the classification procedure described in Section 2.3.4 as a guide,
297 we first obtained the boundaries of all phonemes in the acoustic signal (see Section 2.3.2), extracted the
298 ECoG gamma band power from a 700 ms window centered on the onset of each phoneme, and segmented
299 it in time using 50 ms long windows (25 ms overlap). We then used the mRMR procedure to select
300 10 time segments per phoneme and electrode to minimize the effects of overfitting while training the
301 regularized LDA classifier. A new classifier was trained on each of the speech features to discriminate
302 between the '+' and '-' category members. To analyze the three levels place of articulation features, we
303 computed three binary comparisons: labial+ vs labial-, coronal+ vs coronal- and dorsal+ vs dorsal-. All
304 other features contained only two levels, therefore, only a single binary comparison is needed for each.
305 We then computed an "activation index" that was proportional to the AUC p -value for each tested feature.
306 The activation index (AI) was defined as:

Table 3. AUC cross-validation performances obtained for each subject to classify “spoken word” versus “silence” ECoG segments.

Subject	A	B	C	D	E	F	G
AUC	0.57	0.81	0.51	0.68	0.91	0.87	0.91

$$\psi(p) = \begin{cases} -\log(p) & p < 0.01 \\ 0 & otherwise \end{cases} \quad (1)$$

307 where \log denotes the natural logarithm. These activation indices for each electrode channel were
 308 accumulated across subjects and mapped onto a template brain (Montreal Neurological Institute [MNI];
 309 <http://www.bic.mni.mcgill.ca>) using in-house Matlab routines [18].

310 **Temporal dynamics analysis.** The temporal analysis of speech features over the duration of each data
 311 segment involved similar processing steps used in the spatial analysis. For each subject, we first limited
 312 the temporal analysis to ECoG electrode channels with statistically significant activation indices found
 313 in the spatial topography analysis. For this analysis, we first re-estimated the ECoG gamma band power
 314 using 50 ms time bins, but with 40 ms overlaps (10 ms steps) for use in the LDA procedure. The change in
 315 overlap was used to visualize and analyze the activation index time course with a higher resolution, such
 316 resolution is neither needed nor desired for the spatial topography analysis. The same speech features and
 317 phonetic boundaries used in the spatial analysis were used here as well. Also in this analysis, dimension
 318 reduction and regularization were not required since there was only one data feature (time-binned ECoG
 319 band power) per classification attempt. The average AUC was then used to compute a significance p -
 320 value, corrected for multiple comparisons (subjects, time bins and electrodes with statistically significant
 321 activation indices in the spatial topography analysis) using the FDR, and transformed into an activation
 322 index. The temporal profiles of the activation indices were averaged across subjects and over all electrodes
 323 per speech feature to represent the gross cortical processes involved in the discrimination of speech
 324 articulation features.

3 RESULTS

3.1 SPEECH VERSUS SILENCE

325 We employed a functional screening criteria based on classification results for a speech versus silence
 326 discrimination analysis. These results are summarized in Table 3. Those subjects that did not have neural
 327 responses that consistently responded to the task, and thus had signals that could differentiate between
 328 speech and silence, were excluded from the remainder of the speech feature analysis. Recall, an AUC
 329 value of 0.5 represents chance discrimination, while a value of 1.0 indicates perfect discrimination.
 330 None of our analyses resulted in AUC values less than 0.5, indicating that all classifications were above
 331 chance levels. However, as illustrated in Table 3, our analysis was not able to well-differentiate the
 332 neural activation patterns for the speech versus silence contrast for subjects A, C and D using our higher
 333 screening threshold ($AUC < 0.8$), which would lead to similarly poor results in any subsequent analyses
 334 of articulatory and phonological features. In contrast, the analysis for subjects B, E, F and G resulted in
 335 relatively good differentiation between speech and silence ($AUC > 0.8$). Thus, we included only data
 336 from these subjects in the remainder of our study. The resulting combined electrode locations for these
 337 four subjects can be found in the bottom right of Figure 1.

3.2 CORTICAL MAPPINGS AND TEMPORAL PROFILES

338 Topographical cortical mappings and temporal profiles reported here reflect electrodes, grouped over all
 339 four subjects, with statistically significant differences in ECoG recordings between our speech features
 340 of interest. In our method, each discrimination is along a binary feature dimension and represents a
 341 comparison of neural patterns of activation between pairs of speech features.

342 *3.2.1 Place of articulation* We analyzed ECoG recordings to identify differential neural activity for
 343 three place of articulatory features: labial, coronal, and dorsal representing vocal-tract closures at the lips
 344 (labial), tongue tip & blade (coronal), and tongue dorsum (dorsal). We then used statistically significant,
 345 above-chance LDA classifications ($AUC > 0.5$) as a measure of differential neurological representations
 346 of each speech feature. We generally found statistically significant responses across the sensorimotor
 347 speech production network and auditory feedback processing regions (see left column Figure 4). The
 348 responses superior to the Sylvian fissure are distributed over the primary motor and somatosensory cortices
 349 (sensorimotor cortex for speech), while the responses in the temporal lobe are found in perisylvian
 350 auditory cortex, particularly in the posterior aspects of the superior temporal gyrus (e.g., Wernicke's
 351 area). The coronal feature resulted in the largest topographical montage of statistically significant ECoG
 352 electrodes contributing to differentiation of place of articulation (N=19 of 401 electrodes), followed by
 353 the labial (N=9) and dorsal (N=3) features. A summary of these results is found in Table 4.

Table 4. Summary of results for place of articulation over all sampled electrodes

Place	# electrodes	peak AI	peak latency	local maxima
Labial	9	21.24	25 ms	-185 ms, -75 ms, +105 ms
Coronal	19	13.42	35 ms	-195 ms, -85 ms, +95 ms
Dorsal	3	6.25	45 ms	-165 ms, +115 ms

354 In the temporal dimension group analysis, we found the latency of peak AI for all three place conditions
 355 near the onset of phoneme alignment at 0 ms (see Figure 4, right column and summarized in Table 4).
 356 Specifically, the labial condition is characterized by an overall difference from all other features that rose
 357 markedly to a peak response at +25 ms (with 21.24 peak activation index) and persisted well afterward.
 358 The peak activation index for the coronal condition was 13.42 at +35 ms latency and the dorsal condition
 359 was 6.25 at +45 ms latency. In general, both the labial and coronal temporal profiles indicated prolonged
 360 duration of statistically significant activation indices preceding and following peak response near 25–35
 361 ms while the dorsal condition was much more narrow in its response. We should note that this may be
 362 due to the relatively few sounds with constriction or closure of the tongue along soft palate compared
 363 to those in the anterior portions of the oral cavity. Furthermore, each of the three place conditions had
 364 multiple local maxima throughout the analysis window. Specifically, local maxima were found for the
 365 labial condition at -185 ms, -75 ms and +100 ms, the coronal condition at -195 ms, -85 ms and +100 ms,
 366 and the dorsal condition at -165 ms and +115 ms.

367 *3.2.2 Manner of articulation* The analysis of place of articulation is oriented toward the articulations
 368 and points-of-contact in the oral portion of the upper vocal tract. In contrast, manner of articulation, which
 369 describes airflow resulting from constriction or closure (release) is oriented generally as the muscular
 370 activation of the entire upper vocal tract (larynx, velum and oral structures). In typical definitions of
 371 manner of articulation, categorical features are used to describe the overall airflow. In the present analysis,
 372 we follow this convention and examined two main classes of manner: obstruents and sonorants.

373 The spatial topography of electrodes with differential activity patterns between the two manner
 374 categories are shown in Figure 5, left column. This analysis revealed statistically significant perisylvian
 375 auditory cortex and sensorimotor cortex response contributing to differentiation of the obstruent (N=10
 376 electrodes) and sonorant (N=11 electrodes) features.

