Decoding vowels and consonants in spoken and imagined words using electrocorticographic signals in humans

2011 J. Neural Eng. 8 046028

(http://iopscience.iop.org/1741-2552/8/4/046028)

J. Neural Eng. **8** (2011) 046028 (11pp) doi:10.1088/1741-2560/8/4/046028

Decoding vowels and consonants in spoken and imagined words using electrocorticographic signals in humans

Xiaomei Pei[1], **Dennis L Barbour**[2], **Eric C Leuthardt**[2,3] **and Gerwin Schalk**[1,3,4,5,6,7]

[1] Brain–Computer Interface R&D Program, Wadsworth Center, New York State Department of Health, Albany, NY, USA
[2] Department of Biomedical Engineering, Washington University, St Louis, MO, USA
[3] Department of Neurological Surgery, Washington University, St Louis, MO, USA
[4] Department of Neurology, Albany Medical College, Albany, NY, USA
[5] Department of Biomedical Engineering, Rensselaer Polytechnic Institute, Troy, NY, USA
[6] Department of Biomedical Sciences, State University of New York, Albany, NY, USA

E-mail: schalk@wadsworth.org

Received 10 August 2010
Accepted for publication 2 March 2011
Published 13 July 2011
Online at stacks.iop.org/JNE/8/046028

Abstract
Several stories in the popular media have speculated that it may be possible to infer from the brain which word a person is speaking or even thinking. While recent studies have demonstrated that brain signals can give detailed information about actual and imagined actions, such as different types of limb movements or spoken words, concrete experimental evidence for the possibility to 'read the mind', i.e. to interpret internally-generated speech, has been scarce. In this study, we found that it is possible to use signals recorded from the surface of the brain (electrocorticography) to discriminate the vowels and consonants embedded in spoken and in imagined words, and we defined the cortical areas that held the most information about discrimination of vowels and consonants. The results shed light on the distinct mechanisms associated with production of vowels and consonants, and could provide the basis for brain-based communication using imagined speech.

S Online supplementary data available from stacks.iop.org/JNE/8/046028/mmedia

(Some figures in this article are in colour only in the electronic version)

1. Introduction

Recent studies have shown that brain–computer interface (BCI) systems can use brain signals that are usually related to motor movements or motor imagery [1–5] to select from different characters or words [6–10]. While this approach is effective, it has distinct limitations that include a relatively slow communication rate and extensive subject training. This training requirement could be reduced, and perhaps BCI performance further increased if it was possible to directly, i.e. without the use of intermediate choices (such as selection of characters out of a group), determine what specific word the users wished to communicate through their brain signals [11]. However, compared to brain signals in the motor system, which are often governed by relatively simple (e.g., linear or cosine) relationships with parameters of movements, language processing appears to be more complex. It involves a widely distributed neural network of distinct cortical areas that are engaged in phonological or semantic analysis, speech production, and other processes [12–14]. Nevertheless, recent studies have begun to elucidate the relationship of brain activity with different aspects of a receptive or expressive, auditory or articulatory language

[7] Author to whom any correspondence should be addressed.

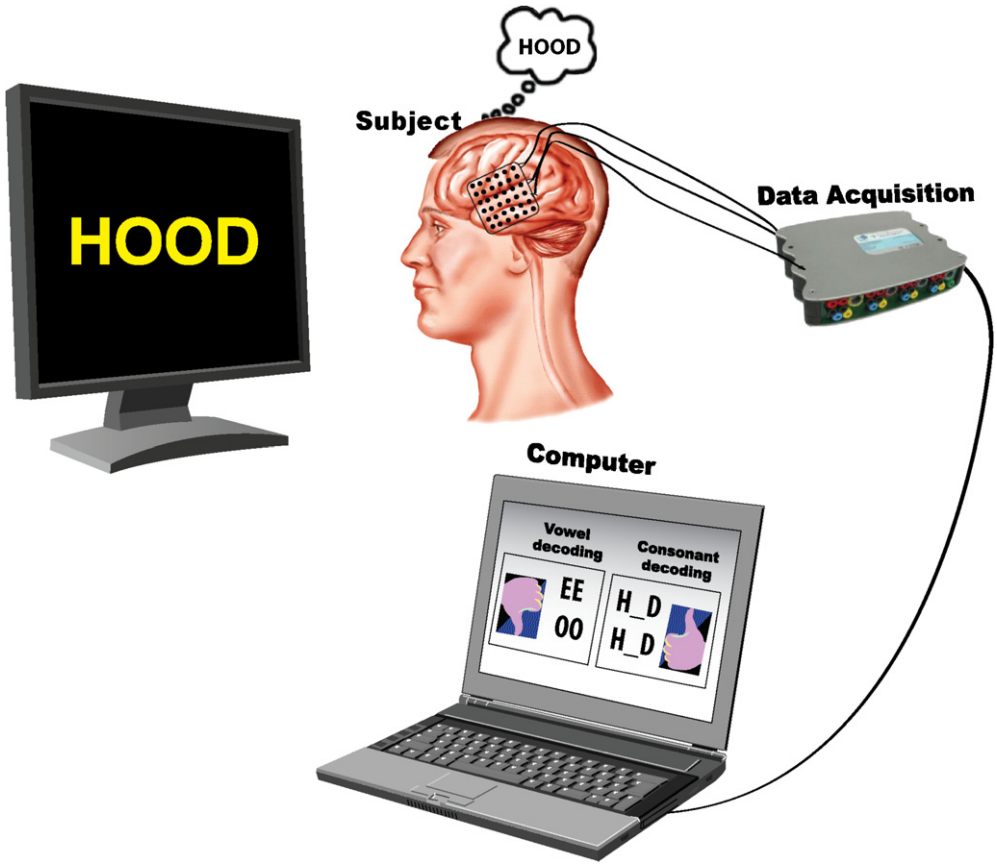

Figure 1. Schematic diagram of the experimental setup.

function [15–19]. For example, functional magnetic resonance imaging (fMRI) of auditory and other cortices was shown to hold information about different monophthongs (i.e. /a/, /i/, /u/) that the subjects listened to [17], scalp-recorded electroencephalography (EEG) was shown to hold information about the rhythm of syllables [19] and some information about individual vowels (i.e. /a/, /u/) [18], and electrocorticography (ECoG) was used to decode several spoken words [20]. However, concrete evidence that brain signals could allow for the identification of components of words has remained largely elusive. Identification of the neural correlates of speech function could allow for determination of those cortical areas that allow for discrimination of words or their components. While previous studies have already demonstrated evidence for the neural basis of differential processing of vowels and consonants [21–25], this evidence has either been indirect or inconsistent, and has not pinpointed the anatomical location of consonant–vowel dissociation.

ECoG recordings from the surface of the brain have recently attracted increasing attention because they combine relatively high spatial with high temporal resolution. Some of these ECoG-based studies [26–33] have begun to investigate neural correlates of speech processing. These and other studies [34–38] consistently showed that ECoG amplitude over anatomically appropriate areas decreased during a task in mu (8–12 Hz) and beta (18–26 Hz) frequency bands and increased in gamma (>40 Hz) bands. Other studies [39–44] have shown that this information in ECoG can be used to reconstruct or map the different aspects of motor or language function.

