Unsupervised Classification of High-Frequency Oscillations in Human Neocortical Epilepsy and Control Patients


1Department of Bioengineering, 2Department of Statistics, The Wharton School, and 3Department of Neurology, University of Pennsylvania, Philadelphia, Pennsylvania; and 4Department of Neurology and 5Department of Neurosurgery, Mayo Clinic, Rochester, Minnesota

Submitted 10 December 2009; accepted in final form 25 August 2010

INTRODUCTION

In the epilepsy research community, there is mounting interest in the idea that high-frequency oscillations (HFOs)—narrowband transients recorded on the intracranial electroencephalogram (iEEG), having predominant frequency between roughly 100 and 500 Hz and lasting on the order of tens of milliseconds—can be used to identify epileptogenic brain tissue. From a clinical perspective, the utility of HFOs as biomarkers lies in their potential to improve the outcomes of surgical resections in patients with medication-resistant epilepsy, either by augmenting or replacing signals currently used to delineate epileptogenic cortex. HFOs might also serve as input signals to controllers within implantable devices designed to deliver therapeutic intervention—for example, targeted electrical stimulation, drug delivery, or tissue cooling—in closed-loop fashion. Such devices, one of which is currently in clinical trials (Barkley et al. 2006), promise to help the large population of medically refractory epilepsy patients who are not candidates for resective surgery, either because their seizures appear to emanate from “eloquent” cortical areas with vital function or because epileptiform activity appears diffuse or multifocal within the brain.

To date, most studies of HFOs in epilepsy rely on highly human intensive methods to extract the signals of interest from multichannel iEEG. With the goals of increasing the tractability of labor-intensive marking and analysis tasks and improving the interpretability of results, investigators typically perform dramatic data reduction steps before committing the data to statistical analysis. Examples of common data-culling measures include preselecting: 1) subjects, electrodes, channels, and time epochs that, on visual preinspection, show prominent examples (e.g., high signal-to-noise ratio) of the activity to be analyzed; 2) electrodes or contacts that, according to magnetic resonance images, appear located in gray matter or specific brain structures; 3) data recorded during certain states of consciousness, typically slow-wave sleep; 4) data recorded during specific brain states such as interictal, preictal, or ictal periods; 5) data recorded from periods deemed artifact-free, typically by human reviewers; and 6) data—or a subsample thereof, such as a 10-min segment—that satisfy some combination of the above-cited criteria. A smaller number of studies use machine-based HFO detection in semiautomated procedures in which similar human screening occurs before the automatic detector is deployed, as well as after, in an attempt to reduce the number of spurious machine detections. Such quasi-automated methods are designed with the primary aim of reducing the workload for human reviewers.

In addition, based in large part on human observation, most studies of HFOs in epilepsy presuppose two discrete categories of these oscillations, distinguished by the subband between 100 and 500 Hz in which the majority of the transient signal’s energy lies. The two types of high-frequency oscillations are termed “ripples” and “fast ripples,” and statistical analyses of HFO properties are typically preceded by the forced classification of all detected events as one of these two types.

Herein, we present a methodology for automatically detecting and classifying high-frequency oscillations in the human intracranial electroencephalogram. Differing from prior approaches to automated detection, we use no data preselection beyond our choice to study patients with epilepsy believed to be of neocortical origin and control patients, analyzing all recorded data from all available channels in all available patients. Furthermore, we do not presuppose the existence of two distinct subpopulations of HFOs. Instead, we treat this idea as a hypothesis, using it to inform the design of some of the
features we use as inputs to our classifier. We develop an unsupervised, rather than supervised, scheme for event classification because 1) we believe ideas about what constitutes a physiologically or clinically meaningful HFO are still evolving and the former framework is more natural for exploratory analysis; and 2) prior studies in which “expert” human reviewers have been asked to mark HFO-like activity on the iEEG have shown poor interrater reliability and reproducibility (Gardner et al. 2007). In a similar vein, we find no good reason to constrain the categorization of events to two types. Thus the classification method we use requires no prespecification of the number of HFO classes present in the data and includes the possibility that the data do not cluster at all.

The primary purpose of this study was to present our method—both its rationale and the technical details to replicate it—along with results that support its utility as a tool for automatically processing and organizing large quantities of neurophysiologic data containing spontaneous, transient events, in a manner that minimizes not only selection bias but also human subjectivity and labor. Further, we show objective evidence for the existence of distinct classes of oscillations that have heretofore been taken largely for granted, strengthening the argument that “ripples” and “fast ripples” are meaningful physiologic labels (even if perhaps not the best terminology for avoiding confusion in the literature), and suggest that a third class of oscillations containing components of both frequencies should be investigated.

METHODOLOGY

Patient population and data acquisition

Nine patients with medically refractory partial epilepsy believed to be of neocortical origin underwent implantation of subdural electrodes (Ad Tech Medical Instruments, Racine, WI) to localize the seizure onset zone after noninvasive monitoring was indeterminate. The location, number, and type of intracranial electrodes (grid and/or strip and/or depth electrodes) were determined by a multidisciplinary team, including neurosurgeons, neurologists, neuroradiologists, and neuropsychologists, as part of routine clinical care. Standard clinical grid and strip electrodes were modified under an Institutional Review Board (IRB)–approved research protocol, by adding arrays of nonpenetrating platinum–iridium microwires (40-μm diameter, with intrarray spacing of 0.5–1 mm center-to-center) between the clinical, 4-mm diameter contacts (Van Gompel et al. 2008). Standard clinical depth leads were similarly modified, in two ways: 1) by embedding microwires around the circumference of the lead-body between the 2-mm long clinical contacts; and 2) by passing a bundle of microwires within the lumen of the lead, so that they protruded by about 7–8 mm from the distal tip (Worrell et al. 2008). Schematic depictions of the electrode modifications and an image of a typical implanted subdural grid are shown in Fig. 1, A and B, respectively.

Two “control” patients, with chronic intractable facial pain but with no history of seizures, were similarly implanted, as part of an unrelated, IRB-approved research protocol investigating electrical stimulation of motor cortex as a potential treatment for their condition.

In addition to intracranial electrodes, a limited montage of standard gold scalp electrodes (EOM1, EOM2, Fp1, Fp2, Fpz, C3, and C4) was placed on all patients, as well as electrodes to record electromyographic signals from the chin and tibialis anterior surface. Two stainless steel surgical sutures (Ethicon, Somerville, NJ) were placed at the vertex region of the head at the time of surgery and used as reference and ground, respectively, for both intra- and extracranial electrodes.

Continuous, long-term data were acquired using the Digital Lynx Data Acquisition System (Neuralynx, Bozeman, MT) at 32,556 samples/s and 20 bits per sample (stored), from ≤144 channels in each patient (Brinkmann et al. 2009). The input dynamic range was ±132 mV and the noise level was ~1.3 μV root-mean-square (RMS), yielding approximately 18 effective bits. Recordings were made using DC-capable amplifiers and a 9-kHz analog low-pass filter was used to minimize aliasing effects. The bandwidth of the raw recordings was thus approximately near–DC–9 kHz. Sleep staging was not part of the experimental protocol.

