Published in final edited form as:
J Neural Eng. ; 19(4): . doi:10.1088/1741-2552/ac8766.

A sparse representation strategy to eliminate pseudo-HFO events from intracranial EEG for seizure onset zone localization

Behrang Fazli Besheli[1], **Zhiyi Sha**[2], **Jay R. Gavvala**[3], **Candan Gurses**[4], **Sacit Karamursel**[5], **Michael M. Quach**[6], **Daniel J. Curry**[7], **Sameer A. Sheth**[8], **David J. Francis**[9], **Thomas R. Henry**[2], **Nuri F. Ince**[1]

[1]Department of Biomedical Engineering, University of Houston, Houston, TX, USA

[2]Department of Neurology, University of Minnesota, Minneapolis, MN, USA

[3]Department of Neurology-Neurophysiology, Baylor College of Medicine, Houston, TX, USA

[4]Department of Neurology, School of Medicine, Koç Üniversitesi, Istanbul, Turkey

[5]Department of Physiology, School of Medicine, Koç Üniversitesi, Istanbul, Turkey

[6]Department of Neurology, Texas Children’s Hospital, Houston, Texas, USA

[7]Department of Neurosurgery, Texas Children’s Hospital, Houston, Texas, USA

[8]Department of Neurosurgery, Baylor College of Medicine, Houston, Texas, USA

[9]Department of Psychology, University of Houston, Houston, TX, USA

Abstract

Objective.—High-frequency oscillations (HFOs) are considered a biomarker of the epileptogenic zone in intracranial EEG recordings. However, automated HFO detectors confound true oscillations with spurious events caused by the presence of artifacts.

Approach.—We hypothesized that, unlike pseudo-HFOs with sharp transients or arbitrary shapes, real HFOs have a signal characteristic that can be represented using a small number of oscillatory bases. Based on this hypothesis using a sparse representation framework, this study introduces a new classification approach to distinguish true HFOs from the pseudo-events that mislead seizure onset zone (SOZ) localization. Moreover, we further classified the HFOs into ripples and fast ripples by introducing an adaptive reconstruction scheme using sparse representation. By visualizing the raw waveforms and time-frequency representation of events recorded from 16 patients, three experts labelled 6400 candidate events that passed an initial amplitude-threshold-based HFO detector. We formed a redundant analytical multiscale dictionary built from smooth oscillatory Gabor atoms and represented each event with orthogonal matching pursuit by using a small number of dictionary elements. We used the approximation error and residual signal at each iteration to extract features that can distinguish the HFOs from any type of

---

nfince@uh.edu .

Conflict of Interests
Sameer A. Sheth is a consultant for Boston Scientific, Zimmer Biomet, NeuroPace. Other authors don’t have any conflict of interests to disclose.

artifact regardless of their corresponding source. We validated our model on sixteen subjects with thirty minutes of continuous interictal iEEG recording from each.

Main Results.—We showed that the accuracy of SOZ detection after applying our method was significantly improved. In particular, we achieved a 96.65% classification accuracy in labelled events and a 17.57% improvement in SOZ detection on continuous data. Our sparse representation framework can also distinguish between ripples and fast ripples.

Significance.—We show that by using a sparse representation approach we can remove the pseudo-HFOs from the pool of events and improve the reliability of detected HFOs in large data sets and minimize manual artifact elimination.

Keywords

Epilepsy; High-Frequency Oscillation; Orthogonal Matching Pursuit; Sparse representation; Pseudo-HFO

1. Introduction

Epilepsy is a condition characterized by recurrent unprovoked seizures (1). It is one of the most common neurological disorders and affects between 0.5–1% of the population worldwide (2). Despite the advancements in pharmaceutical therapy, about 30% of patients remain retractable to antiepileptic drugs (3). For these patients, resective surgery, the removal of a small portion of the brain where seizures are generated, is one of the most effective treatments. The outcome of epilepsy surgery highly depends on the proper localization of the seizure onset zone (SOZ) identified by prolonged intracranial EEG (iEEG) monitoring (4). This stresses the necessity of using biomarkers in iEEG recording to localize the SOZ precisely and improve the surgical outcome accordingly.

During the past two decades, high-frequency oscillations (HFOs) of iEEG have become a promising biomarker of the epileptogenic zone (EZ) (5). HFOs are transients residing between 80–600 Hz with low amplitude that last around 30–100 ms (6). HFOs generally are divided into two categories: ripples (R) with frequencies between 80–250 Hz and fast ripples (FR) with frequencies at 250–600 Hz.

Considering their low amplitude and short duration, visual detection of epileptogenic HFOs in long-term recordings is cumbersome. For this reason, computer-based methods are frequently used to automate HFO detection. Generally, HFO detectors employ either a high-pass or band-pass filtering in the HFO range. They detect the oscillatory events that stand out from the background activity. Earlier investigations showed that iEEG data obtained in prolonged epilepsy monitoring units (EMU) frequently become contaminated with electrical interference and other artifacts (7,8,9). These artifacts have internal and external origins (10) and might heavily corrupt the recordings. Numerous studies (11,12,13) have shown that sharp events or random noise might mimic HFO after high pass filtering (12). The filtering effect creates a spurious oscillation in the HFO range, i.e., 80–600 Hz.

Therefore, pseudo-events with random nature or sharp changes might reduce the SOZ localization accuracy. Moreover, if the artifacts are recorded from channels that overlap with normal or functional cortex, it might have catastrophic results. Consequently, there is a growing interest in distinguishing artifacts and noise from the “real” HFOs for accurate SOZ localization. Due to their random and arbitrary nature, it is difficult to translate different types of noise into handcrafted features. On the other hand, defining large numbers of rules and constraints in the HFO detector might also impair the sensitivity, causing real HFOs to be removed from the event pool. Here we introduce a new method to distinguish between true and pseudo-HFOs by inspecting the raw signal without bias towards conventional HFO features such as time-frequency representation.

In this framework, we employ orthogonal matching pursuit (OMP) to represent initial candidate events using a small number of elements from a redundant dictionary. We hypothesized that, unlike pseudo-HFOs that might be corrupted with sharp transients or arbitrary shapes, real HFOs have a waveform characteristic that can be represented using only a few oscillatory atoms. A perfect local and global representation of events using a small number of Gabor atoms was translated to features in order to shape a classifier distinguishing between all types of noises regardless of their sources from the real HFO. We also validated the importance of using Gabor atoms for HFO classification and representation by repeating the whole pipeline using the discrete cosine dictionary. We tested this method on sixteen subjects and compared the results with a previously published method (14). The results showed a notable improvement in classification accuracy and SOZ localization compared to the existing method.

This paper also describes a new approach to identifying the frequency components of HFO events by exploiting the sparse representation as an adaptive regenerative model to reconstruct the original events. This can be a valuable tool for identifying the frequency of HFOs and splitting the HFOs into R and FR categories.

2. Materials and Methods

2.1 Data Acquisition

In this multi-institutional study, we analyzed iEEG data recorded from sixteen patients with refractory focal epilepsy. All recordings were obtained in the EMU at a sampling frequency equal to or greater than 2 kHz. This study was approved by the Institutional Review Boards (IRB) of the University of Houston. Relevant annotations, including seizure onset zone (SOZ) and surgery outcome, were provided by the clinical team at the affiliated institutes. The study protocol was also approved by the IRB at each site the data were recorded.

For each subject, 30-minute-long interictal iEEG segments were selected for the analysis. The raw iEEG was converted to bipolar derivation for further processing. To investigate the efficacy of the algorithm, all channels were included in the analysis without any pre-processing or manual artifact rejection.

2.2 Signal Processing

As shown in figure 1, in three steps (A, B, and C), the proposed method distinguished between real HFOs and pseudo-HFOs originating from noise/artifacts. First, using an amplitude-threshold-based detector (figure 1(A)), an initial pool of potential HFO events was formed. In the second step (figure 1(B)), each event is represented with OMP in a sparse fashion using atoms from a redundant oscillatory dictionary. The approximation error and the residual signal at each iteration were used to shape a feature vector. Finally, as shown in figure 1(C), events were classified into HFO and pseudo-HFO categories using a random forest (RF) classifier. All algorithms were developed and employed in MATLAB 2019b (MathWorks, Inc., Natick, MA, US) software. Below we describe each stage in detail.

2.2.1 Initial Detection—An amplitude-threshold-based detector described in (14) captured the initial events with high sensitivity but low specificity. Specifically, the initial detection was executed in two separate frequency bands. After filtering the iEEG data in the R [80–250 Hz] and FR [250–600 Hz] ranges using a 64-order FIR digital filter, the standard deviation (SD) of these two bands was computed using a 100-millisecond long sliding window with 50% overlap. To capture the variation of background activity and minimize the effect of outliers with large amplitudes, the SD was computed in each sliding window, and the median of these values was computed in 1 minute. This provided a robust and time-varying estimation of the SD of the background activity. The amplitude threshold is defined as three times the SD estimated from the background activity. 512 samples long iEEG segments locally exceeding this threshold were used to form a pool of events once they satisfied the following constraints:

- **i.** The number of samples that cross the threshold should be more than four to capture at least two complete oscillations above the threshold at each event.
- **ii.** The distance between oscillatory components (at a sampling frequency of f_s) that passed the threshold should be at least $80/f_s$ and $250/f_s$ for R and FR bands respectively.
- **iii.** The amplitude of filtered events at the center should be four times larger than the SD of the signal at the sides.

After the detection of initial events across subjects, three experts jointly labelled 400 events (at least 200 HFOs and 200 pseudo-HFOs) at minimum for each subject. A graphical user interface (GUI) was designed in MATLAB for the visual inspection of raw signal, its filtered version in the HFO band, and its time-frequency plane representation. Experts reviewed the data in time and time-frequency domain using the GUI and annotated the events. (A snapshot of this GUI and more information regarding the annotation process are provided in supplementary figure 7). These labelled events were used to extract informative features, train a classification algorithm, and validate its performance in later stages using the leave one subject out method.

2.2.2 Orthogonal Matching Pursuit and Analytical Redundant Dictionary— After forming an HFO event pool using the initial amplitude-threshold-based detector, each candidate event was represented in a sparse fashion using an overcomplete dictionary. Assume that we want to represent an HFO event $y \in R^{m \times 1}$ using a linear combination of a small number of atoms from a dictionary of $D \in R^{m \times n}$ where the atoms span the columns of matrix D with m, n the size of events and the number of dictionary atoms, respectively:

$$y = D\alpha \tag{1}$$

for $n > m$, in which the dictionary is redundant, this representation can be formulated as a l_0 – minimization problem that penalizes the number of nonzero coefficients in α:

$$\min_{\alpha}\|\alpha\|_0 \ s.t. \ y = D\alpha \tag{2}$$

where $\| \cdot \|_0$ denotes l^0 – norm, which counts the number of nonzero entries in the vector $\alpha \in R^{n \times 1}$. Figure 2(C) shows the matrix formulation of sparse representation of an HFO using predefined waveforms called atoms.

The sparse approximation problem of equation (2) is combinatorial and NP-hard. Thus, approximate but efficient solutions such as MP (15) and OMP (16) are considered, which employ a greedy search in a recursive fashion to select atoms sequentially.