In this study, we show for the first time that it is possible to decode vowels and consonants that are embedded in spoken or imagined monosyllabic words from ECoG signals in humans, and also characterize the cortical substrates involved in the discrimination within distinct vowels and consonants, respectively.

2. Methods

2.1. Subjects

The subjects in this study were eight patients with intractable epilepsy who underwent temporary placement of subdural electrode arrays to localize seizure foci prior to surgical resection. They included two men (subjects F and I) and six women (subjects A, B, C, D, E, and G). (See table 1 for additional information.) All gave informed consent for the study, which was approved by the Institutional Review Board of Washington University School of Medicine and the Human Research Protections Office of the US Army Medical Research and Materiel Command. Each subject had an electrode grid (48 or 64 contacts) placed over frontal, parietal and temporal regions (see figure 1 for general system setup and figure 2 for approximate electrode locations). Grid placement and duration of ECoG monitoring were based solely on the requirements of the clinical evaluation, without

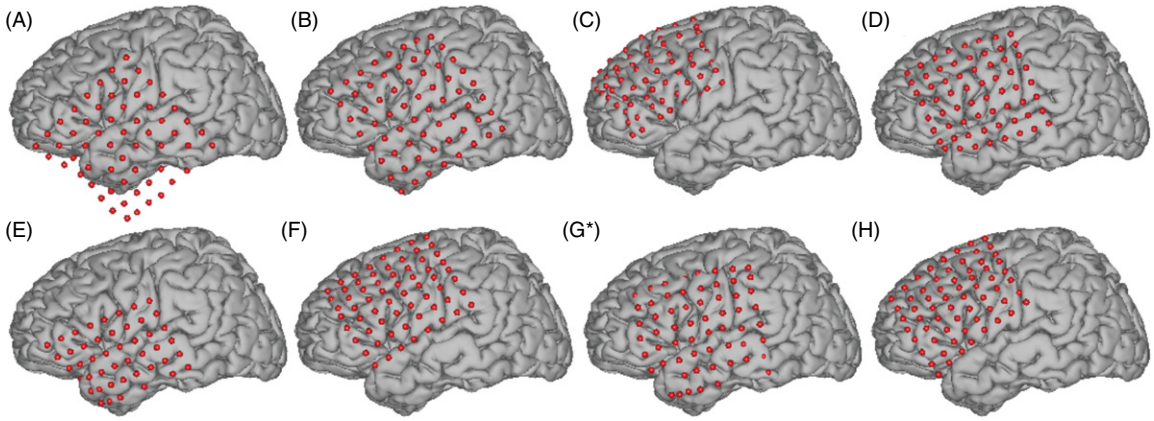

Figure 2. Electrode locations in the eight subjects. Electrodes were projected onto the left hemisphere for subject G.

Table 1. Clinical profiles.

Subject	Age	Sex	Handedness	Grid location	Tasks
A	16	F	R	Left frontal-parietal-temporal	Overt/covert word repetition
B	44	F	L	Left frontal-parietal-temporal	Overt/covert word repetition
C	58	F	R	Left frontal	Overt/covert word repetition
D	48	F	R	Left frontal	Overt/covert word repetition
E	49	F	R	Left frontal-parietal-temporal	Overt/covert word repetition
F	55	M	R	Left frontal-parietal-temporal	Overt/covert word repetition
G	47	F	R	Right frontal-parietal-temporal	Overt word repetition
I	44	M	R	Left frontal	Overt word repetition

any consideration of this study. As shown in figure 2, the location of the implanted grid varied across subjects. These grids consisted of flat electrodes with an exposed diameter of 2.3 mm and an inter-electrode distance of 1 cm, and were implanted for about 1 week. The electrodes for all subjects except subject G were localized over the left hemisphere. Following the placement of the subdural grid, each subject had postoperative anterior–posterior and lateral radiographs to verify grid location.

2.2. Experimental paradigm

During the study, each subject was in a semi-recumbent position in a hospital bed about 1 m from a video screen. In separate experimental runs, ECoG was recorded during two different conditions: overt or covert word repetition in response to visual word stimuli. Throughout the paper, we will refer to these two tasks as 'Overt' and 'Covert.' The visual stimuli consisted of 36 monosyllable words that were presented on a video monitor for 4 s, followed by a break of 0.5 s during which the screen was blank. Each of these 36 words was composed of one of four different vowels (i.e. /ɛ/, /æ/, /i:/ and /u:/, which are well separable in formant space) and one of nine consonant pairs (i.e. /b_t/, /c_n/, /h_d/, /l_d/, /m_n/, /p_p/, /r_d/, /s_t/, /t_n/, which were chosen to create actual words, rather than pseudowords, in combination with the vowels). These vowels and consonants were integrated in a consonant–vowel–consonant (CVC) structure (see table 2 for the list of all words.) This structure allowed us to group the words based on either the vowel within the word (_V_) or the leading/trailing consonant pair (C_C). In other words, each word was uniquely identified by its vowel and consonant pair.

Table 2. List of word stimuli.

CVC	b_t	c_n	h_d	l_d	m_n	p_p	r_d	s_t	t_n
/ɛ/	bet	ken	head	led	men	pep	red	set	ten
/æ/	bat	can	had	lad	man	pap	rad	sat	tan
/i:/	beat	keen	heed	lead	mean	peep	read	seat	teen
/u:/	boot	coon	hood	lewd	moon	poop	rood	soot	toon

2.3. Data collection

In all experiments, we recorded ECoG from the electrode grid using the general-purpose software BCI2000 [45, 46] that was connected to five g.USBamp amplifier/digitizer systems (g.tec, Graz, Austria). Simultaneous clinical monitoring was achieved using a connector that split the cables coming from the subject into one set that was connected to the clinical monitoring system and another set that was connected to the BCI2000/g.USBamp system. Thus, at no time was clinical care or clinical data collection affected. All electrodes were referenced to an inactive electrode. In a subset of subjects (B, C, D, E, F), the verbal response was recorded using a microphone; in the remaining subjects, speech onset was detected using the g.TRIGbox (g.tec, Graz, Austria). ECoG signals and the microphone signal were amplified, bandpass filtered (0.15–500 Hz), digitized at 1200 Hz, and stored by BCI2000. We collected two to seven experimental runs from each subject for each of the two conditions (i.e. overt or covert word repetition). Each run included 36 trials (140 trials total

