Silent Communication Toward Using Brain Signals

ARTICLE *in* IEEE PULSE · JANUARY 2012

Impact Factor: 0.66 · DOI: 10.1109/MPUL.2011.2175637 · Source: PubMed

Silent Communication

Toward Using Brain Signals

By Xiaomei Pei, Jeremy Hill, and Gerwin Schalk

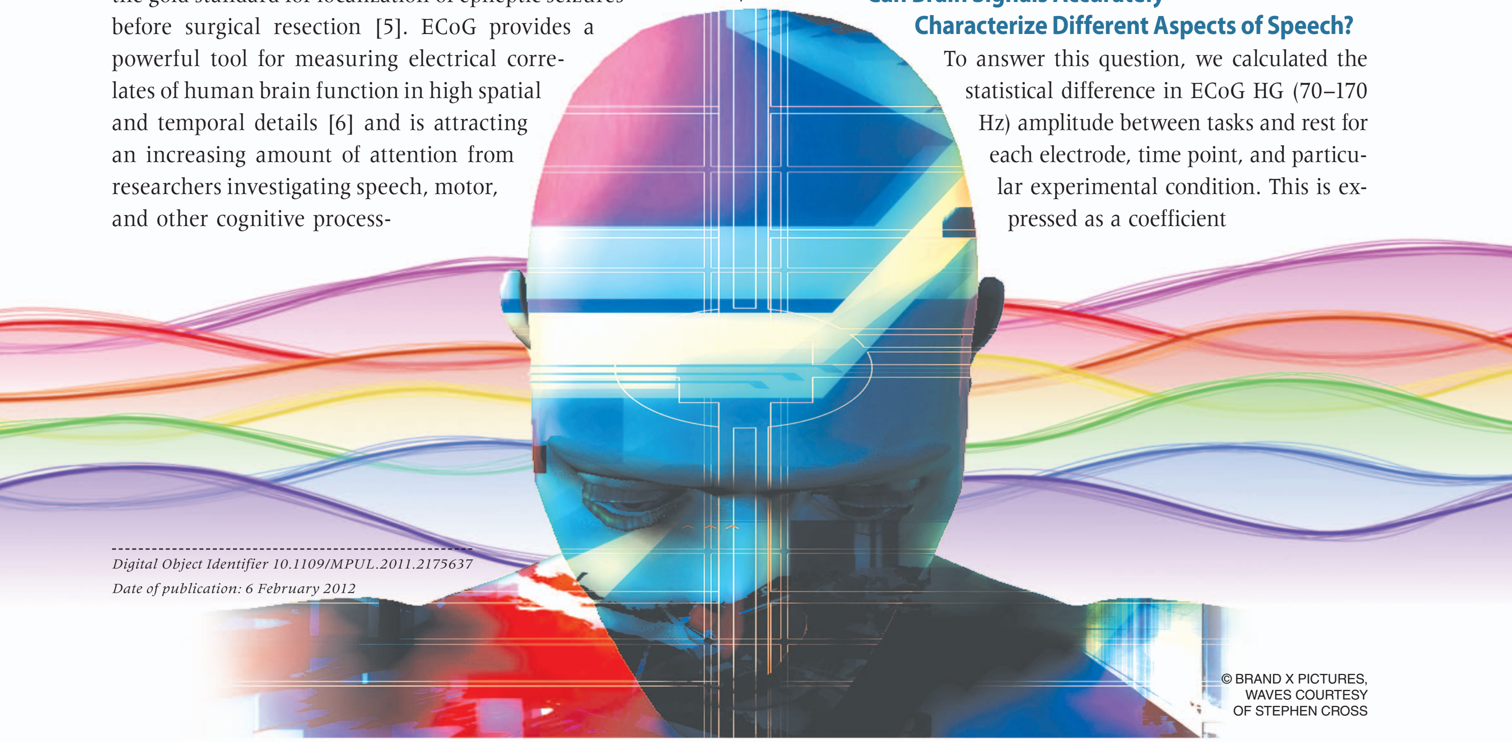
From the 1980s movie *Firefox* to the more recent *Avatar*, popular science fiction has speculated about the possibility of a person's thoughts being read directly from his or her brain. Such brain–computer interfaces (BCIs) [1] might allow people who are paralyzed to communicate with and control their environment, and there might also be applications in military situations wherever silent user-to-user communication is desirable [2]. Previous studies have shown that BCI systems can use brain signals related to movements and movement imagery [3] or attention-based character selection [4]. Although these systems have successfully demonstrated the possibility to control devices using brain function, directly inferring which word a person intends to communicate has been elusive. A BCI using imagined speech might provide such a practical, intuitive device. Toward this goal, our studies to date addressed two scientific questions: 1) Can brain signals accurately characterize different aspects of speech? 2) Is it possible to predict spoken or imagined words or their components using brain signals?

