METHOD FOR CONCURRENT PROCESSING OF EMG SIGNALS FROM MULTIPLE MUSCLES FOR IDENTIFICATION OF SPASMS

F. Sikder
Research & Development
3Z Telecom Inc.
Miramar, FL - 33025
Email: fsikder@3ztelecom.com

D. Sarkar and O. Schwartz
Department of Computer Science
University of Miami
Coral Gables, FL - 33146
Emails: {sarkar, odelia}@cs.miami.edu

C. K. Thomas
The Miami Project to Cure Paralysis
University of Miami
Miami, FL - 33136
Email: cthomas@miami.edu

Abstract—Involuntary contractions are common in muscles paralyzed by spinal cord injury (SCI). These contractions may impact joint movements and upset daily tasks. Existing rule-based algorithms for counting the number of muscle contractions from electromyographic (EMG) signals work on one channel at a time. However, to understand activation of muscles during involuntary contractions, it is important to develop algorithms that can process signals from multiple muscles simultaneously.

To characterize these contractions, EMG signals recorded from paralyzed muscles were analyzed. First, existing single-channel signal processing techniques were applied to each EMG recording. Then an Eigenvalue decomposition technique on a block of signals from all muscles was used to extract features of the signal-block. An extended version of the KL-distance measure (a method to compute the distance between two probability mass distributions) was used to find the distance between two adjacent sets of Eigenvalues. The regions with significant distances were marked as potential areas for identification of muscle contractions. We further developed algorithms to identify co-activation of muscles and the muscle that was activated first, since this muscle may be targeted in interventions that aim to dampen these contractions.

These algorithms were tested on five hours of EMG data (2:00 am to 7:00 am) recorded from five paralyzed (no voluntary control) leg muscles of a person with SCI. Most of the myoclonus spasms that involved contractions containing clonic-like EMG were identified accurately. Although the soleus muscle was often co-active, on average 74.31% of the time, it initiated the contraction in only 60.91% of the cases. In contrast, the tibialis anterior muscle was co-activated 37.92% of the time, on average, but was the first muscle to respond 85.48% of the time. While the proposed approach has shown great potential for concurrent analysis of multi-muscle EMG recordings, we want to continue testing on larger data sets and for other spasm types.

I. INTRODUCTION

Involuntary contractions (spasms) are common in muscles paralyzed by spinal cord injury (SCI). These involuntary muscle contractions can occur at anytime, disrupt daily activities, and sleep. They often lower the quality of life. Most spasms involve co-activation of multiple muscles after SCI [1]. To study the characteristics of these contractions, Electromyographic (EMG) signals recorded from paralyzed muscles have been analyzed by experts. One kind of spasm that is common during sleep is myoclonus [2].

To get further insights into the nature of muscle co-activation during spasms, 24 hour EMG data sets from several SCI participants have been collected [2]. But these data sets are too large for manual analysis [3]. Unfortunately, the methods for analyzing such EMG data sets are not as mature as those for analyzing electrocardiogram (ECG) [4], and electroencephalogram (EEG) [5], [6] data sets. To fill this gap, a rule-based method for identification and classification of spasms has been reported [3]. But this method works on one channel at a time, even though muscle co-activation is common and may result in joint movements and disruption of tasks.

For multiple reasons, concurrent processing of data from all muscles is essential. Consider a simple case where EMG signals from two muscles have been processed sequentially, that is one channel at a time. Suppose that each muscle had 50 spasms. This raises the question: How many spasms did the person with SCI experience? It could be only 50 if each contraction involved co-activations of both muscles. On the other hand, it could be 100 if no spasm had co-activation.

Consider another question: Suppose that the number of spasms experienced by the person with SCI was 75, that is, 25 spasms involved co-activations of both muscles, while 25 spasms involved activation of each muscle independently. Among the 25 spasms where co-activation occurred, did the contractions in one muscle always occur before the activation of the other muscle? Or was it a combination of both? Did the spasms feel stronger when the muscles were co-activated? Understanding this muscle co-activation can help to characterize how individuals count spasms, and when linked to the injured person’s perception of the spasm count, may lead to identification of those spasms that need management.

