Time-domain features for sEMG signal classification: A brief survey

Sidi Mohamed Sid’El Moctar*, Imad Rida*, Sofiane Boudaoud*

*Alliance Sorbonne Université, Université de technologie de Compiègne, Laboratoire de Biomécanique et de Bioingénierie UMR CNRS 7338, Compiègne, France

Correspondance: sidimohamed.sidelmoctar@utc.fr

Abstract

Surface electromyography (sEMG) signals have been widely used in various robotic and medical applications. Thus, Time domain features have been shown to be effective in extracting useful information from these signals for classification purposes. They are mainly based on amplitude, energy, and time statistics. Furthermore, they can be easily implemented in real-time systems and combined to conventional supervised learning algorithms. The proposed study depicts and presents an up-to-date literature review highlighting the significant time-domain features and their applications in the field.

Keywords— sEMG, signal processing, feature extraction, classification

1 Introduction

Electromyography (EMG) is a technique used to measure the electrical activity of muscles during contraction, which is controlled by the nervous system. There are two types of EMG signals: surface EMG (sEMG), which is recorded non-invasively using electrodes placed on the skin, and intramuscular EMG, which requires insertion of a needle electrode directly into the muscle [1]. High Density surface electromyography (HD-sEMG) is a recent technique, combining ambulatory device, that involves recording sEMG signals from a high number of electrodes [2], this technique provides a more detailed and spatially distributed representation of the muscle activity for medical applications using connected devices [3]. The sEMG signals provide valuable information for understanding the musculoskeletal system in normal and pathological conditions. They are used in both medical and engineering fields to generate device control commands for rehabilitation equipment and human-computer interfaces (HCI) [1]. Despite its potential, the analysis and classification of EMG signals can be challenging due to the complex pattern of the signal and the presence of various types of noise during recording. To improve the accuracy of recognizing sEMG signals, researchers have focused on efficient feature extraction task. Feature extraction refers to the process of selecting the most relevant information from the signal to analyze its pattern for recognition purposes. Popular feature extraction methods include time domain features, frequency domain features, and time-frequency features [4]. However, some of these methods only provide limited neural/motor control information, and few features can fully reflect the detailed characteristics of sEMG signals. Pattern recognition based approach is often used and has shown promise as a reliable method for classifying user signals. For example, pattern recognition algorithms can be used to analyze a collection of features that describe the sEMG signals to identify the user’s intended movement for controlling an external device for example [5]. For illustration purpose, Figure 1 shows the process of sEMG signal based pattern recognition. The use of time domain features for the classification of EMG signals dates back to the 2000’s, as in the study of Bekka et al. that combined time domain features with a neural network for the identification of motor unit features [6]. Since then, numerous studies have demonstrated that both offline and online testing of sEMG signals can result in classification accuracies of over 90% [7][8][9][10][11][12]. The purpose of this short survey is to explore the effect of the commonly used sEMG features on the sEMG signal classification and determine which feature sets are the most reliable for sEMG pattern recognition. Our review focused solely on time-domain features, which are straightforward and computationally less demanding and do not require additional signal transformation. The following sections cover sEMG signal preprocessing methods, feature extraction, and results of some sEMG signal classification studies.
2 Signal preprocessing

Despite its numerous advantages, sEMG signals are frequently affected by various forms of noise, including white Gaussian noise, power line interference (PLI), and other electrophysiological signals such as ECG artifact [13]. These noises can affect the signal quality and classification, so it is very important to go through signal processing to eliminate these noises and reduce their effect. In a recent tutorial, techniques and recommendations for sEMG signal detection, conditioning and pre-processing have been presented [14]. For denoising surface electromyography (sEMG) signals, numerous methods have been proposed in the literature, including baseline noise spectrum subtraction and power line noise subtraction methods [15]. However, these techniques often make unrealistic assumptions regarding the line noise and fail to distinguish between different sources of noise. Conventional digital filters have also been proposed, but they tend to corrupt the useful signal due to spectral overlapping [16]. Empirical Mode Decomposition (EMD) [17] followed by Ensemble EMD (EEMD) [18] has emerged as an effective denoising procedure that overcomes the limitations of mode mixing associated with standard EMD. Additionally, wavelet and adaptive wavelet thresholding methods have been utilized [19], which offer precise targeting of white Gaussian noise components through time-frequency representation. The HD-sEMG signals can be denoised using blind source separation (BSS) techniques, such as independent component analysis (ICA) [20]. Canonical correlation analysis (CCA) has been used for denoising biomedical multichannel signals [20] [21]. The primary benefit of CCA is that it ranks the estimated sources based on a correlation coefficient and enables the construction of a relevant thresholding paradigm.

