

Unsupervised Robust Detection of Behavioral Correlates in ECoG

Carlos A. Loza, Jose C. Principe

Abstract—Electrocorticogram (ECoG) based Brain-Computer Interfaces (BCI) provide finer spatial resolution and improved signal-to-noise ratio than its noninvasive counterpart, Electroencephalogram (EEG). This remarkable feature allows for processing in higher spectral bands in order to elucidate more spatially localized encoding mechanisms. We propose an automatic, fully data-driven method to extract relevant neuromodulation events from single-channel, single-trial traces. In particular, our scheme involves two alternating optimizations that resemble k-means; moreover, correntropy is utilized to provide robust estimation and protection against outliers. In this way, we find distinct behavioral correlates in the low-gamma band (76 - 100 Hz) that encode finger flexion movements in a cued task. The results show that correntropy should be used when working with neuronal oscillations due to the high probability of outliers.

Index Terms—BCI, Correntropy, ECoG, Neuromodulation, Transient Model

I. INTRODUCTION

Neuronal oscillations represent the synchronous average activity of neural populations in the brain. In general, they are the result of the spatiotemporal interplay between excitatory and inhibitory postsynaptic potentials [1] (EPSP and IPSP). Hence, unlike action potentials, neuronal oscillations encode information not only via their timings, but also by means of amplitude, phase, and frequency features [2].

One of the paramount properties of neuronal oscillations is its correlation to inputs from the outside world, internal physiological processes, and behavior in general. For instance, ECoG has been utilized in brain mapping, BCIs, and epilepsy studies to name a few [3]–[6]. In the current paper we focus on ECoG-based BCI experiments while, at the same time, strive to expand our previous studies on a transient model for neuronal oscillations [7], [8]. In particular, our approach does not assume stationarity nor ergodicity; rather, it exploits neurophysiological concepts, such as transient synchronization of neuronal assemblies [9], and models a single-channel trace as a noisy spectral superposition of marked point processes that preserve the temporal resolution up to the sampling interval scale. In addition, correntropy is incorporated into the framework in order to provide robust estimation and protection against outliers. In this way, the resulting marked point processes and temporal templates constitute a novel interpretation of ECoG and neuronal oscillations in general.

The rest of the paper is organized as follows: Section 2 briefly reintroduces the transient model for ECoG, Section 3 details the robust estimation methods while Section 4 describes the experiment and results. Lastly, Section 5 concludes the paper and discusses further research.

II. TRANSIENT MODEL FOR ECoG

Following the neurophysiological and clinical interpretation of EEG and neuronal oscillations in general [9], [10], we proposed an anthropomimetic framework that models a single-channel trace as the noisy spectral superposition of marked point processes (MPP) activating patterns over time. In particular, the point processes are segregated according to the well-known clinical EEG bands [2]; in this way, the temporal patterns can be regarded as finite impulse response (FIR) filters with similar central frequencies. The main advantage of the model is that it exploits the reoccurring, transient phasic events inherent to neuronal oscillations.

Fig. 1 depicts the block diagram of the transient model alongside its 3 main building blocks: MPP features, i.e. timing, amplitude and index, the set of FIR filters for each spectral band, and the added noise properties. In equations, the output of the system, $x(t)$, is a single-channel, single-trial ECoG-like signal that displays the ever-changing temporal structure of macroscopic brain activity.

$$x(t) = n(t) + \hat{x}(t) = n(t) + \sum_{i=1}^F y_i(t) \quad (1)$$

$$y_i(t) = \sum_{j=1}^{n_i} \int_{-\infty}^{\infty} \alpha_{i,j} \delta(t - \tau_{i,j}) h_{i,\omega_j} du \quad (2)$$

where F is the number of filters banks, i.e. ECoG rhythms to be analyzed, $n(t)$ is the additive colored noise, α and τ are the phasic event's amplitudes and timings, respectively. $H_i = \{h_{i,\omega_j}\}$ is the i -th filter bank, also known as dictionary. Borrowing terms from sparse modeling, the unit-norm FIR filters are also referred to as atoms. In short, a sample from a MPP will modulate a single temporal template from a filter bank with amplitude α at time τ by means of indexing. Moreover, $H_i \in \mathbb{R}^{M \times K}$, where M represents the duration (in samples) of putative phasic events, and K , the dictionary size. The final goal of an analysis scheme is to obtain the MPP features and FIR filters of a relevant spectral band for a particular behavioral task without imposing the assumption of stationarity nor limiting the putative encoding mechanisms to amplitude modulation alone.