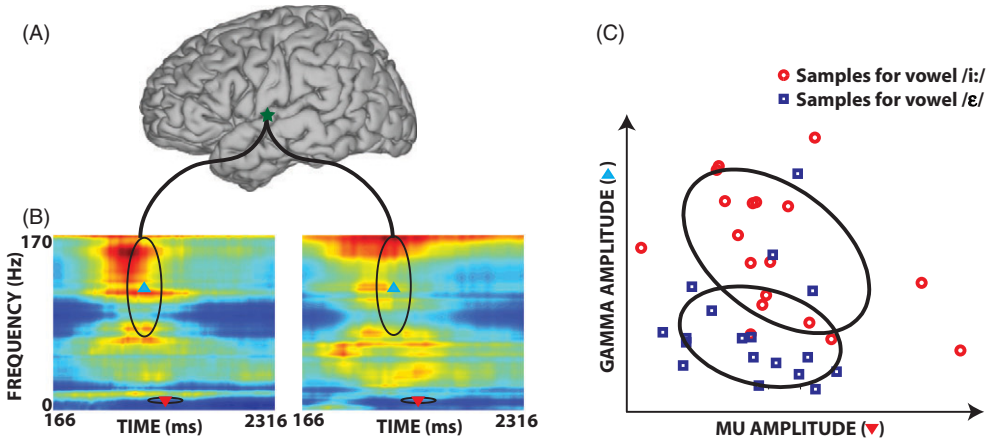

Figure 3. Example of ECoG features from one subject. (A) 3D brain template with one location marked with a green star; (B) normalized spectrogram of ECoG signals recorded at the location marked in (A), averaged across all spoken words with the same vowels, i.e. /i:/ (left) and /ε/ (right); (C) distribution of samples in two-dimensional feature space (i.e. gamma and mu amplitudes outlined by ellipses in (B), marked with upward (blue online) and downward (red online) triangles, respectively) for vowels /i:/ (circles) and /ε/ (squares).

per condition, on average). All eight subjects participated in the experiments using overt word repetition; a subset of six subjects (A, B, C, D, E, F) participated in experiments using covert word repetition. The subjects completed 72–216 trials for overt speech (140 on average) and 72–252 trials for covert speech (126 on average). Each dataset was visually inspected and all channels that did not contain clean ECoG signals (e.g., ground/reference channels, channels with broken connections, etc) were removed, which left 47–64 channels for our analyses.

2.4. 3D cortical mapping

We used lateral skull radiographs to identify the stereotactic coordinates of each grid electrode with software [47] that duplicated the manual procedure described in [48]. We defined cortical areas using Talairach's Co-Planar Stereotaxic Atlas of the Human Brain [49] and a Talairach transformation (http://www.talairach.org). We obtained a 3D cortical brain model from source code provided on the AFNI SUMA website (http://afni.nimh.nih.gov/afni/suma). Finally, we projected each subject's electrode locations on this 3D brain model and generated activation maps using a custom Matlab program.

2.5. Feature extraction and classification

We first re-referenced the signal from each electrode using a common average reference (CAR) montage [39]. Then, for every 50 ms and for each channel, we converted the time-series ECoG signals of the previous 333 ms into the frequency domain using an autoregressive (AR) model [50] and an empirically determined model order (25)[8]. Using this AR model, we calculated spectral amplitudes between 0 and 200 Hz in 2 Hz bins. We then averaged these spectral amplitudes in three different frequency ranges, i.e. 8–12, 18–26 and 70–170 Hz (excluding 116–124 Hz). Figure 3(B) shows an example of normalized ECoG time–frequency spectrograms recorded from the location marked in figure 3(A) (channel 38, superior temporal gyrus, Brodmann Area 22) for subject A. The two spectrograms in this figure were generated across responses for all word stimuli containing /i:/ and /ε/, respectively. Figure 3(C) shows an example of the distributions of samples in two-dimensional feature space (i.e. ECoG spectral amplitude within 70–170 Hz between 900 and 1233 ms, and within 8–12 Hz between 1100 and 1433 ms, respectively). In this study, we chose 70–170 Hz as the gamma frequency band, which is the same band we used in a different study using the same dataset [51]. Different groups or studies have selected different frequency bands (e.g., Crone's group used 80–100 Hz [26, 27], Knight's group used 80–200 Hz [29]). In general, a large number of ECoG studies have shown that functional activation of cortex is consistently associated with a broadband increase in signal power at high frequencies, i.e. typically >60 Hz and extending up to 200 Hz and beyond [52]. These high gamma responses have been observed in different functional domains including motor [53], language [27] and auditory [26, 31], and different choices of frequency bands have yielded comparable results [54, 55]. At the same time, recent evidence suggests that this view of a broadband phenomenon may be an oversimplification [56]. In addition to these frequency-based features, we also derived the local motor potential (LMP) [39], which was calculated as the running average of the raw time-domain signal at each electrode. These four sets of features were derived between 500 and 2500 ms after stimulus onset using a window size of 333 ms (50 ms stepping size). We extracted from each trial a total of 136 features (four different sets of features and 34 samples per feature). Because we only used ECoG information after 500 ms, and because no subject had coverage of visual areas, our ability to infer vowels or consonants was mainly based on interpretation of neural processes involved in overt/covert word repetition rather than of processes involved most directly with stimulus presentation.

Then, separately for each channel and analysis (overt or covert speech, vowels or consonants), we ranked the ECoG

[8] This model order typically maximized identification of task-related ECoG activity in offline analyses of this and other experiments.

features using the MRMR (maximum relevance and minimum redundancy) criterion [57]. We submitted the best (35 or 40 for decoding consonants or vowels, respectively) features at each location to a Naive Bayes classifier and used the optimized features to decode from each trial the vowel and consonant pair group for the target word of that trial, respectively.

We first asked if we could determine from the brain signals in each trial which of the four vowels (i.e. /ɛ/, /æ/, /i:/, /u:/) was present in the spoken or imagined word. For each vowel, the classifier was constructed and evaluated using tenfold cross-validation. To do this, each dataset was divided into ten, and the parameters of the Bayes classifier were determined from 9/10th of the dataset as a training set and tested on the remaining 1/10th test set. This procedure was then repeated ten times—each time, a different 1/10th of the dataset was used as the test set. The decoder for each vowel (i.e. including all nine different words with different consonant pairs) was constructed by modeling ECoG features associated with that vowel using a Naive Bayes classifier.

Then, using the same methods, the classifier identified the consonant pair in each trial (e.g., B_T, H_D, L_D, or P_P). To allow for a better comparison of the results between vowels (four possible vowels) and consonants (nine possible consonant pairs), we selected groups of four different consonant pairs for classification, and repeated the process for all possible combinations of four out of nine pairs (i.e. 126 combinations). The classification result for each combination of consonant pairs was the average result achieved for tenfold cross validation. We reported accuracies of the averaged results across all possible combinations for the best location.

Finally, we determined which vowel or consonant pair was most accurately identified by the procedures above. To do this, we first determined the actual and decoded vowel or consonant for each trial using the analyses described above. Then, we simply tabulated the frequency with which any actual vowel/consonant resulted in decoding of any of the vowels/consonants. The results are shown in the confusion matrices in tables 4–7, separately for vowels and consonant pairs and for overt and covert speech. These confusion matrices give indications which vowels/consonants were most similar or most discriminative.

2.6. Cortical discriminative mapping

For each subject, we derived a measure of classification accuracy from each electrode. Therefore, we were able to ask which cortical locations held the most information about discrimination of the vowels or consonants, i.e. which electrodes had a classification accuracy that was least likely due to chance.