In these investigations, we have been using electrocorticographic (ECoG) signals that are directly recorded from the surface of the brain. Since the 1950s, ECoG has been used as the gold standard for localization of epileptic seizures before surgical resection [5]. ECoG provides a powerful tool for measuring electrical correlates of human brain function in high spatial and temporal details [6] and is attracting an increasing amount of attention from researchers investigating speech, motor, and other cognitive processes. In particular, changes of ECoG activity in the high gamma (HG) frequency band have been found to be a very specific, spatially precise correlate of the performance of such tasks [7], [8]. HG functional mapping results are generally concordant with those identified using electrical cortical stimulation [9], [10] or metabolic imaging [11], [12].

This article summarizes the results of our collaborative efforts to date. These efforts are described in detail in [13] and [14] and are summarized here. In these studies, we evaluated nine subjects temporarily implanted with electrode grids for epilepsy monitoring. We presented words to them either visually or acoustically. Auditory words were presented once at the beginning of each 4-s trial. Visual words were displayed on the screen for the complete duration of a trial. Subjects were asked to either speak aloud (i.e., overtly) or imagine speaking (i.e., covertly) each word. Thus, there were four experimental conditions: visual/actual, audio/actual, visual/imagined, and audio/imagined. Thirty-six words were used with four possible vowel sounds (/ɛ/, /æ/, /i:/, or /u:/) and nine consonant pairs (/b_t/, /c_n/, /h_d/, /l_d/, /m_n/, /p_p/, /r_d/, /s_t/, or /t_n/).

Can Brain Signals Accurately Characterize Different Aspects of Speech?

To answer this question, we calculated the statistical difference in ECoG HG (70–170 Hz) amplitude between tasks and rest for each electrode, time point, and particular experimental condition. This is expressed as a coefficient



Digital Object Identifier 10.1109/MPUL.2011.2175637
Date of publication: 6 February 2012