Carlos A. Loza and Jose C. Principe are with the Computational NeuroEngineering Laboratory (CNEL), University of Florida, Gainesville, FL (email: cloza@ufl.edu, principe@cnel.ufl.edu).

This work was supported by Michael J. Fox Grant 9558

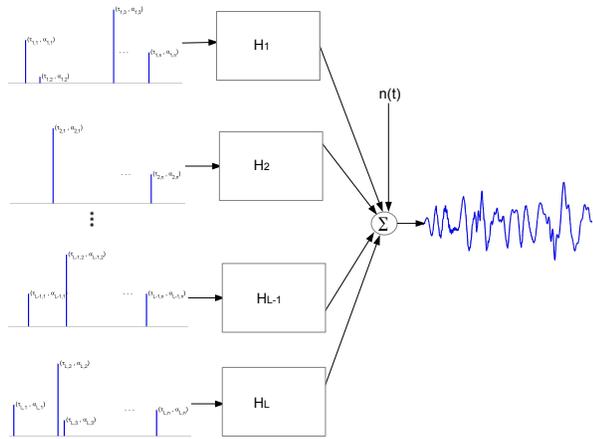


Fig. 1: Transient Model for ECoG. A single-channel, single-trial ECoG trace is modeled as the noisy spectral addition of reoccurring, transient patterns over time.

III. ROBUST DETECTION OF NEUROMODULATIONS

Estimating the model parameters is an example of a blind source separation problem; particularly, isolating one rhythm via bandpass filtering reduces the estimation task to learning one MPP and one dictionary given a single-channel ECoG recording. If no constraints are given, the final solution will not be unique; thus, it is imperative to impose conditions that not only make the problem more accessible, but also that are neurophysiologically sound. The first constraint is the presence of non-overlapping phasic events; this is possible due to the single-scale nature of the bandpassed traces and will assure that a relevant neuromodulation is represented by a scaled version of a single dictionary atom. In contrast, other Time-Frequency decompositions, such as Matching Pursuit (MP) [11], aim to overrepresent phasic events in order to minimize the L_2 norm of the reconstruction error. The second constraint attempts to obtain a decomposition as unique as possible. Specifically, the final estimated dictionary should have the lowest mutual coherence among all the possible filter bank candidates. Mutual coherence is defined as [12]:

$$\mu(H) = \max_{i \neq j} |h_i^T h_j| \quad (3)$$

where h_i is the i -th atom or FIR filter from dictionary H . Two extreme cases are orthogonal bases ($\mu(H) = 0$) and dictionaries with linearly dependent elements ($\mu(H) = 1$). The first scenario would yield a unique solution, while on the other hand, the second decomposition would be ambiguous. The final version of the estimation algorithm involves two clear stages: phasic event assignment and dictionary update. In particular, they resemble the two alternating steps in the clustering algorithm k-means [13] applied to time series.

A. Phasic Event Assignment

Given a dictionary H , we extract relevant phasic events that are maximally correlated to such vector space. This is

efficiently computed via 2-dimensional, Fast Fourier Transform (FFT) based convolutions between each single-channel, single-trial ECoG trace and the 2-dimensional filter bank. Due to the non-overlapping constraint, this convolution is computed only once, unlike MP decompositions, where multiple runs are needed and the computational cost increases considerably. Next, each putative neuromodulation is sequentially extracted in a decreasing order based on the absolute value of the decomposition amplitudes, α 's.

