Spatio-Spectral Analysis of ECoG Signals during Voice Activity

Vasileios G. Kanas, Iosif Mporas, Heather L. Benz, Kyriakos N. Sgarbas, Nathan E. Crone, and Anastasios Bezerianos, *Senior Member, IEEE*

***Abstract*— In this paper, we perform spatio-spectral analysis of the human cortex with implanted electrocorticographic (ECoG) electrodes during the voice production process. For this study, the ECoG signals were recorded while the subject performed two-syllable tasks. Additionally, assuming that the speech activity of a subject is expressed as ECoG signal activity disparately distributed over the space of the electrodes, we examined the spectral information in response to the electrode locations. The study was based on spectral features (power spectral density) estimated for each electrode. Quantitative analysis based on the Relief algorithm was followed to estimate the degree of importance of each electrode for describing the voice activity. The experimental results showed that the spectral analysis with resolution of 8 Hz offers the highest voice discrimination performance (94.2%) using support vector machines as classifier. Finally, our analysis showed that during voice activity the frequency bands [168,208] Hz are mostly affected.**

I. Introduction

Brain machine interfaces (BMIs) aim to rehabilitate paralyzed individuals by exploiting the electrical activity of neural ensembles, allowing direct communication between the human brain and an external machine [1]. BMIs have been used in numerous applications for the rehabilitation of paralyzed individuals, such as the direct control of a prosthesis [2]-[4] or advanced communication with locked-in patients[5]-[6]. A silent speech BMI should perform automatic speech recognition and reconstruction to enable disabled individuals to produce spoken words through neural activity. Pei et al. [7] demonstrated the discrimination of vowels and consonants during overt and covert word production using ECoG recordings. They used spectral amplitudes in different frequency bands, estimated via an autoregressive (AR) model, and the local motor potential (LMP) as features, and a Naïve Bayes model as classifier. In [8], EEG recordings were used to discriminate three different tasks (imagined speech of vowels /a/ and /u/ and a no action state). Features were extracted using the common spatial pattern (CSP) method and the support vector machine (SVM) algorithm to classify the three tasks. In another study [9], fast Fourier transform (FFT) and principal component analysis (PCA) were combined to classify a small set of spoken words using microelectrode arrays on the cortical surface. In [6] they recorded multi-unit activity (MUA) to decode auditory parameters for a real-time speech synthesizer from neuronal activity in the left ventral premotor cortex. Except decoding approaches, in a more recent study, Pasley et al. [10] focused on producing spoken words and sentences from ECoG recordings using a stimulus reconstruction model. However, such prosthetic systems have to meet power constraints to be practical for out of the lab conditions.

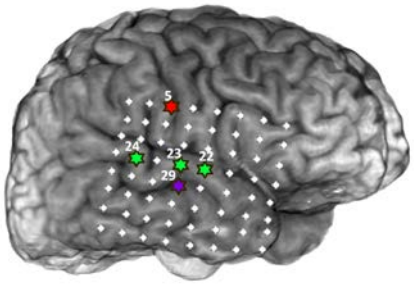

Fig. 1. Electrode locations in the subject. The stars show the five best ECoG electrodes for VAD as calculated by the proposed scheme. Its color correspond to specific brain region (red: ventral sensorimotor cortex, green: superior temporal gyrus, purple: superior temporal sulcus)

A major task in BCI is voice activity detection (VAD). VAD has been studied for more than ten years from the speech technology community [11]. The focus of VAD is to detect the time intervals in which vocal communication is occurring, especially when its acoustic spectrum overlaps that of other sources, such as music or noises. This step aims to the efficient usage of the available resources for speech recognition systems.

Up to now in BCIs the proposed experimental protocols require human intervention to distinguish between speech modes (i.e. speech versus silence), resulting in non-autonomous speech prosthetic systems. However, prosthetic systems will need to determine the voice-epochs autonomously in order to be useful in everyday life. To the authors' best knowledge, no previous work has extensively considered analysis of the human cortex during voice activity for the problem of voice activity detection. In this paper, we study the effect of voice activity both to the spatial and frequency domains. The spatio-spectral analysis is performed with respect to the voice/silence discriminative characteristics of each electrode and frequency band.

