

# Supplemental Material

## 1. Patient Selection Protocol

We reviewed SEEG data from all patients who were evaluated with SEEG in Cleveland Clinic between 2009 and July 2014 ( $n = 280$ ). Patient selection workflow at the time of recruitment (January-February 2016) is shown in Fig. 1.

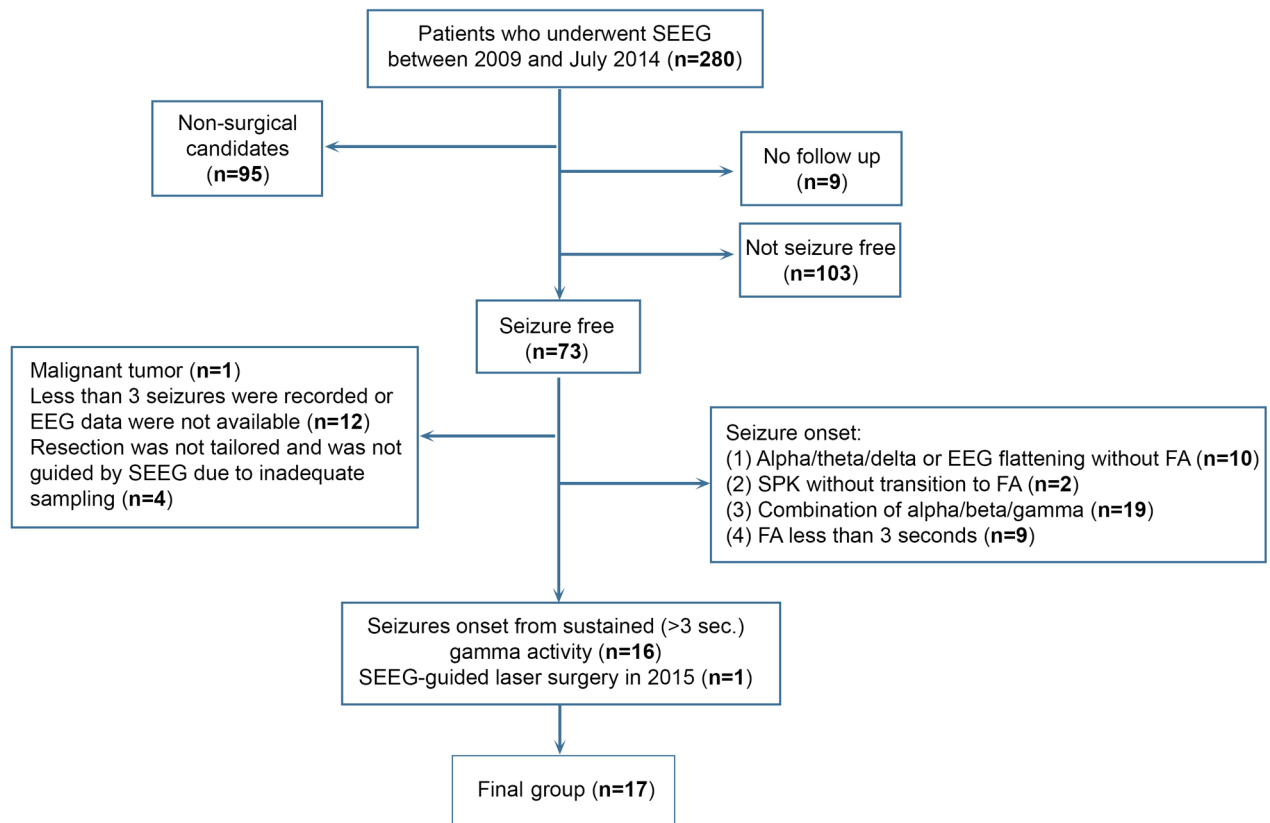
Our inclusion criteria were:

1. Tailored resection or laser ablation guided by SEEG  
95 patients were not surgical candidates after SEEG and were not included in the study.
2. No seizures, including auras, after surgery  
103 patients were not seizure free and another nine were lost from follow up. These patients were not included in the study.
3. Three or more seizures recorded during SEEG that were characterized by sustained (three seconds duration or longer) gamma activity at the onset

Based on this criterion we did not include:

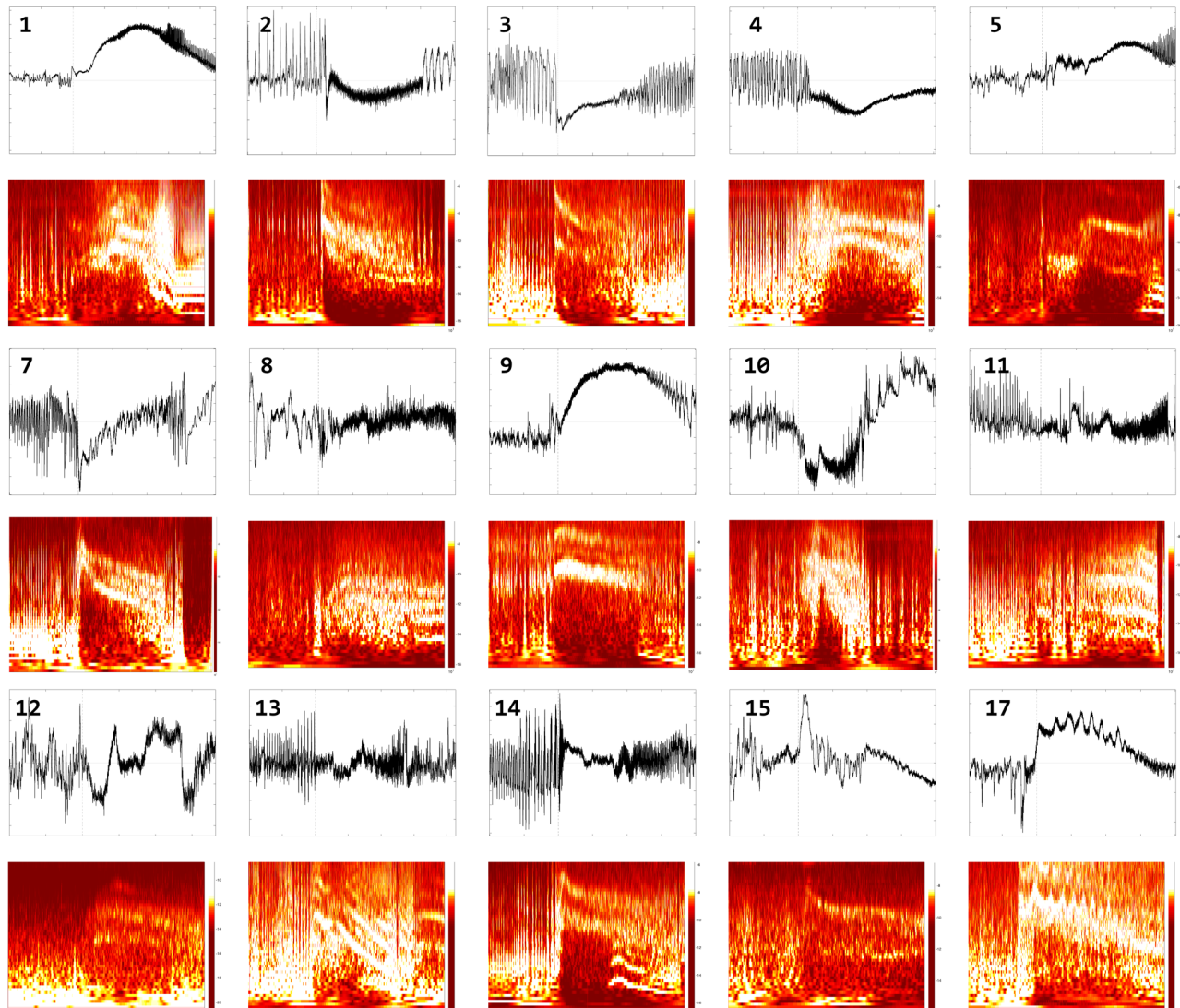
- a. 12 patients that had less than 3 seizures recorded or EEG data were not available for analysis;
- b. 4 patients that had resection that was not tailored and was not guided by SEEG due to inadequate sampling;
- c. 40 patients that had ictal onset patterns with characteristics different than sustained gamma activity: rhythmical spikes without clear transition to sustained FA ( $n = 2$ ); rhythmical oscillations in alpha/theta/delta range or only EEG flattening ( $n = 10$ ); FA less than 3 seconds duration ( $n = 9$ ); FA in alpha/beta/gamma range ( $n = 19$ ).

Finally, 16 patients met all inclusion criteria. We also included one patient (Subject 4) who underwent SEEG-guided laser surgery (“minimal resection”).