During the extraction stage, each temporal snippet contributes with an estimated decomposition timing, τ , amplitude, α , and index, ω . Specifically, the timing is estimated with a precision up to the sampling interval, while ω performs the assignment to the closest dictionary atom. As previously mentioned, this stage evokes the k-means encoding process; however for our case, the membership vector for each putative phasic event counts with a single graded non-zero entry. The specific details regarding this side of the estimation scheme are omitted due to space limitations.

B. Dictionary Update

Given a set of labeled M -dimensional samples, i.e. extracted temporal snippets, we update each element in the filter bank, H . In k-means, this task is simply the mean or median value of the assigned samples for each cluster. Here instead, we are working with oscillating patterns that would most likely yield a close-to-zero value as average cluster centroid, e.g. patterns with opposite phase would cancel each other. For this reason, it is imperative to update each atom according to a sign-invariant low-rank transformation; in particular, Singular Value Decomposition (SVD) provides a principled mechanism for this problem. In addition, robustness against outliers is incorporated by exploiting the properties of the correntropy dependence measure.

Correntropy was proposed in [14] as an alternative beyond second-order statistics and Gaussian conditions. For two random variables X and Y , correntropy is defined as:

$$V_\sigma(X, Y) = \mathbf{E}[G_\sigma(X - Y)] \quad (4)$$

where $G_\sigma(X - Y)$ represents the Gaussian kernel with parameter σ , also known as kernel width. In particular, this parameter will control the metric correntropy will mimic; for instance, a very large value will resemble ℓ_2 interactions, while a very well-tuned low value will provide robustness via ℓ_0 penalty-like behavior. In particular as shown in [15], [16], correntropy can be used as cost function in the SVD problem when outliers are present and might be potentially detrimental. For instance, Fig. 2 illustrates these effects on the estimated principal components for a toy example. The figure shows how a very low density of outliers can bias the first principal eigenvector, i.e. all of the input samples contribute equally to the SVD computation, while on the other hand, a correntropy-based optimization yields the original outlier-free principal component.

The robust estimation method fully based on the Half-Quadratic (HQ) technique [17] is presented in Algorithm 1.

Algorithm 1 Correntropy-based Robust SVD

Input: $X \in \mathbb{R}^{M \times d}$, $\epsilon, \mu \in \mathbb{R}^M$, $U^1 \in \mathbb{R}^M$
Output: $\mu, U \in \mathbb{R}^M$
 $J \leftarrow 1$
while convergence == FALSE **do**
 $r \leftarrow \|(x_i - \mu) - U^J(U^J)^T(x_i - \mu)\|^2 \quad i = 1, \dots, d$
 $\sigma \leftarrow 1.06 \times \min \{\text{std}(r), \text{IQR}(r)/1.34\} \times d^{-1/5}$
 $x_i^\mu \leftarrow x_i - \mu \quad i = 1, \dots, d$
 $p_i \leftarrow -G_\sigma(\sqrt{x_i^{\mu T} x_i^\mu - x_i^{\mu T} (U^J)(U^J)^T x_i^\mu}) \quad i = 1, \dots, d$
 $\mu \leftarrow (\sum_{i=1}^d p_i x_i) / (\sum_{i=1}^d p_i)$
 $X_c \leftarrow [x_1 - \mu, x_2 - \mu, \dots, x_d - \mu]$
 $P \leftarrow \text{diag}(-p)$
 $U^{J+1} \leftarrow \text{SVD}(X_c P X_c^T)$
if $\|U^J - U^{J+1}\| < \epsilon$ **then**

convergence = TRUE

 $U \leftarrow U^{J+1}$
else

convergence = FALSE

 $J \leftarrow J + 1$
end if
end while

Specifically, d represents the number of temporal snippets of duration M associated to a particular cluster centroid, ϵ is a stopping threshold, μ is the estimated mean vector, and U is the final first principal component. As an added feature, there are no extra free parameters associated to this method; singularly, the kernel width, σ , is updated and estimated sequentially alongside the first principal component by exploiting Silverman's rule [18]. Lastly, the initial values for μ and U , i.e. U^1 , can be easily provided via regular average and SVD operations, respectively. It is worth mentioning that Algorithm 1 has to be utilized for each dictionary atom, i.e. each updated atom is the first principal component of the corresponding labeled M -dimensional samples. Readers can refer to [16] for further details regarding the algorithm.