Specifically, we first computed, for a given number of samples, the parameters (i.e. mean and standard deviation) of the normal distribution of accuracy values expected for a four-class problem using a randomization test. In this test, we produced 10 000 subsets of samples, where each sample had one of four random labels. We then calculated the accuracy of each subset of samples by comparing the random labels to the true labels. Based on these distributions (i.e. one distribution for each possible number of samples in our evaluations) of 10 000 accuracy values, we calculated the expected mean and standard deviations of accuracy values. We then derived a z-score for the observed accuracy level as the difference of that accuracy from the mean in units of standard deviation. Finally, we used a custom Matlab program to project the z-scores at all locations on to a three-dimensional template cortical brain model (i.e. cortical discriminative maps).

2.7. Spatial overlap

We then asked to what extent the cortical areas that were involved in the processing of vowels and consonants or of overt and covert speech overlapped with each other. To do this, we quantitatively evaluated the spatial overlap of the respective cortical discriminative maps using the reshuffling technique described in a recent study [58]. Specifically, we first determined the z-score at each location as described above, and set all z-scores below an empirically derived threshold (listed below) to zero. Second, we quantified the spatial overlap between two conditions (e.g., overt and covert) by calculating the dot product of the two sets of z-score values. This calculation was confined to those locations that had non-zero z-scores for at least one of the two conditions. Third, we created a surrogate distribution of dot product values by randomly reshuffling electrode positions for one of the two conditions, calculated the dot product, and repeated this process 10^6 times. Finally, we used this surrogate distribution to estimate the statistical significance of the overlap value that we observed for the correct (i.e. unshuffled) locations. We computed these significance values for individual subjects (using a z-score threshold of 1.64, which corresponds to a p-value of 0.05) and also combined values for all subjects (i.e. simply by concatenating locations and z-scores across all subjects, using a z-score threshold of 2, which corresponds to the threshold shown in figure 5).

3. Results

3.1. Decoding performance

The average classification accuracies for decoding vowels across all subjects were 40.7 ± 2.7% (overt speech) and 37.5 ± 5.9% (covert speech) (see figure 4 and supplementary material video 1 available at stacks.iop.org/JNE/8/046028/mmedia). For decoding consonants, the average classification accuracies across all subjects for the best location were 40.6 ± 8.3% (overt speech) and 36.3 ± 9.7% (covert speech) (see figure 4). These classification accuracies were substantially better than those expected by chance (i.e. 25% for vowels and also for consonants) as evaluated using the parameterized bootstrap resampling test with 10 000 repetitions as described above. In particular, for overt speech, accuracies were significant in all subjects ($p < 0.004$) for vowels and in most subjects (7/8 subjects, $p < 0.0022$) for consonants (see table 3 for details). For covert speech, accuracies were significant in most subjects (5/6 subjects, $p < 0.007$) for vowels and in the majority of the subjects (4/6 subjects, $p < 0.03$) for consonants (see table 3 for details). Statistical analyses using the paired

Table 3. Statistical analysis of the classification accuracy. In this table, 'n' is the number of trials; 'z-scores' indicates how many standard deviations a particular accuracy level is better than that expected by chance.

	Overt speech		Covert speech	
	Vowels (n)/(z-scores)	Consonant (n)/(z-scores)	Vowel (n)/(z-scores)	Consonant (n)/(z-scores)
A	72/3.8	72/5.5	72/4.1	72/5.9
B	108/3.6	108/2.03	108/1.2	108/2.08
C	216/5.4	216/4.8	180/2.8	180/1.3
D	216/3.9	216/7.7	72/3.2	72/2.3
E	216/4.7	216/5.4	252/2.48	252/1.44
F	72/3.2	72/2.1	72/2.7	72/1.91
G	144/5.3	144/5.7		
H	72/2.66	72/1.3		
p-value	$P < 0.004$	$p < 0.022$ (except subject H)	$p < 0.007$ (except subject B)	$p < 0.03$ (except subjects C and E)

Table 4. Confusion matrix of consonant decoding for overt speech. The column labels correspond to the predicted consonant pair in a given trial. The row labels correspond to the correct consonant pair. The values in each cell give frequencies in percent and the standard deviation calculated across subjects. This table represents a composite of frequencies derived from all 126 combinations of 4 of 9 consonants. The best consonant pair (R_D) is marked in bold.

%	B_T	K_N	H_D	L_D	M_N	P_P	R_D	S_T	T_N
B_T	23.2 ± 9.9	8.8 ± 6.5	11.2 ± 4.1	10.6 ± 4.7	9.8 ± 4.1	7.5 ± 3.7	7.2 ± 6.2	9.6 ± 3.9	12.2 ± 3.4
K_N	9.2 ± 4.5	20.9 ± 5.2	9.8 ± 3.5	8.0 ± 4.3	7.5 ± 3.9	12.7 ± 3.4	13.3 ± 1.9	10.5 ± 4.1	8.1 ± 4.5
H_D	9.2 ± 4.9	8.9 ± 4.8	27.2 ± 7.2	10.7 ± 3.9	8.3 ± 5.4	7.4 ± 2.9	8.3 ± 4.1	11.2 ± 4.5	8.8 ± 3.8
L_D	13.8 ± 3.7	8.1 ± 4.3	11.4 ± 5.1	21.8 ± 4.2	8.9 ± 3.4	9.6 ± 2.7	6.2 ± 4.1	8.9 ± 2.2	11.37 ± 3.1
M_N	10.9 ± 2.3	7.9 ± 4.2	11.8 ± 5.6	7.9 ± 3.6	27.6 ± 9.1	7.5 ± 2.9	6.8 ± 4.8	8.2 ± 3.0	11.19 ± 2.6
P_P	6.4 ± 2.2	14.4 ± 2.7	10.8 ± 4.7	9.7 ± 3.9	6.9 ± 2.4	24.3 ± 5.2	10.5 ± 2.1	9.6 ± 2.3	7.2 ± 2.9
R_D	7.9 ± 4.7	13.6 ± 5.8	8.4 ± 4.3	6.2 ± 2.7	6.9 ± 4.8	9.9 ± 3.9	**32.9 ± 12.5**	7.9 ± 3.7	6.3 ± 3.7
S_T	10.9 ± 3.9	11.6 ± 4.8	11.2 ± 3.3	10.0 ± 2.3	7.8 ± 2.7	9.4 ± 3.1	8.4 ± 2.6	22.7 ± 5.2	7.9 ± 1.3
T_N	12.5 ± 3.9	7.1 ± 3.4	12.6 ± 4.8	11.5 ± 2.7	10.9 ± 2.4	6.9 ± 2.2	6.5 ± 4.7	8.1 ± 2.3	23.7 ± 6.0

Table 5. Confusion matrix of consonant decoding for covert speech.