**Figure 1: Patient selection workflow**

## 2. Visual Characteristics of Epileptogenic Zone



**Figure 2: Contacts in the Epileptogenic Zone: one exemplar time series and its corresponding time frequency plot is shown for each patient. The number of each time series plot indicates the patient ID. Each plot shows 10 seconds prior to onset and 20 seconds after, and the frequencies are logarithmically spaced from 1 to 200 Hz. Note the characteristic association of pre-ictal sharp transient, bands of fast activity and suppression is present in each of the time frequency plot, despite its variations across subjects.**

### 3. Statistics of the Frequency of FA and Suppression

#### The Maximum and Minimum Frequency of FA

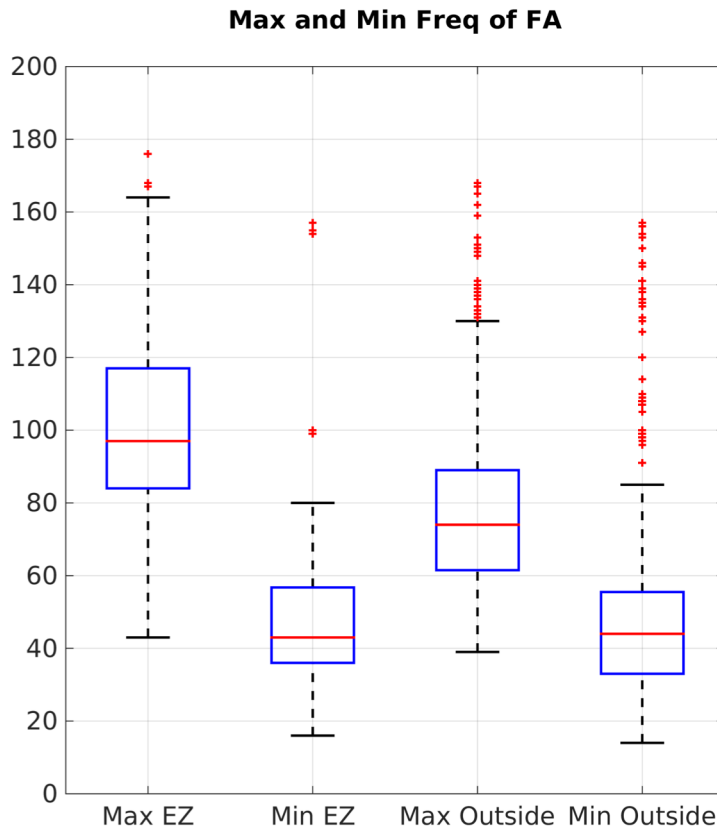
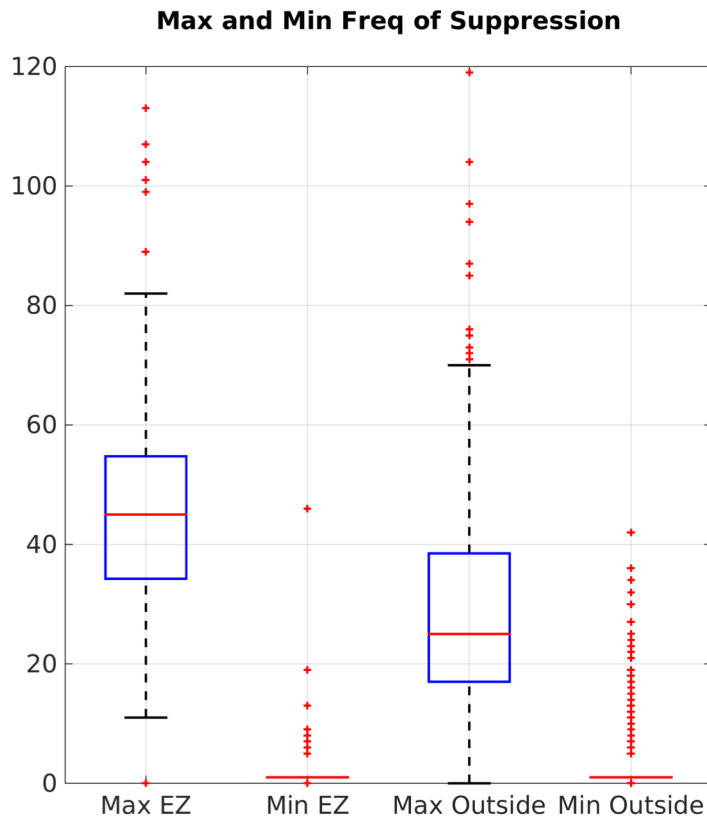


Figure 3: Column 1: The maximum frequency of FA in EZ contacts; Column 2: The minimum frequency of FA in EZ contacts; Column 3: The maximum frequency of FA in contacts outside resection region; Column 4: The minimum frequency of FA in contacts outside resection region.



# The Maximum and Minimum Frequency of Suppression

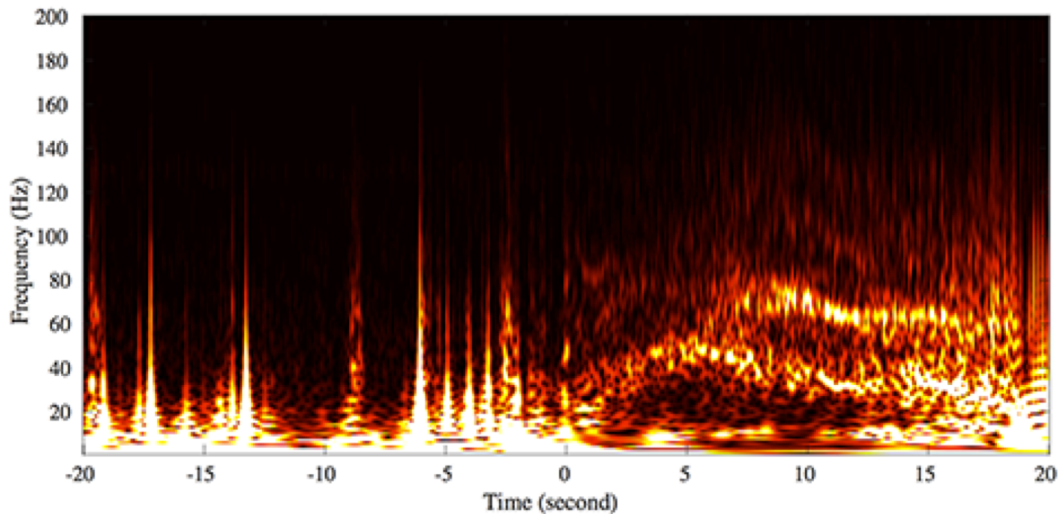


**Figure 4: Column 1: The maximum frequency of suppression in EZ contacts; Column 2: The minimum frequency of suppression in EZ contacts; Column 3: The maximum frequency of suppression in contacts outside resection region; Column 4: The minimum frequency of suppression in contacts outside resection region.**

## 4. Preprocessing

### Time-frequency Decomposition

The continuous Morlet wavelet transform with linear frequency scale from 1 to 200 Hz was applied to the raw SEEG data for each channel. The bipolar channel pair X2-X3 from Subject 1, Seizure 1P was chosen as an example for illustration. The full-width-half-maximum (FWHM) of the Gaussian kernel in the time domain for the mother wavelet was chosen to be 8 cycles as a trade-off between temporal and spectral resolution. The resulting time-frequency (TF) plot is shown in Fig. 5.

