



Classification of intended phoneme production from chronic intracortical microelectrode recordings in speech-motor cortex

Jonathan S. Brumberg^{1*}, E. Joe Wright², Dinal S. Andreasen^{2,3}, Frank H. Guenther^{1,4,5} and Philip R. Kennedy²

¹ Department of Cognitive and Neural Systems, Boston University, Boston, MA, USA

² Neural Signals Inc., Duluth, GA, USA

³ Georgia Tech Research Institute, Marietta, GA, USA

⁴ Department of Speech, Language, and Hearing Sciences, Boston University, Boston, MA, USA

⁵ Division of Health Sciences and Technology, Harvard University–Massachusetts Institute of Technology, Cambridge, MA, USA

Edited by:

Nicholas Hatsopoulos, University of Chicago, USA

Reviewed by:

Jose M. Carmena, University of California Berkeley, USA

Wasim Q. Malik, Harvard Medical School, USA

*Correspondence:

Jonathan S. Brumberg, Department of Cognitive and Neural Systems, Boston University, 677 Beacon Street, Boston, MA 02215, USA.

e-mail: brumberg@bu.edu

We conducted a neurophysiological study of attempted speech production in a paralyzed human volunteer using chronic microelectrode recordings. The volunteer suffers from locked-in syndrome leaving him in a state of near-total paralysis, though he maintains good cognition and sensation. In this study, we investigated the feasibility of supervised classification techniques for prediction of intended phoneme production in the absence of any overt movements including speech. Such classification or decoding ability has the potential to greatly improve the quality-of-life of many people who are otherwise unable to speak by providing a direct communicative link to the general community. We examined the performance of three classifiers on a multi-class discrimination problem in which the items were 38 American English phonemes including monophthong and diphthong vowels and consonants. The three classifiers differed in performance, but averaged between 16 and 21% overall accuracy (chance-level is 1/38 or 2.6%). Further, the distribution of phonemes classified statistically above chance was non-uniform though 20 of 38 phonemes were classified with statistical significance for all three classifiers. These preliminary results suggest supervised classification techniques are capable of performing large scale multi-class discrimination for attempted speech production and may provide the basis for future communication prostheses.

Keywords: locked-in syndrome, speech prosthesis, neurotrophic electrode, chronic recording, motor cortex

INTRODUCTION

The objective of providing reliable control over external devices using neural activity is becoming more a reality. Already, many researchers have shown it is possible for humans and animals to control computer mouse cursors (Kennedy et al., 2000, 2004; Serruya et al., 2002; Taylor et al., 2002; Carmena et al., 2003; Leuthardt et al., 2004; Wolpaw and McFarland, 2004; Hochberg et al., 2006), robotic arms (Chapin et al., 1999; Carmena et al., 2003; Hochberg et al., 2006; Velliste et al., 2008), spelling devices (Farwell and Donchin, 1988; Sutter, 1992; Donchin et al., 2000; Scherer et al., 2004; Blankertz et al., 2007; Friman et al., 2007), and artificial speech synthesizers (Brumberg et al., 2009, 2010; Guenther et al., 2009) using brain machine interfaces (BMIs). The current goal of many such systems is to provide an effective means for severely paralyzed persons to interact with the outside environment and conduct interpersonal communication. In particular, those with locked-in syndrome (LIS; Plum and Posner, 1972) can receive the greatest benefit, as they suffer from complete paralysis of the voluntary motor system, including limb, facial, and vocal movements while maintaining intact cognition, sensation, and perception. Those with LIS often have some voluntary ocular control, though if at all, typically confined to vertical movements only and blinking. Further, persons with LIS resulting from stroke have relatively long survival rates with over 80% survival at 5 and 10 years post-onset (Doble et al., 2003).

Current neurophysiological recording techniques used for BMI applications include: electroencephalography (EEG), intracranial electrocorticography (ECoG), and intracortical microelectrode recordings. Most primate studies use invasive techniques such as implanted microelectrode arrays for use in devices which utilize predicted limb kinematics from primary motor cortical activity to operate computer cursors and robotic arms and legs. In contrast, most human studies are designed for the sole purpose of restoring communication to those who have been rendered mute. Of these, the majority of applications use non-invasive means (i.e., EEG) to record brain potentials for control of spelling devices. These methods can include synchronized feedback event related potentials (P300; Farwell and Donchin, 1988; Donchin et al., 2000), visual evoked potentials (SSVEP; Sutter, 1992; Cheng et al., 2002; Friman et al., 2007), slow cortical potentials (Birbaumer et al., 1999, 2000), and sensorimotor potentials (Neuper et al., 2003; Obermaier et al., 2003; Scherer et al., 2004; Vaughan et al., 2006; Blankertz et al., 2007).

Human volunteers, often undergoing pre-surgical evaluation for intractable epilepsy, have also been implanted with temporary ECoG arrays to study its utility in BMI applications. Of particular interest to the development of BMI for speech communication are ECoG investigations of the neural architecture underlying speech perception and production (Crone et al., 2001a,b; Towle et al., 2008) and those which investigate classification of words and phonemes during speech production attempts (Blakely et al.,

2008; Kellis et al., 2010). A main result of the studies by Towle et al. (2008) and Crone et al. (2001a,b) illustrates the pattern of gamma (>30 Hz) band activity and suppression is consistent with expected locations of motor cortex activity during speech production based on human microstimulation studies (e.g., Penfield and Roberts, 1959) as well as numerous functional magnetic resonance imaging (fMRI) studies (see Guenther et al., 2006 for a summary). Additional human volunteers suffering from quadriplegia, LIS, or ALS have been chronically implanted with microelectrode recording systems for BMI control of spelling devices (Kennedy et al., 2000, 2004; Hochberg et al., 2006) and artificial speech synthesizers (Guenther et al., 2009; Brumberg et al., 2010).

The study reported here specifically complements prior work for decoding phonemes and words from ECoG recordings of neurological activity during speech production (Blakely et al., 2008; Kellis et al., 2010) and our own previous results with a chronic BMI for speech synthesizer control using the Neurotrophic Electrode (Brumberg et al., 2009, 2010; Guenther et al., 2009). The prior ECoG investigations examined neurological activity using closely spaced micro-ECoG arrays (<3 mm spacing) during production of relatively small numbers of speech sounds (four phonemes in Blakely et al., 2008; 10 words in Kellis et al., 2010). The results of these studies illustrate that phoneme and word decoding is possible using invasively recorded neurological activity related to speech production and provides an experimentally sound foundation for the current study. Our prior investigation (Guenther et al., 2009; Brumberg et al., 2010) used a discrete-time filter based approach for prediction of acoustic features of intended vowel sound productions by the BMI user. In that study, the user was able to correctly produce one of three vowel sounds (/ɑ/, /i/, or /u/) with 70% mean accuracy by the end of each recording session. In the current investigation, we explore the usefulness of machine learning techniques for off-line discrete classification of intended phoneme production including both vowels and consonants. In essence this, and prior, work is trying to perform automatic speech recognition (ASR) directly from neurological sources rather from voice measurements or other non-vocal representations used for silent speech interfaces (see Denby et al., 2010 for a review of silent speech interfacing). The results of this study support the hypothesis that neural activity from the motor cortex is correlated with attempted phoneme production. Though preliminary, the present results provide useful insights into the future design of speech BCI using implanted electrodes in the speech-motor cortex and motivates future study and development.

MATERIALS AND METHODS

The methods described below have been approved by the Food and Drug Administration (IDE # G960032) and by the Institutional Review Boards of Neural Signals, Inc. and Gwinnett Medical Center, Lawrenceville, GA, USA. Informed consent was obtained by the participant and his legal guardian for all procedures and experimental studies.

SUBJECT

A single male participant with LIS volunteered for implantation of the Neurotrophic Electrode (Kennedy, 1989; Bartels et al., 2008) device to investigate the neural correlates of speech production for use in a speech production BMI. The participant was 25 years old at

the time of the present study with LIS resulting from a brain–stem stroke of the ventral pons. The stroke occurred 9 years prior as the result of a severe automobile accident. Consistent with LIS, he has near-total paralysis with voluntary motor control of slow and disconjugate eye movements but is otherwise cognitively intact. He occasionally exhibits involuntary muscle spasms and episodes of coughing. At the present, his natural communication is restricted to voluntary vertical eye movements: up to indicate “yes” and down for “no.” He was implanted with the Neurotrophic Electrode device on December 22, 2004 according to previously described methods (Bartels et al., 2008).

Data collection is possible for only an hour or two per recording session limited by the level of fatigue endured by the participant and is terminated if he becomes overly fatigued. In addition, he is administered 200 mg of Provigil (Modafinil, Cephalon Inc.) 30 min before data collection to help ensure a constant state of wakefulness. No additional medications have been administered to the participant during the course of this study.