377 The temporal profile results (right column, Figure 5) indicate very limited differences between
 378 manner categories prior to phoneme onset (speech-leading latencies with negative intervals) and greater
 379 differences at speech-following latencies (positive intervals). Specifically, the peak statistical significance
 380 for differentiating manner features from each other at +85 ms for both obstruents and sonorants. These
 381 differences are largely present during the entire post-onset speech period. These results are summarized
 382 in Table 5.

Table 5. Summary of results for manner of articulation over all sampled electrodes

Manner	# electrodes	peak AI	peak latency
Obstruent	10	18.26	85 ms
Sonorant	11	28.58	85 ms

383 **3.2.3 Voicing** In contrast to both the manner and place features, voicing refers to only one articulatory
 384 structure, the larynx, or more specifically, the vocal folds. The spatial topography of electrodes (left
 385 column, Figure 6) with differential patterns of activity between the voiced and voiceless classes of
 386 phonemes is concentrated in the perisylvian auditory and motor cortex, with additional activation of the
 387 ventral motor cortex. In our analysis, 12 electrodes contributed to differentiation of phonemes along the
 388 voicing dimension. The temporal profile of these activations (right column, Figure 6) indicate a peak
 389 statistical difference at +95 ms with an activation index of 11.45. There was a smaller local peak just prior
 390 to vocalization onset at -25 ms. These results are summarized in Table 6.

Table 6. Summary of results for voicing and phonological category (vowels only) over all sampled electrodes

Manner	# electrodes	peak AI	peak latency	local maxima
Voicing	12	11.45	95 ms	-25 ms
Vowels	8	21.80	95 ms	-105 ms

391 **3.2.4 Vowel versus consonant** We examined the vowel versus consonant contrast to determine whether
 392 differences existed in neural activation patterns between production of sounds varying in phonological
 393 class. The spatial topography and temporal dynamics representing differences between these two classes
 394 were represented by a large region of auditory cortex and a more focal region of sensorimotor cortex.
 395 The temporal patterns of neural activation had peak statistical difference +95 ms, but appear to also show
 396 moderate differentiation at speech-leading intervals, with a local maxima at -105 ms as shown in Figure 6
 397 (right column). A summary of these results can be found in Table 6.

4 DISCUSSION

4.1 GENERAL COMMENTS

398 In this paper, we identified patterns of cortical topographies and temporal dynamics involved in speech
 399 production based on segmental articulatory-acoustic and phonological characteristics. To do this, we
 400 used a classification analysis to identify spatial or temporal neurological activity that best discriminated
 401 between common sets of articulatory and phonological features of continuous speech production. Some
 402 recent studies of speech production using ECoG and intracortical microelectrode recordings have also
 403 examined phonetic content [11], and articulatory-acoustic features [14, 35]. Importantly, our task and
 404 analyses differ from these earlier attempts by first considering fluent, continuous speech production of
 405 whole sentences and paragraphs, which is more natural than isolated utterances and may account for
 406 effects of coarticulation. Second, our signal recordings come from a much larger area of the cortical

407 surface, which enabled us to investigate all of the lateral (perisylvian) regions involved in the motor,
408 perceptual and planning neurological processing components of speech production. Last, our analysis
409 focuses on the determination of the neurological activity that differentiates speech segments (e.g.,
410 phonemes) from one another based on their phonological and articulatory features.

411 The continuous speaking task is doubly advantageous as it allows for acquisition of a large amount of
412 phoneme data in a short amount of time, which is imperative when interacting with patients with an ECoG
413 implant. We are also able to analyze simultaneously overlapping processes of phonological processing,
414 execution of articulatory plans and monitoring of acoustic feedback in a manner. Our technique of
415 machine learning classification for discrimination of speech features via ECoG recordings enable direct
416 inference of the neurological structures and dynamics that dissociate production of phonemes with varying
417 phonological and articulatory characteristics. We discuss the major implications of our results along these
418 themes in the following sections. In general, the neurological structures and dynamics revealed in our
419 study overlapped with many of our expectations [14, 36], but our specific analyses identified some striking
420 differences from prior work.

4.2 MOTOR AND SENSORY PROCESSING

421 Speech articulation is composed of at least two “first-order” processes: motor control and sensory (i.e.,
422 acoustic) feedback, whose functionality is typically reflected by neural activation of the precentral gyrus
423 and superior temporal gyrus, respectively. Though both types of processes are certainly involved in speech
424 production, the relative timing of neural activations, before or following speech, can help to determine
425 whether processing is related to planning and execution of speech sounds (speech-leading) or feedback
426 maintenance (speech-lagging).

427 The design of our analysis procedures allowed us to simultaneously analyze neural recordings of
428 continuous speech production from two separate perspectives. In the *place* and *voicing* analyses, we
429 examine the contribution of neural signals to specific articulatory gestures (just the larynx in the case of
430 voicing), while in the *manner* analysis, the motor response is not differentiated. Without examining both,
431 we would have limited the explanatory potential of the recorded data and miss the observation of a dual-
432 role played by sensory cortex (receptive cortex) in speech production. These results are described in more
433 detail in the following sections.

434 *4.2.1 Place of articulation* Place of articulation is easily interpreted along motor and somatosensory
435 dimensions. The placement of a vocal tract closure or constriction necessarily involves movement of the
436 speech articulators as well as tactile (for closure) and proprioceptive (for constriction) somatosensation.
437 In our analysis, we used the place features labial, coronal and dorsal for discriminating ECoG responses
438 as a result of speech articulation. The sensorimotor interpretation for the labial feature refers to closure
439 of the lips, either against each other (bilabial) or of the lower lip against the maxillary teeth (labiodental),
440 both result from the movement of the lip(s) and / or jaw. Similarly, the sensory interpretation for the
441 feature coronal refers to closures occurring between the tongue, maxilla and hard palate, while the motor
442 interpretation refers to muscular involvement of the tongue tip, tongue body and anterior portions of
443 the tongue body as they contact the teeth (dentals), alveolar ridge (alveolars) and hard palate (palatals).
444 Finally, the sensorimotor interpretation of the dorsal feature refers to a vertical and posterior movement
445 of the tongue dorsum for closure against the soft palate, or velum, resulting in the class of velar
446 sounds. Additionally, vowel sounds can be included in the dorsal feature owing to the motor execution
447 requirements of the tongue, but they are not included in any other place category [23], and we do not
448 include them here.

449 Our analysis revealed a network of neurological structures typically involved in speech motor control
450 with auditory feedback exhibiting patterns of ECoG recordings between three top-level place of
451 articulation categories (labial, coronal, dorsal). These regions included speech sensorimotor cortices,
452 premotor cortex, auditory cortex and Wernicke’s area. The combined contributions of all electrodes

453 over the 700 ms time window place category discrimination indicates a primary role in instantaneous
454 motor execution and sensory processing as evidenced by peak statistically significant responses near
455 zero ms latency relative to speech output. These networks are also likely involved in planning and
456 feedback processing as shown by statistically significant responses with local AI maxima at speech leading
457 latencies (-300 – 0 ms) and speech lagging latencies (0 – 300 ms), respectively. The topography over the
458 primary motor and somatosensory cortices in Figure 4 provide neurophysiological evidence to support this
459 intuitive interpretation. Further, the presence of overlapping sensorimotor locations (defined by electrode
460 placements) suggests the primary motor, premotor and somatosensory cortices are all differentially active
461 across various configurations of the lips and tongue used in speech. The spatial topography also includes
462 perisylvian auditory regions for all feature categories. We interpret these results as representing both
463 prediction of sensory consequences as well as self-perception of vocalized output (e.g., efference motor
464 copy [37]), evidenced by significant contributions preceding and following speech onset, respectively.
465 Like the motor production results, the overlapping auditory cortical responses between conditions indicate
466 that phonemes yield differential ECoG signals during auditory feedback (cf. [10]).

