A Simple Statistical Method for the Automatic Detection of Ripples in Human Intracranial EEG

K. Charupanit, B. A. Lopour*

Department of Biomedical Engineering, University of California, Irvine, CA 92697 USA

* Corresponding author. Address: Department of Biomedical Engineering, Henry Samueli School of Engineering, 3120 Natural Sciences II, University of California, Irvine, CA 92697-2715.
E-mail address:beth.lopour@uci.edu

Highlights

- New automatic HFO detection algorithm requires optimization of only one tunable parameter that is related to the number of allowable false positive events.
- The detector uses an iterative estimate of the background amplitude distribution to choose the threshold.
- Cross-validation tests indicate that sensitivity is superior to a commonly used automated detector based on RMS amplitude.

Abstract

High frequency oscillations (HFOs) are a promising biomarker of epileptic tissue, but detection of these electrographic events remains a challenge. Automatic detectors show encouraging results, but they typically require optimization of multiple parameters, which is a barrier to good performance and broad applicability. We therefore propose a new automatic HFO detection algorithm, focusing on simplicity and ease of implementation. It requires tuning of only an amplitude threshold, which can be determined by an iterative process or directly calculated from statistics of the rectified filtered data (i.e. mean plus standard deviation). The iterative approach uses an estimate of the amplitude probability distribution of the background activity to calculate the optimum threshold for identification of transient high amplitude events. We tested both the iterative and non-iterative approaches using a dataset of visually marked HFOs, and we compared the performance to a commonly used detector based on the root-mean-square. When the threshold was optimized for individual channels via ROC curve, all three methods were comparable. The iterative detector achieved a sensitivity of 99.6%, false positive rate (FPR) of 1.1%, and false detection rate (FDR) of 37.3%. However, in an eight-fold cross-validation test, the iterative method had better sensitivity than the other two methods (80.0% compared to 64.4% and 65.8%), with FPR and FDR of 1.3%, and 49.4%, respectively. The simplicity of this algorithm, with only a single parameter, will enable consistent application of automatic detection across research centers and recording modalities, and it may therefore be a powerful tool for the assessment and localization of epileptic activity.

1. Introduction

High frequency oscillations (HFOs) are a promising biomarker for the delineation of the seizure-onset zone in epileptic patients (Bragin et al. 2010; Zijlmans et al. 2012), and the removal of brain regions exhibiting HFOs has been correlated with a greater likelihood of seizure-free surgical outcome (Jacobs et al. 2010; Fujiwara et al. 2012). Pathological HFOs are broadly defined as spontaneous electrographic events of at least three to four full oscillations that are clearly distinguishable from the background signal, with frequencies ranging from 80-500 Hz (Buzsáki et al. 1992; Jacobs et al. 2012). Visual identification by expert reviewers is a widely used procedure for detecting HFOs in both scalp and intracranial EEG recordings (Jacobs et al. 2014; Ferrari-Marinho et al. 2015), and it is currently considered the gold standard (Worrell et al. 2012). This type of manual detection affords great adaptability to recordings with different baseline levels, and rejection of artifacts can be accomplished simultaneously. However, the results are subjective; Zelman et al. (2012) reported 60.98% agreement between two trained, human reviewers. The process is also highly time-consuming, as visual identification of HFOs in 2 minutes of EEG recordings with 8 channels can take about two hours. For these reasons, visual analysis is infeasible for longer studies. To overcome these disadvantages, there has been a push to develop automatic detection algorithms to improve the speed and accuracy of detection across different recordings and clinical settings, with the long-term goal of eventually replacing traditional visual identification.

HFO detection typically consists of two steps: (1) initial detection of candidate events and (2) rejection of artifacts and other false positives. A majority of published detectors focus primarily on the first step of this process and rely on human visual validation for the rejection of artifacts (Staba et al. 2002; Gardner et al. 2007; Crépon et al. 2010) or accept the results of detection without post-processing (Zelmann et al. 2010; Dümpelmann et al. 2012; Chaibi et al. 2013). To detect high frequency events that stand out from the background, these algorithms filter the raw data in the frequency band of interest and then estimate the energy of the signal using various techniques, including the RMS amplitude (Staba et al. 2002), line length (Gardner et al. 2007), Hilbert transform (Crépon et al. 2010), Hilbert Huang transform (Chaibi et al. 2013), or a combination of techniques (Zelmann et al. 2010; Dümpelmann et al. 2012). Then a threshold is applied, along with other parameters such as minimum duration, minimum number of oscillations, or minimum time between successive events. Increasing the complexity of the algorithm can improve accuracy, but it also makes implementation more difficult. Most algorithms require a complex optimization of three to four parameters, and this optimization step is critical for achieving accurate performance (Zelmann et al. 2012).

More recently, several algorithms for automated artifact rejection have been developed (Burnos et al. 2014; Cho et al. 2014; Amiri et al. 2016; Gliske et al. 2016). These algorithms are independent from the initial detection, which is accomplished through either human visual analysis (Amiri et al. 2016) or an existing detector (Burnos et al. 2014; Cho et al. 2014; Gliske et al. 2016). These methods reduce the number of false detections, but again increase the complexity and number of parameters, making implementation more challenging.

Overall, both detection and artifact rejection algorithms tend to be optimized for the recordings of a specific research group, as they are highly sensitive to characteristics of the data, including the source and prevalence of artifacts. This makes it difficult to directly compare results of different studies and remains a barrier to the translation of automated HFO detection to the clinical setting. We hypothesize that simplifying the detection procedure will aid in overcoming some of these challenges.

Therefore, we propose a new algorithm for automatic detection of HFOs in human intracranial EEG (iEEG) data focusing on simplicity and ease of implementation. This method requires optimization of only one parameter, the amplitude threshold, which is directly related to the sensitivity and specificity of the detection. We propose two approaches for selection of the threshold: an iterative method and a non-iterative method. The iterative algorithm stems from the idea of distinguishing transient burst-like events from background noise, as is done in single-molecule fluorescence experiments (Grange et al. 2008). In those experiments, the fluorescence intensity is analyzed, which is always a positive quantity; therefore, we adapt the rectified filtered iEEG signal to the algorithm. Based on the amplitude of each oscillation in the filtered data, an estimate of the probability distribution of the background activity is obtained via an iterative process and is used to calculate an optimal amplitude threshold. In the non-iterative method, the threshold is calculated directly from the mean and standard deviation of the filtered, rectified signal. We directly compared the performance of our algorithm to the Root Mean Square (RMS) detector (Staba et al. 2002) by testing them on the same dataset of visually identified HFO events from Zelmann et al. 2012. Because our algorithm achieves accurate performance and requires the optimization of only a single parameter, it offers advantages over alternative methods for the initial detection of candidate HFO events.

2. Methods

2.1 Electrophysiological Recordings

All recordings used in this analysis were previously published by Zelmann et al. (2012). The iEEG recordings were collected between September 2004 and April 2008 from 45 patients with medically refractory epilepsy who underwent depth macro-electrode implantation at the Montreal Neurological Hospital. Recordings were sampled at 2 kHz and low-pass filtered at 500 Hz. The dataset contained 1-minute segments of slow wave sleep from 19 patients, each with 10-39 channels. As in the original study, channels with continuous artifacts, channels containing zero visually-detected HFOs, and channels without distinguishable baseline were excluded, leaving 373 channels for analysis.

2.2 Channel selection and event identification

HFOs were visually identified by two experienced reviewers at the Montreal Neurological Institute. In addition to HFOs, reviewers marked baseline segments of data with no oscillatory activity (no possible HFOs) with a maximum length of 400 ms (NegBASE). If baseline segments exceeded the maximum duration, they were split into multiple 200 ms-long events. For this study, automatic detection was performed on the ripple band (80-250 Hz) because there were very few visually marked fast ripples (250-500 Hz).

Two validation schemes introduced by Zelmann et al. (2012), referred to as "strict" and "open" validation were used to evaluate the performance of the automatic detectors. The strict validation was based on HFOs (PosAND) and baseline segments (NegBASE) jointly marked by both reviewers, representing positives and negatives respectively. In the open validation scheme, events marked by either one or both reviewers (PosANY) were considered, since events marked by only one reviewer are possible HFOs, but with lower probability. In this scheme, the negative baseline was defined as any iEEG segment with a minimum of 25 ms separation from any PosANY event. Sensitivity and FPR were evaluated based on strict validation; on the contrary, false detection rate (FDR) employed open validation.

We excluded four channels because they contained only fast ripples (250-500 Hz) with no ripples, leaving us with 369 channels for analysis. Within these channels, 20% (73 channels) were randomly selected to serve as training channels to

optimize the parameters of the RMS detector. The remaining 296 channels were used to evaluate the performance of both automatic detectors, which contained 7561 PosAND HFOs, 12316 PosANY HFOs, and 43101 NegBASE segments that were visually identified.