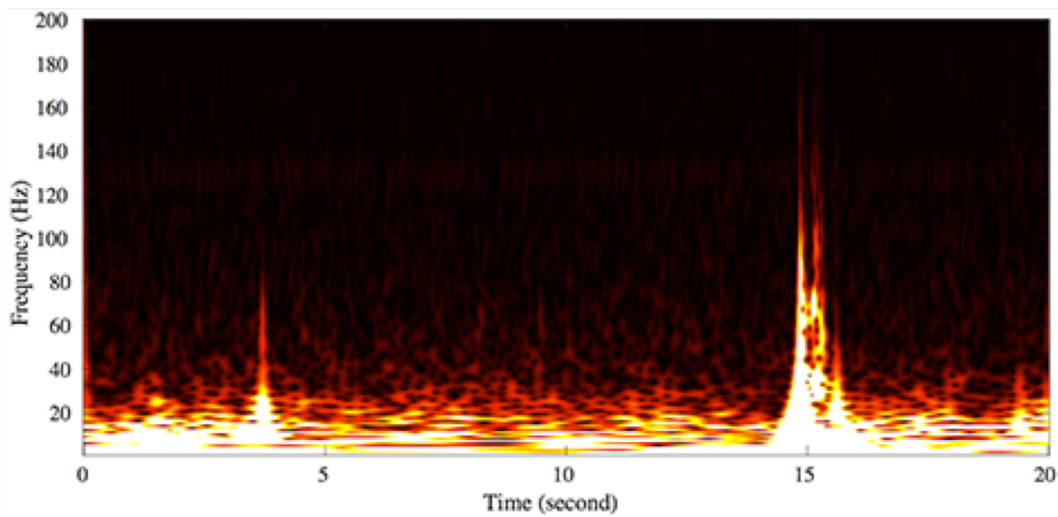


**Figure 5: Time-frequency plot of seizure onset for the selected channel X2-X3 from Subject 1, Seizure 1P.**

### Normalization

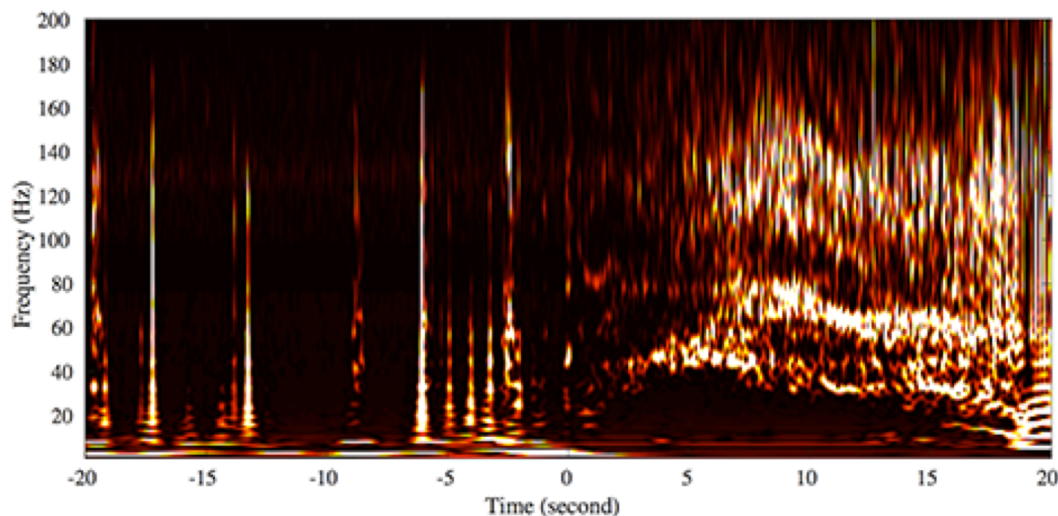
Two minutes before seizure onset, a baseline data set was also recorded for each seizure and transformed in the same way to a TF plot as shown in Fig. 6. The seizure data were then normalized with respect to the baseline segment for each frequency component. Let  $x(t, f)$  be the seizure onset TF data,  $y(t, f)$  the baseline TF data, and  $x_n(t, f)$  the normalized TF data. Then the normalization was performed as

$$x_n(t, f) = \frac{x(t, f) - \text{mean}_t\{y(t, f)\}}{\text{std}_t\{y(t, f)\}} \quad (1)$$



**Figure 6: Time-frequency plot of the baseline data for the selected channel X2-X3 from Subject 1, Seizure 1P.**

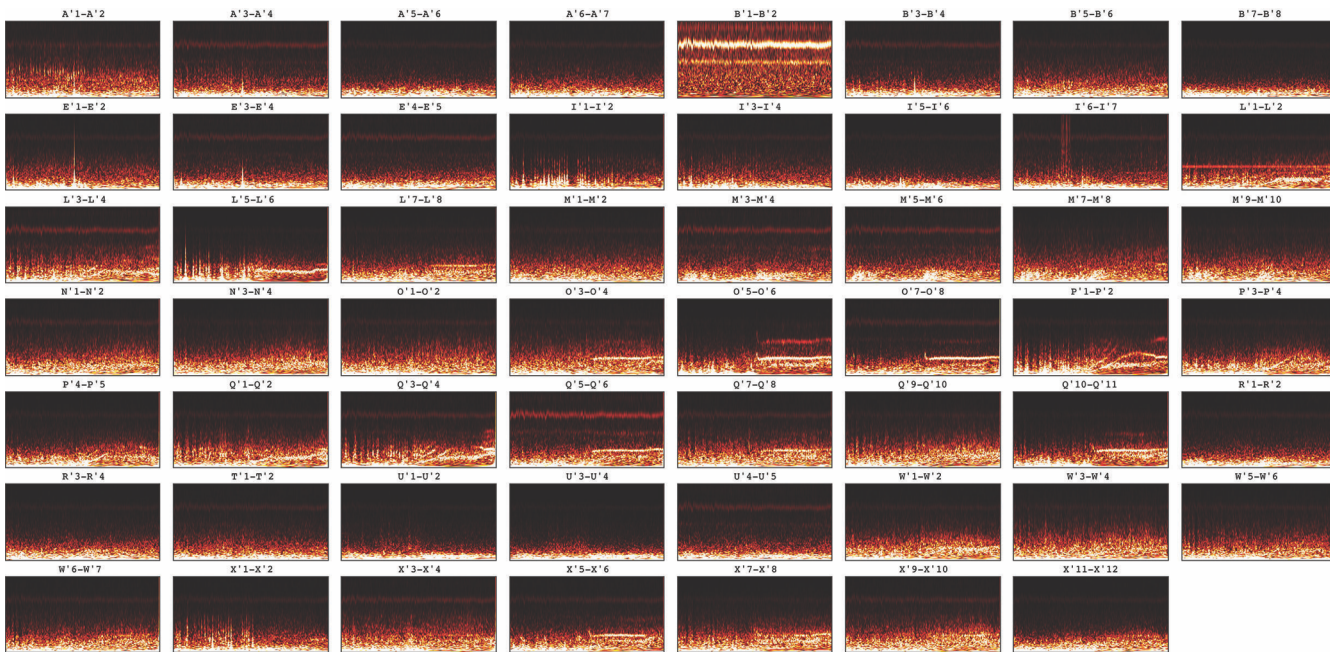
where  $\text{mean}_t\{\cdot\}$  and  $\text{std}_t\{\cdot\}$  denotes the mean and standard deviation with respect to time, respectively. Fig. 7 shows the result of normalization.



**Figure 7: Time-frequency plot of the normalized data for the selected channel X2-X3 from Subject 1, Seizure 1P.**

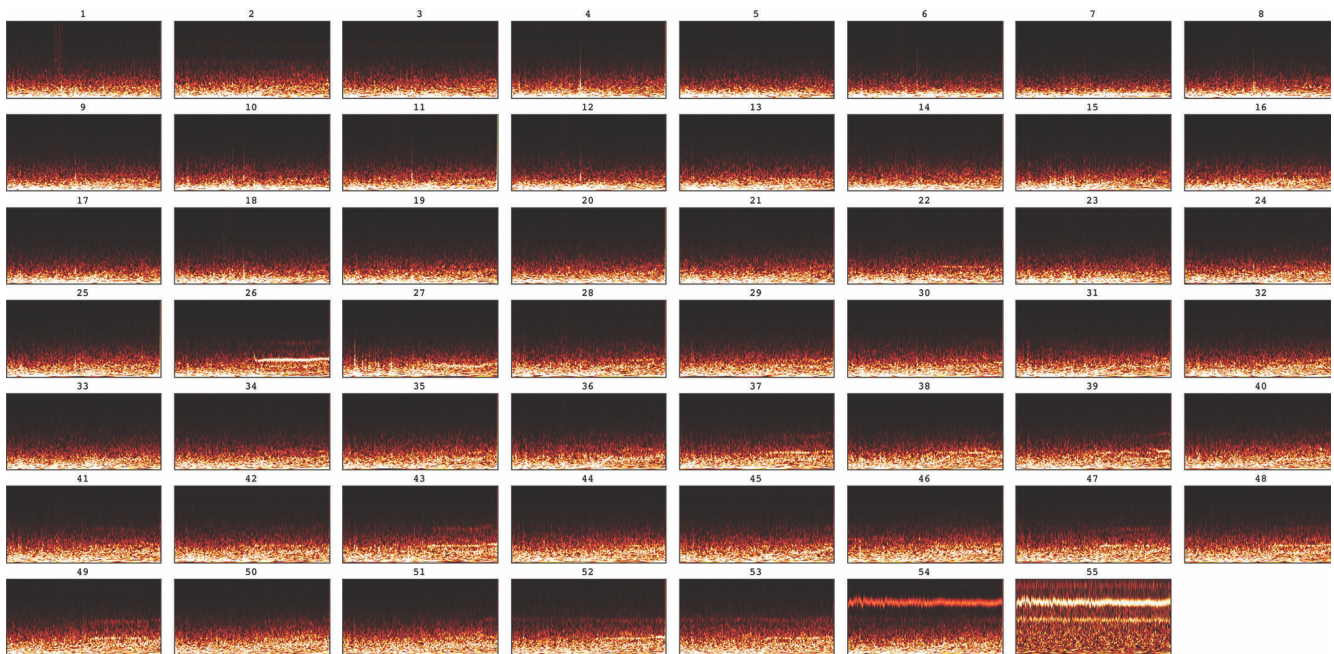
## Artifact Removal

Independent component analysis (ICA) has proven to be an effective algorithm in the analysis of EEG/MEG data (Makeig *et al.*, 1996; Vigário *et al.*, 2000), and is widely used to identify artifacts, such



