 DOI:10.3217/978-3-85125-260-6-37

Using ECoG Gamma Activity to Model the Mel-Frequency Cepstral Coefficients of Speech

S. Chakrabarti[1], J. S. Brumberg[2], A. Gunduz[3], P. Brunner[4], G. Schalk[4], D. J. Krusienski[1]
[1]Old Dominion University, Norfolk, VA, USA; [2]University of Kansas, Lawrence, KS, USA;
[3]University of Florida, Gainesville, FL, USA; [4]Wadsworth Center, Albany, NY, USA

Correspondence: S. Chakrabarti, E-mail: schak001@odu.edu

Abstract: In this study, electrocorticographic (ECoG) data collected from eight subjects was analyzed to obtain the spatiotemporal dynamics of the correlations between the high gamma band (70-170 Hz) and the mel frequency cepstral coefficients (MFCC), an important set of speech features used for vowel and word recognition in Automatic Speech Recognition systems. Significant correlations were found between the high gamma band and the mel frequency cepstral coefficients, which resulted in activations in the expected language-related brain areas. Subject-specific spatial linear regression models were then determined to predict the mel frequency cepstral coefficients from the high gamma band power.

Keywords: ECoG, Gamma, Mel Frequency Cepstral Coefficients, Speech

1. Introduction

High gamma activity from ECoG signals has been found to be highly correlated with speech and motor movements in humans. There have been numerous studies investigating cortical activity using ECoG during various speech tasks to identify and characterize the areas of the cortex most involved in speech production and perception [Crone et al., 2001; Towe et al., 2008; Edwards et al., 2010; Pei et al., 2011; Leuthardt et al., 2012], and even decoding of speech from brain activity [Pasley et al., 2012; Blakely et al., 2008; Kellis et al., 2010; Pei et al., 2011]. In automatic speech recognition systems, well-defined features can be extracted from speech signals to aid in the classification of phonemes, vowels, consonants, and words. A popular set of features used for automatic speech recognition (ASR) are the mel frequency cepstral coefficients (MFCCs) [Vergin et al., 1999], which are a parameterization of the spectral content of speech based on perceptual characteristics of human hearing. By parameterizing the speech spectrum using MFCCs, we accomplish two goals: 1) the speech content dimensionality is effectively reduced to the critical acoustic components and 2) the lower dimensionality aids statistical modeling (by alleviating data size demands) and future applications for ECoG-based speech prostheses. The aim of this analysis is to determine the extent to which high gamma band activity from ECoG is correlated with the MFCCs extracted during concurrent speech. We additionally developed spatial linear regression models for predicting MFCCs from the ECoG gamma band as a step toward neural speech decoding.

2. Material and Methods

ECoG activity was recorded during an overt speech task from eight epileptic patients. The number of electrodes per patient ranged from 58-120. For the task, each subject was presented with scrolling text on a computer monitor which they spoke aloud. The ECoG and speech signals were simultaneously recorded at a sampling rate of 9600 Hz using g.USBamps and BCI2000 software. The data was analyzed as follows: (i) Gamma power envelope: The ECoG signals were highpass filtered with a cutoff frequency of 0.01 Hz and re-referenced using a common average reference (CAR) montage. The resulting signals were lowpass filtered and decimated to 400 Hz. A zero-phase FIR filter was applied to the signals to extract the gamma band activity (70–170 Hz; excluding a window of 116–124 Hz to avoid power line interference). The gamma band power envelope was computed using the Hilbert transform. (ii) MFCCs: The speech signal was divided into overlapping frames of length 256 samples, each of which was then windowed by multiplying it with the Hamming window. The powers of the obtained spectrum were mapped onto the mel scale, a transformation that accounts for non-linear pitch perception. The logarithm of the power was then computed from the spectrum and the discrete cosine transform taken from 12 mel frequency bands. The amplitudes of the resulting spectrum are taken as the twelve MFCCs. (iii) Correlations: The spatiotemporal correlations between the gamma band powers (decimated to 20 Hz to match the MFCCs) and the MFCCs were then obtained at temporal lags between -300 ms and 300 ms. The silence periods were removed from the MFCCs and the corresponding lagged gamma signals prior to computing the correlations. Spatial linear regression models were computed to predict the twelve MFCCs from the gamma band power envelopes at zero lag. The *p*-values for the correlations were computed using a randomization test.

3. Results

With modeling, 66.67% of the correlations were statistically significant after Bonferroni correction. No clear modeling trends were apparent across MFCCs or subjects. Fig. 1 shows the –log (p) values of the spatial correlations between the high gamma band (70–170 Hz) power envelope and one of the representative MFCCs (MFCC 3), at lags from -300 ms to 300 ms, and the comparison of the actual MFCC 3 and the predicted MFCC 3 using spatial linear regression for one of the subjects (Subject H).

