Chapter 4

Segmentation and Clustering of MEPs

This chapter describes a series of steps for processing the EMG signal in order to extract useful information. The EMG signal was recorded from patients with spinal cord injury during their rehabilitation described in [24]. For a complete description on the physiology and characteristics of the said EMG signal, please refer to Section 2.1 and 2.2.

During the rehabilitation training in [24], electrical stimulation is applied to the spinal cord. The stimulation parameters (such as amplitude, frequency, and electrode configurations) are varied manually by the researchers in order to study the different effects of the electrical simulation on the spinal neurons, and to find the optimal set of parameters. The stimulation with one fixed set of stimulation parameters is called one "event". Every event lasts for a couple of seconds. As a result, the processing of the EMG signal is carried out on every event individually. There are several benefits in doing so. Firstly, by performing EMG processing on a short interval, it is safe to assume the statistics of the noise are stationary. Secondly, the EMG signal from different events is likely to show different statistics, and the performance is good when the signal in regard has a stationary statistics. Last but not least, it's desired to compare the characteristics of the EMG signal between different events.

The muscle responses in the EMG signal are called Motor Evoked Potentials (MEPs). MEPs can be classified into two subgroups: monosynaptic MEPs and polysynaptic MEPs. They are also sometimes referred to as the early responses and late responses, respectively, due to the differences of their arrival times after the electrical stimulus. After the peaks of MEPs are detected using the peak detection methods described in Chapter 3, a series of interesting tasks can be further carried out in order to expose finer structures of the MEPs. This information can assist the neurophysiologists in assessing the underlying neural activities of the spinal cord.

The fundamental two tasks are to segment and cluster the MEPs. In this thesis, the goal

of segmentation is to identify MEP segments from the EMG signal sequence. More specifically, the boundaries of an MEP need to be obtained. Suppose an EMG sequence of N samples, $\mathbf{X} = \{x_1, x_2, \cdots, x_N\}$, contains one MEP. The goal of segmentation is to obtain index a and b ($b > a$) , such that segment, $\mathbf{w} = \{x_a, x_{a+1}, \cdots, x_b\}$ contains all the samples from the MEP, but no other samples. Clustering, in general, is the task of partitioning a set of objects into different groups, so that two objects within one group are more similar than two objects from two different groups. In the task of MEP clustering, the goal is to group similar MEPs together. This is motivated by the fact that similar MEPs are likely to come from the same neural physiological process. By identifying different clusters of MEPs, different neural processes can be inferred by the physiologists. In particular, monosynaptic MEPs are significantly different from polysynaptic MEPs because of the distinct neural pathways. As a result, the clustering is carried out in a sequential manner: firstly monosynaptic MEPs are separated from polysynaptic MEPs via an initial, crude clustering; then clustering is performed within monosynaptic MEPs and polysynaptic MEPs so as to discover finer structures. The overall procedure is first stated here. Detailed explanations on these step are described in the following sections.

1. Find segments of MEPs from the detected peaks in the EMG signal.
2. Find majority of the monosynaptic MEPs via hierarchical clustering on all segments of MEPs.
3. Further cluster monosynaptic MEPs into subgroups based on clustering with Gaussian mixtures.
4. Decompose segments that contain both polysynaptic MEPs and monosynaptic MEPs
5. Clustering on polysynaptic MEPs with Gaussian mixtures

To demonstrate the proposed method, the same event of the EMG signal is used throughout this chapter. The EMG signal was recorded from left medial gastrocnemius (L MG) of a patient with clinically motor complete spinal cord injuries while lying in the supine position with electrical stimulation applied to the spinal cord [24]. The stimulation amplitude is 7.2V and frequency is 10Hz. This event lasts 30 seconds with 300 simulation intervals (A simulation interval is the time period between two stimuli).

4.1 Segmentation of MEP Waveforms

After the peaks of the MEPs are detected from the EMG signal, MEP waveforms are segmented based on peak information. As observed from the EMG signal, MEPs are composed of consecutive transient peaks. As a result, if two peaks are "next to" each other, then they are likely to be peaks from the same MEP, or vice versa. It's not a good idea to choose a hard threshold because the

distance between two peaks varies for different shapes of MEPs, and is not a known priori. A better way is to incorporate the shape information from the wavelet transform of the peaks.

In the proposed wavelet-based peak detection algorithm in Chapter 3, every detected peak corresponds to one ridge in the wavelet space (the time-scale space). The wavelet used is the Mexican hat wavelet, which naturally resembles a peak. The peak location is estimated from the translation of the maximum wavelet coefficient along the ridge, as the magnitude of the wavelet coefficient measures the resemblance between the signal and the wavelet function (See Eq. (3.3.6.2) and Eq. (3.3.6.3)). For convenience, the equations are copied here.

$$(r\hat{x}_i, r\hat{y}_i) = \arg\max_{(x,y)\in\mathcal{R}_i} X(x,y)$$

$$\hat{p}_i = r\hat{y}_i, \qquad \text{for} \quad i = 1, 2, \cdots, N_R$$

where $X(i,j) : (i,j) \mapsto X_{i,j}$, which is the wavelet coefficient matrix defined by:

$$X_{j,k} \stackrel{\text{def}}{=} (\mathbf{X})_{j,k} = \sum_{n=-\infty}^{+\infty} x[n]\psi_{s_j,k}[n] \qquad j = 0, 1, \ldots, J, \quad k = 0, 1, \ldots, N;$$

$\{\mathcal{R}_i, \quad i = 1, 2, 3, \cdots, N_R\}$ is the set of ridges detected, where N_R is the total number of ridges detected, and $\mathcal{R}_i$ is a set of pair of indices: $\mathcal{R}_i = \{(rx_{i,j}, ry_{i,j}), \quad j = 1, 2, \cdots, L_i\}$, where $rx_{i,j}$ and $ry_{i,j}$ give the row and column index (as in $\mathbf{X}$) of the j-th peak maxima along ridge $\mathcal{R}_i$, respectively, and $L_i = |\mathcal{R}_i|$ is the number of peak maxima along the ridge.

The scale of the maximum coefficient along the ridge $\mathcal{R}_i$ is therefore:

$$\hat{rs}_i = s_{r\hat{x}_i}$$

since $r\hat{x}_i$ gives the index in the scale set $\mathcal{S} \stackrel{\text{def}}{=} \{s_0, s_1, \ldots, s_j, \ldots, s_J\}$.

Now the effective support of the peak at $\hat{p}_i$ is defined with respect to the wavelet function at scale $\hat{rs}_i$. Although the theoretical mother wavelet function is not time-limited, its most energy is confined only within a small region around the origin. Define the effective support of the Mexican Hat mother wavelet in time domain as w (as shown in Figure 4.1-1), then the Mexican Hat wavelet at any scale s has effective support of $w \cdot s$. There are multiple choices of the values of w. In this thesis, w is chosen to be $[-4, 4]$, and the signal energy within the support occupies almost 100% of the total energy.

The effective support of a detected peak at $\hat{p}_i$ is estimated to be the same as the wavelet at scale $\hat{rs}_i$:

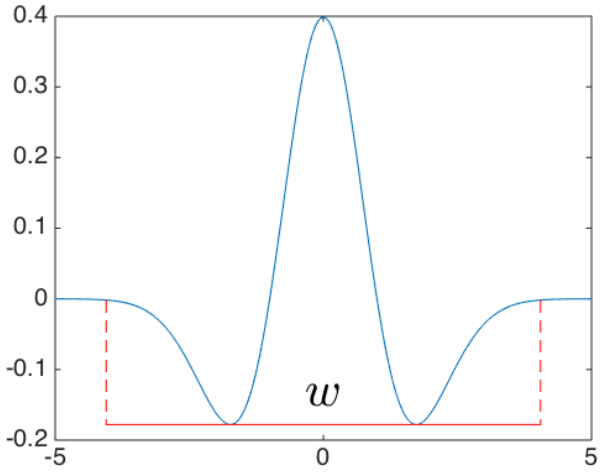


Figure 4.1-1: Mexican Hat mother wavelet $\psi(t)$ (given by Eq. 3.3.1.1) and its effective support (denoted by red line and w)

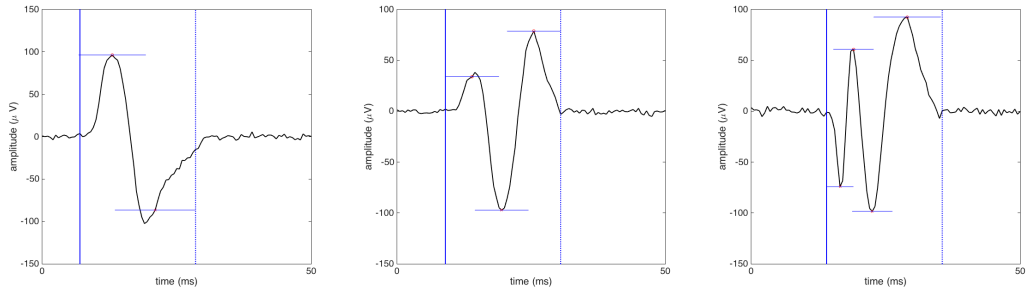
$$w(p_i) = w \cdot \hat{rs}_i \tag{4.1.0.1}$$

If the supports of two consecutive peaks overlap, then these two peaks are claimed to be from the same MEP. More formally, $\hat{p}_i$ and $\hat{p}_{i+1}$ are judged to be from the same MEP, if and only if:

$$\hat{p}_i + w(p_i)/2 > \hat{p}_{i+1} - w(p_{i+1})/2 \tag{4.1.0.2}$$

Example results are shown in Figure 4.1-2. The figure contains three EMG slices containing different MEPs to demonstrate the effectiveness of this method on different shapes of MEPs. For every peak estimated, an estimated effective support is calculated using Eq. (4.1.0.1). After this is done for all the peaks, peaks are grouped into MEPs based on the criterion in Eq. (4.1.0.2). The onset of an MEP is estimated as the beginning of the effective support of its first peak, while the end of an MEP is estimated as the end of the effective support of its last peak, as shown in Figure 4.1-2.

