Time-varying parametric modeling of ECoG for syllable decoding

Vasileios G. Kanas[1], Iosif Mporas[1,2], Griffin W Milsap[3], Kyriakos N Sgarbas[1], Nathan E Crone[4] and Anastasios Bezerianos[5,6]

[1] Dept. of Electrical and Computer Engineering, University of Patras, Patras, Greece
[2] Computer Informatics Engineering Dept., TEI of Western Greece, Patras, Greece
[3] Dept. of Biomedical Engineering, Johns Hopkins University, Baltimore, MD 21205 USA
[4] Dept. of Neurology, Johns Hopkins University, Baltimore, MD 21205 USA
[5] Dept of Medical Physics, School of Medicine, University of Patras, Patras, Greece
[6] Singapore Institute for Neurotechnology, National University of Singapore, Singapore
Email: vaskanas@upatras.gr (V.G. Kanas)

Abstract. As a step toward developing neuroprostheses, the purpose of this study is to explore syllable decoding in a subject with implanted electrocorticographic (ECoG) recordings. For this study, we use ECoG signals recorded while a subject volunteered to perform a task in which the patient has been visually cued to speak isolated consonant-vowel syllables varying in their articulatory features. We propose a recursive estimation method to calculate the parametric model coefficients in each time instant and band power features from individual ECoG sites are extracted to decode the articulated syllables. Our findings may contribute to the development of brain machine interface (BMI) systems for syllable-level speech rehabilitation in handicapped individuals.

Keywords: electrocorticography, time-varying autoregressive model, speech rehabilitation, brain machine interface

1 Introduction

Speech decoding directly or indirectly from neural activity has been extensively studied for several years. The motivation of such approaches is, on the one hand, to examine the underlying functionality of the human brain during speech articulation and, on the other hand, to restore speech capabilities in severely handicapped individuals, such those suffering from disorders of consciousness [1]. Currently, electroencephalography (EEG), intracranial electrocorticography (ECoG), and intracortical microelectrode recordings have been utilized as neurophysiological recording techniques speech restoration.

Due to its non-invasive nature, EEG has been widely used in human studies to record brain activity for indirect control of spelling devices. In particular, approaches based on letter or word selection have been proposed utilizing specific brain waveforms, such as slow cortical potentials [2-4], the P300 event-related potential (ERP)

[5-8], steady state visual evoked potentials (SSVEP) [9, 10], sensorimotor rhythms (SMR) [11] and event-related (de)synchronization (ERD/ERS) [12-14]. On the contrary, relevant studies have focused on the challenging task of decoding continuous, dynamic speech, at various levels including formants [15], phonemes [12, 16], words [17], and even sentences [18]. However, speech decoding at the level of syllables has been paid less attention in the literature. In a related study [19], the authors support that speech rhythms can be identified using scalp-recorded EEG brain signals. In their work, Hilbert spectra are applied to decode three rhythmic structures for two different syllables (e.g. /ba/ and /ku/), without decoding the syllables itself. The aforementioned approaches for speech rehabilitation have been implemented using either time invariant parametric models, or filtering techniques, or short time spectral representations to isolate spectro-temporal features for classifying speech related events. In the current study, we focus on modeling the non-stationary ECoG brain activity proposing a time-varying autoregressive method (TVAR) for deriving ECoG spectral features to be used in the decoding of twelve individually spoken syllables.

Since brain signals are non-stationary containing numerous time-variant and transient components associated with underlying physiological activities, parametric modeling is usually used to study such data, as it allows the understanding of the underlying brain dynamics and mechanisms [20]. The main advantage of these models is that they avoid spectral leakage compared to non-parametric methods that use windowing like the fast Fourier transform. In addition, time-varying parametric models generate instantaneous estimates of power spectrum to perform non-stationary spectral analysis, providing higher time-frequency resolution [21]. While a detailed review of the methods for estimation and/or identification of time-varying model parameters can be found in [22], here, we describe briefly the main categories.