Overview of HFO detection and classification

The automated HFO detection and classification algorithm is comprised of three major stages, depicted in Fig. 2. In the first stage, candidate HFO events are detected in the band-pass filtered iEEG using a method designed by Staba et al. (2002) to be sensitive, but which is highly nonspecific. This initial pass over the data set dramatically reduces its size. In the second stage, a statistical model of the local background iEEG surrounding each candidate event is built and events bearing too large a spectral similarity to the background activity according to the model are discarded from candidacy. In the final stage, computational features are extracted from the retained candidates (which we label simply “anomalies,” since they may or may not represent clinically meaningful events) and these features are
### Innovative Methodology

By “local” we mean the amount of time typically displayed when screening for HFOs in raw or modestly filtered data (e.g., 1 s/screen with ~0.5 – 500 Hz bandwidth, ~1.2-kHz sampling rate, and 7 μV/mm vertical scaling, on a monitor with 1,280 × 1,024 pixel resolution).

---

used, after a dimensionality reduction step, as inputs to a classifier. Importantly, the classifier uses an automatic method of determining the number of clusters in the data and the possibility that the data do not cluster at all (i.e., there is a single cluster) is not excluded.

In the following text, we describe the method in detail. All analyses were carried out using custom Matlab (The Mathworks, Natick, MA) scripts run on a computer with dual quad-core Intel Xeon E5430 CPUs at 2.66 GHz and 16 GB of RAM.

#### Stage 1: detection of candidate HFOs

HFOs are initially detected using methods closely adapted from those by Csicsvari et al. (1999a,b) and Staba et al. (2002). Raw, referential iEEG data are processed on a per-channel basis in 10-min nonoverlapping segments. Data are first decimated by a factor of 12 (8th-order low-pass Chebyshev type I filter, cutoff frequency of 1,085 Hz, forward and reverse filtered to eliminate phase distortion), to 2,713 Hz, before band-pass filtering between 100 and 500 Hz (20th order Cauer filter, forward and reverse filtered; specifications: 65 dB minimum lower/upper stop band attenuation, 0.5 dB maximum pass band ripple, 25 Hz lower/upper transition width). Filters were designed and tested for stability using Matlab’s filter design toolbox.

Using a 3-ms sliding window, a running RMS signal is computed from the band-pass data. The RMS signal is then compared against a threshold (the mean of the RMS signal plus 5SDs) and successive samples exceeding this threshold for a minimum duration of 6 ms are delimited by their upward and downward threshold crossing times. Marked events are then subject to the additional criterion that they must have at least six peaks >3SDs from the mean of the rectified band-pass signal. Finally, retained events separated by <10 ms are merged to generate a candidate set of HFOs.

#### Stage 2: retention of anomalies

Staba et al. (2002) reported high sensitivity for their detector when compared with a ground truth set of human markings, but the method suffers high false-positive rates (Gardner et al. 2007), especially when no preselection of data is performed. In pilot studies, we observed that a common putative failure mode for the detector was retaining transients with amplitudes larger than the average global background signal (on which the algorithmic threshold is based) but similar in morphology to the local background iEEG. An example of these signal (on which the algorithmic threshold is based) was observed in Fig. S2. In summary, false positives are common in the initial detection stage and this step can be improved by imposing additional constraints.

Candidate HFOs passing through stage 1 are processed individually in stage 2, which operates on the decimated data. Twelve hundred milliseconds (after a 5-ms guard time) on either flank of a given candidate are used to build a model of the local background iEEG. Background data are segmented into clips of length equal to the candidate detection.

For each background segment, as well as the candidate detection, the best linear fit is found using least-squares regression and removed by subtraction (i.e., the signal is detrended). Each segment is then energy-normalized by dividing by its Euclidean length and a power spectral density estimate is computed using a multitaper technique from Thomson (1982) (three tapers; 512-point Discrete Fourier Transform (DFT) (zero-padded); adaptive nonlinear combination method). The method in general involves multiplying the signal by n orthogonal data tapers having optimal spectral concentration properties—discrete prolate spheroidal sequences (Slepian and Pollak 1961)—before computing the squared magnitude of the DFT in each case. This produces n independent spectral estimates, which are then additively combined using weighting factors that are adaptively chosen depending on local characteristics of the spectrum.

The nonstationary local background iEEG is modeled using a mixture of Gaussians, under the assumption that each background segment is an observation emitted from one of a small number (one to three) of multivariate Gaussian components. Thus the probability of observing any local background segment p(x), is represented by the density function.

---

2. In the supplemental material (Figs. S1 and S2; Tables S1–S3) we compare the performance of the automated method with that of three expert human reviewers on an HFO verification task.

3. To simplify notation, in Eqs. 1, 2, and B1–B5, and supplemental material Eqs. S1–S7 that follow, we have suppressed the indexing subscripts that indicate that all functions and variables are specific to a particular candidate HFO, but we here emphasize that these subscripts are implicit.

4. Detections exceeding 50 ms (~14%) are first truncated to 50 ms to place a lower limit on the number of background segments available.

5. Multitaper techniques are particularly well suited to the analysis of short duration signals, as well as signals with spectra having a large dynamic range of amplitudes (Mitra and Pesaran 1999; Percival and Walden 1993), making them an appropriate choice for our application.

6. In the Gaussian Mixture Model (GMM) discussion in the following text, we use the notation style of Bishop (2006).
UNIVERSAL CLASSIFICATION OF HFOs IN EPILEPSY

FIG. 3. Macroelectrode (right frontal lobe) recording examples showing the lack of specificity of the stage 1 detector: partial motivation for the design of stages 2 and 3. A: a run of putative false positive stage 1 detections, none of which appears visually well distinguished from the surrounding intracranial electroencephalographic (iEEG) recording. Individual detections are delimited by solid (start) and dotted (stop) red lines. These detections should be contrasted with the single detection in B, a putatively valid high-frequency oscillation (HFO). C: typical sharp transient artifact detected by stage 1; unfiltered waveform. D: band-pass filtered version of C; this filter artifact is what the stage 1 detector actually “sees,” giving insight into why the event is flagged as an oscillation. The algorithm we present is designed to discard in stage 2 events like those in A and aggregate into a single cluster in stage 3 events like the one represented by C and D. More interesting events like those in B should group naturally into one or more clusters in stage 3.

Given the small number of background segments relative to the data, to combat the curse of dimensionality (Bellman 1961) the number of dimensions of each background segment (and the candidate segment) is reduced to two (i.e., $D = 2$), prior to learning the Gaussian Mixture Model (GMM). This is done using principal components analysis (PCA) (Hotelling 1933; Pearson 1901) on the frequency-domain representations of the background segments discussed earlier, after removing the mean and dividing by the SD of the background segments.