2.3 Automatic detection algorithm

The detection procedure starts with a band-pass filter (80-250 Hz; finite impulse response filter, fstop1 = 70 Hz; fpass1 = 80 Hz; fpass2 = 250 Hz; fstop2 = 260 Hz; stopband attenuation = -60 dB) applied to the iEEG recordings. The signals are filtered forward and backward to obtain zero-phase distortion. The filtered signals are then rectified, and the amplitude of each oscillatory cycle is measured by identifying each "peak" (local maximum) in the rectified data (Figure 1). Our algorithm then consists of choosing a threshold for the local maxima and identifying events that have a minimum number of oscillations above that threshold. More specifically, we look for a number of rectified peaks that exceed the threshold within a window of time, e.g. for 6 consecutive peaks, at least 5 are above the threshold.

Here, we propose two methods for choosing the threshold: (1) calculated based on the mean and standard deviation of the peak amplitudes and (2) determined via an iterative process to estimate the amplitude distribution of the background activity. We will refer to these as the "non-iterative" and "iterative" methods, respectively. In the non-iterative method, the threshold is calculated directly from the peak amplitudes (local maxima) of the rectified filtered signal, defined as the mean plus a number of standard deviations. In the iterative process for threshold selection, we first create a histogram of the peak amplitudes (Figure 1E). We model the background activity of the channel with a gamma distribution, f(x), which can be estimated as

$$f(x) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-\frac{x}{\theta}} , \qquad x \in (0, \infty)$$

where k and θ are the shape and scale parameters of the gamma probability distribution, respectively. In channels with HFOs, the presence of the high amplitude activity will cause an observable long tail at the upper end of this distribution, superimposed on the background activity. In the first iterative step, k and θ are estimated using the height of all peaks (via MATLAB function "gamfit"), which will include both background activity and HFOs. The shape and scale parameters are used to construct an estimated probability distribution f(x) to represent the amplitude of background activity (Figure 2A). We then define a cutoff of $F(x) > 1 - \alpha$, where F(x) is the cumulative probability distribution function of f(x) (Figure 2B). Peaks with amplitude above the cutoff are excluded from the distribution and then k and θ are recalculated for the new improved estimate of the background activity, f(x). Note that the cutoff value is tuned by a single parameter, α, related directly to the number of peaks that are removed in each step of the process; for example, if α = 0.01, peaks that fall within the top 1% of the estimated f(x) will be removed during each iteration. The threshold based on alpha is determined relative to the amplitude of the signal, rather than being an absolute threshold in terms of microvolts. For example, with α = 0.01, a signal with a mean amplitude of 100 µV will have a higher threshold than a signal with a mean amplitude of 50 µV. The process is then repeated with the new estimate of f(x) until the values of k and θ converge and no additional peaks are removed with each iteration. At this point, f(x) is the estimate of the underlying background distribution given a tolerance for false positives (α). We then define the threshold for detection based on the cutoff value of the final iteration (Figure 2B). The threshold will depend on the magnitude of α, with higher values of α leading to a lower threshold. In both approaches, the user must only provide one parameter to set the threshold, either the number of standard deviations above the mean or the tolerance for false positive

detections, respectively. After threshold selection, all peaks with amplitudes above the threshold are marked, and events with at least 5 out of 6 consecutive peaks above the threshold are defined as HFOs.

We tested the detector performance as a function of the threshold (Section 2.5), and we also measured the robustness of the algorithm to changes in the shape of the amplitude probability distribution, the number of iterations for threshold convergence, and the required number of consecutive peaks above threshold (see Section 3.3).

2.4 RMS detector

The RMS detector (Staba et al. 2002) is based on the energy of the moving average of the root mean square amplitude of the filtered signal. In the original publication, the signal was bandpass filtered from 100-500 Hz, and segments of data in which the RMS value exceeded a threshold for at least 6 ms were marked as pre-qualified events. The threshold was defined as five standard deviations above the mean RMS value. Events less than 10 ms apart were joined together and considered as a single event. Finally, to be qualified as HFOs, the events were required to have at least six rectified peaks above a second threshold, which was three standard deviations above the mean of the rectified filtered signal. Originally, the detector was designed to identify HFOs in hippocampus and entorhinal cortex microwire recordings in humans (Staba et al. 2002), and it was used in the studies of microelectrode recordings in temporal regions (Staba et al. 2004; Staba et al. 2007). The RMS detector is one of the most widely used automatic HFO detectors in published studies (Gardner et al. 2007; Blanco et al. 2010; Zelmann et al. 2012; Gliske et al. 2016).

Before applying the RMS detector to our dataset, we used the training data to optimize all parameters, in order to ensure the best possible performance. We implemented the same filter for both detectors (see Section 2.3), while the following parameters required optimization: duration of moving window to calculate RMS amplitude, minimum event duration, RMS threshold (1st threshold), rectified threshold (2nd threshold), and minimum time gap between consecutive candidate events. All parameters except the RMS threshold, which was selected from the ROC curve, were optimized via the training channels (20% of all channels). We tested 4-10 ms for RMS window size; 4-24 ms for minimum event duration; 1-3.5 times the standard deviation above the mean for the rectified threshold; and 4-13 ms for minimum time gap between consecutive events. The HFO detection was performed in the training channels, and a single ROC curve was constructed for each parameter combination by varying the number of standard deviations for the RMS threshold. The parameter set that gave the best performance at the optimum RMS threshold (the point on the ROC curve that was closest to the upper left corner of the plot) was chosen as default: 10 ms for RMS window size, 6 ms for minimum event duration, one standard deviation for rectified peak threshold, and 13 ms minimum time gap between consecutive events. Then the RMS detector with the optimized parameters was applied to the 296 channels reserved for testing and the RMS threshold was optimized using an ROC curve in the same way as α for our detector (see Section 2.5). Note that these parameters were different from the original study of Staba et al. 2002.

2.5 Evaluation of detector performance
Performance measured from ROC curve

The performances of both automatic detectors were evaluated via an ROC curve, which is plotted using sensitivity and FPR under the strict validation scheme. To calculate sensitivity and FPR, true positives (TP) were defined as detected

events that matched PosAND. False positives (FP) were automatically detected events that corresponded to a visually marked baseline. True negatives (TN) referred to the NegBASE segments containing no automatically-detected HFOs. False negatives (FN) were undetected PosAND. Then sensitivity was defined as TP/(TP+FN), and FPR was defined as $1 -$ TN/(FP+TN). The ROC curves were created by varying α for our detector and varying the first RMS threshold (specifically the number of standard deviations) for the RMS detector. The optimum values of α and the threshold were chosen by identifying the best balance between sensitivity and FPR, the point on the ROC curve that was closest to the upper left corner of the plot. The performance at this optimum value illustrates the best possible performance of the detector. In addition, the area under the ROC curve (AUC), which ranges from zero to one (one indicating perfect performance), is used to represent the detector's overall performance.

Note that FPR reflects only the falsely detected events within the visually-marked baseline segments; therefore, this measure does not account for the events that are neither HFOs (PosANY) nor baseline (NegBASE), which are likely to be moderate amplitude events. Accordingly, the false detection rate (FDR) was introduced as the number of detected events that did not overlap with PosANY events, including false positives, divided by the total number of detected events.

Optimization of α or the threshold was performed in three different ways: (1) across channels, with α or the threshold optimized for each individual channel, representing the best possible outcome of the detector, (2) across patients with a single value for all channels within each patient, illustrating the robustness of the detector when applied to different patients, and (3) grouping all events together, using a single value for all channels.

Performance measured from cross-validation

Cross-validation was used as a more stringent evaluation of the detection performance. Ideally, the detector would be optimized for each channel based on visual markings from a short amount of data within that particular channel, at least one minute free of artifact, and those settings would be used for the remaining data. However, this was not possible here because the dataset contained only one minute of data for each channel. We therefore measure the performance of the detector using independent subsets of channels for training (optimizing) and testing. This implementation method would only be used if visual detection is not available for all channels. It illustrates the performance of the detector when a single parameter setting is applied to new data, without taking the characteristics of the new signals into account.

We apply an eight-fold cross-validation technique, where 296 channels are randomly assigned to eight subsample groups with 37 channels each. For each iteration, one subsample group is selected for testing and the seven remaining groups are used for optimizing. The average optimum α value from all seven training groups (derived from ROC curves) is used to perform automatic detection in the remaining testing group. This procedure is repeated eight times, so each subsample group is tested once.