In summary, the estimation process involves two alternat-

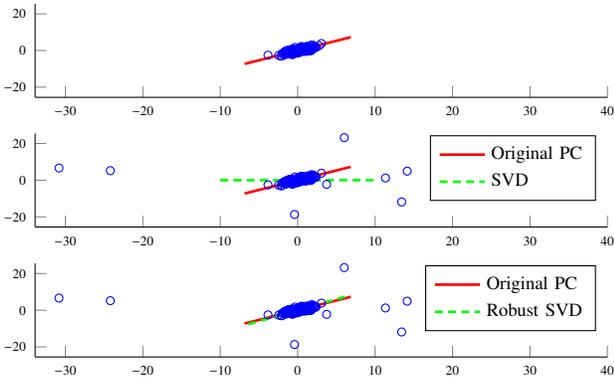


Fig. 2: Correntropy-based Robust SVD example. From top to bottom: Original Gaussian pdf (First Principal Component in red). 10 outliers (5%) are added and SVD is computed. Same distribution with Correntropy-based Robust SVD eigenvector.

ing stages: phasic event assignment and dictionary update. The first one is performed via fast FFT-based convolutions, while the second one exploits correntropy to compute principal components as updated dictionary atoms. Both stages are ran until a particular stopping criterion is met, e.g. minimum variance of estimated atoms. Finally, the optimal dictionary is chosen according to a minimal coherence criterion.

IV. BEHAVIORAL CORRELATES IN ECoG

An ECoG-based BCI was utilized to show the plausibility of our methods and algorithms. The dataset can be found in the BCI competition website [19], [20]; namely, 64 channels (for subject 3) are recorded subdurally in epileptic patients receiving ECoG monitoring for the localization of seizure foci. The task consists on cued finger movements (3-5 each time) with 2-second rest intervals. In order to find behavioral correlates, the finger flexion traces are recorded using a data glove, digitized and provided alongside the ECoG data.

The ECoG traces were downsampled to 500 Hz and bandpassed around the low-gamma rhythm (76 - 100 Hz). However when dealing with data-driven frameworks and EEG oscillating activity, it is imperative to select the band-pass filter parameters in a principled manner, i.e. a compromise between stopband attenuation and ringing artifacts; [21] stated that a quality factor, (Q), close to one 1 is ideal for such trade-off. Thus, we utilized a 6-th order Butterworth filter with $Q = 1$ and cut-off frequencies 44 and 132 Hz.

Furthermore, after visual inspection of the ECoG recordings, we set the duration of putative phasic events to 0.1 seconds; next, the dictionary size was chosen from a range of values, and, lastly, the alternating optimizations were ran for a maximum of 20 iterations each with 5 different initial conditions. Fig. 3 illustrates the learned FIR filters after convergence for the best case scenario and $K = 12$. It is evident that the data displays phasic events with diverse modulation patterns and frequencies. The most striking observation is the presence of outliers in the distribution of extracted temporal snippets (red crosses on boxplot insets); hence, our choice of correntropy as SVD cost function is justified and appropriate.

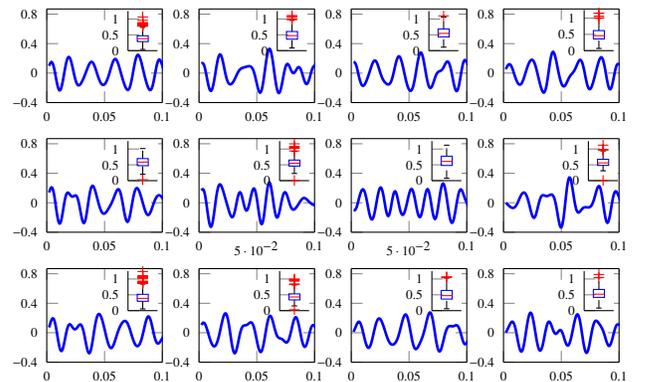


Fig. 3: Learned dictionary atoms for channel 9 and $K = 12$. Insets show the boxplot of the normalized Euclidean distances from each labeled snippet to the particular cluster centroid.