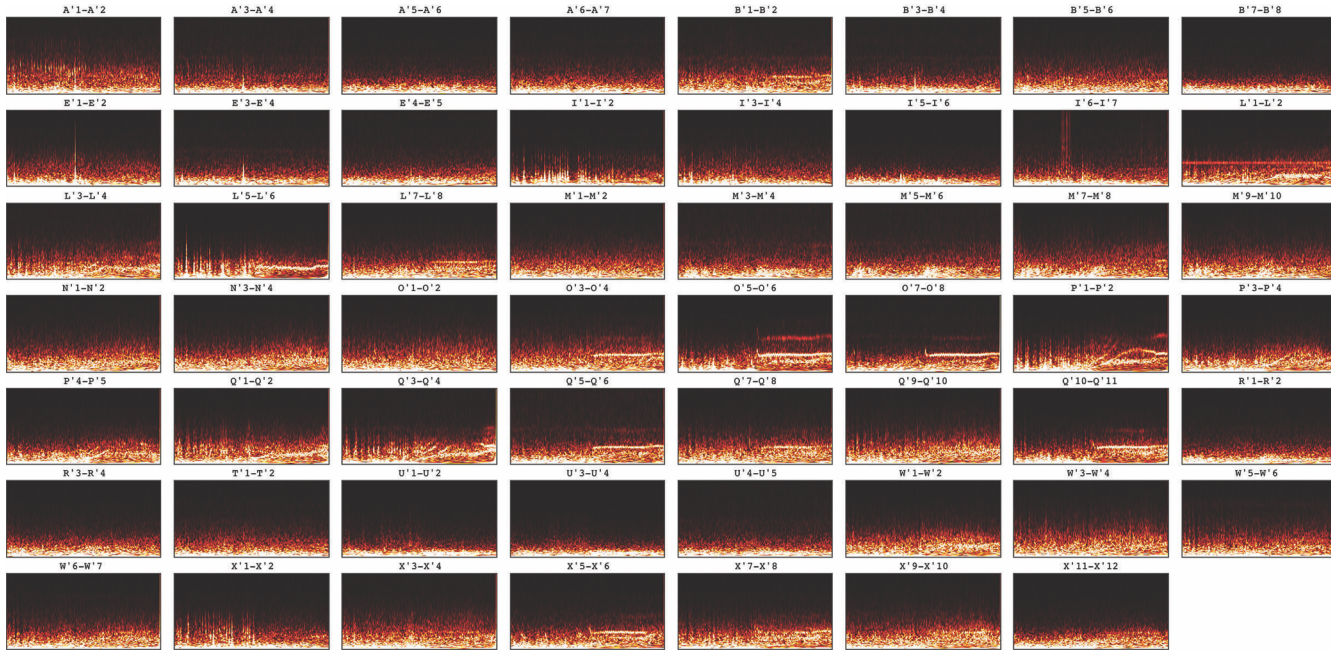
**Figure 8: Original TF plots of all channels of Subject 6, Seizure 2P with artifacts present**

as eye movement, eye blink and electrical artifact (Urrestarazu *et al.*, 2004; Delorme *et al.*, 2007). Here we applied complex ICA (cICA) (Bingham & Hyvärinen, 2000) to the TF data to identify and remove



**Figure 9: Plots of independent components obtained from cICA of all channels of Subject 6, Seizure 2P.**





**Figure 10: TF plots of all channels of Subject 6, Seizure 2P with artifact components removed.**

artifacts. This procedure was applied separately to the seizure and baseline data before normalization. Fig. 8 shows an example of the TF plots of all channels for Subject 6, Seizure 2P. We observed that two horizontally-structured artifacts are present across almost all channels. A three-dimensional tensor was formed from the TF data as channel  $\times$  time  $\times$  frequency by concatenating all channels along the first dimension, denoted as  $Y(c, t, f)$ , where  $c \in S_C = [1, 2, \dots, N]$ ,  $t \in S_T = [-20, \dots, 20]$ ,  $f \in S_F = [1, \dots, 200]$  and  $N$  is the total number of channels for that subject. Let  $X(c, t, f)$  be the source signals in the same tensor format and let  $M \in \mathbb{R}^{N \times N}$  be the un-mixing matrix. The cICA decomposition can be expressed as

$$X(c, t, f) = M \underset{1}{\times} Y(c, t, f) \quad (1)$$

where  $\underset{i}{\times}$  is the standard matrix-tensor multiplication in mode  $i$  (see Kolda & Bader 2009). Fig. 9 shows the independent components matricized from  $X(c, t, f)$ . Let  $S_A$  be the set of indices of the identified artifact components and let  $s_k, k \in S_A$ , be the  $k^{\text{th}}$  artifact component. For example, in this case the last two components,  $s_{54}$  and  $s_{55}$  captured the artifacts we observed in Fig. 8. Then those artifact components can be removed by zeroing out the  $k^{\text{th}}$  column of  $M$  for  $k \in S_A$  and the "artifact-free" TF plots, denoted as  $Y'(c, t, f)$ , can be reconstructed as

$$Y'(s, t, f) = M'_{1 \times} X(s, t, f) \quad (2)$$

where  $M'(i, j) = M^{-1}(i, j)$  for all  $i$  and  $j \notin S_A$  and  $M'(i, j) = 0$  for all  $i$  and  $j \in S_A$ . Fig. 10 shows the TF plots when two artifact components were removed from the original TF plots. It can be seen that the artifacts are suppressed for most channels.

## Reference

BINGHAM E, HYVÄRINEN A. a Fast Fixed-Point Algorithm for Independent Component Analysis of Complex Valued Signals [Internet]. *Int. J. Neural Syst.* 2000; 10: 1–8. Available from: <http://www.worldscientific.com/doi/abs/10.1142/S0129065700000028>

Delorme A, Sejnowski TJ, Makeig S. Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. *Neuroimaging* 2007; 34: 1443–1449.

Kolda TG, Bader BW. Tensor decompositions and applications. *SIAM Rev.* 2009; 51: 455–500.

Makeig S, J. Bell. A, Jung T-P, Sejnowski TJ. Independent Component Analysis of Electroencephalographic Data. *Adv. Neural Inf. Process. Syst.* 1996; 8: 145–151.

Urrestarazu E, Iriarte J, Alegre M, Valencia M, Viteri C, Artieda J. Independent component analysis removing artifacts in ictal recordings. *Epilepsia* 2004; 45: 1071–1078.

Vigário R, Särelä J, Jousmäki V, Hämäläinen M, Oja E. Independent component approach to the analysis of EEG and MEG recordings. *IEEE Trans. Biomed. Eng.* 2000; 47: 589–593.

## 5. Feature Extraction

Identification of each of the three features that characterize the EZ fingerprint is described in the following subsections with the following two steps: (i) candidate extraction, which checks for the existence of the feature and finds its location, and (ii) descriptor extraction, which generates a numerical description of the candidates. We begin with the TF plots after artifact removal, as described in Supplemental Material IV. We use channel X2-X3 from Subject 1, Seizure 1P, as shown in Fig. 7, for illustrative purposes.