Given the EMG signal of N samples: $\mathbf{X} = \{x_1, x_2, \cdots, x_N\}$. The result of this segmentation procedure is a set of MEPs: $\mathcal{W}_{MEP} = \{\mathbf{w}_i\}_{i=1}^{N_M}$, where $\mathbf{w}_i$ is the waveform i-th MEP: $\mathbf{w}_i = w_{i,1}, w_{i,2}, \cdots, w_{i,D_i}$, and N_M denotes the total number of MEPs. Extra information about the MEPs are also obtained to assist further analysis. The onset locations of MEPs are denoted as $\mathbf{b}_{MEP} = \{b_1, b_2, \cdots, b_{N_M}\}$ where $b_i \in \{1, 2, \cdots, N\}$ is the index of $w_{i,1}$ as in $\mathbf{X}$. If one MEP is caused by the electrical stimulus, then this stimulus is said to be "associated" with this MEP. This stimulus associated with MEP $\mathbf{w}_i$ is simply obtained by finding the closest stimulus that comes before it. Given an array of the locations of N_s stimuli: $\mathbf{T}_{stim} = \{t_1, t_2, \cdots, t_{N_s}\}$ where $t_i \in \{1, 2, \cdots, N\}$



(a) R MH (b) L MG (c) L VL

Figure 4.1-2: Examples of segmenting MEPs from effective supports of peaks: Each subplot shows one piece of the EMG signal simulated from one of the MEP waveforms in Figure 3.4-2. The black line is the simulated EMG signal containing one MEP. The red dots/crosses are estimated positive/negative peak locations from the proposed wavelet method. The effective support of every peak is plotted as a short blue horizontal line centering at the peak. The blue vertical line denotes the onset of an MEP, while the dotted vertical line denotes the end. R MH = right medial hamstrings. L MG = left medial gastrocnemius. L VL = left vastus lateralis.

gives the index of i-th stimulus in $\mathbf{X}$, then the stimulus associated with $\mathbf{w}_i$, a_i follows:

$$a_i = \underset{t \in \mathbf{T}_{stim} \cap t < b_i}{\arg\max} \quad t - b_i \tag{4.1.0.3}$$

The result is a set of indices $\mathbf{a} = \{a_i\}_{i=1}^{N_M}$.

Running the peak detection algorithm proposed in this thesis followed by the above segmentation method on the example event of the EMG signal yields 469 MEPs. The EMG signal is recorded from left medial gastrocnemius (L MG) of a patient with clinically motor complete spinal cord injuries while lying in the supine position with electrical stimulation applied to the spinal cord. The stimulation amplitude is 7.2V and frequency is 10Hz. This event lasts 30 seconds with 300 simulation intervals.

4.2 Identify Cluster of Monosynaptic MEPs

As discussed in Section 2.1.2.2, MEPs can be classified as monosynaptic MEPs and polysynaptic MEPs. Monosynaptic MEPs are usually the direct responses of muscles after electrical stimulation, while polysynaptic MEPs are typically indirect responses. It takes different times for the effect of the electrical stimulation to show up in the muscle cells. The *latency* of an MEP is defined as the time between the electrical stimulation and the onset of the MEP. An example of the latencies of two MEPs is shown in Figure 4.2-1.

The latency of a monosynaptic MEP is normally larger than that of a polysynaptic one. Hence,

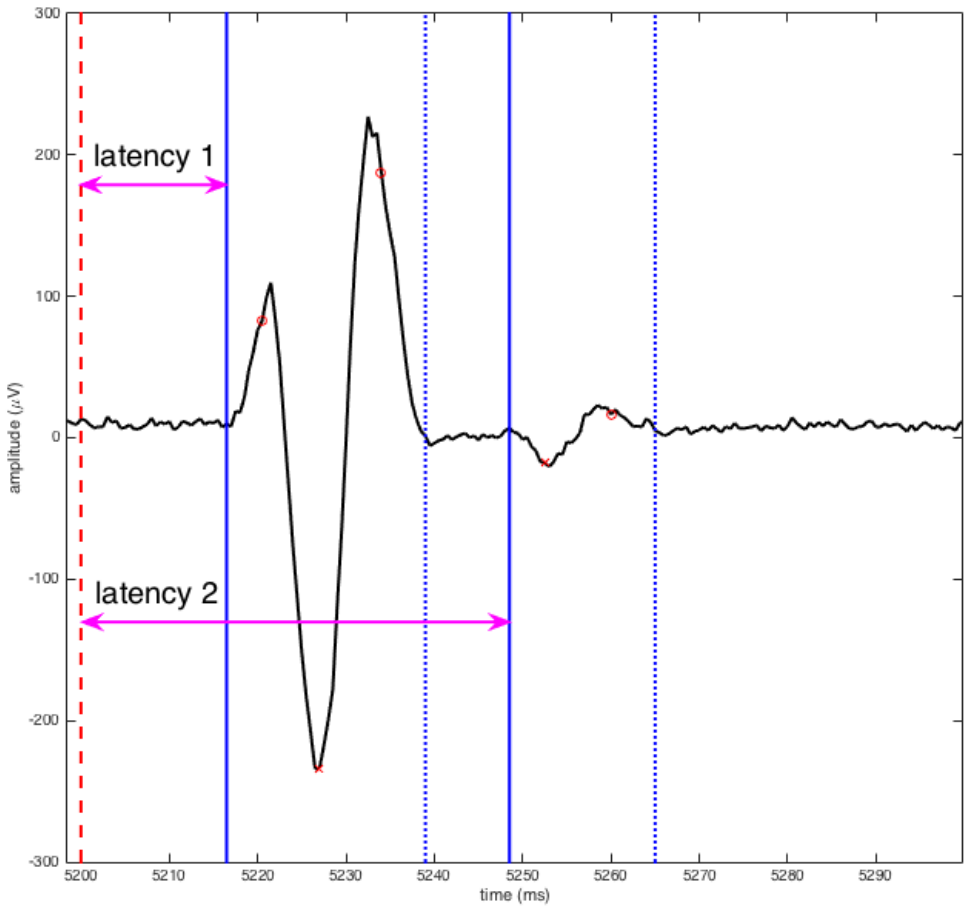


Figure 4.2-1: Example of latency of an MEP when occurrence of MEP is defined as its onset: The vertical red dotted line on the far left indicates the occurrence of one stimulus. *latency 1* is the latency of the monosynaptic MEP, while *latency 2* is the latency of the polysynaptic MEP.

sometimes monosynaptic MEPs are also called early responses while polysynaptic MEPs are called late responses. Monosynaptic MEPs typically have similar latencies among them, while monosynaptic MEPs don't. In addition, monosynaptic MEPs are normally much stronger than polysynaptic MEPs. For a complete explanation of the underlying physiology, please refer to Section 2.1.2.2.

The ultimate goal is to divide MEPs into monosynaptic ones and polysynaptic ones, and perform further analysis (such as clustering) within each group. To achieve this goal, the cluster that contains the majority of the monosynaptic MEPs is identified in this section, because monosynaptic MEPs have a relatively predictable structure, and hence easier to identify. Then the remaining MEPs are processed to extract remaining monosynaptic MEPs, and absorb them to the cluster of monosynaptic MEPs. The final remaining MEPs are grouped into the cluster of polysynaptic MEPs.

4.2.1 Feature Extraction

Based on these prior knowledge on the key differences between polysynaptic MEPs and monosynaptic MEPs, the following features are extracted for every MEP:

Strength: the strength of i-th MEP $\mathbf{w}_i$ is defined as its energy:

$$strength(i) = \|\mathbf{w}_i\|^2 = \sum_{j=1}^{D_i} w_{i,j}^2 \tag{4.2.1.1}$$

Latency: the latency of i-th MEP[1].

$$latency(i) = b_i - a_i \tag{4.2.1.2}$$

Duration: the length of i-th MEP: $duration(i) = D_i$

As a result, each MEP $\mathbf{w}_i$ is associated with a vector v_i in its feature space, comprised of its strength, latency and duration: $v_i = [strength(i) \quad latency(i) \quad duration(i)]$. The set of all features are $\mathbf{v} = \{v_i\}_{i=1}^{N_M}$. Denote the k-th features of all MEPs as $\mathbf{v^k} = \{v_{i,k}\}_{i=1}^{N_M}$.