467 *4.2.2 Manner of articulation* Like place, the manner of articulation also results from muscular
468 contraction of the vocal tract, but is used to describe the quality of vocal airflow during speech production.
469 In the present analysis, we focus on two major feature descriptions of phonemes: obstruent and sonorant.
470 Obstruent sounds are characterized by a blockage of the oral cavity that prohibits sustained voicing,
471 while sonorants facilitate sustained voicing through a relatively open vocal tract. Obstruents include stops
472 (/b/), fricatives (/f/) and affricates (/tʃ/) while sonorants include nasals (/m/), liquids (/l/), glides (/w/) and
473 vowels. It is possible to examine neurological responses to each of the manner subtypes. However, for this
474 analysis we chose to focus on the top-level categories to boost the feature sample size given our phoneme
475 data taken from continuous speaking of paragraph scripts.

476 The spatial topography and temporal dynamics of statistically significant differences in neural activity
477 between manner features revealed a network involving the premotor cortex, auditory cortex and the
478 posterior superior temporal gyrus (i.e., Wernicke's area) for obstruent and sonorant features. The
479 perisylvian auditory regions were activated to a larger spatial extent compared to the more focal premotor
480 contribution. The temporal dynamics reach peak levels between 65 – 145 ms following acoustic output of
481 the phoneme and persists throughout the speech production window (up to 300 ms). These observations
482 of spatial and temporal results have three implications: 1) motor and sensory processes are involved in
483 the production of requisite airflow for different classes of phonemes (obstruents and sonorants), 2) that
484 discriminating auditory feedback of manner is represented over a relatively large region of perisylvian
485 auditory cortex, and 3) the differences in motor production of manner is represented by a focal region of
486 motor cortex.

487 *4.2.3 Voicing* Voicing reflects both the laryngeal muscular contractions needed to configure the larynx
488 for phonation as well as the acoustic perception of phonated speech (i.e., contains vocal fold oscillation).
489 The voicing feature is separated into just two classes, voiced and voiceless, and therefore can be
490 represented in our analysis by a single voicing feature. All of the sonorant sounds used in this analysis are
491 included in the [voiced] feature, as are those obstruents that are produced with vocal fold oscillation (e.g.
492 /b/ and /v/). The remaining obstruents are included in the [voiceless] feature.

493 The spatial topography analysis revealed a network of perisylvian regions extending into both the motor
494 and auditory cortices, and was similar to the patterns found in the manner condition analysis. The peak
495 response occurred at 95 ms post-vocalization, which suggests that this network is primarily involved in
496 the acoustic perception of voicing in self-produced speech. There is, however, a small pre-vocalization
497 response at -25 ms that may be interpreted as involved in the preparation or execution of laryngeal
498 commands for initiating (voiced), or preventing (voiceless) vocal fold oscillation.

499 4.2.4 *Summary of acoustic-articulatory features* The sensorimotor contribution for discriminating
500 manner of articulation and voicing is subdued and focal compared to responses in the place of articulation
501 analysis. According to the analysis of place, widespread activity over the precentral gyrus was likely
502 related to discriminating the three classes of articulation according to different lip, jaw and tongue
503 configurations. In contrast, the focal sensorimotor response observed in the manner and voicing analyses
504 indicates that there is less overall differential sensorimotor activation between the production of obstruent
505 and sonorant phonemes and those with and without voicing. Interestingly, the location of the *manner* and
506 *voicing* sensorimotor response is similar to a region recently proposed to represent laryngeal muscular
507 activation during phonation [38, 39]. The larynx, with the respiratory system, is critical for phonation
508 and generation of acoustic signals in the vocal tract. Our result supports the hypothesis that this region
509 is involved in the planning and execution of laryngeal movements used to separately produce voiced and
510 voiceless speech. That a putative neural correlate of laryngeal excitation may be useful for discrimination
511 of obstruents from sonorants potentially implicates a fundamental role of the larynx for planning and
512 executing different manners of articulation as well. Last, recent evidence has also shown this region
513 responds to auditory processing during perception of music [40]. These combined observations suggest
514 that portions of the motor cortex may be involved in both motor and auditory processing. With the
515 limited number of subjects meeting our screening criteria, we were unfortunately unable to complete
516 a combined spatio-temporal analysis with the statistical power necessary to precisely determine the role
517 of the sensorimotor activity. Future work with an increased sample size will be required to fully investigate
518 these effects.

4.3 EXAMINING PHONOLOGICAL DISCRIMINATION

519 We last examined differences in ECoG recordings between production of consonants and vowels. The
520 category of vowel versus consonant is mutually exclusive and binary. As seen in Figure 6 (bottom),
521 portions of the speech production and auditory feedback processing networks are differentially active for
522 production of consonants versus production of vowels, with similar spatial topography as observed in the
523 analysis of manner and voicing. The similarity between these and our previous manner and voicing results
524 is not surprising, as the consonant-vowel, obstruent-sonorant and voiced-voiceless classes encompass
525 nearly the same distribution of phonemes. The main difference between the two features is that certain
526 sonorants are included as consonants, but not obstruents (e.g., nasals, liquids and glides); similarly, some
527 consonants are included in the voiced category largely consisting of vowels. The consonant-vowel contrast
528 is represented by a primary peak in statistically significant differences in activation indices at +95 ms, with
529 a secondary increase in the range -110 – 0 ms relative to onset of speech output. This bimodal response
530 is different than the observed response for manner and place, and likely reflects the complex motor-
531 sensory dynamics involved in the production of all speech sounds, which are particularly intertwined
532 when considering a higher level, phonological concept. In contrast, the manner feature appears to be
533 solely determined by neural analysis of resulting auditory streams.

4.4 MORE FEATURES

534 In the present study, we examined differential neural representations of high-level articulatory-acoustic
535 (place and manner of articulation) and phonological characteristics during speech production. In
536 particular, we focused on the *places*: labial, coronal and dorsal, the *manners*: obstruent and sonorant
537 and *voicing*: voiced and voiceless. As noted previously, the place and manner factors have additional
538 sublevels of increasing refinement (e.g., bilabial & labiodental for place; nasal & fricative for manner).
539 With the present sample size, and the limited amount of time available with each patient, we were not
540 able to examine these additional features. For those factors that we did test, but did not report (e.g., vowel
541 tongue position), we believe that movements of the tongue for vowels are so common to all production
542 attempts that there were no differentially distinguishing features in the ECoG recordings. For results such
543 as these, prior investigations of the overall neural activations found during speech production adequately
544 describe these processes. Future studies with additional subjects and stimuli may help to pick up where

545 this study leaves off. In particular, new studies may optimize speech stimuli selected for representation
546 of as many phonemes and articulations as possible, while maintaining low user effort requirements. In
547 addition, it is possible that the electrode size and spacing in this study was too coarse to disambiguate the
548 fine distinctions between all possible features of speech articulation [11, 14]. Advances in micro-ECoG
549 [11, 12] and additional studies employing such preparations should be able to more comprehensively
550 investigate additional features.

4.5 POTENTIAL APPLICATIONS FOR BRAIN-COMPUTER INTERFACING

551 Another guiding principle of this work concerned potential application to a neural speech prosthesis,
552 which can interpret brain activity for generating speech output, or a “brain-to-text” device. Our techniques
553 are directly applicable to a motor-speech brain-computer interface (BCI) as the major observations were
554 all based on machine learning classification of speech sounds, which is alternately known as prediction or
555 decoding. Martin and colleagues [41] have recently developed a similar method that attempts to predict
556 actual speech acoustic output from recorded ECoG signals. Our work is distinguished from the Martin
557 et al. technique by the adoption of articulatory gestures as the classification basis as opposed to direct
558 acoustic prediction. However, both methods are advantageous as they limit the required classification
559 dictionary (cf. thousands of words needed for word prediction versus a dozen of articulatory features
560 or acoustic bases) and offer a generative means for word and sentence prediction. In other words, by
561 classifying or predicting a small set of place, manner, voicing and phonological features, it is possible
562 to represent any phoneme, combinations of phonemes (i.e., syllables, words), or even sentences. By
563 considering continuous speech, our methods are also capable of keeping pace with speaking rates observed
564 during natural communication, which would be a marked advancement in the field of augmentative
565 and alternative communication as well as brain-computer interfacing. In contrast, classifying individual
566 discrete words from brain signals would require a prohibitively large data set to select the correct word
567 from the thousands of words used in language.