Since there is no a priori reason to prefer a specific number of mixture components in the GMM, the parameters for 1, 2, and 3 component mixtures are learned and the Bayesian Information Criterion (BIC) (Schwarz 1978) is used to make the final model selection. A metric based on the Mahalanobis distances (McLachlan 1999) from the HFO candidate to each component of the mixture model is then used to decide whether to reject the candidate or to pass it on to stage 3 (see Supplemental Material for more detail).7

Stage 3: unsupervised classification

FEATURE EXTRACTION. Before clustering, seven features are computed on the data corresponding to each event output from stage 2. All features were designed or chosen with one of two goals in mind: 1) to permit separation of “ripples” and “fast ripples,” should there indeed be evidence for such compartmentalization (and not otherwise); 2) to aggregate highly stereotypical, sharp transient recording artifacts that were observed in pilot studies (e.g., Fig. 3, C and D). Features in category 1 were conceived a priori (blinded to all patient data), whereas features in category 2 were developed and refined using labeled artifacts from two patients (CT01 and SZ05). The features used are described in detail in APPENDIX A. The set of feature vectors \( \{F_1, \ldots, F_I\} \), where \( I \) is the total number of anomalies passing through stage 2 (across all patients and intracranial electrodes; in our case, \( I = 290,273 \)), is the data set that is clustered.

7 The online version of this article contains supplemental data.
K-MEDIODS CLUSTERING. Clustering was accomplished using a variant of the $k$-means algorithm (MacQueen 1967). The $k$-means algorithm seeks to minimize the quantity

$$ W_K = \sum_{k=1}^{K} \frac{1}{2N_k} \sum_{i \in C_k} d_{ii} $$

(3)

where $K$ is the number of clusters, $C_k$ denotes the set of indices of the observations in cluster $k$ (with each observation assigned to a single cluster), and $N_k$ denotes the number of observations in cluster $k$. $d_{ii}$ is the squared Euclidean distance between $P$-dimensional observations $x_i$ and $x_j$:

$$ d_{ii} = \sum_{p=1}^{P} (x_{ip} - x_{jp})^2 = \|x_i - x_j\|^2 $$

(4)

The quantity $W_K$ can be viewed as a modified version of the within-cluster point scatter (Hastie et al. 2001) in which each cluster’s contribution is now weighted inversely by its membership. Thus diffuse clusters with few members are not counted as energetically favorably as compact clusters with many. Minimizing this metric is exactly equivalent to minimizing the quantity

$$ \sum_{k=1}^{K} \sum_{i \in C_k} \|x_i - \bar{x}_k\|^2 $$

(5)

where $\bar{x}_k$ is the mean of the elements in cluster $k$. An iterative descent algorithm for minimizing the above-cited criterion can be obtained by first defining clusters via prototypes $\{u_{ik}\}_{i=1}^{K}$ and attempting to minimize, over parameters $u$ and $Y$, the error function

$$ e_k(u, Y) = \sum_{i,j} Y_{ij}\|x_i - u_{jk}\|^2 $$

(6)

where $u$ is the $P \times K$ matrix of prototypes and $Y$ is a $K \times l$ indicator matrix of cluster assignments, with $l$ the total number of data observations. That is, $Y_{ai} = 1$ whenever $x_i$ is assigned to cluster $k$, and 0 otherwise. The algorithm proceeds as shown in APPENDIX B.

Instead of using the L2 norm (Euclidean distance), the classification scheme uses the L1 norm (Manhattan distance) as distance metric, which yields the $k$-medoids algorithm. In step 2 of the algorithm (Eq. B6 in APPENDIX B), a given point is assigned to the prototype that is closest in the L1 sense, and in step three (Eq. B7 in APPENDIX B) the appropriate minimization is obtained by computing the mediod, rather than the centroid, for each newly formed cluster. The $k$-medoids method is used to achieve a clustering that is more robust against outliers than $k$-means, albeit at computational expense (Hastie et al. 2001).

The gap statistic

In the preceding text, $W$ has been subscripted with “$K$,” the number of clusters, to emphasize that the $k$-means and $k$-medoids algorithms require that the number of clusters in the data be prespecified. This is an obvious limitation in situations like the present study, in which the number of clusters (i.e., “types” of anomalous events passing through stage 2) is a priori unknown and is itself a major goal of the exploratory analysis.

$W_k$ is a monotonically decreasing function of $k$; this can be intuited by considering the projection to the extreme case, in which each observation constitutes its own cluster and thus $W_k$ is equal to zero. Plots of $W_k$ versus $k$ often show a characteristic “elbow point,” a value for $k$ at which previously large decreases in the function become markedly flattened. Traditionally, this point has been taken to represent the transition between decreases that represent an evolution toward the data’s true underlying structure and decreases that are simply attributable to the inevitably greater compactness of clusters that can be achieved by adding more cluster centroids. This elbow value of $k$ is thus typically taken as the “optimal” number of clusters.

Tibshirani et al. (2001) addressed the issue of estimating the optimal number of clusters in a data set given a similarity metric and error measure, presenting a statistical procedure to formalize the above-cited elbow heuristic concept. Their idea was to compare the graph of $\log (W_k)$ versus $k$ with its expectation, $E_k[\log (W_k)]$, under a suitable null reference distribution of the data. They defined the “gap statistic” for sample size $I$, $G(k)$, as

$$ G(k) = E_k[\log (W_k)] - \log (W_k) $$

(7)

and proposed to choose the optimal $k$ as that for which the gap statistic is largest after taking its sampling distribution into account. The null model assumed is a single-component model and the statistical procedure is designed to reject it in favor of a $k$-component model ($k > 1$), if the strongest evidence for any such $k$ under consideration merits it. The choice of the number of $k$ values over which to probe is of course still up to the experimenter. As recommended by Tibshirani et al. (2001), a uniform distribution over a box aligned with the principal components of the data is used as the form of the single-component reference distribution.

The clustering procedure begins with the data set $\{\tilde{F}_i, \ldots, \tilde{F}_I\}$: the PCA-reduced feature-vector representations of anomalies 1 through $I$ that have passed through stage 2 (see APPENDIX A). For number-of-clusters values $k = 1$ through 20, the $k$-medioids algorithm is run using 10 random initializations and the run that yields the lowest $\log (W_k)$ is chosen. Then, again for each $k$, an expected value for $\log (W_k)$ under the null reference is computed, using 20 reference data sets for each value of $k$. A given reference set is created by performing a PCA rotation of the data, sampling $I$ vectors uniformly over the range of values along each dimension, and then back-rotating the samples into the original coordinate frame. Each reference set is then clustered in an identical manner to the true data set and the expected value for each $\log (W_k)$ is taken as the average over the reference sets. In accordance with Tibshirani et al. (2001) a corrected SD $s_k$, which accounts for the simulation error is then computed as

$$ s_k = sd_k \sqrt{1 + \frac{1}{B}} $$

(8)

where $sd_k$ is the SD and $B$ is the number of reference sets, in our case 20. Finally, again following Tibshirani et al. (2001), the optimal clustering of the data is taken as that which meets the following criterion

$$ k = \min \{k | \text{Gap}(k) \geq \text{Gap}(k + 1) - s_{k+1}\} $$

(9)

where $k$ is the estimated optimal number of clusters. (If the preceding set were empty, $k$ would be set equal to the largest value tested, which in our case is 20.)