TABLE I
SYSTEM PERFORMANCE (%) USING THE N-BEST ECOG FEATURES FOR DIFFERENT FREQUENCY RESOLUTIONS

Frequency resolution	*N-best ECoG Features*								
	10	***50***	***100***	***500***	***1000***	***3000***	***5000***	***10000***	***All***
256 Hz (q = 0)	82.5	84.62	-	-	-	-	-	-	85.99
128 Hz (q = 1)	87.1	88	88.45	-	-	-	-	-	88.25
64 Hz (q = 2)	83.05	88.21	88.53	-	-	-	-	-	89.92
32 Hz (q = 3)	86.07	87.12	88.34	-	-	-	-	-	89.39
16 Hz (q = 4)	88.85	88.8	89.98	90.2	-	-	-	-	90.28
8 Hz (q = 5)	87.5	88.96	91.16	**94.2**	93.88	-	-	-	92.88
4 Hz (q = 6)	87.53	87.59	88.52	91.03	93.92	92.68	-	-	92.25
2 Hz (q = 7)	84.28	85	86.54	87.09	90.16	92.51	91.82	-	90.46
1 Hz (q = 8)	83.1	83.11	83.36	84.1	85	86.04	86.51	87.63	86.76

The remainder of this paper is organized as follows. In Section 2, the data used in the analysis are presented. Next, we describe the parameterization of the ECoG signals, along with the feature selection and classification algorithms which were used. Finally, in Sections 4 and 5 the results and conclusions are respectively presented.

II. DATA COLLECTION

One male patient with intractable epilepsy participated in this study. ECoG electrodes were implanted for one week to localize his seizure focus for resection. The experimental protocol was approved by the Johns Hopkins Medicine Institutional Review Board, and the patient gave informed consent for this research. The subdural array contained 64 electrodes (Ad-Tech, Racine, Wisconsin; 2.3 mm exposed diameter, with 1 cm spacing between electrode centers) was placed according to clinical requirements. Electrodes in the array, shown in Fig. 1, covered portions of the frontal, temporal, and parietal lobes. Data were amplified and recorded through a NeuroPort System (Blackrock Microsystems, Salt Lake City, Utah) at a sampling rate of 10 kHz, low pass filtered with a cutoff frequency of 500 Hz. The patient's spoken responses were recorded by a Zoom H2 recorder (Samson Technologies, Hauppauge, New York), also at 10 kHz and time-aligned with ECoG recordings.

Two syllable tasks were performed by the patient during ECoG recording. The patient was seated in a hospital bed with a computer screen in front of him on a hospital table. Syllable stimuli were presented to the patient using E-Prime software (Psychology Software Tools, Inc., Sharpsburg, Pennsylvania). The patient was instructed to speak each syllable as it was presented. The syllables were constructed from two vowels ("ah" and "ee") and six consonants, which varied by place of articulation and voiced or voiceless manner of articulation ("p", "b", "t", "d", "g", "k"). Each of the 12 syllables was presented 10 times, for a total of 120 trials in each task. Between trials a fixation cross was displayed on the screen for 1,024 ms. In the visual phoneme task, in each trial the patient was presented with one of the syllables spelled on a computer screen. The syllable was presented for 3,072 ms. In the auditory phoneme task, each syllable was sounded for the patient by the computer, after which the patient repeated the syllable. Each trial was 4,000 ms long.

III. SPATIO-SPECTRAL ANALYSIS PROTOCOL

A. Feature Extraction

Prior to any other processing, each recorded dataset was visually inspected and all channels that did not contain clean ECoG signals were excluded, leaving 55 channels for our analysis. To eliminate noise common to all channels, recorded data from each ECoG electrode were re-referenced by subtracting the common average (CAR) [12], as follows,

$$X_{CAR}^{ch} = X^{ch} - \frac{1}{N_{ch}} \sum_{m=1}^{N_{ch}} X^{m} \qquad (1)$$

where X^{ch} and X_{CAR}^{ch} are the ECoG and CAR referenced ECoG amplitudes on the *ch-th* channel out of a total of N_{ch} recorded channels. ECoG signals of each channel were also normalized by subtracting the average value and dividing with the standard deviation.

