Published in final edited form as:

Conf Proc IEEE Eng Med Biol Soc. 2019 July ; 2019: 91–94. doi:10.1109/EMBC.2019.8856494.

Data-driven separation and estimation of atrial dynamics in very high-dimensional electrocardiograms from epilepsy patients

Catherine Stamoulis¹ [Member IEEE], Jack Connolly², Frank H. Duffy³

¹Faculty of Medicine, Harvard Medical School, Department of Medicine, Boston Children's Hospital, Boston, MA 02115, USA

²Department of Neurology, Boston Children's Hospital, Boston, MA 02115, USA

³Faculty of Neurology, Harvard Medical School, Boston Children's Hospital, Boston, MA 02115, USA

Abstract

Across biomedical areas, there is a significant unmet need for multimodal biomarkers that can improve prediction of abnormal events such as seizures, heart and asthma attacks or stroke. These markers may be multimodal and may include electrophysiological measures estimated from noninvasive, routinely collected clinical data, such as electroencephalograms (EEG) and electrocardiograms (ECG). In epilepsy, seizure detection and prediction from noninvasive data remains a difficult problem in need of novel approaches and markers. The inherent noise in high-dimensional EEG signals and artifact contamination often severely impacts the sensitivity and specificity of otherwise promising biomarkers. Long-term epilepsy clinical studies typically collect continuous ECG from which additional features may be estimated and combined with EEG measures to improve sensitivity to ictogenesis and seizure specificity. Prior work has focused on ventricular activity and features of the QRS complex, but atrial activity may also be modulated by seizure evolution. Given the high dimension of the ECG collected continuously over several days, an entirely data-driven approach is proposed, based on which ECG signals may be separated into ventricular and atrial contributions and studied separately. The relationship of atrial dynamics to seizure occurrence is assessed in a small number of pediatric epilepsy patients with high-dimensional ECG.

I. INTRODUCTION

A broad range of human studies routinely collect electrocardiograms (ECG), often over long periods of time (days), either as the primary or secondary data modality of interest. These include studies that specifically aim to evaluate cardiovascular health, for which the ECG is the primary diagnostic tool, polysomnography studies where the ECG is one of several physiological signals, or other studies, e.g., in epilepsy, where the primary modality is the electroencephalogram (EEG) but ECGs are simultaneously recorded and available for analysis. Multimodal physiological studies are of particular interest across biomedical areas. Given the complexity and heterogeneity of many diseases and disorders, such as

cardiovascular disease, Autism Spectrum Disorders (ASD), neuropsychiatric disorders and epilepsy, there is a significant need to identify and combine robust biomarkers from multiple data types that can lead to improved diagnoses, risk assessment, outcome and event prediction (e.g., heart attacks, asthma attacks, seizures and surgical, cognitive and other outcomes) and the development of promising next-generation therapies [1], [2], [3], [4], [5]. In epilepsy, a relatively common neurological disorder that affects ~1% of the US population, including ~500,000 children, integration of multimodal physiological measures may significantly improved seizure predictability [6], [7].

Almost 200,000 children in the US suffer from medically refractory epilepsy and have few treatment options. Surgical resection of the epileptogenic brain tissue is available to a select small group of patients with unequivocally localizable focal seizures. More recently, neurostimulation and targeted drug delivery have emerged as promising treatments for a potentially larger number of patients with intractable seizures. However, unless noninvasive seizure detection can be improved, the potential clinical utility of these treatments will likely remain limited. Thus, improved and potentially heterogeneous biomarkers of ictogenesis are needed. Almost all noninvasive long-term monitoring studies in epilepsy collect both scalp EEG and ECG. Prior work has reported frequent heart rate changes prior to or around ictal onset and may, in some cases, precede clinical or electrographic changes [8], [9]. However, the majority of prior studies have focused on the QRS complex and its features and ventricular rhythms. Atrial signals remain relatively unexplored, except in sudden unexpected death in epilepsy (SUDEP) [10].

Given the complexity of high-dimensional physiological recordings, including true biological contributions, noise and artifacts, an efficient data-driven algorithm is proposed, to separate the low-amplitude ventricular from atrial activity in continuous ECG from epilepsy patients and assess their dynamic association with seizure occurrence. Prior studies have proposed various algorithms for decomposing the two signals, including Independent Component Analysis approaches [11], [12] and template-based methods [13], [14]. With one- or two-lead ECG, ICA and its variants may not accurately separate the two signals. Template methods are more promising but several use either model-based on averaged QRS signals. In long-term recordings, there may be some variability in the QRS complex characteristics, both in amplitude and duration. Thus, a data-driven and adaptive approach may be more appropriate. The proposed method uses data-derived templates of the QRS complex, estimated through clustering of multiple estimated waveforms, to locally adapt them to the QRS morphology and subtract them from the raw ECG and obtain the atrial signals. It is applied to a small dataset of 6 pediatric epilepsy patients with continuous ECG (2 leads) spanning at least 24 h. The atrial dynamic complexity is then quantified by each signal's fractal dimension [15], estimated sequentially in time. We show that seizures typically occur in periods of low atrial signal complexity.