%	B_T	K_N	H_D	L_D	M_N	P_P	R_D	S_T	T_N
B_T	22.3 ± 8.1	11.4 ± 6.1	10.1 ± 3.7	8.6 ± 4.7	9.6 ± 1.2	7.2 ± 3.1	8.9 ± 4.2	11.9 ± 4.9	9.9 ± 1.6
K_N	8.2 ± 4.3	24.2 ± 6.2	8.7 ± 5.9	11.8 ± 2.8	6.9 ± 3.3	11.3 ± 3.6	7.6 ± 3.3	12.6 ± 3.4	8.7 ± 2.9
H_D	9.7 ± 5.3	8.3 ± 3.2	18.3 ± 3.7	10.4 ± 2.3	9.2 ± 2.4	10.3 ± 3.7	11.3 ± 2.5	10.1 ± 3.1	12.5 ± 1.8
L_D	6.7 ± 4.4	10.3 ± 4.7	10.5 ± 2.4	24.7 ± 7.0	8.9 ± 3.9	9.7 ± 1.2	8.8 ± 2.3	10.9 ± 1.8	9.4 ± 1.7
M_N	13.4 ± 9.2	8.3 ± 4.5	8.4 ± 2.8	11.3 ± 3.8	23.8 ± 6.4	8.1 ± 3.2	7.8 ± 3.6	9.7 ± 3.2	9.2 ± 1.9
P_P	7.9 ± 3.5	10.5 ± 3.0	9.9 ± 3.3	12.0 ± 1.1	7.8 ± 3.9	25.3 ± 9.0	9.1 ± 1.2	8.6 ± 3.4	8.8 ± 4.1
R_D	7.6 ± 4.5	9.3 ± 2.7	12.8 ± 0.8	9.5 ± 5.4	7.9 ± 4.5	8.1 ± 3.2	**27.8 ± 16.9**	8.6 ± 2.2	8.2 ± 2.9
S_T	9.0 ± 3.5	11.4 ± 3.1	10.7 ± 1.3	10.5 ± 2.9	8.4 ± 3.5	8.1 ± 1.9	8.2 ± 2.8	23.7 ± 6.2	9.9 ± 2.7
T_N	10.6 ± 3.4	8.7 ± 4.1	12.5 ± 3.6	9.3 ± 4.8	7.1 ± 3.2	8.0 ± 4.6	7.7 ± 2.1	10.8 ± 5.3	25.2 ± 11.8

Table 6. Confusion matrix of decoding vowels for overt speech. The column labels correspond to the predicted vowel pair in a given trial. The row labels correspond to the correct vowel pair. The values in each cell give frequencies in per cent and the standard deviation calculated across subjects. The best vowel pair /u:/ is marked in bold.

Accuracy (%)	/ɛ/	/æ/	/i:/	/u:/
/ɛ/	27.72 ± 7.92	18.76 ± 6.27	24.43 ± 8.29	29.09 ± 4.55
/æ/	21.38 ± 11.83	34.09 ± 16.00	23.79 ± 9.45	20.74 ± 5.78
/i:/	19.82 ± 6.92	17.88 ± 8.22	38.80 ± 7.71	23.48 ± 8.31
/u:/	19.74 ± 6.38	19.64 ± 4.77	21.70 ± 7.36	**38.91 ± 7.68**

Table 7. Confusion matrix of decoding vowels for covert speech.

Accuracy (%)	/ɛ/	/æ/	/i:/	/u:/
/ɛ/	40.49 ± 12.32	19.16 ± 4.01	19.01 ± 6.09	21.33 ± 12.76
/æ/	27.93 ± 6.20	26.63 ± 4.72	23.70 ± 6.01	21.74 ± 8.71
/i:/	22.19 ± 9.88	27.03 ± 7.96	28.69 ± 7.08	22.09 ± 9.98
/u:/	20.82 ± 9.59	16.00 ± 7.76	20.45 ± 9.25	**42.72 ± 9.35**

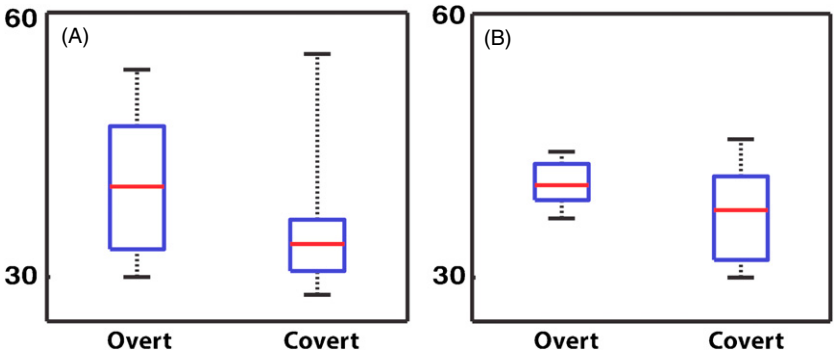

Figure 4. Classification accuracies of the ECoG-based decoding of vowels and consonants during overt and covert speech, respectively. On each box, the central mark is the median value, the edges of the box are the 25th and 75th percentiles and the whiskers extend to the maximum/minimum. Chance accuracy is 25%. (A) Consonant decoding accuracy for overt/covert word repetition; (B) vowel decoding accuracy for overt/covert word repetition.

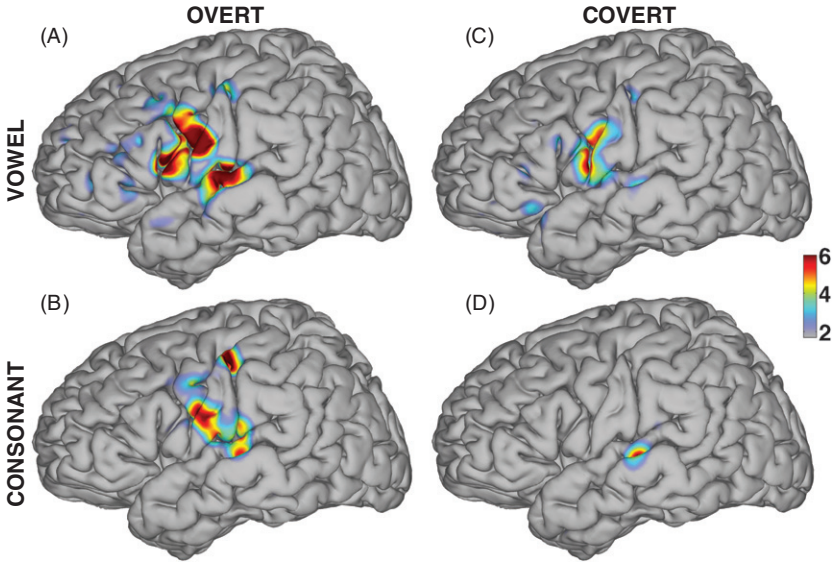

Figure 5. Color-coded cortical discriminative maps for vowels or consonants and for actual or imagined speech, respectively. The color-coded (see color bar) cortical patterns show the locations with the best decoding performance and were superimposed for all left-hemisphere subjects (seven for actual speech and six for imagined speech, respectively). Color gives z-scores indicating how much better accuracy at the respective location was compared to chance (p-value: 0.0023 at a z-score of 2). (A), (B) Discriminative maps for decoding vowels and consonants during overt word repetition. (C), (D) Discriminative maps for decoding vowels and consonants during covert word repetition.