ELECTRODE PLACEMENT AND TARGET DETERMINATION

The purpose of the implantation and experimental study was to develop a BMI for speech production. Therefore, we used fMRI prior to implantation to determine the stereotactic locations of brain areas active during speech production attempts. Brain images were collected using a 1.5 Tesla scanner as the participant performed a picture naming task in which he was asked to attempt to speak, rather than simply imagine speaking, the names of the shown items. We used an active speech production protocol based on evidence that motor execution elicits greater neurological activity than imagery alone. Specifically, motor execution shows approximately 30% more activation than imagery in fMRI studies and similar spatial activity patterns (Roth et al., 1996; Jeannerod and Frak, 1999). The fMRI protocol was expected to highlight the brain regions related to speech articulation as relative increases in blood oxygenation level dependent (BOLD) response compared to baseline. A single area on the left precentral gyrus with peak BOLD response was selected and was subsequently stereotactically implanted with the Neurotrophic Electrode. Functionally, this area lies on or near the border between pre-motor and primary motor cortex. A further description of the electrode placement and target determination can be found elsewhere (Bartels et al., 2008, Guenther et al., 2009).

NEUROTROPHIC ELECTRODE DESIGN

Extracellular microelectrodes and microelectrode arrays have been used for chronic intracortical recording of neural activity in many animal studies. Two well known designs include the planar (or Michigan) array (Wise et al., 1970; Hoogerwerf and Wise, 1994) and microwire array (Williams et al., 1999; Taylor et al., 2002; Nicoletis et al., 2003), though these devices are not approved for human use. Currently, the only microelectrodes for use in human studies are the Utah, or NeuroPort array (Jones et al., 1992; Maynard et al., 1997; Hochberg et al., 2006; Donoghue et al., 2007) and Neurotrophic Electrode (Kennedy, 1989; Kennedy et al., 1992a; Kennedy and Bakay, 1998). Both have been previously used in human and animal investigations. A detailed description of the planar array, microwire array and Utah array are beyond the scope of this paper. See Brumberg et al. (2010), Schwartz (2004) and Ward et al. (2009) for more complete reviews.

The design and construction of the Neurotrophic Electrode are described in detail elsewhere (Kennedy, 1989; Bartels et al., 2008). Briefly, it consists of three coiled wires that end inside a conical glass tip with one wire as a reference signal, resulting in two recording channels. The glass cone tip is 1.5 mm in length, 0.1 mm in diameter at the narrow end and 0.3 mm in diameter at the wide end where the wires enter. The wires are glued to the inside of the glass using methyl methacrylate glue. Surgical implantation techniques follow those previously reported (Kennedy et al., 1992a).

The proximal ends of the wires are connected to implanted electronics for pre-amplification and wireless telemetry (Bartels et al., 2008). The implanted electronic device consists of a power induction coil (rather than implanted batteries), amplifiers for each recording channel, and frequency-modulated (FM) transmitters for wireless data transmission across the scalp. The telemetry system is tuned and signal amplitudes are calibrated utilizing a sine wave signal generated by the implanted hardware system. The sine wave is emitted by the implanted electronics upon receiving power from the induction system which operates at a transmission frequency near 1.1 MHz. The implanted electronics are contoured to the shape of the skull by applying a three-layer coating of insulation over a skull model. The first two layers consist of Parylene (Sadhir et al., 1981) and Elvax, an Ethylene-Vinyl Acetate Copolymer resin (E.I. DuPont de Nemours, Inc.), providing excellent electrical insulation. The outer layer of Silastic (MED-6607, NuSil Silicone Technology) is applied for mechanical protection.

The glass cone is filled with a neurotrophic growth factor which promotes neurite growth from nearby (~600 μm ; Kennedy, 1989) neurons into the hollow glass cone. Histological analysis of recovered implants from post-mortem rats and monkeys indicates myelinated axons stream through the glass cone, providing mechanical stability within the cortex (Kennedy et al., 1992b). Further, the electrode wires are tightly coiled between the cortical surface and the implanted electronics to reduce mechanical strain. These features and the electrode geometry help minimize the effects of micro-movements and related glial scarring which both impair long-term chronic recordings. Consequently, the Neurotrophic Electrode has been successfully used to record from human neocortex for over 4 years in three subjects (Kennedy, 2006).

RECORDING SYSTEM

Neural recordings begin once the surgical incision has completely healed, typically 3 weeks after implantation, with the neural signal stabilizing after approximately 3 months. The external power induction coil is approximated to the scalp over the inner coil, while receiving antennae are placed over the subdural transmission coils as shown in **Figure 1** (right side coil: power induction, left side coils: reception antennae). All coils are placed close to the shaved scalp and secured in place by water soluble EC2 electrode cream (Grass Technologies). Recorded neural signals are 100 \times amplified via implanted hardware and FM for wireless data transmission. FM signals from the wireless telemetry hardware are acquired and tuned using WinRADiO antennae and demodulation systems (WinRADiO Communications, Adelaide, Australia). Next, the received FM signals are routed to an analog amplifier (model BMA 831, CWE, Inc., USA) with 10 \times gain and band-pass filter settings of 1 Hz to 10 KHz and archived on a DDS tape recorder (Cygnum

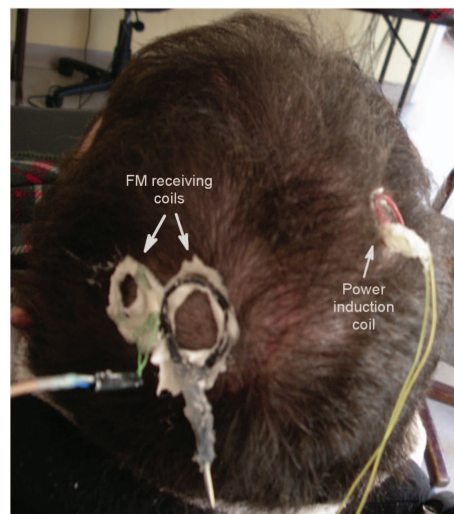
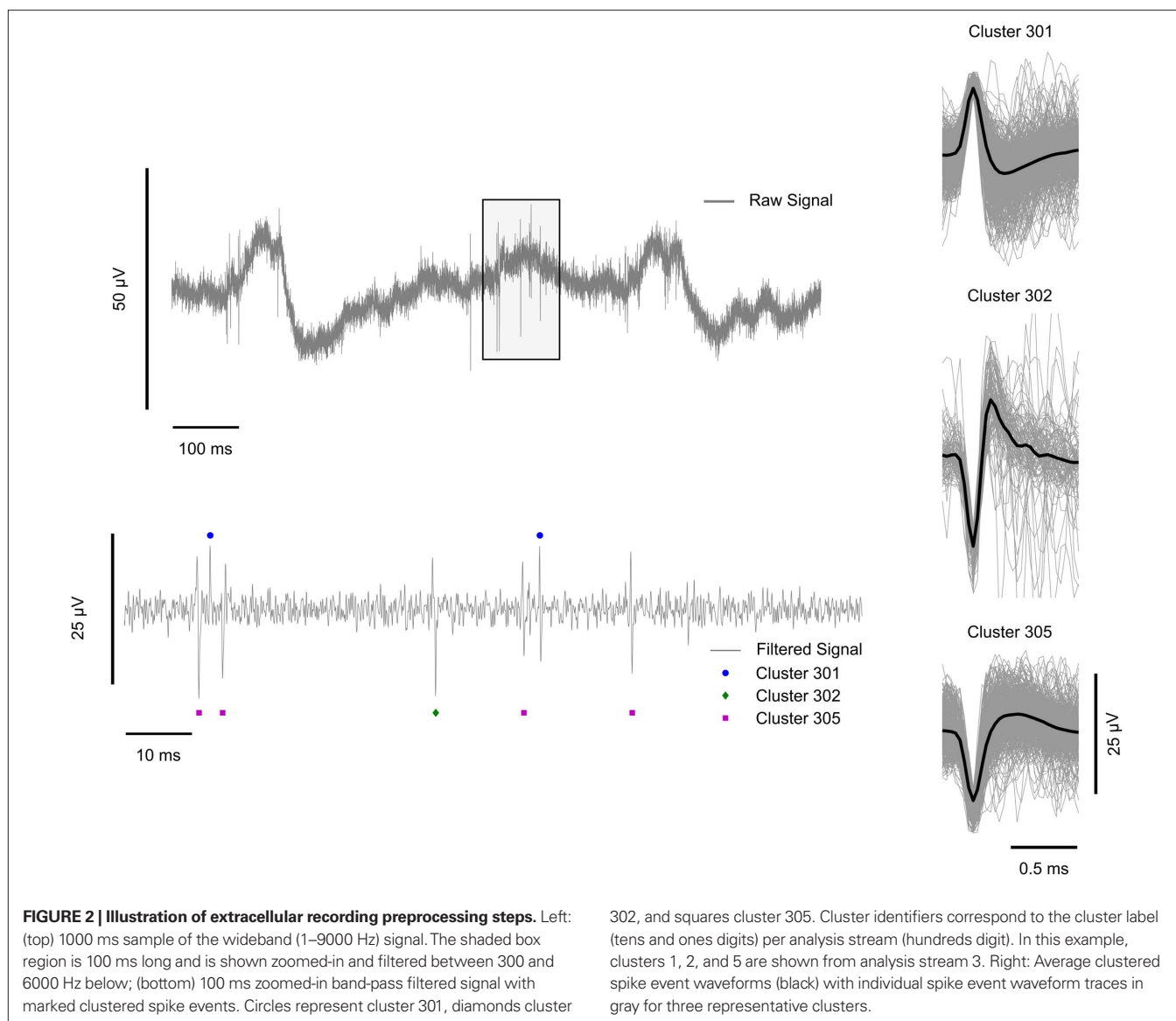


FIGURE 1 | The power induction coil is on the right side, with the two FM receiving coils on the left. All coils are held in position with EC2 electrode paste.