3 Feature extraction

sEMG pattern recognition systems typically utilize a collection of standard functions to extract features from the time domain. These features are well-suited for real-time systems that must meet specific constraints, and can be easily implemented using basic hardware. Conventional time domain features are typically derived from statistical analyses of either the sEMG signal or its first derivative. The time domain features are extracted from the amplitude of the signal, which changes over time. The amplitude of the signal is influenced by the type of muscle and the observation conditions. Time domain features have several advantages over other types of features. For example, they do not require high computational complexity, can be used in real-time applications, are easy to implement, and perform well in low noise environments. Additionally, these features do not require any additional signal transformation [22]. Below, we will introduce the various time domain features. These handcrafted features are typically integrated with conventional supervised learning algorithms, such as random forest (RF), support vector machines (SVM), Decision Tree (DT), Linear Discriminant Analysis (LDA), and k-nearest neighbors (kNN). The combination between time-domain features and these classifiers gave good results in many different type of applications. Table 1 resume the results of different studies.

Prior to exploring the specifics of the time domain features, we will be using the following notations: $\mathbf{x} \in \mathbb{R}^N$ an sEMG signal, $x_i$ its $i^{th}$ entry and $|.|$ the absolute value. $a_i$ is the linear predictive coefficients. $\hat{x}$ is the predicted sEMG signal value.

- **Mean Absolute Value (MAV)**: It represents the average of the absolute value of sEMG signal amplitude. It is calculated by taking the average value of the absolute entries of the sEMG signal. MAV features are widely used to detect and measure the muscle contraction levels in hand/finger recognition applications. It is defined by:

$$MAV = \frac{1}{N} \sum_{i=1}^{N} |x_i| \quad (1)$$

- **Waveform Length (WL)**: It is intuitively the cumulative length of the waveform over the sample. It indicates a measure of the waveform related to time and amplitude providing a measure of the signal complexity. It is given by:

$$WL = \sum_{i=1}^{N-1} |x_{i+1} - x_i| \quad (2)$$
where $x_i$ and $x_{i+1}$ are two consecutive samples entries of the sEMG signal $x$.

- **Modified Waveform Length (MWL)**: It is a modified and improved version of the aforementioned Waveform Length (WL) features. Indeed, it is calculated based on the first derivative of the sEMG signal and formulated as follows:

$$MWL = \sum_{i=1}^{N} |x'_{i+1} - x'_{i}|$$ (3)

where $x'$ is the first derivative of the signal.

- **Slope Sign Changes (SSC)**: It counts the number of times the sign in the slope of the signal changes. It should be noted that the threshold function seeks to minimize the interference. It is given by:

$$SSC = \sum_{i=2}^{N-1} f[(x_i - x_{i-1}) \cdot (x_i - x_{i+1})]$$ (4)

where $f(x) = \begin{cases} 1 & \text{if } x \geq \epsilon \\ 0 & \text{otherwise} \end{cases}$

- **Zero Crossing (ZC)**: It counts the number of times the amplitude value of the sEMG signal crosses zero. The threshold condition seeks to avoid the background noise. It is mathematically calculated as follows:

$$ZC = \begin{cases} \sum_{i=1}^{N-1} \text{sgn}(x_i \cdot x_{i-1}) & (x_i - x_{i-1}) \geq \epsilon \\
(x_i - x_{i-1}) \geq \epsilon & \text{otherwise} \end{cases}$$ (5)

where $\text{sgn}(.)$ is the sign function

$$\text{sgn}(x) = \begin{cases} 1 & \text{if } x < \epsilon \\ 0 & \text{otherwise} \end{cases}$$

- **Integrated EMG (IEMG)**: also known as the Integral of Absolute Value (IAV), it represents the summation of the absolute values of the sEMG signal amplitude given by:

$$IEMG = \sum_{i=1}^{N} |x_i|$$ (6)

- **Simple Square Integral (SSI)**: It uses the energy of the sEMG signal as a feature. It formulated as follows:

$$SSI = \sum_{i=1}^{N} |x_i|^2$$ (7)

- **Mean**: It represents the mean amplitude value of an sEMG signal. It is expressed by:

$$\text{Mean} = \frac{1}{N} \sum_{i=1}^{N} x_i$$ (8)

- **Variance (VAR)**: It stands to the variance value of the signal given by:

$$\text{VAR} = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \mu)^2$$ (9)

where $\mu$ is the mean value.