Wilcoxon signed-rank test for eight subjects (overt speech) and six subjects (covert speech) did not reveal significant differences in accuracy for consonants and vowels or overt and covert speech.

The classification accuracies for each vowel and consonant pair are shown in confusion matrices that are presented in tables 4–7. These results show that the best averaged classification performance across all subjects, for both overt and covert speech, was achieved for the vowel /u:/ and the consonant pair 'R_D'. The corresponding accuracies were 39%, 43%, 33%, 28% (i.e. vowels/overt, vowels/covert, consonants/overt, consonants/covert, respectively), all of which were above the chance level (i.e. 25% for four-vowel matched groups classification and for nine-consonant-pair matched groups classification).

These results demonstrate that it is possible to infer the vowels and consonant pairs independently in spoken and imagined words using ECoG signals in humans.

3.2. Cortical discriminative maps

Figure 5 shows the cortical discriminative maps that indicated the areas that held the most information about vowels and consonants from all subjects. Figures 6 and 7 show the same results for individual subjects. The results shown in figure 5 demonstrate that the cortical areas that best discriminated vowels or consonants in the overt speech tasks were located in primary motor cortex (PMC, Brodmann's area (BA) 4), premotor cortex (BA 6), Broca's area (BA 44/45) and also posterior superior temporal gyrus (STG, i.e. the posterior part of BA 22). For covert speech tasks,

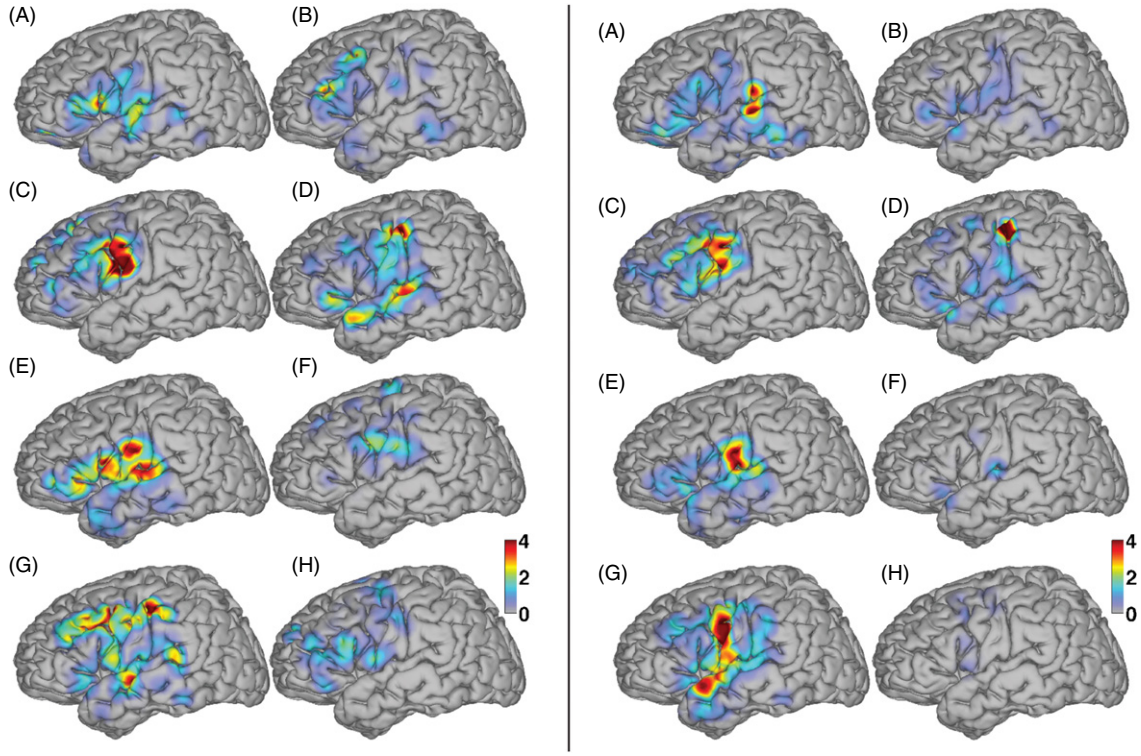

Figure 6. Cortical discriminative maps for individual subjects and for vowels (left panel) and consonants (right panel) in overt speech.

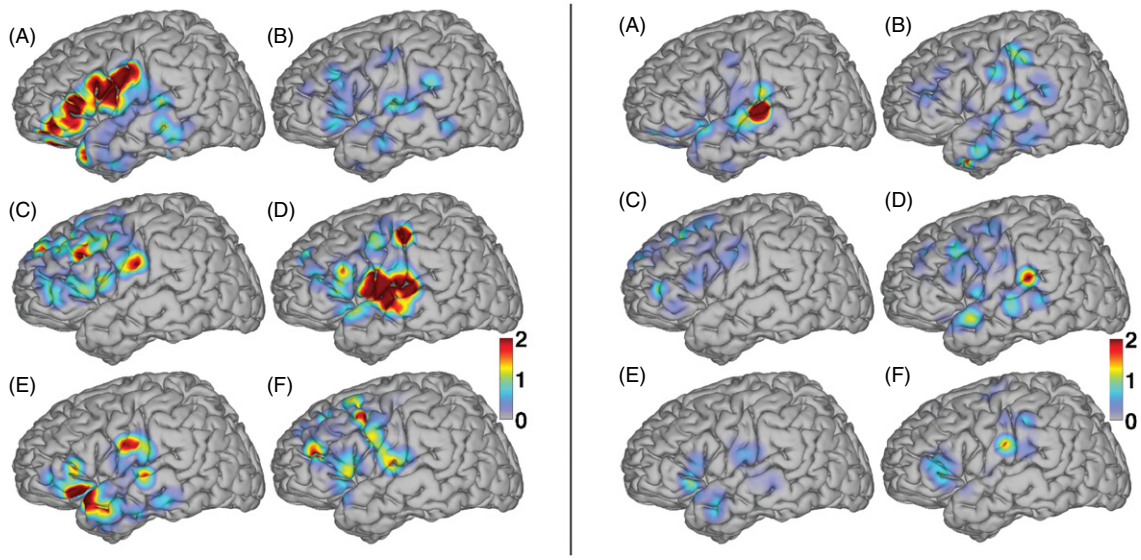

Figure 7. Cortical discriminative maps for individual subjects and for vowels (left panel) and consonants (right panel) in covert speech.

the best cortical areas were localized over small foci in temporal and frontal regions. In addition, for decoding consonants, optimal sites tended to be located surrounding Wernicke's area, whereas for vowels, optimal sites tended to be centered surrounding the premotor regions with smaller involvements of Broca's and Wernicke's areas. The results shown in figure 6 indicate that the discriminative maps are relatively consistent for overt word repetition. Discriminative information is mainly located in the frontal lobe (premotor cortex and Broca's area) and also in superior temporal regions. In contrast, the discriminative maps shown in figure 7 are more distributed and variable across subjects for covert word repetition.