In our study, to represent the events, we employed OMP and enforced sparsity in α, by forming a rich Gabor dictionary resembling the oscillatory nature of the HFO components. The Gabor basis is defined as the product of a Gaussian function with the cosine function:

$$g_{u,\sigma}(t) = 1 \Big/ \sqrt{2\pi\sigma} e^{-\frac{(t-u)^2}{2\sigma^2}} . \cos\left(2\pi\frac{\omega}{N}(t-u)\right) \tag{3}$$

which is described by parameters σ: time spread, u: shift in time, ω: frequency.

The dictionary $D = [d_1, d_2, d_3, \ldots, d_n]$ where $d_i \in R^{m \times 1}$ is the ith dictionary atom then can be formulated as Gabor atoms with different frequencies and time supports:

$$D = \left[g_{(\sigma_1, u_1, \omega_1)}\middle|g_{(\sigma_2, u_2, \omega_2)}\middle|\cdots\middle|g_{(\sigma_n, u_n, \omega_n)}\right]. \tag{4}$$

We varied the parameters based on the definition of HFOs in terms of their time support (see figure 2(B)) and frequency content (see figure 2(A)). The frequency of atoms, w, ranged between 0–600 Hz to represent R and FR components and the slow waves below 80 Hz. As implemented originally by *Mallat and Zhang* in (15), the parameter sigma, which corresponds to an atom's width in time, is chosen from dyadic sequences of integers ($\sigma = 2^{-j}$) and j (integer) were varied in a systematic fashion below and above 80Hz. Specifically, $j \in \{0,1\}$ for $w < 35\ Hz$ and $35 < w < 80$ (figure 2(A) first and second rows). For $80 < w < 200\ Hz,\ j \in \{2\}$ (figure 2(A) 3rd row). For $150 < w < 500\ Hz\ j \in \{3\}$ (figure 2(A) 4th row) and for $300 < w < 600\ Hz\ j \in \{4\}$ (figure 2(A) 5th row). Consequently, atoms with localized high-frequency oscillations were constructed through such a strategy.

Similarly, the shifts in time are chosen from dyadic sequences of integers. As we go to the higher frequencies in our dictionary, the scale of the atoms becomes smaller and more shifts in time are included in the dictionary to have localized atoms in higher frequencies. Overall, our Gabor dictionary included 5600 atoms. Due to their analytical design, we stored the frequency information of all atoms in a vector and used this information in feature extraction and later in R and FR categorization.

2.2.3 Feature Extraction

We extracted simple informative features from the residual of the signal and selected atoms during the iterations of OMP. Let r_i represent the residual of candidate event y at ith iteration,

$$r_i = y - \sum_{k=1}^{i} a_k g_k \tag{5}$$

where g_k is the selected atom at iteration k, a_k is its coefficient. We computed the residual signal over successive iterations and extracted the following features from the residual signal and selected atoms:

Approximation Error (L^2 Error): Approximation error is defined as the ratio of the energy of the residual to the energy of the event y:

$$\varepsilon_i = \frac{\|r_i\|_2}{\|y\|_2}. \tag{6}$$

Here ε_i revealed the ability to reconstruct an event at a specific iteration globally.

Variation Factor (V-Factor): The range of residual error was divided by the standard deviation of residual error at iteration i:

$$V_i = \frac{\max(r_i) - \min(r_i)}{std(r_i)}. \tag{7}$$

We named V_i as variation factor at iteration i. This feature captured the local behaviour of the residual error with respect to the overall error. If the signal is corrupted with a localized sharp artifact, the OMP needs a large number of atoms to reconstruct such segments. Thus, it leaves the large localized sharp artifacts untouched or poorly represents within the first few iterations as the dictionary does not include sharp atoms. Therefore, while the overall error is small, which is captured by the standard deviation of r_i, the V-Factor will be large. This feature plays a vital role in distinguishing between sharp artifacts and HFOs and describes the quality of reconstruction of events locally (Supplementary figures 1 and 2 present examples of the OMP process, selected atoms, and extracted features for real and pseudo-HFOs. See (17) for a demo).

Line Noise: The iEEG signals can often be exposed to power line interference at 50 or 60 Hz. In such a case, a small signal distortion or discontinuity caused fake oscillations mimicking HFO. Such events can be easily represented with oscillatory Gabor atoms, but the selected atoms repeat themselves at the power line frequency. On this account, the number of selected atoms around the line noise frequency might serve as an important feature. Consequently, since we know the frequency index of the analytical Gabor atoms, the number of atoms located 60 Hz over iterations was used as another feature. A representative example regarding this feature can be found in supplementary figure 6.

Other Features: The range and standard deviation of the raw event and in the HFO range above 80 Hz were computed and used as additional features.

2.2.4 Classification of Events and Model Validation—We utilized the random forest classifier to assign the events into HFO and noise (pseudo-HFO) categories. RF is an ensemble learning method that fits multiple decision tree models on different subsets of a training dataset, then combines the predictions from all models. Specifically, it contains different individual trees where each tree uses a random subset of the feature space in each split. (18). In this way, the correlation between trees is reduced, and the prediction power increases. The final decision of the RF was computed based on the majority voting scheme. We utilized the RF toolbox (19) to construct the trees and employed the Gini impurity index (18) to find the best feature in each node and do the splits with a smaller Gini index. The Gini index is defined as

$$G = \sum_{i = 1}^{C} p(i)(1 - p(i)) \tag{8}$$

where C is the number of total classes and $p(i)$ denotes the probability of picking up data from class i. In our problem, we wish to classify the events into two classes, i.e., real HFO and pseudo-HFO.

Feature Selection: At this stage, one critical question is how to choose the number of iterations to represent the HFO events in a sparse fashion, which will influence the approximation error and V-Factor values and their discrimination power. Rather than computing these two features at a specific iteration, we continued to approximate each event up to 50 iterations. Then, corresponding features, including approximation error, V-Factor, line noise, range, and SD of raw event and high-frequency components, were concatenated across iterations to form a 154-dimensional feature vector. The computed features were fed to the RF classifier, and the learned model was used on a test subject who was not included in the learning stage. More specifically, the efficacy of the model is validated using leave -one-subject-out: One subject was used to test the model, while the rest of the subjects were used as training data to learn the informative features. This process is repeated for all subjects, and the mean accuracy is reported as the overall classification accuracy. We opted to use 50 trees with 5 minimum leaf and 10 minimum parent sizes with the square root of the total number of features as the number of the predictor to sample.

Consequently, we compared our detector with the previously published method (14), which uses time-frequency features, including sub-band power ratio, the frequency with maximum peak to notch ratio, and spectral entropy to form clusters of events using the Gaussian mixture model (GMM). In order to compare these two methods, we used the labelled events of each subject. The time-frequency features were computed, and the noise cluster(s) was selected. The accuracy was obtained using the ground truth labels of events in each subject.

We also inspected the efficacy of using the redundant Gabor dictionary against the complete discrete cosine dictionary by repeating the same feature extraction and classification scheme.

2.2.5 HFO Categorization via Sparse Representation—

Once the classification is done and pseudo-HFOs are removed from the pool of events, we use the sparse model to further classify the HFO events into the Rs and FRs. The schematic of sparse representation and a typical example of the original and reconstructed HFOs are shown in figure 3.

Each event goes into band-pass filters in the range of R (80–250 Hz) and FR (250–600 Hz). In the following step, R and FR components enter the amplitude-threshold-based detector, counting the number of threshold crossings. The detector accepts an event as HFO if it passes the threshold at least four positive or negative cycles. If the detector confirms an existing oscillatory component that pops out of the background as an HFO component, the sparse representation will be conducted on that specific signal, and the residual signal will be replaced with the original signal. The amplitude-threshold-based detection and representation as a joint procedure will continue to represent each band until no component from the residual passes the amplitude-threshold-based detector. This framework acts as a regenerative model that reconstructs HFOs without adapting to the local background by using the initial detector as a stopping criterion. Finally, the event will be reconstructed by mixing the low-frequency signal below 80 Hz, and R/FR bands reconstructed using the adaptive sparse representation.

The classification of R and FR is then conducted based on the highest frequency among atoms that have been used during the adaptive sparse representation process (see figure 3).

2.2.6 Validation through SOZ identification—

After training a classification model using labelled events, as a next step, we validated its efficacy over long-term iEEG data towards SOZ identification. We executed the whole classification process on detected events in 30 minutes of continuous interictal iEEG recording of each subject. The delineation of the SOZ is assessed as the rate of classified HFOs coming from SOZ channels as follows:

$$Accuracy = \frac{Events\ Coming\ from\ SOZ}{Total\ Number\ of\ Events} \tag{9}$$

We evaluated the efficacy of the classification method by calculating the rate of HFOs coming from SOZ before and after applying the noise removal method. We then use a paired Wilcoxon rank test to show the statistical significance of SOZ localization after removing pseudo-HFO events from the pool of candidates.

The agreement between SOZ and Rs/FRs is evaluated by applying the adaptive HFO sparse representation. We realized that the FRs were rare compared to Rs, and a few subjects might have a small amount of FRs compared to the size of the entire pool. To fairly compare the SOZ accuracy estimated from Rs and FRs, we excluded subjects with an FR rate below 3% of the total HFO pool in this part of the study. The SOZ localization of Rs and FRs was computed, and the Wilcoxon rank test was applied to evaluate the statistical difference between these two groups.

3. Results

Across all sixteen subjects, around 3200 HFO and 3200 pseudo-HFO events that passed the amplitude-threshold-based HFO detector were labelled. We annotated an equal number of real and pseudo-HFO events among all subjects to avoid bias toward a specific subject. Figure 4 (A) shows the continuous iEEG recording and the corresponding HFO band data. Typical annotated HFO and pseudo-HFO events that passed the initial detector are provided in figure 4 (B, C). More examples on real and pseudo-HFOs are also provided in supplementary figure 8 and 9. When the filtered continuous trace is inspected visually, one can get easily mislead by these pseudo HFO events. When inspected closely, the raw pseudo-HFOs have no clear oscillation but some random waveform that passed the initial detector.

3.1 Characteristic of Features over Iterations

We executed the sparse reconstruction process up to 50 iterations. Figures 5(A) and 5(B) visualize the residual signal, the change in approximation error, and V-Factor over iterations for various HFO and pseudo-HFO events. We observed that for the HFO events, the approximation error decreased rapidly within the first 15–20 iterations and then almost reached a plateau. This plateau is likely due to the white noise structure of the residual background, which could not be represented by local oscillatory atoms efficiently. We observed flat behaviour for the V-Factor of the HFO samples across iterations. Compared to HFO events, the pseudo-HFOs could not be represented efficiently within the first 15–20 iterations. For those pseudo-HFO events with localized sharp artifacts, the V-Factor values were also larger than those of real HFOs. These sample events provide evidence that pseudo-events cannot be represented efficiently locally or globally in a sparse fashion.

The approximation error and V-Factor characteristic for representative and all subjects are given in figure 5(C–F). Figures 5(C–F) illustrate that HFOs have low approximation error and V-Factor, which confirms our previous assumption. In patients 3 and 10, the approximation error (ε) was a distinguishable feature as most artifacts fell into a category of pseudo-HFOs with a high background noise level and, as a result, high approximation error. As sharp artifacts exist in patients 4 and 6, the V-Factor (V) will be a vital feature (see figure 5(C–D)).