- **Standard Deviation (SD)**: It is the square root of the variance and calculated as follows:

$$SD = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (x_i - \mu)^2}$$ (10)

- **Root Mean Square (RMS)**: As mentioned by its name, it stands to the square root of the mean square of the signal amplitude. It is given by:

$$RMS = \sqrt{\frac{1}{N} \sum_{i=1}^{N} x_i^2}$$ (11)

- **Willison amplitude (WAMP)**: It counts the number of times the changes in the sEMG signal amplitude exceeds a predefined threshold in order to reduce noise effect. It is given by:

$$WAMP = \sum_{i=1}^{N-1} f(|x_i - x_{i+1}|)$$ (12)

where $f(x) = \begin{cases} 1 & \text{if } x \geq \epsilon \\ 0 & \text{otherwise} \end{cases}$

- **Average Amplitude Change (AAC)**: It corresponds to the averaged WL, it is given by:

$$AAC = \frac{1}{N} \sum_{i=1}^{N-1} |x_{i+1} - x_i|$$ (13)

- **Skewness**: It is a measure of the symmetry of a dataset. It corresponds to the ratio of the average deviation
from the mean cubed divided by the standard deviation cubed. It is defined as:

$$\text{Skew} = \frac{1}{N\sigma^3} \sum_{i=1}^{N} (x_i - \bar{x})^3$$  \hspace{1cm} (14)$$

where $N$, $\mu$ and $\sigma$ correspond to the signal length, the mean and the standard deviation respectively.

- **Kurtosis**: It is a measure of the signal steepness or flatness relative to a normal distribution. It is a robust metric for detecting impulsive content within the data and it is defined as:

$$\text{Kurtosis} = \frac{1}{N\sigma^2} \sum_{i=1}^{N} (x_i - \bar{x})^4$$  \hspace{1cm} (15)$$

- **Difference absolute standard deviation value (DASDV)**: It corresponds to the standard deviation value of the difference between the adjacent samples and calculated by

$$\text{DASDV} = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N-1} (x_{i+1} - x_i)}$$  \hspace{1cm} (16)$$

- **Modified Mean Absolute Value 1 (MMAV1)**: It is an extension of the Mean Absolute Value using a weighting window function $w_i$ and defined as

$$\text{MMAV1} = \frac{1}{N} \sum_{i=1}^{N} w_i |x_i|$$  \hspace{1cm} (17)$$

where

$$w_i = \begin{cases} 
1 & \text{if } 0.25N \leq i \leq 0.75N \\
0.5 & \text{otherwise}
\end{cases}$$

- **Modified Mean Absolute Value 2 (MMAV2)**: It is also an extension of the Mean Absolute Value with an improved continuous weighting window function. $w_i$. It is expressed as follows:

$$\text{MMAV2} = \frac{1}{N} \sum_{i=1}^{N} w_i |x_i|$$  \hspace{1cm} (18)$$

where

$$w_i = \begin{cases} 
1 & \text{if } 0.25N \leq i \leq 0.75N \\
\frac{4i}{N} & \text{if } 0.75N < i \\
\frac{4(i - N)}{i} & \text{if } 0.25N > i
\end{cases}$$

There are other time domain features that have been used for classification of sEMG signals. Among them the integrated absolute value [23], EMG histogram (Hist) [24], difference between moments [25], Co-Contraction Index [26], Myopulse percentage rate (MYOP) [27], Sample Entropy [28], log-Detector (LD), v-Order [29], Cepstral Coefficients [30] and Guided Under-determined Source Signal Separation [31].