To extract the spectral features, each ECoG channel was segmented by applying a sliding Hamming window with length 256 samples and shifting step 128 samples. For each frame and channel the power spectral density (PSD) was estimated based on the FFT [13]. In this study, we computed the PSD for frequencies between 0 Hz and 256 Hz in 1 Hz bins. Power estimates in the whole frequency range were log-transformed to approximate normal distributions. To examine the optimal spectral resolution necessary for the VAD task, the PSD values were averaged in 2^q frequency bands to obtain the final spectral features per ECoG channel, with $0 \le q \le 8$, resulting in nine different experimental setups. Subsequently, a total of 55-14080 spectral features (depending on the number of the averaged frequency bands) were used during the analysis.

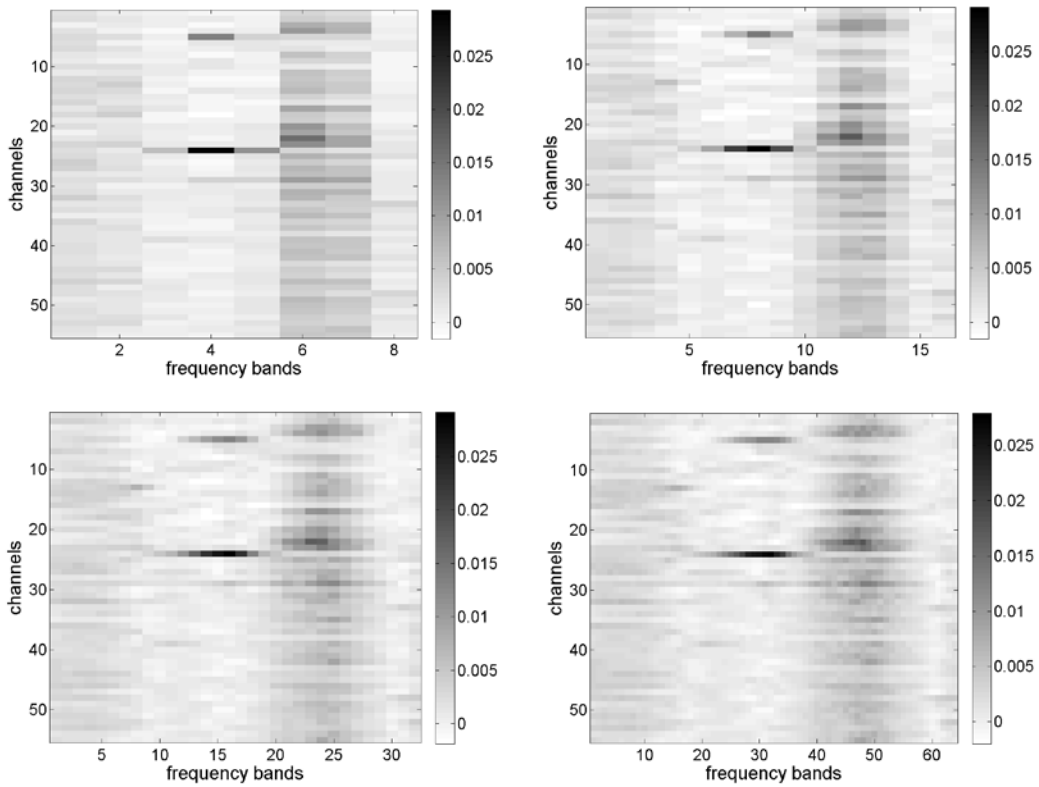

Fig. 2. Feature ranking maps as calculated by the Relief algorithm. From top to bottom and from left to right the feature ranking maps represent the ranking scores per channel and frequency band for q= 3, 4, 5, and 6.

B. Feature Evaluation and Classification

The spatio-spectral analysis is performed in terms of the voice discrimination ability of each extracted feature. Separately for each of the feature vector sets the corresponding PSD features were evaluated using the Relief algorithm [14]. The Relief algorithm evaluates the worth of a feature by iteratively sampling an instance and considering the value of the given feature for the nearest instance of the same and different class. Using the feature ranking results, we evaluated the N-best features for the VAD task, i.e. the binary classification problem between voice/silence.