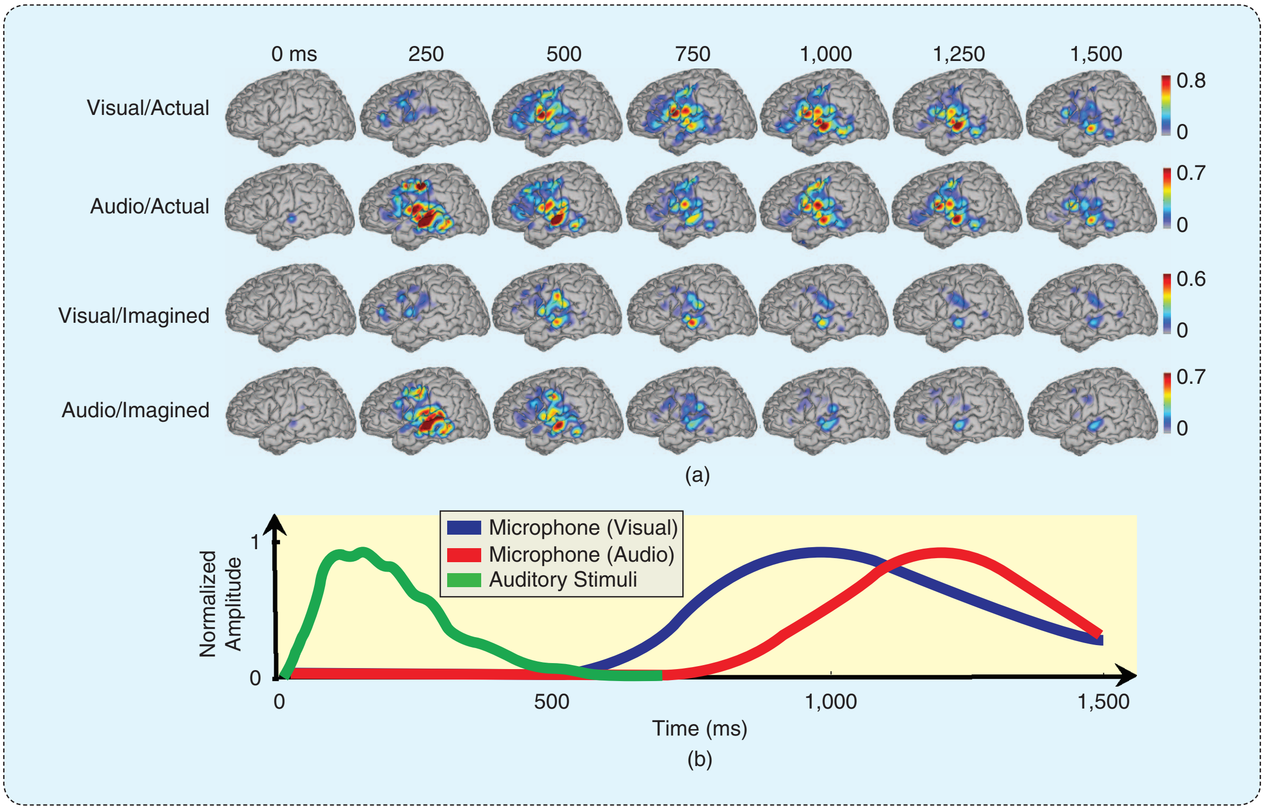


FIGURE 1 (a) Spatiotemporal dynamics of ECoG HG activity superimposed across all subjects and characterized by the color-coded r^2 (a value between zero and one) during the four conditions. (b) Normalized and averaged temporal envelopes of the auditory stimuli (green) and microphone signals reflecting the subjects' verbal responses following visual (blue) or auditory (red) presentation of the word on the same time scale as (a). (Reprinted from [13], with permission from Elsevier.)

of determination, otherwise known as an r^2 value [15], from zero to one. In Figure 1, r^2 is projected onto a three-dimensional (3-D) template brain model across all eight subjects who had a grid on the left hemisphere, to derive the spatial distributions of task-related cortical activations. For comparison, we show the averaged temporal envelopes of the auditory stimulus and microphone signals.

We can see that the auditory stimulus induces strong cortical activations that spread across temporal, parietal, and frontal lobes. In contrast, the primary cortical responses to the visual stimuli are not detected (as subjects did not have coverage of occipital cortex). Together with the temporal envelopes of the microphone signals following visual (blue) and auditory (red) stimuli, the patterns of cortical activations are qualitatively similar between the visual/actual and audio/actual conditions. Compared with actual speech, the imagined conditions resulted in weaker activations and with two foci: the superior temporal gyrus (STG) and Wernicke's area. Figure 2 shows the averaged time courses of r^2 across all electrodes and subjects. The auditory stimulus produces a peak neural response at 100 ms for actual and imagined tasks. Similarly, the subjects' verbal outputs (captured by the microphones) produce the neural responses at corresponding times. These results demonstrate that ECoG HG activities are well coupled, in spatially distinct functional areas, to the temporal evolution of the stimuli and subjects' responses.

For comparison, we show the averaged temporal envelopes of the auditory stimulus and microphone signals.

Is It Possible to Predict Spoken or Imagined Words or Their Components Using Brain Signals?

To answer this question, we determined whether we could decode, from ECoG signals, the vowels and consonants of spoken or imagined words. Recent studies have begun to elucidate the relationship of brain activity with different aspects of language function. For example, functional magnetic resonance imaging measurements have been shown to contain information about different individual vowels that subjects listened to [16], electroencephalography (EEG) signals have been found to contain information about the rhythm of syllables [17] and individual vowels [18], and ECoG was used to decode several spoken words [19]. However, no previous study had shown evidence of decoding vowels and consonants in actual (overt) as well as imagined (covert) speech.