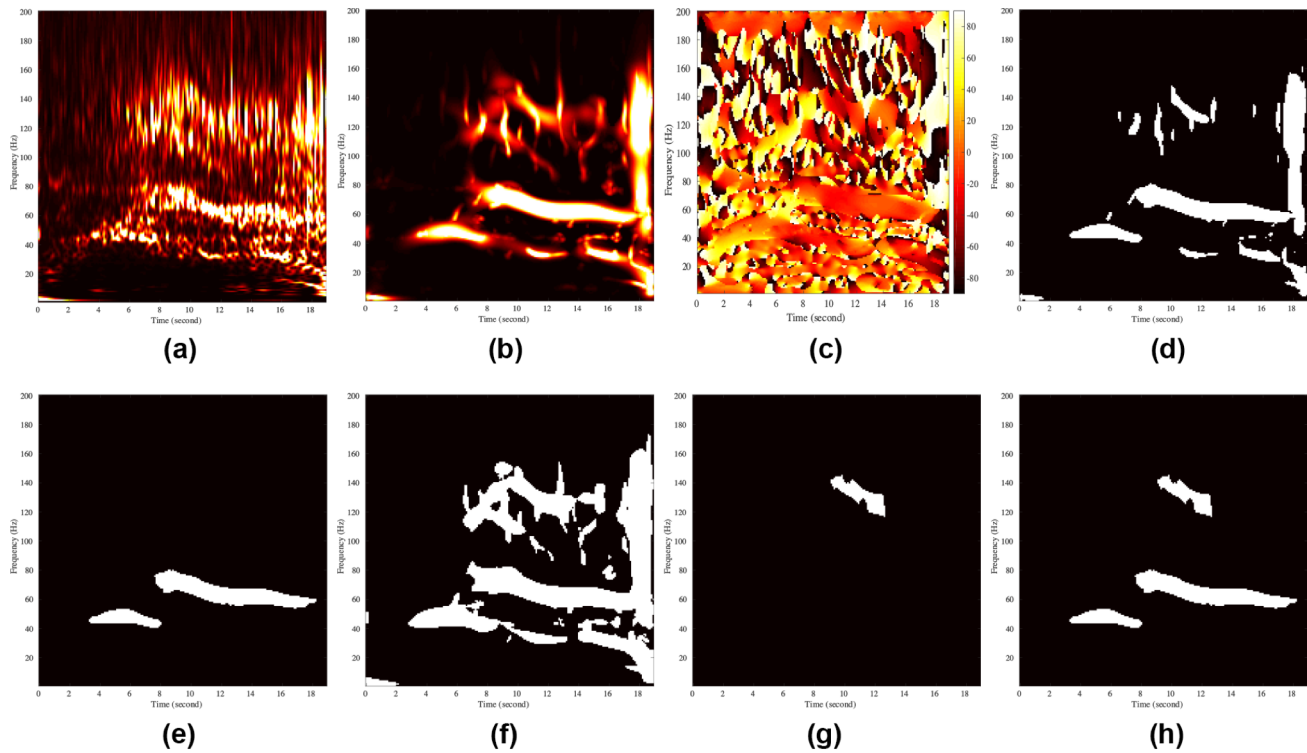
### Fast Activity

Fast Activity (FA), also known as banding, is defined as one or more high frequency oscillations (HFOs) after seizure onset. Fig. 11 (a) shows the TF plot from Fig. 7 from seizure onset. FA is highly variable across subjects and channels. In some instances, only one band may be present. In others, multiple bands can occur simultaneously as in this example. Moreover, the FA often exhibits “down-chirping” in which the frequency bands shift lower as a function of time.

FA detection uses Algorithm 1, for which we include pseudocode below. In order to capture the ridge-shaped structures, a novel multiscale Frangi filter algorithm was developed. Frangi *et al.* (1998) described a multiscale Hessian-based filter mainly for blood vessel enhancement. We use the Matlab implementation by Kroon (2010). Frangi’s method uses different scales of Gaussian smoothing to allow detection of vessels of different sizes. Our modification of this approach applies different scales of thresholding to identify FAs of differing strengths. At each scale, false detections are eliminated using information about the orientation, area, position and eccentricity of the FA feature. The final FA mask is generated by combining all candidates from different scales of detection. Finally, numerical descriptors, as a representation of FA, are extracted from the mask.

### FA Candidate Extraction

The procedure `FASTACTIVITYDETECTION` in Algorithm 1 takes a normalized artifact-cleaned post-onset TF plot, denoted as  $TF_{\text{Right}}$ , as input (Fig. 11 (a)) and the Frangi filter is applied. This yields a filter output, denoted by  $TF_{\text{F}}$ , and a direction map, denoted by  $D$ . Each pixel in  $TF_{\text{F}}$  reflects the degree of “banding” and each pixel in  $D$  represents the direction or orientation of these bands. Fig. 11 (b) shows an example of  $TF_{\text{F}}$  and Fig. 11 (c) shows an example of  $D$  where the angle ranges from -90 degree (negative y-axis)



**Figure 11: Fast Activity (FA) extraction example. (a) TF plot of Subject 1 Seizure 1P after onset. (b) Frangi filtering output. (c) The direction map. (d) First level initial mask. (e) The refined first level mask. (f) Second level initial mask. (g) The refined second level mask. (h) The final combined mask.**

to 90 degree (positive y-axis), rotating anti-clockwise as indicated by the color bar. For each level of thresholding, the following operations are performed:

1. Binarize  $TF_F$  to generate the initial mask, denoted as  $M_{new}$  by thresholding using the current  $level_k$ .
2. For any pixel in  $M_{new}$ , if the corresponding pixel in  $D$  exceeds some limit on the maximum angle, it is flagged to be false in  $M_{new}$ .
3. Find all the connected components (CCs) in  $M_{new}$ . For each CC, if any of the following criteria is satisfied, then that CC is removed from  $M_{new}$ .
  - a. If the area of the CC is smaller than some threshold  $areaTh$
  - b. If the center of the CC is outside the region restricted by some threshold  $bdrTh$
  - c. If the eccentricity of the CC is smaller than some threshold  $eccenTh$
4. Remove any banding area that has already been detected from the previous iteration and only keep the newly detected band(s).
5. Merge the mask into the mask from the previous iteration



**Algorithm 1 FA Detection****Procedure FASTACTIVITYDETECTION ( $TF_{Right}$ )**

```
Initialize  $M_{old}$ 
Initialize levels
 $N \leftarrow$  number of levels
Initialize angleTh, areaTh, bdrTh, eccenTh
 $(TF_F, D) \leftarrow$  FrangiFilter( $TF_{Right}$ )
for  $k = 1, 2, \dots, N$  do
     $M_{new} \leftarrow$  threshold( $TF_F$ , level $_k$ )
    for each pixel with index  $(i, j)$  in  $M_{new}$  do
        if  $D(i, j) >$  angleTh then
             $M_{new}(i, j) \leftarrow 0$ 
        end if
    end for
    for each CC with index  $i$  in  $M_{new}$  do
        if the area of  $CC_i <$  areaTh then
            remove  $CC_i$  from  $M_{new}$ 
        end if
        if the center of  $CC_i$  outside bdrTh then
            remove  $CC_i$  from  $M_{new}$ 
        end if
        if the eccentricity of  $CC_i <$  eccenTh then
            remove  $CC_i$  from  $M_{new}$ 
        end if
    end for
    for each CC with index  $i$  in  $M_{new}$  do
        if  $CC_i \cap M_{old} \neq \emptyset$  then
            remove  $CC_i$  from  $M_{new}$ 
        end if
    end for
     $M_{old} \leftarrow M_{new} \cup M_{old}$ 
end for
return  $M_{old}$ 
end procedure
```

The initial mask from the first level thresholding is shown in Fig. 11 (d), with refinement based on step 2 and 3 resulting in Fig. 11 (e). Step 4 and 5 are not involved in this first iteration. During the second iteration, the initial mask obtained from the second level thresholding is shown in Fig. 11 (f). After step 3, 4, an additional weaker banding region is found successfully as shown in Fig. 11 (g). Finally, the two masks are merged as shown in Fig. 11 (h).