In general, the parametric methods can be categorized into three distinct classes according to the type of structure imposed on the evolution of the time-varying model parameters: 1) unstructured parameter evolution methods 2) stochastic parameter evolution methods and 3) deterministic parameter evolution methods (basis expansion modeling). The first category imposes no constraints to the parameter evolution, with short-time (sliding window based) and recursive approaches being prime methods within this class. However, the selection of the window size is vital to the performance of the sliding-window approach and, until now, there is no automatic method to guarantee an optimal selection. The second category of parameter estimation methods imposes stochastic smoothness constraints to the parameter evolution; approaches based on Kalman filtering constitute this category. Finally, the approaches belong to the third category aim to represent the evolution of the model parameters by deterministic functions belonging to known functional subspaces. In real-world applications, however, the model parameters of Kalman filtering or the basis functions in basis expansion modeling are often unknown and they have to be estimated or chosen by trial and error [23-25]. Therefore, the above categories are highly dependent on these parameters, and their performance will be considerably degraded when the parameters are inappropriately chosen. Motivated by this, in this paper, an adaptive unstructured approach, namely recursive maximum likelihood (RML), to estimate the parameters of the TVAR model is employed.

2 Materials and Methods

2.1 Neurosurgical patient and experimental design

One patient with medically refractory epilepsy participated in this study. ECoG electrodes (Ad-Tech, Racine, Wisconsin; 2.3 mm exposed diameter, with 1 cm spacing between electrode centers) were surgically implanted on the brain surface (subdural, epipial) to map eloquent cortex and localize each patient's ictal onset zone prior to resective surgery. Electrode placement was dictated entirely by clinical needs. The patient provided informed consent to participate in the experimental protocol, which was approved by the Johns Hopkins Medicine Institutional Review Board. Localization of ECoG electrodes after surgery was performed using Bioimage by co-registration of pre-implantation volumetric MRI with post-implantation volumetric CT [26] (shown in Figure 1). An 8 x 8 grid of electrodes covered portions of the right frontal, temporal, and parietal lobes. There were no known auditory or language deficits in the patient.

The subject was seated in a hospital bed facing a laptop sitting on a hospital table in front of them. Textual representations of syllable stimuli (TAH, DEE, etc.) were presented to the patients using E-Prime software (Psychology Software Tools, Pittsburgh, PA) (Psychology Software Tools, Inc., Sharpsburg, Pennsylvania). The patients were instructed to speak each syllable as it was presented. The syllables were constructed from two vowels (/a/ and /i/) and six consonants (/p/, /b/, /t/, /d/, /g/, /k/) which varied by place of articulation and voiced or voiceless manner of articulation. 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. Each syllable was presented for 3,072 ms.

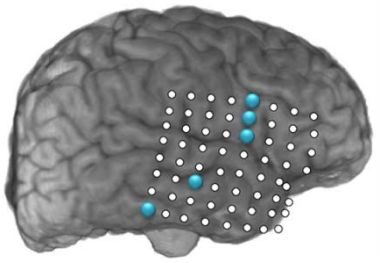

Fig. 1. ECoG electrode locations for the subject. The electrode grid covered areas of temporal, parietal, and frontal lobe in the right hemisphere that are analogues of language-processing areas in the left hemisphere. The blue circles represent the most informative channels as calculated by the Relief algorithm (see Section 4). The five best-ranked electrodes, are located in cortical areas relevant to the speech task.

2.2 ECoG recording and preprocessing

ECoG data was amplified and digitized with a sampling rate of 10 kHz using a 128-channel NeuroPort (Blackrock Microsystems, Salt Lake City, Utah). The ECoG re-

cordings were pre-processed with a low-pass filter at 500 Hz before down-sampling to 1000 Hz. Notch filtering was also applied to attenuate interference at 60 Hz and its harmonics from power lines. Afterwards, the recorded signals were visually inspected for noise and motion artifacts. Noisy ECoG channels were excluded, leaving channels $M = 55$ (out of 64). To reduce the common-mode noise introduced by electrode referencing, the recordings were re-referenced by subtracting the mean of all commonly-referenced electrodes, a method called common average referencing (CAR) [27].