The features need to be scaled to have a common value range so that all features have equal impact on the computation of the distance. Two scaling schemes are employed.

[1]The neurophysiologists define latency to be $b_i - a_i$, as shown in Figure 4.2-1. Alternatively, the weighted mean (defined in Eq. (4.3.1.8) as τ_i, and used as the alignment mark) is less sensitive to the noise and the non-stationarity of MEPs than the onset of the MEP, b_i. As a result, weighted mean is instead used to calculate the latency feature in the hierarchical clustering.

Interval scaling: the interval-scaled feature of $\mathbf{v^k}$ is:

$$\hat{v_{i,k}} = \frac{v_{i,k} - \min_i v_{i,k}}{\max_i v_{i,k} - \min_i v_{i,k}} \tag{4.2.1.3}$$

Ratio scaling: the ratio scaling is performed simply by doing a log transform followed by an interval scaling:

$$\hat{v_{i,k}} = \frac{\log(v_{i,k}) - \min_i \log(v_{i,k})}{\max_i \log(v_{i,k}) - \min_i \log(v_{i,k})} \tag{4.2.1.4}$$

Both scaling are used for numeric features. The interval scaling is the most common one, and ratio scaling is used when the feature takes exponential values. In the proposed method, the latency and duration features are interval-scaled. The strength feature is ratio-scaled so as to minimize the distance between monosynaptic MEPs as the energy of an MEP takes value in a wide range.

Different weights can be put on different features so as to put more emphasis on certain features. Suppose the weights vector is $\mathbf{c} = [c_1\ c_2\ c_3]^T$, then the k-th feature is replaced by $\mathbf{v^k}$ multiplied by c_k. In the proposed method, $c_1 = 1, c_2 = 1, c_3 = 2$ to highlight the difference of duration between monosynaptic MEPs and polysynaptic MEPs.

4.2.2 Hierarchical Clustering

The objective of the clustering is to partition MEPs into two groups: the monosynaptic MEPs and the polysynaptic MEPs. There are a number of clustering techniques employed for the EMG signal, such as hierarchical clustering [31, 42, 15, 8], k-means clustering [52]. k-means clustering is simple and efficient, but it is sensitive to outliers and initial seeds, and it's not suitable for discovering clusters that are not hyper-ellipsoids (or hyper-spheres). Since the detection and segmentation are not perfect, there will be outliers. In addition, as one can see when the features are plotted, the clusters of monosynaptic MEPs are not hyper-spheres at all. Because of the weaknesses of k-means clustering, hierarchical clustering is chosen for this task. Although hierarchical clustering is less efficient, the number of MEPs detected from one event is not big (typically $100 - 1000$), so efficiency is not a concern in this application.

Hierarchical clustering involves following three steps:

1. compute the distance between all pairs of objects, and stores the result in a distance matrix.
2. group the objects into a binary, hierarchical cluster tree, called *dendrogram*. Initially every object forms a cluster. Then the two closest pair of clusters gets merged into one new cluster. Repeat the linking until only one cluster remains, resulting a binary tree with root being the final cluster with all objects and leaves being every individual objects.

3. determine where to cut the hierarchical tree into clusters.

Each step can be implemented in various ways. In the proposed method, the following options are chosen.

- the distance measure between two objects is chosen to be the L_1 distance. A L_1 distance between two vectors $\mathbf{p}$ and $\mathbf{q}$ is defined as:

$$d_1(\mathbf{p}, \mathbf{q}) = \|p - q\|_1 = \sum_{i=1}^{n} |p_i - q_i| \tag{4.2.2.1}$$

  This distance is chosen so as to minimize the distance within monosynaptic MEPs while maximizing the distance between monosynaptic MEPs and polysynaptic MEPs.

- In the linking stage, the distance between two clusters is defined as the shortest distance between any two objects in the two clusters, one from each cluster.

- When cutting the hierarchical tree, the criterion is chosen to be the distance between the clusters. If the distance between two clusters are above the threshold, then they are two separate clusters in the final result. The distance is chosen to be 0.1.

The time complexity for the proposed clustering is $O(N_M^2)$ for N_M MEPs.

From the physiology of MEPs, monosynaptic MEPs tend to have similar latencies and durations, and much larger energy than polysynaptic ones. The cutoff threshold is intentionally chosen to be a small value to ensure that there is at least one cluster full of pure monosynaptic MEPs. The potential drawback is that other clusters may contain monosynaptic ones, too. This problem will be solved later. The cut of the dendrogram results in many clusters. One of them contains only the monosynaptic MEPs. The next task is to identify the largest cluster that contains only the monosynaptic MEPs. That cluster of monosynaptic MEPs are plotted with their associated stimuli aligned in Figure 4.2-2. The method of finding this cluster is described as follows.

The idea is to assign a score to every cluster that measures the effective standard deviation of the latencies of monosynaptic MEPs within the said cluster. Then the cluster with the lowest effective standard deviation is chosen to be the cluster of monosynaptic MEPs. Suppose there are N_C clusters $\mathcal{C} = \{\mathbf{C}_i\}_{i=1}^{N_C}$, where $\mathbf{C}_i = \{c_{i,1}, c_{i,2}, \cdots, c_{i,n_i}\}$ gives the set of indices of MEPs in the i-th cluster ($c_{i,1} \in \{1, 2, \cdots, N_M\}$ are indices in the set $\mathcal{W}_{MEP}$). The latency feature is $\mathbf{v^2} = \{v_{i,2}\}_{i=1}^{N_M}$, the second features of all MEPs in the feature space. Denote the total number of events as N_E. Then the effective standard deviation of i-th cluster, $\sigma(i)$, is calculated following:

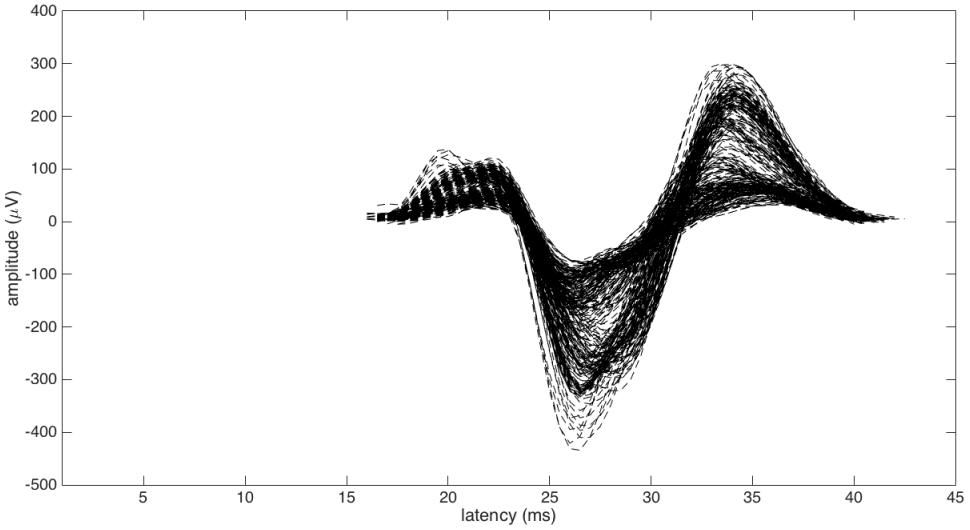


Figure 4.2-2: Example of a cluster with 285 monosynaptic MEPs from the initial hierarchical clustering of 469 MEPs detected and segmented from 30 seconds of the EMG signal. The MEP waveforms are aligned to their associated electrical stimuli at the origin, so the x axis is the latencies of the MEPs. The EMG signal was recorded from left medial gastrocnemius (L MG) of a patient with clinically motor complete spinal cord injuries while lying in the supine position with EES. The stimulation amplitude is 7.2V and frequency is 10Hz.

$$\sigma(i) = std(\{v_{c_{i,j},2}\}_{j=1}^{n_i}) \cdot \frac{|N_E - n_i|}{N_E} \tag{4.2.2.2}$$

where $std(\mathbf{S})$ gives the standard deviation of the set $\mathbf{S}$. The above formula is formulated with following rationale. The monosynaptic MEPs have very consistent latencies, so the standard deviation of the latencies of the cluster of monosynaptic MEPs should be very small (close to 0). However, the standard deviation alone does not always work because small clusters could have smaller standard deviations than the large cluster of pure monosynaptic MEPs. Based on physiology, there is approximately 1 monosynaptic MEP in one event, so theoretically there should be N_E monosynaptic MEPs in the cluster. As a result, the closer that n_i is to N_E, the higher confidence that $\mathcal{C}_i$ contains the monosynaptic MEPs. As a result, a factor $\frac{|N_E - n_i|}{N_E}$ is multiplied to the standard deviation to get rid of the small clusters. The index of the identified cluster of monosynaptic MEPs is:

$$\hat{I}_M = \underset{i \in \{1,2,\cdots,N_C\}}{\arg\min} \sigma(i) \tag{4.2.2.3}$$

In the example shown in Figure 4.2-2, the identified cluster of monosynaptic MEPs contain 285 MEPs and have a effective standard deviation of 0.000371. MEPs of the second largest cluster from hierarchical clustering is also plotted in Figure 4.2-3 to give readers some insight on the hierarchical clustering result. It contains 46 polysynaptic MEPs.