In general, the spasm count from EMG recordings exceeds the count-scales used clinically [2]. The spasm frequency ratings people give (a measure of spasm count) correlate to EMG duration and/or the time of agonist/antagonist muscle co-activity. Thus, people likely do not count what is happening in one muscle at a time, but rather the movement that is happening about a joint or limb [1].
Over the years, various techniques have been developed to analyze EMG signals to count the number of muscle contractions. Both time domain and frequency domain (as well as a combination of the time and frequency domains) analysis techniques have been used [2], [3], [7]–[12].

Time domain data analysis is a very conventional way to analyze EMG signals and it is associated with the amplitude of the signals. Also, rectification, integration or computation of root-mean-square values of EMG signals is often used to extract information from EMG signals [3], [12]. To analyze EMG recordings in the frequency-domain, the Fourier transform has been the most widely used method.

On the other hand, short-term Fourier transforms (STFTs) and wavelets process the data in both the time and frequency domain. The Morlet wavelets have been used for analysis of EMG data. They are scaled linearly when the time resolution of events is unknown. Vincent Von Tscharner [9] developed a technique for non-linear scaling of wavelets, which is very useful and effective to locate the timing of events in the EMG signals [12]. In our work, we use this method as one of several steps for preparing the EMG signals. After preparing the EMG signals, we extract features from these (prepared) signals and use these features in our main algorithm (see Section II-B3c).

A. Problem Statement

A spinal cord injured individual can experience myoclonus (rhythmic, repetitive involuntary muscle contractions) in only one muscle or across multiple muscles at the same time, particularly during sleep. In EMG recordings from five muscles of an individual with SCI, we address two problems: (i) How can the presence of contractions in one or more muscles during myoclonus be detected in the EMG recordings? (ii) If an identified region has clonic-like contractions in multiple muscles, how can the muscle that initiates the contractions be identified?

The proposed solution to the first problem has three steps: (a) data preparation for feature extraction (Section II-B3a), (b) feature extraction (Section II-B3b), and (c) Log-Sum distance computation and its application to identify locations of EMG bursts (Section II-B3c). Our algorithm for solving the second problem — identification of the muscle that contracts first — utilizes envelopes of EMG signals [12] and the Log-Sum distance [13].

In this research, we focus first on detection of the contractions that are characteristic of myoclonus. In the rest of the paper, we examine the clonic-like bursts of the EMG that occur with contractions that make up the myoclonus (see Fig. 1 for a detailed illustration).

For data preparation, we utilize observations reported in [2], [3], [11], [12]. Similar to them, we use non-linearly scaled Morlet wavelet filters for retaining signals in the 74.8-193.9 Hz frequency band. Also, to identify clonic-like EMG in individual contractions in myoclonus spasms we use envelopes of the EMG signals that 1) have frequencies between 4Hz and 12Hz and 2) have EMG burst durations between 40ms to 90ms, because they are typical clonus characteristics [14].

II. MATERIALS AND METHODS

A. Materials

For detailed descriptions of the data collection procedures and equipment, readers are directed to [2]. Here is a brief summary. The EMG was sampled at 1000Hz from eight muscles over 24-hours, but only data from five different leg muscles were analyzed here (Medial gastrocnemius, MG; Tibialis anterior, TA; Biceps femoris, BF; Vastus lateralis, VL; Soleus, SL; referred to as muscles 1,2,3,4, and 8 in the graphs and tables). For the evaluation of our proposed methods, we used only 5 hours (2:00 am to 6:59 am) of the EMG data from each muscle obtained while one individual with chronic (> 1 year) SCI at C8 was asleep. All five muscles were paralyzed (under no voluntary control) so all of the EMG was involuntary.