5 CONCLUSIONS

568 In the present study, we examined speech production in the human brain as a sequence of articulatory
569 movements. These sequences have been alternately proposed in the literature to arise from phonetic
570 transcriptions from phonological representations (e.g., phonemes and syllables) [1, 2, 42], or theorized as
571 the basis for speech planning and production (e.g., gestural scores) [43]. The present study brings us closer
572 to resolving this debate by first determining whether fundamental articulatory features are identifiable
573 from electrocorticographic recordings in human subjects. The shift toward articulation changes the
574 paradigm of functional neural analysis toward understanding invariant motor outputs of language and
575 away from abstract representations of speech motor control (e.g., phonemes, syllables and words). The
576 combined analysis of motor sequences and phonological representations will provide the requisite means
577 for confirming or rejecting these two different theories of speech production.

DISCLOSURE/CONFLICT-OF-INTEREST STATEMENT

578 The authors declare that the research was conducted in the absence of any commercial or financial
579 relationships that could be construed as a potential conflict of interest.

AUTHOR CONTRIBUTIONS

580 Gerwin Schalk designed the research. Anthony Ritaccio, Peter Brunner and Aysegul Gunduz conducted
581 the research protocol. Fabien Lotte and Cuntai Guan contributed data analysis. Jonathan Brumberg and

582 Gerwin Schalk contributed to the interpretation of the results. Fabien Lotte, Jonathan Brumberg and
583 Gerwin Schalk wrote the paper.
584 † These authors contributed equally.

ACKNOWLEDGEMENTS

585 This work was supported by grants from the US Army Research Office (W911NF-07-1-0415, W911NF-
586 08-1-0216, W911NF-12-1-0109), the NIH/NIBIB (EB006356 and EB000856) and the NIH/NIDCD
587 (R03-DC011304). The authors would like to thank Dr. Cheung Chi Leung, Dr. Bin Ma, Christian Herff
588 and Dr. Haihong Zhang for their help in the segmentation of the audio signal.

REFERENCES

- 589 [1]P. Indefrey and W. J. M. Levelt. The spatial and temporal signatures of word production components.
590 *Cognition*, 92(1-2):101–44, 2004.
- 591 [2]Willem J. M. Levelt, Ardi Roelofs, and Antje S. Meyer. A theory of lexical access in speech
592 production. *The Behavioral and Brain Sciences*, 22(1):1–75, February 1999.
- 593 [3]Stéphanie Riès, Niels Janssen, Borís Burle, and F-Xavier Alario. Response-locked brain dynamics of
594 word production. *PLoS one*, 8(3):e58197, January 2013.
- 595 [4]V. L. Towle, H. A. Yoon, M. Castelle, J. C. Edgar, N. M. Biassou, D. M. Frim, J. P. Spire, and M. H.
596 Kohrman. ECoG gamma activity during a language task: differentiating expressive and receptive
597 speech areas. *Brain*, 131(Pt 8):2013–2027, 2008.
- 598 [5]Erik Edwards, Srikantan S Nagarajan, Sarang S Dalal, Ryan T Canolty, Heidi E Kirsch, Nicholas M
599 Barbaro, and Robert T Knight. Spatiotemporal imaging of cortical activation during verb generation
600 and picture naming. *NeuroImage*, 50(1):291–301, March 2010.
- 601 [6]N. E. Crone, D. Boatman, B. Gordon, and L. Hao. Induced electrocorticographic gamma activity
602 during auditory perception. *Clinical Neurophysiology*, 112(4):565–582, 2001.
- 603 [7]Edward F. Chang, Jochem W. Rieger, Keith Johnson, Mitchel S. Berger, Nicholas M. Barbaro, and
604 Robert T. Knight. Categorical speech representation in human superior temporal gyrus. *Nature*
605 *Neuroscience*, 13(11):1428–32, November 2010.
- 606 [8]Xiaomei Pei, Dennis L. Barbour, Eric C. Leuthardt, and Gerwin Schalk. Decoding vowels and
607 consonants in spoken and imagined words using electrocorticographic signals in humans. *Journal of*
608 *Neural Engineering*, 8(4):046028, August 2011.
- 609 [9]Xiaomei Pei, Eric C. Leuthardt, Charles M. Gaona, Peter Brunner, Jonathan R. Wolpaw, and Gerwin
610 Schalk. Spatiotemporal dynamics of electrocorticographic high gamma activity during overt and
611 covert word repetition. *NeuroImage*, 54(4):2960–72, February 2011.
- 612 [10]Brian N. Pasley, Stephen V. David, Nima Mesgarani, Adeen Flinker, Shihab A. Shamma, Nathan E.
613 Crone, Robert T. Knight, and Edward F. Chang. Reconstructing speech from human auditory cortex.
614 *PLoS Biology*, 10(1):e1001251, January 2012.
- 615 [11]Timothy Blakely, Kai J. Miller, Rajesh P. N. Rao, Mark D. Holmes, and Jeffrey G. Ojemann.
616 Localization and classification of phonemes using high spatial resolution electrocorticography
617 (ECoG) grids. In *IEEE Engineering in Medicine and Biology Society*, volume 2008, pages 4964–7,
618 January 2008.
- 619 [12]Spencer Kellis, Kai Miller, Kyle Thomson, Richard Brown, Paul House, and Bradley Greger.
620 Decoding spoken words using local field potentials recorded from the cortical surface. *Journal of*
621 *Neural Engineering*, 7(5):056007, October 2010.
- 622 [13]Eric C. Leuthardt, Xiao-Mei Pei, Jonathan Breshears, Charles Gaona, Mohit Sharma, Zac
623 Freudenberg, Dennis Barbour, and Gerwin Schalk. Temporal evolution of gamma activity in human
624 cortex during an overt and covert word repetition task. *Frontiers in Human Neuroscience*, 6:99,
625 January 2012.