3. Results

3.1 Detector performance

Comparing the performance across channels

The ROC curves for each channel were created using a range of α values or the number of standard deviations to determine the amplitude threshold. The optimum threshold for each channel (the threshold on the ROC curve closest to the upper left-hand corner, representing the best performance) was selected, and then the associated performance was vertically

averaged across all 296 channels (Figure 3). Note that this performance, in which the threshold was optimized for each channel, represents the best possible detection performance for the detector. For our detector with iterative threshold selection, the averaged AUC was 0.995. The average sensitivity was 99.6%, and the FPR and FDR were 1.1% and 37.3%, respectively (Table 1). For the non-iterative method, the sensitivity, FPR, FDR and AUC were 99.5%, 1.1%, 36.9%, and 0.995, respectively (Table 2). The RMS detector had 98.9% sensitivity, 0.7% FPR, and 34.6% FDR, with an area under the ROC curve of 0.995 (Table 3). The difference in detection performance was tested with a Wilcoxon signed rank test. The test failed to reject the null hypothesis when comparing the two different thresholding schemes for our detector. This was expected, as the only difference was how the threshold was selected. However, the sensitivity and AUC of our detector, with the iterative method, were statistically significantly higher than the RMS detector, at $p<0.01$ and $p<0.05$, respectively, but the FDR was significantly higher, $p<0.05$. Our detector, with the iterative method, exhibited error-free performance in a large number of channels. In 257 of 296 total channels (86.8%), all visually marked PosAND events were identified at the optimum threshold (100% sensitivity), and 187 of them (63.2% of all channels) had all PosAND events detected with zero FPR, representing perfect performance (AUC =1). Only 9 of the 296 channels (3.0%) had sensitivity below 95%.

The optimum RMS threshold ranged from -1.7 to 8.5 (mean 2.15) standard deviations above the mean RMS value. Similarly, in our detector with the non-iterative approach, the standard deviation of the rectified peak amplitude ranged -0.5 to 10.0 (mean 2.41). The negative values of standard deviation for the optimum threshold need to be interpreted with caution because this indicates that the optimum threshold was lower than the mean value for the signal.

Comparing the performance across patients

We then assessed the performance of each detector when a single value of α or number of standard deviations was used for all channels within a patient's dataset. This illustrates the robustness of the detector when it is applied to channels with different amplitudes and background characteristics. All ROC curves within each patient were averaged together with equal weighting, and the average AUC, sensitivity, FPR, and FDR were obtained for each of the 19 subjects (Figure 4). Here, a single α value or number of standard deviations cannot optimally represent all channels due to the variation in background activity in each channel. In channels with an active background, the detector requires a higher α (lower threshold) to capture more peaks because there are random peaks from background activity mixed in with the detected peaks of HFOs, while HFOs can be detected with a lower α (higher threshold) in quiet channels. Therefore, the performance of our detector with iterative threshold selection decreased (0.980 AUC, 93.6% sensitivity, 5.2% FPR, and 66.9% FDR; see Table 1) compared to the case where it was optimized for every single channel. A similar trend can be observed for both the non-iterative method (0.976 AUC, 92.7% sensitivity, 6.0% FPR, and 66.6% FDR; see Table 2) and the RMS detector (0.973 AUC, 92.2% sensitivity, 5.4% FPR, and 64.0 % FDR; see Table 3).

Comparing the performance across all events

Both detectors were tested when a single α value or number of standard deviations was used across all 296 channels. Again, all channels were weighted equally. As in the previous section, we expected a decrease in detection performance because the level of background activity varies across channels and subjects. For our detector with iterative threshold selection, the optimum value of α was 0.065. At this parameter setting, it identified 75.3% of PosAND events (5696 out of 7561) and 65.4% of PosANY events (8051 out of 12316); further, 95.1% of baseline segments (NegBASE, 40985 out of 43101) were free of detected HFO events. The performance indicates that reliable detection was achieved by using one single

value of α to iteratively select a threshold in all channels, with an average sensitivity of 93.1%, FPR of 6.4%, FDR of 71.6%, and AUC of 0.981. Similar outcomes were found for our detector with the non-iterative approach (92.3%, FPR of 6.1%, FDR of 68.0%, and AUC of 0.979 at the optimum threshold, which was 0.80 standard deviations above the mean of rectified peak amplitude). The RMS detector also illustrated comparable accuracy, with an average sensitivity of 92.0%, FPR of 7.0%, FDR of 66.2% and AUC of 0.976 at a threshold equal to the mean plus 0.63 standard deviations of RMS amplitude. The ROC curves for both detectors are shown in Figure 5.

Cross-validation of the automated detection algorithm

In the previous sections, optimization of the threshold was based on the prior knowledge of all visually detected events as a "true" reference via ROC curve. However, in practice, it may not be realistic to complete visual detection on all channels, especially in the case of intracranial recordings, which often have over one hundred channels. Cross-validation techniques serve as a measure of this worst-case scenario, in which the detector is optimized on one set of channels and the parameters are applied to a different set of channels. This is the most stringent test of performance, as each channel's background characteristics and HFO amplitude are independent from those of the other channels. We implemented eight-fold cross-validation, in which seven subgroups of channels were used to optimize the threshold and subsequently tested on the remaining channels. Note that this is not typically how the detector would be implemented. In general, it is best to use a portion of the data within each particular channel to optimize the parameters for that channel, e.g. use the first minute of a ten-minute recording.

For our detector with iterative threshold selection, the optimum α value of the eight cross-validation iterations ranged from 0.036 to 0.038. The average sensitivity, FPR, and FDR were 80.0%, 1.3%, and 49.4%, respectively. In the case of our detector paired with non-iterative thresholding, the performance decreased to 64.4%, 0.4%, and 31.2% for sensitivity, FPR, and FDR respectively. The optimum number of standard deviations for the threshold ranged from 2.35 to 2.5. Similarly, the RMS detector had 65.8% sensitivity, 0.5% FPR, and 33.4% FDR. The optimal thresholds for the RMS detector ranged from 2.10 to 2.20 standard deviations above the mean RMS. There was large decrease in sensitivity for both detectors when the threshold was determined using the amplitude of the signal (mean plus number of standard deviations). These results indicate that the iterative method for threshold selection is more robust than directly using the statistics of the amplitude because it adapts to the characteristics of each channel.

3.2 Parameter optimization to reduce the false detection rate

When optimizing α individually for each channel, our detector demonstrated very high sensitivity, but the FDR was 37.3%. False detection of events remains a barrier to the practical implementation of automatic algorithms. There are several means to reduce the FDR, e.g. applying post-processing steps (Burnos et al. 2014; Cho et al. 2014; Amiri et al. 2016; Gliske et al. 2016) or using human validation (Staba et al. 2002; Gardner et al. 2007; Crépon et al. 2010). Here we propose another approach, in which α is optimized based on FDR instead of FPR.

More specifically, we optimized the threshold using the precision and recall curve, which is the plot of sensitivity and 1-FDR, rather than the ROC curve. Here, the optimum threshold was defined as the point on the curve that was closest to the upper right corner of the plot. Our detector with iterative threshold selection achieved a sensitivity of 89.1% (originally 99.6%), but the FDR was reduced by more than half to 16.5% (from 37.3%; see also Table 1). There were only 20 channels (6.8%) in which the FDR exceeded 50% and only 2 channels (0.7%) with sensitivity less than 50%. Both the non-iterative

method and RMS detector performed similarly; see also Table 2 and 3. Note that, as with all other results presented here, the only parameter involved in the optimization of our detector was α or the number of standard deviations. Therefore, adjusting this single parameter enables us to choose whether we prioritize high sensitivity or high specificity.

3.3 Algorithm design

In addition to the parameter α, the design of the automatic detection algorithm involved several important elements. For example, we also chose a probability distribution to model the amplitude of the background activity, the maximum number of iterations for convergence of the threshold, and the number of consecutive oscillations required to define an HFO event. However, we found that changes to these elements did not have a significant effect on the detection performance, as demonstrated below.

Probability distribution for the background activity

A key step in the detection algorithm is the estimation of the amplitude probability distribution for the background activity. To estimate the true distribution with minimal influence from HFOs, the 33 channels with the fewest visually detected HFOs (i.e. PosANY was less than three events) were used as representative of background activity. We compared gamma distributions to the amplitude histograms of the rectified band-pass filtered iEEG recordings. A Kolmogorov-Smirnov test showed that 31 channels were consistent with a gamma distribution ($p < 0.05$). Furthermore, a Q-Q plot showed that the peak amplitude distribution was a good match to the reference distribution at low quantiles (data not shown). At high peak amplitudes, however, the sample distributions deviated from the reference. Therefore, both the Kolmogorov-Smirnov test and the Q-Q plot confirmed our hypothesis that without HFOs, the gamma distribution would be a good model for the peak amplitude distribution of the background activity. We performed the same tests with a zero-truncated normal distribution to determine whether it could be used as a model of the probability distribution of the peak amplitude. Interestingly, the two distributions gave comparable performance, which suggests that the detection outcome is robust to a change of the model distribution. For all results presented here, we selected a gamma distribution as the default model for the background activity.