All procedures had approval from the University of Miami Institutional Review Board and the participant gave informed written consent to participate.

B. Methods

For completeness, the steps for data preparation are outlined in Section II-B3a. The data obtained were then divided into smaller blocks for Eigenvalue-decomposition. We used the singular value decomposition (SVD) method to calculate Eigenvalues (see Section II-B2), because this leads to faster computation [15]. The Eigenvalues were then used to identify areas of the EMG that have potential muscle contractions. Then a recently developed extension of the KL-distance measure was used to compute the distance between two sets of Eigenvalues obtained from adjacent blocks of preprocessed data (see Section II-B1).
1) Log-Sum Distance Measure: Our recently developed Log-Sum distance measure is an extended version of the KL-distance measure [13]. It measures the distance between two sequences of positive numbers (including two sets of probability mass distributions), whereas the KL-distance measures the distance between two probability mass distributions. Since the Eigenvalues obtained after feature extraction are a sequence of positive numbers and not a probability mass distribution, KL-distance cannot be used. A formal definition of Log-Sum distance is provided next.

**Definition 1.** Let \( U = < u_1, u_2, \ldots, u_m > \) and \( V = < v_1, v_2, \ldots, v_m > \) be two sequences of positive numbers. Log-sum distance, \( LD(U||V) \), between them is defined as,

\[
LD(U||V) = \sum_{i=1}^{m} u_i \log \frac{u_i}{v_i} + \sum_{i=1}^{m} v_i \log \frac{v_i}{u_i}.
\]  
\[ (1) \]

Similar to KL-distance, it has been shown in [13] that \( LD(U||V) \) is non-negative. Formally,

**Property 1.** If \( U = < u_1, u_2, \ldots, u_m > \) and \( V = < v_1, v_2, \ldots, v_m > \) are two sequences of positive numbers, then

\[
LD(U||V) \geq 0
\]  
\[ (2) \]

Equality holds if and only if \( u_i = v_i \) for all \( 1 \leq i \leq m \).

For a proof see [13].

2) Computation of Eigenvalues using SVD: Let \( X = [X_1, X_2, \ldots, X_m] \) be a matrix of \( n \) rows and \( m \) columns, where each \( X_j = [x_{1j}, x_{2j}, \ldots, x_{nj}]^T \), for \( 1 \leq j \leq m \), is a column of \( n \) elements.

Singualr value decomposition of \( X(n \times m) \) can be written as:

\[
X = USV^T
\]  
\[ (3) \]

For \( n > m, U \) (\( n \times m \)) and \( V \) (\( m \times m \)) are orthogonal matrices and \( S \) (\( m \times m \)) is a diagonal matrix.

Let us construct a matrix \( A \) from \( X \) as follows:

\[
A = X^TX; A \in \mathbb{R}^{m \times m}
\]  
\[ (4) \]

\( A \) is symmetric, because

\[
A^T = (X^TX)^T = (X)^T(X^T)^T = X^TX = A.
\]

Since every square symmetric matrix of real-value elements is orthogonally diagonalizable, if we decompose \( A \) for Eigenvalues we have,

\[
A = Z\Lambda Z^T,
\]  
\[ (5) \]

where \( Z \) and \( \Lambda \) represent orthogonal and diagonal matrices, respectively. Since \( A \) is a positive symmetric matrix, its Eigenvalues are nonnegative and ordered from high to low, and the first Eigenvalue is the largest.

For computing \( Z \) and \( \Lambda \), we can use singular value decomposition of \( X \). Noting that \( A = X^TX \), and then using Equation 3, we have,

\[
A = X^TX = V\Lambda SV^T = VS^2V^T
\]  
\[ (6) \]

Now comparing Equations (5) and (6), we get the following relationship between singular values and Eigenvalues [15]:

\[
S^2 = \Lambda.
\]  
\[ (7) \]

Since in our case \( n >> m \), we compute \( S \) and then square each value to get the desired Eigenvalues. This approach reduces the computation time significantly.