Figure 5(F) shows the AUC versus the number of iterations for ε_i, V_i, and line noise (L_i) features. While the ε_i and L_i features have a larger AUC before iteration 15th, the V-Factor reached a maximum value around iteration 30th.

3.2 Classification Accuracy and Selected Features

Figure 6(A) shows the confusion matrix of our proposed method vs. other strategies. The proposed method using OMP, and redundant Gabor dictionary reached 96.65% accuracy in discriminating the real and pseudo-HFOs.

When we replaced the redundant Gabor dictionary with the complete discrete cosine transform (DCT) dictionary, we could reach 93.53% mean accuracy. Finally, the STFT-

GMM method (14) had 80.09% classification accuracy. In order to have a fair comparison between the features extracted from time-frequency analysis and our method, we used the same supervised classifier method (Random Forest classifier) in conjunction with the STFT derived features. We called this method STFT-RF and it reached 82.3% classification accuracy, which was significantly lower than the classification accuracy of our proposed method. These results suggest that more than the classifier, the features extracted with the sparse signal representation strategy plays a critical role in distinguishing between real and pseudo-HFOs (a comparison between the ROC curves of features extracted from time-frequency/spectral analysis and sparse representation is provided in supplementary figure 3). The classification accuracy of real and pseudo-HFOs for individual subjects is provided in figure 6(B). A box plot comparing the accuracy of each method is given in figure 6(C). We noted that the OMP-Gabor method was significantly better than the other techniques ($P < 0.001$ Wilcoxon signed-rank test between Gabor-OMP and other techniques). As shown in figure 6(B), in some cases (P-9, P-10, P-11, and P-13), the STFT-GMM method noticeably failed to distinguish between real and pseudo-HFOs (P-9: out of 236 pseudo-HFO events, 136 are misclassified as real-HFO, P-10: out of 217 pseudo-HFO events, 127 are misclassified as real-HFO, P-11: out of 202 pseudo-HFO events, 123 are misclassified as real-HFO, and P-13: out of 217 real-HFO events, 127 are misclassified as pseudo-HFO) while our proposed method had an accuracy above 90% (P-9: out of 236 pseudo-HFO events, 12 are misclassified as real-HFO, P-10: out of 217 pseudo-HFO events, 5 are misclassified as real-HFO, P-11: out of 202 pseudo-HFO events, 6 are misclassified as real-HFO , and P-13: out of 217 real-HFO events, 16 are misclassified as pseudo-HFO) across these subjects. Overall, the most significant improvement of the proposed method was in detection of pseudo-HFO with high accuracy among all subjects (4.2%, i.e., 142 false-positive events in total) compared to the GMM/STFT method (24.8%, i.e., 846 false-positive events in total).

Figure 7(A) shows the frequency distribution of selected atoms over iterations for real and pseudo-HFOs. We observed that the OMP algorithm picks atoms with lower frequencies for the first 10–15 iterations. However, as we go to the higher iterations, the algorithm picks atoms with higher frequencies from the R and FR bands. This trend is almost similar between real and pseudo-HFOs. However, for the pseudo-ones, the OMP starts to select the higher frequency atoms in earlier iterations. This is likely that many pseudo-HFOs have sharp changes which has either broad-band characteristics or reside in higher frequencies. Therefore, the OMP tends to pick more high-frequency atoms in the first few iterations for pseudo HFOs compared to real ones.

The features selected by RF over iterations are depicted in figure 7 (B, first row). We noticed that the RF picks approximation error feature between iterations 15–30 and line-noise feature around iteration 15. The V-Factor was selected at iterations above 22. The importance factor of features is also investigated in figure 7 (B, second row). It was observed that the most important iterations are ε_{18-22}, V_{28-32}, and L_{12-16} suggesting that the sparse solution is more critical for classification.

3.3 Spatial Distribution of HFO

After exploring the efficacy of our proposed method using labelled events, we further tested this method on continuous iEEG data to see whether removing pseudo-HFOs from the pool of events could improve the identification of HFO to SOZ. Figure 8(A) illustrates the distribution of HFOs over implanted electrodes for patients 1, 2, 5, and 8 before (upper row) and after (lower row) pseudo-HFO removal. The SOZ identified by the epileptologist is marked with red arrows for each patient. As shown in figure 8(A), the existence of noises and artifacts in the initial pool of HFO might be misleading. The amplitude-threshold-based detection results in P-2, 5, and 8 also identified some busy channels out of SOZ, which are pseudo-HFOs and removed from the pool of events by using the proposed method. Additionally, we demonstrated the changes in SOZ localization accuracy for each subject in figure 8(B). P-9 had the lowest improvement accuracy among all these patients with 1.75% SOZ localization improvement, while P-15 had the highest with a 37.67% SOZ localization improvement.

Across all sixteen subjects, 62.5% (10/16) had busy channels laden with artifacts out of SOZ in their initial pool of events. However, most of these pseudo-HFOs were eliminated from the pool after applying the proposed method. In the continuous iEEG data, a total of 171,665 events passed the initial detector. Before the removal of pseudo-HFOs, 57.7% of the initially detected events were localized in the SOZ. After applying our method, 34.51% (59,242) of events were removed from the initially detected pool. Of these events, 75.6% (44,814) were out of the SOZ, and 24.4% (14,428) were from the SOZ. After removing pseudo-HFOs, 75.3% (84,630) of the remaining events were in the SOZ. As shown in figure 8(C), compared to the initial 57.7% accuracy, the proposed method showed a significant 17.6% improvement in the SOZ localization accuracy ($P < 0.001$ Wilcoxon signed-rank test).

3.4 HFO Representation and Spatial Distribution of Ripples and Fast Ripples

To assess the SOZ localization accuracy of R and FR separately, those events that are classified as real HFOs were further categorized into R and FR groups based on the frequency of atoms used in their adaptive representation as described in section 2.2.5. Using the adaptive sparse representation, we were able to find all high-frequency components in HFOs that pass the initial detector. We then used the highest frequency of the selected atoms to create a two-dimensional distribution of events over different frequencies and channels (figure 9(A)). Finally, we divided them into two categories (below 250 Hz as R and above 250 Hz as FR). We have provided the distribution of R and FR events for each subject and their SOZ localization accuracy in figure 9. Across all subjects, 24.5% of real HFO events were categorized as FR through our adaptive representation approach. In some subjects (P3, P-7, P-15, and P16), the number of FR was smaller than 3% of total HFOs (see figure 9(B)) and therefore discarded from the SOZ analysis. As shown in figure 9(C), we noticed that although the FR rate was noticeably lower than Rs, they were more specific to the SOZ channels than the Rs (figure 9(C) and 9(D)). Specifically, SOZ accuracy of Rs (75.23% ± 16.46) was significantly lower than FRs (82.81% ± 17.24) shown in figure 9(D).

4. Discussion

Prolonged iEEG recordings obtained in the clinical setting frequently are contaminated with different noises and artifacts (7,8,9). Generally, HFO detectors employ either high-pass or band-pass filtering. Numerous studies (11,12,13) have shown that sharp artifacts or random noise with arbitrary non-sinusoidal waveshape might mimic HFO after high pass filtering. To date, most studies on HFOs in epilepsy are based on highly human-intensive methods to extract the signals of interest from multichannel iEEG (20). Due to difficulties associated with the visual inspection of data over multiple channels and recordings lasting hours to days, investigators typically perform dramatic data reduction steps preceding the detailed analysis, such as pre-selecting electrodes or limiting the analysis to short segments of iEEG without any artifacts. Such data reduction techniques and the complexity associated with visual inspection limit the practical use of HFOs in clinical practice.

There is no standardized definition of pseudo-HFO events due to their random origins. They vastly differ among recordings due to the distinct data acquisition setting and environmental factors causing interference in the assessed signal. Various algorithms have recently been introduced to detect different types of pseudo-HFOs in iEEG data. Some of the early work utilized the comparison of the power spectrum of the background activity and HFO peak frequencies and comparison of HFO with respect to background activity (21). Gliske *et al.* removed non-neural artifacts, including fast transient DC shifts and noise appearing jointly across channels (22). Their recent work relied on extracting statistical features and fuzzy clustering to detect spurious HFOs (23). They also focused on removing pseudo HFOs resulting from EMG artifacts (24), which are biased toward a particular type of noise. Recent studies investigated time-frequency analysis of HFOs fused with an unsupervised clustering method (14) and convolutional neural network (CNN) (25,26) and reported a dramatic increase in HFO identification accuracy compared to straightforward filtering. However, in such methods exploring the time-frequency (TF) content of the signal with short-time-Fourier-transform, continuous wavelet transforms, or Stockwell transform might cause pseudo-HFOs to still remain in the pool of HFO waveforms and affect the SOZ localization accuracy. Specifically, in previous publications (27,28,29,30,14), including the authors' own work (14), the real HFO is assumed to be an oscillatory pattern that appears as an isolated region in time-frequency representation. However, as we have visualized in Fig. 4, the non-sinusoidal artifacts might appear as isolated regions in TF maps and frequently get classified as true HFOs. During the annotation stage of events, although the TF domain representation of some of the artifacts was similar to the real HFOs, we noted that through visual inspection of the raw events, human investigators could easily recognize these events as artifacts suggesting that waveform morphology plays a critical role. It is likely that during mapping the raw 1D signal into a 2D time-frequency map, essential characteristics of the signal, such as its morphology, are lost.

This study introduces a novel sparse signal representation approach founded on the hypothesis that real HFOs can be represented by a linear weighted combination of a small number of oscillatory time-frequency atoms selected from a redundant dictionary. Unlike real HFOs, pseudo-HFOs have either an arbitrary nature in the form of white noise or are very sharp, which are difficult to represent with a small number of oscillatory bases.

Consequently, over several iterations, the approximation error of these events was larger than those of real HFOs. Furthermore, we noted that artifacts with sharp spiking transients could not be represented locally with our predefined oscillatory Gabor atoms. In those cases, where the pseudo-HFOs were represented globally but not locally, their approximation error was similar to real HFOs. To overcome this limitation, we introduced the V-Factor, which computes the local error ratio to the global error. Therefore, a very localized artifact is expected to have a large V-Factor. This feature is particularly good for classifying sharp tiny non-oscillatory artifacts that might pass the amplitude-threshold-based detector.

Finally, the idea of reconstructing all initially detected HFOs in an iterative fashion using a predefined oscillatory dictionary allowed us to see any monotonic activity such as line noise that has a component at either 50 or 60 Hz. Therefore, we could detect corrupted events with persistent line noise using the number of selected atoms at a specific frequency (50 or 60 Hz).