II. METHODS

A. Electrophysiological Data

ECG data were recorded at the Comprehensive Epilepsy Center at Boston Childrens Hospital as part of inpatient, noninvasive electrophysiological studies. Signals from two

ECG leads (left and right) were obtained simultaneously with 24-channel clinical EEG, all sampled at 1024 samples/s. Data from 6 consecutive pediatric patients was used to test the proposed algorithm. Patient characteristics are summarized in Table I. Recording length was 38.3 - 96.9 h. Only ECG data were analyzed in this study.

a) Signal Pre-processing: To suppress power-line noise at 60 Hz and its harmonics, ECG signals were filtered using a third-order elliptical filter with 1 Hz bandwidth, 20 dB attenuation in the stopband and 0.5 dB ripple in the passband. The filter was applied in both directions to eliminate potential phase distortions associated with its non-zero phase. Signals were also low-pass filtered with the same type of IIR filter and a cutoff at 100 Hz.

III. ALGORITHM

The proposed algorithm was developed to locally denoise the ECG signals, separate ventricular (the QRS complex) from atrial activity and estimate a robust template for the former. It is summarized in Figure 2 and involves a multistep iterative approach. Global signal statistics may fluctuate significantly in long ECG recordings, which may be contaminated by various high-amplitude artifacts. An example is shown in Figure 1. Various steps of the algorithm e.g., peak finding and related adjustments, depend on these statistics. Given the overall data-driven approach in this work, a window of analysis was estimated from the data. A sliding 30-s window was used to estimate the median absolute ECG amplitude. A root-mean-squared (RMS) envelope was fitted to the resulting median signals and time points at which signal amplitude was above the *global median + inter-quartile range (IQR)* were estimated. A changepoint detection approach was used for this purpose. The minimum distance between change points was used as the analysis unit in the algorithm. In this dataset, it was ~30-50 min.

Within each analysis window, ECG signal peaks were detected, assuming that both RR peaks and artifacts represent extreme amplitude outliers, i.e., at or above the threshold of *median + 3×IQR* [16]. After the first pass of peak estimation (which assumed an arbitrary inter-peak distance), the thresholds for amplitude and peak-to-peak distance were adjusted based on the statistics of these parameters and peak estimation was repeated. Each QRS complex was then extracted locally based on peak locations assuming a threshold for its duration of 0.12 s (typical duration of this complex).

In the classification stage of the algorithm, QRS signals were assembled into a matrix. The *gap statistic* [17] was used to estimate the optimum number of clusters in the data. The within-cluster dispersion (a variance measure) W_k of cluster k is defined as the normalized sum of intra-cluster distances D_k , i.e., for cluster C_k containing n_k points:

$$D_k = 2n + k \sum_{x_i \in C_k} \|x_i - \mu_k\|^2 \quad (1)$$

where $\|\cdot\|^2$ is the Euclidean distance of each point from its cluster's mean.

$$W_k = \sum_{p=1}^k \frac{1}{2n_p} D_p \quad (2)$$

Across ECG signals and datasets, the estimated optimal number of clusters was typically small, 3-8 clusters. A simple k-means clustering approach using this optimum number of clusters was then used to classify the QRS waveforms. An example of waveform distribution in these clusters is shown in Figure 3. Given that across recordings, the QRS complex is the most frequently encountered stereotypical (in morphology) signal, the cluster(s) containing the highest number of waveforms and lowest variance between them were selected for further analysis. Typically, the same cluster had both properties but if not (as in the example in Figure 3), waveforms from the two clusters were merged.

The last step of the algorithm involved the template estimation. This was achieved by fitting an optimal polynomial to the data in the cluster meeting the above criteria of membership and variance (in a least-squares sense as well as based on the Akaike Information Criterion (AIC) for model fit). This represented the template to be subtracted from the raw ECG signals to remove the QRS signal. At each previously identified peak, the template was adapted to the peak amplitude prior to subtraction. The process was repeated to remove additional peaks, until the residual signal amplitude statistics were below the threshold for extreme outliers. An example of raw ECG and separate ventricular and atrial signals is shown in Figure 4.