RESULTS

The number of channels and total number of iEEG hours processed per subject are shown in Table 1, along with the number of events detected in stage 1 and retained after stage 2. Per-channel computation time, estimated using the data from subject SZ01, was about 1.94 s/min (1.2 s/min for stage 1, 0.7 s/min for stage 2, and 0.04 s/min for stage 3).

Across all 11 subjects, a total of 1,423,741 HPO candidates were detected in stage 1. The summed duration of all stage 1 detections was 13.7 h, or 0.04% of the total data processed, giving an indication of the rarity of stage 1 detections. Among these stage 1 candidates, 290,273 (−20%) were flagged as
anomalous events in stage 2 and passed on to stage 3 for clustering. To verify that the stage 2 anomaly detection algorithm was not performing the trivial task of taking a random subsample of post stage 1 events, we tested the null hypothesis that the proportion of events retained after stage 2 was the same for all patients, finding, as we expected, evidence to reject it ($\chi^2(10, N = 1,423,741) = 57,088.23, P < 0.0001$). To examine whether electrode size influenced the per-channel event detection rate, we tested the null hypothesis that the cumulative distribution functions for event detection rate were identical for macro- and microelectrodes. There was a trend toward higher detection rates for macroelectrodes, but a two-sided test was not significant at the 0.05 level (Mann–Whitney U test, $z = 1.84, n_{macro} = 479, n_{micro} = 622, P = 0.07$). To examine whether electrode size influenced the proportion of retained events after stage 2, we tested the null hypothesis that the cumulative distribution functions for event retention fraction were identical for macro- and microelectrodes. Microelectrodes tended to have lower event retention rates and a two-sided test was highly significant (Mann–Whitney U test, $z = -6.05, n_{macro} = 479, n_{micro} = 622, P \ll 0.001$). (The latter two tests assume homogeneity across patients.)

Four clusters were found in stage 3 by the k-mediods/gap-statistic method. The results of clustering are illustrated in Figs. 4 and 5. Figure 4A shows a clustering of 1,000 events selected randomly across all patients, plotted in the space defined by $f_1$, $f_2$, and $f_3$ and colored according to class label. Figure 4B gives an alternative view of the clustering, projecting the points onto the planes defined by all possible pairs of principal components. Boundaries are apparent between the red, the green, and the two blue clusters taken together, while the separation between the light and dark blue clusters is less marked in these limited two- and three-dimensional representations of the four-dimensional clustering.

Figure 5 shows the cluster prototypes for each of the four clusters found by applying our algorithm to all 290,273 events passing through stage 2. The prototypes are found by computing the class members nearest (in the Mahalanobis sense) to the cluster mediods in feature space. In Fig. 5A, both the raw (top) and the 100–500 Hz band-pass filtered (bottom) detection for each prototype are displayed. In Fig. 5B, we show five randomly selected members (band-passed version only) from each cluster. Supplemental Figs. S3–S6 show the latter events within the context of surrounding iEEG.

Qualitatively speaking, cluster 1 ($n = 70,671; n_{macro} = 47,893, n_{micro} = 22,778$) is comprised of irregular, mixed-frequency events; cluster 2 ($n = 92,744; n_{macro} = 34,665, n_{micro} = 58,079$) is comprised of sharp-rising transients that appear to be artifacts (i.e., whose oscillatory characteristics come from filtering the sharp rise); cluster 3 ($n = 38,945; n_{macro} = 10,529, n_{micro} = 28,416$) is comprised of fast, regular waveforms of relatively pure frequency; and cluster 4 ($n = 87,913; n_{macro} = 48,190, n_{micro} = 39,723$) is comprised of slow, regular waveforms of relatively pure frequency. We rejected the null hypothesis that the proportions of events detected by macro- and microelectrodes were the same across clusters [$\chi^2(3, N = 290,273) = 23,678.65, P < 0.0001$].

Supplemental Table S5 gives median event rates and amplitudes (excluding putative artifacts), broken down by subject and electrode type; Supplemental Table S6 gives event amplitudes broken down by cluster.

To investigate the stability of the clustering algorithm and test whether the aggregated clustering results we show in Figs. 4 and 5 were heavily influenced by individual subjects, we performed two additional experiments.

---

**Table 1. Raw data summary**

<table>
<thead>
<tr>
<th>Subject</th>
<th>Number of Channels</th>
<th>Total iEEG Hours</th>
<th>Number of Stage 1 Events</th>
<th>Number of Stage 2 Events</th>
</tr>
</thead>
<tbody>
<tr>
<td>CT01</td>
<td>144</td>
<td>1,176</td>
<td>44,819</td>
<td>14,768</td>
</tr>
<tr>
<td>CT02</td>
<td>128</td>
<td>1,877</td>
<td>49,576</td>
<td>12,595</td>
</tr>
<tr>
<td>SZ01</td>
<td>45</td>
<td>480</td>
<td>11,436</td>
<td>3,255</td>
</tr>
<tr>
<td>SZ02</td>
<td>126</td>
<td>1,890</td>
<td>86,958</td>
<td>17,647</td>
</tr>
<tr>
<td>SZ03</td>
<td>69</td>
<td>977</td>
<td>79,138</td>
<td>18,220</td>
</tr>
<tr>
<td>SZ04</td>
<td>90</td>
<td>2,122</td>
<td>64,617</td>
<td>29,215</td>
</tr>
<tr>
<td>SZ05</td>
<td>84</td>
<td>5,460</td>
<td>472,351</td>
<td>77,684</td>
</tr>
<tr>
<td>SZ06</td>
<td>115</td>
<td>10,139</td>
<td>132,336</td>
<td>36,254</td>
</tr>
<tr>
<td>SZ07</td>
<td>86</td>
<td>5,615</td>
<td>429,010</td>
<td>62,615</td>
</tr>
<tr>
<td>SZ08</td>
<td>104</td>
<td>812</td>
<td>34,312</td>
<td>14,167</td>
</tr>
<tr>
<td>SZ09</td>
<td>110</td>
<td>735</td>
<td>19,188</td>
<td>3,853</td>
</tr>
<tr>
<td>All</td>
<td>1,101</td>
<td>31,283</td>
<td>1,423,741</td>
<td>290,273</td>
</tr>
</tbody>
</table>

Total iEEG hours is the sum over all channels of the length of time recorded for each channel.

---

**Fig. 4.** A: clustering of 1,000 randomly selected events (across all patients) passing through stage 2, in the 3-dimensional space defined by $f_1$, $f_2$, and $f_3$. Each of the 4 clusters found by the algorithm is colored uniquely. B: the points on the left projected onto all possible pairs of principal components; axes are scaled to enable better visualization of cluster boundaries, so a small number of outlying points are not visible.
In the first, we applied the clustering algorithm to individual subjects, clustering 100 random samples drawn from the set of all events passing through stage 2 for each subject. The size of each of the 100 random samples was 25% of all the subject’s post stage 2 events. Results are shown in Fig. 6, which contains one panel for each patient: a histogram summarizing the outcome of the 100-iteration sampling experiment. Figure 6, bottom right represents the results of a similar random sampling experiment performed on the data aggregated across subjects, with the only difference being that sample sizes of 5% (15,000 of all 290,273 events passing through stage 2 across all subjects) were used, due to the computational expense of clustering 25% 100 times. Each individual subject’s histogram is colored in proportion to the relative number of events passing through stage 2 for that subject. Thus histogram color reflects the weight of the subject’s contribution to the aggregated histogram in the bottom right panel. The number in red indicates the mode of the distribution, which can be interpreted as the number of clusters the algorithm yields for that patient. The degree of concentration of the distribution for a given individual reflects the stability of the algorithm for that subject, which is a function of the variability and shape of the subject’s events in feature space, as well as the sample size clustered.

