30th Annual International IEEE EMBS Conference
Vancouver, British Columbia, Canada, August 20-24, 2008

Localization and classification of phonemes using high spatial resolution electrocorticography (ECoG) grids

Timothy Blakely*[1], *Member, IEEE,* Kai J. Miller*[2], *Member, IEEE,* Rajesh P.N. Rao[3], *Member, IEEE*
Mark D Holmes[4], Jeffrey G. Ojemann[4]

Abstract— **We present results of cortical activity during phoneme pronunciation, recorded using miniaturized electrocorticography grids with high spatial resolution. A patient implanted with the miniature grid was instructed to audibly pronounce one of four phonemes. For each phoneme, we observed distinct spatial correlation patterns at the 3mm electrode spacing. We applied a support vector machine classification scheme and, for the first time, were able to distinguish discrete phonemes with high accuracy. In addition, we found that sub-regions of our miniature array were specific for distinct pairs of phonemes, showing that cortical phoneme processing occurs at a higher resolution than previously though.**

I. INTRODUCTION

A significant amount of human communication is done through speech. Language assimilation begins at an early stage in human development and is reinforced throughout the lifetime. However, the exact representation of this speech is yet unknown[1, 2]. One of the initial forays into the area of speech representation in human cortex was the use of tonotopic maps[3-5]. Further investigation by Diesch & Luce illustrated that tonotopic organization is insufficient to explain specific phoneme representation. They showed that the N100m component responded differently when exposed to two-formant vowels and their decomposed formants separately and could not find sufficient evidence to suggest that the vowel source locations were linear combinations[6].

Diesch and Luce then investigated whether vowel formants F_1 and F_2 can interact at one or more early stages of the auditory pathway[7]. Neurons in the auditory pathway demonstrated the ability to respond to narrow information bands, effectively band-passing the auditory information to lower auditory areas. Investigating the possibility of early formant interactions in gerbils, Ohl *et. Al* found interactions as far away as 2000Hz and as close as 500Hz[8]. Wang *et. Al* approached the possibility of formant interactions from a spatial perspective in marmosets. They observed distributed neuronal discharge patterns in the marmosets' primary auditory cortexes from behaviorally relevant, species-specific vocalizations[9]. Due to the limitations in techniques available on humans at the time, it was not possible to get the same spatial resolution in humans. Diesch proposed to use Euclidean distances between from evoked magnetic field measurements and found that the distance between the vowels 'I' and 'u' was further than the distance between vowels 'a' and 'e' in more than 60% of trials[10]. In 2003, Eulitz et al investigated magnetoencaphalography (MEG) evoked N100m responses due to auditory queues. Like the results observed by Diech, they determined that the acoustic vowel sounds generated by 'a' and 'i' were more spatially separated in comparison with the similar sounding vowels 'e' and 'i'.

In this paper we present a look at direct cortical surface potentials acquired via an electrocorticography (ECoG) microarray. This microarray allowed us to look at frontal processing regions with a much greater spatial resolution than currently used ECoG grids (10mm spatial resolution). Surface potential changes were recorded during overt pronunciation of language-base phonemes at high spatial resolution to determine if core language phoneme pronunciation can be separated and classified.

II. METHODS

The experimental methods used rely on the BCI2000 program. The reader may find it useful to examine existing papers [11, 12].

A. Subject

The subject that participated in the experiment was a 46 year old left handed, female, epileptic patient, who received placement of a large scale sub-dural electrocorticographic array (figure 1) for the clinical purpose of localizing the seizure focus prior to resection. The seizure focus was determined to be in frontal cortex, significantly dorsal and anterior to the location of the miniaturized experimental ECoG array.

Manuscript received April 16, 2008. The authors would like to thank the staff and patients at Harborview Hospital (Seattle, WA) for their time, consideration, and effort on behalf of this research. This research was supported by grants NSF 0622252 and 0130705, NIH T32-NS07144, US Army Research Office W911NF-07-1-041 (GS), NIH/NIBIB EB006356 (GS) and the Poncin Foundation (KJM).