Technology, Inc., USA) for off-line analysis. The amplified signals are then sampled at 30 kHz using the Neuralynx, Inc. (Bozeman, MT, USA) Cheetah data acquisition system for A/D conversion and further amplification (digital 2 \times gain). Each signal is divided into two processing streams; one is band-pass filtered from 1 to 9000 Hz for digital storage of the wideband signal and the other is filtered at 300–6000 Hz, beyond the LFP range, for further analysis.

Spike events were first obtained by positive and negative voltage threshold crossings at $\pm 10 \mu\text{V}$. A 32-point (~1 ms) waveform was taken around the threshold crossing, peak-aligned to the eighth sample and saved for further analysis. Observed signal magnitudes ranged from 12 to 60 μVpp . Total acquisition system setup time is less than 10 min.

Thresholded spike event waveforms were sorted first manually off-line with the Neuralynx, Inc. SpikeSort3D program, utilizing a standard manual convex-hull cluster-cutting algorithm according to five features: peak/valley amplitude, height (peak-to-peak amplitude), energy (area under the spike), and amplitude at the spike alignment point (See Lewicki, 1998 for review of spike sorting techniques). The eighth sample amplitude, which ideally corresponds to the maximum peak or valley amplitude, was used as the primary separation feature and clustered. The clustered spike events represent the complex or compound waveforms of many extracellular potentials occurring within the glass cone. For this reason, we do not differentiate between those spike clusters which are possibly single-units and those which are not resolvable. Spike timestamps were recorded and stored according to the alignment point of the clustered waveform. A total of 31 spike clusters were obtained using to this procedure. The final feature space parameters derived from this off-line analysis were saved for reuse over all recording sessions in this study. **Figure 2** shows an example of the neural recordings used in this study, including a 1000-ms wideband signal trace, a 100-ms filtered signal trace and examples of clustered spike events with individual and mean waveform shapes.



EXPERIMENTAL PROTOCOL

The participant was presented with acoustic stimuli of 39 English phonemes (only 38 were used in the following analysis). The phonemes used are listed in **Tables 1** (vowels and diphthongs) and **2** (consonants). All phonemes were 200 ms samples taken from exemplar words spoken by the participant's father and recorded as 44.1 kHz wav-formatted sound files. The stimuli consist of nearly all steady English vowels, common diphthongs, and balanced voiced/voiceless consonants (where possible and appropriate). A potential confound is present with consonant stimuli. Specifically, stop consonants cannot be produced in the absence of a trailing or leading vowel. In this study, all consonants are presented as a consonant–vowel (CV) utterance, with the trailing vowel minimized. Additionally, not all trailing vowel combinations were used.

Experimental protocol execution was controlled via custom software written in the Microsoft .Net Framework running on a computer integrated with the data acquisition system. Each trial began with an auditory LISTEN instruction followed by a “ding”

and playback of a phoneme sound file. Following the LISTEN period, an auditory SPEAK instruction was given followed by a “ding.” During the LISTEN period, the participant was told to passively listen to the presented phoneme and to attempt to produce the sound during the SPEAK period. Importantly, he was instructed to not merely visualize the production task, but to try to physically move his face and vocal tract, as required, despite his paralysis. We hypothesized that motor cortex activity would be greatest for attempted speech production rather than imagined speech movements, though this was not explicitly tested in the current study. This hypothesis was supported by evidence of greater motor activation during motor execution compared with imagery in fMRI (Roth et al., 1996) and EEG (Neuper et al., 2005). Other evidence suggests similar activations are elicited in people with quadriplegia during both motor imagery and attempted motor execution (Lacourse et al., 1999) implying imagery may be as effective as attempted motor execution. However, the instruction to attempt the speech productions was intended to evoke, at

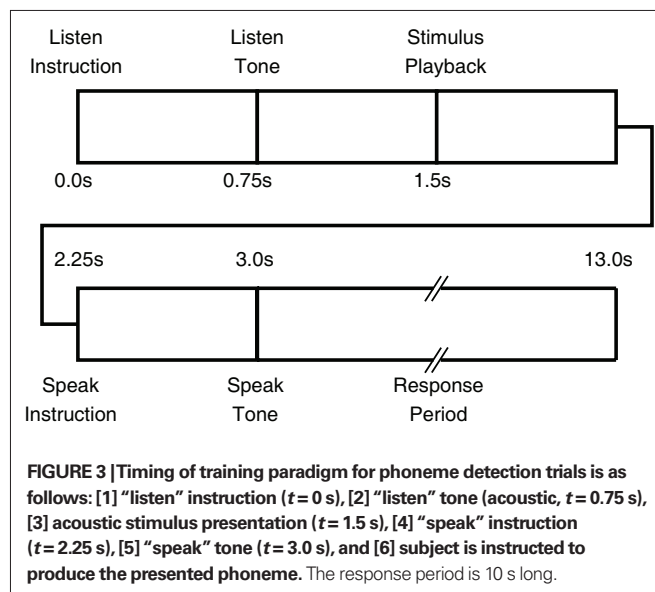
Table 1 | Vowel and diphthong stimuli used for the phoneme production experiment.

Phoneme #	ARPABET	IPA	Example	Type	Articulatory features		
					Height	Location	Rounding
1	IY	i	bEAd	Vowel	High	Front	Spread
2	IH	ɪ	bld	Vowel	High	Front	Spread
3	EH	ɛ	bEd	Vowel	Middle	Front	Neutral
4	AE	æ	bAd	Vowel	Low	Front	Neutral
5	AA R	ɑr	bARd	Diphthong: rhotic	NA	NA	NA
6	AA	ɑ	pOt	Vowel	Low	Back	Neutral
7	AO R	ɔr	pORt	Diphthong: rhotic	NA	NA	NA
8	UH	ʊ	pUt	Vowel	High	Back	Rounded
9	UW	u	bOOt	Vowel	High	Back	Rounded
10	AH	ʌ	bUd	Vowel	Middle	Central	Neutral
11	ER	ɜr	bIRd	Vowel	Middle	Central	Neutral
12	EY	eɪ	bAlt	Diphthong	NA	NA	NA
13	AY	aɪ	blte	Diphthong	NA	NA	NA
14	OY	ɔɪ	bOY	Diphthong	NA	NA	NA
15	AW	aʊ	cOW	Diphthong	NA	NA	NA

minimum, kinesthetic motor imagery which has been shown to outperform visual–motor imagery in BMI tasks (Neuper et al., 2005). Each instruction (LISTEN and SPEAK) was 750 ms in duration, the LISTEN period 1500 ms and the SPEAK period 10000 ms. The time-course of each trial is shown in **Figure 3**. Neural signals from the Neurotrophic Electrode were continuously sampled during the entire session and stored for off-line analysis. Unique trial identifiers were generated by the paradigm control computer and sent as analog input to the data acquisition system to synchronize the experimental protocol with neural recordings.

The experiment was completed over 10 months involving 59 sessions in which the participant passively listened and/or imagined speaking with and without biofeedback. For the current analysis, only a subset of sessions having no data collection technical problems and no biofeedback were investigated. The resulting data set consists of 431 trials over five sessions with neural firing rate estimated in 300 ms bins during the 10-s SPEAK period ($N = 20$ bins per trial). There were between 10 and 12 repetitions of each phoneme randomly presented across all sessions, and phoneme 24 (/f/) was not used due to insufficient trials. Rest periods were inserted between blocks of trials. Relatively large, 300 ms bins were used, rather than shorter bins (cf. 15 ms bins in our previous studies; Guenther et al., 2009; Brumberg et al., 2010) due to the nature of the classification task in the current study. In particular, we were interested in classification of entire phoneme productions, each of which we assumed was close to the auditory stimulus length (200 ms). We therefore chose to examine the bin-counts for a time range that would hopefully encompass an entire production period.

A potential limitation in any study involving subjects with LIS is the lack of instantaneous, observable behavior. For instance, in a prior ECoG study of speech production, subjects’ overt speech productions were directly compared to neurological activity (Kellis et al., 2010). Such comparisons are not possible when investigating the intended behavior of LIS subjects. Therefore, we were not able to determine precisely when the study participant began his



production attempts, though he was asked afterward to confirm his attempt, as instructed, during post-trial debriefing. On rare occasions when he denied trial attempts the data were rejected.