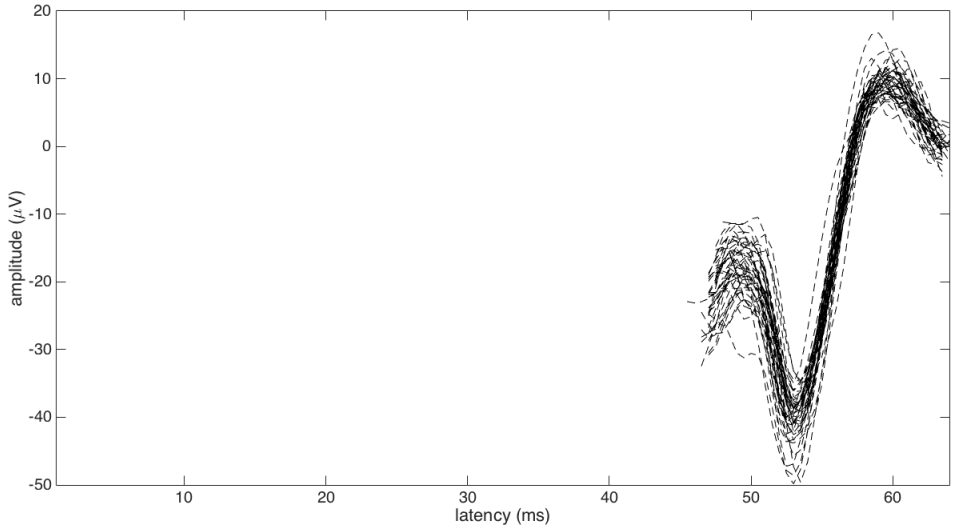


Figure 4.2-3: Example of the second largest cluster with 46 MEPs from the initial hierarchical clustering of 469 MEPs detected and segmented from 30 seconds of the EMG signal. The MEP waveforms are aligned to their associated electrical stimuli at the origin, so the x axis is the latencies of the MEPs. The EMG signal was recorded from left medial gastrocnemius (L MG) of a patient with clinically motor complete spinal cord injuries while lying in the supine position with EES. The stimulation amplitude is 7.2V and frequency is 10Hz.

4.3 Clustering of Monosynaptic MEPs

In the hierarchical clustering, the features are chosen so as to maximize the difference between monosynaptic MEPs and polysynaptic MEPs, while at the same time minimizing the difference within monosynaptic MEPs. Because both monosynaptic and polysynaptic MEPs can have various waveforms and the exact shapes are not a known priori, shape information is not used in the hierarchical clustering. After the majority of the monosynaptic MEPs are extracted during the hierarchical clustering, a further clustering within them is performed to find intrinsic structures of the monosynaptic MEPs.

As discussed in the physiology of MEPs in Section 2.1.2.2, MEPs are the activities of the muscle cells in response to the electrical stimulation provided by the implanted electrodes on the spinal cord. The neurons in the spinal cord are excited and their activities are modulated by the electric field. Different neural pathways would result in different MEPs shown in the EMG signal of the lower-body muscles. The goal of the clustering within monosynaptic MEPs is to infer the different neural pathways or activities from the different structures or shapes of the MEPs.

4.3.1 Feature Extraction: Principal Component Analysis

In order to cluster with the shape information, the entire waveform of an MEP should be used. However, the MEP waveforms contain about $\sim 10 - 100$ samples, and many of them are highly redundant. The redundant features add unnecessary complexity and computation time to the clustering task, and would shadow the most differentiating features. As a result, the first step before performing any kinds of clustering is to reduce the number of features, and principal component analysis (PCA) is the most common way of doing this.

PCA is a technique that is widely used for applications such as dimensionality reduction, lossy data compression, feature extraction, and data visualization. PCA can be defined as the orthogonal projection of the data onto a lower dimensional linear space, known as the *principal subspace*, such that the variance of the projected data is maximized [4]. Suppose the data is $\{\mathbf{x}_n\}$, where $n = 1, 2, \cdots, N$, and $\mathbf{x}_n = [x_{n1}\ x_{n2}\ \cdots\ x_{nD}]^T$ is a (column) vector in the Euclidean space with dimensionality D. Assume the mean has already been subtracted from the data such that $\bar{\mathbf{x}} = 0$. The subtraction of the mean is a preprocessing of the data before performing PCA. With $\bar{\mathbf{x}} = 0$, the meaning of maximum variance in projected data is achieved.

The goal is to find a subspace with dimensionality $M < D$ such that the projected data contains most of the variance in the original data. Suppose the subspace is given by M orthonormal vectors $\{\mathbf{u}_i\}$ where $i = 1, 2, \cdots, M$. $\mathbf{u}_i = [u_{i1}\ u_{i2}\ \cdots\ u_{iD}]$ is a column vector in original space with following properties:

$$\mathbf{u}_i^T \mathbf{u}_j = \delta_{ij} \tag{4.3.1.1}$$

$$\|\mathbf{u}_i\| = 1 \tag{4.3.1.2}$$

where δ_{ij} is the Kronecker delta function.

The projection of a given data $\mathbf{x}_n$ onto the subspace $\{\mathbf{u}_i\}_{i=1}^M$ is therefore $\mathbf{z}_n = [z_{n1}\ z_{n2}\ \cdots\ z_{nM}]^T$ where $z_{ni} = \mathbf{x}_n^T \mathbf{u}_i$. Let $\mathbf{U} = [\mathbf{u}_1\ \mathbf{u}_2\ \cdots\ \mathbf{u}_M]$ be the matrix with columns being the orthonormal vectors $\mathbf{u}_i$. Then, the projection of $\mathbf{x}_n$ can compactly be expressed as:

$$\mathbf{z}_n = \mathbf{U}^T \mathbf{x}_n \tag{4.3.1.3}$$

Let matrix $\mathbf{X} = [\mathbf{x}_1\ \cdots\ \mathbf{x}_N]^T$ represent the N data with rows being the observations $\mathbf{x}_n$, and let matrix $\mathbf{Z} = [\mathbf{z}_1\ \cdots\ \mathbf{z}_N]^T$ represent the N projections with rows being the projections $\mathbf{z}_n$. Then:

$$\mathbf{Z} = \mathbf{X}\mathbf{U} \tag{4.3.1.4}$$

From the projection $\mathbf{Z}$, original data can be reconstructed with:

$$\tilde{\mathbf{X}} = \mathbf{Z}\mathbf{U}^T \tag{4.3.1.5}$$

$\tilde{\mathbf{X}}$ would be the same as $\mathbf{X}$ if $M = D$. For $M < D$, there is a distortion or error. The distortion measure is the squared distance between the original data and the reconstructed data.

$$J = \frac{1}{N}\sum_{n=1}^{N} \|\mathbf{x}_n - \tilde{\mathbf{x}}_n\| \tag{4.3.1.6}$$

It turns out PCA minimizes J for a fixed $M < D$ with respect to different choice of the M orthonormal vectors $\{\mathbf{u}_i\}_{i=1}^{M}$ [4].

Theoretically, PCA is achieved by evaluating the covariance matrix $\mathbf{S}$ of the data set $\mathbf{X}$ and then finding the M eigenvectors of $\mathbf{S}$ corresponding to the M largest eigenvalues. The covariance matrix is calculated by: $\mathbf{S} = 1/N \sum_{n=1}^{N} \mathbf{x}_n^T \mathbf{x}_n$.[2] In practice, many algorithms have been developed to compute the M eigenvectors efficiently.

The mathematical derivation can be found in [4]. Here is the summary of the conclusions. Suppose $\{\lambda_k\}$ and $\{\mathbf{u}_k\}$ are the eigenvalues and eigenvectors of $\mathbf{S}$, where $k = 1, 2, \cdots, D$. If $\{\mathbf{u}_k\}_{k=1}^{M}$ are used as the bases of the subspace, then the variance of the projected data onto this subspace is given by $\sum_{k=1}^{M} \lambda_k$. In addition, the distortion measure J can also be derived as:

$$J = \sum_{k=M+1}^{D} \lambda_k \tag{4.3.1.7}$$

Therefore, by choosing the M eigenvectors of $\mathbf{S}$ corresponding to the M largest eigenvalues, the variance of the projected data is maximized while the distortion measure is minimized.

In practice, the choice of M depends on how much of the total variance is desired to be kept in the principal subspace. 90% is typically considered to be a good approximation. In fact, beyond that, the remaining 10% of the variance is mainly composed of noise, so PCA also helps reduce the effect of the noise.

One typical issue when applying PCA to the waveforms is alignment. Suppose $\mathbf{x}_n = [x_{n1}\ x_{n2}\ \cdots\ x_{nD}]^T$ gives the feature vector of length D. Then, the i-th feature x_{ni} should be the same type of feature for all $\mathbf{x}_n$ where $n = 1, 2, \cdots, N$. Unfortunately, MEP waveforms don't satisfy this. First of all, the segmented waveforms of MEPs don't have the same length. Secondly, the amplitude at every sampled time is simply a measurement of the voltage, and there is no information about how the

[2] This is why data $\{\mathbf{x}_n\}$ needs to be centered. Only when $\bar{\mathbf{x}} = 0$ will $\mathbf{S}$ given by above equation give the covariance matrix.