IV. RESULTS

The proposed algorithm was applied to continuous noninvasive ECG data from 6 pediatric epilepsy patients. Following decomposition of the ECG signals, atrial dynamics were further examined. In addition to simple cross-correlations between the two atrial signals, estimated using a 5-s sliding window, their complexity (or irregularity) was also investigated. The Higuchi fractal dimension (HFD) [18], [19] of each signal was estimated using a 5-s sliding window. Seizure consistently occurred in intervals with decreased or low fractal dimension in one or both atrial signals. Instead, other measures such as peak cross-correlation between these signals varied non-specifically as a function of seizures (Figure 5 bottom panel). Examples of atrial fractal dimensions in two patients are shown in Figures 5 and 6.

Table II summarizes the median (and IQR) atrial fractal dimension 5 min before to 5 min after a seizure, as well as 6 h removed from a seizure (true interictal interval) for each patient and statistical significance of their comparison (based on non-parametric comparisons of fractal dimensions in these intervals). Across patients, these differences were statistically significant in one or both leads. The ratio of fractal dimensions was significant only when dynamic changes in this measure for one of the two signals were non-significant across time. Other measures, such as signal cross-correlation, was consistently high (as expected) and changed non-significantly with time for all patients. This was also the case for signal power at dominant frequencies in the atrial spectrum. These findings suggest that atrial dynamics could be useful for estimating relevant measures that are sensitive to the underlying seizure dynamics. However, given the complexity of cardiovascular dynamics

over long periods of time, the specificity of atrial fractal dimension as a marker of seizure evolution needs to be further evaluated.

V. DISCUSSION

We have proposed a data-driven method for extracting atrial activity from very high-dimensional ECG signals. This approach uses the data structure and statistics to separate this activity from the high-amplitude QRS complex and various artifacts that typically contaminate noninvasive physiological signals recorded over long periods of time. Although this method has been applied to ECG data from 6 pediatric epilepsy patients in this study, it may be widely applicable to any study that collects this type of signals, e.g. polysomnography, and uses a small number of leads (1-2). Following separation of atrial and ventricular activity, we estimated several measures of atrial dynamics in the two leads, including their cross-correlation, fractal dimension, dominant frequency and spectral power. Only fractal dimension in one or both leads varied significantly in peri-ictal intervals (5 min prior to and following seizures) compared to interictal epochs. Across patients, seizures consistently occurred in periods of low fractal dimension and thus low atrial signal complexity. However, these findings as well as this measure's sensitivity and specificity needs to be rigorously evaluated in a large cohort to assess its potential utility as a marker of seizure evolution. To evaluate its performance, the separation method should also be compared to model-based approaches.

ACKNOWLEDGMENT

The authors would like to thank Sheryl Manganaro and William White at the Epilepsy Center at Boston Children's Hospital for their help with the data and Dr Philip Pearl for overall support of this work. This research is, in part, supported by the National Science Foundation, Division of Advanced Cyber-Infrastructure (grant ACI # 1649865).

REFERENCES

- [1]. Perrin RJ, Fagan AM, Holtzman DM, Multi-modal techniques for diagnosis and prognosis of Alzheimers disease, *Nature*, 461(7266): 916–922. [PubMed: 19829371]
- [2]. Libero LE, DeRamus TP, Lahti AC, et al., Multimodal neuroimaging based classification of autism spectrum disorder using anatomical, neurochemical, and white matter correlates, *Cortex*, 66:49–59, 2015.
- [3]. Qui S, Calhoon VD, van Erp TGM et al., Multimodal Fusion with Reference: Searching for Joint Neuromarkers of Working Memory Deficits in Schizophrenia, *IEEE Tans Med Imaging*, 37(1): 93–105.
- [4]. de Lemos JA, Ayers CR, Levine BD, et al., Multimodality Strategy for Cardiovascular Risk Assessment: Performance in 2 Population-Based Cohorts, *Circulation*, 135(22): 2119–2132.
- [5]. Mormann F, Andrzejak RG, Elger CE et al., Seizure prediction: the long and winding road, *Brain*, 130:314–333, 2007. [PubMed: 17008335]
- [6]. Kuhlmann L, Lehnertz K, Richardson MP et al., Seizure prediction ready for a new era, *Nat Rev Neurol*, 618–630, 2018. [PubMed: 30131521]
- [7]. Dumanis SB, French JA, Bernard C et al., Seizure Forecasting from Idea to Reality. Outcomes of the My Seizure Gauge Epilepsy Innovation Institute Workshop, *eNeuro*, 4(6):1–5, 2017.
- [8]. Zijlmans M, Flanagan D, Gotman J, Heart Rate Changes and ECG Abnormalities During EpilepticSeizures: Prevalence and Definition of an ObjectiveClinical Sign, *Epilepsia*, 43(8): 847–854. [PubMed: 12181003]