- 626 [14]Kristofer E Bouchard, Nima Mesgarani, Keith Johnson, and Edward F Chang. Functional
627 organization of human sensorimotor cortex for speech articulation. *Nature*, 495(7441):327–32, March
628 2013.
- 629 [15]Emily M. Mugler, James L. Patton, Robert D. Flint, Zachary A. Wright, Stephan U. Schuele, Joshua
630 Rosenow, Jerry J. Shih, Dean J. Krusienski, and Marc W. Slutzky. Direct classification of all American
631 English phonemes using signals from functional speech motor cortex. *Journal of Neural Engineering*,
632 11(3):035015, June 2014.
- 633 [16]William J. Hardcastle and Nigel Hewlett, editors. *Coarticulation: Theory, data and techniques*.
634 Cambridge University Press, 1999.
- 635 [17]Daniel Recasens. Lingual coarticulation. In William J. Hardcastle and Nigel Hewlett, editors,
636 *Coarticulation: Theory, data and techniques*, chapter 4, pages 80—104. Cambridge University Press,
637 1999.
- 638 [18]Jan Kubanek and Gerwin Schalk. NeuralAct: A Tool to Visualize Electro cortical (ECoG) Activity on
639 a Three-Dimensional Model of the Cortex. *Neuroinformatics*, November 2014.
- 640 [19]J L Lancaster, M G Woldorff, L M Parsons, M Liotti, C S Freitas, L Rainey, P V Kochunov,
641 D Nickerson, S A Mikiten, and P T Fox. Automated Talairach atlas labels for functional brain
642 mapping. *Human Brain Mapping*, 10(3):120–31, Jul 2000.
- 643 [20]Gerwin Schalk, Dennis J. McFarland, Thilo Hinterberger, Niels Birbaumer, and Jonathan R.
644 Wolpaw. BCI2000: a general-purpose brain-computer interface (BCI) system. *IEEE Transactions*
645 *on Biomedical Engineering*, 51(6):1034–1043, 2004.
- 646 [21]Jürgen Mellinger and Gerwin Schalk. Brain-Computer Interfaces. In Bernhard Graimann, Gert
647 Pfurtscheller, and Brendan Allison, editors, *Brain-Computer Interfaces*, The Frontiers Collection,
648 pages 259–279. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010.
- 649 [22]International Phonetic Association. *Handbook of the International Phonetic Association: A Guide to*
650 *the Use of the International Phonetic Alphabet*. Cambridge University Press, 1999.
- 651 [23]T. A. Hall. Segmental features. In Paul de Lacy, editor, *The cambridge handbook of phonology*, page
652 324. Cambridge University Press, 2007.
- 653 [24]Lawrence R. Rabiner and Bing-Hwang Juang. *Fundamentals of speech recognition*. Prentice Hall,
654 1993.
- 655 [25]S Young, G Evermann, M Gales, T Hain, D Kershaw, XA Liu, G Moore, J Odell, D Ollason,
656 D Povey, V Valtchev, and P Woodland. *The HTK Book (for HTK version 3.4)*. Cambridge University
657 Engineering Department, 2006.
- 658 [26]Andreas Antoniou. *Digital filters*. McGraw Hill, New York, 2nd edition, 1993.
- 659 [27]H. Peng, F. Long, and C. Ding. Feature selection based on mutual information: criteria of max-
660 dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and*
661 *Machine Intelligence*, 27(8):1226–1238, 2005.
- 662 [28]F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi. A review of classification algorithms
663 for EEG-based brain-computer interfaces. *Journal of Neural Engineering*, 4:R1–R13, 2007.
- 664 [29]O. Ledoit and M. Wolf. A well-conditioned estimator for large-dimensional covariance matrices.
665 *Journal of Multivariate Analysis*, 88(2):365–411, 2004.
- 666 [30]F. Lotte and C.T. Guan. Learning from other subjects helps reducing brain-computer interface
667 calibration time. In *International Conference on Audio, Speech and Signal Processing*
668 *(ICASSP'2010)*, pages 614–617, 2010.
- 669 [31]B. Blankertz, S. Lemm, M.S. Treder, S. Haufe, and K.-R. Müller. Single-trial analysis and
670 classification of ERP components a tutorial. *Neuroimage*, 2010.
- 671 [32]T. Fawcett. An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8):861–874, 2006.
- 672 [33]JA Hanley and B.J. McNeil. A Method of Comparing the Areas under ROC curves derived from same
673 cases. *Radiology*, 148(3):839–843, 1983.
- 674 [34]W.S. Noble. How does multiple testing correction work? *Nature Biotechnology*, 27:1135–1137, 2009.
- 675 [35]Jonathan S. Brumberg, E. Joe Wright, Dinal S. Andreasen, Frank H. Guenther, and Phillip R.
676 Kennedy. Classification of intended phoneme production from chronic intracortical microelectrode
677 recordings in speech-motor cortex. *Frontiers in Neuroscience*, 5(65), 2011.

- 678 [36]Wilder Penfield and Lamar Roberts. *Speech and brain-mechanisms*. Princeton University Press,
679 Princeton, NJ, 1959.
- 680 [37]John F Houde, Srikantan S Nagarajan, Kensuke Sekihara, and Michael M Merzenich. Modulation of
681 the auditory cortex during speech: an MEG study. *Journal of Cognitive Neuroscience*, 14(8):1125–38,
682 November 2002.
- 683 [38]Kristina Simonyan and Barry Horwitz. Laryngeal motor cortex and control of speech in humans. *The*
684 *Neuroscientist*, 17(2):197–208, April 2011.
- 685 [39]Steven Brown, Elton Ngan, and Mario Liotti. A larynx area in the human motor cortex. *Cerebral*
686 *Cortex*, 18(4):837–45, April 2008.
- 687 [40]Cristhian Potes, Aysegul Gunduz, Peter Brunner, and Gerwin Schalk. Dynamics of
688 electrocorticographic (ECoG) activity in human temporal and frontal cortical areas during music
689 listening. *NeuroImage*, 61(4):841–8, July 2012.
- 690 [41]Stéphanie Martin, Peter Brunner, Chris Holdgraf, Hans-Jochen Heinze, Nathan E Crone, Jochem
691 Rieger, Gerwin Schalk, Robert T Knight, and Brian N Pasley. Decoding spectrotemporal features of
692 overt and covert speech from the human cortex. *Frontiers in Neuroengineering*, 7(May):14, January
693 2014.
- 694 [42]Frank H. Guenther, Satrajit S. Ghosh, and Jason A. Tourville. Neural modeling and imaging of the
695 cortical interactions underlying syllable production. *Brain and Language*, 96(3):280–301, March
696 2006.
- 697 [43]Elliot L. Saltzman and Kevin G. Munhall. A Dynamical Approach to Gestural Patterning in Speech
698 Production. *Ecological Psychology*, 1(4):333–382, December 1989.

Figure 1. Locations of implanted grids on individual subject cortical models based on co-registered pre-op MR and post-op CT data. The bottom right figure shows the electrode locations projected on an average brain for those four subjects (B, E, F and G) that passed initial screening (see Section 2.3.5). Each subject's electrodes are represented with a different color.

Figure 2. Experimental setup

Figure 3. Example of the ECoG gamma envelope from the two electrodes circled in green and blue, for the production and perception of the words “abolish all.” The transcription of these two words into phonemes is all also provided, together with the corresponding class label for the articulatory feature “vowel” (‘+’: the phoneme is a vowel, ‘-’: the phoneme is a consonant).

Figure 4. The spatial topography and temporal dynamics are shown in the left and right columns, respectively, for electrode locations with significant machine learning classification for the ‘place’ category levels: labial, coronal, and dorsal.

Figure 5. The spatial topography and temporal dynamics are shown in the left and right columns, respectively, for electrode locations with significant machine learning classification for the ‘manner’ category levels: obstruent and sonorant.

FIGURES

Figure 6. The spatial topography and temporal dynamics are shown in the left and right columns, respectively, for electrode locations with significant machine learning classification for the ‘voicing’ (i.e., voiced vs. voiceless) and ‘phonological’ (i.e., consonant vs. vowel) categories.