CLASSIFICATION ANALYSIS

We compared the performance of linear discriminant analysis (LDA; Fisher, 1936; Rao, 1973), support vector machine (SVM; Vapnik, 1995), and flexible discriminant analysis (FDA; Hastie et al., 1994) classifiers for prediction of the intended phoneme attempted by the participant during the response period. Both SVM and FDA classifiers operate by transforming the feature set into a higher dimensional space via basis-set expansion. In this way, linear discriminant boundaries can be computed in the higher dimensional space which often corresponds to non-linear

boundaries in the true feature space. In addition, both SVM and FDA are particularly suited to high dimensional classification problems and for calculation of decision boundaries with overlapping class distributions (Hastie et al., 2001). FDA is based on a generalization of standard LDA as a series of linear regressions and allows substitution of the regression procedure. For this analysis, we used multivariate adaptive regression splines (MARS; Friedman, 1991) as a non-parametric regression alternative. See Hastie et al. (2001) for a detailed discussion of LDA and generalizations for FDA and SVM algorithms. We used radial-basis kernels for the SVM analysis with cost parameter = 1, and $\gamma = 1/31$ (or 0.032). For all analyses, the response period was divided into 20 equally spaced bins, 300 ms in duration, and the number of spikes were counted per bin and spike event cluster. Ten-fold cross-validation was used to evaluate classifier performance. Specifically, the spike cluster bin-counts were randomly partitioned into 10 mutually exclusive data sets, each with approximately equal phoneme occurrence probabilities, resulting in between 180 and 216 data points per-phoneme. Then, 9 of the 10 data sets were used as input features for all classification algorithms. The final data set was used to obtain untrained test predictions. The weightings associated with phoneme data point imbalance were accounted for in subsequent analyses using bootstrapping methods. This procedure was repeated 10 times, for each combination of nine training sets. All algorithms, LDA, SVM, and FDA were performed using the MASS (Idea; Venables and Ripley, 2002), mda (fda; Leisch

et al., 2009), and e1071 (svm; Dimitriadou et al., 2010) packages of the R statistical programming language (R Development Core Team, 2009).

STATISTICAL ANALYSIS

We computed overall cross-validated test-set classification accuracy and confusion matrices for all classifiers. Additionally, we used a one-tailed (i.e., greater-than) exact binomial test, which uses bootstrapping to obtain empirical probability distributions, to determine which of the 38 phonemes were classified significantly above per-phoneme chance-levels since there was a slight imbalance in class probabilities. Specifically, chance-level performance was approximately 1/38, or 0.0263, though the exact value varied depending on the number of phoneme repetitions (10–12). Phonemes were significantly predicted above chance if Bonferroni-corrected binomial test *p*-values were less than 0.05. All procedures were performed using R and Matlab (Mathworks, Inc., Natick, MA, USA).

The phonemes varied considerably in terms of their phonetic characteristics. In particular we used a combination of monophthong vowels, diphthong and rhotacized vowels, and consonants (stop, fricative, affricate, liquids, and glides). A summary of the characteristics of each phoneme used in this study is found in **Tables 1 and 2**.

We sought to uncover which, if any, articulatory characteristics were important for classifying intended phonemes from neural activity. For instance, steady vowel sounds can be defined according

Table 2 | Consonant stimuli used for the phoneme production experiment.

Phoneme #	ARPABET	IPA	Example	Type	Place	Articulatory features		
						Voicing	Manner	Articulator
16	P	p	Pop	Consonant	Labial	Voiceless	Stop	Lips
17	B	b	Bib	Consonant	Labial	Voiced	Stop	Lips
18	T	t	Teat	Consonant	Alveolar	Voiceless	Stop	Tongue blade
19	D	d	Died	Consonant	Alveolar	Voiced	Stop	Tongue blade
20	K	k	Cake	Consonant	Velar	Voiceless	Stop	Tongue body
21	G	g	Go	Consonant	Velar	Voiced	Stop	Tongue body
22	CH	tʃ	CHin	Consonant	Palatal	Voiceless	Affricate	Tongue blade
23	JH	dʒ	Judge	Consonant	Palatal	Voiced	Affricate	Tongue blade
24	F	f	Fine	Consonant	Labiodental	Voiceless	Fricative	Lips
25	V	v	Vine	Consonant	Labiodental	Voiced	Fricative	Lips
26	TH	θ	THink	Consonant	Dental	Voiceless	Fricative	Tongue blade
27	DH	ð	THen	Consonant	Dental	Voiced	Fricative	Tongue blade
28	S	s	See	Consonant	Alveolar	Voiceless	Fricative	Tongue blade
29	Z	z	Zoo	Consonant	Alveolar	Voiced	Fricative	Tongue blade
30	SH	ʃ	SHy	Consonant	Palatal	Voiceless	Fricative	Tongue blade
31	CH R	tʃr	TReasure	Consonant	Palatal	Voiceless	Affricate	NA
32	M	m	My	Consonant	Labial	Voiced	Nasal	Lips
33	N	n	Near	Consonant	Alveolar	Voiced	Nasal	Tongue blade
34	NG	ŋ	siNG	Consonant	Velar	Voiced	Nasal	Tongue body
35	L	l	Low	Consonant	Lingua-alveolar	Voiced	Liquid: lateral	Tongue blade
36	R	r	Raw	Consonant	Lingua-alveolar	Voiced	Liquid: rhotic	Tongue blade
37	W	w	West	Consonant	Labial-velar	Voiced	Glide	Lips
38	J	j	Year	Consonant	Palatal	Voiced	Glide	Tongue body
39	HH	h	High	Consonant	Glottal	Voiceless	Fricative	NA

to the height and location of the tongue position during vocalization and whether the lips were rounded (as in /u/). Similarly, consonants can be defined by the place and manner of articulation and voicing. The vowel sound /i/ (as in bead) is an example of a vowel sound with high tongue height, front tongue placement and spread lips. Similarly, the consonant /b/ (as in bib) is a bilabial, voiced, stop consonant. For our analysis, we chose to use the primary articulator (e.g., lips for /b/) rather than the place of articulation (e.g., bilabial for /b/) as a factor under the assumption that primary motor and pre-motor cortex are more related to the execution of a speech movement, not necessarily the place of articulation. The primary motor cortex is known to be involved in the recruitment of motor neurons for activating peripheral muscles used during movement execution, in this case the muscles innervating the vocal tract. The pre-motor cortex, according to a neurocomputational model of speech production (Directions into Velocities of Articulators model, Guenther et al., 2006), is involved in the encoding of speech-motor programs. In particular, the pre-motor cortex is used in planning upcoming speech-motor movements. Therefore, our understanding is that both primary motor and pre-motor cortex are involved in various aspects of production rather than tactile feedback (e.g., place of articulation).

We performed a three-way ANOVA using logistic regression of cross-validated prediction successes of the vowel sounds on three factors: height (three levels), location (three levels), and rounding (four levels). We performed the same analysis using prediction successes of consonants on the three factors: manner (four levels; liquids and glides were combined as semivowels), voicing (two levels), and articulator (three levels). It is difficult to characterize diphthong phonemes according to scalar factors; therefore, we did not include them in this analysis. For both ANOVA analyses, a Tukey HSD multiple comparisons test was used to determine which factor levels most contributed to increased classification performance.

RESULTS

LDA CLASSIFICATION PERFORMANCE

The LDA classifier was able to predict the correct phoneme with 16.9% mean accuracy, which is statistically above the theoretical chance-level performance (binomial test, $p < 0.05$). Individually, we found

24 phonemes (shown in **Figure 4**, marked with *) predicted above chance-levels (Bonferroni-corrected $p < 0.05$) determined using the binomial test detailed in Section “Statistical Analysis.” The classifier confusion matrix (**Figure 5**, left) shows non-uniform distribution of accuracy scores among the tested phonemes with low to high accuracy scores indicated in grayscale from white to black. In terms of phonemic category, three of nine vowels, three of six diphthongs, and 18 of 23 consonants were classified above chance performance.

In addition, a separate three-way ANOVA characterizing the articulatory features of the tested phonemes (vowels and consonants only) found significant main effects of tongue height [$F(2,1929) = 32.39, p < 0.001$], tongue location [$F(2,1929) = 41.06, p < 0.001$], and lip rounding [$F(2,1929) = 30.91, p < 0.001$] for vowels; voicing [$F(1,4842) = 105.56, p < 0.001$], manner of articulation [$F(5,4842) = 30.04, p < 0.001$], and primary articulator [$F(2,4842) = 8.67, p < 0.001$] for consonants. These results are summarized in **Tables 3 and 4**.

SVM CLASSIFICATION PERFORMANCE

The overall cross-validation classification accuracy for the SVM classifier was 21.0%, also statistically greater-than the theoretical chance-level performance (binomial test, $p < 0.05$). Inspection of the classifier confusion matrix (shown in **Figure 5**, middle) reveals a number of phonemes were correctly classified at much higher rates. Specifically, we found 24 phonemes were classified significantly greater-than chance ($p < 0.05$, Bonferroni-corrected). The significant phonemes are indicated with an asterisk in **Figure 4**. For this analysis, five of nine vowels, three of six diphthongs, and 16 of 23 consonants were classified above chance performance.

Separate three-way ANOVAs of the articulatory features for tested phonemes (vowels and consonants only) found significant main effects of tongue height [$F(2,1929) = 29.46, p < 0.001$], tongue location [$F(2,1929) = 32.20, p < 0.001$], and lip rounding [$F(2,1929) = 43.07, p < 0.001$] for vowels. The results were the same for consonants, with significant main effects of voicing [$F(1,4842) = 375.56, p < 0.001$], manner of articulation [$F(5,4842) = 122.44, p < 0.001$], and primary articulator [$F(2,4842) = 25.70, p < 0.001$] for consonants. These results are summarized in **Tables 3 and 4**.