$$S_{CAR}^{m}[t] = S^{m}[t] - \frac{1}{M}\sum_{i=1}^{M} S^{i}[t]$$

S^{n} and S_{CAR}^{n} are the raw and CAR referenced ECoG amplitudes, respectively, on the *m-th* channel out of a total of *M* recorded channels. These re-referenced recordings were also normalized per-channel by subtracting the mean and dividing by the standard deviation of the time series [28]. Moreover, spoken responses were recorded in parallel through auxiliary Neuroport channels and a Zoom H2 recorder (Samson Technologies, Hauppauge, New York). Since the ECoG and speech recordings were time-aligned, the open source Praat software [29] was used by a speech technology engineer to manually segment each speech recording into intervals labeled as *silence*, *speech* and *noise*. The *noisy* intervals were excluded and the *speech* intervals, which corresponded to periods of articulation, were used for evaluating the classification models.

2.3 TVAR model and RML estimation method

Parametric approaches assume the time series under analysis to be the output of a given linear mathematical model. As such, following the time-varying autoregressive moving average (TARMA) model, a nonstationary discrete-time signal $S_{CAR}^{m}[t]$ is modeled as follows:

$$S_{CAR}^{m}[t] = e[t] + \sum_{i=1}^{n_a} a_i[t] S_{CAR}^{m}[t-i] + \sum_{i=1}^{n_c} c_i[t] e[t-i]$$

with t designating discrete time, $e[t]$ an (unobservable) uncorrelated (white) noise sequence with zero mean and time-dependent variance $\sigma_e^2[t]$ that "generates" $S_{CAR}^{m}[t]$, and $a_i[t], c_i[t]$ the model's time-dependent AR and MA parameters, respectively. In this study, in order to estimate the TARMA model parameters at each time instant the recursive maximum likelihood is performed, imposing no structure on the signal dynamics. Here, the parameter n_c is set to zero. Let us denote the instantaneous parameter vector at time t as

$$\vartheta[t] = [\ a_1[t] \ \ldots \ a_{n_a}[t]\ ]$$

then $\vartheta^t = [\ \vartheta^{Tr}[1] \ldots \vartheta^{Tr}[t]\ ]$ stands for AR parameters up to time t, where transposition is designated by the superscript Tr. Based on the following exponentially weighted prediction error criterion,

$$\hat{\vartheta}[t] = \arg\min_{\vartheta[t]} \sum_{\tau=1}^{t} \lambda^{t-\tau} e^2[\tau, \vartheta^{\tau-1}]$$

$$e[t, \theta^{t-1}] = S_{CAR}^m[t] - \sum_{i=1}^{n_a} a_i[t-i] S_{CAR}^m[t-i]$$

the recursive estimation of $\vartheta[t]$ is accomplished via the RML method [30], with $\arg\min$ denoting the minimizing argument, and $e[t, \theta^{t-1}]$ the model's one-step-ahead prediction error (residual) made at time $t-i$ without knowing the parameter values at time t. The term $\lambda^{t-\tau}$, $\lambda \in (0,1)$ is a windowing function (known as the "forgetting factor") that assigns more weight to more recent errors. The AR parameter estimates are obtained using the RML method given as:

$$\hat{\vartheta}[t] = \hat{\vartheta}[t-1] + k \cdot \hat{e}[t \mid t-1]$$

$$\hat{e}[t \mid t-1] = S_{CAR}^m[t] - \xi^{Tr}[t] \cdot \hat{\vartheta}[t-1]$$

$$k[t] = \frac{P[t-1] \cdot \zeta[t]}{\lambda + \zeta^{Tr}[t] \cdot P[t-1] \cdot \zeta[t]}$$

$$P[t] = \frac{1}{\lambda}\left(P[t-1] - \frac{P[t-1] \cdot \zeta[t] \cdot \zeta^{Tr}[t] \cdot P[t-1]}{\lambda + \zeta^{Tr}[t] \cdot P[t-1] \cdot \zeta[t]} \right)$$