In the next section, we use these Eigenvalues as the feature vector in our proposed Log-Sum Distance based EMG bursts algorithm.

3) Algorithms to Detect EMG Bursts: There are four steps:

(a) Data preparation, (b) Eigendecomposition, (c) Log-Sum distance computation, and (d) identification of contractions that have EMG bursts.

a) Data Preparation for Feature Extraction: In the data preparation phase we apply short-term Fourier transforms (STFTs) and a wavelet filter to each individual channel of EMG. First, we use STFTs to filter out high and low frequencies from our input EMG signals. Then we use a non-linearly scaled Morlet wavelet filter to obtain the EMG envelope.

![Fig. 2: An example of envelopes for five muscles that were calculated using the Morlet wavelet filter for a segment of EMG data. Three muscles show clear EMG bursts (bottom to top: MG (C=1); TA (C=2); SL (C=8)). A few potentials are seen in BF (C=3), while VL (C=4) is inactive.](image)

As in [12], the wavelets we use have a pass-band of 74 - 194 Hz. Figure 2 shows an example of the enveloped EMG signals on five muscles.

Note that the original EMG signals are not used in the next step. We use envelopes of the EMG for computing Eigenvalues, as discussed in the next section.

b) Feature Extraction from envelopes of EMG-Signals: From the EMG recording envelopes of \( m \) channels, \( n \) consecutive samples are selected from each muscle to create a \( n \times m \) matrix. In our case \( n >> m \). We use an overlapping window of \( w \) to get the next matrix, that is, two adjacent matrices that share \( w \) columns form the envelopes of \( m \) EMG muscles.

Algorithm 1 computes features from envelopes of EMG data sets. The input to the algorithm is a sequence of \( p \)
matrices, $X = <X^{(1)}, X^{(2)}, \ldots, X^{(p)}>$ of size $n \times m$ each, constructed from envelopes of EMG recordings. In step 4, the algorithm uses Equation (3) to compute $S^{(i)}$, and then Equation (7) to compute the diagonal matrix $\Lambda$. The values of the diagonal are the Eigenvalues. The algorithm forms a column vector $z^{(i)}$ from the Eigenvalues of the data matrix $X^{(i)}$. These column vectors are stored in $z$ for use in Algorithm 2, which computes Log-Sum distances.

$$z = <z^{(1)}, z^{(2)}, \ldots, z^{(i)}, \ldots, z^{(p)}>.$$  

**Algorithm 1 Calculate Eigenvalues ($z$)**

1. procedure EXTRACTFEATURES
2. Input: $X = <X^{(1)}, X^{(2)}, \ldots, X^{(p)}>$
3. Output: $z = \{z^{(2)}, z^{(2)}, \ldots, z^{(p)}\}$
4. for each $X^{(i)} \in X$
5. Compute $S^{(i)}$ from $X^{(i)}$ using Equation (3)
6. Compute $\Lambda^{(i)}$ from Equation (7)
7. Extract Eigenvalues from $\Lambda^{(i)}$ and assign them to $z^{(i)}$
8. end procedure

**Algorithm 2 Compute Log-Sum Distance ($D_{LD}$)**

1. procedure COMPUTELOGSUMDISTANCE
2. Input: $z = \{z^{(1)}, z^{(2)}, \ldots, z^{(p)}\}$
3. Output: $D_{LD} = <d^{(1)}_{LD}, d^{(2)}_{LD}, \ldots, d^{(p-1)}_{LD}>$
4. for $i = 1:p-1$
5. Compute $d^{(i)}_{LD}$ using Log-Sum Distance
6. Equation (1) from $z^{(i)}$ and $z^{(i+1)}$
7. end procedure

c) Algorithm to Detect EMG Burst Locations using Log-Sum Distances: In the first phase of EMG Bursts location detection, we use Algorithm 2 to get the Log-Sum Distance between two adjacent sets of Eigenvalues $z^{(i)}$ and $z^{(i+1)}$. The algorithm takes Eigenvalue sequence $z$ as an input and outputs $D_{LD}$, a sequence of distances. Two adjacent Eigenvalues, $z^{(i)}$ and $z^{(i+1)}$, are used to compute their Log-Sum distance $d^{(i)}_{LD}$. Thus, a sequence of Log-Sum distances, $D_{LD}$, are obtained.