For the classification, we relied on the support vector machines (SVMs) method implemented using the sequential minimal optimization algorithm. In this study, for the SVM kernel we used the radial basis function (RBF). The RBF values C =10.0 and γ =0.01 were found to offer optimal classification performance after a grid search at C = {1.0, 5.0, 10.0, 20.0} and γ = {0.001, 0.01, 0.1, 0.5, 1.0, 2.0}. The evaluation of the results was performed using 10-fold cross validation, in order to avoid overlapping between the training and the test data.

IV. Analysis of Results

As a first step in our analysis we examined the voice discrimination performance for the nine experimental setups using the N-best features is shown in Table I. The best classification accuracy for different number of N-best features was achieved for $q = 5$, demonstrating that the highest spectral information could be utilized when PSD is averaged in 32 frequency bands, i.e. with frequency resolution of 8Hz. The use of more or fewer frequency bands resulted in suboptimal performance. Additionally, the use of the 500-best ECoG features provided the optimal classification accuracy (94.20%).

Using the ranking scores calculated by the Relief algorithm, we constructed, for each experimental setup, the feature ranking maps (Fig. 2), which depict the ranking scores per channel and frequency band. These plots indicate the channels and frequency bands which underlie most of the information about the speech activity. For q = 5, it is observed that the most information is held in high frequencies ranging between 152-216 Hz. Channel 24 held information in lower frequencies ranging between 88-144 Hz. The most informative feature is the average 120-128 Hz high gamma power spectrum of channel 24, with a ranking score of 0.029 as calculated by the Relief algorithm. Moreover, in order to distinguish the most informative channels and frequencies, we averaged the feature ranking map, corresponding to the optimal accuracy (q=5), across the different ECoG channels and frequencies (Fig. 3). Fig. 3(a) illustrates the average ranking scores per frequency band. It is noticeable that there are two distinct informative regions. The first region, having the highest ranking scores, is traced in the high frequency bands (168-208 Hz) while the

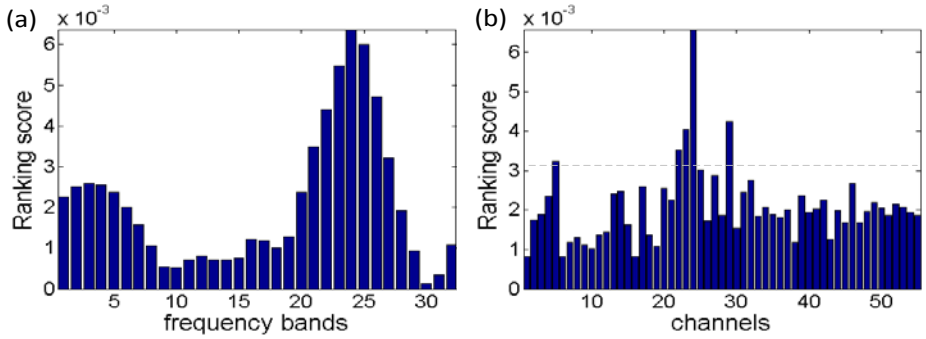

Fig. 3. (a) The average ranking scores per frequency for q = 5. The spectral information is located in low (0-48 Hz) and high (168-208 Hz) frequencies. (b)The average ranking scores per channel for q = 5. The five best channels are 24, 29, 23, 22 and 5 (The dashed line shows the threshold).

other one is traced in lower frequencies (0-48 Hz). Fig. 3(b) shows the average ranking scores per channel. The five dominant electrodes are the 24, 29, 23, 22 and 5, marked in Fig. 1 with color-coded stars. These electrodes were located in cortical areas typically involved in speech and language processing. In detail, channel 5 was located over the ventral sensorimotor cortex Channels 24, 23 and 22 were located over the posterior superior temporal gyrus (STG), which contains auditory association cortex and is part of Wernicke's area, typically important for speech perception. Channel 29 was located over superior temporal sulcus, which is vital to the processing of speech.