[1]*T. Blakely is with the University of Washington, Bioengineering department
[2]*K. J. Miller is with the University of Washington, Physics dept
[3]RPN Rao is with the University of Washington, Computer Science and Engineering department.
[4] J. Ojemann and M. Holmes are with the University of Washington, Neurological Surgery department
* These writers contributed equally to this work

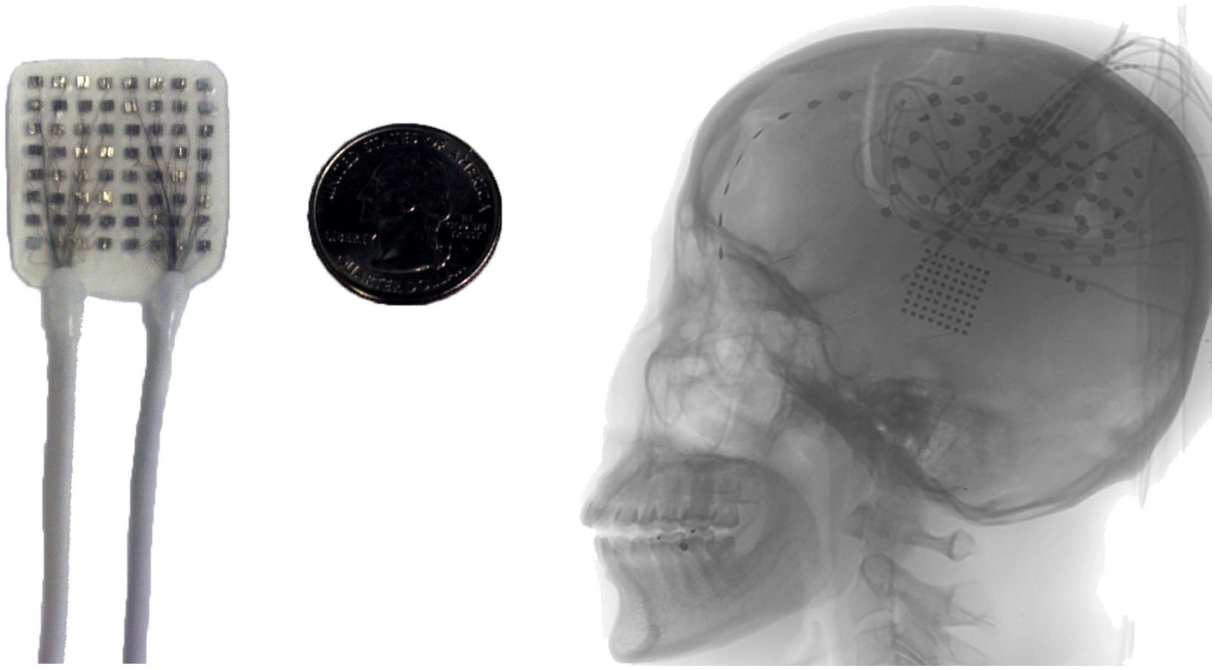

Figure 1: Left – The miniature ECoG array can be seen on the left, with a quarter to demonstrate the scale. The electrodes were spaced 3mm apart from center to center. Right – The subject's x-ray demonstrates the clinical array (larger electrodes, spaced 10mm) and, within frontotemporal area, the microgrid with a spacing of 3mm. The higher spatial resolution of the grid allows each electrode to average over a smaller number of neurons, allowing more distinct potential recordings from smaller populations of neurons.

B. Recordings

A miniature 8 x 8 platinum electrode was placed sub-durally in the perisylvian area in left inferior frontal cortex during the placement of a diagnostic clinical array (figure 1). The electrodes were spaced 3mm apart from each other, from center to center (manufactured by Ad-Tech, Racine, WI). Previously used ECoG grids have had electrode spacings of 10mm [13-16]. Potentials were sampled at 1000Hz, using Synamps2 amplifiers (Neuroscan, TX), using a clinical sub-dural electrode as ground. There was an instrument-imposed band-pass filter from 0.15-200Hz.

C. Task

In response to a visual cue on a monitor at the bedside, the subject would repeat a phoneme 3 times within the 2 second cue period. Each experimental run consisted of 2 paired phonemes, "BA/WA" or "RA/LA", and there were 30 cues for each phoneme, presented in random order. Between each 2 second cue, there was a 2 second rest period.

D. Signal Processing

The electrode recording was transformed from the 64 channel array, to 112 pairs of nearest-neighbor pairwise difference channels. This re-referencing was performed so that any fluctuations in potential found were as local as possible, minimizing both common-mode input and noise.

Signal processing was performed offline after data collection. The data were notch filtered for 60, 120, and 180 Hz to eliminate line noise, using a 3rd order Butterworth filter. We computed the Fast Fourier Transform (FFT) for each 2 second cue interval and each 2 second rest interval. The data from these epochs were transformed using overlapping 0.256s (256 sample) windows with 0.1s step sizes between them. A Hann window was imposed on each data window to attenuate edge effects. Each spectrum from each phoneme cue and rest period was averaged by the mean spectrum throughout the trial, for each electrode independently. More information on the technique can be found in previously published papers[16-18].