Figure 3 displays an example of 8 seconds of EMG from five muscles, envelopes of the signals, and the corresponding Log-Sum distances (bottom green trace). All the readings in the figure are normalized to ‘one’ for the ease of presentation. It can be observed that Log-Sum distance values are higher where EMG bursts occur. In the next phase, we search through all the Log-Sum Distances, $D_{LD}$, to find these higher Log-Sum distance regions and to thus locate the EMG bursts.

In the second and final phase, we search through each $d^{(i)}_{LD}$ within the areas where EMG bursts were located. This search leads us to the location of the beginning and end of the EMG bursts locations. All the steps involved in detecting the EMG bursts locations from Log-Sum distance values $D_{LD}$ are shown in Fig. 4. For this purpose, we compute the average Log-Sum distance for each 5-minute window of data and use it to set thresholds for the Log-Sum Distance values. All the values $d^{(i)}_{LD}$ above the threshold are the regions where we have active muscles (EMG signals). We mark all values above the threshold ‘one’ (to specify that these are potential areas for EMG bursts) and mark all values below the threshold ‘zero’. In the case of EMG bursts, all of the higher Log-Sum Distance values are closer to each other, because several muscles are usually active when there are EMG bursts (see Fig. 3).

Now, among those values that have passed the threshold, we search for clusters of locations that are positive (marked ‘one’). We do this for segments of one second at a time. If one second of Log-Sum Distance values have at least 15 locations marked
'one', we detect that region as an EMG burst region\(^1\). For the next cluster search, we keep an overlap of half a second as in [3]. This increases the resolution of the search and helps us to detect long-duration EMG bursts. Figure 5 shows a detected EMG burst region inside the red curve.

![Fig. 5: A region with EMG bursts from the 5th hour, between 2,194 and 2,299 seconds.](image)

\(d\) Identifying the muscle that has the First EMG Burst: In the previous Section, we identify the EMG bursts locations in one or more of the EMG recordings. Now an algorithm is presented to identify the muscle that responds first in the contraction. To identify the start of the EMG burst, we take one identified EMG burst region at a time and then process the data from each muscle. If a muscle is active and the largest peak in the envelope is within the area where the frequency range of the EMG bursts is from 4 to 12 Hz, we identify that this muscle has EMG bursts. For each active muscle, we mark the start time of the first EMG burst. Finally, we identify the muscle that has the lowest start time as the muscle that initiates the contraction in that region. Figure 6 shows the marked start positions for each active muscle in one contraction during a myoclonus spasm. Here the \textit{tibialis anterior} (marked as \(C=2\)) initiates the contraction.

![Fig. 6: Start locations of EMG bursts marked in each muscle. This data is taken from the 5th hour, between 1,273 and 1,278 seconds.](image)

**III. Results**

We evaluate the performance of the proposed algorithm in Section III-A and then report execution time in Section III-B.

For the results reported here, the values of number of samples \(n\), number of channels \(m\), and overlap-window \(w\) are 100, 5, and 80 respectively.