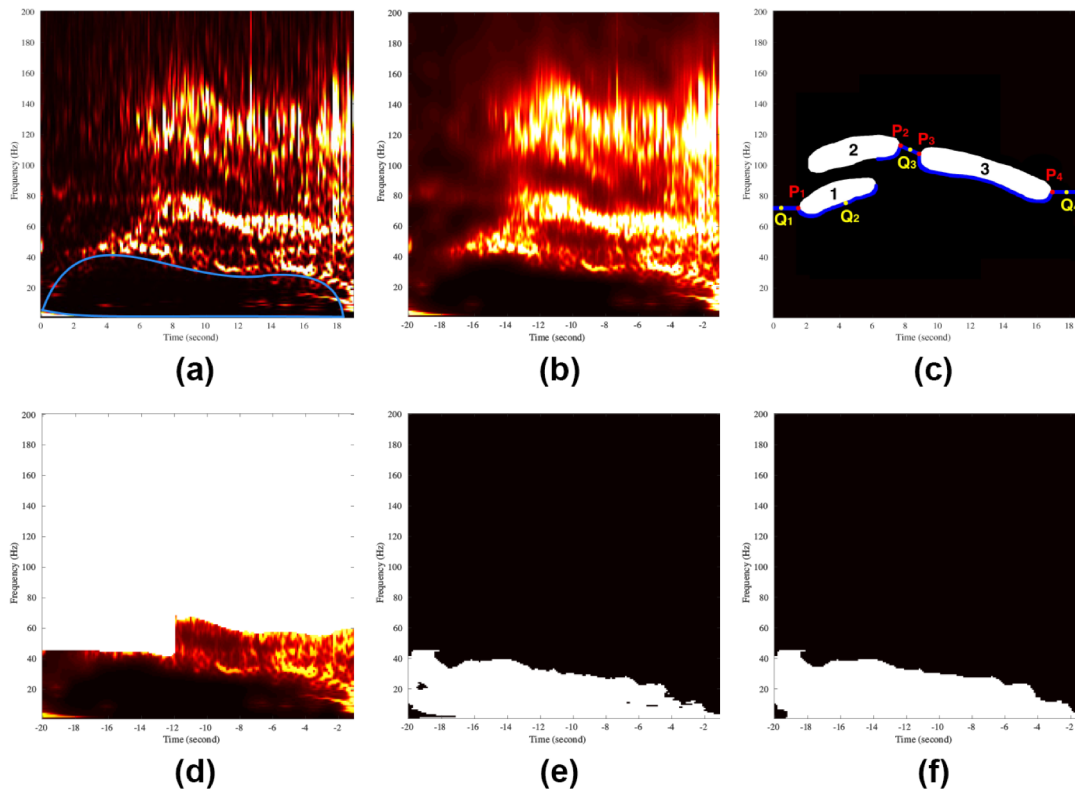
## FA Descriptor Extraction

There are 17 numerical descriptors extracted from the mask. The first descriptor is simply the total number of “banding” features found in the candidate extraction step above, denoted as  $N_B$ . In our chosen example, this number is 3. Among all detected bandings, only the two largest are retained and sorted based in order of ascending center frequency. For each banding feature the following 8 descriptors are extracted:

- $A_B$  – the area of the banding
- $O_B$  – the orientation of the banding
- $T_{BS}$  – the starting time of the banding with respect to onset time
- $T_{BE}$  – the ending time of the banding with respect to onset time
- $F_{BS}$  – the corresponding average frequency at time  $T_{BS}$
- $F_{BE}$  – the corresponding average frequency at time  $T_{BE}$
- $F_{BMax}$  – the maximum frequency that the banding reaches
- $F_{BMin}$  – the minimum frequency that the banding reaches

## Suppression

Low frequency suppression is apparent immediately after seizure onset, shown as a dark region in the lower frequencies of the TF plot in 7. Although the level of suppression plays an important role in characterizing the suppression region, determining the existence of suppression and describing its appearance numerically remains challenging. Fig. 12 (a) shows the ideal suppression region for this example, where the boundary is shown as the blue curve. A simple thresholding will not work well in most cases for the following reasons. First, for those channels where FA exists, it is not uncommon that the FA only partially spans the suppression region. In these cases, thresholding will incorporate some non-suppression dark regions, where the frequency is higher than the frequency of FA bands, into the suppression region. This phenomenon could occur, for example, above the suppression region in the left side of Fig. 12 (a). Second, for those channels where FA does not exist, there is no upper boundary to define the edge of the suppression region. Third, the intensity inside the suppression region varies considerably. As a result, the thresholded suppression region would be broken into several sub-regions. This phenomenon could occur, for example, in the lower right corner of the suppression region in Fig. 12 (a). Hence, we proposed a suppression detection pipeline to remedy the difficulties discussed above as shown in Algorithm 2. The TF plot was smoothed by a guided filter (He *et al.*, 2010) and a boundary was



**Figure 12: Suppression extraction example. (a) TF plot of Subject 1, Seizure 1P after onset with ideal suppression region. (b) Guided filtered TF plot. (c) A phantom of FA mask. (d) Guided filtered TF plot with upper bound set. (e) Initial mask from thresholding. (f) The final refined mask.**

set based on the banding information to restrict the detection region. Then a thresholding was performed followed by morphological operations in order to retain the topological integrity of the detected mask. Finally, similar to the banding detection procedure, numerical descriptors were extracted from the mask.

### Suppression Candidate Extraction

First, a guided filter (He *et al.*, 2010) is applied to the original TF plot for the sake of edge-preserving smoothing. The resulting TF plot, denoted as  $TF_{GF}$  in the algorithm, can be seen in Fig. 12 (b), where the region inside the suppression is well smoothed but the boundary between the suppression region and the FA is kept clear.

Second, we define an upper bound for the suppression due to the partial coverage of the banding discussed above. This is shown in the sub-procedure `BOUNDARYDETECTION` in Algorithm 2 and illustrated in Fig. 12 (c). We number each three bandings (1, 2 and 3 respectively in Fig. 12 (c) in black). Then we define four sentinel points: P1 is the left most point of banding 1. P2 is the right most point of banding 2. P3 is

## Algorithm 2 Suppression Detection

### Procedure SUPPRESSIONDETECTION ( $\mathbf{TF}_{\text{Right}}$ , $\mathbf{Mask}_B$ )

```
Initialize Th
 $\mathbf{TF}_{\text{GF}} \leftarrow \text{GuidedFilter}(\mathbf{TF}_{\text{Right}})$ 
 $\text{maxVal} \leftarrow \max(\mathbf{TF}_{\text{GF}})$ 
 $\mathbf{Bdr} \leftarrow \text{BOUNDARYDETECTION}(\mathbf{Mask}_B)$ 
 $T \leftarrow$  number of time samples along x-axis
 $N \leftarrow$  maximum frequency along y-axis
for  $k = 1, 2, \dots, T$  do
     $\mathbf{TF}_{\text{GF}}(k, [\mathbf{Bdr}(k), \mathbf{Bdr}(k) + 1, \dots, N]) \leftarrow \text{maxVal}$ 
end for
 $\mathbf{TF}_{\text{GFT}} \leftarrow \text{threshold}(\mathbf{TF}_{\text{GF}}, \text{Th})$ 
 $\mathbf{TF}_{\text{GFTM}} \leftarrow \text{M}_{\text{erode}}(\text{M}_{\text{fill\_holes}}(\text{M}_{\text{dilate}}(\mathbf{TF}_{\text{GFT}})))$ 
end procedure
```

### Procedure BOUNDARYDETECTION( $\mathbf{Mask}_B$ )

```
if  $\mathbf{Mask}_B \neq \emptyset$  then
    for  $k = 1, 2, \dots, T$  do
        if  $k < x_{P1}$  then
             $\mathbf{Bdr}(k) \leftarrow y_{P1}$ 
        else if  $x_{P1} \leq k \leq x_{P2}$  or  $x_{P3} \leq k < x_{P4}$  then
             $\mathbf{Bdr}(k) \leftarrow \min(\mathbf{Mask}_B(k, :))$ 
        else if  $x_{P2} \leq k \leq x_{P3}$  then
             $\mathbf{Bdr}(k) \leftarrow$  linear interpolation of  $y_{P2}$  and  $y_{P3}$ 
        else if  $k > x_{P4}$  then
             $\mathbf{Bdr}(k) \leftarrow y_{P4}$ 
        end if
    end for
    return  $\mathbf{Bdr}$ 
else
    return default boundary
end if
end procedure
```

the left most point of banding 3 and  $P4$  is the right most point of banding 3. Finally, we also define four exemplar points  $Q1$  to  $Q4$  (yellow in Fig. 12 (c)). The coordinates of point  $P_i$  will be denoted as  $x_{P_i}$  and  $y_{P_i}$ . If there is at least one banding found in the FA detection step, there are four different situations to consider when determining the boundary for each time point  $k$ . They will be discussed as follows:

- Case 1: When there is no banding at time  $k$  and all the bandings are on the right side of time  $k$ , i.e. when  $k < x_{P1}$ , e.g. point  $Q1$ . Then the boundary  $\mathbf{Bdr}(k)$  is defined as  $y_{P1}$ .

- Case 2: When there is a banding or multiple bandings at time  $k$ , i.e. when  $x_{P1} < k < x_{P2}$  or  $x_{P3} < k < x_{P4}$ , e.g. point Q2. Then the boundary  $Bdr(k)$  is defined as the minimum frequency that the bandings reach at time  $k$ .
- Case 3: When there is no banding at time  $k$  and time  $k$  is in a gap between bandings, i.e. when  $x_{P2} < k < x_{P3}$ , e.g. point Q3. Then the boundary  $Bdr(k)$  is defined as the linear interpolation of  $y_{P2}$  and  $y_{P3}$ .
- Case 4: When there is no banding at time  $k$  and all the bandings are on the left side of time  $k$ , i.e. when  $k > x_{P4}$ , e.g. point Q4. Then the boundary  $Bdr(k)$  is defined as  $y_{P4}$ .

If there is no banding detected at all, a default boundary is used. The default boundary is defined as the minimum of all boundaries across those channels where at least one banding is found and hence the boundary can be well defined.

