34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

Spoken Sentences Decoding Based on Intracranial High Gamma Response Using Dynamic Time Warping

Dan Zhang, Enhao Gong, Wei Wu, Jiuluan Lin, Wenjing Zhou and Bo Hong*, *Member, IEEE*

Abstract— **In this study, we explore the discriminability of high gamma activities from speech production cortex during the overt articulation of two sentences. Neural activities were recorded from one intracranial electrode placed approximately over the posterior part of the inferior frontal gyrus. By employing a dynamic time warping (DTW) method to realign single-trial high gamma response during speech productions, averaged temporal activation patterns corresponding to the two spoken sentences were obtained. Single-trial ECoG responses were subsequently classified according to their correlations with these two temporal activation patterns. On average, 77.5% of the trials were correctly classified, which was much higher than the chance-level performance of the SVM classifier without DTW. Our preliminary results shed light on the construction of cortical speech brain-computer interfaces on the sentence level.**

I. Introduction

Brain-computer interfaces (BCIs) aim at helping the severely motor disabled people to communicate with the external world [1-2]. Till now, self-regulation of motor related functions and attention modulation of sensory responses have been extensively studied for BCI control, with the majority using the non-invasive electroencephalography (EEG) technology. In contrast, although speech is the most effective communication channel for human beings, BCI studies using speech-related brain activities are limited: it is difficult to characterize the complex and distributed speech network using EEG due to its relatively low spatial resolution.

Recently, it has been suggested that intracranial EEG (electrocorticography, ECoG) could be a suitable candidate toward building speech BCIs [3-6]. ECoGs are recorded directly from the surface of the human cortex, having both high spatial resolution (~mm) and high temporal resolution (~ms). In addition, low voltage, high frequency oscillations (i.e. >40 Hz) that cannot be easily seen in EEG can be clearly observed using ECoG. It has been demonstrated that the production of different phonemes, including vowels and consonants, can be discriminated using ECoG in the high gamma range (70-170 Hz) from a variety of brain regions, including the superior and middle part of temporal lobe, Wernicke's area, Broca's area, premotor cortex, etc. [5, 7-8].

While previous speech BCI studies focused on the phoneme-level speech production, BCI classifications can also be carried out on the sentence level. The sentence level BCI paradigm may facilitate BCI performance for the following reasons: 1) better user experience is expected as it is more natural to speak a complete and meaningful sentence than single phonemes; 2) the rich temporal information of sentences may provide additional information for classification [9].

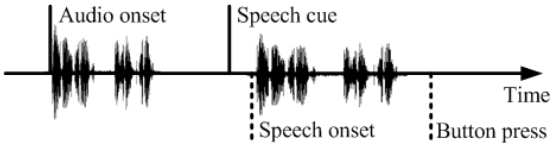

Figure 1. Scheme of one experiment trial

In this study, we investigated whether different spoken sentences can be distinguished using the simultaneously recorded ECoG signals. We hypothesized that the temporal structures of different spoken sentences could result in discriminable temporal activation patterns of high gamma oscillations at brain areas related to speech production, providing information for BCI classification. ECoG data were obtained from one epilepsy patient with subdural electrodes placed over the left frontal and temporal lobes. During the experiment, the patient was asked to overtly speak one of two 8-character sentences in Chinese. We analyzed the ECoG signals from one subdural electrode placed approximately over the posterior part of the inferior frontal gyrus, which showed the largest speech-related high gamma responses. Specifically, in consideration of the large variability in time and speed of the single-trial speech articulation, the dynamic time warping (DTW) method that has been widely employed in speech signal processing [13] was introduced for constructing the BCI classifier articulation. By using the high gamma oscillations from one subdural electrode, 77.5% of the trials were correctly recognized.

II. Methods

A. Paradigm and Procedure

A delayed sentence-repeating paradigm was used in our experiment (Fig. 1). At the beginning of each trial, the patient first heard one of two 8-character sentences (denoted as sentence A and B) in Chinese through computer speakers while he was fixating on a white cross presented on the computer screen. The two sentences were famous Chinese proverbs: A) 种瓜得瓜，种豆得豆 (As a man sows, so he shall reap); B) 十年树木，百年树人 (it takes a decade to become a tree, and a century to become a man). After a delay of 1.2 s to 1.5 s following the offset of the auditory presentation, the patient received a visual cue (the color of the fixation cross changed from white to red) instructing him to repeat the heard sentence verbally. The patient was required to

*Research supported by the National Natural Science Foundation of China under grant #61071003 and #61101151.