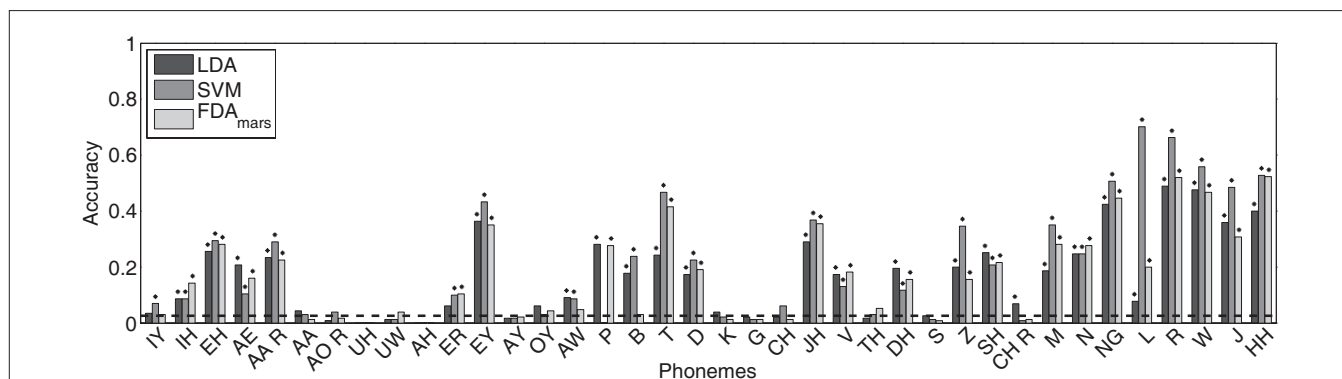


FIGURE 4 | The classification accuracy for each phoneme using the LDA, SVM, and FDA classifiers is shown with the theoretical chance performance level (dashed line, ~1/38). Phonemes with classification rates statistically above chance (binomial test, Bonferroni-adjusted $p < 0.05$) are marked with an asterisk (*). Twenty of 38 phonemes were classified above chance for all methods, and 24 phonemes for at least two of the three classification methods and 26 phonemes for any.

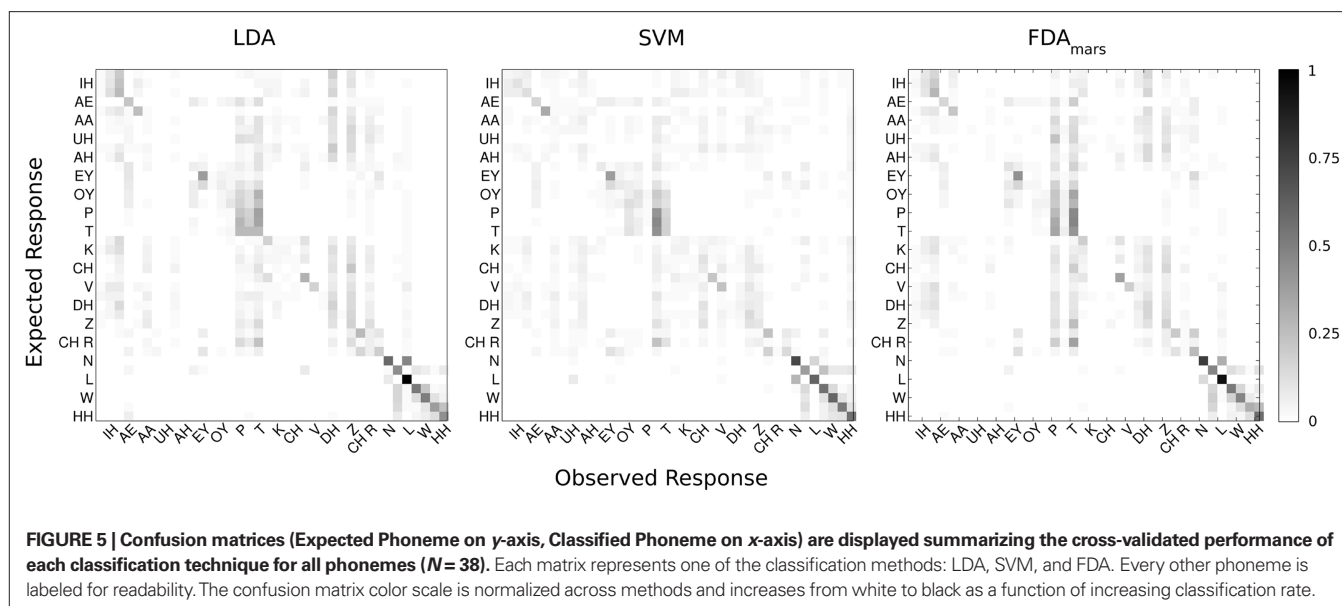


Table 3 | Summary of performance for each classification method.

Number of correctly classified phonemes	LDA	SVM	FDA
Overall accuracy	16.9%	21.0%	17.6%
Vowels (N = 9)	3	5	4
Diphthongs (N = 6)	3	3	2
Consonants (N = 23)	18	16	16
Total (N = 38)	24	24	22

FDA CLASSIFICATION PERFORMANCE

Flexible discriminant analysis classification performance was between that of LDA and SVM with approximately 17.6% accuracy, which is significantly above chance (binomial test, $p < 0.05$). Like both LDA and SVM, a number of phonemes were classified better than others, evidenced by individual accuracy rates (Figure 4) and confusion matrices (Figure 5). According to individual binomial tests, 22 phonemes were classified significantly above chance (Bonferroni-corrected, $p < 0.05$) and are shown in Figure 3. The classification groupings were four of nine vowels, two of six diphthongs, and 16 of 23 consonants classified above chance.

Additional three-way ANOVA of vowel articulatory features resulted in significant main effects of tongue height [$F(2,1929) = 18.46, p < 0.001$], tongue location [$F(2,1929) = 38.31, p < 0.001$], and lip rounding [$F(2,1929) = 35.14, p < 0.001$]. A three-way ANOVA analysis of consonant articulatory features found significant main effects of voicing [$F(1,4842) = 87.19, p < 0.001$], manner of articulation [$F(5,4842) = 44.05, p < 0.001$], and primary articulator [$F(2,4842) = 21.28, p < 0.001$]. These results and subsequent multiple comparisons per level are shown in Tables 3 and 4.

CLASSIFICATION SUMMARY

Classification accuracy between the three methods ranged from 16 to 21% correct (LDA: 16.9%; SVM: 21.0%; FDA: 17.6%); accounting for performance across all 38 phonemes tested in this study. In

addition, classification agreement was high among all three methods with 24 phonemes classified above chance for the LDA and SVM methods and 22 phonemes for the FDA method. This resulted in an overlap of 20 phonemes classified above chance common among all three methods and 26 phonemes for any combination of methods. Details of the classification analysis are shown in Table 3. Examinations of the confusion matrices reveal some systematic misclassification using both the LDA and FDA methods (false positives for phonemes /p/, /t/, /δ/, and /z/). These misclassifications appear to be minimized using the SVM technique.

Based on our understanding of motor cortical representations of speech, it is not reasonable to expect all 38 phonemes to be represented equally in the implanted region. Prior studies confirm our assumption with evidence for a topographical distribution of activities related to the production of different speech sounds (Blakely et al., 2008; Kellis et al., 2010). Our results reflect this inequality in neural representations of speech with non-uniform classification quality across phonemes. We believe future study with additional electrodes will improve spatial coverage of the motor cortex and aid classification of greater numbers of phonemes, resulting in higher overall classification rates.

A closer look at the performance results reveal two primary classification groups – those significantly classified above chance (N = 18) and those below (N = 20). This “pass/fail” perspective can be explained by the cortical representation justification given above, but are also likely influenced by the presence of noise in our processing stream, including spike clustering errors and uncertainty of user response. Improved spike sorting will likely reduce any potential mis-identification of potential spike events. Uncertainty of user response is a difficult problem, as is noted in Section “Experimental Protocol.” It is not possible to obtain instantaneous verification that the subject attempted a trial production, and when the attempt occurred relative to the trial onset. Rather, we can only ask him after the end of a trial whether he attempted a speech production or not. Depending on the response we decide to accept or reject trial data. Even when the subject confirms his production attempt, our classifi-

Table 4 | Vowel and consonant articulatory features are summarized with respect to classification accuracy.

	Tongue height	Tongue location	Lip rounding
VOWEL ARTICULATORY FEATURES			
LDA	Mid OR low > high	Front > central OR back	Neutral OR rounded > spread
SVM	Mid > high OR low	Front > central OR back	Neutral > spread
FDA	Mid > high OR low	Front > central OR back	Neutral OR rounded > spread
CONSONANT ARTICULATORY FEATURES			
	Voicing	Manner of articulation	Primary articulator
LDA	Voiced > voiceless	Glide > nasal OR liquid* > affricate OR fricative OR stop	Lips OR tongue blade > tongue body
SVM	Voiced > voiceless	Liquid > glide > nasal > fricative OR affricate OR stop	Tongue blade > lips OR tongue body
FDA	Voiced > voiceless	Glide OR liquid OR nasal > affricate OR fricative OR stop	Lips OR tongue blade > tongue body

*Difference and directionality are statistically significant (multiple comparisons t-test, Tukey HSD corrected $p < 0.05$). Features joined by "OR" were not statistically different. There were no differences between the "rounded" level and the remaining two levels for the SVM analysis (Vowels: Lip Rounding). *For the LDA analysis, there was no statistical difference between liquids and affricates, but the non-significant effect is in the direction reported.*

cation methods are likely being applied during time segments when he has not yet started, or has already finished the trial. These issues are endemic to experimental study with subjects who are not capable of overt motor behavior. During these time segments, we do not expect the classifiers to correctly predict the appropriate phoneme. That our methods, including potential noise sources (spikes and user response), is capable of predicting phonemes at above-chance rates is impressive and the rates reported represent a conservative, lower-bound on the possible classification/prediction performance.