In our study, we evaluated data from actual speech in eight subjects and from imagined speech in six subjects following visual stimuli. From each electrode location, we used ECoG signals between 500 and 2,500 ms after visual stimulus onset and extracted spectral amplitudes from those signals within the

8–12, 18–30, and 70–170-Hz frequency bands, as well as a time-domain feature called the *local motor potential* (LMP) [20]. In each trial, a Naive Bayes classifier was used to determine which of the four vowels and which of the four consonant pairs were present in the spoken or imagined word. (Each classification attempt was a four-way choice between the correct consonant pair and three others drawn from the overall pool of nine. All possible four-way choices were tested for a given word.) Figure 3(a) shows the decoding performance, which was estimated using tenfold cross-validation and averaged across subjects from the best single location. The average accuracies for decoding vowels and consonants in both spoken and imagined words were significantly above the level expected by chance (25%). Particular conditions exceeded 55%. Figure 3(b) shows a cortical discriminative mapping, i.e., a map of those locations that are most predictive of the particular vowel/consonant. The colors correspond to classification accuracy expressed as a *z* score, i.e., as a number of standard deviations above chance, as derived using a randomization test for a four-class problem. This reflects the significance of classification accuracy relative to chance level (see [14] for details). We found that information for decoding vowels or consonants in actual speech tasks is richest in primary motor cortex, premotor cortex, Broca's area, and posterior STG. For imagined speech, discriminative information was localized over two small foci in the temporal and frontal regions (see details in [14]).

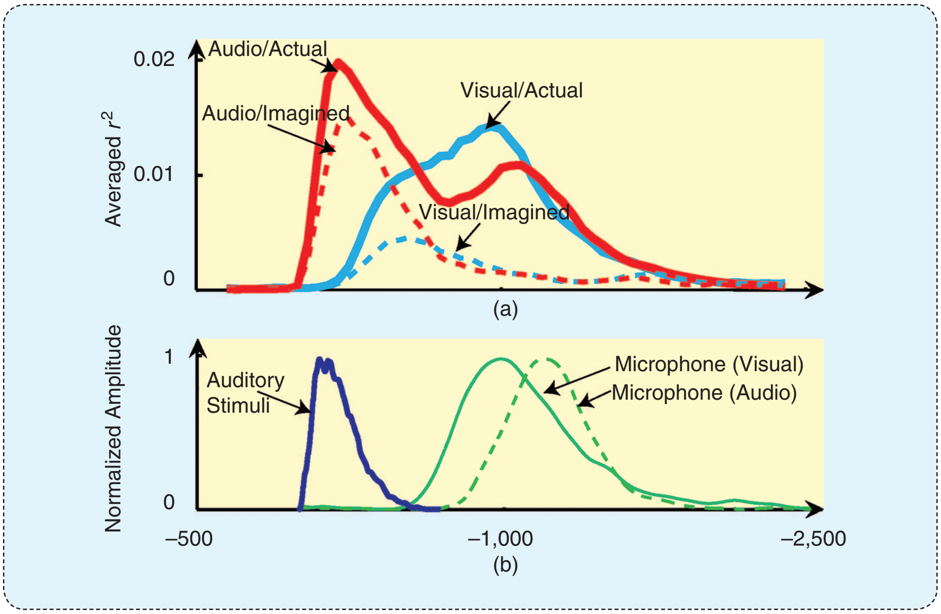


FIGURE 2 (a) Averaged time courses of ECoG HG activity across all electrodes and subjects. (b) Normalized and averaged temporal envelopes of auditory stimuli and microphone signals following visual and auditory stimuli. (Reprinted from [13], with permission from Elsevier.)

Conclusions and Future Directions

Taken together with previous studies, our work shows that it is possible to use brain signals to predict the vowels and consonants in spoken or imagined words. Furthermore, this decoding is possible based on single utterances rather than requiring brain signals to be averaged over hundreds of task repetitions, as in more traditional brain imaging. This is an encouraging sign that it may be feasible in the near future to decode complete words in real time, thereby allowing users to silently communicate without the need for muscle movements. It further underscores the potential usefulness of ECoG as a basis for neural engineering applications, both within clinical contexts (e.g., as a speech prosthesis for paralyzed users) and outside (e.g., as a silent communication method in military and security applications). To build a practical brain-based communication system using silent speech, some questions remain unanswered. For example, can we train