In the 6 of 11 subjects for whom the algorithm yielded 4 or fewer clusters, the clusters generally appeared to be subsets of those found in the aggregated data. In the remaining 5 subjects, for whom there were greater than 4 clusters, the additional clusters generally appeared to have resulted from oversplitting some combination of the 4 clusters found in the aggregated data. These trends are illustrated in Fig. 7, which shows examples of prototypes derived from clustering individual subject data. Prototypes in Fig. 7 are given a color corresponding to the aggregate prototype to which they bear the closest visual resemblance (cf. Fig. 5, bottom set of waveforms). Subject SZ01, for whom the algorithm yields strong evidence for three clusters, appears to exhibit the artifact, mixed, and slower oscillation clusters, but not the faster oscillations of the aggregated clustering. Subject SZ08, for whom the algorithm weakly suggests two (over three) clusters, prominently exhibits the artifact and slower oscillation clusters. Subject SZ03, for whom the algorithm performs relatively poorly, also appears to have clusters that are a subset of those found in the aggregate—i.e., the mixed frequency and slower oscillation clusters—but they are
often highly split by the algorithm. Figure 7 shows one particular iteration in which six clusters were obtained from what appears to be these two.

To quantify these observations about the stability of the algorithm, we used Rand’s C-statistic (Rand’s index; Rand 1971), a common measure of agreement between two partitionings of a data set. Rand’s index is simply the fraction of all pairs of observations that are treated similarly by the two clusterings, and one, when the clusterings are identical.

We were interested in how similar one partitioning was to another, within a subject, for iterations on which the algorithm returned more than a single cluster. Because each of the 100 clusterings for a given subject was based on an independent random sample, the samples on which any two clusterings $K$ and $K'$ were based almost always contained some nonoverlapping observations. Because the Rand index has nothing to say in regard to such unique observations, we considered only the intersection of the two clustered samples. For each individual subject, as well as for the aggregated clustering, we computed Rand’s index for all
these four-cluster partitionings were highly similar to one another. Figure 8 shows that Shannon entropy \cite{12} decreases with sample size for the aggregated data. There is a negative linear relationship (over the range of sample sizes considered) between the Shannon entropy of the distribution of number of clusters returned by the algorithm and sample size \( [\beta = -4.09 \times 10^{-5}, r(6) = -4.45, P = 0.004; \rho^2 = 0.77] \). Distributions were formed at each sample size by making 100 random draws from all post stage 2 detections. This result provides further quantitative evidence that the stability of the algorithm improves with the addition of data.

In the second experiment, to examine the influences of individual subjects on the aggregate result we performed “leave-one-out” clustering. Each subject, in turn, had his/her events removed from the pool available for clustering and 5,000 events were randomly drawn without replacement from the remaining events, in 25 iterations. Figure 9 shows that only subjects SZ05 and SZ07 alter the aggregate result of four clusters, suggesting that they were the predominant contributors to at least one cluster. Indeed, subjects SZ05 and SZ07 account for 73% of all events in cluster 3, whereas they make 

\[
\text{Shannon entropy} = S = -\sum_i p_i \log_2 p_i, \quad \text{where } p_i = \text{the probability of being in state } i, \text{ and } P \log_2 P := 0 \text{ when } P = 0.
\]

**TABLE 2. Rand’s C-statistic**

<table>
<thead>
<tr>
<th>Subject</th>
<th>( \bar{c} )</th>
<th>sd</th>
<th>( n )</th>
<th>( \bar{s} )</th>
<th>( \bar{s}_r(4, 950) )</th>
<th>sd(( \bar{s}_r(4, 950) ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>CT01</td>
<td>0.843</td>
<td>0.114</td>
<td>2,080</td>
<td>923</td>
<td>0.610</td>
<td>0.314</td>
</tr>
<tr>
<td>CT02</td>
<td>0.873</td>
<td>0.060</td>
<td>4,656</td>
<td>787</td>
<td>0.830</td>
<td>0.182</td>
</tr>
<tr>
<td>SZ01</td>
<td>0.787</td>
<td>0.129</td>
<td>4,851</td>
<td>204</td>
<td>0.778</td>
<td>0.146</td>
</tr>
<tr>
<td>SZ02</td>
<td>0.893</td>
<td>0.092</td>
<td>4,950</td>
<td>1,103</td>
<td>0.894</td>
<td>0.092</td>
</tr>
<tr>
<td>SZ03</td>
<td>0.936</td>
<td>0.020</td>
<td>4,005</td>
<td>1,139</td>
<td>0.782</td>
<td>0.329</td>
</tr>
<tr>
<td>SZ04</td>
<td>0.923</td>
<td>0.024</td>
<td>4,753</td>
<td>1,826</td>
<td>0.891</td>
<td>0.159</td>
</tr>
<tr>
<td>SZ05</td>
<td>0.995</td>
<td>0.002</td>
<td>946</td>
<td>4,856</td>
<td>0.670</td>
<td>0.329</td>
</tr>
<tr>
<td>SZ06</td>
<td>0.935</td>
<td>0.054</td>
<td>3,160</td>
<td>2,266</td>
<td>0.797</td>
<td>0.209</td>
</tr>
<tr>
<td>SZ07</td>
<td>0.895</td>
<td>0.085</td>
<td>105</td>
<td>3,919</td>
<td>0.793</td>
<td>0.350</td>
</tr>
<tr>
<td>SZ08</td>
<td>0.934</td>
<td>0.054</td>
<td>4,950</td>
<td>885</td>
<td>0.934</td>
<td>0.054</td>
</tr>
<tr>
<td>SZ09</td>
<td>0.843</td>
<td>0.114</td>
<td>4,950</td>
<td>241</td>
<td>0.843</td>
<td>0.115</td>
</tr>
<tr>
<td>All</td>
<td>0.950</td>
<td>0.074</td>
<td>3,403</td>
<td>775</td>
<td>0.760</td>
<td>0.311</td>
</tr>
</tbody>
</table>

Rand’s C-statistic (Rand’s index) for all pairs of clusterings in which \( (K, K') \geq 1 \) (columns 2–5). Shown for each subject are the mean \( \bar{c} \), SD sd, size of the sample (i.e., number of clustering pairs) on which the mean was based \( n \), and the average number of overlapping data observations in each pair of samples \( \bar{s} \). A value of 1 for any given \( \bar{c} \) would indicate a perfect match between all \( n \) pairs of clusterings. Columns 6 and 7 included for reference, give the mean value for the Rand index when all 100-choose-2 (4,950) pairs of clusterings are considered and its SD, respectively. The average number of overlapping data observations in this case is so close to the value in column \( \bar{s} \) (i.e., \( \pm 1 \) observation, typically less) for all subjects that we have chosen not to replicate it.