DISCUSSION

ELECTRODE PLACEMENT AND RECORDINGS

The purpose of this study is to determine whether intended speech production attempts can be accurately classified using neural activity by standard supervised classification techniques (cf. discrete-time filtering; Guenther et al., 2009; Brumberg et al., 2010). In this study, the location of the implant was chosen to maximize the probability that recorded neural activity was related to speech production. This was accomplished by specifying the implant area according to the brain region with highest BOLD response during a pre-surgical fMRI study of picture naming speech production tasks (Bartels et al., 2008). The results of the present study support the use of spike cluster firing rates for discrete classification of intended phonemes.

A major benefit of intracortical recordings is the lack of contamination by motor potentials due to inadvertent movements in contrast to non-invasive methods, such as EEG, which are extremely susceptible to motor and other non-neural artifacts. However, in this study we were concerned about the possibility of motor cortical activity and disruptions to the wireless transmission of neural activity as a result of the participant's uncontrolled muscle spasms. Therefore, we rejected trials in which overt muscle spasms were observed. In addition, our recording setup gave us a unique ability to scan for other more subtle changes in recording potential. Specifically, the participant's involuntary head movements placed strain on the recording setup resulting in a change in signal quality, namely the continuous signal would saturate. Any railing events were automatically detected and trial information rejected for further analysis.

Last, the Neurotrophic Electrode is designed to record extracellular potentials, some of which can be resolved as single-units while others can not, leaving many unresolved multi-unit potentials. Intuitively, single-unit spikes are considered the primary signal of interest in neuroscience and BCI development as they represent the fundamental unit of information transfer in the brain. However, devices for BCI are capable of utilizing any, functionally relevant, information source. In this way, we do not make any distinctions between single and multi-unit sources and rely solely on clustered spike events. Inputs such as these have already been shown to mediate accurate neural control over a continuous vowel synthesizer in a different study with the current participant (Guenther et al., 2009; Brumberg et al., 2010). Our previous results, combined with the statistically significant classification performance reported in the current study, lead us to believe the motor cortical signals described in this study are capable of representing specific speech-motor information related to phoneme productions.

PHONETIC AND ARTICULATORY RELATIONSHIPS

The statistical analysis of classifier performance was divided into two major themes: general accuracy (overall classification and individual phoneme performance) and functional accuracy (performance related to speech sound features). In the first case, we consider the task stimuli as generic discrete items and seek only to determine the overall classifier performance and to find those items classified greater-than chance. In the second, we sought to relate the degree of classification success to stimulus phonetic and articulatory features. The results of the first-pass analysis showed that each method was able to make new classifications (via cross-validation test data set) between 16 and 21% accuracy, which is an order of magnitude greater-than the chance-level performance expected from a 38-class prediction protocol (~2.6%). Further, confusion matrices qualitatively illustrated that a number of stimulus phonemes were well predicted, indicated by large values along the diagonal of the confusion matrix. Finally, we tested each phoneme against the number of successes expected by chance (binomial test) given the number of true occurrences of each. From this analysis, we found that 24 of 38 phonemes were predicted above chance using the LDA classification algorithm, 24 of 38 for SVM and 22 of 38 for FDA. These results indicate that the classification techniques used in this study are capable of

predicting many phonemes from recorded neural data. Further, these results help prove the feasibility of discrete classification techniques for future speech production brain–computer interfaces.

Importantly, these individual phoneme results do not consider any common factors between speech sounds. The second-pass analysis used phonetic and articulatory features of speech sounds to evaluate commonalities underlying classifier performance. In particular, we used two major axes of distinction to help understand the possible similarities between sounds that were classified with high accuracy. The first axis was used to group stimulus phonemes into three major categories: vowels, diphthongs, and consonants. In this nomenclature, vowels are monophthongs with static articulatory configurations and sustained phonation. Diphthongs typically refer to sustained phonation, vowel–vowel combinations such as the /aɪ/ in bite. This diphthong is produced by an initial /a/ which quickly transitions into an /ɪ/. Consonants in this context are actually CV pairs, as many are difficult to pronounce without a trailing vowel sound. Consonants require higher degrees of articulatory movements to form constrictions, or closures of the vocal tract. Typically, consonants are short duration, ballistic movements, but some types (fricatives and nasals) can be produced with sustained phonation. We grouped the phonemes classified above chance according to phonetic type and found, across classifiers, a larger percentage of consonants were predicted than vowels and diphthongs.

A further breakdown of articulatory features common among the phonemes is possible. These categories can be refined almost *ad infinitum*, but such groupings would leave just one example phoneme per category. Therefore, we used a compromise grouping definition. For vowels, the articulatory features were broken into tongue height, tongue location, and lip rounding categories. Tongue height has three levels: high, mid, or low, tongue location three levels: front, central and back, and lip rounding three levels: rounded, neutral, and spread. Consonants were grouped according to voicing, manner of articulation, and primary articulator. The voicing factor has two levels, voiced and voiceless while the manner factor has five levels: stop, fricative, affricate, nasal, glide, and liquid. The final consonant category, primary articulator, is not conventional for standard articulatory phonetics analysis. More common is the place of articulation, or where the articulators meet to create obstructions or constrictions. However, we wanted to emphasize the motor execution involved in speech production, rather than the spatial, tactile consequences. Therefore, the primary articulator factor has three levels: tongue body, tongue blade, and lips. Voiced and voiceless consonants typically occur in pairs for each manner and place of articulation (e.g., /b/ and /p/), with voiced consonants characterized by a pre-release vocalization (see Stevens, 2000 for a more complete discussion). Diphthongs were not grouped according to articulatory features since it is difficult to characterize dynamic speech sounds using the categories used for vowels and consonants. For instance, any diphthong will require at least two classifications for the beginning and end of the utterance.

Statistical analysis of these three factors for vowels and consonants showed that all were significant main effects. In particular, three comparisons were most interesting. First, for vowel sounds, those with a front tongue position were classified best. Second, voiced consonants were classified better than voiceless. Last, lip movements appear to be important for both consonants and

vowels. Specifically, rounded vowels and those consonants primarily involving the lips were among those classified with the highest rates. Further analysis of the tongue placement, lip movement, and voicing effects are needed before any conclusions can be made. However, this initial finding suggests that the implant location may represent lip movements (among other movement types).

COMPARISON OF CLASSIFICATION TECHNIQUE

The three classification methods used in this study resulted in very similar classification rates and were able to correctly classify greater-than expected due to chance, 24, 24, and 22 individual phonemes for LDA, SVM, and FDA respectively. Twenty of 38 phonemes were successfully recognized above chance by all three classifiers, while 24 phonemes were common amongst at least two of the methods. The most important finding here is that standard classification techniques can be reasonably applied to the problem of predicting attempted productions of phonemes from neural activity. All three classification methods employed in this study are capable of real-time use assuming model parameters are estimated beforehand. The task of assigning class labels is not computationally intensive and is well within the limits of modern computing. Future studies using multiple Neurotrophic Electrodes (increasing the signal dimensionality) may benefit from sophisticated approaches such as SVM classifiers. However, if future results continue to show that linear methods perform as well as more sophisticated methods, then such results may help streamline algorithm development in favor of the most parsimonious with the highest potential for efficient real-time performance.

UTILITY FOR FUNCTIONAL BMI FOR SPEECH SOUND PREDICTION

The goal of the presented study was to examine the usefulness of supervised classification techniques for off-line prediction of speech sounds from neural activity in the speech-motor cortex recorded using the Neurotrophic Electrode. On its own, this study did not evaluate the tested methods for usefulness in a functional BMI. However, the results presented here are immediately applicable to use in online applications. A slightly more nuanced question can be posed with respect to this type of analysis for a viable speech production device. Though the results presented here show statistically significant performance for classifying speech sounds, the accuracy rates are truly not great enough to be useful as a communication device. However, we believe this concern can be alleviated by increasing the number of recording channels and/or features used for classification. For instance, implantation of multiple electrodes along the motor cortex (possibly other locations in the speech production network as well) will both increase the number of neural signals and broaden the scope of representation beyond just a single point in motor cortex. This is supported by evidence of local topographical variability in motor cortex activations found correlated to different phoneme and word productions in prior micro-ECoG decoding investigations (Blakely et al., 2008; Kellis et al., 2010). An increase of the number of recording surfaces within each electrode cone will also provide greater numbers of recording channels. We are currently investigating improvements to the electrode design increasing the number of recording contacts, and wireless telemetry for transmission of recordings from multiple electrode implants.

The results presented here, in combination with these previous findings of word and phoneme decoding research validate the approach for discrete speech sound prediction as a potential mode for BMI-based

communication. An often cited, major drawback of speech sound classification is the dictionary size needed to represent all words and phrases used in conversational speech (Rabiner and Juang, 1993). Our prior speech BMI research avoids this limitation by directly decoding the dynamic features needed for control of a speech synthesizer (Guenther et al., 2009; Brumberg et al., 2010). The approach taken here and in Blakely et al., 2008 attempts to resolve the problem of dictionary size for classification methods by decoding phonemes, the smallest unit of speech from which all other speech units are built. Decoding at the level of phonemes allows for generalized speech outputs from combinations of phonemes including syllables, words, and phrases.