$$\zeta[t] + \hat{c}_1[t-1] \cdot \zeta[t-1] + \ldots + \hat{c}_{n_c}[t-n_c] \cdot \zeta[t-n_c] \square\ \xi[t]$$

$$\xi[t] \square\ [-S_{CAR}^m[t-1]\ \ldots\ -S_{CAR}^m[t-n_a]\ \vdots\ \hat{e}[t-1 \mid t-1]\ \ldots\ \hat{e}[t-n_c \mid t-n_c]]^{Tr}$$

$$\hat{e}[t \mid t] = S_{CAR}^m[t] - \xi^{Tr}[t] \cdot \hat{\vartheta}[t]$$

The term $S_{CAR}^m[t \mid t-1]$ indicates the one-step-ahead prediction of the signal at time t made at time t-1, and the term $\hat{e}[t \mid t-1]$ corresponds to the prediction error. For the initialization of the method it is customary to set $\hat{\vartheta}[0] = 0, P[0] = aI$ (where a stands for a large positive number and I the unity matrix), and the initial signal and a posteriori error values to zero. Finally, in order to reduce the effects of the arbitrary initial conditions, we applied the recursions on each signal in sequential phases (for instance a forward pass, a backward pass and a final forward pass). In general, the optimum order of the model can be considered as a tradeoff between maximizing the model's fitness while limiting its complexity. In this study, the model order and forgetting factor term are optimized for the recognition of task-related ECoG activity. Specifically, the model order and the forgetting factor are set to 30 and 0.999, accordingly.

3 Feature extraction and syllable decoding

After estimating the model parameters, the power spectrum for each time instant t is defined as follows:

$$PS = \left| \frac{1}{1 + \sum_{i=1}^{n_a} \hat{a}_i[t] e^{-j\omega T_s i}} \right|^2 \hat{\sigma}_e^2[t]$$

In this formulation, ω denotes frequency in rad/time unit, j the imaginary unit, and T_s the sampling period. Using the above model, we calculate spectral amplitudes between 0 and 200 Hz in 2 Hz bins. The calculated values (spectro-temporal ECoG features) in each time and frequency point result in a $T \times F$ representation, where T denotes the samples of the ECoG signal and F denotes the number of frequency bins. Afterwards, the spectro-temporal ECoG features were averaged across time and frequency (separately for each channel and trial). In the frequency domain, features were averaged over five ECoG frequency bands: delta-theta (1-7 Hz), alpha (8-12 Hz), beta (18-26 Hz), gamma (30-70 Hz) and high-gamma (80-200 Hz). Similar to our previous work [31], we used 80-200 Hz high gamma activity to track the spatiotemporal dynamics of word processing [32].

Finally, two classifiers were employed to decode the isolated spoken syllables from ECoG signals, namely the k-nearest neighbors kNN) algorithm [33], and support vector machines (SVM) using the sequential minimal optimization algorithm [34]. For the kNN classifier, the Euclidean distance was selected as the distance metric. After testing the parameter spacef, $k = 5$ was chosen empirically. Moreover, the Gaussian radial basis function (RBF) for the SVM kernel was used. Polynomial-based kernels were also considered, but their performance was considerably lower than the RBF kernel. The values of the soft margin parameter C =20.0 and the scaling factor $\gamma = 1.0$ were found to offer optimal classification performance after a grid search at all combinations of $C = \{1.0,\ 5.0,\ 10.0,\ 20.0,\ 30.0\}$ and $\gamma = \{0.1,\ 0.5,\ 1.0,\ 2.0\}$. For evaluation of the results, a 5-fold cross validation was applied by utilizing 80% of the whole data to train our models and the remaining data for the test phase.