B. Hong is with the Department of Biomedical Engineering, Tsinghua University, Beijing 100084, China (e-mail: hongbo@tsinghua.edu.cn).

D. Zhang, E. Gong, and W. Wu are with the Department of Biomedical Engineering, Tsinghua University, Beijing 100084, China.

J. Lin and W. Zhou are with the Department of Neurology, Affiliated Yuquan Hospital, Tsinghua University, Beijing 100084, China.

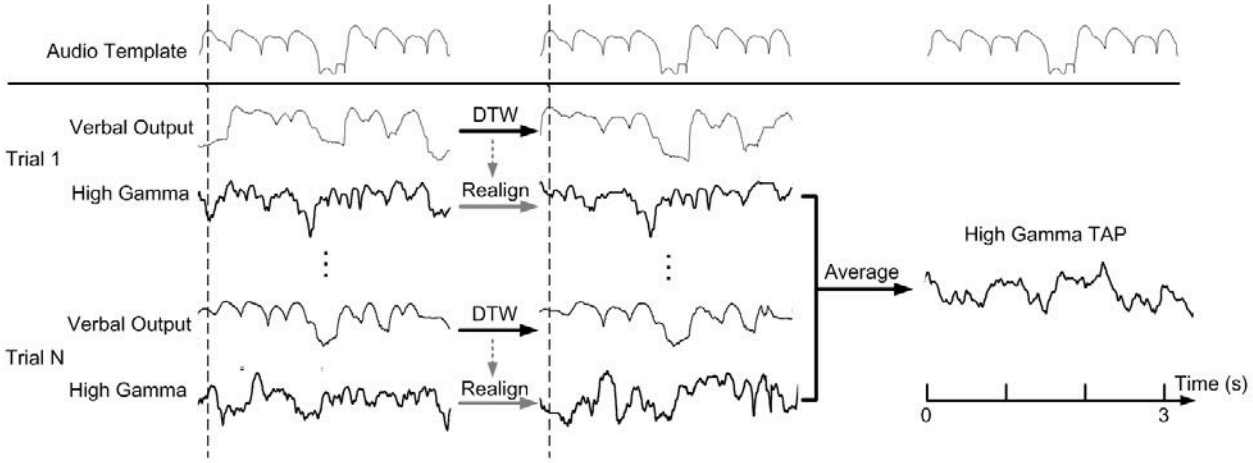

Figure 2. Procedure for obtaining the averaged temporal activation pattern (TAP) using dynamic time warping (DTW).

press a button when the verbal task was finished. 40 trials were run in total, with 20 trials per sentence presented in a random order. The experiment program was implemented in Matlab (the Mathworks, USA) using Psychophysics Toolbox 3.0 extensions [10].

B. Data Preprocessing

In this study, we focused on the power envelop of the high gamma oscillations, as the high gamma oscillations were previously reported to be highly correlated with speech functions [5, 7]. Specifically, the ECoG data were first band-pass filtered to 60-90 Hz and then transformed into analytic signals by Hilbert transform. The time-varying high gamma power envelops were obtained by taking the amplitudes of the Hilbert transformed data. The high gamma power envelops were log-transformed to follow approximately normal distributions [11]. Afterwards, the log transformed power envelops of 1 s duration prior to the speech cue were used as the baseline to normalize the power envelops following the speech cue: the power envelops following the speech cue were transformed into z-scores by subtracting the mean and being divided by the standard deviation of the baseline data segment. The subdural electrode with the largest high gamma responses was chosen for further analysis.

Since we hypothesized that the neural activities of the two spoken sentences can be discriminated by their temporal activation patterns of the high gamma responses at brain areas related to speech production, it was therefore likely for these high gamma responses to follow the verbal output. However, it was unlikely for high gamma activities at the cortex level to represent the verbal output in every detail [9]. Rather, an optimal correlation should presumably be achieved at a coarse time scale, using a properly defined sliding time window to smooth the data. As the possible difference between the time courses of the high gamma responses was considered to originate from the verbal responses, the time window width showing the maximal correlation between high gamma power and verbal output was deemed as the optimal scale for classification. To determine the optimal time window width, the high gamma power envelopes and the corresponding verbal output envelops were each smoothed by a sliding window with the same time width ranging from 50 ms to 1000 ms; the correlation coefficient between the smoothed envelopes was then calculated. The z-score transformed high gamma power envelops were smoothed at the optimal time window width before the following analysis.