Finally, we asked whether the discriminative cortical maps for overt and covert speech or for vowels and consonants were different from each other. Using the same reshuffling technique described above, our results demonstrate that for both individual subjects and also for all subjects combined, the cortical discriminative maps did not overlap between vowels and consonants ($p > 0.26$ and $p = 1$, for overt and covert speech, respectively), or between overt and covert speech ($p > 0.94$ and $p > 0.37$, for vowels and consonants (except for

subject A, $p < 0.05$ when evaluating subjects individually), respectively). These results suggest that the neural substrates involved in discrimination of vowels and consonants during overt and covert speech tasks are different. However, the fact that these areas are statistically different does not preclude the possibility that they do not share some commonalities. In fact, figure 5 shows that the cortical patterns for vowels and consonants during overt speech tasks involve some common areas over premotor cortex and parts of middle temporal regions.

4. Discussion

In this study, we showed that ECoG signals can be used to decode vowels and consonants in spoken or imagined words. Discriminating different components of speech, such as phonemes, vowels, consonants or words, could provide an approach for rapid selection of one of multiple choices that may be quite intuitive. Thus, the results presented here may ultimately lead to speech-based BCI systems that may provide effective communication with only a little training.

In this context, it is interesting to note that the decoding accuracies that we reported in this study were similar for overt and covert speech tasks. At the same time, these decoding accuracies reflect accuracies achieved for individual locations. The discriminative maps shown in figure 5 suggest that overt speech allows for discrimination in larger cortical areas (i.e. not only auditory areas, but also motor areas) than does covert speech. This is consistent with a recent report that found that covert speech does not engage motor areas [51]. The fact that multiple locations, and not just one, hold information about vowels or consonants also points to a straightforward way to further improve the classification accuracies shown here.

We also began to elucidate the neural substrate associated with vowel and consonant production. Previous behavioral studies based on lesion cases and lexical decision tasks have shown that the brain dissociates vowel and consonant processing [21, 22], which could be explained by the differential demands on prosodic and lexico-semantic processing placed by vowels and consonants [21], respectively. Our results give quantitative evidence that production of different vowels and consonants is associated with different ECoG activity patterns. These differential patterns are predominantly located in premotor and motor areas for spoken vowels/consonants, and in different speech-related areas for imagined vowels/consonants. This finding supports the notion that overt word repetition is composed partly of motoric processes of speech production [29, 30, 59–61] that contribute less to covert word repetition. Moreover, these results suggest that covert word repetition consists at least in part of imagining the perceptual qualities of the word (i.e. imagining what the word sounds like) rather than of processes that simulate the motor actions necessary for speech production. This is in marked contrast to recent findings [58] that demonstrated that overt and covert motor performances result in similar ECoG activation patterns. One should keep in mind, however, that it is unclear whether our results will generalize to speech tasks other than the word repetition task used here.

No previous study demonstrated that different vowels or consonants that are embedded in different spoken or imagined words can be discriminated using brain signals. Our results show that vowels and consonants can be decoded independently, and thus provide additional evidence for the dissociation within different vowels or consonants. Decoding vowels or consonants across groups of words is a more complicated problem than, for example, simply decoding one of four spoken vowels, and also complicates the corresponding interpretations (e.g., misclassification of /ɛ/ as /u:/ in table 6 and /æ/ as /ɛ/ in table 7). At the same time, our results show that the vowel /u:/ overall provided the best classification rates for both overt and covert speech. This observation supports the hypothesis that formant-based features may play an important role in brain-based discrimination of the spoken/imagined different vowels.

In conclusion, the results shown in this paper may ultimately lead to BCI systems based on overt or covert speech. Furthermore, our findings add empirical evidence that there is cortical dissociation not only between processing of different vowels or consonants in spoken and imagined words, but also between processing of vowels and consonants. Further research is needed to improve detection accuracy and/or extend these results to more vowel/consonant categories. In particular, use of information from different locations, and not just individual locations as done here, should prove useful.

Acknowledgments

This work was supported by grants from the US Army Research Office (W911NF-07-1-0415 (GS), W911NF-08-1-0216 (GS)), the NIH/NIBIB (EB006356 (GS) and EB000856 (JRW and GS)), and the James S McDonnell Center for Higher Brain Function (ECL). P Brunner provided technical assistance.

References

[1] 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. USA.* **101** 17849–54
[2] Serruya M *et al* 2002 Instant neural control of a movement signal *Nature* **416** 141–2
[3] Santhanam G *et al* 2006 A high-performance brain–computer interface *Nature* **442** 195–8
[4] Velliste M *et al* 2008 Cortical control of a prosthetic arm for self-feeding *Nature* **453** 1098–101
[5] Leuthardt E C *et al* 2004 A brain–computer interface using electrocorticographic signals in humans *J. Neural Eng.* **1** 63–71
[6] Wolpaw J R *et al* 2002 Brain–computer interfaces for communication and control *Clin. Neurophysiol.* **113** 767–91
[7] Wolpaw J R 2007 Brain–computer interfaces as new brain output pathways *J. Physiol. (Lond.)* **579** 613–9
[8] Sellers E and Donchin E 2006 A P300-based brain–computer interface: initial tests by ALS patients *Clin. Neurophysiol.* **117** 538–48
[9] Hochberg L R *et al* 2006 Neuronal ensemble control of prosthetic devices by a human with tetraplegia *Nature* **442** 164–71

[10] Blankertz B *et al* 2006 The Berlin brain–computer interface: EEG-based communication without subject training *IEEE Trans. Neural Syst. Rehabil. Eng.* **14** 147–52

[11] Guenther F H *et al* 2009 A wireless brain–machine interface for real-time speech synthesis *PLoS ONE* **4** e8218

[12] Price C J 2000 The anatomy of language: contributions from functional neuroimaging *J. Anat.* **197** (Pt 3) 335–59

[13] Fiez J A and Petersen S E 1998 Neuroimaging studies of word reading *Proc. Natl Acad. Sci. USA* **95** 914–21

[14] Hickok G and Poeppel D 2007 The cortical organization of speech processing *Nat. Rev. Neurosci.* **8** 393–402

[15] Mitchell T M *et al* 2008 Predicting human brain activity associated with the meanings of nouns *Science* **320** 1191–5

[16] Suppes P and Han B 2000 Brain-wave representation of words by superposition of a few sine waves *Proc. Natl Acad. Sci. USA* **97** 8738–43

[17] Formisano E *et al* 2008 'Who' is saying "what"? Brain-based decoding of human voice and speech *Science* **322** 970–3

[18] DaSalla C S *et al* 2009 Single-trial classification of vowel speech imagery using common spatial patterns *Neural Netw.* **22** 1334–9

[19] Deng S *et al* 2010 EEG classification of imagined syllable rhythm using Hilbert spectrum methods *J. Neural Eng.* **7** 046006