4 Experimental Results

Figure 2 depicts the syllable classification performance with respect to the individual frequency bands and the examined classifiers. The highest performance (19.5%) was achieved using features derived from the gamma band using the kNN classifier, while the second best accuracy (18.5%) achieved using the feature vector from the beta band and the SVM as classifier. These scores were higher than the chance decoding level (8.3%), as evaluated by a bootstrap resampling test with 20,000 repetitions while being statistically significant ($p < 0.009$). At this point, the Wilcoxon signed-

rank test was applied to reject the null hypothesis that the two classifiers performed equally well on the whole collection of data sets. The null hypothesis was rejected with the Wilcoxon statistic being smaller than the critical value for a two-tailed test with a significance level of 0.05 ($p < 0.05$).

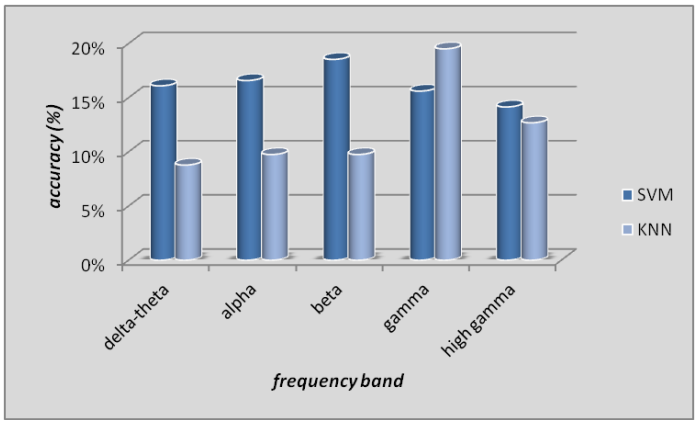

Fig. 2. Classification results (accuracy in percentage) for the subject with respect to each classifier and ECoG frequency band. Separately for each frequency band, each feature vector was used to drive our classification models.

In a further step, the Relief algorithm [35] is used to investigate the channels' significance in relationship to their location on the brain. The feature ranking scores depicted the discriminative ability of each ECoG feature for each frequency band. The feature ranking scores were averaged across each ECoG channel to reveal the most informative channels, as shown in Figure 1 (blue circles). The figure shows the five most informative ECoG sites (across frequency bands) for the syllable classification task. Our findings are consistent with the literature on the neuroanatomy of speech production. More specifically, the ECoG sites (as depicted in Figure 1) suggest that widely distributed areas, contain discriminative information for overtly articulated syllables [36]. Furthermore, many of the other most highly ranked electrodes (as calculated by the Relief algorithm) are located in cortical areas also relevant to speech processing, such as superior temporal gyrus (STG) and right ventral sensorimotor cortex (vSMC) [37-39].

5 Conclusion

Previous ECoG studies on speech decoding at levels including formants, phonemes, words, and even sentences. Here, the problem of syllable decoding from ECoG neural activity is studied. A recursive parameter estimation approach is proposed to calculate the time-varying autoregressive model coefficients and consequently to extract the power spectrum from the raw ECoG signal from different frequency bands. The highest accuracy achieved for classifying twelve overtly articulated syllables is 19.5%. In

conclusion, to our best knowledge, this study has validated for the first time the feasibility of syllable classification from ECoG signals. Thus, no direct comparison with other approaches is feasible. Additional experiments on channel reduction, based on the information acquired from causal interactions between cortical areas, may also prove useful in the future.

Acknowledgement

This research has been co-financed by the European Union (European Social Fund, ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF); Research Funding Program: THALES. Investing in knowledge society through the European Social Fund.

References

[1] O. Gosseries, A. Vanhaudenhuyse, M.-A. Bruno, A. Demertzi, C. Schnakers, M. M. Boly, *et al.*, "Disorders of consciousness: coma, vegetative and minimally conscious states," in *States of Consciousness*, ed: Springer, 2011, pp. 29-55.

[2] N. Birbaumer, N. Ghanayim, T. Hinterberger, I. Iversen, B. Kotchoubey, A. Kübler, *et al.*, "A spelling device for the paralysed," *Nature*, vol. 398, pp. 297-298, 1999.