4 **Discussion and conclusion**

From Table 1, the exposed features with conventional supervised learning algorithms has been shown to achieve high classification accuracy in various applications, including neuromuscular system monitoring, prosthetic control, rehabilitation, and human-computer interaction. We see that the MAV, RMS, WL, ZC and SSC features are often used for applications related to motion classification. This is mainly due to their low computational complexity, which makes them well-suited for real-time applications. In [32], [33] we see that time domain features have a better performance than frequency domain features for long term and provides a more consistent method for electrode selection. The selection of appropriate features requires a precise application knowledge and further research is needed to improve sEMG signal classification accuracy by proposing other possible low complexity features with possible interpretability to encourage medical applications. For robotic applications, that are dominant in this survey, this relationship knowledge is not needed. Applications using HD-sEMG signals allow the access to spatial information enhancing classification task. Few studies are related to musculoskeletal system evaluation for medical applications.
Table 1: Overview of handcrafted features based classification of sEMG signal.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Year</th>
<th>Application</th>
<th>Features</th>
<th>Classifier</th>
<th>Accuracy</th>
<th>Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Venugopal et al [34]</td>
<td>2014</td>
<td>Muscle fatigue</td>
<td>MAV</td>
<td>KNN, SVM, NB, LDA</td>
<td>93%</td>
<td>Private</td>
</tr>
<tr>
<td>Soman et al [35]</td>
<td>2016</td>
<td>Wrist/finger movement</td>
<td>RMS</td>
<td>SVM</td>
<td>95%</td>
<td>Private</td>
</tr>
<tr>
<td>Jose et al [36]</td>
<td>2017</td>
<td>Forearm movement</td>
<td>IEMG, ZC, SSC, WL, WAMP</td>
<td>RF, NN</td>
<td>97.7%</td>
<td>Private</td>
</tr>
<tr>
<td>Bhattacharya et al [37]</td>
<td>2017</td>
<td>Hand movement</td>
<td>AR, MAV, RCS, ZC, SSC, WL</td>
<td>KNN, LDA, QDA</td>
<td>83.33%</td>
<td>Private</td>
</tr>
<tr>
<td>Wei et al [38]</td>
<td>2017</td>
<td>Hand movement</td>
<td>RMS</td>
<td>RF, SVM, KNN, NN</td>
<td>98.06%</td>
<td>NinaPro</td>
</tr>
<tr>
<td>Li et al [39]</td>
<td>2017</td>
<td>Hand/wrist movement</td>
<td>MAV, WL, ZC, SSC</td>
<td>LDA</td>
<td>91.7%</td>
<td>Private</td>
</tr>
<tr>
<td>Wahid et al [40]</td>
<td>2018</td>
<td>Hand movement</td>
<td>MAV, ZC, WL, SSC</td>
<td>KNN, DA, RF, SVM</td>
<td>96.38%</td>
<td>Private</td>
</tr>
<tr>
<td>Rescio et al [26]</td>
<td>2018</td>
<td>Fall risk detection</td>
<td>IEMG, CCI, MAV, RMS, VAR, WL, ZC, SSI, SSC, WAMP</td>
<td>LDA</td>
<td>90%</td>
<td>Private</td>
</tr>
<tr>
<td>Ga et al [41]</td>
<td>2018</td>
<td>Vocal fatigue</td>
<td>MAV, ZC, SSC, WL, WAMP, RMS, AR</td>
<td>LDA</td>
<td>93.9%</td>
<td>Private</td>
</tr>
<tr>
<td>Morbidoni et al [42]</td>
<td>2019</td>
<td>Gait recognition</td>
<td>SD, RMS, MAV, IEMG, WL</td>
<td>NN</td>
<td>94.9%</td>
<td>Private</td>
</tr>
<tr>
<td>Freitas et al [8]</td>
<td>2019</td>
<td>Hand movement</td>
<td>VAR, RMS, MAV, IEMG</td>
<td>NN, KNN, LDA, QDA, DT, NB</td>
<td>99%</td>
<td>Private</td>
</tr>
<tr>
<td>Leone et al [10]</td>
<td>2019</td>
<td>Hand/wrist Movement</td>
<td>MAV, RMS, SSC, WL, VAR</td>
<td>LDA, NLR</td>
<td>98.7%</td>
<td>Private</td>
</tr>
<tr>
<td>Reference</td>
<td>Year</td>
<td>Application</td>
<td>Features</td>
<td>Classifier</td>
<td>Accuracy</td>
<td>Dataset</td>
</tr>
<tr>