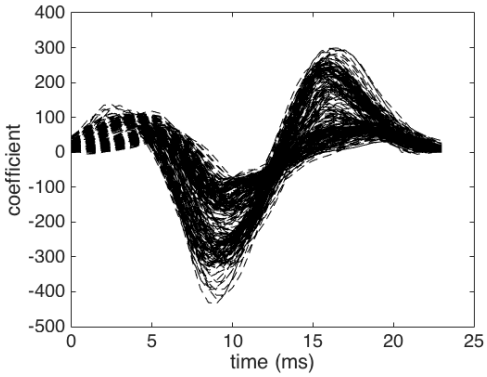
waveforms of two MEPs are correlated with each other. Normally, the alignment is performed by aligning the most significant feature of the waveform. For example, the largest peak of a MUAP is used as the alignment mark. This is not possible for MEPs because one MEP can have multiple significant peaks, and there is no single most significant peak. As a result, MEPs can't be aligned based on their peaks.

In this thesis, the alignment mark is chosen to be the weighted average of the time indices with weight being the amplitude square, as defined in Eq. (4.3.1.8). This definition removes the dependency on the peak locations of the MEPs, and therefore can be applied to MEPs with various shapes. In addition, this definition makes use of the shape information so that similar MEPs will be aligned correctly and consistently. After all the MEPs are aligned with the weighted time mean, the MEPs are adjusted to have the same length. The length is chosen to be the average of all MEPs. Hence, longer MEPs are truncated while shorter MEPs are padded with 0s.

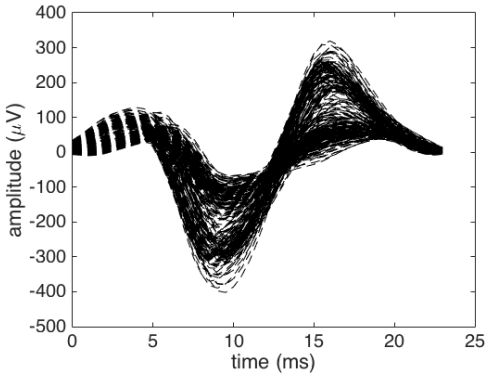
$$\tau_i = \frac{\sum_{j=1}^{D_i} j * w_{i,j}^2}{\sum_{j=1}^{D_i} w_{i,j}^2} \tag{4.3.1.8}$$

where $\tau_i \in \{1, 2, \cdots, N\}$ is the alignment mark of the i-th MEP, of which the time indices are $\{1, 2, \cdots, D_i\}$, and the waveform is $\mathbf{w}_i = \{w_{i,1}, w_{i,2}, \cdots, w_{i,D_i}\}$.

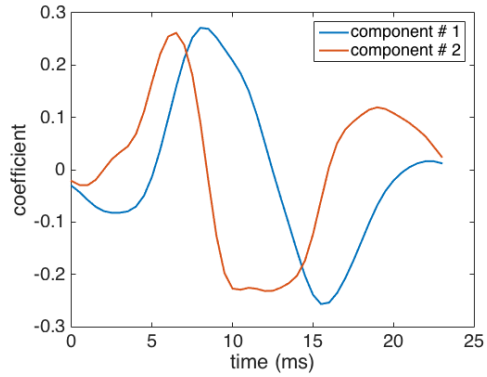
To illustrate the PCA process applied to the monosynaptic MEPs, the MEPs in Figure 4.2-2 are used for demonstration. The original MEP waveforms, aligned with respect to their weighted time means, are plotted in Figure 4.3-1a. Figure 4.3-1e shows the percentage of the variance each principal component (eigenvector) explains. The percentage of variance is simply the ratio of the eigenvalue corresponding to each eigenvector over the sum of all eigenvalues. As shown in Figure 4.3-1e, the first two eigenvectors already explain more than 95% of the total variance, so two principal components are good enough to retain most of the useful information in the original waveforms (See Figure 4.3-1c). Please note that the PCA is performed on the centered data (eg. data with mean subtracted). To reconstruct the original data, the mean has to be added back (See Figure 4.3-1d). The projection of the centered, aligned monosynaptic MEPs onto the principal subspace of dimension 2 is plotted in Figure 4.3-1f. This shows a side benefit of PCA, which is to help visualize high-dimensional data. Finally, the reconstructed waveforms are plotted in Figure 4.3-1b. Compare it with Figure 4.3-1a, one can tell the reconstructed waveforms are very close to the original waveforms, and hence a two-dimensional subspace is enough to capture most of the useful information in the original high-dimensional space.



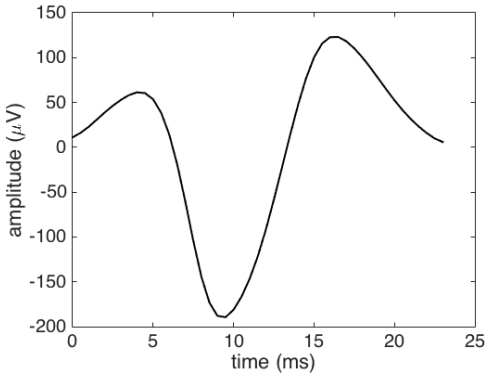
(a) original monosynaptic MEP waveforms



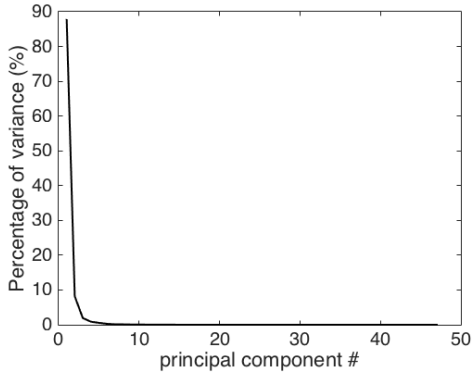
(b) reconstructed monosynaptic MEP waveforms



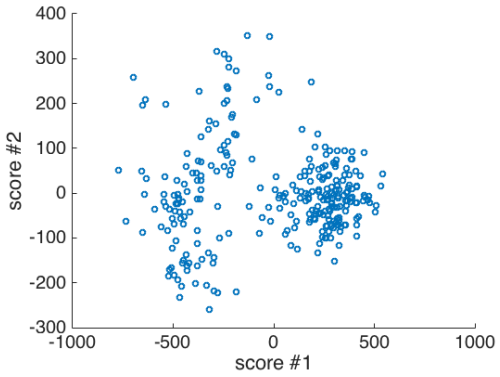
(c) coefficients of the first two principal components from PCA



(d) point-average waveform of the original monosynaptic MEPs



(e) percentage of the variance explained by every component from PCA



(f) projection of the original monosynaptic MEPs

Figure 4.3-1: PCA result on the 285 MEPs in the monosynaptic MEP cluster obtained from the initial clustering. (a) gives the original MEP waveforms aligned with respect to their weighted time means. The common length of all waveforms is 47 samples. (b) gives the MEP waveforms reconstructed from their PCA projections with a two-dimensional principal subspace. (c) gives the coefficients of the first two principal components. (d) gives the point-average waveform of the original MEP waveforms. (e) gives the percentage of the variance explained by every principal component. (f) gives the projection of the original MEP waveforms onto the two-dimensional principal subspace. Score 1 and score 2 give the coordinates in the principal subspace when projecting the original waveforms onto the first and second component, respectively.

4.3.2 Clustering with Gaussian Mixure Model

Clustering with Gaussian Mixture Model (GMM), or Mixtures of Gaussians, is a very popular and powerful technique. K-means clustering is actually a special case of GMM clustering. Unlike K-means clustering, which only works well if clusters of data form hyper-spheres, GMM can model clusters with hyper-ellipsoids of various orientations. In addition, K-means gives hard thresholding while GMM assigns the probability to every data point that measures the likelihood that each cluster explains the data. GMM has been used to cluster neural signals successfully [65]. A Gaussian distribution can accounts for the variability in the MEP waveforms. As a result, GMM has been used to cluster monosynaptic MEPs in this thesis.

A Gaussian Mixture Model is a probability distribution which is a linear superposition of Gaussian distributions with different weights:

$$p(\mathbf{x}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \tag{4.3.2.1}$$

where $\mathbf{x}$ is a random vector, π_k are called *mixing coefficients* and satisfy $\sum_{k=1}^{K} \pi_k = 1$, and $\mathcal{N}(\mathbf{x}|\mu_k, \Sigma_k)$ gives the multivariate normal distribution with mean μ_k and covariance matrix Σ_k. The above GMM has K components. For convenience, let $\boldsymbol{\pi} = \{\pi_k\}_{k=1}^{K}$, $\boldsymbol{\mu} = \{\boldsymbol{\mu}_k\}_{k=1}^{K}$, and $\boldsymbol{\Sigma} = \{\boldsymbol{\Sigma}_k\}_{k=1}^{K}$. Given a fixed number of components K, the set of parameters of a GMM is therefore $\boldsymbol{\Theta} = \{\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}\}$.