Another potential concern, raised by Kellis et al. (2010), suggests relatively long prediction delays (500 ms in that study, 300 ms in the present case) may be too slow for conversational speech. They further report smaller time windows resulted in lower classification accuracy but hypothesize that more sophisticated methods including SVM classifiers and hidden Markov models (HMMs) may overcome these performance deficits. HMMs have great potential given their common usage in standard ASR (Rabiner and Juang, 1993) and are routinely used in real-time transcription of voiced speech. We agree dynamic models (those that incorporate prior state information) may improve classification accuracy, especially in conversational situations, and are an active area of future work (Matthews et al., 2010). A combination of the classification techniques presented here and filter-style vowel synthesis previously investigated by our group (Guenther et al., 2009; Brumberg et al., 2010) may lead to even greater potential in a functional, real-time BMI for speech-based communication.

REFERENCES

- Bartels, J. L., Andreasen, D., Ehirim, P., Mao, H., Seibert, S., Wright, E. J., and Kennedy, P. R. (2008). Neurotrophic electrode: method of assembly and implantation into human motor speech cortex. *J. Neurosci. Methods* 174, 168–176.
- Birbaumer, N., Ghanayim, N., Hinterberger, T., Iversen, I., Kotchoubey, B., Kubler, A., Perelmouter, J., Taub, E., and Flor, H. (1999). A spelling device for the paralysed. *Nature* 398, 297–298.
- Birbaumer, N., Kubler, A., Ghanayim, N., Hinterberger, T., Perelmouter, J., Kaiser, J., Iversen, I., Kotchoubey, B., Neumann, N., and Flor, H. (2000). The thought translation device (TTD) for completely paralyzed patients. *IEEE Trans. Rehabil. Eng.* 8, 190–193.
- Blakely, T., Miller, K. J., Rao, R., Holmes, M., and Ojemann, J. G. (2008). “Localization and classification of phonemes using high spatial resolution electrocorticography (ECoG) grids,” in *Conference Proceedings of the IEEE Engineering in Medicine and Biology Society, 2008. 30th Annual International Conference of the IEEE EMBS 2008*, Vancouver, August 20–24, 2008, 4964–4967.
- Blankertz, B., Krauledat, M., Dornhege, G., Williamson, J., Murray-Smith, R., and Müller, K.-R. (2007). “A note on brain actuated spelling with the Berlin brain-computer interface,” in *Universal Access in Human-Computer Interaction. Ambient Interaction*, ed. C. Stephanidis (Heidelberg: Springer), 759–768.
- Brumberg, J. S., Kennedy, P. R., and Guenther, F. H. (2009). “Artificial speech synthesizer control by brain-computer interface,” in *Proceedings of the 10th Annual Conference of the International Speech Communication Association*. (Brighton: International Speech Communication Association).
- Brumberg, J. S., Nieto-Castanon, A., Kennedy, P. R., and Guenther, F. H. (2010). Brain-computer interfaces for speech communication. *Speech Commun.* 52, 367–379.
- Carmena, J. M., Lebedev, M. A., Crist, R. E., O’Doherty, J. E., Santucci, D. M., Dimitrov, D. F., Patil, P. G., Henriquez, C. S., and Nicolelis, M. A. L. (2003). Learning to control a brain-machine interface for reaching and grasping by primates. *PLoS Biol.* 1, e42. doi: 10.1371/journal.pbio.0000042
- Chapin, J. K., Moxon, K. A., Markowitz, R. S., and Nicolelis, M. A. L. (1999). Real-time control of a robot arm using simultaneously recorded neurons in the motor cortex. *Nat. Neurosci.* 2, 664–670.
- Cheng, M., Gao, X., Gao, S., and Xu, D. (2002). Design and implementation of a brain-computer interface with high transfer rates. *IEEE Trans. Biomed. Eng.* 49, 1181–1186.
- Crone, N. E., Boatman, D., Gordon, B., and Hao, L. (2001a). Induced electrocorticographic gamma activity during auditory perception. *Clin. Neurophysiol.* 112, 565–582.
- Crone, N. E., Hao, L., Hart, J., Boatman, D., Lesser, R. P., Irizarry, R., and Gordon, B. (2001b). Electrocorticographic gamma activity during word production in spoken and sign language. *Neurology* 57, 2045–2053.
- Denby, B., Schultz, T., Honda, K., Hueber, T., Gilbert, J. M., and Brumberg, J. S. (2010). Silent speech interfaces. *Speech Commun.* 52, 270–287.
- Dimitriadou, E., Hornik, K., Leisch, F., Meyer, D., and Weingessel, A. (2010). *e1071: Misc Functions of the Department of Statistics (e1071)*, TUWien. Available at: <http://cran.r-project.org/web/packages/e1071/index.html>
- Doble, J. E., Haig, A. J., Anderson, C., and Katz, R. (2003). Impairment, activity, participation, life satisfaction, and survival in persons with locked-in syndrome for over a decade: follow-up on a previously reported cohort. *J. Head Trauma Rehabil.* 18, 435–444.
- Donchin, E., Spencer, K., and Wijesinghe, R. (2000). The mental prosthesis: assessing the speed of a P300-based brain-computer interface. *IEEE Trans. Rehabil. Eng.* 8, 174–179.
- Donoghue, J. P., Nurmikko, A., Black, M., and Hochberg, L. R. (2007). Assistive technology and robotic control using motor cortex ensemble-based neural interface systems in humans with tetraplegia. *J. Physiol.* 579, 603–611.
- Farwell, L., and Donchin, E. (1988). Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalogr. Clin. Neurophysiol.* 70, 510–523.
- Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Ann. Eugen.* 7, 179–188.
- Friedman, J. H. (1991). Multivariate adaptive regression splines. *Ann. Stat.* 19, 1–67.
- Friman, O., Luth, T., Volosyak, I., and Graser, A. (2007). “Spelling with steady-state visual evoked potentials,” in *2007 3rd International IEEE/EMBS Conference on Neural Engineering (Kohala Coast, HI: IEEE)*, 354–357.
- Guenther, F. H., Brumberg, J. S., Wright, E. J., Nieto-Castanon, A., Tourville, J. A., Panko, M., Law, R., Siebert, S. A., Bartels, J. L., Andreasen, D. S., Ehirim, P., Mao, H., and Kennedy, P. R. (2009). A wireless brain-machine interface for real-time speech synthesis. *PLoS ONE* 4, e8218. doi: 10.1371/journal.pone.0008218
- Guenther, F. H., Ghosh, S. S., and Tourville, J. A. (2006). Neural modeling and

CONCLUSION

The study reported here was conducted to determine if classification of intended phoneme productions was possible using neural activity from a chronically implanted microelectrode in human speech-motor cortex. The results show that such classification is indeed possible with overall classification rates of between 16 and 21% (above chance for a 38-class problem). Further, we found that certain phonemes were classified at greatly higher rates than others, though these results are not completely understood at this time. An exploratory analysis of articulatory features found that certain articulators may have played a role in phoneme classification (in this case, the lips). These findings may be specific to the current subject and study, and may not generalize well to others; however, they indicate the importance of articulatory analysis of putative speech sound stimuli as well as the representation of potential implantation sites.

ACKNOWLEDGMENTS

Supported in part by the National Institute of Deafness and other Communication Disorders (R44 DC007050-02; RO1 DC002852) and in part by CELEST, a National Science Foundation Science of Learning Center (NSF SMA-0835976). The authors thank the participant and his family for their continued participation. The authors also thank Jess Bartels and Steven Seibert for their assistance in administering the experimental paradigm, Dr. Princewill Ehrim for conducting the implant surgery, and Dr. Hui Mao for radiological evaluation.