pairs of clusterings in which both numbers of clusters were \( \geq 1 \). Results are shown in Table 2. For completeness, we also show the results when the condition \( (K, K') > 1 \) is removed.\(^9\)

The high values of \( \bar{c} \) in the second column of Table 2 (and the accompanying low SDs in column 3) indicate that in cases where the algorithm returns a number of clusters \( > 1 \) on different iterations, the clustering tends to have considerable agreement. This is consistent with the idea that the number-of-cluster variability for individual subjects comes largely from the bulk splitting and merging of the clusters that were found in the aggregate.\(^10,11\)

Despite the expected inconsistency on the individual subject data apparent in Fig. 6, the algorithm performs robustly on the aggregated data (bottom right panel), as well as on most patients for whom a relatively large number of events were clustered. Specifically, 77% of the clustering iterations performed on the aggregated data yielded four clusters; columns two and three of Table 2 (last row) show that, furthermore, these four-cluster partitionings were highly similar to one another. Figure 8 shows that Shannon entropy \cite{12} decreases with sample size for the aggregated data. There is a negative linear relationship (over the range of sample sizes considered) between the Shannon entropy of the distribution of number of clusters returned by the algorithm and sample size \( [\beta = -4.09 \times 10^{-5}, r(6) = -4.45, P = 0.004; \rho^2 = 0.77] \). Distributions were formed at each sample size by making 100 random draws from all post stage 2 detections. This result provides further quantitative evidence that the stability of the algorithm improves with the addition of data.

In the second experiment, to examine the influences of individual subjects on the aggregate result we performed “leave-one-out” clustering. Each subject, in turn, had his/her events removed from the pool available for clustering and 5,000 events were randomly drawn without replacement from the remaining events, in 25 iterations. Figure 9 shows that only subjects SZ05 and SZ07 alter the aggregate result of four clusters, suggesting that they were the predominant contributors to at least one cluster. Indeed, subjects SZ05 and SZ07 account for 73% of all events in cluster 3, whereas they make

\[
\text{Shannon entropy} = S = -\sum_i p_i \log_2 p_i, \quad \text{where } p_i = \text{the probability of being in state } i, \text{ and } P \log_2 P := 0 \text{ when } P = 0.
\]
an average combined contribution of only 45% in clusters 1, 2, and 4 (and 48% to the total number of events clustered).

Taken collectively, the results of both the individual patient and aggregated clustering experiments are objective evidence in support of the idea that human neocortex is capable of generating distinct classes of transient oscillations within the 100–500 Hz frequency band, and that there are three such classes (excluding the putative artifact class).

**DISCUSSION**

**Main contributions**

The first contribution of this work is automation. We have introduced a new methodology, based on prior work, for automatically detecting and clustering high-frequency oscillations on the human intracranial electroencephalogram. We believe this is significant for several reasons. Perhaps the most pedestrian, but nonetheless valuable, is the fact that it substantially decreases human labor. Researchers have reported that it can take up to 3 h to mark HFOs in a single channel of EEG that is just 1 min in duration (Urrestarazu et al. 2007). Assuming a generous per-channel marking rate of 1 h per 10 min of data, it would take a superhuman reviewer (who did nothing but mark EEG all day at the same rate for 8 continuous hours, never taking a weekend or holiday) over 64 yr to manually mark the quantity of data processed for this study.

Although the number of subjects in our sample is comparable to that of many previous studies of HFOs in epilepsy, the total amount of iEEG data we analyzed, to the best of our knowledge, is vastly larger. Analyzing similarly large databases of iEEG—ideally with an even greater number of subjects—will, in our opinion, be required to make definitive statistical statements about the relationship between HFOs and seizure generation going forward. We consider the present effort a step in the direction of such large-scale analyses.

Importantly, in our automated approach we have done away with all data preselection. One could argue that this is not as beneficial as we would like to believe. After all, reducing the human workload is not the only motivation for performing data preselection; one might claim, for example, that choosing channels in the appropriate brain regions or epochs during certain states of wakefulness is critical to the interpretability of findings, or that interesting results will be entirely masked in the absence of these steps. However, in fact, increasingly researchers are finding exactly the opposite: that many of the preselection steps previously thought to be critical are most likely not. In the following text, we list just a few examples of beliefs that have restricted past HFO analyses but are no longer as widely held (in parentheses, we list examples of studies that have cast doubt on the original belief): 1) that HFOs occur only in hippocampal and parahippocampal areas (they also are seen in neocortex; Grenier et al. 2001, 2003; Jirsch et al. 2006); 2) that HFOs occur only during periods of slow-wave sleep (they also are observed during the active states of waking and REM sleep; Bagshaw et al. 2009; Grenier et al. 2001); 3) that HFOs occur only interictally (they are observed ictally; Jirsch et al. 2006) and preictally (Jacobs et al. 2009) as well; and 4) that HFOs can be recorded only with microelectrodes [they can also be recorded with standard clinical macroelectrodes (Jirsch et al. 2006), albeit perhaps with less fidelity (Worrell et al. 2008)].

Moreover, our aim is precisely to investigate the generalizability of some of the enticing findings that have recently been reported in the study of HFOs in human epileptic brain. We are concerned with how, if at all, conclusions about the practical utility of HFOs for localization of epileptogenic regions are modified when the many conditions implied by stringent data preselection are removed. If the answer turns out to be “not much,” this will be particularly encouraging for the prospects of using HFOs as control signals on board implantable seizure therapy devices, for example, given that these devices, at least in early generations, will not have the same luxuries as human researchers when it comes to preidentifying optimal data for processing.

The second major contribution of this work is that we have provided objective evidence for the existence, in the neocortex of epilepsy and control patients, of distinct classes of transient oscillations in the range between 100 and 500 Hz. We do not think this is splitting hairs. Most studies of HFOs in epilepsy of which we are aware simply presuppose the existence of discrete entities typically termed “ripples” and “fast ripples” (albeit with experimental evidence, usually of the observational variety, for doing so). Moreover, at least one group has reported that their data did not clearly show distinct ripple and fast ripple modes (Worrell et al. 2008). One study that attempted to quantitatively verify the existence of discrete populations (Staba et al. 2002) did so by fitting functions to the distribution of peak frequency values of detected events, and arguing that the bimodality of the best-fitting function was telling. The authors acknowledged that their analysis ignored the possibility of events that themselves had bimodal or more complex frequency spectra. Here, we show evidence that there actually is a cluster of mixed-frequency events and, furthermore, that it appears to be separate from two other clusters that are comprised more purely of a single frequency.