The goal is to estimate the parameters of a GMM from a given set of observations $\mathbf{X} = \{\mathbf{x}_n\}_{n=1}^{N}$. The most common approach is to maximize the log of the likelihood function (commonly referred as Maximum Likelihood, or ML) with respect to the model parameter $\boldsymbol{\Theta}$:

$$\ln p(\mathbf{X}|\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \sum_{n=1}^{N} \ln \left\{ \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \right\} \tag{4.3.2.2}$$

Maximizing the above function is a very complex task and does not yield a closed form solution. The most commonly used alternative approach for finding maximum likelihood solutions to GMM is called the *expectation-maximization* algorithm, or *EM* algorithm [4]. Various constraints can be put on the parameter $\boldsymbol{\Theta}$ when maximizing the likelihood function. The most common one is the free form optimization as described in [4], in which no constraints are put on the parameters. For covariance matrix, a parsimonious model can be used to apply different constraints, as discussed in details in [6]. For example, the covariance matrix $\boldsymbol{\Sigma}$ can be constrained to be diagonal. The parameters of all components can be constrained to be the same. Adding constraints to the model reduces the model complexity, and therefore reduces overfitting when the amount of data is small.

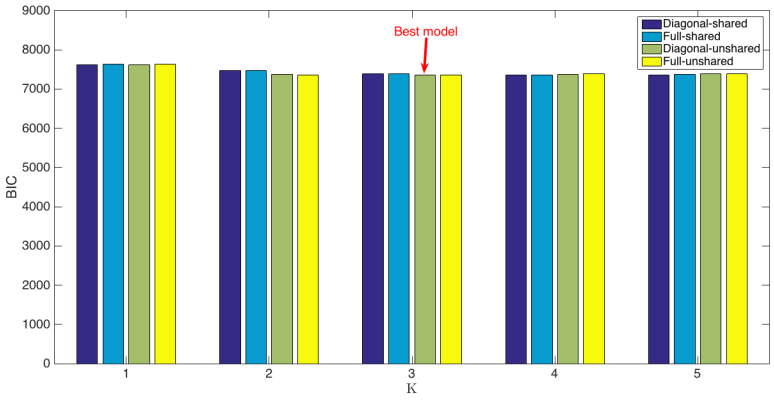


Figure 4.3-2: BIC for various K and Σ choices: Run clustering with GMM on the monosynaptic MEPs from Figure 4.2-2. The best model has 3 Gaussians ($K = 3$) with diagonal covariance matrix and non-shared parameters. The BIC associated with the best model is 7360, the smallest of all.

Model Class

In this thesis, the following models of Gaussian mixtures are considered:

- The number of Gaussian mixture components $K = 1$, 2, 3, 4, or 5.
- The covariance matrix $\mathbf{\Sigma}_k$ is diagonal or full.
- The parameters $\boldsymbol{\mu}_k$ and $\mathbf{\Sigma}_k$ are shared or not.

Overall there are $5 \cdot 2 \cdot 2 = 20$ models to choose from.

The more complex the model is, the better it fits the data, and the bigger the likelihood is. However, a complex model tends to overfit the data. As a result, a penalty has to be given based on the complexity of the model. *Bayesian information criterion* (BIC) is commonly used as a criterion for model selection among a finite set of models. The model with the lowest BIC is preferred. The BIC is defined as:

$$\text{BIC} = -2\ln(\hat{L}) + M\ln(N) \tag{4.3.2.3}$$

where $\hat{L}$ is the maximized value of the likelihood function of the model, M is the number of free parameters to be estimated, and N is the number of data points. The BIC calculated under different models for the clustering of monosynaptic MEPs is shown in Figure 4.3-2. For that particular data set, the best model has three Gaussians, with all having diagonal covariance matrices and non-shared parameters.

After the model is identified from the data, the clustering is done by evaluating the *responsibility*,

γ_{nk}, that component k takes for explaining data $\mathbf{x}_n$, as defined by:

$$\gamma_{nk} = \frac{\pi_k \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)} \tag{4.3.2.4}$$

$\mathbf{x}_n$ is assigned to cluster k if cluster k has the largest responsibility for explaining it. For a total of K clusters, the cluster index c_n of data $\mathbf{x}_n$ is given by:

$$c_n = \underset{k \in \{1,2,\cdots,K\}}{\arg\max} \ \gamma_{nk} \tag{4.3.2.5}$$

The clustering result is shown in Figure 4.3-3. From simple visual inspection, the clustering result looks good as it finds the subtle differences between different clusters of MEPs, and MEPs from every cluster share similar shapes. Of course, to accurately interpret the clustering result, one needs to consult neurophysiologists for expertise.

4.4 Decomposition of Overlapping MEPs

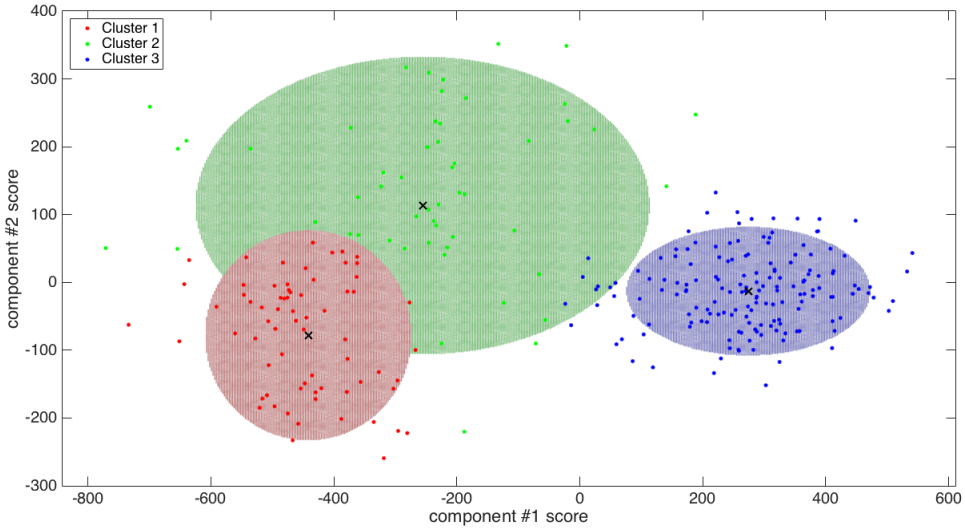
In the initial hierarchical clustering, the cutoff threshold is intentionally chosen to be small so that there is at least one cluster of only monosynaptic MEPs. As a result, some of the monosynaptic MEPs are "left out". This is especially true when a monosynaptic MEP is very close to, or even overlaps with, a polysynaptic MEP. In this case, the segmentation step produces a long segment with both monosynaptic MEPs and polysynaptic MEPs. The task in this step is to extract the remaining monosynaptic MEPs when two or more MEPs overlap in time.

The overall procedure is first stated, followed by a detailed discussion of every step. For every remaining MEP,

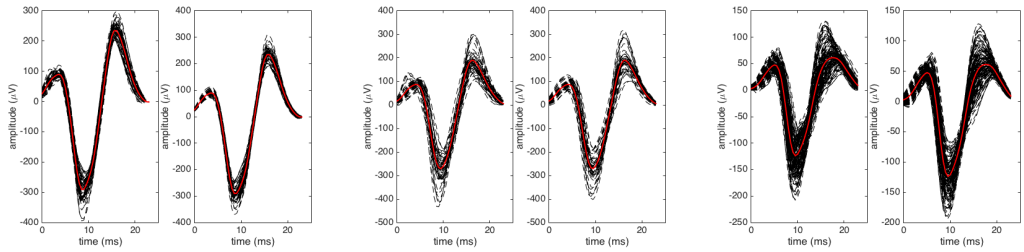
1. Find a segment of the waveform that gives the best matching with the cluster of monosynaptic MEPs. This segment is regarded as a candidate monosynaptic MEP.

2. Calculate the likelihood of given candidate MEP being from the clusters of monosynaptic MEPs using information of the shape and the latency. If the likelihood is above a certain threshold, the candidate MEP is labeled as a monosynaptic MEP.

3. Classify the newly added monosynaptic MEP using the Gaussian mixture model identified in Section 4.3.2.



(a) Projections of the MEPs and results of clustering with GMM



(b) cluster 1 (c) cluster 2 (d) cluster 3

Figure 4.3-3: Clustering result on monosynaptic MEPs with GMM: the monosynaptic MEPs are from Figure 4.2-2. PCA is performed to reduce the dimensionality before clustering (See Figure 4.3-1f). The GMM has three Gaussians with diagonal covariance matrices. There are 285 MEPs in total: cluster 1 has 163 MEPs; cluster 2 has 51 MEPs; cluster 3 has 71 MEPs. (a) gives the scatter plot of the projections of the MEPs onto a two-dimensional principal subspace. Each projection is color-coded to give its membership. For every component, the Gaussian distribution is also plotted with mean given by cross (X) and covariance given by ellipsoids that specifies a 99% probability threshold for confidence region; (b) - (d) Monosynaptic MEPs in Cluster 1 - 3 with original waveforms on the left and reconstructed waveforms on the right. The points average is plotted (in red) within each group of waveforms.