The idea of sparse representation of HFOs was first introduced by *Bénar et al.* (12). However, some challenges restricted them from using the OMP-based sparse representation in HFO reconstruction. The main limitation of the process (12) was that one could generally control the sparsity levels using two different parameters, including approximation error and the number of atoms. Setting both parameters to a fixed number a priori might be problematic as, across different subjects, the background noise level might be different, and the number of HFO components can also vary. Therefore, in the pseudo vs. real HFO discrimination scenario, we let the RF select the most discriminative approximations over 50 iterations to distinguish between the two classes. We visualized the frequency distribution of selected atoms over iterations for each class in figure 7(A). In figure 4 of supplementary material, we also visualized the number of selected atoms from low, ripple and fast ripple bands. As expected, since the iEEG activity follows 1/f power spectrum, the initial atoms were generally selected from the lower frequency (<80Hz). As the iterations increased (>10), the atoms were selected not only below but also from frequencies above 80Hz for the real-HFO events. Although rare, for the pseudo-HFO events, atoms above 80Hz were used within the first 5–10 iterations. But once again, the initial atoms were generally selected from a frequency range below 80Hz. As the high frequency components vary from event to event (and also from subject to subject) in real-HFO category, no particular frequency above 80Hz was a clear determinant in discrimination. Therefore, we extracted the approximation error and V-factor as features without having any bias to the frequency of the selected dictionary atoms. We also observed that the approximation error between iterations 15–30 are the most important features (see figure 7(B)). As we expected, because of the concentration of energy in low band in raw iEEG, the OMP method starts to reconstruct components with higher energy in the low band first. As we move to the higher iterations, the algorithm starts to pick higher frequencies. As we showed in Figure 7(B), the sweet-spot for approximation error is between iterations 15–30, for V-factor after iteration 20, and line noise feature between iteration 10–15. These iterations generally correspond to the stage when the OMP method starts to represent the high frequency component of the events.

The evaluation of classification accuracy of real and pseudo-HFOs recorded from different types of electrodes shows no significant difference between these two electrodes (supplementary table 3). The result confirms that the proposed method is not biased towards one type of electrode over another.

Later, in order to further categorize real HFOs into Rs and FRs, we defined a new adaptive solution (see figure 3) to reconstruct each event as long as the residual signal passed the initial amplitude-based detector in the R and FR range over iterations. We used the highest frequency indices of selected atoms to assign an event into the R and FR categories. It should be noted that by using a fixed level (either approximation error or the number of approximations) to control sparsity level, one might either fit the background noise by using unnecessary atoms residing in the R or FR range or might use an insufficient number of atoms to reconstruct the R or FR component(s). However, our sparse/regenerative solution made it possible to use different numbers of atoms for different events and prevent overfitting the background noise. Furthermore, it was shown that, regardless of the existence of wideband HFOs, the solution to reconstruct high-frequency bands is sparse (supplementary figure 4 (B)). By inspecting the frequency content of the selected atoms, we could further categorize real HFOs as R and the FR events. We noticed that although the FR rate is lower than the R, the SOZ localization accuracy of FRs was significantly higher than the Rs.

The computational time for the whole process shows this method is also fast enough (0.09 s per event) to be applied for prolonged iEEG data (more information regarding the mean computational time exists in supplementary table 2).

Although the specificity of HFO to SOZ is critical, the type of surgical procedure and post-surgical outcome are important endpoints in HFO related studies. Out of 16 patients used in this study, 2 of them did not proceed with any surgical intervention (P5, and P9). Out of the remaining 14 patients, 10 of them had Engel class 1, 3 with Engel class 2, and 1 with Engel class 3 (supplementary table 1). In P10 and P13, RNS (responsive neurostimulation) was used as surgical therapy (Engel class 2 and 3 respectively). In all other cases, surgical therapy was the resection of electrode sites at target regions. A complete statistical analysis of the remaining HFOs in resection volume is proposed in supplementary figure 5. In all cases, the proposed method improves the localization of resected region of brain except P16. However, evaluation of the raw iEEG of P16 reveals that this subject had a noisy recording with only a small amount of non-artifact laden iEEG recording available at the selected interictal periods from the electrodes of concern. Thus, the percentage of HFOs falling into the resection volume of these subjects decreased (from 67.75% to 64.73%) after applying the proposed method.

5. Conclusion

This study introduces a sparse signal representation-based approach to detect the pseudo-HFO events in long-term iEEG recordings. Instead of biasing toward a specific type of noise or using conventional time-frequency features, we show that real HFOs, unlike pseudo-HFOs, can be represented with a limited number of oscillatory components. The algorithm successfully discriminated real from pseudo-HFOs with an accuracy of 96.65% and further classified real HFOs into R and FR categories using an adaptive sparse representation strategy.

Compared to our previous work (14), the proposed sparse method surpassed the STFT-GMM fusion in classification accuracy (96.65% vs. 80.09% respectively) for two reasons: First, the STFT-GMM uses the time-frequency plane features that are blinded to the waveform morphology. Moreover, the GMM process is an unsupervised clustering method. In other words, GMM requires large data for the reliable estimation of the density/ distribution of clusters in an automated fashion. If the fraction of pseudo-HFOs vs. real HFOs is small, then captured pseudo-HFO events might be insufficient to form a cluster and could get merged with real HFOs.

One of the limitations of this work is the lack of physiological and pathological HFO discrimination using sparse representation. Earlier work showed that HFOs could also be found in the healthy brain (31,32,33,34). These oscillatory events, so-called physiological HFOs, are generated in functional tissue and might be confused with epileptogenic HFO so-called pathological HFO. Although different studies have been performed to understand the mechanism of physiological HFOs (35,36,37) and the differences between pathological and physiological oscillations (34,38,39,40,41,42,43,44), the complete answer is lacking, and this topic is still an open question for further research. One suggestion is to detect stereotypical HFOs (39) as pathological events. However, the purpose of this study is to remove pseudo-HFOs, and we did not focus on the discrimination of physiological and pathological HFOs.

Moreover, one of the limitations of the study is that since we employed a binary classification scheme, all events including ambiguous ones are mapped either to the real or pseudo-HFO categories by RF model. In a future work, rather than being binary, the inter-rater variability can be incorporated into the study and one can assign a variable that represents the confidence of the experts with their annotation and the model might produce a decision that associated with the probability of an event being a real or pseudo-HFO. This can be helpful to see the difference between more obvious pseudo-HFO examples and difficult ones.

Finally, a future application of this study could be to extend the analysis to the iEEG recorded in the operating room during electrode implantation. Due to the existence of different devices in surgery, data recorded intraoperatively might be corrupted with a massive amount of noise and artifacts. Since there is no need to wait for the accumulation of a large number of events to build clusters as in *Liu et al.* (14) work, this new sparse method can be executed in a real-time fashion intraoperatively over brief recording periods towards early detection of the SOZ before the prolonged monitoring.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We cordially give our thanks to sixteen patients, the neurological teams at the University of Minnesota, Texas Children's hospital, Baylor St. Luke's Medical center, and Istanbul University, who made this research possible. We are also grateful to all resources provided by TIMES.

Funding

This study was supported by grants R01NS112497 (NFI) and UH3NS117944 (NFI, JG) from the National Institutes of Health—National Institute of Neurological Disorders and Stroke.

References

1. Hauser WA, and Kurland LT. The epidemiology of epilepsy in Rochester, Minnesota, 1935 through 1967. Epilepsia. 1975; 16(1): 1–66. [PubMed: 804401]
2. Shorvon JW Sander SD Epidemiology of the epilepsies. Journal of Neurology, Neurosurgery and psychiatry. 1996; 61(5): 433–443. [PubMed: 8965090]
3. Habibi M Refractory Epilepsy. U.S Pharmacist. 2009; 34(3): 8–14.
4. Najm I, Jehi L, Palmini A, Gonzalez-Martinez J, Paglioli E, Bingaman W. Temporal patterns and mechanisms of epilepsy surgery failure. Epilepsia. 2013; 54(5): 772–782. [PubMed: 23586531]
5. Engel J, Pitkänen A, Loeb JA, Dudek FE, Bertram EH, Cole AJ, Moshé SL, Wiebe S, Jensen FE, Mody I, Nehlig A, and Vezzani A. Epilepsy Biomarkers. Epilepsia. 2013; 54: 61–69.
6. Worrell GA, Jerbi K, Kobayashi K, Lina JM, Zelmann R, Le Van Quyen M. Recording and analysis techniques for high-frequency oscillations. Progress in Neurobiology. 2012; 98(3): 265–278. [PubMed: 22420981]
7. Kovach CK, Tsuchiya N, Kawasaki H, Oya H, Howard MA, and Adolphs R. Manifestation of ocular-muscle EMG contamination in human intracranial recordings. NeuroImage. 2011; 54(1): 213–233. [PubMed: 20696256]
8. Tuyisenge V, Trebaul L, Bhattacharjee M, Chanteloup-Forêt B, Saubat-Guigui C, Mîndruţă I, Rheims S, Maillard L, Kahane P, Taussig D, and David O. Automatic bad channel detection in intracranial electroencephalographic recordings using ensemble machine learning. Clinical NeuroPhysiology. 2018; 129(3): 548–554. [PubMed: 29353183]
9. Ball T, Kern M, Mutschler I, Aertsen A, and Schulze-Bonhage A. Signal quality of simultaneously recorded invasive and non-invasive EEG. NeuroImage. 2009; 46(3): 708–716. [PubMed: 19264143]
10. Islam MK, Rastegarnia A, and Sanei S. Signal Artifacts and Techniques for Artifacts and Noise Removal. In Signal Processing Techniques for Computational Health Informatics.: Springer, Cham; 2020. p. 23–79.
11. Lee S, Issa NP, Rose S, Tao JX, Warnke PC, Towle VL, van Drongelen W, & Wu S DC shifts, high frequency oscillations, ripples and fast ripples in relation to the seizure onset zone. Seizure. 2020; 77: 52–58. [PubMed: 31101405]
12. Bénar CG, Chauvière L, Bartolomei F, and Wendling F. Pitfalls of high-pass filtering for detecting epileptic oscillations: A technical note on. Clinical Neurophysiology. 2010; 121(3): 301–310. [PubMed: 19955019]
13. Thomschewski A, Hincapié A-S, and Frauscher B. Localization of the Epileptogenic Zone Using High Frequency Oscillations. Frontiers in Neurology. 2019; 10.
14. Liu S, Sha Z, Sencer A, Aydoseli A, Bebek N, Abosch A, Henry T, Gurses C, & Ince NF Exploring the time–frequency content of high frequency oscillations for automated identification of seizure onset zone in epilepsy. Journal of Neural Engineering. 2016; 13(2): 026026.
15. Zhang SG Mallat and Zhifeng. Matching pursuits with time-frequency dictionaries. IEEE transaction on Signal Processing. 1993 December; 41(12): 3397–3415.
16. Pati YC, Rezaiifar R, and Krishnaprasad PS. Orthogonal Matching Pursuit: Recursive function approximation with applications to wavelet decomposition. Proceedings of 27th Asilomar Conference on Signals, Systems and Computers.. 1993.
17. "Pseudo-HFO Elimination with Sparse Signal Representation", GitHub. [Online].; 2022 [cited 2022 7. Available from: https://github.com/InceLab/JNE_Pseudo_HFO_Elimination.