- imaging of the cortical interactions underlying syllable production. *Brain Lang.* 96, 280–301.
- Hastie, T., Tibshirani, R., and Buja, A. (1994). Flexible discriminant analysis by optimal scoring. *J. Am. Stat. Assoc.* 89, 1255–1270.
- Hastie, T., Tibshirani, R., and Friedman, J. H. (2001). *The Elements of Statistical Learning*. New York: Springer, 560.
- Hochberg, L. R., Serruya, M. D., Friehe, G. M., Mukand, J. A., Saleh, M., Caplan, A. H., Branner, A., Chen, D., Penn, R. D., and Donoghue, J. P. (2006). Neuronal ensemble control of prosthetic devices by a human with tetraplegia. *Nature* 442, 164–171.
- Hoogerwerf, A., and Wise, K. (1994). A three-dimensional microelectrode array for chronic neural recording. *IEEE Trans. Biomed. Eng.* 41, 1136–1146.
- Jeannerod, M., and Frak, V. (1999). Mental imaging of motor activity in humans. *Curr. Opin. Neurobiol.* 9, 735–739.
- Jones, K., Campbell, P., and Normann, R. (1992). A glass/silicon composite intracortical electrode array. *Ann. Biomed. Eng.* 20, 423–437.
- Kellis, S., Miller, K., Thompson, K., Brown, R., House, P., and Greger, B. (2010). Decoding spoken words using local field potentials recorded from the cortical surface. *J. Neural Eng.* 7, 056007.
- Kennedy, P. R. (1989). The cone electrode: a long-term electrode that records from neurites grown onto its recording surface. *J. Neurosci. Methods* 29, 181–193.
- Kennedy, P. R. (2006). “Comparing electrodes for use as cortical control signals: tiny tines or tiny cones on wires: which is best?” in *The Biomedical Engineering Handbook*, 3rd Edn, ed. J. D. Brazino (Boca Raton: CRC/Taylor & Francis), 32.1–32.14.
- Kennedy, P. R., and Bakay, R. A. E. (1998). Restoration of neural output from a paralyzed patient by direct brain connection. *Neuroreport* 9, 1707–1711.
- Kennedy, P. R., Bakay, R. A. E., Moore, M. M., Adams, K., and Goldwaite, J. (2000). Direct control of a computer from the human central nervous system. *IEEE Trans. Rehabil. Eng.* 8, 198–202.
- Kennedy, P. R., Bakay, R. A. E., and Sharpe, S. M. (1992a). Behavioral correlates of action potentials recorded chronically inside the cone electrode. *Neuroreport* 3, 605–608.
- Kennedy, P. R., Mirra, S. S., and Bakay, R. A. E. (1992b). The cone electrode: ultrastructural studies following long-term recording in rat and monkey cortex. *Neurosci. Lett.* 142, 89–94.
- Kennedy, P. R., Kirby, T., Moore, M. M., King, B., and Mallory, A. (2004). Computer control using human intracortical local field potentials. *IEEE Trans. Neural Syst. Rehabil. Eng.* 12, 339–344.
- Lacourse, M. G., Cohen, M. J., Lawrence, K. E., and Romero, D. H. (1999). Cortical potentials during imagined movements in individuals with chronic spinal cord injuries. *Behav. Brain Res.* 104, 73–88.
- Leisch, F., Hornik, K., and Ripley, B. D. (2009). *Mda: Mixture and Flexible Discriminant Analysis. S Original by Trevor Hastie and Robert Tibshirani*. Available at: <http://cran.r-project.org/web/packages/mda/index.html>
- Leuthardt, E. C., Schalk, G., Wolpaw, J. R., Ojemann, J. G., and Moran, D. W. (2004). A brain-computer interface using electrocorticographic signals in humans. *J. Neural Eng.* 1, 63–71.
- Lewicki, M. S. (1998). A review of methods for spike sorting: the detection and classification of neural action potentials. *Network* 9, R53–R78.
- Matthews, B., Kim, J., Brumberg, J. S., and Clements, M. (2010). “A probabilistic decoding approach to a neural prosthesis for speech,” in *Proceedings of the 2010 4th Annual Conference on Bioinformatics and Biomedical Engineering (iCBBE)*, Chengdu, June 18–20, 2010.
- Maynard, E. M., Nordhausen, C. T., and Normann, R. A. (1997). The Utah intracortical electrode array: a recording structure for potential brain-computer interfaces. *Electroencephalogr. Clin. Neurophysiol.* 102, 228–239.
- Neuper, C., Müller, G. R., Kübler, A., Birbaumer, N., and Pfurtscheller, G. (2003). Clinical application of an EEG-based brain-computer interface: a case study in a patient with severe motor impairment. *Clin. Neurophysiol.* 114, 399–409.
- Neuper, C., Scherer, R., Reiner, M., and Pfurtscheller, G. (2005). Imagery of motor actions: differential effects of kinesthetic and visual-motor mode of imagery in single-trial EEG. *Cogn. Brain Res.* 25, 668–677.
- Nicolelis, M. A. L., Dimitrov, D., Carmena, J. M., Crist, R., Lehw, G., Kralik, J. D., and Wise, S. P. (2003). Chronic, multisite, multi-electrode recordings in macaque monkeys. *Proc. Natl. Acad. Sci. U.S.A.* 100, 11041–11046.
- Obermaier, B., Müller, G., and Pfurtscheller, G. (2003). “Virtual keyboard” controlled by spontaneous EEG activity. *IEEE Trans. Neural Syst. Rehabil. Eng.* 11, 422–426.
- Penfield, W., and Roberts, L. (1959). *Speech and Brain Mechanisms*. Princeton, NJ: Princeton University Press.
- Plum, F., and Posner, J. B. (1972). The diagnosis of stupor and coma. *Contemp. Neurol. Ser.* 10, 1–286.
- R Development Core Team. (2009). *R: A Language and Environment for Statistical Computing*, Vienna. Available at: <http://www.R-project.org>
- Rabiner, L., and Juang, B. H. (1993). *Fundamentals of Speech Recognition*. Upper Saddle River, NJ: Prentice-Hall, Inc.
- Rao, C. R. (1973). *Linear Statistical Inference and its Applications*. New York: Wiley.
- Roth, M., Decety, J., Raybaudi, M., Massarelli, R., Delon-Martin, C., Segebarth, C., Morand, S., Gemignani, A., Decors, M., and Jeannerod, M. (1996). Possible involvement of primary motor cortex in mentally simulated movement: a functional magnetic resonance imaging study. *Neuroreport* 7, 1280–1284.
- Sadhir, R., James, W., Yasuda, H., Sharma, A., Nichols, M., and Haln, A. (1981). The adhesion of glow-discharge polymers, Silastic and Parylene to implantable platinum electrodes: results of tensile pull tests after exposure to isotonic sodium chloride. *Biomaterials* 2, 239–243.
- Scherer, R., Müller, G., Neuper, C., Graimann, B., and Pfurtscheller, G. (2004). An asynchronously controlled EEG-based virtual keyboard: improvement of the spelling rate. *IEEE Trans. Biomed. Eng.* 51, 979–984.
- Schwartz, A. B. (2004). Cortical neural prosthetics. *Annu. Rev. Neurosci.* 27, 487–507.
- Serruya, M. D., Hatsopoulos, N. G., Paninski, L., Fellows, M. R., and Donoghue, J. P. (2002). Instant neural control of a movement signal. *Nature* 416, 141–142.
- Stevens, K. N. (2000). *Acoustic Phonetics*. Cambridge, MA: MIT Press.
- Sutter, E. E. (1992). The brain response interface: communication through visually-induced electrical brain responses. *J. Microcomput. Appl.* 15, 31–45.
- Taylor, D. M., Tillery, S. I., and Schwartz, A. B. (2002). Direct cortical control of 3D neuroprosthetic devices. *Science* 296, 1829–1832.
- Towle, V. L., Yoon, H., Castelle, M., Edgar, J. C., Biassou, N. M., Frim, D. M., Spire, J., and Kohrman, M. H. (2008). ECoG gamma activity during a language task: differentiating expressive and receptive speech areas. *Brain* 131, 2013–2027.
- Vapnik, V. N. (1995). *The Nature of Statistical Learning Theory*. New York: Springer Verlag.
- Vaughan, T., McFarland, D., Schalk, G., Sarnacki, W., Krusienski, D., Sellers, E., and Wolpaw, J. (2006). The wadsworth BCI research and development program: at home with BCI. *IEEE Trans. Neural Syst. Rehabil. Eng.* 14, 229–233.
- Velliste, M., Perel, S., Spalding, M. C., Whitford, A. S., and Schwartz, A. B. (2008). Cortical control of a prosthetic arm for self-feeding. *Nature* 453, 1098–1101.
- Venables, W. N., and Ripley, B. D. (2002). *Modern Applied Statistics with S. Fourth Edition*. New York: Springer.
- Ward, M. P., Rajdev, P., Ellison, C., and Irazoqui, P. P. (2009). Toward a comparison of microelectrodes for acute and chronic recordings. *Brain Res.* 1282, 183–200.
- Williams, J. C., Rennaker, R. L., and Kipke, D. R. (1999). Long-term neural recording characteristics of wire microelectrode arrays implanted in cerebral cortex. *Brain Res. Brain Res. Protoc.* 4, 303–313.
- Wise, K. D., Angell, J. B., and Starr, A. (1970). An integrated-circuit approach to extracellular microelectrodes. *IEEE Trans. Biomed. Eng.* 17, 238–247.
- Wolpaw, J. R., and McFarland, D. J. (2004). Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans. *Proc. Natl. Acad. Sci. U.S.A.* 101, 17849–17854.

Conflict of Interest Statement: Philip Kennedy is a 98% owner and Dinal Andreasen is a 2% owner of Neural Signals Inc., and may derive some financial benefit from this work at some time in the future. Authors Jonathan Brumberg, Frank Guenther and E. Joe Wright have no financial relationships that could be construed as a potential conflict of interest.

Received: 17 September 2010; accepted: 24 April 2011; published online: 12 May 2011.
Citation: Brumberg JS, Wright EJ, Andreasen DS, Guenther FH and Kennedy PR (2011) Classification of intended phoneme production from chronic intracortical microelectrode recordings in speech-motor cortex. *Front. Neurosci.* 5:65. doi: 10.3389/fnins.2011.00065

This article was submitted to *Frontiers in Neuroprosthetics*, a specialty of *Frontiers in Neuroscience*.

Copyright © 2011 Brumberg, Wright, Andreasen, Guenther and Kennedy. This is an open-access article subject to a non-exclusive license between the authors and Frontiers Media SA, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and other Frontiers conditions are complied with.