Waveforms with similar time-domain characteristics to our mixed-frequency class prototype can be generated by applying a 100–500 Hz band-pass filter to a white Gaussian noise burst, and the features we use for classification will likely not disambiguate these types of broadband transients from truly bimodal ones. Having examined the spectra of these mixed-frequency events, often observing two prominent peaks, however, we feel confident suggesting that they should be further investigated as a legitimately hybrid class of oscillations. As for the other three clusters, we believe, as mentioned, that one is comprised largely of clinically uninteresting recording artifacts. The other two bear sufficient resemblance to what have been labeled ripples and fast ripples to persuade us that we are recording the oscillations that have been recognized by other researchers in the community.

In this work, we were interested in whether 100–500 Hz HFOs cluster in a frequency-based feature space. We focused on 100–500 Hz oscillations because the largest body of HFO research in epilepsy to date considers this range and our goal was to probe the generalizability of this prior research using automated techniques. We do not wish to suggest that 100–500 Hz is a definitive range; indeed, researchers are beginning to report HFOs of potential clinical interest >500 Hz and up to around 1,000 Hz (Kobayashi et al. 2010; Worrell et al. 2008) and a natural extension of the present work would be to broaden the analysis bandwidth. We also note that frequency alone may not distinguish pathological from physiological oscillations (Engel Jr et al. 2009); studies that seek to incor-
porate our method in correlating HFO activity with pathology will likely benefit from explicitly including other variables (e.g., anatomic coordinates, electrode size, presence of lesions, etc.) in postclustering analyses.

Conclusion

A flurry of recent studies suggests that the relationship between subclasses of high-frequency oscillations recorded using nonstandard, high bandwidth data acquisition systems could be a more reliable indicator of seizure-generating cortex than the waveform morphologies typically used by clinicians for seizure localization. In this study, we introduced an algorithm for detecting and classifying these signals in an automatic fashion and demonstrated the tractability of exploring a data set of unprecedented volume using the method. Using an unsupervised approach that did not presuppose a specific number of clusters in the data, we showed direct evidence for the existence of distinct classes of oscillations within the 100–500 Hz frequency range in a population of neocortical epilepsy patients and controls, addressing a current controversy in the field. The number of classes we found, four (three plus one artifact cluster), is consistent with prior studies that distinguish between ripple and fast ripple oscillations and, additionally, suggests that a class comprised of hybrid ripple/fast-ripple events exists.

By minimizing selection bias and other human judgment in detecting and classifying HFOs, we believe that it will be possible to answer the important question of whether they can practically be used to delineate epileptogenic brain. This work also demonstrates the tractability of using machine learning techniques to analyze large streams of high-resolution neurophysiologic data, which may become standard in the future as new devices and techniques for mapping and modulating function and dysfunction in neurological disorders, on multiple brain scales, evolve.

APPENDIX A

HFO distinguishing features

1. Power band ratio: (250–500 Hz)/(100–200 Hz). Computed on the band-passed data; frequency-domain input. The ratio of estimated power in the hypothesized “fast-ripple” band to that in the hypothesized “ripple” band. If there are indeed two distinct populations of HFOs whose respective energy is concentrated in these two bands, one would expect the distribution of data along this feature to be strongly bimodal.

\[ f_1 = \frac{\hat{P}_{[250,500]}}{\hat{P}_{[100,200]}} \]  

(A1)

where

\[ \hat{P}_{[a,b]} = \sum_{k \in \mathbb{Z}, \{4 \leq k \leq \tau \}} |M[k]|^2 \quad (a < b) \in \{0, Fs/2\} \]  

(A2)

M[k] can be a multitaper power spectral density estimate. In our case, to lessen the computational burden we use the modified periodogram

estimate, \( M[k] = \sum_{n=0}^{N-1} w[n]\tilde{x}[n]e^{-j2\pi nk/L} \), \( k \in \{0, 1, \ldots, N - 1\} \), with an N-point Discrete Fourier Transform (DFT). \( \{w[n]\} \) is the Hanning window (Oppenheim et al. 1999) of length L and \( \{\tilde{x}[n]\} \) is the detrended signal, obtained by subtracting the least-squares fit of a line from the original length L signal. \( \{x[n]\} \). Note the signal \( m[n] = w[n]\tilde{x}[n], n \in \{0, 1, \ldots, L - 1\} \) is padded with N-L zeros before computing the DFT. \( g(n) = (0.5 + NTu), \) where T is the sampling period and \( u \) is the “floor” operator, which as used in this context accomplishes rounding to the nearest integer. This handles cases when a and b do not coincide precisely with the frequency values sampled by the DFT. Fs is the sampling frequency.

2. Spectral centroid. Computed on the band-passed data; frequency-domain input. The frequency corresponding to the “center of mass” of the spectral content. This feature should distinguish events whose energy is relatively concentrated in either the ripple or fast ripple subband alone from those with narrow components in both bands, with relatively broadband spectra, or with predominant energy in an intermediate band

\[ f_2 = \frac{\sum_{k=0}^{N/2} k |M[k]|^2}{\sum_{k=0}^{N/2} |M[k]|^2} \]  

(A3)

3. Spectral peak. Computed on the raw data; frequency-domain input. Frequency corresponding to the peak of the estimated power spectral density. This feature should capture some differences in the spectral characteristics of the background signal on which the detections are superimposed

\[ f_3 = \frac{1}{NT} \arg \max_k |M[k]|^2 \quad k \in \{0, 1, \ldots, N/2\} \]  

(A4)

4. Line length. Computed on the raw data, after first-order backward differencing (Gardner et al. 2007; Usui and Amidror 1982); time-domain input. Detections are first detrended and energy normalized by dividing by their Euclidean lengths. Because detections are also of different durations, we normalize by the signal length (i.e., number of samples) as well. This feature was included because it has proven broadly applicable in automatic EEG discrimination tasks, such as seizure (Esteller et al. 2001) and gamma oscillation (Gardner et al. 2007) detection.

\[ f_4 = \frac{1}{L} \sum_{k=0}^{L-1} |x[n+1] - x[n]| \]  

(A5)

Artifact distinguishing features

1. Global average-local peak ratio. Computed on the band-passed data; time-domain input. Ratio of the global maximum value to the average of all other local maxima (i.e., excluding the global maximum). Detections are first detrended and energy normalized using a three-point moving average filter.

\[ f_5 = \frac{g^*}{\sum_{\{g_k \in [G]: g_k < g^*\}} g_k} \]  

(A6)

where \( G = \{g_1, \ldots, g_J\} \) is the set of all local maxima and \( g^* = \max G \).

2. Entropy of the squared and normalized Teager Energy vector. Computed on the band-passed data; time-domain input. Detections are first detrended and energy normalized by dividing by their Euclidean length. For a signal composed of a single time-varying (or not) frequency and sampled at a sufficiently high rate, the Teager Energy sequence has a value at any time that approximates, to within a constant, the energy required to “generate” the signal using a mechanical spring-and-mass analogy. It is given by (Kaiser 1990).