18. Breiman L Random Forests. Machine Learning. 2001; 45(1): 5–32.
19. Erogol. GitHub. [Online].; 2021 [cited 2021 December 27. Available from: https://github.com/erogol/Random_Forests.
20. Blanco Justin A., Stead Matt, Krieger Abba, Viventi Jonathan, Marsh W. Richard, Lee Kendall H., Worrell Gregory A., and Litt Brian. Unsupervised Classification of High-Frequency Oscillations in Human Neocortical Epilepsy and Control Patients. Journal of neurophysiology. 2010 Nov; 104(5): 2900–2912. [PubMed: 20810694]
21. Cho JR, Koo DL, Joo EY, Seo DW, Hong SC, Jiruska P, & Hong SB Resection of individually identified high-rate high-frequency oscillations region is associated with favorable outcome in neocortical epilepsy. Epilepsia. 2014; 55(11): 1872–1873. [PubMed: 25266626]
22. Gliske SV, Irwin ZT, Davis KA, Sahaya K, Chestek C, Stacey WC. Universal automated high frequency oscillation detector for real-time, long term EEG. Clinical Neurophysiology. 2016 February; 127(2): 1057–1066. [PubMed: 26238856]
23. Gliske SV, Qin Z, Lau K, Alvarado-Rojas C, Salami P, Zelmann R, & Stacey WC Distinguishing false and true positive detections of high frequency oscillations. Journal of Neural Engineering. 2020; 17(5): 056005.
24. Ren S, Gliske SV, Brang D, and Stacey WC. Redaction of false high frequency oscillations due to muscle artifact improves specificity to epileptic tissue. Clinical Neurophysiology. 2019; 130(6): 976–985. [PubMed: 31003116]
25. Lai D, Zhang X, Ma K, Chen Z, Chen W, Zhang H, Yuan H, and Ding L. Automated detection of high Frequency oscillations in Intracranial EEG using the combination Of Short-time energy and convolutional neural networks. IEEE Access. 2019; 7: 82501–82511.
26. Nejedly P, Cimbalnik J, Klimes P, Plesinger F, Halamek J, Kremen V, Viscor I, Brinkmann BH, Pail M, Brazdil M, Worrell G, and Jurak P. Intracerebral EEG Artifact identification using convolutional neural networks. Neuroinformatics. 2018; 17(2): 225–234.
27. Fedele Tommaso, van ‘t Klooster Maryse, Burnos Sergey, Zweiphenning Willemiek, van Klink Nicole, Leijten Frans, Zijlmans Maeike, Sarnthein Johannes. Automatic detection of high frequency oscillations during epilepsy surgery predicts seizure outcome. Clinical Neurophysiology. 2016; 127(9): 3066–3074. [PubMed: 27472542]
28. Amiri M, Lina J-M, Pizzo F, and Gotman J. High frequency oscillations and spikes: Separating real hfos from false oscillations. Clinical Neurophysiology. 2016; 127(1): 187–196. [PubMed: 26100149]
29. Baotian Zhao, Wenhan Hu, Chao Zhang, Xiu Wang, Yao Wang, Chang Liu, Jiajie Mo, Xiaoli Yang, Lin Sang, Yanshan Ma, Xiaoqiu Shao, Kai Zhang, Jianguo Zhang. Integrated automatic Detection, classification and imaging of high Frequency oscillations With Stereoelectroencephalography. frontiers in Neuroscience. 2020; 14.
30. Donos C, Mîndruţă I, and Barborica A. Unsupervised detection of high-frequency oscillations using time-frequency maps and computer vision. Frontiers in neuroscience. 2020; 14.
31. Kucewicz Michal T., Cimbalnik Jan, Matsumoto Joseph Y., Brinkmann Benjamin H., Bower Mark R., Vasoli Vincent, Vlastimil Sulc, Fred Meyer WR. Marsh SM. Stead, Worrell Gregory A.. High frequency oscillations are associated with cognitive processing in human recognition memory. Brain. 2014 August; 137(8): 2231–2244. [PubMed: 24919972]
32. Matsumoto A, Brinkmann BH, Matthew Stead S, Matsumoto J, Kucewicz MT, Marsh WR, Meyer F, & Worrell G Pathological and physiological high-frequency oscillations in focal human epilepsy. Journal of neurophysiology. 2013; 110(8): 1958–1964. [PubMed: 23926038]
33. Nagasawa T, Juhász C, Rothermel R, Hoechstetter K, Sood S, & Asano E Spontaneous and visually driven high-frequency oscillations in the occipital cortex: intracranial recording in epileptic patients. Human brain mapping. 2012; 33(3): 569–583. [PubMed: 21432945]
34. Cimbalnik J, Brinkmann B,Kremen V,Jurak P, Berry B JV Gompel M. Stead G. Worrell. Physiological and pathological high frequency oscillations in focal epilepsy. Annals of Clinical and Translational Neurology. 2018; 5(9): 1062–1076. [PubMed: 30250863]
35. Jefferys JG, Menendez de la Prida L, Wendling F, Bragin A, Avoli M, Timofeev I, & Lopes da Silva FH Mechanisms of Physiological and epiletic HFO generation. Progress in neurobiology. 2012; 98(3): 250–264. [PubMed: 22420980]

36. Gulyás AI, & Freund TT Generation of physiological and pathological high frequency oscillations: the role of perisomatic inhibition in sharp-wave ripple and interictal spike generation. Current opinion in neurobiology. 2015; 31: 26–32. [PubMed: 25128735]

37. Alvarado-Rojas C, Huberfeld G, Baulac M, Clemenceau S, Charpier S, Miles R, de la Prida LM, & Le Van Quyen M Different mechanisms of ripple-like oscillations in the human epileptic subiculum. Annals of neurology. 2015; 77(2): 281–290. [PubMed: 25448920]

38. Pail M, Cimbálník J, Roman R, Daniel P, Shaw DJ, Chrastina J, & Brázdil M High frequency oscillations in epileptic and non-epileptic human hippocampus during a cognitive task. Scientific reports. 2020; 10(1): 18147. [PubMed: 33097749]

39. Liu S, Gurses C, Sha Z, Quach MM, Sencer A, Bebek N, Curry DJ, Prabhu S, Tummala S, Henry TR, & Ince NF Stereotyped high-frequency oscillations discriminate seizure onset zones and critical functional cortex in focal epilepsy. Brain : a journal of neurology. 2018; 141(3): 713–730. [PubMed: 29394328]

40. Bruder JC, Dümpelmann M, Piza DL, Mader M, Schulze-Bonhage A, & Jacobs-Le Van J Physiological Ripples Associated with Sleep Spindles Differ in Waveform Morphology from Epileptic Ripples. International journal of neural systems. 2017; 27(7).

41. Minthe Annika, Janzarik Wibke G., Lachner-Piza Daniel, Reinacher Peter Christoph, Schulze-Bonhage Andreas, Dümpelmann Matthias and Jacobs Julia. Stable high frequency background EEG activity distinguishes epileptic from healthy brain regions. Brain Communications. 2020; 2.

42. Nonoda Y, Miyakoshi M, Ojeda A, Makeig S, Juhász C, Sood S, & Asano E Interictal high-frequency oscillations generated by seizure onset and eloquent areas may be differentially coupled with different slow waves. Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology. 2016; 127(6): 2489–2499. [PubMed: 27178869]

43. Frauscher B, von Ellenrieder N, Ferrari-Marinho T, Avoli M, Dubeau F, & Gotman J Facilitation of epileptic activity during sleep is mediated by high amplitude slow waves. Brain: a journal of neurology. 2015; 6: 138.

44. Frauscher B, von Ellenrieder N, Zelmann R, Rogers C, Nguyen DK, Kahane P, Dubeau F, & Gotman J High-Frequency Oscillations in the Normal Human Brain. Annals of neurology. 2018; 84(3): 374–385. [PubMed: 30051505]

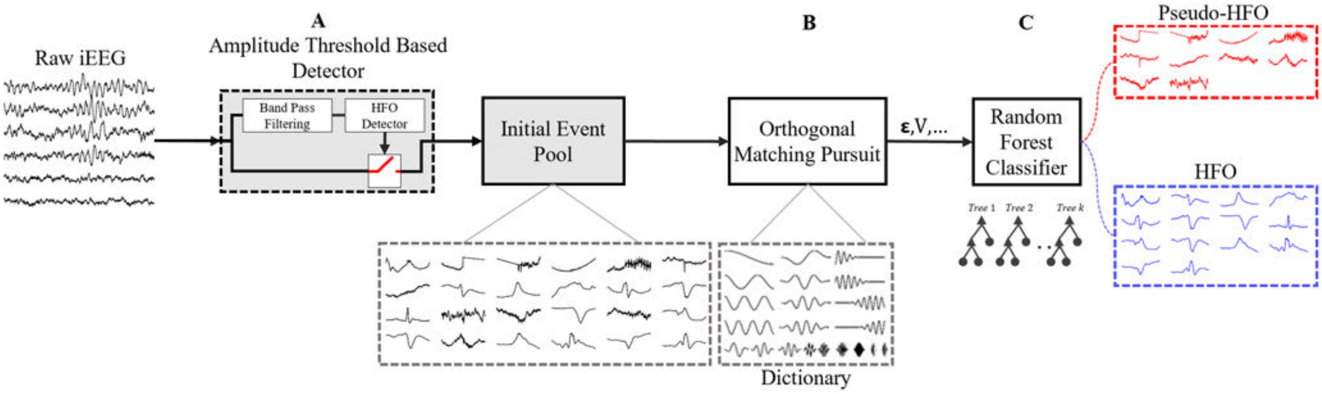

Figure 1.

Schematic diagram of the proposed method. (A) The raw iEEG data first goes into the amplitude-threshold-based HFO detector which implements sub-band filtering (in ripple and fast ripple bands separately). The raw event will be captured if the HFO detector accepts that event and it will be placed into the initial pool of HFOs including both real and pseudo-ones. (B) All initially detected events enter the OMP algorithm, and they are reconstructed using the oscillatory atoms selected from a redundant multi-resolution dictionary. Features $(\varepsilon, V, \ldots)$ are extracted at each iteration of this representation process from the residual signal and the selected atoms. (C) The binary classification is performed using a random forest classifier, and the pseudo-HFOs (red) are eliminated from the pool of candidate events.