V. Conclusions

In this paper, we studied the spatial and spectral characteristics of the cortex, as described by ECoG signals, during voice activity. The analysis was based on the assumption that different channels, i.e. cortex locations, contain distinct information in specific frequencies. The spatial analysis showed that channels 24, 29, 23, 22 and 5 hold most of the information for the voice activity. These electrodes correspond to the cortical areas responsible for speech perception and articulation. In the voice/silence separation task the highest classification accuracy (94.2%) was achieved using the 500-best ECoG from 32 frequency bands spanning 0-256 Hz and 55 channels, which corresponds to frequency resolution of 8Hz. Furthermore, the feature ranking maps suggested that distributed locations hold information about speech activity. That is, language processing involves large-scale cortical areas that are engaged in phonological analysis, speech articulation and other processes, which is in agreement with [15].

More extensive training using larger datasets acquired from different subjects is expected to further improve the analysis of the effect of voice activity on the cortex and consequently offer the knowledge to accurately detect and analyze voice activity, and thus advance automated natural speech BMIs.

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, Jun. 2002, pp. 767–791.

[2] H. Benz, H. Zhang, A. Bezerianos, S. Acharya, N.E. Crone, X. Zheng, and N.V. Thakor, "Connectivity analysis as a novel approach to motor decoding for prosthesis control," IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 20, Mar. 2012, pp. 143-152.

[3] M.S. Fifer, S. Acharya, H.L. Benz, M. Mollazadeh, N.E. Crone, N.V. Thakor, "Toward Electrocorticographic Control of a Dexterous Upper Limb Prosthesis: Building Brain-Machine Interfaces," IEEE Pulse, vol. 3, Jan. 2012, pp. 38-42.

[4] M.L. Stavrinou, L. Moraru, L. Cimponeriu, S. Della Penna, and A. Bezerianos, "Evaluation of Cortical Connectivity During Real and Imagined Rhythmic Finger Tapping," Brain topography, vol. 19, Mar. 2007, pp. 137-145.

[5] X. Pei, J. Hill, and G. Schalk, "Silent communication: Toward using brain signals," IEEE Pulse, vol. 3, Jan. 2012, pp. 43-46.

[6] F.H. Guenther, J.S. Brumberg, E.J. Wright, A. Nieto-Castanon, J.A. Tourville et al., "A wireless brain–machine interface for real-time speech synthesis," PloS Biology, vol. 4, Dec. 2009, e8218

[7] X. Pei, D.L Barbour, E.C. Leuthardt, and G. Schalk, "Decoding vowels and consonants in spoken and imagined words using electrocorticographic signals in humans," Journal of Neural Engineering, vol. 8, Aug. 2011, 046028.

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

[9] 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," Journal of Neural Engineering, vol. 7, Oct. 2010, 056007.

[10] B.N. Pasley, S.V. David, N. Mesgarani, A Flinker, S.A. Shamma, N.E. Crone, R.T. Knight, and E.F. Chang, "Reconstructing Speech from Human Auditory Cortex," PloS Biology, vol. 10, Jan. 2012, e1001251.

[11] J. Chang, N.S Kim, S.K. Mitra, "Voice Activity Detection Based on Multiple Statistical Models, " IEEE Trans on Signal Processing, vol. 54, Jun. 2006, pp. 1965-1976.

[12] D. Goldman, "The clinical use of the 'average' reference electrode in monopolar recording," Electroencephalogr. Clin. Neurophysiol., vol. 2, May 1950, pp. 209-212.

[13] R.N. Bracewell, "The Fourier transform," Sci. Am., vol. 260, Jun. 1989, pp. 86-9, 92-5.

[14] I. Kononenko. "Estimating Attributes: Analysis and Extensions of RELIEF," In Proc. of the European Conference on Machine Learning, 1994, pp. 171-182.

[15] A. Korzeniewska, P.J. Franaszczuk, C.M. Crainiceanu, R. Kuś, and N.E. Crone, "Dynamics of large-scale cortical interactions at high gamma frequencies during word production: Event related causality (ERC) analysis of human electrocorticography (ECoG)," NeuroImage, vol. 56, Jun. 2011, pp. 2218–2237.