C. Extracting the Temporal Activation Patterns (TAPs) Using Dynamical Time Warping (DTW)

Ideally, there are temporal activation patterns (TAPs) that represent the stereotypical ECoG response patterns for each of the two sentences. Single-trial ECoG signals could then be classified according to their correlations with the TAPs. Nonetheless, the variability in time and speed of single-trial speech production posed a significant challenge for TAP estimation; computing the TAPs by simply averaging the ECoG signals across trials was likely to be detrimental to the original temporal structure of the spoken sentences. To remedy this issue, the DTW method was utilized to first realign different trials in time according to the patient's verbal responses. Briefly speaking, with the acoustic signal of each presented sentence as the template, we employed the DTW to find a nonlinear transformation solution in the time domain in order to achieve the optimal match of single-trial verbal responses to the audio template. The same transformation solution was identically applied to the corresponding ECoG high gamma envelope signal, resulting in realigned neural responses. The TAPs of the two sentences used in the experiment were then calculated as the average of the realigned single-trial cortical responses in the corresponding trials. The procedure is shown in Fig. 2.

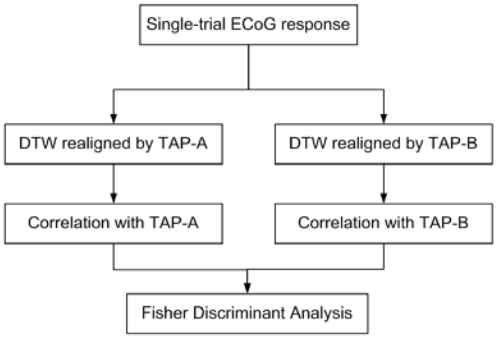

Figure 3. The classification procedure.

D. DTW based Classification

In the previous section, DTW was applied on audio signals (i.e. patient's verbal responses) and the resulted transformation solutions were used to realign the single-trial ECoG responses before averaging. For classification, a second round of DTW was applied to realign single-trial ECoG responses to the two TAPs corresponding to the two sentences. Then the correlation coefficients between the DTW-realigned single-trial responses and the two TAPs were calculated and taken as the features for BCI classification, thus forming a two-dimension feature vector (i.e. single-trial correlation with TAPs of sentence A and B). Fisher Linear Discriminant Analysis was subsequently employed for feature classification. The discriminability between the two sentences was evaluated using leave-one-out cross validation, where the TAPs were re-estimated from the training dataset for each classification. The classification procedure is summarized in Fig. 3.

To validate the effectiveness of the DTW-based classification method, another classifier was constructed using the original ECoG high gamma envelop as features to train a support vector machine (SVM) using a linear kernel function, without temporal adjustment using DTW.

E. Patient and ECoG Recordings

The experiment was conducted with one patient (male, 38 years old) who suffered from intractable epilepsy and underwent temporary placement of intracranial ECoG electrode arrays to localize seizure foci prior to surgical resection. Prior to the implantation of electrodes, the patient gave written informed consent for his involvement in research. This study was approved by the Research Ethics Committee of Tsinghua University and the affiliated Yuanquan Hospital.

For this patient, two 32-electrode grids (4 mm electrode diameter and 1 cm inter-electrode distance) were placed over the frontal and temporal lobe (see Fig. 3a). Grid placement and duration of ECoG monitoring were entirely based on the clinical requirements, without any consideration of this study.

ECoG signals were recorded using a g.USBamp amplifier/digitizer system (g.tec, Graz, Austria). The amplifier sampled the 64-channel (2 × 32-electrode grids) signal at 1200 Hz with a high-pass filter of 0.1 Hz cutoff frequency and a notch filter at 50Hz to remove the power line noise. Four inactive epidural electrodes facing the skull were employed as the ground and the reference. In addition, the patient's verbal responses were synchronously recorded as one channel in the g.USBamp system.

For localizing the ECoG electrodes, the stereotactic coordinates of the electrodes were identified based on the patient's lateral skull radiographs (acquired by Siemens SOMATOM Sensation 64 CT), using the LOC toolbox [12].

III. Results

The patient performed all 40 trials of sentence-repeating task correctly without mistake. On average, the patient spent approximately the same time to speak both the two sentences (sentence A vs. B: 2.8 ± 0.2 s vs. 2.6 ± 0.2 s, $p>0.05$, t-test). During the sentence articulation period, strong high gamma responses were found in the posterior part of the inferior frontal gyrus (Fig. 4a). The electrode with the strongest high gamma response (marked by the arrow) was chosen for the following analysis. From its time-frequency plot (Fig. 4b), significant increase of high gamma power was observed.