Number of consecutive peaks above threshold

For the purposes of detection, the definition of an HFO may vary slightly depending on the automatic detector algorithm and the study. Generally, an HFO is required to have 3-5 consecutive cycles above a threshold (Staba et al. 2002; Jacobs et al. 2008). Here we used the number of local maxima ("peaks") in the rectified filtered iEEG recordings to count the oscillations, with one oscillation consisting of two consecutive rectified peaks. We tested performance while varying this parameter from 4 to 8 consecutive peaks. When a high number of consecutive peaks was required, the optimum α value was increased (indicating a lower threshold value), enabling events with more consecutive peaks to be detected. Our results indicated that detector performance was robust to changes in the required number of oscillations; because the threshold was recalculated each time, α increased as a function of the number of consecutive peaks. As shown in Figure 6, the difference between each condition was primarily in the FDR value (more than 3% difference between the top three conditions with highest sensitivity, 42.5%, 39.3%, and 40.3%), while the sensitivity difference among the best conditions was less than 0.5% (99.7%, 99.4%, and 99.2% in the 4 out of 5, 5 out of 6, and 6 out of 7 conditions, respectively). We chose 5 out of 6 consecutive peaks within the test window, i.e. for a group of 6 consecutive peaks, at least 5 needed to exceed the threshold, as a default setting for our detector because it offered the best balance of sensitivity and FDR at the optimum threshold.

Number of iterations

Our algorithm employed an iterative process together with the estimate of the peak amplitude distribution, which enabled us to optimally adjust the threshold based on the selected α. After a number of iterations, e.g. up to 10-11 iterations in channels with very high α value, the estimate of the peak amplitude distribution converged and the stable cutoff value was used as the detection threshold. Once this occurred, additional iterations of the algorithm did not affect the final estimate of the distribution. The rate of convergence varied and depended on the magnitude of α and the peak amplitude probability distribution. In our study, 15 iterations were sufficient for all channels while still minimizing the computational cost. With our desktop system (CPU: intel i7-4790k, 16 GB of RAM), one minute of iEEG with 20 channels recorded at a 2 kHz sampling rate required approximately 1-2 seconds for filtering and rectification; after that, the detection procedure took an additional 2-3 seconds per channel for a single value of α. To calculate the ROC curves, 40-60 different values of α were tested, which took 4-6 minutes for each channel.

3.4 Optimum alpha values

Alpha (α) was the only parameter in our algorithm that had a substantial impact on the accuracy of HFO detection. Since α is related to the threshold, it affects the number of detected events. Lowering α, which raises the threshold, reduces the detection sensitivity together with the FPR and FDR. In this case, the detected events have a higher probability of being real HFOs, but overall fewer HFOs are detected. On the contrary, increasing α, which decreases the threshold, allows more falsely detected events which results in an increase in sensitivity; however, the FDR and FPR also increase as more peaks are detected.

In section 3.1, we showed that the detection performance was best when α was individually optimized for each channel, rather than using a single value of α for multiple channels. Recall that alpha determines the percentage of peaks in each iteration that are marked by the detector as being possibly associated with an HFO. Therefore, in a channel with a quiet background and prominent HFOs, we can select very few peaks (high threshold, low alpha) and still detect all HFOs. On the other hand, when there is an active background, the peaks associated with HFOs are mixed with "noisy" peaks, which may be approximately the same height. In this case, the detector must mark many more peaks as possibly associated with HFOs (high alpha) in order to detect the events. Across all channels, the optimum alpha values ranged from 0.0001 to 0.135. Most channels had α values on the order of 10^{-3} when we optimized the detector with an ROC curve (Figure 7A). When α was optimized using a precision and recall curve, its value tended to be lower, which increased the threshold and reduced the number of falsely detected events, (Figure 7B); however, there were some channels in which α was significantly higher. Overall, detection performance in those channels was poor, but these channels tended to have a very few or only a single visually marked HFO.

4. Discussion

Here we have presented a simple automatic detection algorithm for HFOs that consists of filtering and rectifying the data, then counting consecutive peaks above a threshold. An optimum threshold can be determined using an iterative process to estimate the statistical properties of the background activity, or it can be based on the mean and standard deviation of the peak amplitudes. The algorithm requires optimization of only a single parameter, greatly simplifying the implementation of automatic detection. Importantly, we have shown that the performance of this detector is comparable to or better than the most commonly used published algorithm, which makes it a suitable technique for the initial detection of candidate HFOs. It could also be paired with an algorithm for the automatic rejection of artifacts and false positive detections; because these post-processing steps increase the number of parameters and overall complexity, it is advantageous to use the simplest possible approach for the initial detection of candidate events.

The iterative and non-iterative methods used in our detection algorithm gave comparable results in most cases, but sensitivity using iterative threshold selection was superior in a cross-validation test. When the threshold was optimized for each individual channel, both approaches performed similarly: average sensitivity, FPR, and FDR were approximately 99%, 1%, and 37%, respectively. This result was expected, as the only difference was the method of threshold selection; therefore, the ROC curves of both cases were theoretically the same. A slight difference in the observed performance between the two approaches, which was not statistically significant, resulted from differences in the smoothness of the ROC curve. Using a fixed threshold for all channels caused a decrease in detection performance, due to the variation between channels. However, our detector still provided satisfactory results with approximately 93% sensitivity, 6% FPR, and 70% FDR for both the iterative and non-iterative approaches. The biggest difference between the two approaches occurred for the cross-validation test. The iterative method yielded better sensitivity (80.0% compared to 64.4% for our detector with the non-iterative method). While the non-iterative method resulted in lower FPR and FDR, the sensitivity is the most crucial factor, as false detections can be reduced by post-processing procedures. Therefore, when optimizing channels independently, the iterative method is more robust to variation across channels compared to the non-iterative method in which the threshold is directly calculated from the mean and standard deviation of the amplitude.

The two thresholding schemes each have strengths and weaknesses. The non-iterative method is the simplest method we tested, and it is also the least computationally expensive. However, our results showed that there is no intuitive way to choose the number of standard deviations for the threshold. When the threshold was individually optimized for each channel, the optimum number of standard deviations ranged from -0.5 to 10.0 (mean 2.41). Thirteen of 296 channels had a negative number of standard deviations as the optimum threshold, resulting in a threshold that was below the mean. In those channels, the overall amplitude of the signal was significantly higher than other channels, causing the mean and standard deviation to be much larger than usual. The amplitude distribution was non-normal and highly skewed; therefore, the mean and standard deviation were not appropriate metrics for the selection of a threshold. Overall, the wide range and possibility of needing negative values presents an obstacle for appropriate selection of a threshold. On the other hand, the iterative method is more complex and computationally intensive, but it is able to adapt to the characteristics of the data in each channel. It therefore performs better when the detector is optimized on one dataset and then applied to an independent set of data. Moreover, the optimal α values ranged from 0.0005 to 0.145 (mean 0.037); selection of this parameter is easier, as its range is narrower and its value is always positive.

The detection performance of our algorithm was either comparable or superior to the RMS detector when tested using the same iEEG data, visually-marked HFOs, and evaluation scheme. It was comparable when testing the best possible performance using an ROC curve, but the sensitivity for the iterative method was superior in the cross-validation test (80.0% for iterative method, 65.8% for RMS). This is somewhat expected because the RMS detector was initially designed for use with microwire data, and it has been shown that the characteristics of high frequency activity depend on electrode size (Worrell et al. 2008). However, the RMS detector has been successfully implemented in various recording schemes, including microwire electrodes (Staba et al. 2002; Staba et al. 2004) and macro- and microelectrodes (Zelmann et al. 2012; Gliske et al. 2016). Importantly, the results of our tests demonstrate that good performance is dependent on proper optimization of all parameters, not just the threshold. Optimization of four parameters using the training channels was necessary to achieve detection performance similar to our detector. Here, most of the optimum parameters that were used for the RMS detector were different from the original study, especially the RMS threshold. The average RMS threshold across the 296 test channels was only 0.80 standard deviations above the mean, while the original study recommended 5 standard deviations. Similar to the non-iterative method, the optimal RMS threshold had a wide range of values, including negative values of the standard deviation. If we had limited the standard deviation to positive values to ensure that the HFO amplitude was greater than the mean, the performance of the RMS detector would have been significantly worse.

Taking all of these results into consideration, we recommend the following guidelines for implementation of an automated detector. The best performance will be achieved if visual detection is first performed on a small segment of data from each channel, approximately one minute. These visual detections can be used to optimize α or the number of standard deviations for each channel, which can then be applied to detection in the remainder of the dataset. In this situation, we anticipate that our detector and the RMS detector will perform similarly, but the four parameters of the RMS detector must be optimized on an independent dataset prior to this procedure. Our detector does not require this initial optimization step. On the other hand, if visual detection is only available for a subset of channels, the parameter α can be optimized using that group of channels and then applied to the rest of the dataset. This is analogous to the cross-validation test we performed. Therefore, in this case, the best performance will be achieved with our detector, using the iterative method of threshold selection.