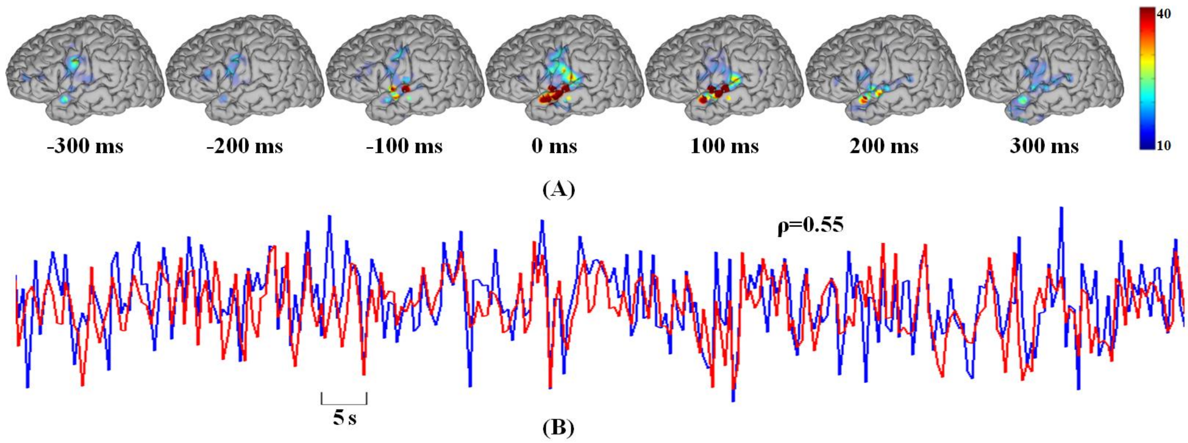

Figure 1. *(A)The spatiotemporal topography of statistically significant gamma-power correlations to MFCC 3 across the entire overt speech production task are shown in 100 ms increments. Color values represent the –log(p) of the correlation. The lower limit is based on statistical significance assessed after a p = 0.05 Bonferroni correction. (B) Blue: Representative time course of MFCC 3 for Subject H. Red: The prediction of MFCC 3 obtained from the linear spatial gamma power model. The Pearson correlation between the two signals over the entire run (178 s) is ρ = 0.55.*

4. Discussion

The results indicate that gamma band power over language-related cortical areas is significantly correlated with MFCCs, perceptual representations of speech. Simple linear spatial models can be used to further enhance the correlations between the gamma activity and the MFCCs. By representing MFCCs using the ECoG signal, a more robust speech representation is utilized as compared to a simple speech power envelope feature or filter bank features that are not perceptually based. MFCCs predicted using our spatial linear regression model can further be used for real-time resynthesis of speech signals. This work develops a natural extension of automatic speech recognition techniques to neurological data, and provides a step toward a neural speech prosthesis.

References

Blakely T, Miller KJ, Rao RPN, Holmes MD, Ojemann JG. Localization and classification of phonemes using high spatial resolution electrocorticography (ECoG) grids. *Conference proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 4964–4967, 2008.

Crone NE, Boatman D, Gordon B, Hao L. Induced electrocorticographic gamma activity during auditory perception. *Clin Neurophysiol*, 112:565-582, 2001.

Edwards E, Nagarajan SS, Dalal SS, Canolty RT, Kirsch HE, Barbaro NM, Knight RT. Spatiotemporal imaging of cortical activation during verb generation and picture naming. *NeuroImage*, 50:291-301, 2010.

Kellis S, Miller KJ, Thomson K, Brown R, House P, Greger B. Decoding spoken words using local field potentials recorded from the cortical surface. *J Neural Eng*, 7:056007, 2010.

Leuthardt EC, Pei X, Breshears J, Gaona C, Sharma M, Freudenberg Z, Barbour D, Schalk G. Temporal evolution of gamma activity in human cortex during an overt and covert word repetition task. *Front Hum NeuroSci*, 6:99, 2012.

Pasley BN, David SV, Mesgarani N, Flinker A, Shamma SA, Crone NE, Knight RT, Chang EF. Reconstructing speech from human auditory cortex. *PLoS Biology*, 10(1):e1001251, 2012.

Pei X, Barbour DL, Leuthardt EC, Schalk G. Decoding vowels and consonants in spoken and imagined words using electrocorticographic signals in humans. *J Neural Eng*, 8:046028, 2011.

Pei X, Leuthardt EC, Gaona CM, Brunner P, Wolpaw JR, Schalk G. Spatiotemporal dynamics of ECoG high gamma activity during overt and covert word repetition. *NeuroImage*, 54:2960-2972, 2011.

Towe VL, Yoon HA, Castelle M, Edgar JC, Biassou NM, Frim DM, Spire JP, Kohrman MH. EcoG gamma activity during a language task: differentiating expressive and receptive speech areas. *Brain*, 131:2013-2027, 2008.

Vergin R, O'Shaughnessy D, Farhat A. Generalized Mel Frequency Cepstral Coefficients for Large-Vocabulary Speaker-Independent Continuous-Speech Recognition. *IEEE Trans Speech Audio Proc*, 7(5):525-532, 1999.