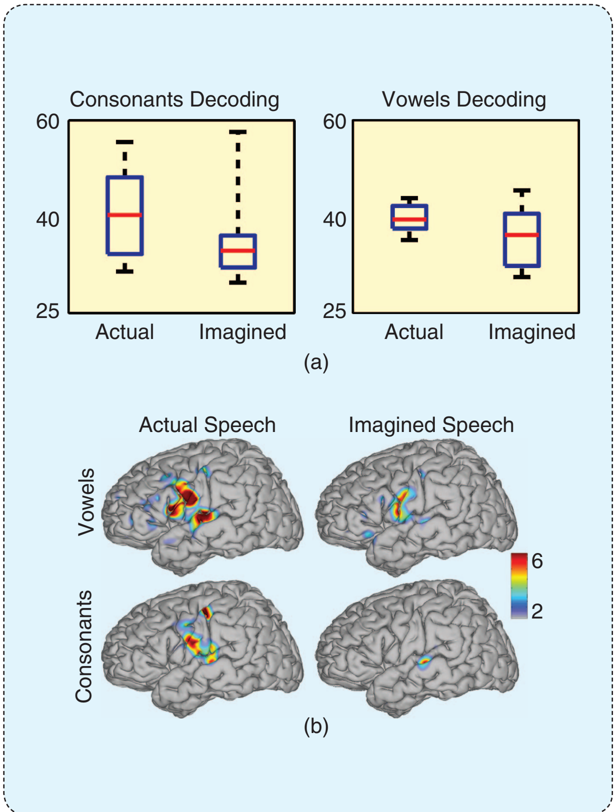


FIGURE 3 (a) Classification performance in which red lines correspond to the median value of accuracy across subjects, the edges of the box are the 25th and 75th percentiles, and the whiskers extend to a maximum/minimum. The chance level is 25%. (b) Cortical discriminative mapping for decoding vowels and consonants during actual/imagined speech tasks (color bar denotes *z* score, indicating how many standard deviations the classification accuracy is above chance level). (Reprinted from [14], with permission from IOP Publishing.)

our system on the neural correlates of spoken words and transfer the classifier directly to imagined words? To what extent can decoding performance be improved by using brain signals from multiple electrodes simultaneously and adapting to the differences in spatial activity patterns? Are the brain signal patterns that discriminate words or their components consistently structured (i.e., across subjects) across particular perceptual or expressive categories? Will this categorization allow us to extend the system to a larger number of vowel and consonant categories? Finally, to achieve a large-scale deployable communication system, the extension of our methods and findings to noninvasive methods such as EEG requires investigation.

For imagined speech, discriminative information was localized over two small foci in the temporal and frontal regions.

Acknowledgments

This work was supported by grants from the U.S. Army Research Office [W911NF-07-1-0415 (GS), W911NF-08-1-0216 (GS)] and the National Institutes of Health (NIH)/National Institute of Biomedical Imaging and Bioengineering (NIBIB) [EB006356 (GS) and EB000856 (JRW and GS)].

Xiaomei Pei (peixiaomei@gmail.com), Jeremy Hill (jezhill@gmail.com), *and* ***Gerwin Schalk (schalk@wadsworth.org)*** *are with Brain–Computer Interface R&D Program, Wadsworth Center, Albany, New York.*

References

[1] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interfaces for communication and control," *Clin. Neurophysiol.*, vol. 113, pp. 767–791, June 2002.

[2] I. S. Kotchetkov, B. Y. Hwang, G. Appelboom, C. P. Kellner, and E. S. Connolly, Jr., "Brain-computer interfaces: Military, neurosurgical, and ethical perspective," *Neurosurg. Focus*, vol. 28, p. E25, May 2010.

[3] J. R. Wolpaw and D. J. McFarland, "Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans," *Proc. Nat. Acad. Sci. USA*, vol. 101, pp. 17849–17854, Dec. 2004.

[4] E. W. Sellers and E. Donchin, "A P300-based brain-computer interface: Initial tests by ALS patients," *Clin. Neurophysiol.*, vol. 117, no. 3, pp. 538–548, 2006.

[5] W. Penfield and T. Rasmussen, *The Cerebral Cortex of Man*. Oxford, England: Macmillan, 1950.