The high number of falsely detected events remains a challenge in the implementation of automated detection. While threshold optimization based on the ROC curve yields high sensitivity and extremely low FPR, this low FPR must be interpreted with care because it does not reflect all falsely detected events. The FPR includes only the detected events that lay within visually marked baseline segments. Detected events that occurred within unmarked sections of the signal (not clearly baseline and not an HFO) were not included in this number, but instead contributed to the FDR. The FDR for both detectors was as high as 36.1%, which means that approximately one third of detected HFO events occurred outside of the PosANY visual markings. Several techniques can be implemented to reduce this value. Here, we optimized the threshold based on a precision and recall curve, rather than using an ROC curve. When we did this for the iterative method, the average sensitivity across all channels decreased from 99% to 89%, but the FDR showed a much greater decrease, from 36% to 17%. This trade-off between sensitivity and specificity may be desirable in some applications, such as the analysis of long datasets in which visual rejection of falsely detected events is not feasible. Note that the detection performance in several channels suffered due to very few HFOs (there were 17 channels with only one visually-marked HFO), a very active background with continuous

high frequency activity, and the detection of artefactual waveforms (the sensitivity was below 50% in two channels, and the FDR exceeded 50% in 27 out of 296 channels). Another option to reduce the FDR and detection of artifacts is to apply a post-processing step to eliminate falsely detected events and leave only "true" HFOs. This can be done either automatically, using an artifact rejection algorithm (Burnos et al. 2014; Cho et al. 2014; Amiri et al. 2016; Gliske et al. 2016) or data classification via clustering (Blanco et al. 2010; Malinowska et al. 2015), or manually with supervision by experts.

There are two possible limitations to the automatic algorithm presented here. First, for the iterative threshold selection, we assumed that HFOs were rare and that the high amplitudes associated with these events would be superimposed on a stable distribution of the background amplitude. If HFOs occur frequently, the estimation of the background activity will be inaccurate and will result in a drop in performance. However, channels with this characteristic are typically quite difficult to interpret, even for human reviewers, so it is perhaps not surprising that it also provides a challenge for automatic detectors. Moreover, the dataset used for testing contained such channels, and this method still demonstrated good overall performance. Second, because the detection relies on identification of local maxima in the rectified filtered EEG signals, the accuracy may be sensitive to high amplitude events, including artifacts. On the other hand, the requirement that each event have a number of consecutive peaks above the threshold will help reduce the detection of short, sharp transients, as they usually cause only a few full oscillations in which the amplitude is high enough to exceed the threshold in the filtered data. It is important to note that the dataset we used for performance testing contained epileptiform discharges and other artifacts, and we compared the automated detection results to a gold standard of visual detection.

The iterative procedure presented here has some similarities to the one utilized in the MNI detector (Zelmann et al. 2012), however there are several important differences. First, the MNI detector incorporated two different algorithms for HFO detection, depending on whether or not the iEEG channel contained continuous high frequency activity. The iterative procedure for determining the threshold was used only in cases where an insufficient amount of baseline activity was detected. Here, we use a single algorithm for all channels. Second, the iterative procedure in the MNI detector is based on the moving average of the RMS within a window of data. In our algorithm, the iterative choice of threshold is based on detection of local maxima in the filtered, rectified data. This enabled us to detect high amplitude oscillations directly, rather than indirectly through energy measures. Third, the MNI detector requires the optimization of many parameters. The baseline detection requires selection of a time window, amount of overlap, and a threshold for the wavelet entropy; the HFO detection relies on selection of time windows for calculation of the RMS and the moving average of the RMS, a threshold for the cumulative distribution function, an energy threshold, and a minimum amount of time between events. Here, our simplified detection algorithm used only four parameters (α, the choice of background probability distribution, the max number of iterations, and the number of required consecutive oscillations), and we were able to show that the choice of α was the only one that affected the performance of the detector. While our algorithm provides levels of accuracy that are similar to other published detectors, we feel that the value of our detector lies in the simplicity of the optimization, which is based on only one parameter. This will enable it to be applied broadly to a variety of data types.

Roehri et al. (2016) also reported a technique to estimate the background activity for the purposes of HFO detection. A continuous wavelet transform was applied to the data and distributions of the real and imaginary coefficients were created at each frequency. It was then assumed that the background noise could be represented by a Gaussian fit to the central portion of those distributions. This enabled estimation of the background activity without requiring manual selection of a baseline

segment of data. The distributions for each frequency were normalized using this technique, thus whitening the data. Whitening of electrophysiological data can prevent HFO detection from being dominated by low frequencies, which naturally have higher amplitude. This procedure is related to the non-iterative approach presented here, but we create amplitude distributions using local maxima ("peaks") in the filtered data, rather than using all data points. Moreover, our results suggest that iterative estimation of the background provides a distinct advantage in choosing a threshold; therefore, combining the whitening methods suggested by Roehri et al. with iterative background estimation may be a powerful technique for HFO detection. On the other hand, if the goal is to match visually marked HFOs, whitening may not be desirable, as humans viewing bandpass filtered data will naturally be biased toward low frequency events.

There are several remaining barriers to achieving automated detection accuracy which is sufficient for clinical use. The physiological mechanism underlying HFOs is not fully understood, and therefore detection is based on empirical definitions (Engel and da Silva 2012; Jefferys et al. 2012). To complicate matters further, the shape of the HFO waveform can vary depending on the position of the electrode and its distance from the generating tissue (Buzsáki et al. 2012). This implies that a rigid template of HFO shape may be insufficient for detection. While these ideas can be loosely factored into visual detection, a precise physiological definition of HFOs will be crucial for the next generation of automatic detection. Until that definition is understood, however, visually-detected HFO events will act as the gold standard for evaluating the performance of automated algorithms.

Furthermore, the distinction between physiological and pathological HFOs is not understood. There is evidence that a high rate of ripples and fast ripples is indicative of the seizure onset zone, even without attempting to distinguish between pathological and physiological events (Bragin et al. 2010; Zijlmans et al. 2012). Both ripple and fast ripple rates are higher in seizure onset zones, confirmed by surgical outcome (Jacobs et al. 2010). Other studies have suggested that HFOs should be divided into two categories based on frequency (Bragin et al. 1999; Staba et al. 2002), where ripples are considered to be physiological events and fast ripples are associated with epileptogenicity (Staba et al. 2007). However, to differentiate between pathological and physiological HFOs, frequency alone is insufficient (Matsumoto et al. 2013). In the present study, the default parameters were configured for the detection of ripples, but we expect the algorithm to work for other frequency bands as well, simply by changing the bandpass filter. By iteratively estimating the background amplitude distribution and choosing the threshold, the procedure adapts to the characteristics of the data, regardless of the frequency band or recording modality.

In conclusion, this simple algorithm can be used to automatically detect HFOs with a high degree of accuracy, confirmed by comparison to visually marked events. When compared directly using the same dataset, our detector's performance equaled or exceeded the most commonly used HFO detection method based on the RMS amplitude. The statistical detection of HFOs at different confidence levels was done by estimating the peak amplitude probability distribution of the background activity and then identifying oscillatory iEEG events with amplitudes that were statistically higher than the background. Because the algorithm requires the optimization of only one parameter, related to the percentage of allowable false positive events, this technique can be applied consistently across data from different research centers and different recording modalities. Due to its high detection sensitivity and simple optimization procedure, our detector provides advantages over other techniques for the initial detection of candidate HFO events. It is suitable to be paired with an algorithm for the automatic rejection of artifacts and false positives, or it may be used with human validation. Overall, this

type of automated algorithm is less subjective and much faster than visual HFO detection, and it has the potential to be a powerful tool for the assessment and localization of epileptic activity.

MATLAB code for HFO detection

Code to perform HFO detection in MATLAB using the iterative method described here will be provided as supplementary material.

Acknowledgements

5. References

Amiri M, Lina JM, Pizzo F, Gotman J (2016) High Frequency Oscillations and spikes: Separating real HFOs from false oscillations. Clin Neurophysiol 127:187–196. doi: 10.1016/j.clinph.2015.04.290

Blanco JA, Stead M, Krieger A, et al (2010) Unsupervised classification of high-frequency oscillations in human neocortical epilepsy and control patients. J Neurophysiol 104:2900–2912. doi: 10.1152/jn.01082.2009

Bragin A, Engel J, Staba RJ (2010) High-frequency oscillations in epileptic brain. Curr Opin Neurol 23:151–156. doi: 10.1097/WCO.0b013e3283373ac8

Bragin A, Engel J, Wilson CL, et al (1999) High-frequency oscillations in human brain. Hippocampus 9:137–142. doi: 10.1002/(SICI)1098-1063(1999)9:2<137::AID-HIPO5>3.0.CO;2-0

Burnos S, Hilfiker P, Sürücü O, et al (2014) Human Intracranial High Frequency Oscillations (HFOs) Detected by Automatic Time-Frequency Analysis. PLoS One 9:e94381. doi: 10.1371/journal.pone.0094381

Buzsáki G, Anastassiou CA, Koch C (2012) The origin of extracellular fields and currents — EEG, ECoG, LFP and spikes. Nat Rev Neurosci 13:407–420.