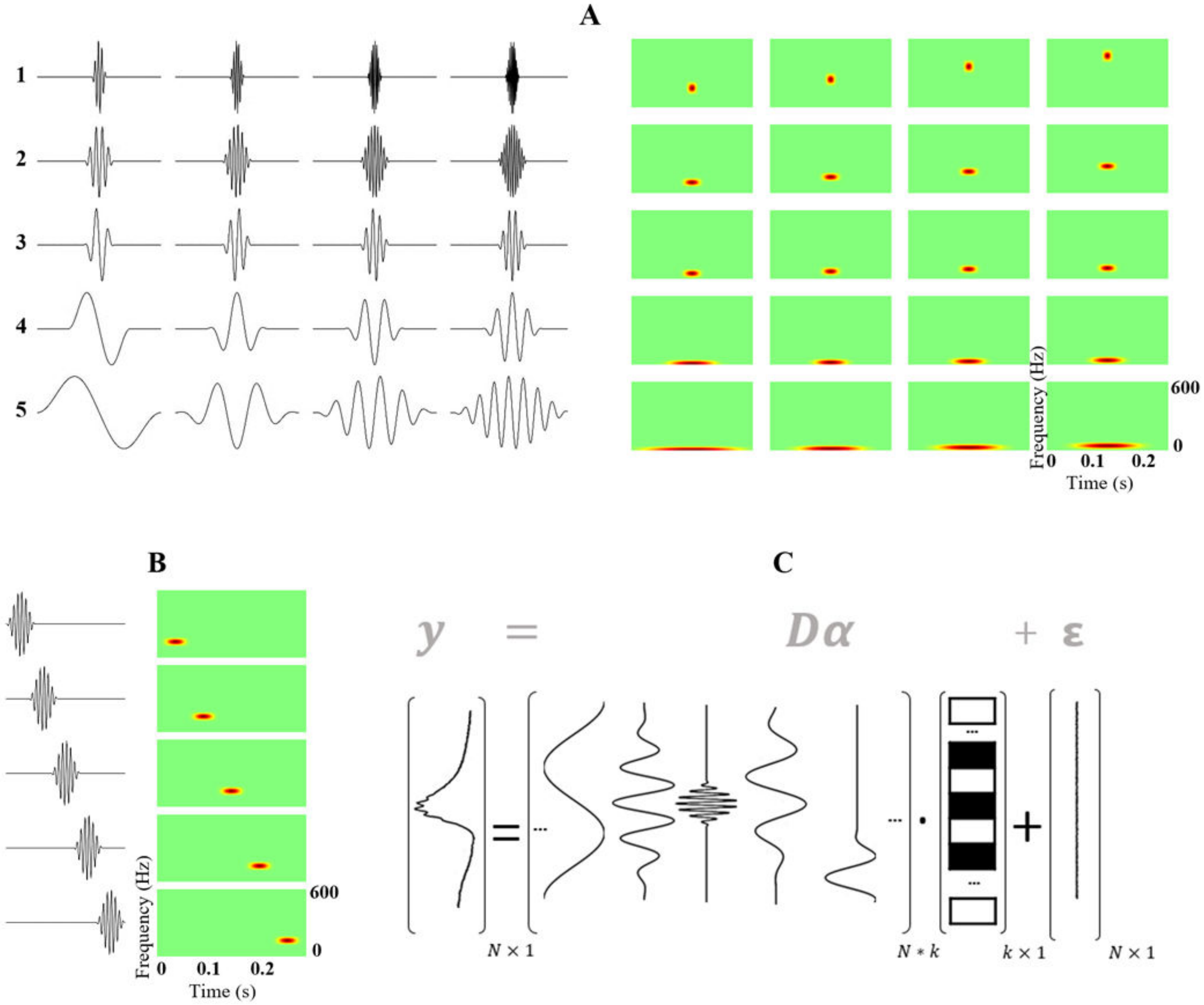

Figure 2.

Section (A) shows examples of analytical dictionary atoms and their corresponding Wigner-Ville transform. Line 1–5 represents the different dyadic dictionary $\left(\sigma = 2^{-j}\right)$ starting from $j = 4$ to $j = 0$. Section (B) represents an HFO atom with different shifts in our dictionary. (C) A schematic description of sparse representation using the Gabor atoms in dictionary. The HFO event y is represented by the combination of a small number of dictionary atoms where their indices are coded in sparse vector α. The nonzero entries are marked by black rectangles, while white rectangles present the zero coefficient. The residual signal is represented with r.

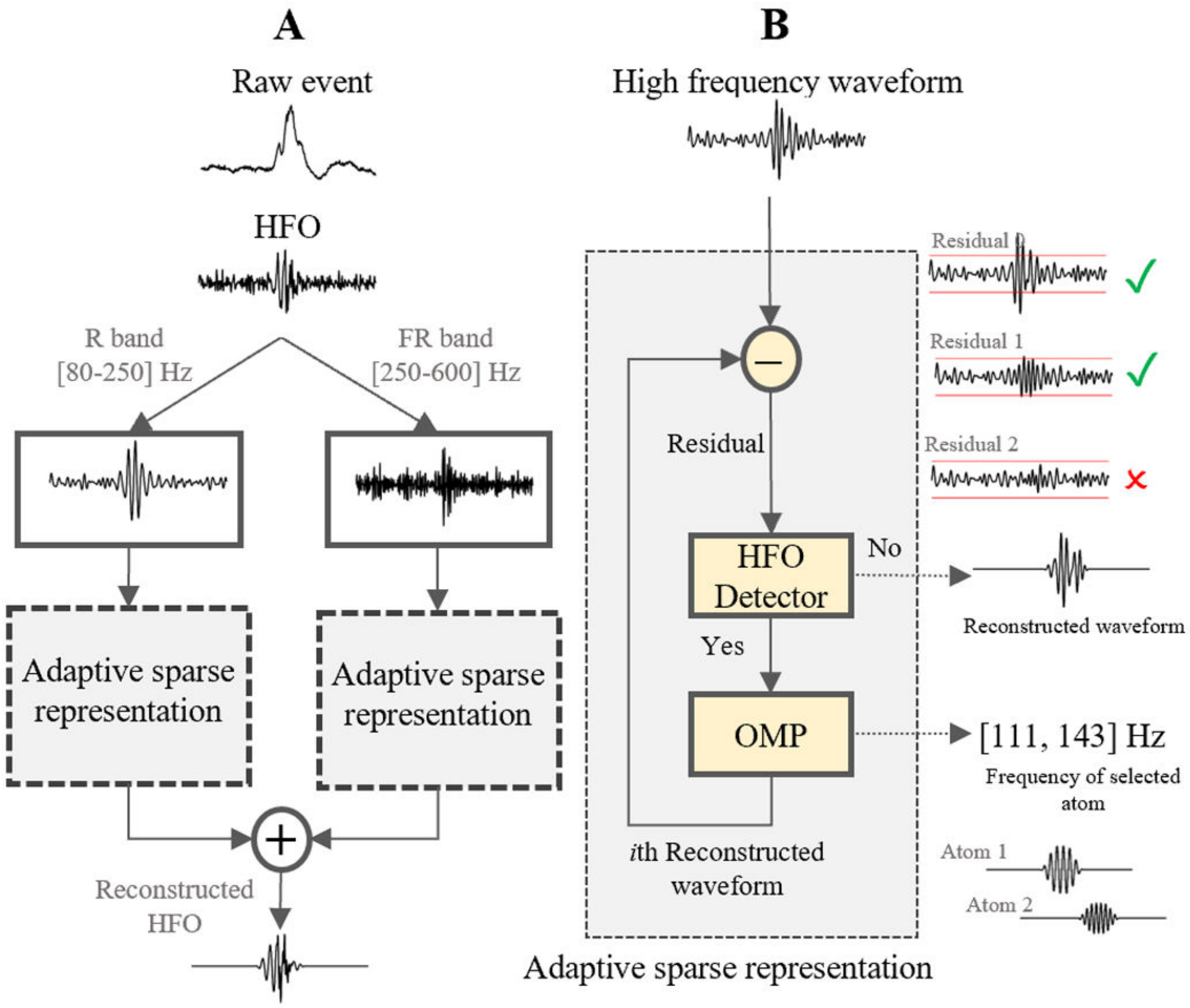

Figure 3.

The schematic diagram of HFO representation using the OMP procedure. (A) The HFO event divides into two different bands: ripple band (80:250 Hz) and fast ripple band (250:600 Hz). We reconstruct the ripple and fast ripple waveforms separately using adaptive sparse representation (gray box). Finally, we find the reconstructed HFO events by merging these two waveforms. (B) Adaptive sparse representation: The HFO waveforms (R and FR bands) go through a regenerative process to find the real HFO components. We continue to reconstruct these filtered bandpass events as long as the residual waveform satisfies the amplitude-threshold-detector’s constraints. The outputs of this box are the reconstructed waveform and the frequency of atoms used in the reconstruction process. The highest frequency of selected atoms will be used to divide the HFO events into the Rs and FRs.

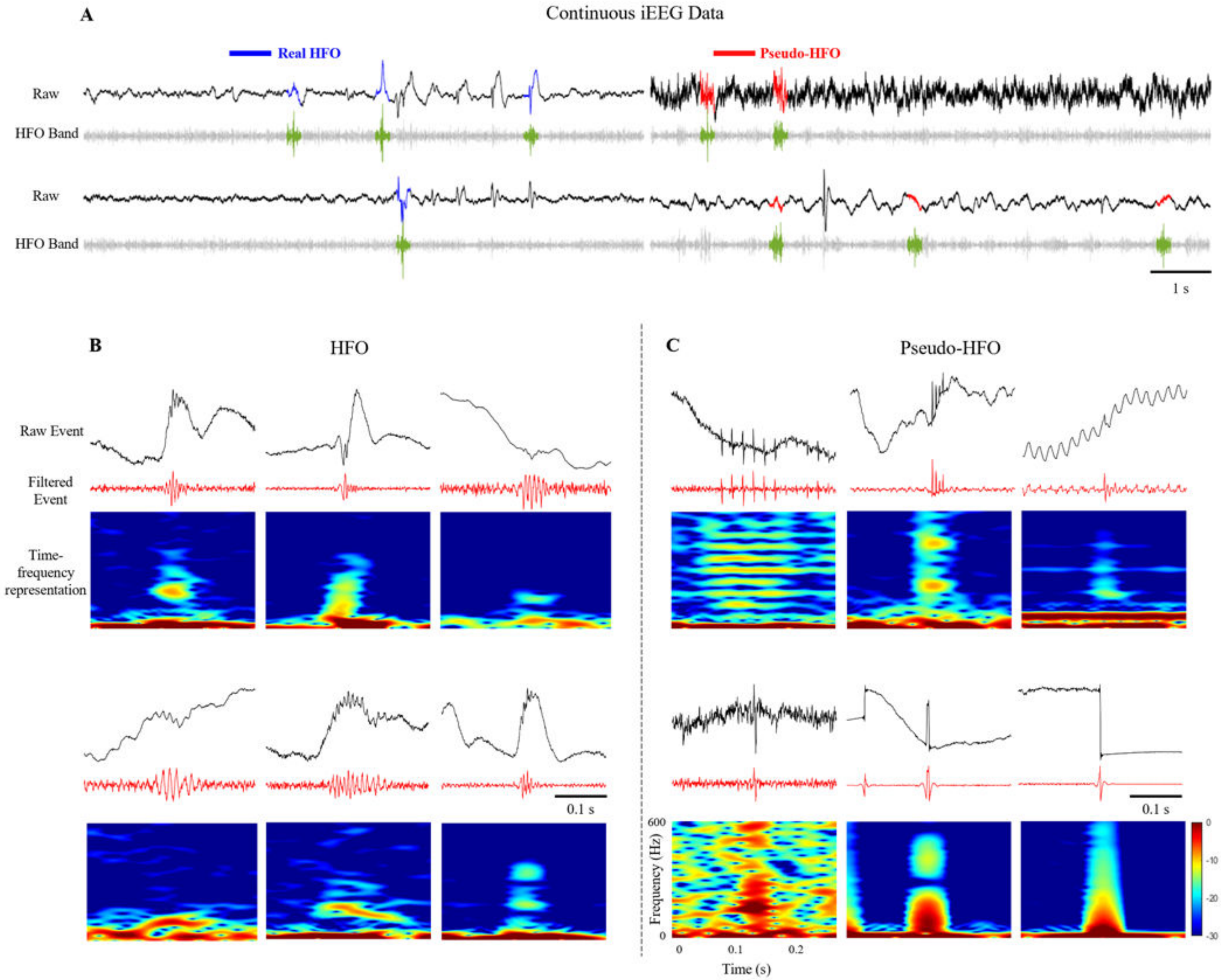

Figure 4.