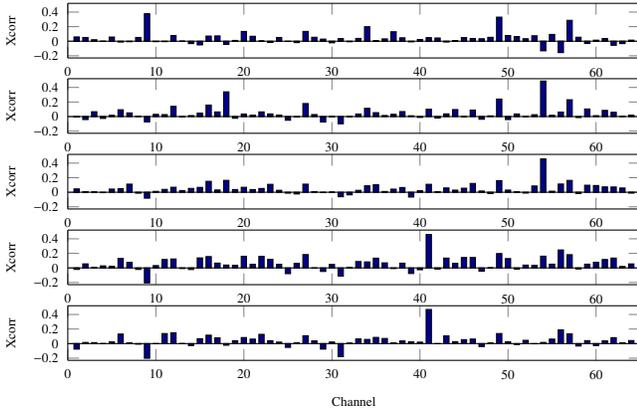


Fig. 4: Encoding of finger flexion is spatially sparse and well-localized. Average maximum cross-correlation between each finger flexion trace and a smoothed version of the low-gamma MPP from each channel. From top to bottom: Finger 1,2,3,4,5.

Next, the MPP features are utilized to find correlates to the finger flexion traces. A Gaussian kernel is placed on top of each MPP sample in order to create a continuous signal; then, we compute the cross-correlation between each finger flexion trace and the smoothed MPP from each sensor. Fig. 4 depicts the final result as average maximum cross-correlations. Particularly, each finger seems to have few relevant areas that are strongly correlated to its movement, while, most of the remaining sensors do not contribute to this motor output. As mentioned in [20], this high-frequency activity serves as a proxy of the broadband average firing rate of the population in proximity to the electrode that is correlated to the task. Similar results are observed for Subjects 1 and 2, but they are omitted due to space limitations.

Lastly, we chose various dictionary sizes ranging from $K = 2$ to $K = 100$. The main findings are summarized in Table I as the average maximum cross-correlation of the single channel mostly associated to each finger flexion. In general, the values increase steadily and reach their maximum at $K = 75$; this trend is the result of a richer, more diverse dictionary, i.e. more distinct, higher-frequency, discriminant patches are extracted. If modeling is the final goal, then the selection of the number of clusters becomes more crucial, e.g. underestimation would hinder the discovery of discriminant phasic events in terms of modulatory activity, and overestimation would derive in biased results due to overfitting.

TABLE I: Average maximum cross-correlation of sensor mostly associated to each finger flexion for several dictionary sizes

K	2	5	12	25	50	75	100
Finger 1	0.29	0.30	0.31	0.32	0.35	0.38	0.37
Finger 2	0.44	0.45	0.45	0.47	0.48	0.49	0.49
Finger 3	0.41	0.42	0.45	0.45	0.46	0.46	0.46
Finger 4	0.40	0.42	0.43	0.43	0.46	0.46	0.44
Finger 5	0.40	0.42	0.43	0.44	0.45	0.47	0.46

V. CONCLUSIONS AND FURTHER WORK

We have used a robust, data-driven, unsupervised, alternating optimization scheme that learns ECoG phasic events

positively correlated to finger flexion tasks. In the future, we intend to analyze additional spectral bands, study principled techniques to choose the dictionary size parameter, utilize the MPP samples to assess spatial interactions and address the noise term in the transient model.