4.4.1 Identify Candidate Monosynaptic MEPs

The waveform of a remaining MEP is denoted as a vector of length D, $\mathbf{w} = [w_1\ w_2\ \cdots\ w_D]^T$. From the means of the K components in the GMM, cluster template waveforms can be reconstructed by Eq. (4.3.1.5). Denote the K cluster template waveforms as $\mathbf{w}^{(k)}$, where $k = 1, 2, \cdots, K$, and:

$$\mathbf{w}^{(k)} = \mathbf{U}\boldsymbol{\mu}_k$$

where $\boldsymbol{\mu}_k$ is the mean of the k-th component in GMM, and $\mathbf{U}$ is the base of principal subspace. Note that all vectors $\mathbf{w}^{(k)}$ and $\boldsymbol{\mu}_k$ are column vectors.

Cross-correlate $\mathbf{w}$ with $\mathbf{w}^{(k)}$ and find the best matching point, n_k, the point where the cross-correlation coefficient takes its maximum value, ρ_k.

$$n_k = \arg\max_{n\in\mathbb{Z}} (\mathbf{w} \star \mathbf{w}^{(k)})[n] \tag{4.4.1.1}$$

$$\rho_k = (\mathbf{w} \star \mathbf{w}^{(k)})[n_k] \tag{4.4.1.2}$$

where $(f \star g)[n]$ is the cross-correlation between two real-valued discrete signals $f[n]$ and $g[n]$:

$$(f \star g)[n] = \sum_{m=-\infty}^{+\infty} f[m+n]g[m] \tag{4.4.1.3}$$

which gives the dot product of $f[m]$ with lag n and $g[m]$.

Then find the component $\hat{k}$ that gives the maximum value of ρ_k: $\hat{k} = \arg\max_{k=1,2,\cdots,K} \rho_k$. Suppose the length of the aligned monosynaptic MEPs is $\bar{D}$, then the candidate monosynaptic MEP is $\mathbf{w}' = \{w_j\}_{j=n_{\hat{k}}}^{n_{\hat{k}}+\bar{D}-1}$. The other segments of $\mathbf{w}$ might contain polysynaptic MEPs, so create new MEP segments, $\mathbf{w}^{(1)}$ and $\mathbf{w}^{(2)}$:

$$\mathbf{w}^{(1)} = \{w_j\}_{j=n_1}^{n_{\hat{k}}-1} \tag{4.4.1.4}$$

$$\mathbf{w}^{(2)} = \{w_j\}_{j=n_{\hat{k}}+\bar{D}}^{D} \tag{4.4.1.5}$$

where $\mathbf{w}^{(1)}$ is the segment before $\mathbf{w}'$, while $\mathbf{w}^{(2)}$ is the segment after $\mathbf{w}'$. If there are none or very few samples on either sides, then the candidate intervals $\mathbf{w}^{(1)}$ or $\mathbf{w}^{(2)}$ are not meaningful. Figure 4.4-1 shows an example of breaking down a long segment into a candidate monosynaptic MEP and a candidate polysynaptic MEP. In that case, no segments exist before $\mathbf{w}'$, and only one segment exists after $\mathbf{w}'$. In this example, $\mathbf{w}'$ is successfully accepted to the cluster of monosynaptic MEPs in next

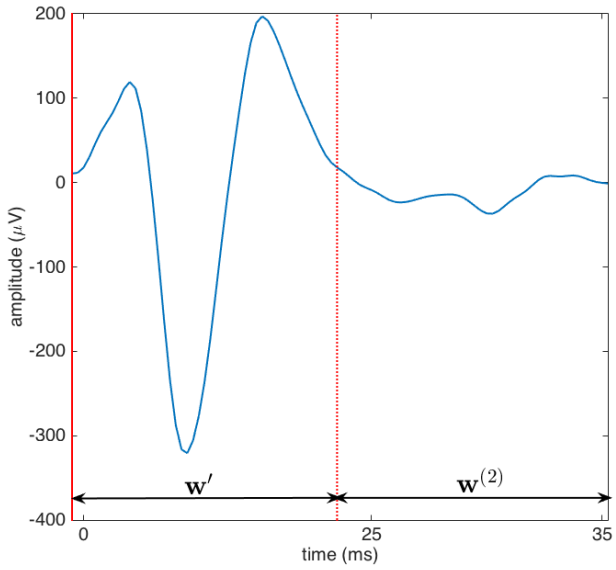


Figure 4.4-1: Breaking segment into candidate monosynaptic and polysynaptic MEP: $\mathbf{w}'$ is the candidate monosynaptic MEP found by cross-correlation. There is not a segment before $\mathbf{w}'$. $\mathbf{w}^{(2)}$ is the remaining segment after $\mathbf{w}'$, and will be analyzed as a potential polysynaptic MEP: red solid line and red dotted line indicate the location of $n_{\hat{k}}$ and $n_{\hat{k}} + \bar{D} - 1$, the beginning and end of the candidate monosynaptic MEP $\mathbf{w}'$, respectively

step.

4.4.2 Determine if the Candidate is a Monosynaptic MEP

After a segment of candidate MEP is identified, the likelihood of it being from the monosynaptic MEP cluster needs to be quantitatively evaluated. The Gaussian mixture model can give the likelihood based on the waveform of the MEP, but it does not use the latency information, which is crucial in determining the type of the MEP. In the following, the waveform and the latency information are combined to give a quantitative measurement of the likelihood.

For convenience, some of the notations introduced in Section 4.1 are restated here, together with some new notations. An EMG signal takes the format of N samples (discrete-time signal): $\mathbf{X} = \{x_1, x_2, \cdots, x_N\}$. Suppose the cluster of monosynaptic MEPs contains N_M MEPs, $\mathcal{W}_{MEP} = \{\mathbf{w}_i\}_{i=1}^{N_M}$, where $\mathbf{w}_i$ is the waveform of the i-th MEP of length D_i: $\mathbf{w}_i = w_{i,1}, w_{i,2}, \cdots, w_{i,D_i}$. The timing of the stimulus which is presumed to elicit the response in $\mathbf{w}_i$ is denoted as $a_i \in \{1, 2, \cdots, N\}$ that represents the (discrete-time) index in the EMG signal $\mathbf{X}$. The latency of the i-th MEP, which is the time from the occurrence of the electrical stimulus (a_i) to the MEP ($\mathbf{w}_i$), is denoted as d_i. The complete information about the i-th MEP is thus $\mathcal{M}_i = \{\mathbf{w}_i, a_i, d_i\}$, which includes the waveform, the associated stimulus, and its latency to the stimulus.

The probability model for the waveform can be given by the Gaussian mixture model, $\mathbf{\Theta} = \{\boldsymbol{\pi}, \boldsymbol{\mu}, \mathbf{\Sigma}\}$. The probability model for the latency is given by the Gaussian distribution, $\mathbf{\Theta}_d = \{\mu_d, \sigma_d^2\}$ where subscript d indicates latency. The principal subspace of the MEP waveform space (the Euclidean space with the dimension of the common length of the MEP waveforms) is given by $\mathbf{U}$ (Refer to Section 4.3.1 for feature extraction via PCA). Assume the waveform is independent of the latency, and then the probability of a given MEP being a monosynaptic MEP is:

$$p(\mathbf{w}_i, d_i; \mathbf{\Theta}, \mathbf{\Theta}_d, \mathbf{U}) = p(\mathbf{U}^T\mathbf{w}_i|\mathbf{\Theta}) \cdot \mathcal{N}(d_i|\mu_d, \sigma_d^2) \tag{4.4.2.1}$$

where $p(\mathbf{U}^T\mathbf{w}_i|\mathbf{\Theta})$ is the Gaussian mixture model given by Eq. (4.3.2.1). Then the log likelihood is:

$$\ln\left\{p(\mathbf{w}_i, d_i; \mathbf{\Theta}, \mathbf{\Theta}_d, \mathbf{U})\right\} = \ln\left\{p(\mathbf{U}^T\mathbf{w}_i|\mathbf{\Theta})\right\} + \ln\left\{\mathcal{N}(d_i|\mu_d, \sigma_d^2)\right\} \tag{4.4.2.2}$$

The parameters of the Gaussian distribution of the latency $\mathbf{\Theta}_d$ are estimated by:

$$\hat{\mu}_d = \frac{1}{N_M}\sum_{i=1}^{N_M} d_i \tag{4.4.2.3}$$

$$\hat{\sigma}_d = \sqrt{\frac{1}{N_M - 1}\sum_{i=1}^{N_M}(d_i - \hat{\mu}_d)^2} \tag{4.4.2.4}$$

which gives the unbiased estimation of the population mean and the population variance.

The threshold for accepting a given MEP as a monosynaptic MEP is given by:

$$T = \gamma\left(\min\left\{\ln\left\{p(\mathbf{U}^T\mathbf{w}_i|\mathbf{\Theta})\right\}\right\} + \min\left\{\ln\left\{\mathcal{N}(d_i|\mu_d, \sigma_d^2)\right\}\right\}\right) \tag{4.4.2.5}$$

where γ is a factor that can be adjusted to either tighten or relax the threshold. For example, in the thesis, $\gamma = 1.5$ is found to yield good results. Note that the log likelihood is a negative number, so a $\gamma > 1$ relaxes the threshold.