E. Feature Selection

The mean power was calculated in each electrode in 6 discrete frequency ranges, the low alpha (7-12 Hz), high alpha (10-13Hz), beta (14-25Hz), low gamma (26-35Hz), high gamma (36-70Hz), and chi (70-150Hz). The chi values were used to select the ten best channels, by r2 value, for discriminating "BA" vs. "WA", and "RA" vs. "LA", independently. Interestingly, these channels were different for each pair, suggesting that the miniature array is capturing the level of "phonemotopy" [10], a finding which was supported strongly by classification result.

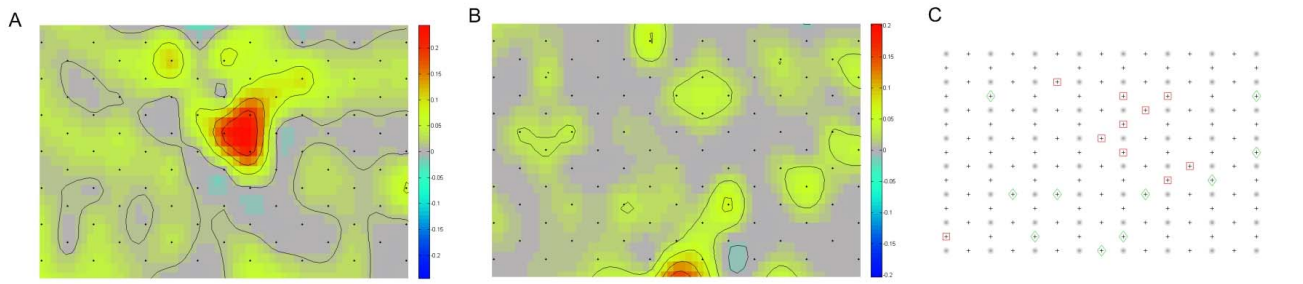

Figure 2: A) Topographic signed r-squared cross-correlation maps showing significant differences for phonemes "RA" and "LA" within the center of the grid. Spatial resolutions as small as a single virtual electrode (created by pairwise re-referencing) show significant correlations. B) Topographic signed r-squared cross-correlation map for "BA" vs. "WA", showing significant differences towards the bottom of the grid, spatially separate from the significant electrodes in "RA" vs. "WA". C) Each plus represents the virtual electrode generated by pairwise re-referencing the electrodes(shown as circles). The 10 virtual electrodes that were used for classification for "RA" vs. "WA" are marked as red squares, with the virtual electrodes used for classifying "BA" vs. "WA" outlined by green diamonds.

F. Classification

We performed binary classification between paired phonemes, using the 6 frequency ranges, in 10 different channels at a time which are shown in red and green in figure 2. This was done using a linear support vector machine with 4-fold nested cross-validation[19].

III. Results and Discussion

As Using the red electrodes shown in figure 2, "RA" and "LA" could be discriminated with 75% accuracy, but "BA" and "WA" could only be discriminated with 48% (chance) accuracy. Using the green electrodes, shown in figure 2, "RA" and "LA" could only be discriminated with 47% (chance) accuracy, but "BA" and "WA" could be discriminated with 70% accuracy. This demonstrates that phonemes can be classified and discriminated using a high resolution array and that discrete sub-regions, as small as 3mm away from each other, have discrete phonemotopic representation in the frontotemporal area. It should be noted that the orientation of the grid in figure 2 flipped about the horizontal axis as compared to the x-ray images.

Figure 2A shows differentiation between phonemes "RA" and "LA" clustered towards the center of the grid while electrodes showing statistical differences between "BA" and "WA" are spatially distant towards the bottom of the grid in figure 2B. This observation has two implications. It is possible that phonemes are represented in distinct spatial locations on the cortex. Discrete phonemes have, until now, been too spatially close to differentiate using common ECoG grid resolution or low resolution of electroencephalograms (EEG). Using high spatial resolution ECoG grids, we have shown that there is separable information within the cortical potential at electrode spacing resolutions of 3mm. Future investigation of higher resolution grids may show further separability.

IV. Acknowledgements

We would like to acknowledge the patients who donated their time and enthusiasm to be subjects in a time of great personal difficulty. We would also like to acknowledge the clinical staff at Harborview hospital (Seattle, WA).