[3] N. Birbaumer, T. Hinterberger, A. Kubler, and N. Neumann, "The thought-translation device (TTD): neurobehavioral mechanisms and clinical outcome," *Neural Systems and Rehabilitation Engineering, IEEE Transactions on,* vol. 11, pp. 120-123, 2003.

[4] N. Birbaumer, A. Kübler, N. Ghanayim, T. Hinterberger, J. Perelmouter, J. Kaiser, *et al.*, "IV. FUTURE WORK," *IEEE Transactions on rehabilitation Engineering,* vol. 8, p. 191, 2000.

[5] E. Donchin, K. M. Spencer, and R. Wijesinghe, "The mental prosthesis: assessing the speed of a P300-based brain-computer interface," *Rehabilitation Engineering, IEEE Transactions on,* vol. 8, pp. 174-179, 2000.

[6] D. J. Krusienski, E. W. Sellers, D. J. McFarland, T. M. Vaughan, and J. R. Wolpaw, "Toward enhanced P300 speller performance," *Journal of neuroscience methods,* vol. 167, pp. 15-21, 2008.

[7] F. Nijboer, E. Sellers, J. Mellinger, M. Jordan, T. Matuz, A. Furdea, *et al.*, "A P300-based brain–computer interface for people with amyotrophic lateral sclerosis," *Clinical neurophysiology,* vol. 119, pp. 1909-1916, 2008.

[8] E. W. Sellers and E. Donchin, "A P300-based brain–computer interface: initial tests by ALS patients," *Clinical neurophysiology,* vol. 117, pp. 538-548, 2006.

[9] M. Cheng, X. Gao, S. Gao, and D. Xu, "Design and implementation of a brain-computer interface with high transfer rates," *Biomedical Engineering, IEEE Transactions on,* vol. 49, pp. 1181-1186, 2002.

[10] O. Friman, T. Luth, I. Volosyak, and A. Graser, "Spelling with steady-state visual evoked potentials," in *Neural Engineering, 2007. CNE'07. 3rd International IEEE/EMBS Conference on,* 2007, pp. 354-357.

[11] T. M. Vaughan, D. J. McFarland, G. Schalk, W. A. Sarnacki, D. J. Krusienski, E. W. Sellers, *et al.*, "The wadsworth BCI research and development program: at home with BCI," *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, vol. 14, pp. 229-233, 2006.

[12] 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, p. 046028, 2011.

[13] C. Neuper, G. R. Müller-Putz, R. Scherer, and G. Pfurtscheller, "Motor imagery and EEG-based control of spelling devices and neuroprostheses," *Progress in brain research*, vol. 159, pp. 393-409, 2006.

[14] R. Scherer, G. Muller, C. Neuper, B. Graimann, and G. Pfurtscheller, "An asynchronously controlled EEG-based virtual keyboard: improvement of the spelling rate," *Biomedical Engineering, IEEE Transactions on*, vol. 51, pp. 979-984, 2004.

[15] F. H. Guenther, J. S. Brumberg, E. J. Wright, A. Nieto-Castanon, J. A. Tourville, M. Panko, *et al.*, "A wireless brain-machine interface for real-time speech synthesis," *PloS one*, vol. 4, p. e8218, 2009.

[16] 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, pp. 1334-1339, 2009.

[17] 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, p. 056007, 2010.

[18] D. Zhang, E. Gong, W. Wu, J. Lin, W. Zhou, and B. Hong, "Spoken sentences decoding based on intracranial high gamma response using dynamic time warping," in *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, 2012, pp. 3292-3295.

[19] S. Deng, R. Srinivasan, T. Lappas, and M. D'Zmura, "EEG classification of imagined syllable rhythm using Hilbert spectrum methods," *Journal of neural engineering*, vol. 7, p. 046006, 2010.

[20] O. Bai, M. Nakamura, A. Ikeda, and H. Shibasaki, "Nonlinear Markov process amplitude EEG model for nonlinear coupling interaction of spontaneous EEG," *Biomedical Engineering, IEEE Transactions on*, vol. 47, pp. 1141-1146, 2000.