[6] E. C. Leuthardt, G. Schalk, J. R. Wolpaw, J. G. Ojemann, and D. W. Moran, "A brain-computer interface using electrocorticographic signals in humans," *J. Neural Eng.*, vol. 1, pp. 63–71, June 2004.

[7] K. J. Miller, G. Schalk, E. E. Fetz, M. den Nijs, J. G. Ojemann, and R. P. Rao, "Cortical activity during motor execution, motor imagery, and imagery-based online feedback," *Proc. Nat. Acad. Sci. USA*, vol. 107, pp. 4430–4435, Mar. 2010.

[8] N. E. Crone, A. Sinai, and A. Korzeniewska, "High-frequency gamma oscillations and human brain mapping with electrocorticography," *Prog. Brain Res.*, vol. 159, pp. 275–295, Oct. 2006.

[9] P. Brunner, A. L. Ritaccio, T. M. Lynch, J. F. Emrich, J. A. Wilson, J. C. Williams, E. J. Aarnoutse, N. F. Ramsey, E. C. Leuthardt, H. Bischof, and G. Schalk, "A practical procedure for real-time functional mapping of eloquent cortex using electrocorticographic signals in humans," *Epilepsy Behav.*, vol. 15, pp. 278–286, July 2009.

[10] E. C. Leuthardt, K. Miller, N. R. Anderson, G. Schalk, J. Dowling, J. Miller, D. W. Moran, and J. G. Ojemann, "Electrocorticographic frequency alteration mapping: A clinical technique for mapping the motor cortex," *Neurosurgery*, vol. 60, pp. 260–270; discussion pp. 270–271, Apr. 2007.

[11] J. P. Lachaux, P. Fonlupt, P. Kahane, L. Minotti, D. Hoffmann, O. Bertrand, and M. Baciu, "Relationship between task-related gamma oscillations and BOLD signal: New insights from combined fMRI and intracranial EEG," *Hum. Brain Mapp.*, vol. 28, pp. 1368–1375, Dec. 2007.

[12] J. Niessing, B. Ebisch, K. E. Schmidt, M. Niessing, W. Singer, and R. A. Galuske, "Hemodynamic signals correlate tightly with synchronized gamma oscillations," *Science*, vol. 309, pp. 948–951, Aug. 2005.

[13] X. Pei, E. C. Leuthardt, C. M. Gaona, P. Brunner, J. R. Wolpaw, and G. Schalk, "Spatiotemporal dynamics of electrocorticographic high gamma activity during overt and covert word repetition," *Neuroimage*, vol. 54, pp. 2960–2972, Feb. 2011.

[14] X. Pei, "Decoding vowels and consonants in spoken and imagined words using electrocorticographic signals in humans," *J. Neural Eng.*, vol. 8, no. 4, p. 046028, 2011.

[15] T. H. Wonnacott and R. Wonnacott, *Introductory Statistics*. New York: Wiley, 1977.

[16] E. Formisano, F. De Martino, M. Bonte, and R. Goebel, 'Who' is saying 'what'? Brain-based decoding of human voice and speech," *Science*, vol. 322, pp. 970–973, Nov. 2008.

[17] S. Deng, R. Srinivasan, T. Lappas, and M. D'Zmura, "EEG classification of imagined syllable rhythm using Hilbert spectrum methods," *J. Neural Eng.*, vol. 7, p. 046006, June 2010.

[18] C. S. DaSalla, H. Kambara, M. Sato, and Y. Koike, "Single-trial classification of vowel speech imagery using common spatial patterns," *Neural Netw.*, vol. 22, pp. 1334–1339, Nov. 2009.

[19] S. Kellis, K. Miller, K. Thomson, R. Brown, P. House, and B. Greger, "Decoding spoken words using local field potentials recorded from the cortical surface," *J. Neural Eng.*, vol. 7, p. 056007, Oct. 2010.

[20] G. Schalk, J. Kubánek, K. J. Miller, N. R. Anderson, E. C. Leuthardt, J. G. Ojemann, D. Limbrick, D. Moran, L. A. Gerhardt, and J. R. Wolpaw, "Decoding two-dimensional movement trajectories using electrocorticographic signals in humans," *J. Neural Eng.*, vol. 4, pp. 264–275, Sept. 2007.