[1] J. J. Eggermont, "Between sound and perception: reviewing the search for a neural code," *Hear Res,* vol. 157, pp. 1-42, Jul 2001.

[2] H. Sussman, "Phonemic representation: a twenty-first century challenge," *Brain Lang,* vol. 71, pp. 237-40, Jan 2000.

[3] G. Langner, M. Sams, P. Heil, and H. Schulze, "Frequency and periodicity are represented in orthogonal maps in the human auditory cortex: evidence from magnetoencephalography," *J Comp Physiol [A],* vol. 181, pp. 665-76, Dec 1997.

[4] C. Pantev, O. Bertrand, C. Eulitz, C. Verkindt, S. Hampson, G. Schuierer, and T. Elbert, "Specific tonotopic organizations of different areas of the human auditory cortex revealed by simultaneous magnetic and electric recordings," *Electroencephalogr Clin Neurophysiol,* vol. 94, pp. 26-40, Jan 1995.

[5] C. M. Wessinger, J. VanMeter, B. Tian, J. Van Lare, J. Pekar, and J. P. Rauschecker, "Hierarchical organization of the human auditory cortex revealed by functional magnetic resonance imaging," *J Cogn Neurosci,* vol. 13, pp. 1-7, Jan 1 2001.

[6] E. Diesch and T. Luce, "Magnetic fields elicited by tones and vowel formants reveal tonotopy and nonlinear summation of cortical activation," *Psychophysiology,* vol. 34, pp. 501-10, Sep 1997.

[7] E. Diesch and T. Luce, "Topographic and temporal indices of vowel spectral envelope extraction in the human auditory cortex," *J Cogn Neurosci,* vol. 12, pp. 878-93, Sep 2000.

[8] F. W. Ohl and H. Scheich, "Orderly cortical representation of vowels based on formant interaction," *Proc Natl Acad Sci U S A,* vol. 94, pp. 9440-4, Aug 19 1997.

[9] X. Wang, M. M. Merzenich, R. Beitel, and C. E. Schreiner, "Representation of a species-specific vocalization in the primary auditory cortex of the common marmoset: temporal and spectral characteristics," *J Neurophysiol,* vol. 74, pp. 2685-706, Dec 1995.

[10] E. Diesch, C. Eulitz, S. Hampson, and B. Ross, "The neurotopography of vowels as mirrored by evoked magnetic field measurements," *Brain Lang,* vol. 53, pp. 143-68, May 1996.

[11] 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, Jun 2004.

[12] E. Leuthardt, K. Miller, G. Schalk, R. Rao, and J. Ojemann, "Electrocorticography-Based Brain Computer Interface - The Seattle Experience," *IEEE Trans Neural Syst Rehabil Eng,* vol. 14, 2006.

[13] K. J. Miller, M. denNijs, P. Shenoy, J. W. Miller, R. P. Rao, and J. G. Ojemann, "Real-time functional brain mapping using electrocorticography," *Neuroimage,* vol. 37, pp. 504-7, Aug 15 2007.

[14] K. Miller, E. Leuthardt, G. Schalk, R. Rao, N. Anderson, D. Moran, J. Miller, and J. Ojemann, "Spectral changes in cortical surface potentials during motor movement.," *J Neurosci,* vol. 27, pp. 2424--32, Mar 2007.

[15] K. J. Miller, Rao, R.P.N., and Ojemann, J.G., "The Behavioral Split in the Gamma Band," *Neural Engineering, CNE '07. 3rd International IEEE/EMBS Conference on (DOI: 10.1109/CNE.2007.369710);,* pp. p 465-468, 2007.

[16] E. Leuthardt, G. Schalk, J. W. JR, J. Ojemann, and D. Moran, "A brain-computer interface using electrocorticographic signals in humans," *JNEURALENG,* vol. 1, pp. 63-71, 2004.

[17] E. Leuthardt, K. Miller, G. Schalk, R. Rao, and J. Ojemann, "Electrocorticography-based brain computer interface -- the {S}eattle experience.," *{IEEE} Transactions Neur Sys Rehab Eng,* vol. 14, pp. 194--8, Jun 2006.

[18] K. J. Miller, Schalk, G., Leuthardt, E.C., Shenoy, P., Rao, R.P.N., Ojemann, J.G., "Correlation in Paired One-Dimensional, Closed Loop, Overt, Motor Controlled BCI " *Journal of Technical University of Graz, Special Issue: Brain Computer Interfaces,* 2007.

[19] V. Vapnik, *The Nature of Statistical Learning Theory*. New York, 1995.