<td>--------------------</td>
<td>------</td>
<td>-----------------</td>
<td>--------------------------------------------------------------------------</td>
<td>------------------</td>
<td>----------</td>
<td>-------------</td>
</tr>
<tr>
<td>Too et al [9]</td>
<td>2019</td>
<td>Hand Movement</td>
<td>MAV, WL, ZC, SSC, AAC, LD, RMS, DASDV, MYOP, SSI, WA, VAR, MMAV1, MMAV2</td>
<td>KNN, LDA, NB, SVM</td>
<td>97.56%</td>
<td>NinaPro</td>
</tr>
<tr>
<td>Kaur et al [11]</td>
<td>2019</td>
<td>Shoulder Movement</td>
<td>MAV, Median, Mean, MAV, RMS, VAR</td>
<td>SVM, DT, RF, NB</td>
<td>98.8%</td>
<td>Private</td>
</tr>
<tr>
<td>Cai et al [12]</td>
<td>2019</td>
<td>Upper-limb movement</td>
<td>RMS, WL, VAR, MAV, SSI</td>
<td>SVM</td>
<td>98.89%</td>
<td>Private</td>
</tr>
<tr>
<td>Marcovic et al [43]</td>
<td>2019</td>
<td>Wrist Movement</td>
<td>RMS, ZC, SSC, WL</td>
<td>LDA</td>
<td>90%</td>
<td>Private</td>
</tr>
<tr>
<td>Fazeli et al [44]</td>
<td>2020</td>
<td>Hand Movement</td>
<td>MAV, WL, ZC, SSC</td>
<td>SVM, LDA, QDA, DT, KNN</td>
<td>87.3%</td>
<td>Private</td>
</tr>
<tr>
<td>Nsugbe et al [45]</td>
<td>2020</td>
<td>Hand Movement</td>
<td>MAV, RMS, ZC, SSC, WAMP, AR, SampEN, Cepstrum, VAR</td>
<td>NN</td>
<td>83%</td>
<td>Private</td>
</tr>
<tr>
<td>Devaraj and Nair [46]</td>
<td>2020</td>
<td>Hand Movement</td>
<td>IEMG, MAV, MMAV, MAV, VAR, RMS, WL</td>
<td>KNN, SVM</td>
<td>93.92%</td>
<td>NinaPro</td>
</tr>
<tr>
<td>Vijayvargiya et al [27]</td>
<td>2020</td>
<td>Limb Movement</td>
<td>MAV, RMS, ZC, SSC, VAR, WAMP, MYOP, DASDV, AAC, Skewness, Kurtosis</td>
<td>KNN, SVM, DT, RF, Extra Tree (ET)</td>
<td>91.3%</td>
<td>UCI</td>
</tr>
<tr>
<td>Sattar et al [47]</td>
<td>2020</td>
<td>Arm Movement</td>
<td>Mean, VAR, Skewness, Kurtosis, SSC, MAV, RMS, WL, WAMP, ZC</td>
<td>LDA, KNN, SVM, QDA</td>
<td>95.8%</td>
<td>Private</td>
</tr>
<tr>
<td>Ahlawat et al [48]</td>
<td>2021</td>
<td>Hand Movement</td>
<td>RMS, MMAV1, MMAV2</td>
<td>SVM, NN</td>
<td>99.3%</td>
<td>Private</td>
</tr>
<tr>
<td>Vijayvargiya et al [49]</td>
<td>2021</td>
<td>Lower limb Movement</td>
<td>MAV, RMS, ZC, SSC, VAR, WAMP, MYOP, AAC, DASDV, Skewness, Kurtosis</td>
<td>Regression Tree, GB, RF, Extra Tree</td>
<td>91.9%</td>
<td>UCI</td>
</tr>
<tr>
<td>Khairuddin et al [50]</td>
<td>2021</td>
<td>Upper limb Movement</td>
<td>WL, MAV, RMS, SD, MIN, MAX</td>
<td>LDA, LR, DT, SVM, KNN</td>
<td>99%</td>
<td>Private</td>
</tr>
</tbody>
</table>
However, despite the success of time domain features in sEMG signal classification, several challenges still exist. One of the main challenges is the variability of SEMG signals between individuals, which can affect the accuracy of the classification. Furthermore, noise and artifacts can impact the quality of the extracted features and the performance of the classifier. Thus, denoising task must be done with caution to ensure noise suppression without sEMG signal corruption. Deep learning-based classification, on the other hand, can automatically learn features from the raw signal and does not require manual feature engineering and can handle complex signal patterns. It has shown superior performance compared to handcrafted feature-based classification, especially in cases where the signal is noisy, or there are multiple classes to be distinguished. However, deep learning-based classification requires a large amount of data and computing resources to train the neural network models, this may explain the use up to these days of handcrafted features.

Conflict of interest: The authors declare no conflict of interest.

References