Third, we mask off the region outside the boundary by setting its intensity to the maximum intensity of the  $TF_{GF}$  as shown in Fig. 12 (d) to ensure it is excluded from the suppression region in the subsequent thresholding.

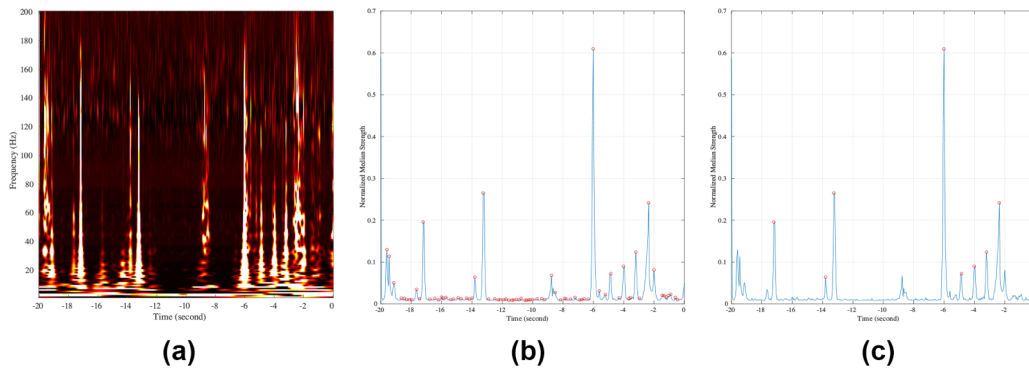
Fourth, a thresholding is applied to obtain a binary initial mask for the suppression region, as shown in Fig. 12 (e)

Next, some morphological operations (dilation, hole-filling and erosion, denoted  $M_{dilate}$ ,  $M_{fill\_holes}$  and  $M_{erode}$  respectively) are performed sequentially to ensure the topological integrity of the detected region. Finally, only the largest connected component is returned as the detected suppression mask as shown in Fig. 12 (f)

### Suppression Descriptor Extraction

There are 9 numerical descriptors extracted from the mask:

- $A_S$  – the area of the suppression area
- $V_S$  – the variance of the intensity of the suppression area
- $IMed_S$  – the median intensity of the suppression area
- $IMax_S$  – the maximum intensity of the suppression area
- $T_{SS}$  – the starting time of the suppression with respect to onset time
- $T_{SE}$  – the ending time of the suppression with respect to onset time
- $F_{SMax}$  – the maximum frequency that the suppression reaches



**Figure 13: Pre-ictal spiking extraction example. (a) TF plot of Subject 1, Seizure 1P before onset. (b) Plot of the median of upper 25% quantile of the intensity distribution in (a) for each time point with initial local maxima. (c) Same plot as (b) with detected spike candidates.**

- $F_{SMin}$  – the minimum frequency that the suppression reaches
- $RtP_S$  – the ratio between  $IMed_S$  and the median intensity of the pre-ictal region. The pre-ictal region is a rectangle area where the -20 s and 0 s define the boundary along the temporal axis and the  $F_{SMin}$  and  $F_{SMax}$  define the boundary along the spectral axis.

## Pre-ictal Spikes

The third feature is spiking during the pre-ictal period, from -20 s to onset. The TF plot for the pre-ictal period for our exemplar channel, i.e. the left side of 7, and denoted  $TF_{Left}$ , is re-drawn here in Fig. 13 (a). We define a spike as activity of very short duration that spans almost all frequencies. This can be well characterized by two properties: the strength of the spike and the sharpness of the spike.

## Pre-ictal Spiking Candidate Extraction

First, for each time point, the upper 25% quantile of the intensity distribution of  $TF_{Left}$  is found and its median determined. This results in a one-dimensional signal, denoted as  $P$ , as shown in Fig. 13 (b). Second, the locations of the local maxima, denoted  $maxPos$ , and their values  $maxVal$ , are found (shown as red circles in Fig. 13 (b)). Third, we compute the ratio of peaks to their corresponding neighbors as a measure of the sharpness of the spikes as follows:

- For each peak  $i$  at  $maxPos(i)$ , the  $k$  neighbors of that peak are found, denoted as  $nbPos$ .
- The median of the response  $P$  at  $nbPos$  is computed
- The ratio between the  $maxPos(i)$  and the median value we obtained from step two is calculated

### Algorithm 3 Preictal Spike Detection

#### Procedure PREICTALSPIKEDETECTION ( $TF_{Left}$ )

```
Initialize number of neighbors  $k$ 
Initialize ratio threshold  $Th$ 
 $T \leftarrow$  number of time samples along x-axis
for  $i = 1, 2, \dots, T$  do
     $P(i) \leftarrow$  median of upper 25% quantile of  $TF_{Left}(i, :)$ 
end for
 $maxVal \leftarrow$  local maximum of  $P$ 
 $maxPos \leftarrow$  time points where  $maxVal$  found
 $N \leftarrow$  number of  $maxVal$  found
for  $i = 1, 2, \dots, N$  do
     $nbPos \leftarrow$  neighbors of  $maxPos(i)$ 
     $ratio \leftarrow maxVal(i) / median(P(nbPos))$ 
    if  $ratio < Th$  then
        remove  $maxPos(i)$  and  $maxVal(i)$ 
    end if
end for
return  $maxVal$  and  $maxPos$ 
end procedure
```

- If the ratio is less than some desired threshold, the candidacy of the peak  $i$  is removed

The final spike candidates are shown as red circles in Fig. 13 (c). It can be seen that only those large and sharp spikes are left after the extraction procedure.

### Pre-ictal Spiking Descriptor Extraction

There are total 3 numerical descriptors extracted from the spike candidates.

- $N_S$  – the total number of spike candidates found
- $M_S$  – the mean of the ratios we obtained from the second stage above
- $T_{PE}$  – the time the last spike occurs with respect to the onset.

### Reference

Chapman BE, Parker DL. 3D multi-scale vessel enhancement filtering based on curvature measurements: Application to time-of-flight MRA. In: Medical Image Analysis. Springer; 2005. p. 191–208.

He K, Sun J, Tang X. Guided Image Filtering (ECCV). Eur. Conference Comput. Vis. 2010; 35: 1–14.

Kroon D-J. Hessian based Frangi Vesselness filter [Internet]. 2010 Available from: <http://www.mathworks.com/matlabcentral/fileexchange/24409-hessian-based-frangi-vesselness-filter>



## 6. Classification

### Features and Classifier

In contrast to the usage of "feature" in Supplemental Material V, here we use "feature", as is standard in machine learning, to refer to the (concatenated) numerical descriptors of FA, suppression and pre-ictal spikes. Let  $N_F$  be the total number of features available for each sample. Here,  $N_F = 17 + 9 + 3 = 29$ . Let  $N_S$  be the total number of samples we have for all channels and all 3 seizures from each of 17 subjects. Let  $C_i$  be the number of channels for subject  $i$ . Therefore,  $N_S = (\sum_i C_i) \times 3 \times 17$ . Then the feature matrix, denoted as  $F \in \mathbb{R}^{N_S \times N_F}$ , along with the labels described in the following section, was used as the input to a support vector machine (SVM). An RBF (Gaussian) kernel with empirical kernel scale 7.5 was applied and the regularization parameter was chosen to be 3.5 for penalizing those samples that violate the margin. Therefore, a soft-margin kernel SVM was formed, computed using standard convex quadratic programming.

### Subject-based cross validation

In order to prevent overfitting the data and to test accuracy, we use cross-validation. Two major cross-validation methods are commonly used: random split  $k$ -fold cross-validation and leave-one-sample-out cross-validation.  $K$ -fold cross-validation divides the dataset into  $k$  folds with equal size via random sampling without replacement. For each fold  $i$ , the other  $k - 1$  folds are combined together for training the classifier. Then the prediction for the  $i^{\text{th}}$  fold is performed based on the trained classifier. Leave-one-sample-out cross-validation can be treated as a special case of  $k$ -fold cross-validation where  $k$  equals the total number of samples  $N_S$ . Therefore,  $k$ -fold cross-validation is more computationally efficient than leave-one-sample-out for  $k < N_S$ .

Unfortunately, standard  $k$ -fold cross-validation is not applicable to our dataset because of the strong dependency across contact pairs and seizures for each subject. i.e. an assumption of independence between samples does not hold. Also, it has been shown that leave-one-sample-out cross-validation tends to introduce (optimistic) bias due to the insufficient testing data (Varoquaux *et al.*, 2017). Therefore, we use leave-one-subject-out cross-validation method: let  $M$  be the total number of subjects, where in our case  $M = 17$ . For each subject  $i$ , we first gather the data from all other  $M - 1$  subjects and train the SVM. We

use the resulting SVM to identify putative EZ was carried out for subject  $i$ . We then repeat for each of  $M = 17$  subjects.