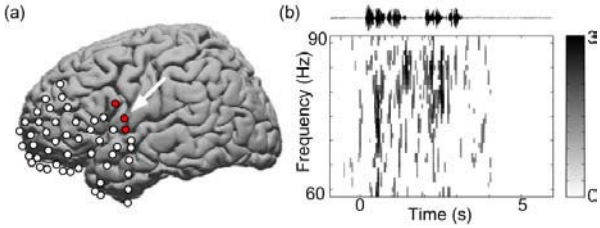

Figure 4. (a) Approximate electrode locations; the 3 electrodes with significant high gamma power increases were marked in red; the electrode pointed by the white arrow was chosen for the following analysis. (b) the time-frequency plot of the chosen electrode during sentence articulation; high gamma power changes were shown in z-score, only time-frequency bins with significant changes were shown ($p<0.05$, false detection rate corrected).

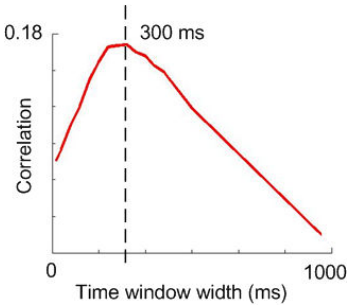

Figure 5. Correlation between the verbal output and the high gamma power as a function of the time window width at the chosen electrode.

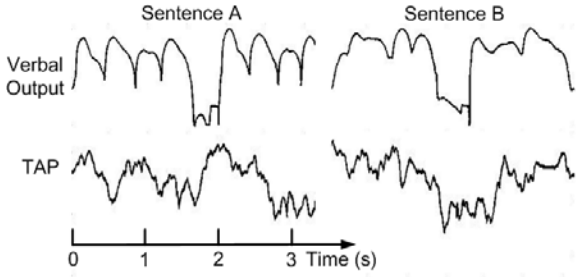

Figure 6. The verbal output and the temporal activation pattern of the two spoken sentences

Fig. 5 shows the correlation between the verbal output and the high gamma power at the chosen electrode as a function of the time window width. The maximal correlation was found at the time window width of 300 ms (0.16, $p<0.0001$), which was used as the optimal time window for smoothing the ECoG data.

Based on the selected electrode and time window width, the temporal activation patterns for the two spoken sentences are illustrated in Fig. 6. From the verbal output, it was clear that the two sentences were of different temporal structures, and the TAP from the ECoG data showed similar temporal structures with the verbal output.

Using the correlation coefficients between the DTW realigned single-trial ECoG response and the two temporal

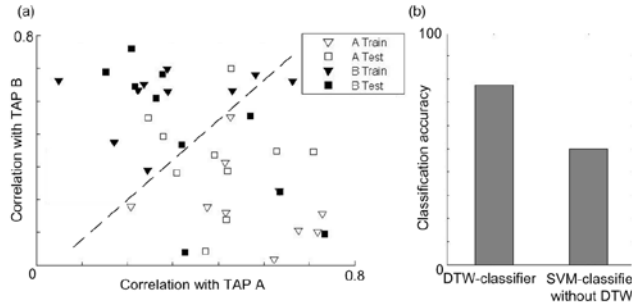

Figure 7. (a) Representation of the two classes in the feature space. (b) The classification accuracies of algorithms with / without DTW.

activation patterns as features, the cross validation revealed that 75% of sentence A and 80% of sentence B were correctly recognized. The discriminability of the two sentences is illustrated in Fig. 7a, in which half of the trials were used for obtaining the TAPs and the other half for testing. It was clear that "A" trials were of higher correlations with TAP A and vice versa. In contrast, the SVM classifier without realignment only performed at the chance level (50% by leave-one-out cross validation, Fig. 7b).

IV. Discussion and Conclusion

In this study, the temporal activation patterns during the articulation of different sentences were extracted from the high gamma response recorded from an ECoG electrode at the posterior part of the inferior frontal gyrus. By applying the DTW method to realign the single-trial ECoG responses, temporal activation patterns of the two spoken sentences were obtained. A DTW-based classifier was then constructed on the basis of the temporal activation patterns. For the two -sentence classification problem, an average classification accuracy of 77.5% was achieved.