[21] C.-M. Ting, S.-H. Salleh, Z. Zainuddin, and A. Bahar, "Spectral estimation of nonstationary EEG using particle filtering with application to event-related desynchronization (ERD)," *Biomedical Engineering, IEEE Transactions on*, vol. 58, pp. 321-331, 2011.

[22] A. Poulimenos and S. Fassois, "Parametric time-domain methods for non-stationary random vibration modelling and analysis—a critical survey and comparison," *Mechanical Systems and Signal Processing*, vol. 20, pp. 763-816, 2006.

[23] A. Schlögl, *The electroencephalogram and the adaptive autoregressive model: theory and applications*: Shaker Germany, 2000.

[24] M. E. Khan and D. N. Dutt, "An expectation-maximization algorithm based Kalman smoother approach for event-related desynchronization (ERD) estimation from EEG," *Biomedical Engineering, IEEE Transactions on*, vol. 54, pp. 1191-1198, 2007.

[25] M. Niedzwiecki, *Identification of time-varying processes*: Wiley New York, 2000.

[26] J. S. Duncan, X. Papademetris, J. Yang, M. Jackowski, X. Zeng, and L. H. Staib, "Geometric strategies for neuroanatomic analysis from MRI," *Neuroimage*, vol. 23, pp. S34-S45, 2004.

[27] D. Goldman, "The clinical use of the “average” reference electrode in monopolar recording," *Electroencephalography and clinical neurophysiology*, vol. 2, pp. 209-212, 1950.

[28] T. Pistohl, A. Schulze-Bonhage, A. Aertsen, C. Mehring, and T. Ball, "Decoding natural grasp types from human ECoG," *Neuroimage*, vol. 59, pp. 248-260, 2012.

[29] P. Boersma and D. Weenink, "Praat, a system for doing phonetics by computer," 2001.

[30] L. Ljung, "System identification: theory for the user," *PTR Prentice Hall Information and System Sciences Series*, vol. 198, 1987.

[31] V. G. Kanas, I. Mporas, H. L. Benz, K. N. Sgarbas, A. Bezerianos, and N. E. Crone, "Joint spatial-spectral feature space clustering for speech activity detection from ECoG signals," *Biomedical Engineering, IEEE Transactions on*, vol. 61, pp. 1241-1250, 2014.

[32] R. T. Canolty, M. Soltani, S. S. Dalal, E. Edwards, N. F. Dronkers, S. S. Nagarajan, *et al.*, "Spatiotemporal dynamics of word processing in the human brain," *Frontiers in neuroscience*, vol. 1, p. 185, 2007.

[33] D. W. Aha, D. Kibler, and M. K. Albert, "Instance-based learning algorithms," *Machine learning*, vol. 6, pp. 37-66, 1991.

[34] J. Platt, "Fast training of support vector machines using sequential minimal optimization," *Advances in kernel methods—support vector learning*, vol. 3, 1999.

[35] I. Kononenko, "Estimating attributes: analysis and extensions of RELIEF," in *Machine Learning: ECML-94*, 1994, pp. 171-182.

[36] M. Vigneau, V. Beaucousin, P.-Y. Herve, H. Duffau, F. Crivello, O. Houde, *et al.*, "Meta-analyzing left hemisphere language areas: phonology, semantics, and sentence processing," *Neuroimage*, vol. 30, pp. 1414-1432, 2006.

[37] P. Indefrey, "The spatial and temporal signatures of word production components: a critical update," *Frontiers in psychology*, vol. 2, 2011.

[38] P. McGuire, D. Silbersweig, and C. Frith, "Functional neuroanatomy of verbal self-monitoring," *Brain*, vol. 119, pp. 907-917, 1996.

[39] S. S. Shergill, M. J. Brammer, R. Fukuda, E. Bullmore, E. Amaro, R. M. Murray, *et al.*, "Modulation of activity in temporal cortex during generation of inner speech," *Human brain mapping*, vol. 16, pp. 219-227, 2002.