For a given new MEP $\mathcal{M}_{new} = \{\mathbf{w}_{new}, a_{new}, d_{new}\}$, $\mathcal{M}_{new}$ is said to be monosynaptic MEP if:

$$\ln\left\{p(\mathbf{w}_{new}, d_{new}; \mathbf{\Theta}, \mathbf{\Theta}_d, \mathbf{U})\right\} \geq T \tag{4.4.2.6}$$

4.4.3 Classify the New Monosynaptic MEP

The newly added monosynaptic MEP, $\mathbf{w}_{new}$, is classified to one of the K clusters determined by the K components of Gaussian mixtures. The posterior probability of the given MEP being from the k-th component is γ_k:

$$\gamma_k = \frac{\pi_k \mathcal{N}(\mathbf{U}^T \mathbf{w}_{new} | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\mathbf{U}^T \mathbf{w}_{new} | \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)} \tag{4.4.3.1}$$

This is the adaptation of Eq. (4.3.2.4) by replacing $\mathbf{x}_n$ with the PCA projection of the new MEP waveform, $\mathbf{U}^T \mathbf{w}_{new}$.

$\mathbf{w}_{new}$ is assigned to the component with the largest posterior probability for the observation.

$$\hat{k} = \underset{k \in \{1,2,\cdots,K\}}{\arg\max} \ \gamma_k \tag{4.4.3.2}$$

In Section 4.1, 469 MEPs were detected and segmented from 30 seconds of the EMG signal which was recorded from left medial gastrocnemius (L MG) of a patient with clinically motor complete spinal cord injuries while lying in the supine position with EES. The stimulation amplitude is 7.2V and frequency is 10Hz. Among the 469 MEPs, 285 MEPs were placed to the monosynaptic MEP cluster in the initial clustering in Section 4.2.2. After running the above decomposition procedure on the remaining 184 MEPs, additional 15 MEPs were placed to the monosynaptic MEP cluster, so the monosynaptic MEP cluster has 300 MEPs. This is expected as there are 300 stimulation intervals from 30 seconds of the EMG signal with stimulation frequency being 10Hz, and normally there is one monosynaptic MEP within one stimulation interval. The remaining 184 MEPs are considered candidate polysynaptic MEPs and will be clustered via GMM in the next section.

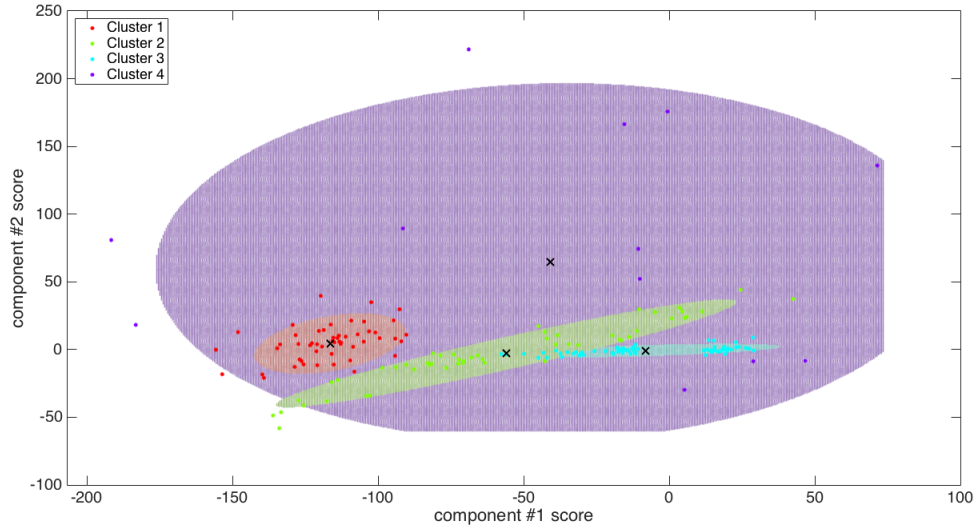
4.5 Clustering of Polysynaptic MEPs

Clustering of polysynaptic MEPs undergoes the same set of procedures as the monosynaptic MEPs as described in Section 4.3. Here is a brief recap of the procedures.

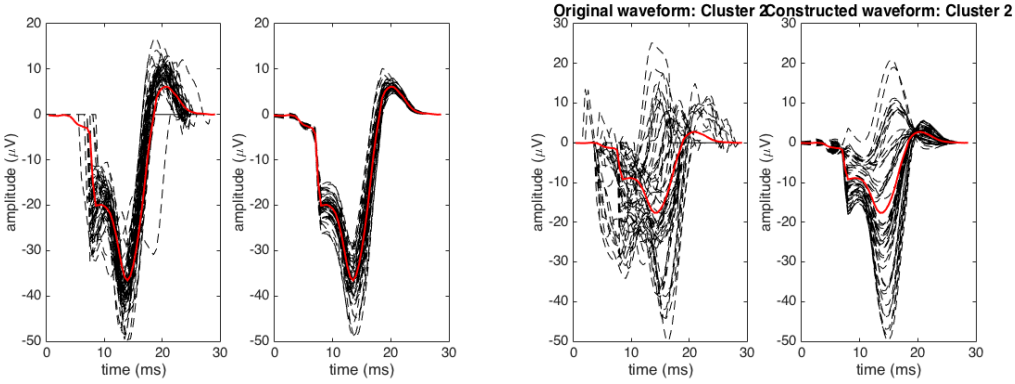
1. Align the MEP waveforms based on its weighted time mean. Run PCA and project waveform onto the principal subspace (with dimensionality of 2).
2. Perform maximum likelihood (ML) optimization on a set of Gaussian mixture models via EM algorithm. Select the best model.
3. Cluster the PCA projections of the MEPs using the selected Gaussian mixture model.

An example clustering result is shown in Figure 4.5-1. PCA is performed to reduce the dimensionality before clustering. The best model has four Gaussians with full covariance matrices. There are 184 polysynaptic MEPs in total: cluster 1 has 52 MEPs; cluster 2 has 52 MEPs; cluster 3 has 68 MEPs; cluster 4 has 12 MEPs.

Although polysynaptic MEPs are very weak, and do not have consistent latencies or shapes even within an event, the clustering successfully partitions them into different groups (See Figure 4.5-1). Because the SNR of the polysynaptic MEPs is low, the real MEP waveforms are tempered by

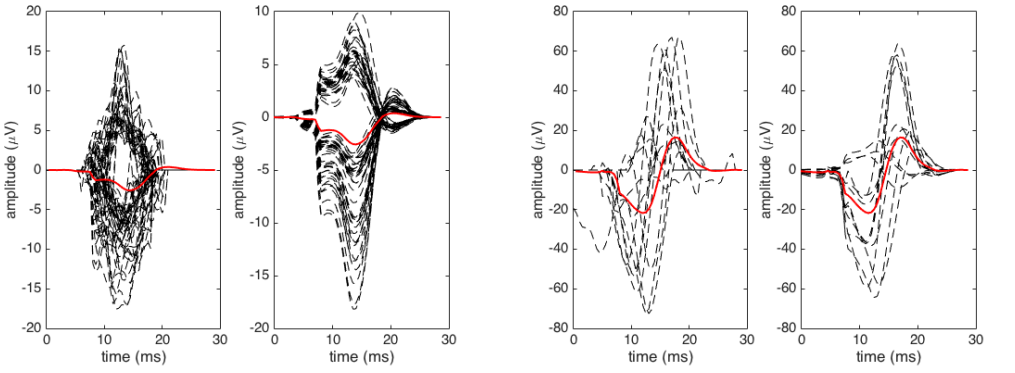


(a)



(b) cluster 1

(c) cluster 2



(d) cluster 3

(e) cluster 4

Figure 4.5-1: Clustering result on polysynaptic MEPs with GMM: the polysynaptic MEPs are from the same EMG signal as the monosynaptic MEPs in Figure 4.2-2. (a) gives the scatter plot of the projections of the MEPs onto a two-dimensional principal subspace. Each projection is color-coded to give its membership. For every component, the Gaussian distribution is also plotted with mean given by cross (X) and covariance given by ellipsoids that specifies a 99% probability threshold for confidence region; (b) - (d) Monosynaptic MEPs in Cluster 1 - 4 with original waveforms on the left and reconstructed waveforms on the right. The points average is plotted (in red) within each group of waveforms.

noise severely, which make them visually hard to analyze by physiologists. The use of PCA not only reduces the feature dimensionality, but also reduces the effect of the noise. In Figure 4.5-1b to Figure 4.5-1e, the reconstructed waveforms are shown on the right hand side of the original waveforms. The reconstructed waveforms are much smoother, and the structure or shape of the MEP waveforms is more prominent.

The clustering result is especially satisfying considering that these clusters don't visually separate from each other in the scatter plot. As a result, the K-means clustering would obtain completely wrong results. This example shows the flexibility and powerfulness of the GMM.

Because polysynaptic MEPs are weak, there are more false positives in detecting transient peaks of the polysynaptic MEPs. The error in the peak detection will be carried over to the segmentation phase and eventually the clustering phase. The threshold in the peak detection step is intentionally chosen to be small enough to yield a high recall, but a low precision. A high recall guarantees that most of the true MEPs are detected. A low precision implies more noise samples are picked up as MEPs, but this can be corrected in the following clustering phase, because noise samples have similar statistics and tend to show up in one cluster. The cluster of noise samples can be removed upon inspection by the physiologists.