Figure 1.TIF

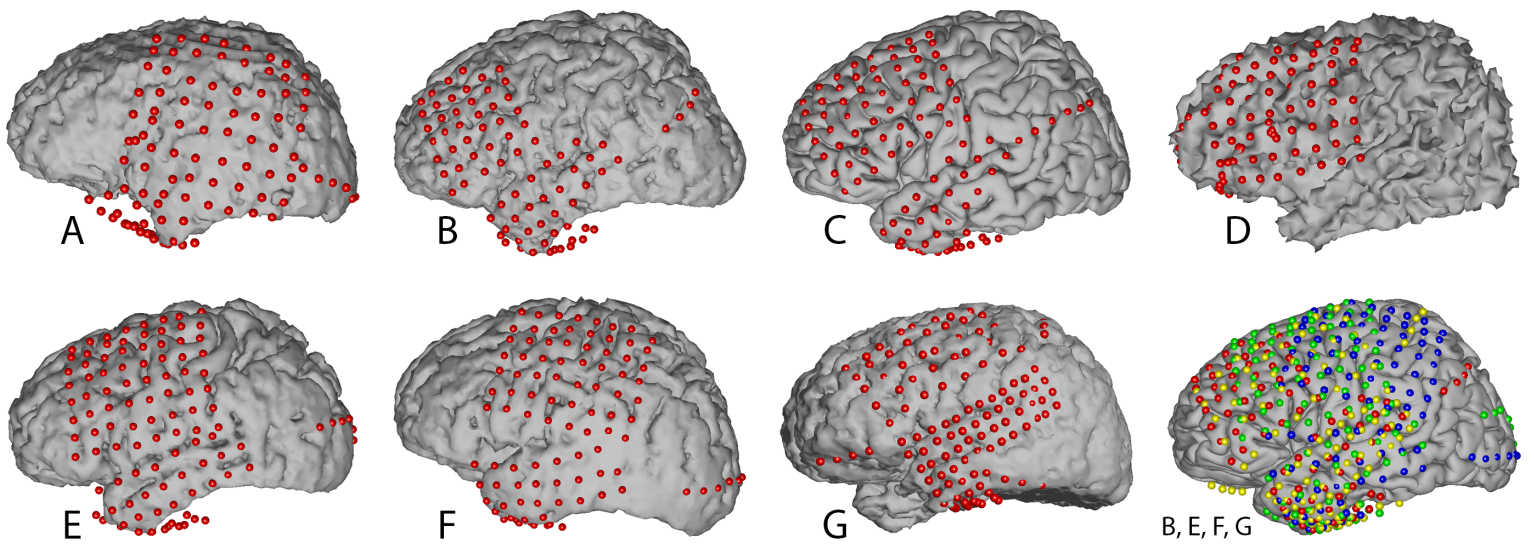


Figure 2.TIF

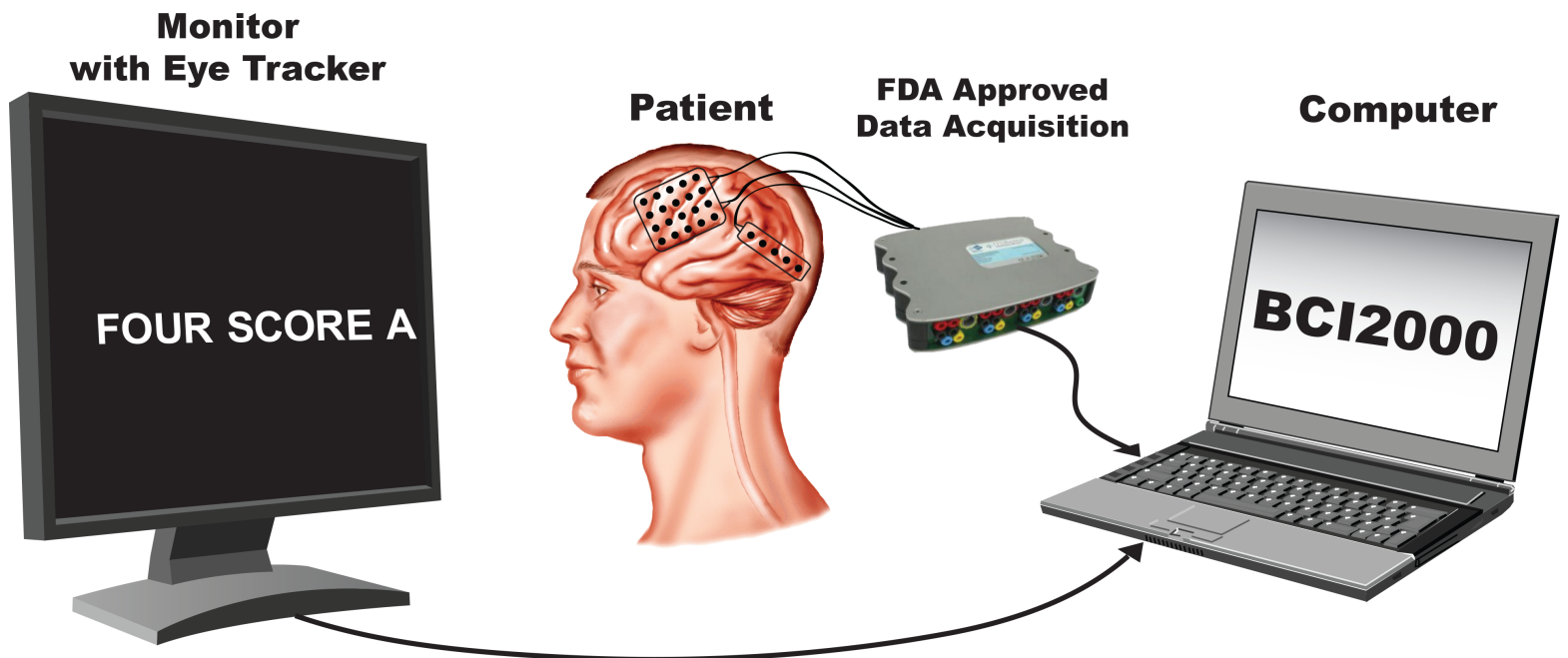


Figure 3.TIF

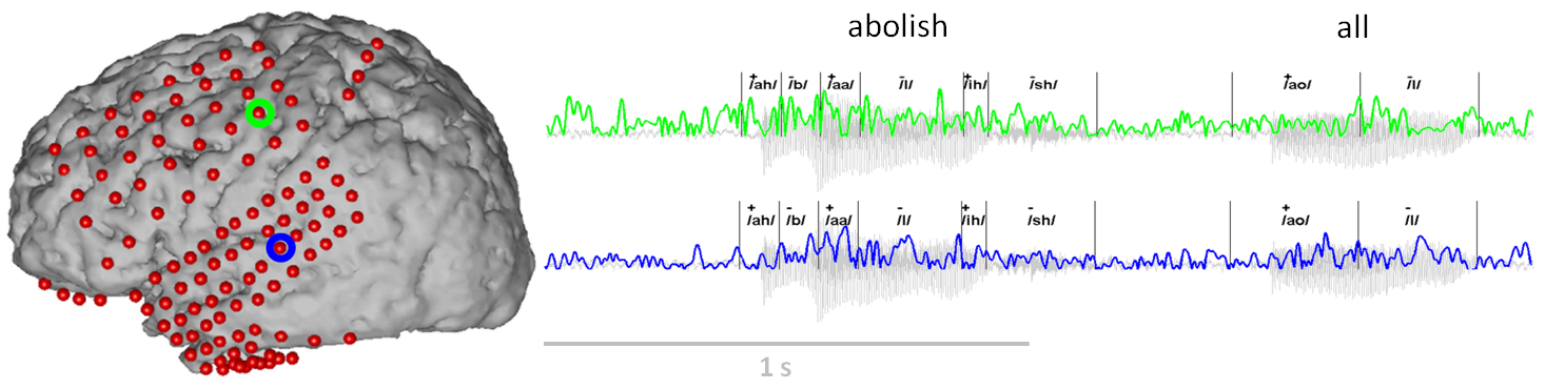
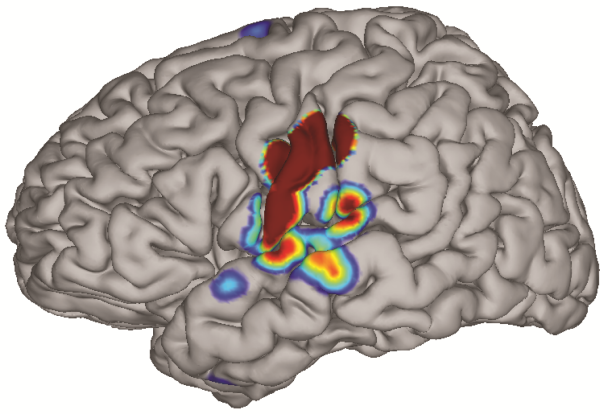
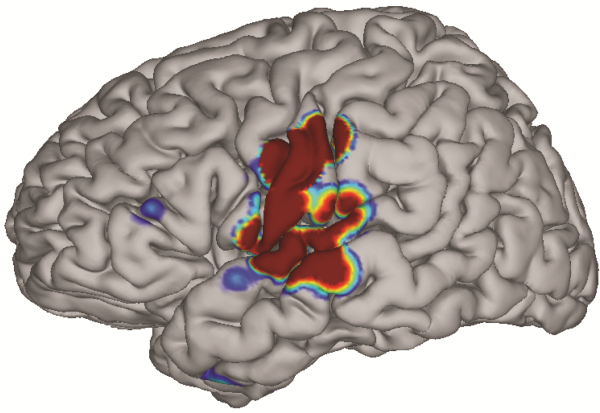
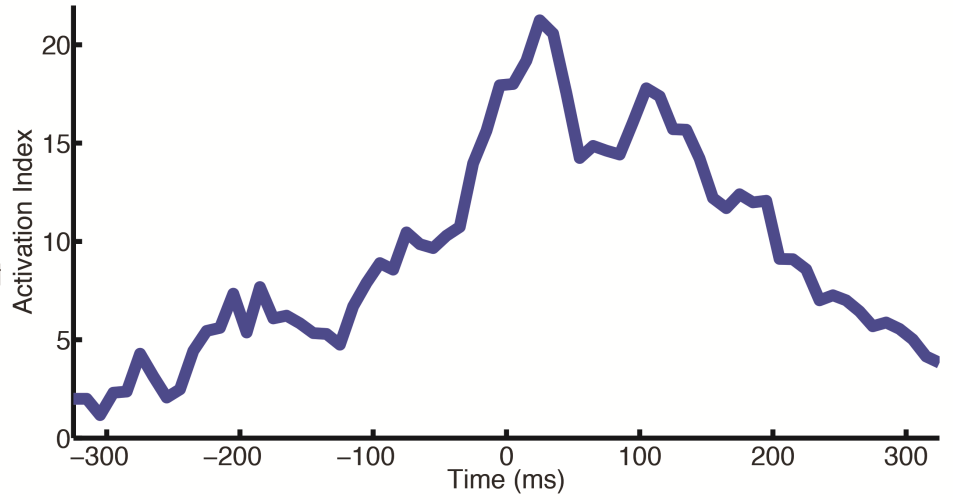