J Neurophysiol • VOL 104 • NOVEMBER 2010 • www.jn.org
UNSUPERVISED CLASSIFICATION OF HFOs IN EPILEPSY

\[ T[n] = \begin{cases} x[n]^2 - x[n+1]x[n-1] & n = \{1, \ldots, L-2\} \\ 0 & \text{otherwise} \end{cases} \]  

(A7)

where we have imposed our own definition by necessity at the endpoints. We square the signal, making it everywhere positive, and normalize by its sum to make it a pseudoprobability mass function before computing its entropy, so that

\[ S[n] = \frac{T[n]}{\sum_{n=0}^{L-1} T[n]} \quad n = \{0, \ldots, L-1\} \]  

(A8)

\[ f_i = -\sum_{n=0}^{L-1} S[n] \log_2 S[n] \]  

(A9)

3. Wavelet packet-based energy feature. Computed on the band-passed data; time-domain input. The signal is first made even length, if necessary, by appending a single sample equal in value to the last sample (sample L-1). It is then made length-512 by either truncation or periodized extension on both sides (Strang and Nguyen 1996) and energy-normalized by dividing by its Euclidean length. A four-level wavelet packet decomposition using the wavelet packet system generated from the Daubechies \( \phi_{2N} \) scaling function (Wickerhauser 1994) is performed. Finally, for each of the 31 nodes in the binary decomposition tree (i.e., each sequence of coefficients that is the output of high- or low-pass filtering followed by downsampling of the signal at the parent node), we compute the ratio of the node’s total energy to the cumulative energy of all nodes at the same level. An energy vector \( \mathbf{E} \) is then formed by concatenating these node relative energies, traversing the tree from top to bottom and left to right. The feature returned, \( f_7 \), is the projection of this energy vector along a predetermined direction in 31-dimensional space, \( \mathbf{V}_1 \), found via PCA on a training set comprised of a subset of detections from two patients with putative artifacts labeled. An energy vector was formed for each detection using the method described earlier; the first principal component \( \mathbf{V}_1 \) was then found and stored for later use in computing \( f_7 \)

\[ f_7 = \mathbf{E}^T \mathbf{V}_1 \]  

(A10)

Similar metrics have been used in developing features for the classification of acoustic transients (Learned and Willsky 1995). To ease the computational burden in the clustering stage, after forming the seven-dimensional feature vectors \( \{\mathbf{F}_1, \ldots, \mathbf{F}_I\} \), we form \( \{\mathbf{F}_1, \ldots, \mathbf{F}_I\} \) by projecting onto the first four principal components. This set of four-dimensional feature vectors is the input to the clustering algorithm described in the main text.

APPENDIX B

Update equations for the Expectation-Maximization algorithm

E-STEP:

\[ \gamma^{(n)}(z_{nk}) = \frac{\pi^{(n)}(z_{nk})}{\sum_{j=1}^{K} \pi^{(n)}(z_{nk})} \]  

(B1)

M-STEP:

\[ u^{(t+1)} = \frac{1}{N} \sum_{n=1}^{N} \gamma^{(t+1)}(z_{nk}) \mathbf{x}_n \]  

(B2)

\[ \sum_{nk} \gamma^{(t+1)}(z_{nk}) = 1 \]  

(B3)

\[ \pi^{(t+1)}_k = \frac{N_k}{N} \]  

(B4)

where

\[ N_k = \sum_{n=1}^{N} \gamma^{(n)}(z_{nk}) \]  

(B5)

where \( z \) is a \( K \)-dimensional random variable (called a “latent” variable because it is not explicitly observed) having \( K \) possible states, such that whenever a particular element is equal to 1, all others are equal to zero. Its marginal distribution is defined in terms of the mixture component weights, such that \( p(z_k = 1) = \pi_k \). It is introduced to achieve a formulation of the likelihood function that is amenable to maximization using EM. \( \gamma(z_{nk}) \) is then defined as the posterior probability that \( z_k = 1 \), given that data segment \( \mathbf{x}_n \) has been observed; that is, the probability that mixture component \( k \) is “responsible” for generating observation \( \mathbf{x}_n \). The EM algorithm allows us to learn the parameters of the model from the data. When it converges, it is guaranteed to converge to (at least) a local maximum of the likelihood function.\footnote{We take the algorithm to have converged when successive computed values of the likelihood function differ by less than or equal to \( 1 \times 10^{-5} \). In cases where numerical issues arise or where convergence is not achieved within a prespecified number of iterations (500)—either due to encountering singularities of the likelihood function or to otherwise slow rates of convergence—the algorithm returns an “indeterminate” flag, and the HFO candidate in question is passed to the third stage of processing without interruption of the program flow.}

K-means algorithm

1. Initialize \( \mathbf{u} \) randomly.\footnote{We use a randomly chosen \( k \) observations from the data set \( \mathbf{X} \).}

2. For each \( \mathbf{x}_n \), compute the nearest prototype

\[ \text{set } Y_{ki} = \begin{cases} 1 & \text{if } k = \arg \min \|\mathbf{u}_j - \mathbf{x}_n\|^2 \\ 0 & \text{otherwise} \end{cases} \]  

(B6)

3. Recompute cluster prototypes

\[ \text{set } \mathbf{u}_k = \frac{\sum \mathbf{Y}_{ki} \mathbf{x}_i}{\sum \mathbf{Y}_{ki}} \]  

(B7)

4. Return to step 2. Repeat until convergence. Because step 2 minimizes the error function with respect to \( \mathbf{Y} \) and step 3 minimizes the error function with respect to \( \mathbf{u} \), \( e \) is guaranteed to decrease with each iteration so that convergence, at least to a local minimum, is ensured.

Rand’s C-statistic

\[ C = \frac{N}{2} - \frac{1}{2} \left( \sum \left( \sum_{i=1}^{K} n_{ik} \right)^2 + \sum \left( \sum_{i=1}^{K} n_{ik} \right)^2 \right) - \sum \sum n_{ik} \]  

(B8)

where \( K \) is the number of clusters in the first partitioning, \( K' \) is the number of clusters in the second partitioning, and \( n_{ik} \) is the number of observations that were simultaneously in cluster \( k \) in the first partitioning and cluster \( k' \) in the second partitioning.

ACKNOWLEDGMENTS

We thank D. Maus for computing support and S. Isard and D. Maus for valuable technical discussions.

We take the algorithm to have converged when successive computed values of the likelihood function differ by less than or equal to \( 1 \times 10^{-5} \). In cases where numerical issues arise or where convergence is not achieved within a prespecified number of iterations (500)—either due to encountering singularities of the likelihood function or to otherwise slow rates of convergence—the algorithm returns an “indeterminate” flag, and the HFO candidate in question is passed to the third stage of processing without interruption of the program flow.

We use a randomly chosen \( k \) observations from the data set \( \mathbf{X} \).
GRANTS

This work was supported by grants from the National Institute of Neurological Disorders and Stroke Grants R01 NS-041811 and R01 NS-48598 to B. Litt and R01 NS-630391 to G. A. Worrell; an Epilepsy Foundation grant to J. A. Blanco; and a Klingenstein Foundation grant to B. Litt and National Institute of Health Grant K08 NS-52232 and Mayo Foundation Research Early Career Development Award for Clinician Scientists to K. H. Lee.

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the author(s).

REFERENCES


Pearson K. On lines and planes of closest fit to systems of points in space. Philosophical Magazine 2: 559–572, 1901.