**A. Performance Evaluation**

The accuracy of our algorithm is calculated using Equation (9). One of the authors manually counted the number of contractions containing clonic-like EMG bursts in the five-hour data. It is important to note that EMG recordings from all five muscles were displayed simultaneously on a computer screen. If one or more muscles had clonic-like EMG bursts within a contraction, it was included in the count. If both our algorithm and the person identified EMG bursts in the same area, we counted this event as a \textit{True Positive (TP)}. If the algorithm identified a contraction containing EMG bursts but the person did not believe the area had EMG bursts, we counted this as a \textit{False Positive (FP)}. When the algorithm failed to identify a contraction that the person believed had EMG bursts, we counted this event as a \textit{Missed positive (M)}.

\[
\text{Accuracy} = \frac{N_{TP}}{N_{TP} + N_{FP} + N_{M}} \quad (9)
\]

1) Total Number of Contractions containing EMG Bursts: Table I shows the total number of contractions identified as including EMG bursts in one or more muscles using our algorithm for the five hours of data. From the table, we find that our algorithm achieved an accuracy of 79.3%. It missed 11 out of 279 contractions that the person identified as having EMG bursts, which is about 4%. Our algorithm had 59 false positives, which is only 18% of the contractions with EMG bursts detected by a person.

2) Identification of the Muscle that Started the Contraction: Table II shows the number of times each muscle had clonic-like EMG bursts during a contraction. It also shows other related statistics, including the number of times a muscle with EMG bursts was the first muscle activated. Some of the salient observations are discussed next.

From the first row of the table, we observe that SL (\(C=8\)) had the highest number of contractions, including clonic-like

\[^1\]The number of EMG bursts in one second of EMG data is a tuning parameter of our algorithm. For the results reported here, 15 EMG bursts in a second is quite effective, but we plan to do more studies to understand why it works. Also, we plan to develop a method for automatic selection of this parameter.
EMG bursts (243). The second row shows the percentage of time a muscle was activated during contractions containing EMG bursts. For example, out of 327 contractions that had EMG bursts detected by the algorithm, SL had 243 of the contractions, which is 74.31%. One should note that the sum of values in row two is much higher than 100%, because multiple muscles were active during some contractions. From the third row, we find that TA also had the highest number of contractions during which it was the first muscle to respond — 148. But the muscle that was most often activated first during the contractions was TA (C=2) — 85.48%.

B. Execution Time

The computer we used has one Intel Core i7-6800 K CPU with 3.40 GHz clock, 64GB of memory, and was running on a 64-bit Windows 7 Enterprise operating system. Our programs were developed in the MATLAB R2010b platform.

For double precision floating point representation 24-hour of data from five muscles would require about 3.5 GB of memory. To make our programs memory-efficient, one hour of data from five muscles would require about 3.5 GB of memory, and was running on a

<table>
<thead>
<tr>
<th>Channel Number</th>
<th>C = 1</th>
<th>C = 2</th>
<th>C = 3</th>
<th>C = 4</th>
<th>C = 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Muscle</td>
<td>MG</td>
<td>TA</td>
<td>BF</td>
<td>VL</td>
<td>SL</td>
</tr>
<tr>
<td>Number of Contractions per Muscle</td>
<td>152</td>
<td>124</td>
<td>95</td>
<td>6</td>
<td>243</td>
</tr>
<tr>
<td>Percentage of Contractions in which each Muscle has EMG bursts</td>
<td>46.48%</td>
<td>37.92%</td>
<td>29.05%</td>
<td>1.83%</td>
<td>74.31%</td>
</tr>
<tr>
<td>Number of times a muscle responded first in a Contraction</td>
<td>148</td>
<td>148</td>
<td>61</td>
<td>61</td>
<td>148</td>
</tr>
<tr>
<td>Percentage of Contractions a Muscle responded first</td>
<td>7.89%</td>
<td>85.48%</td>
<td>64.21%</td>
<td>0%</td>
<td>60.91%</td>
</tr>
</tbody>
</table>

IV. DISCUSSION

This work focused on developing techniques for concurrent processing of EMG recordings from multiple muscles active during myoclonus spasms, which often disrupt sleep in people with SCI. If both counts match closely, and individuals indicate which contractions are problematic, this is a strategy to identify those contractions that need management clinically.

REFERENCES