Real HFO and pseudo-HFO/artifacts in an iEEG recording. (A) Black waveforms: continuous iEEG recording, grey waveform: continuous filtered data above 80 Hz, blue events: real HFO events, red events: pseudo-HFO events. (B, C) Examples of events captured by the initial HFO detector are shown in this figure. For each one, the raw event (black waveform), filtered event above 80 Hz (red waveform) using a 4th order Butterworth high pass filter, and the short-time-Fourier-transform of the raw event are shown.

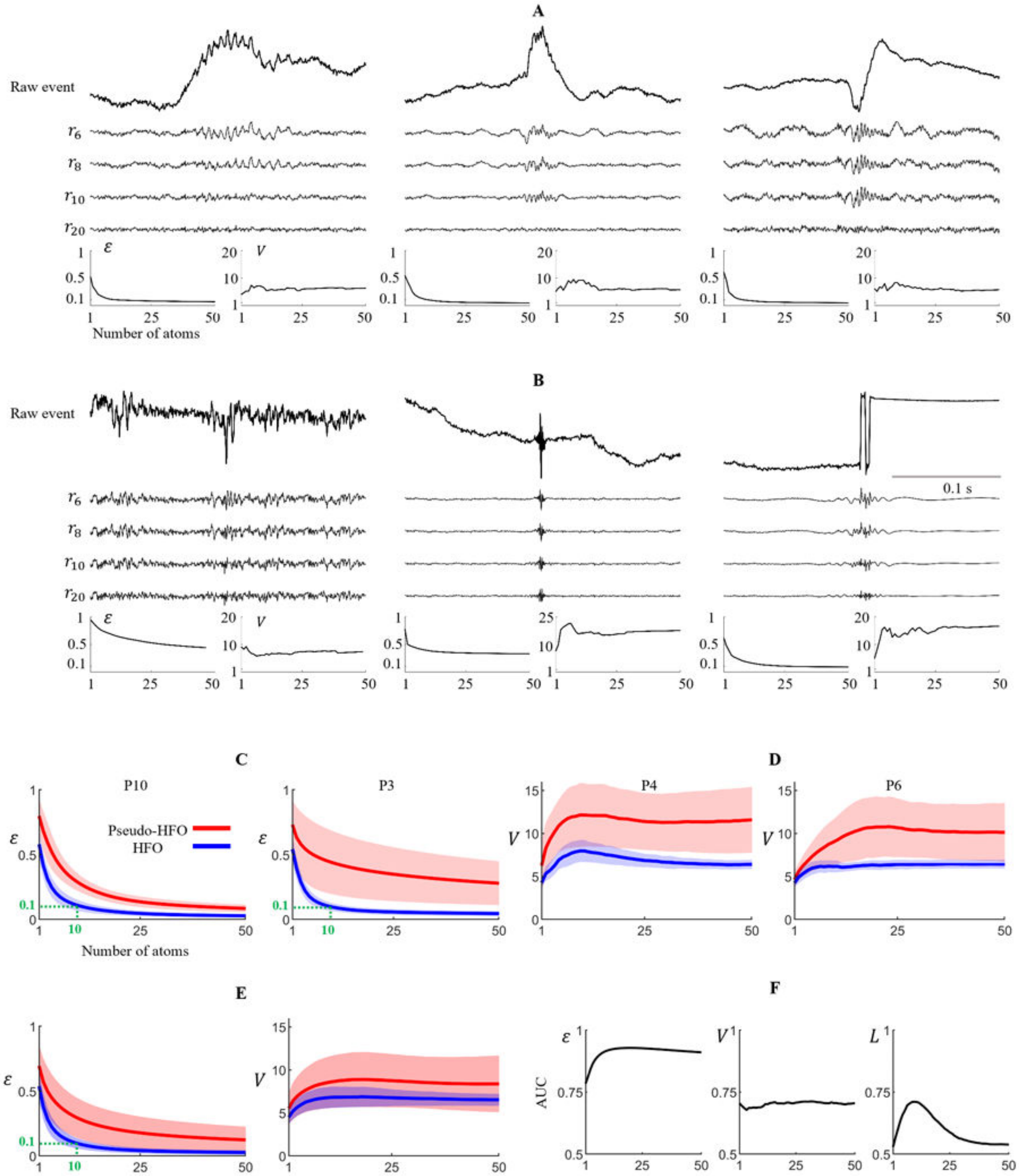

Figure 5.

Illustration of extracted features for HFO (A) and pseudo-HFO (B) examples with their corresponding residual signal at different iterations (6, 8, 10, and 20) and the extracted approximation error, V-Factor features at iteration 1–50. (C) The shaded plot of the approximation error for P10 and P3. (D) The shaded plot of V-Factor for P4 and P6. In P3 and P10, approximation error is an important feature to distinguish between HFO and pseudo-HFO, while in P4 and P6, the V-Factor is a critical feature for classification. (E) The shaded plot of approximation error and V-Factor across all subjects at different sparsity levels. (F) Area under the curve (AUC) values of the approximation error, V-Factor, and line noise at different sparsity levels.

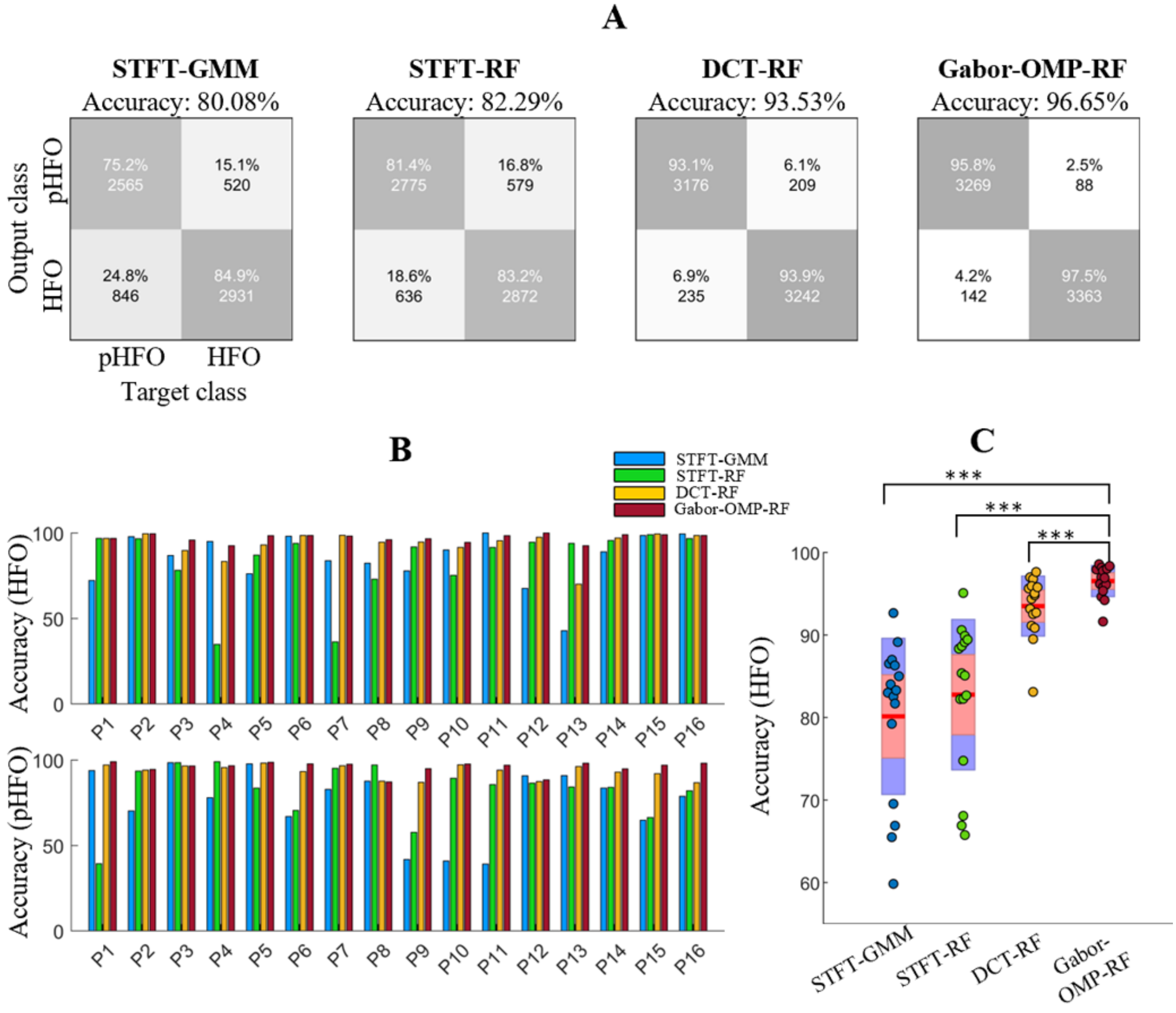

Figure 6.

(A) Comparison of the proposed method with STFT-GMM published by *Liu et al* (14), STFT-RF which is using RF classifier, and discrete cosine transform reconstruction. The proposed method had 96.65% classification accuracy, while the DCT-RF, STFT-RF, and GMM-STFT had 93.53%, 82.29%, and 80.08% classification accuracy, respectively. (B) Classification accuracy of HFO events (top) classification accuracy of pseudo-HFO events (bottom) across subjects. (C) Boxplot of the overall accuracy of different methods (all tests: p-value<0.001, Wilcoxon signed-rank test, *** indicates the significance level p<0.001).