DTW method has been widely used for speech processing [13]. Here we used DTW for realigning not only the audio signals (i.e. verbal responses) but also the single-trial ECoG responses. Compared to the SVM classifier without realignment of the ECoG data, the DTW-based classifier obtained much higher classification accuracy (77.5% vs. 50%). These findings suggested that the single-trial ECoG responses of the same sentence indeed shared similar temporal structures that can be captured by the DTW method.

Here the classification was performed using one electrode from the posterior part of the inferior frontal gyrus, which was within the Broca's territory [9]. Therefore, we conjecture that the ECoG activities used for classification reflected the motor control for speech production. Given the fact that the classification results were achieved using a classifier based on the temporal activation patterns for feature extraction, the discriminability of the two sentences was likely to rely on the different temporal structure of the sentences. The 300 ms time window used for data preprocessing thus might indicate that the differences between the two classes originated from the syllable level speech motor controls.

While previous ECoG studies on speech decoding focused on phoneme level processing [5, 8, 15], our results for the first time showed the possibility to 'decode' the spoken sentences using their temporal structures as a new feature for classification. The proposal of using such a feature was based on the observations that such temporal structures were partly preserved on the high gamma oscillations of the Broca's area. The articulation of sentences also involves other brain regions such as the premotor cortex, cerebellum [14], which were not seen due to the limited coverage of the ECoG electrode in the present study. With more information from different speech-related brain regions available, speech BCIs may assist people to communicate in a natural way by directly using their speech related brain activities.

References

[1] J. Wolpaw, *et al.*, "Brain-computer interfaces for communication and control," *Clin Neurophysiol,* vol. 113, pp. 767-91, Jun 1 2002.

[2] M. Lebedev and M. Nicolelis, "Brain-machine interfaces: past, present and future," *Trends Neurosci,* vol. 29, pp. 536-46, Sep 1 2006.

[3] E. C. Leuthardt, *et al.*, "A brain-computer interface using electrocorticographic signals in humans," *Journal of Neural Engineering,* vol. 1, pp. 63-71, Jul 01 2004.

[4] K. J. Miller, *et al.*, "Cortical activity during motor execution, motor imagery, and imagery-based online feedback," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 107, pp. 4430-4435, Apr 02 2010.

[5] X. Pei, *et al.*, "Decoding vowels and consonants in spoken and imagined words using electrocorticographic signals in humans.," *Journal of Neural Engineering,* vol. 8, p. 046028, Aug 2011.

[6] G. Schalk, *et al.*, "Decoding two-dimensional movement trajectories using electrocorticographic signals in humans," *Journal of Neural Engineering,* vol. 4, pp. 264-275, Sep 01 2007.

[7] E. C. Leuthardt, *et al.*, "Using the electrocorticographic speech network to control a brain-computer interface in humans," *Journal of Neural Engineering,* vol. 8, p. 036004, May 07 2011.

[8] S. Kellis, *et al.*, "Decoding spoken words using local field potentials recorded from the cortical surface," *Journal of Neural Engineering,* vol. 7, p. 056007, Sep 01 2010.

[9] G. Hickok and D. Poeppel, "The cortical organization of speech processing," *Nature Reviews Neuroscience,* vol. 8, pp. 393-402, Jun 01 2007.

[10] D. H. Brainard, "The psychophysics toolbox," *Spatial Vision,* vol. 10, pp. 433-436, 1997.

[11] T. Gasser, *et al.*, "Transformations towards the normal distribution of broad band spectral parameters of the EEG," *Electroencephalogr Clin Neurophysiol,* vol. 53, pp. 119-24, Jan 1982.

[12] K. J. Miller, *et al.*, "Cortical electrode localization from X-rays and simple mapping for electrocorticographic research: The "Location on Cortex" (LOC) package for MATLAB," *J Neurosci Methods,* vol. 162, pp. 303-8, May 15 2007.

[13] L. Rabiner and B. H. Juang, *Fundamentals of speech recognition* vol. 103: Prentice hall, 1993.

[14] M. G. Peeva, *et al.*, "Distinct representations of phonemes, syllables, and supra-syllabic sequences in the speech production network," *Neuroimage,* vol. 50, pp. 626-38, Apr 1 2010.

[15] B. N. Pasley, *et al.*, "Reconstructing Speech from Human Auditory Cortex," *Plos Biol*, vol. 10, no. 1:e1001251, Jan 2012.