Buzsáki G, Horváth Z, Urioste R, et al (1992) High-frequency network oscillation in the hippocampus. Science 256:1025–1027. doi: 10.1126/science.1589772

Chaibi S, Lajnef T, Sakka Z, et al (2013) A comparaison of methods for detection of high frequency oscillations (HFOs) in human intacerberal EEG recordings. Am J … 3:25–34. doi: 10.5923/j.ajsp.20130302.02

Cho JR, Koo DL, Joo EY, et al (2014) Resection of individually identified high-rate high-frequency oscillations region is associated with favorable outcome in neocortical epilepsy. Epilepsia 55:1872–1883. doi: 10.1111/epi.12808

Crépon B, Navarro V, Hasboun D, et al (2010) Mapping interictal oscillations greater than 200 Hz recorded with intracranial macroelectrodes in human epilepsy. Brain 133:33–45.

Dümpelmann M, Jacobs J, Kerber K, Schulze-Bonhage A (2012) Automatic 80-250Hz "ripple" high frequency oscillation detection in invasive subdural grid and strip recordings in epilepsy by a radial basis function neural network. Clin Neurophysiol 123:1721–1731. doi: 10.1016/j.clinph.2012.02.072

Engel J, da Silva FL (2012) High-frequency oscillations - Where we are and where we need to go. Prog Neurobiol 98:316–318. doi: 10.1016/j.pneurobio.2012.02.001

Ferrari-Marinho T, Perucca P, Mok K, et al (2015) Pathologic substrates of focal epilepsy influence the generation of high-frequency oscillations. Epilepsia 56:592–598. doi: 10.1111/epi.12940

Fujiwara H, Greiner HM, Lee KH, et al (2012) Resection of ictal high-frequency oscillations leads to favorable surgical outcome in pediatric epilepsy. Epilepsia 53:1607–1617. doi: 10.1111/j.1528-1167.2012.03629.x

Gardner AB, Worrell GA, Marsh E, et al (2007) Human and automated detection of high-frequency oscillations in clinical intracranial EEG recordings. Clin Neurophysiol 118:1134–1143. doi: 10.1016/j.clinph.2006.12.019

Gliske S V, Irwin ZT, Davis KA, et al (2016) Universal automated high frequency oscillation detector for real-time, long term EEG. Clin Neurophysiol 127:1057–1066. doi: 10.1016/j.clinph.2015.07.016

Grange W, Haas P, Wild A, et al (2008) Detection of transient events in the presence of background noise. J Phys Chem B 112:7140–7144. doi: 10.1021/jp7114862

Jacobs J, Golla T, Mader M, et al (2014) Electrical stimulation for cortical mapping reduces the density of high frequency oscillations. Epilepsy Res 108:1758–1769. doi: 10.1016/j.eplepsyres.2014.09.022

Jacobs J, LeVan P, Chander R, et al (2008) Interictal high-frequency oscillations (80-500 Hz) are an indicator of seizure onset areas independent of spikes in the human epileptic brain. Epilepsia 49:1893–907. doi: 10.1111/j.1528-1167.2008.01656.x

Jacobs J, Staba R, Asano E, et al (2012) High-frequency oscillations (HFOs) in clinical epilepsy. Prog Neurobiol 98:302–15. doi: 10.1016/j.pneurobio.2012.03.001

Jacobs J, Zijlmans M, Zelmann R, et al (2010) High-frequency electroencephalographic oscillations correlate with outcome of epilepsy surgery. Ann Neurol 67:209–220. doi: 10.1002/ana.21847

Jefferys JGR, Menendez de la Prida L, Wendling F, et al (2012) Mechanisms of physiological and epileptic HFO generation. Prog Neurobiol 98:250–264. doi: 10.1016/j.pneurobio.2012.02.005

Malinowska U, Bergey GK, Harezlak J, Jouny CC (2015) Identification of seizure onset zone and preictal state based on characteristics of high frequency oscillations. Clin Neurophysiol 126:1505–1513. doi: 10.1016/j.clinph.2014.11.007

Matsumoto A, Brinkmann BH, Matthew Stead S, et al (2013) Pathological and physiological high-frequency oscillations in focal human epilepsy. J Neurophysiol 110:1958–64. doi: 10.1152/jn.00341.2013

Roehri N, Lina J, Mosher JC, et al (2016) Time-frequency strategies for increasing high frequency oscillation detectability in intracerebral. 2:1–12. doi: 10.1109/TBME.2016.2556425

Staba RJ, Frighetto L, Behnke EJ, et al (2007) Increased fast ripple to ripple ratios correlate with reduced hippocampal volumes and neuron loss in temporal lobe epilepsy patients. Epilepsia 48:2130–2138. doi: 10.1111/j.1528-1167.2007.01225.x

Staba RJ, Wilson CL, Bragin A, et al (2002) Quantitative Analysis of High-Frequency Oscillations (80-500 Hz) Recorded in Human Epileptic Hippocampus and Entorhinal Cortex. J Neurophysiol 88:1743–1752.

Staba RJ, Wilson CL, Bragin A, et al (2004) High-frequency oscillations recorded in human medial temporal lobe during sleep. Ann Neurol 56:108–115. doi: 10.1002/ana.20164

Worrell G a., Gardner AB, Stead SM, et al (2008) High-frequency oscillations in human temporal lobe: Simultaneous microwire and clinical macroelectrode recordings. Brain 131:928–937. doi: 10.1093/brain/awn006

Worrell GA, Jerbi K, Kobayashi K, et al (2012) Recording and analysis techniques for high-frequency oscillations. Prog Neurobiol 98:265–78. doi: 10.1016/j.pneurobio.2012.02.006

Zelmann R, Mari F, Jacobs J, et al (2010) Automatic detector of High Frequency Oscillations for human recordings with macroelectrodes. 2010 Annu Int Conf IEEE Eng Med Biol Soc EMBC'10 2329–2333. doi: 10.1109/IEMBS.2010.5627464

Zelmann R, Mari F, Jacobs J, et al (2012) A comparison between detectors of high frequency oscillations. Clin Neurophysiol 123:106–16. doi: 10.1016/j.clinph.2011.06.006