## Labels

Ground truth labels are crucial for supervised classification problems. For the EZ identification purpose, the ideal binary class label should be positive for those contacts inside the EZ and negative outside the EZ. However, this ideal ground truth is not known precisely with our data. Rather we know which contacts lie within the resected zone and, since these patients were all seizure free, we can assume that these are a superset of the electrodes in the EZ. Therefore, for those contacts that were inside the resected region, the resection label was defined as positive and for those outside the resected region, the resection label was defined as negative.

## Classification with partially certain labels

There is an asymmetry in the labels we have: we know that outside the resection zone there are no EZ contacts, but within it there are a mix of EZ and non-EZ contacts. This appears to be a novel problem in classification from partially uncertain labelled data. We use a clustering-guided algorithm to reduce uncertainty in the partially certain labels. Let  $S$  be the entire data set. Before we train the SVM we first perform the following clustering procedure. This is repeated separately for each of the 17 cross-validation runs.

First, an unsupervised  $k$ -means clustering was performed on all samples from  $S$ , clustering into two sets,  $P_U$  and  $N_U$ . The subscript  $\cdot_U$  stands for "unsupervised". We know that if none of the desired features (FA, suppression and pre-ictal spike) are detected, then the concatenated descriptors formed is a zero vector. So let  $x \in S$  be the sample that is closest to a zero vector in its Euclidean norm. Thus,  $x$  is the most unlikely sample to lie in the EZ and therefore least likely to be misclassified in the  $k$ -means clustering. Then  $P_U$  is chosen as the set that does not contain  $x$  and  $N_U$  is the negative set containing  $x$ . Note that  $P_U$  and  $N_U$  are mutually exclusive, i.e.  $P_U \cap N_U = \emptyset$  and  $P_U \cup N_U = S$ . Also, the resection labels, denoted as  $L_R$ , naturally contained two groups as we discussed before. Let  $P_R$  and  $N_R$  be the positive and negative resection group respectively, where the subscript  $\cdot_R$  stands for "resection".  $P_R$  and  $N_R$  are also mutually exclusive. Then,  $P_S = P_R \cap P_U$  is defined as the EZ-positive set, and  $N_S = N_R$  as the EZ-negative set for the supervised training of the SVM. In short, this pre-labelling step can be viewed as a filtering procedure

because the set  $P_R \cap N_U$  is eliminated from  $S$  as most likely to have incorrect labels before the supervised SVM training.

## Voting

The prediction label, denoted as  $L_P$  was obtained from the classification. Although  $L_P$  contains the labels for all channels from all subjects, it still does not directly indicate the estimated location of the EZ. This is because for each subject, we have 3 seizures recorded and the prediction labels across 3 seizures for a particular channel of a subject may not be consistent. Therefore, we use a majority voting mechanism across the three seizures to draw the final conclusion.

### Algorithm 4 Classification of EZ

#### Procedure EZCLASSIFICATION( $F, L_R$ )

```

for  $i = 1, 2, \dots, M$  do
     $P_U, N_U \leftarrow \text{kmeans}(F_{-i,:})$ 
     $P_R, N_R \leftarrow L_{R_{-i,:}}$ 
     $P_S \leftarrow P_R \cap P_U$ 
     $N_S \leftarrow N_R$ 
    Model  $\leftarrow \text{SVMTraining}(F_{-i,:}, P_S, N_S)$ 
     $L_{P_{i,:}}, \text{Score}_{i,:} \leftarrow \text{Predict}(\text{Model}, F_{i,:})$ 
    for  $k = 1, 2, \dots, C_i$  do
        if  $\sum_j \text{Score}_{i,j,k} > 0$  or  $\sum_j L_{P_{i,j,k}} \geq 2$  then
             $L_{V_{i,k}} \leftarrow \text{True}$ 
        else
             $L_{V_{i,k}} \leftarrow \text{False}$ 
        end if
    end for
end for
return  $L_V$ 
end procedure

```

## Classification pipeline

In summary, the entire pipeline of the classification is shown in Algorithm 4. Note that the subscripts  $\cdot_{i,j,k}$  following variable  $F, L_R, L_P$  and Score represent the indices for subject  $i \in [1, 2, \dots, M]$ , seizure  $j \in [1, 2, 3]$  and channel  $k \in [1, 2, \dots, C_i]$ . Colon is used when all indices are queried. A minus sign before an index indicates all indices but that one. For example,  $F_{i,3,:}$  is the feature matrix for all channels in the 3<sup>rd</sup> seizure of the  $i^{\text{th}}$  subject.  $L_{R_{-j,:}}$  is the resection label for all channels and all seizures of all subjects

excluding the  $j^{\text{th}}$  one. One exception is that the subscript  $\cdot_{i,j}$  for variable  $L_V$  represents the indices for subject  $i \in [1, 2, \dots, M]$  and channel  $j \in [1, 2, \dots, C_i]$  without the indices for seizures as we sum over results across all 3 seizures. Moreover, when counting the number of positive prediction labels for a channel, the prediction label  $L_P$  is treated numerically as 1 when true and 0 when false. Finally, we summarize all variables and notations used in this Supplemental Material VI below for easy reference and understanding of Algorithm 4:

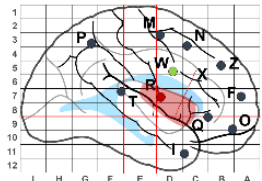
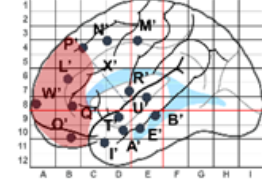
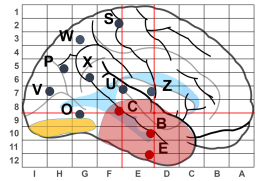
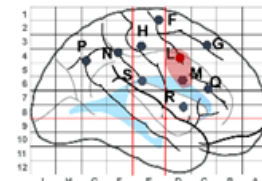
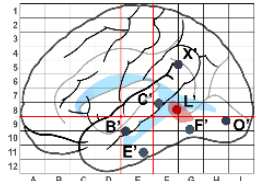
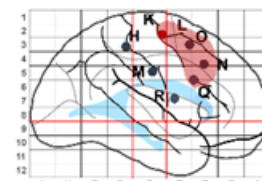
- $F \in \mathbb{R}^{N_S \times N_F}$  – The data matrix with  $N_S$  number of samples and  $N_F$  number of features
- $L_R$  – Resection labels
- $M$  – Total number of subjects
- $P_U, N_U$  – Positive and negative label set obtained from  $k$ -means unsupervised clustering
- $P_R, N_R$  – Positive and negative label set obtained directly from  $L_R$
- $P_S, N_S$  – Positive and negative label set used for SVM supervised training
- $L_P$  – Prediction labels
- Score – Scores indicating the distance from the decision boundary for each sample
- $L_V$  – Voting labels

## Reference

Varoquaux G, Raamana PR, Engemann DA, Hoyos-Idrobo A, Schwartz Y, Thirion B. Assessing and tuning brain decoders: Cross-validation, caveats, and guidelines [Internet]. *Neuroimage* 2017; 145: 166–179. Available from: <http://dx.doi.org/10.1016/j.neuroimage.2016.10.038>

# 7. Classification Result with Lesion Information

**Table 1: Implantation Maps with Schematic Representation of the Resection Margins (Shaded in Red) for Patients with Confirmed or Suspected Lesions. Bipolar SEEG Channels Inside Epileptogenic Lesion Identified (EZ) and Not Identified (Non-EZ) by the Algorithm as Epileptogenic are Named for Each Individual Patient\*.**

ID	Map**	EZ Contacts Inside Lesion	Non-EZ Contacts Inside Lesion	ID	Map**	EZ Contacts Inside Lesion	Non-EZ Contacts Inside Lesion
1		R1-R2, R2-R3, R3-R4, X1-X2, X2-X3, X3-X4, X5-X6		6			Possible MRI lesion P'1-P'2 (not confirmed by pathology)
3		B1-B2, B2-B3, B3-B4, C2-C3		11		L6-L7	
5		L'1-L'2	L'3-L'4	12		K4-K5, K5-K6, K7-K8	K9-K10

\*Only patients who had lesion(s) are shown here.

\*\*Electrodes on the maps are marked: (1) as red, if they contain True Positive (TP) channels (potentially epileptogenic inside the resection); (2) as green if they contain False Positive (FP) channels (potentially epileptogenic outside the resection); (3) as black if they contain only True Negative (TN) channels (not potentially epileptogenic outside the resection); (4) as black in the red-shaded area if they contain False Negative (FN) channels (not potentially epileptogenic inside the resection). Boundaries of prior resection are schematically shaded in yellow (only patient 3 and 9 had a previous resection).