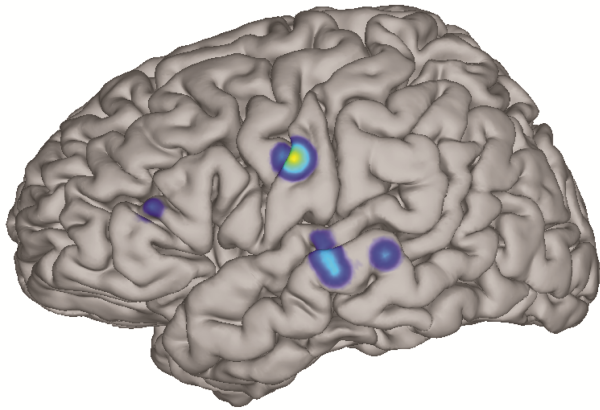
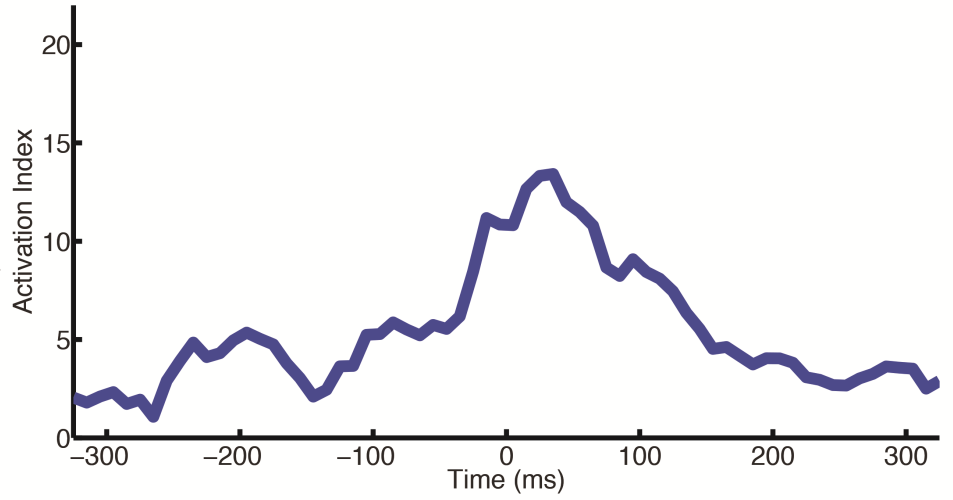
Figure 4.TIF



Labial



Coronal



Dorsal

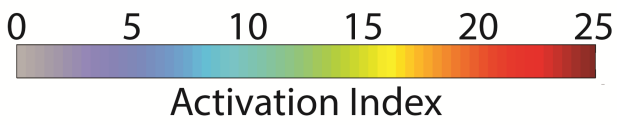
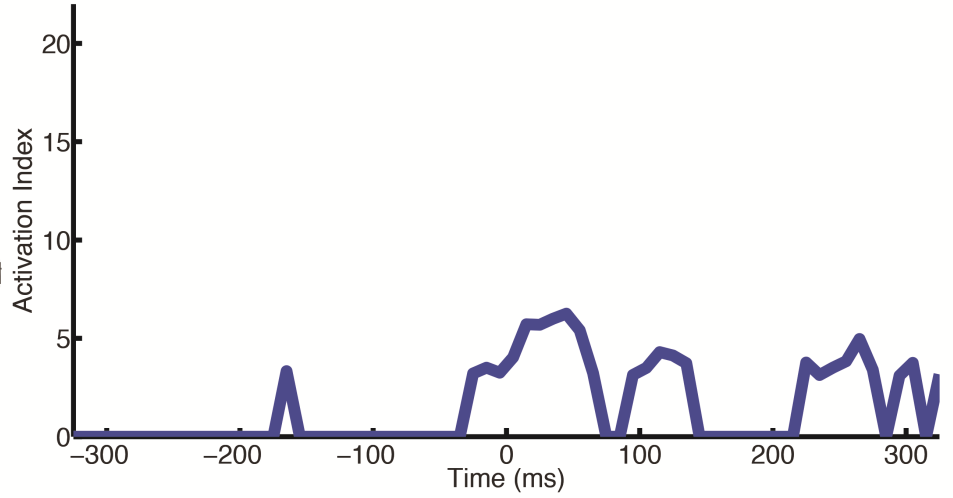
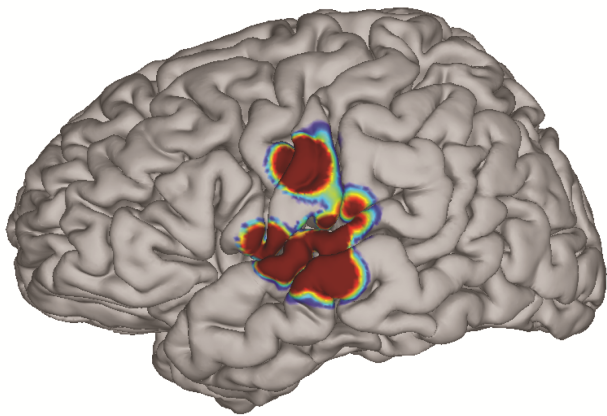
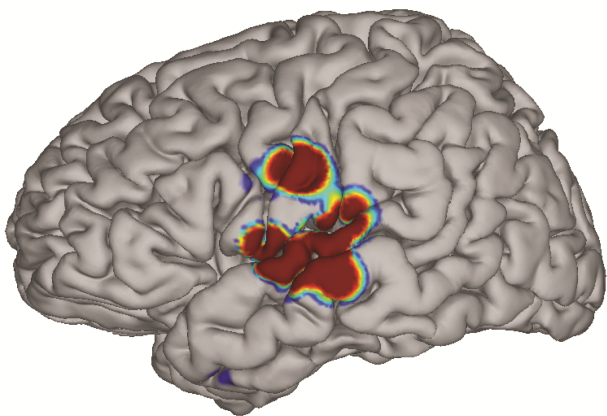
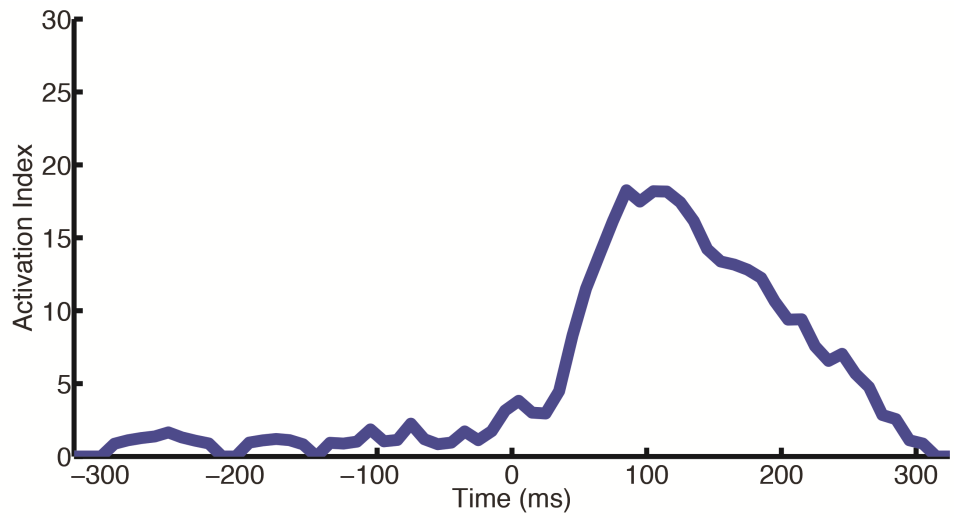