Zijlmans M, Jiruska P, Zelmann R, et al (2012) High-frequency oscillations as a new biomarker in epilepsy. Ann Neurol 71:169–178. doi: 10.1002/ana.22548
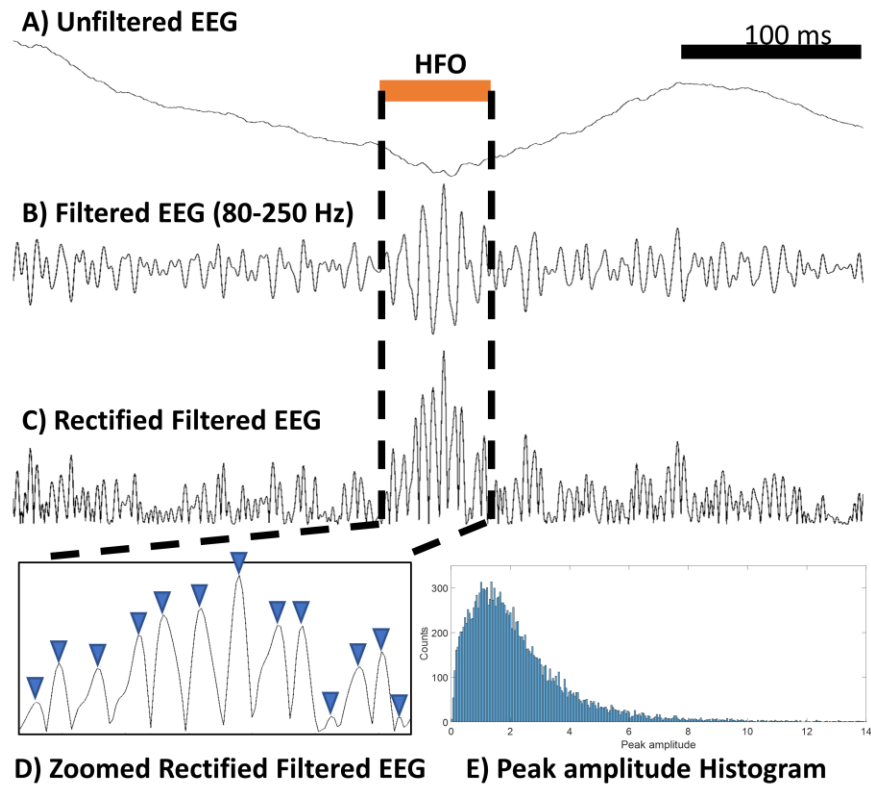
Figure 1. Detection procedure and an example of an HFO event (orange bar). (A) The algorithm starts with unfiltered EEG signal. (B) Then the EEG signal is bandpass filtered from 80-250 Hz. (C) The filtered signal is then rectified. (D)We define "peaks" as local maxima in the rectified, band-pass filtered data (blue triangles). (E) A histogram of peak amplitudes resembles a gamma distribution. HFOs will be superimposed upon this distribution, causing an extraordinarily long tail on the right side.
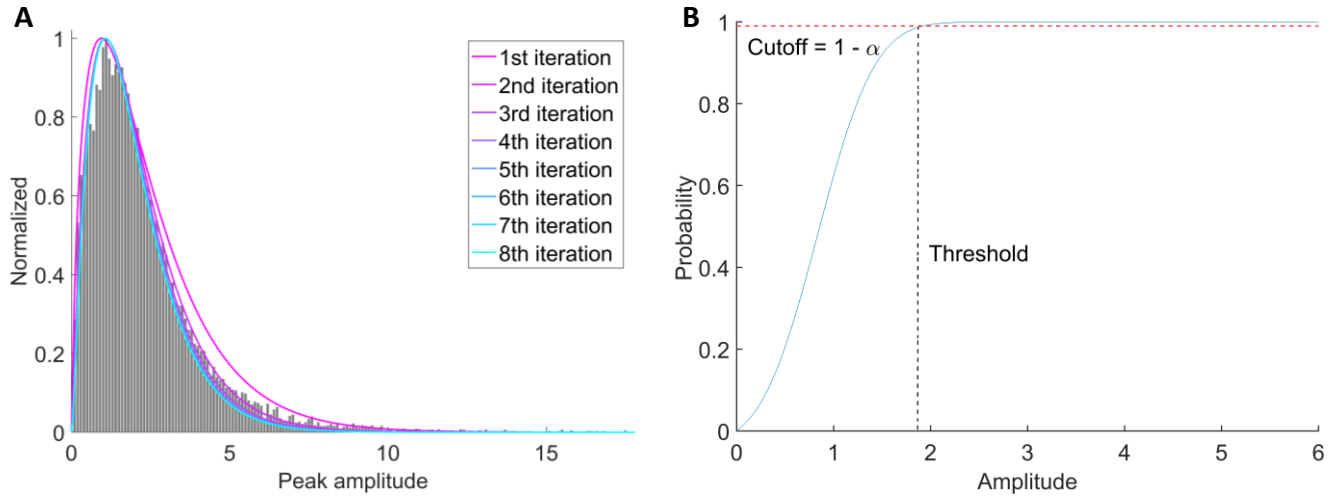
Figure 2. Iterative procedure for threshold optimization. (A) The amplitude probability distribution for a single channel, including both HFOs and background activity (gray bars). Lines represent the estimated background distribution f(x) over 8 iterations. For each iteration, the amplitude values that are higher than the cutoff F(x) > 1- α are removed before re-estimation of the new distribution. After a number of iterations, the estimate of the background distribution converges, providing an optimum threshold for the detection of HFOs. (B) The cutoff value is based on the cumulative probability distribution function F(x) (blue line), and it is directly related to the amplitude threshold in the iEEG data.
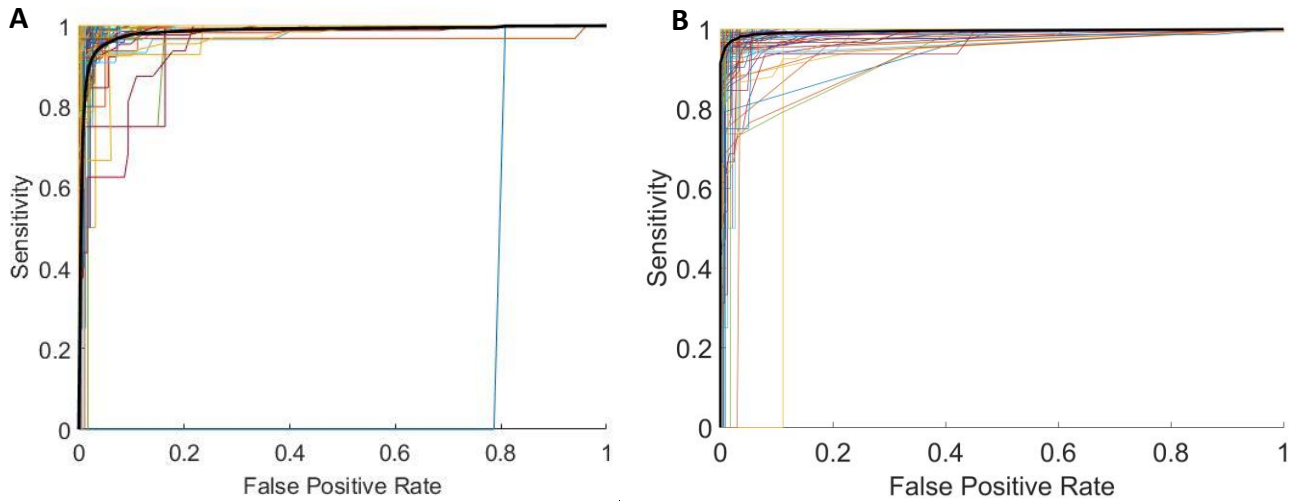
Figure 3. Detector performance based on ROC curves when the threshold was individually optimized to each channel. The subfigures show ROC curves for all 296 channels (thin lines) and the vertical average ROC curve (thick black line) of (A) our detector with iterative threshold selection and (B) RMS detector. The reported optimum threshold was the point on each ROC curve that was closest to the upper left corner.
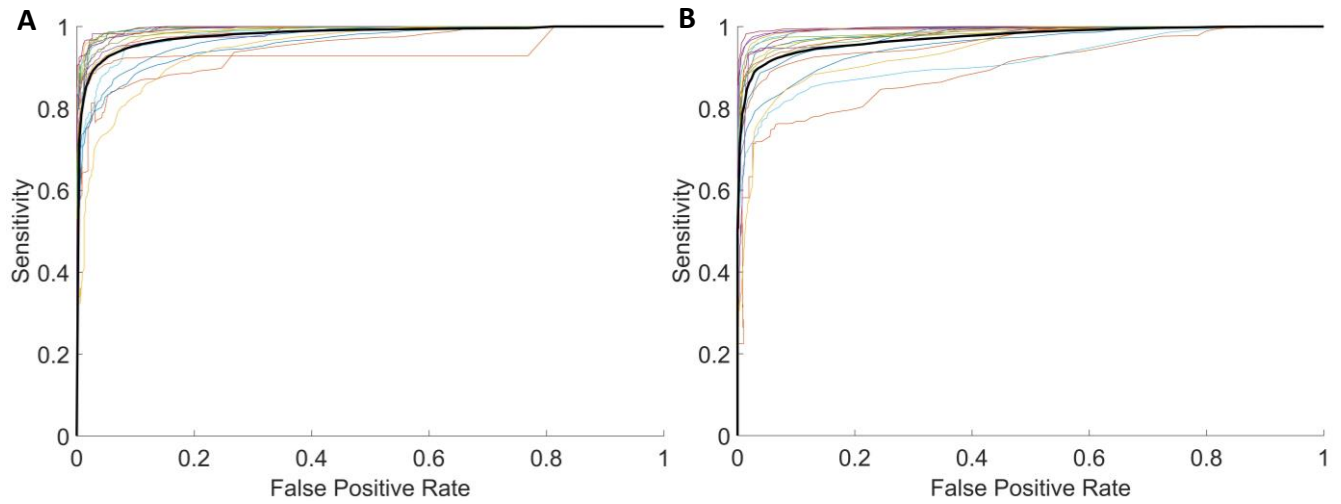
Figure 4. The ROC curves when a single α value was applied to all channels within the same subject in (A) our detector with iterative threshold selection and (B) the RMS detector. The thin lines represent the average of all channels within a subject, and the thick black line is the vertical average of all equally weighted 19 patients.
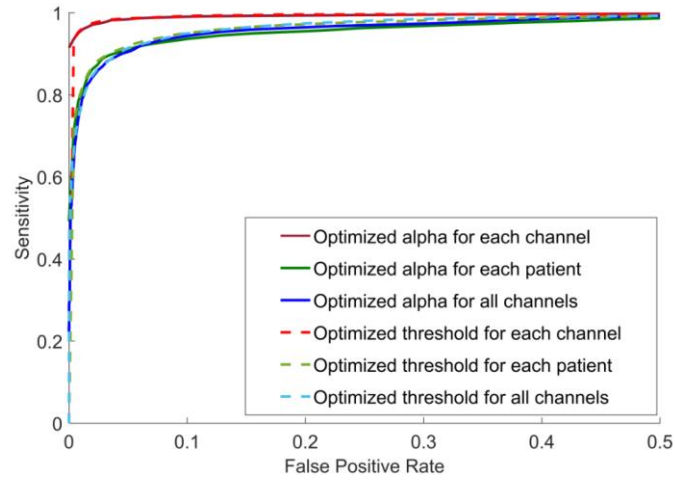
Figure 5. A comparison of ROC curves when both automatic detectors are tested in three different ways. The solid lines represent the ROC curves for our detector, while dashed lines show the results for the RMS detector. First, the threshold was optimized for each individual channel and averaged across all channels (red lines). Second, averaged across patients with a constant threshold for all channels within each patient (green lines). Third, a single threshold value for all channels (blue lines).
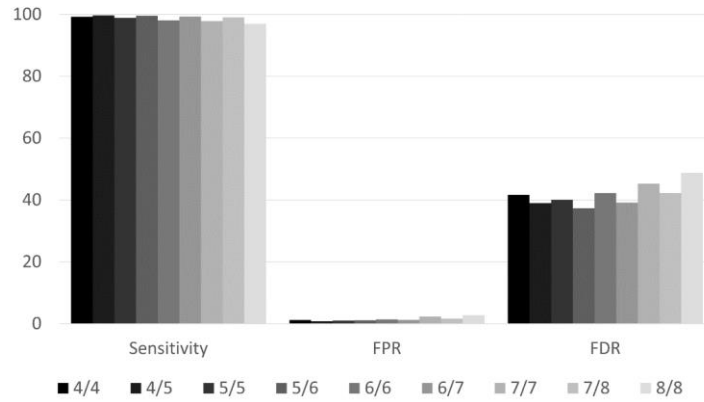
Figure 6. A comparison of Sensitivity, FPR, and FDR when α was individually optimized for each channel, as a function of the number of consecutive peaks for HFO detection. Results are shown for nine different criteria for the number of peaks required for each HFO, including 4 consecutive peaks (4/4), 4 out of 5 consecutive peaks (4/5), and up to 8 consecutive peaks (8/8). Recall that two "peaks" (local maxima) in the rectified, filtered data constitute one full oscillation. As a default configuration, we chose 5 out of 6 consecutive peaks above threshold (5/6) as it resulted in marginally better performance than the other choices. Note that better detection performance corresponds to high sensitivity, but low FPR and FDR.
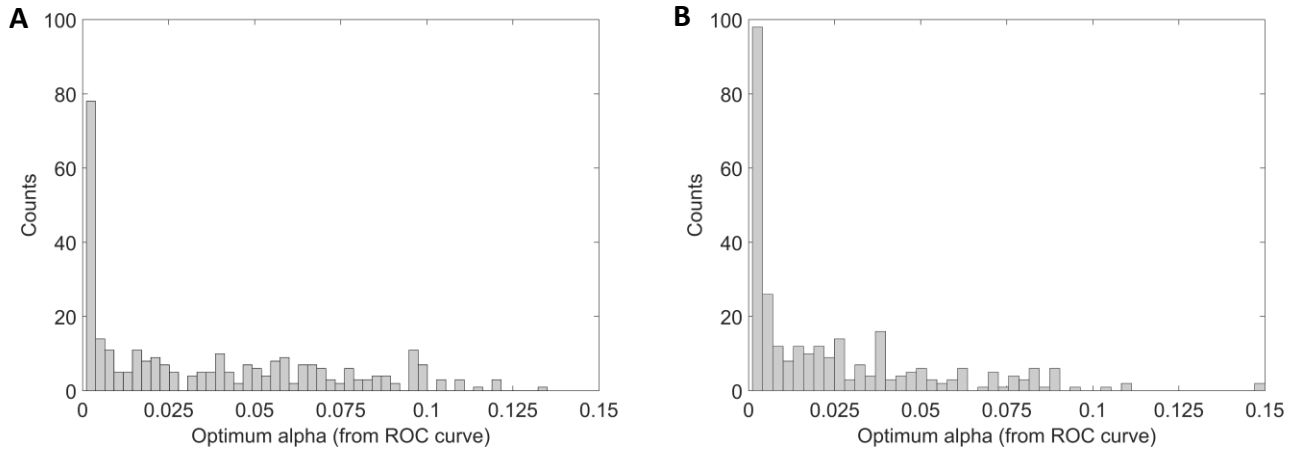
Figure 7. Histogram of α values at optimum thresholds for all 296 channels. (A) α values when the threshold was optimized using sensitivity and FPR. (B) α values when the threshold was optimized using sensitivity and FDR.