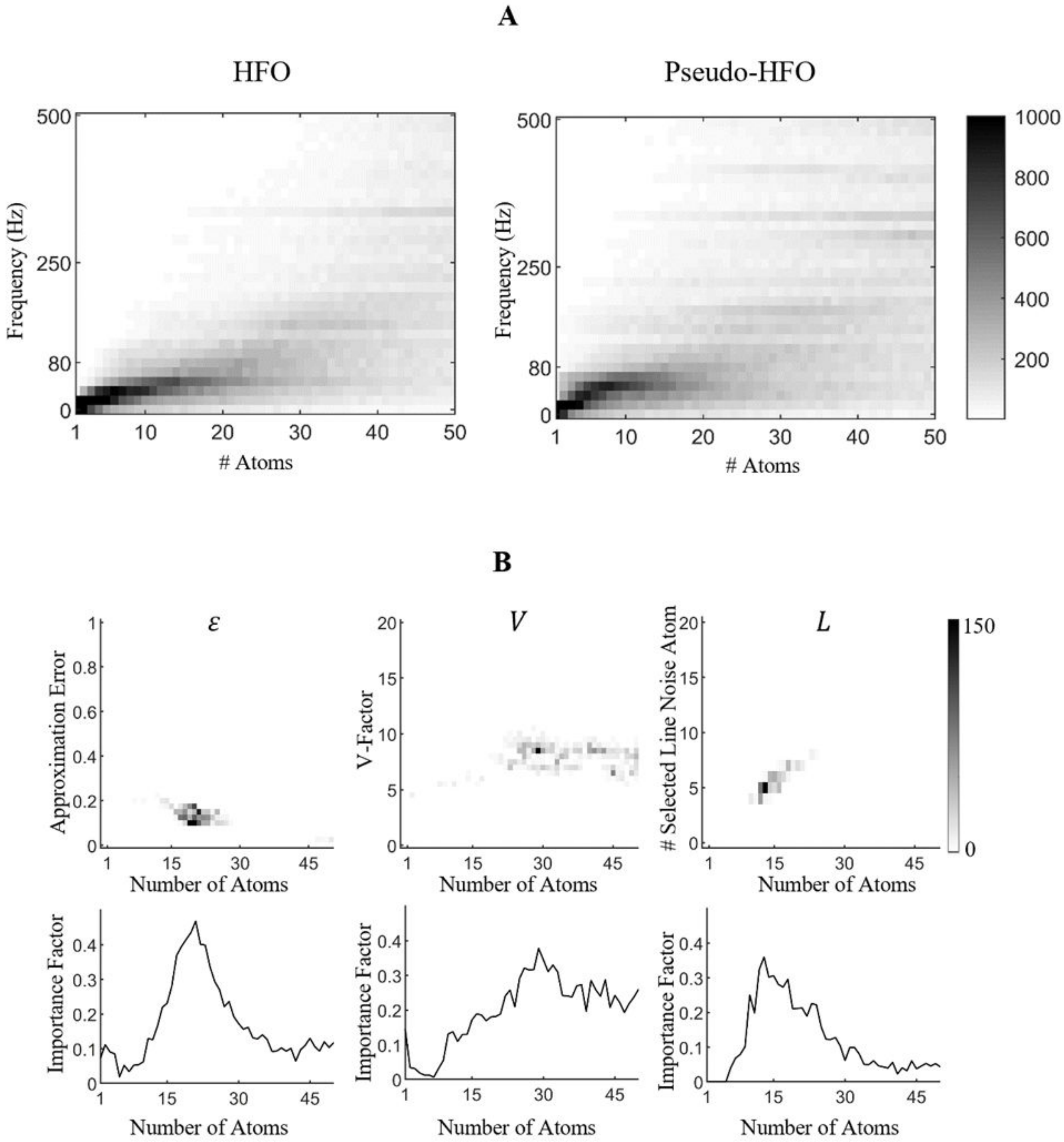

Figure 7.

(A) the frequency of selected atoms by OMP over iterations. (B, first row) Selected features by the random forest classifier for the discrimination of real vs. pseudo-HFOs over 50 iterations. The indices of selected features (x-axis) and their corresponding thresholds (y-axis) are shown for approximation error (ε), variation factor (V) and line noise (L). This figure illustrated the sparsity of approximation error and line noise features. (B, second row) The importance factor of each feature: Each figure represents the importance factor of one feature, i.e., approximation error, V-Factor, and line noise accordingly.

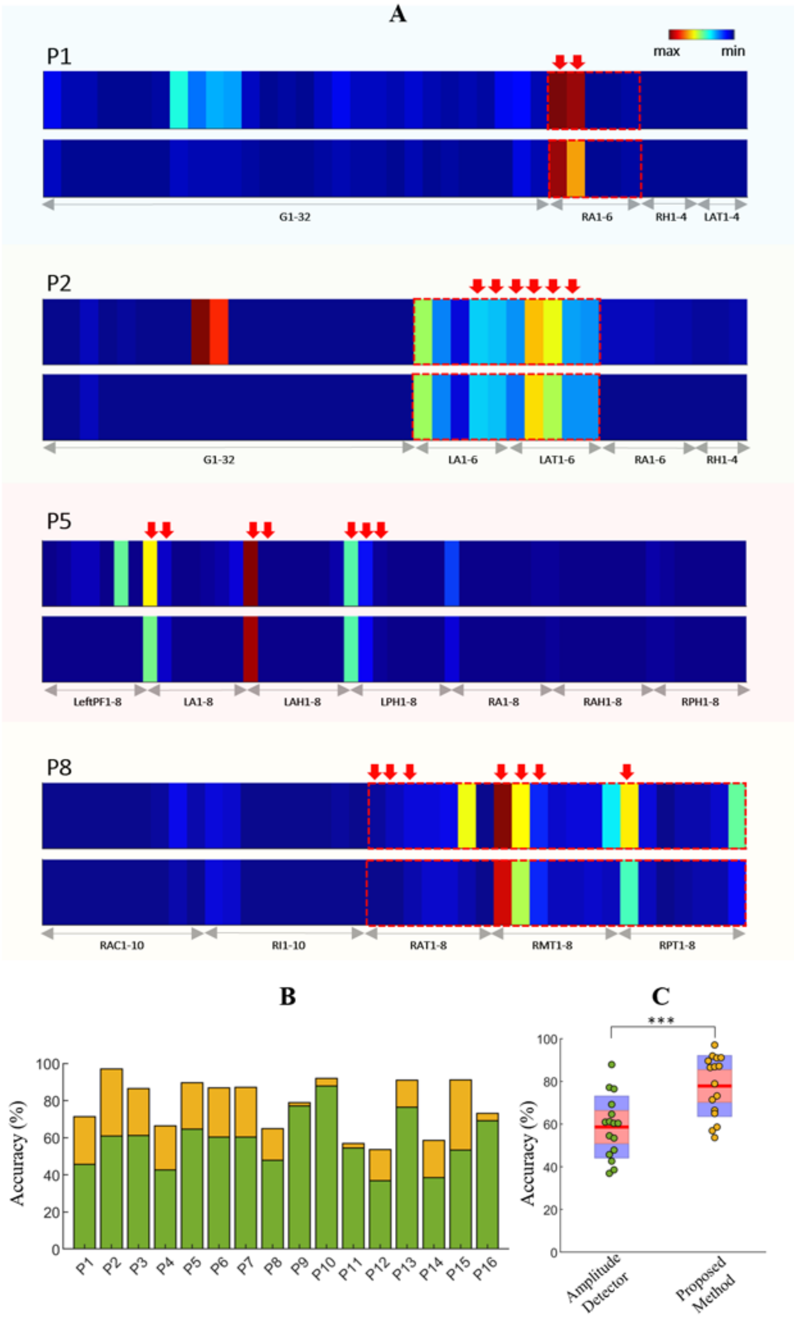

Figure 8.

(A) Spatial distribution of HFO before and after noise/pseudo-HFO removal using the proposed method. The first row of each figure represents the distribution of HFO detected by the initial detector. The second row represents the distribution of remaining events after the second step pseudo-HFO removal using the proposed method. For each patient, SOZ channels were pointed with a red arrow and resected volume (in case that patient went through the second surgery) represented by red dot-dashed boxes. HFO coming from each channel was color-coded based on the maximum amount of HFO from a single channel.

Channel labels (abbreviation) are noted beneath each subplot. (B) SOZ improvement across all subjects. The blue bar represents the SOZ accuracy of the HFO detector, and the added orange bar shows the improvement is obtained after pseudo-HFO removal using the proposed method. (C) Boxplot of the SOZ localization accuracy before and after using the proposed method. There is a significant improvement in SOZ accuracy across all subjects (p-value<0.001, Wilcoxon signed-rank test, *** indicates the significance level p<0.001).

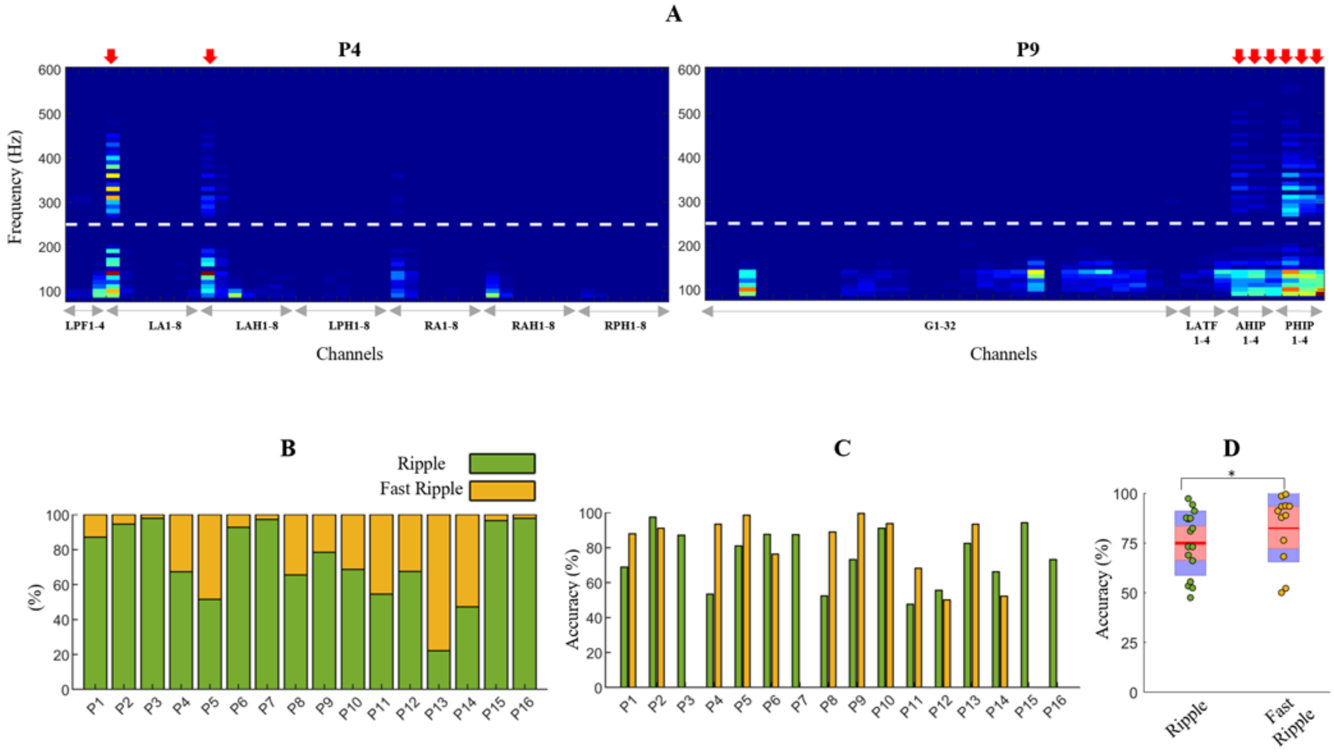

Figure 9.

(A) 2-Dimensional histogram of highest frequency components detected in HFOs passed the algorithm over channels. First column shows the distribution of HFO for P-4 and second column shows the distribution of P9. For each patient, SOZ channels were pointed with a red arrow. (B) The percentage of Rs and FRs in each patient. This bar plot shows that compared to Rs, FRs are more scarce events in iEEG recording, and in some cases, they might be smaller than 3% of the total events (P3, P7, P15, P16). (C) Bar plot of SOZ accuracy across each subject. As P3, P7, P15, and P16 did not have a sufficient amount of fast ripple, i.e., smaller than 3% of the total number of events, we removed them from the analysis. (D) Delineation of R and FR to SOZ localization (p-value<0.05 Wilcoxon signed-rank test, * indicates the significance level 0.01<p<0.05).