REFERENCES

- [1] E. Niedermeyer and F. L. da Silva, *Electroencephalography: basic principles, clinical applications, and related fields*. Lippincott Williams & Wilkins, 2005.
- [2] G. Buzsáki, *Rhythms of the Brain*. Oxford University Press, 2006.
- [3] N. E. Crone, A. Sinai, and A. Korzeniewska, “High-frequency gamma oscillations and human brain mapping with electrocorticography,” *Progress in brain research*, vol. 159, pp. 275–295, 2006.
- [4] E. C. Leuthardt, G. Schalk, J. Wolpaw, J. Ojemann, and D. Moran, “A brain-computer interface using electrocorticographic signals in humans,” *Journal of neural engineering*, vol. 1, no. 2, p. 63, 2004.
- [5] J. C. Sanchez, A. Gunduz, P. R. Carney, and J. C. Principe, “Extraction and localization of mesoscopic motor control signals for human ECoG neuroprosthetics,” *Journal of neuroscience methods*, vol. 167, no. 1, pp. 63–81, 2008.
- [6] N. Nakasatp, M. F. Levesque, D. S. Barth, C. Baumgartner, R. L. Rogers, and W. W. Sutherling, “Comparisons of MEG, EEG, and ECoG source localization in neocortical partial epilepsy in humans,” *Electroencephalography and clinical neurophysiology*, vol. 91, no. 3, pp. 171–178, 1994.
- [7] C. A. Loza and J. C. Principe, “Transient model of EEG using Gini index-based matching pursuit,” in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2016, pp. 724–728.
- [8] C. A. Loza and J. C. Principe, “Estimation and modeling of EEG amplitude-temporal characteristics using a marked point process approach,” in *Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the*. IEEE, 2016, pp. 3720–3723.
- [9] G. Buzsáki and A. Draguhn, “Neuronal oscillations in cortical networks,” *Science*, vol. 304, no. 5679, pp. 1926–1929, 2004.
- [10] W. O. Tatum IV, *Handbook of EEG interpretation*. Demos Medical Publishing, 2014.
- [11] P. Durka and K. Blinowska, “Analysis of EEG transients by means of matching pursuit,” *Annals of biomedical engineering*, vol. 23, no. 5, pp. 608–611, 1995.
- [12] D. L. Donoho and X. Huo, “Uncertainty principles and ideal atomic decomposition,” *IEEE Transactions on Information Theory*, vol. 47, no. 7, pp. 2845–2862, 2001.
- [13] A. Gersho and R. M. Gray, *Vector quantization and signal compression*. Springer Science & Business Media, 2012, vol. 159.
- [14] W. Liu, P. P. Pokharel, and J. C. Principe, “Correntropy: properties and applications in non-gaussian signal processing,” *IEEE Transactions on Signal Processing*, vol. 55, no. 11, pp. 5286–5298, 2007.
- [15] R. He, B. Hu, W. Zheng, and X. Kong, “Robust principal component analysis based on maximum correntropy criterion,” *IEEE Transactions on Image Processing*, vol. 20, no. 6, pp. 1485–1494, 2011.
- [16] C. A. Loza and J. C. Principe, “A robust maximum correntropy criterion for dictionary learning,” in *Machine Learning for Signal Processing (MLSP), 2016 IEEE 26th International Workshop on*. IEEE, 2016, pp. 1–6.
- [17] R. T. Rockafellar, “Convex analysis (princeton mathematical series),” *Princeton University Press*, vol. 46, p. 49, 1970.
- [18] B. W. Silverman, *Density estimation for statistics and data analysis*. CRC press, 1986, vol. 26.
- [19] G. Schalk, J. Kubanek, K. Miller, N. Anderson, E. Leuthardt, J. Ojemann, D. Limbrick, D. Moran, L. Gerhardt, and J. Wolpaw, “Decoding two-dimensional movement trajectories using electrocorticographic signals in humans,” *Journal of neural engineering*, vol. 4, no. 3, p. 264, 2007.
- [20] M. Tangermann, K.-R. Müller, A. Aertsen, N. Birbaumer, C. Braun, C. Brunner, R. Leeb, C. Mehring, K. J. Miller, G. Mueller-Putz *et al.*, “Review of the bci competition iv,” *Frontiers in neuroscience*, vol. 6, p. 55, 2012.
- [21] B. H. Fadlallah, S. Seth, A. Keil, and J. C. Principe, “Robust EEG preprocessing for dependence-based condition discrimination,” in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2011, pp. 1407–1410.