Table 1. Detection statistics of our detector with iterative threshold selection at the optimum threshold. Sensitivity, FPR, and AUC were computed based on strict validation (PosAND and NegBASE); FDR was calculated with open validation (PosANY).

	Sensitivity (%)			FPR (%)			FDR (%)			AUC
	Mean ± SD	Median	Min, Max	Mean ± SD	Median	Min, Max	Mean ± SD	Median	Min, Max	
Average across 296 channels (Single α for each channel)	99.6 ± 1.4	100.0	87.5, 100.0	1.1 ± 5.1	0.0	0.0, 80.8	37.3 ± 30.6	35.7	0.0, 96.5	0.995
Average across 19 patients (Single α for each patient)	93.6 ± 3.7	94.7	86.0, 98.2	5.2 ± 2.9	4.4	1.7, 1189	66.9 ± 10.3	68.7	49.1, 79.7	0.980
All events pooled together (Single α for all channels)	93.1	-	-	6.4	-	-	71.6	-	-	0.981
Average across 296 channels (Optimized threshold by FDR)	89.1 ± 13.8	94.4	33.3, 100.0	6.1±23.6	0.0	0.0, 100.0	16.5 ± 20.6	9.0	0.0, 94.4	-
Eight-fold Cross-validation	80.0 ± 3.4	80.3	74.9, 84.8	1.3 ± 0.6	1.1	0.6, 2.6	49.4 ± 5.1	48.4	41.8, 55.6	-

Table 2 Detection statistics for our detector with non-iterative threshold selection at the optimum threshold. Sensitivity, FPR, and AUC were computed based on strict validation (PosAND and NegBASE); FDR was calculated with open validation (PosANY).

	Sensitivity (%)			FPR (%)			FDR (%)			AUC
	Mean ± SD	Median	Min, Max	Mean ± SD	Median	Min, Max	Mean ± SD	Median	Min, Max	
Average across 296 channels (Threshold optimized for each channel)	99.5 ± 1.4	100.0	87.5, 100.0	1.1 ± 5.0	0.0	0.0, 79.5	36.9 ± 29.2	36.1	0.0, 96.5	0.995
Average across 19 patients (Single number of SD used for threshold for each patient)	92.7 ± 4.3	94.7	85.0, 98.8	6.0 ± 3.8	4.7	1.5, 16.0	66.6 ± 9.7	67.5	46.9, 83.2	0.976
All events pooled together (Single number of SD used for threshold for all channels)	92.3	-	-	6.1	-	-	68.0	-	-	0.979
Average across 296 channels (Optimized threshold by FDR)	88.7 ± 13.6	92.7	33.3, 100.0	1.0 ± 9.4	0.0	0.0, 79.5	16.9 ± 20.5	9.7	0.0, 94.6	-
Eight-fold Cross-validation	64.4 ± 4.3	64.4	58.9, 71.2	0.4 ± 0.3	0.3	0.2, 1.2	31.2 ± 6.4	30.4	22.9, 42.0	-

Table 3. Detection statistics of RMS detector at the optimum threshold (Staba et al. 2002)*. Sensitivity, FPR, and AUC were computed based on strict validation (PosAND and NegBASE); FDR was calculated with open validation (PosANY).*

	Sensitivity (%)			FPR (%)			FDR (%)			AUC
	Mean ± SD	Median	Min, Max	Mean ± SD	Median	Min, Max	Mean ± SD	Median	Min, Max	
Average across 296 channels (Threshold optimized for each channel)	98.9 ± 4.3	100.0	48.3, 100.0	0.7 ± 2.0	0.0	0.0, 18.8	34.6 ± 29.0	33.3	0.0, 97.5	0.995
Average across 19 patients (Single number of SD used for threshold for each patient)	92.2 ± 5.1	93.7	76.3, 98.2	5.4 ± 3.4	4.7	1.5, 12.8	64.0 ± 10.1	66.0	36.3, 78.5	0.973
All events pooled together Single number of SD used for threshold for all channels)	92.0	-	-	7.0	-	-	66.2	-	-	0.976
Average across 296 channels (Optimized threshold by FDR)	88.6 ± 14.4	95.5	33.3, 100.0	7.3 ± 25.6	0.0	0.0, 100.0	16.9 ± 21.1	9.8	0.0, 97.5	-
Eight-fold Cross-validation	65.8 ± 4.7	65.0	58.1, 71.7	0.5 ± 0.4	0.3	0.2, 1.5	33.4 ± 7.2	30.4	24.0, 44.4	-