[20] Kellis S *et al* 2010 Decoding spoken words using local field potentials recorded from the cortical surface *J. Neural Eng.* **7** 056007

[21] Carreiras M and Price C J 2008 Brain activation for consonants and vowels *Cereb. Cortex* **18** 1727–35

[22] Caramazza A *et al* 2000 Separable processing of consonants and vowels *Nature* **403** 428–30

[23] Ferreres A R *et al* 2003 Phonological alexia with vowel–consonant dissociation in non-word reading *Brain Lang.* **84** 399–413

[24] Boatman D, Hall C, Goldstein M H, Lesser R P and Gordon B 1997 Neuroperceptual differences in consonant and vowel discrimination: as revealed by direct cortical electrical interference *Cortex* **33** 83–98

[25] Sharp D J *et al* 2005 Lexical retrieval constrained by sound structure: the role of the left inferior frontal gyrus *Brain Lang.* **92** 309–19

[26] Crone N E *et al* 2001 Induced electrocorticographic gamma activity during auditory perception. Brazier award-winning article 2001 *Clin. Neurophysiol.* **112** 565–82

[27] Crone N, Hao L, Hart J, Boatman D, Lesser R P, Irizarry R and Gordon B 2001 Electrocorticographic gamma activity during word production in spoken and sign language *Neurology* **57** 2045–53

[28] Sinai A *et al* 2005 Electrocorticographic high gamma activity versus electrical cortical stimulation mapping of naming *Brain* **128** 1556–70

[29] Canolty R T *et al* 2007 Spatiotemporal dynamics of word processing in the human brain *Front Neurosci* **1** 185–96

[30] Towle V L *et al* 2008 ECoG gamma activity during a language task: differentiating expressive and receptive speech areas *Brain* **131** 2013–27

[31] Edwards E *et al* 2005 High gamma activity in response to deviant auditory stimuli recorded directly from human cortex *J. Neurophysiol.* **94** 4269–80

[32] Edwards E *et al* 2009 Comparison of time-frequency responses and the event-related potential to auditory speech stimuli in human cortex *J. Neurophysiol.* **102** 377–86

[33] Chang E F *et al* 2010 Categorical speech representation in human superior temporal gyrus *Nat. Neurosci.* **13** 1428–32

[34] Aoki F *et al* 1999 Increased gamma-range activity in human sensorimotor cortex during performance of visuomotor tasks *Clin. Neurophysiol.* **110** 524–37

[35] Aoki F *et al* 2001 Changes in power and coherence of brain activity in human sensorimotor cortex during performance of visuomotor tasks *Biosystems* **63** 89–99

[36] Leuthardt E C *et al* 2007 Electrocorticographic frequency alteration mapping: a clinical technique for mapping the motor cortex *Neurosurgery* **60** 260–70 discussion 270–1

[37] Miller K J *et al* 2007 Spectral changes in cortical surface potentials during motor movement *J. Neurosci.* **27** 2424–32

[38] Brunner P *et al* 2009 A practical procedure for real-time functional mapping of eloquent cortex using electrocorticographic signals in humans *Epilepsy Behav.* **15** 278–86

[39] Schalk G *et al* 2007 Decoding two-dimensional movement trajectories using electrocorticographic signals in humans *J. Neural Eng.* **4** 264–75

[40] Kubanek J *et al* 2009 Decoding flexion of individual fingers using electrocorticographic signals in humans *J. Neural Eng.* **6** 066001

[41] Pistohl T *et al* 2008 Prediction of arm movement trajectories from ECoG-recordings in humans *J. Neurosci. Methods* **167** 105–14

[42] Blakely T, Miller K J, Rao R P, Holmes M D and Ojemann J G 2008 Localization and classification of phonemes using high spatial resolution electrocorticography (ECoG) grids *Conf. Proc. IEEE Eng. Med. Biol. Soc.* **2008** 4964–7

[43] Chao Z C, Nagasaka Y and Fujii N 2010 Long-term asynchronous decoding of arm motion using electrocorticographic signals in monkeys *Front. Neuroeng.* **3** 3

[44] Gunduz A *et al* 2009 Mapping broadband electrocorticographic recordings to two-dimensional hand trajectories in humans Motor control features *Neural Netw.* **22** 1257–70

[45] Schalk G *et al* 2004 BCI2000: a general-purpose brain–computer interface (BCI) system *IEEE Trans. Biomed. Eng.* **51** 1034–43

[46] Mellinger J *et al* 2007 An MEG-based brain–computer interface (BCI) *Neuroimage* **36** 581–93

[47] Miller K J *et al* 2007 Cortical electrode localization from x-rays and simple mapping for electrocorticographic research: the 'Location on Cortex' (LOC) package for MATLAB *J. Neurosci. Methods* **162** 303–8

[48] Fox P *et al* 1985 A stereotactic method of anatomical localization for positron emission tomography *J. Comput. Assist. Tomogr.* **9** 141–53

[49] Talairach J and Tournoux P 1988 *Co-Planar Sterotaxic Atlas of the Human Brain* (New York: Thieme Medical)

[50] Marple S L 1987 *Digital Spectral Analysis: with Applications* (Englewood Cliffs, NJ: Prentice-Hall)

[51] Pei X *et al* 2011 Spatiotemporal dynamics of electrocorticographic high gamma activity during overt and covert word repetition *Neuroimage* **54** 2960–72

[52] Crone N E *et al* 2006 High-frequency gamma oscillations and human brain mapping with electrocorticography *Prog. Brain Res.* **159** 275–95

[53] Crone N E *et al* 1998 Functional mapping of human sensorimotor cortex with electrocorticographic spectral analysis: II. Event-related synchronization in the gamma band *Brain* **121** (Pt 12) 2301–15

[54] Miller K J *et al* 2008 Beyond the gamma band: the role of high-frequency features in movement classification *IEEE Trans. Biomed. Eng.* **55** 1634–7

[55] Ball T *et al* 2009 Differential representation of arm movement direction in relation to cortical anatomy and function *J. Neural Eng.* **6** 016006

[56] Gaona C M *et al* 2011 Non-uniform high-gamma (60–500 Hz) power changes dissociate cognitive task and anatomy in human cortex *J. Neurosci.* **31** 2091–100

[57] Peng H C *et al* 2005 Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy *IEEE Trans. Pattern Anal. Mach. Intell.* **27** 1226–38

[58] Miller K J *et al* 2010 Cortical activity during motor execution, motor imagery, and imagery-based online feedback *Proc. Natl Acad. Sci. USA* **107** 4430–5

[59] Pulvermuller F *et al* 2006 Motor cortex maps articulatory features of speech sounds *Proc. Natl Acad. Sci. USA* **103** 7865–70

[60] Pulvermuller F and Fadiga L 2010 Active perception: sensorimotor circuits as a cortical basis for language *Nat. Rev. Neurosci.* **11** 351–60

[61] Huang J *et al* 2002 Comparing cortical activations for silent and overt speech using event-related fMRI *Hum. Brain Mapp.* **15** 39–53