Figure 5.TIF



Obstruent



Sonorant

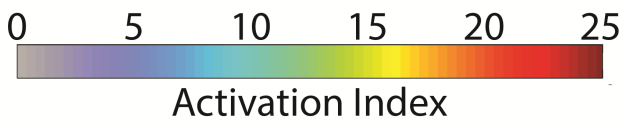
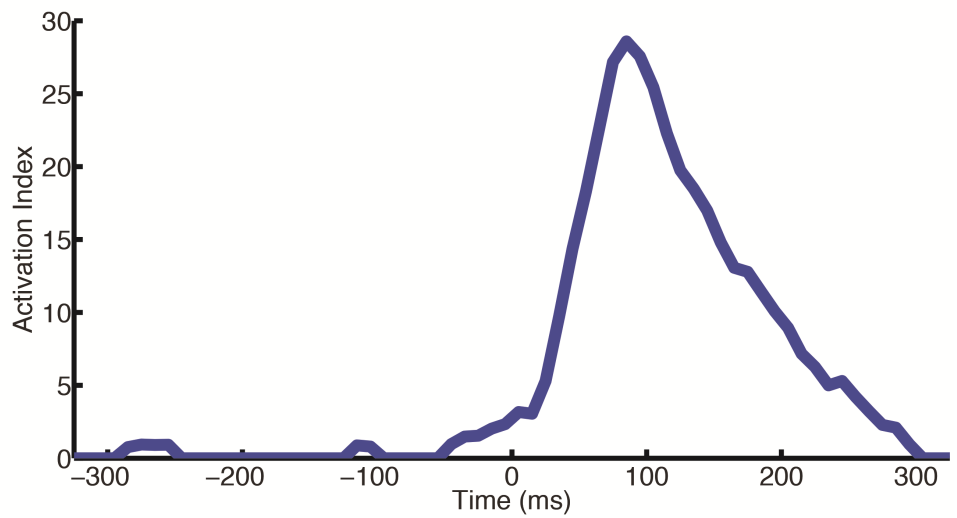
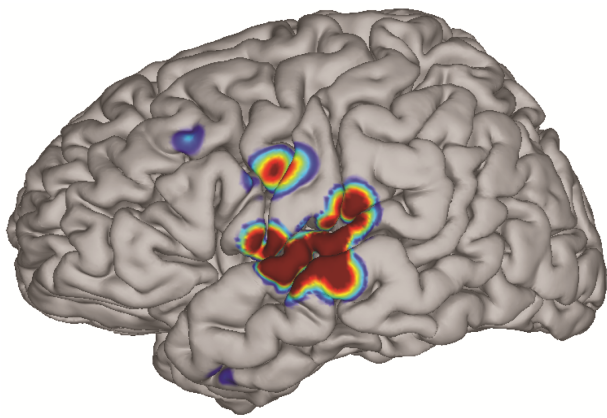
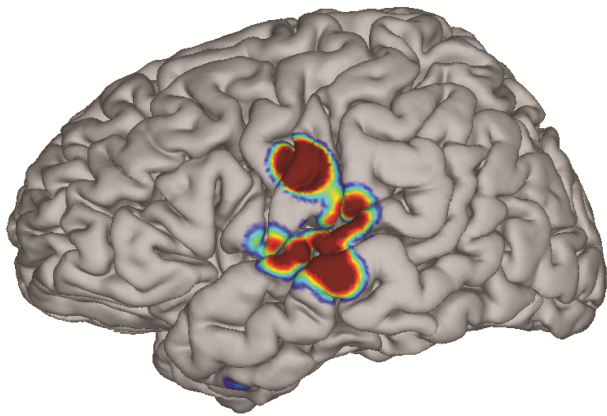
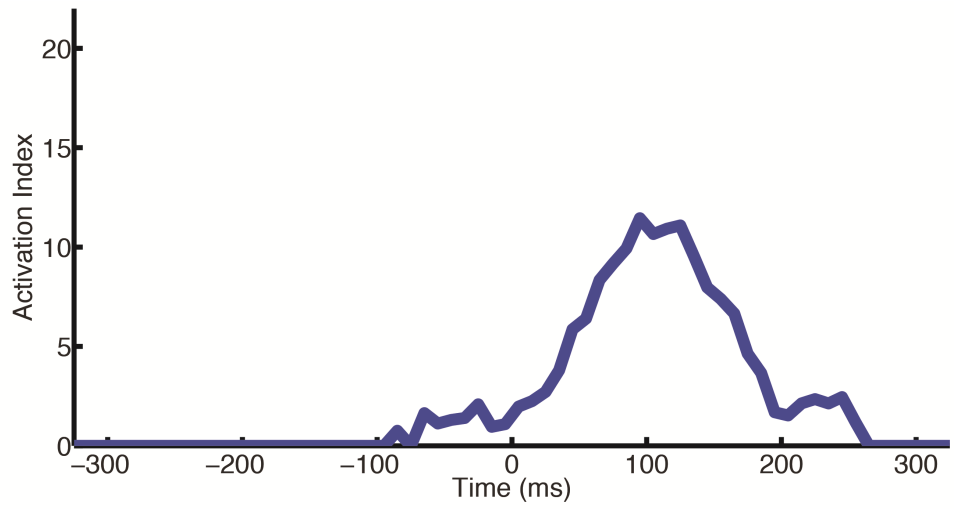


Figure 6.TIF



Voicing



Vowel