- [9]. Leutmezer F, Schernthaner C, Lurger S et al., Electrocardiographic Changes at the Onset of Epileptic Seizures, *Epilepsia*, 44(3): 348–354. [PubMed: 12614390]
- [10]. van der Lende M, Surges R, Sander JW, et al., Cardiac arrhythmias during or after epileptic seizures, *J Neurol, Neurosurg Psychiatry*, 87:69–74, 2016. [PubMed: 26038597]
- [11]. Rieta JJ, Castells F, Sanchez C et al., Atrial activity extraction for atrial fibrillation analysis using blind source separation, *IEEE Trans Biomed Eng*, 51(7): 1176–1186, 2004. [PubMed: 15248534]
- [12]. Castells F, Igual J, Rieta JJ et al., Atrial fibrillation analysis based on ICA including statistical and temporal source information, *ICASSP V93–V96*, 2003.
- [13]. Escoda O Divorra, Granai L, Lemay M et al., Ventricular and Atrial Activities Estimation Through Sparse ECG Signal Decompositions, *ICASSP*, 1–4, 2006.
- [14]. Lemay M, Jacquemet V, Forclaz A, et al., Spatiotemporal QRST cancellation method using separate QRS and T-waves templates, *Computers in Cardiology, IEEE* 2005.
- [15]. Mandelbrot B, How Long is the Coast of Britain? Statistical Self-Similarity and Fractional Dimension, *Science*, 156(3775): 636638, 1967.
- [16]. Tukey JW, W J, 1977, *Exploratory data analysis*, Addison-Wesley.
- [17]. Tibshirani R, Walther G, Hastie T, Estimating the number of clusters in a data via the gap statistic, *J. R. Statist. Soc. B*, 63(2): 411–423, 2001.
- [18]. Higuchi T, Approach to an irregular time series on the basis of fractal theory, *Physica D*, 277–283, 1988.
- [19]. Kecic S, Spacic SZ, Application of Higuchi's fractal dimension from basic to clinical neurophysiology: A review, *Comp Meth Biomed*, 133: 55–70, 2016.

Fig. 1.
Example of ECG signal statistics (median) over the course of 24 h and their respective root-mean-square (RMS) envelopes.

Fig. 2.
Schematic representation of data-driven algorithm for separation and sparse representation of ventricular and atrial activity in the ECG.

Fig. 3. Example of QRS waveform clustering for estimation of QRS template from the high-dimensional ECG signals. Cluster #4 had the lowest variance and cluster #8 the higher membership.

Fig. 4.

Example of raw ECG and separated atrial and ventricular activity (20 s segment).

Fig. 5. Fractal dimension over a period of ~38 h from Patient #3 for each atrial signal (left and right leads), the ratio of these dimensions and their peak cross-correlation. The onset time for 7 seizures is superimposed (red).

Fig. 6. Fractal dimension over a period of ~51 h from Patient #1 for each atrial signal (left and right leads) and the ratio of these dimensions. The onset time for 8 seizures is superimposed (red).

TABLE I**PATIENT CHARACTERISTICS**

Patient No.	Origin (Lobe)	Etiology	Age (years)	Sex
1	Right Temporal	Cerebellar Tumor	10	M
2	Bilateral Frontal	Cryptogenic	10	M
3	Right Temporal	Perinatal Stroke	22	F
4	Left Temporal	Encephalomalacia	14	F
5	Left Frontal	Cryptogenic	15	F
6	Right Frontal	Cortical Dysplasia	3	M

TABLE II

PERI-ICTAL AND ICTAL STATISTICS OF ATRIAL FRACTAL DIMENSIONS (FD-P AND FD-I)

Patient No.	Median FD-P (IQR) Peri-ictal	Median FD-I (IQR) Interictal	P-value
1	1.17 (0.08)	1.29 (0.10)	<0.05
2	1.06 (0.13)	1.47 (0.12)	<0.01
3	1.03 (0.11)	1.25 (0.09)	<0.01
4	1.14 (0.06)	1.33 (0.15)	<0.05
5	1.27 (0.09)	1.41 (0.10)	<0.01
6	1.06 (0.03)	1.28 (0.13)	<0.01