

Article

Discussion of the Influence of Multiscale PCA Denoising Methods with Three Different Features

Chizhou Zhang¹ and Tao Sun^{2,*}

¹ Electronic & Computer Science School, University of Shouthampton, Southampton SO17 1BJ, UK; cabbagekue@163.com

² Electronic Information School, Wuhan University, Wuhan 430072, China

* Correspondence: suntao@whu.edu.cn

Abstract: Bioinformation is information generated from biological movement. By using a variety of modern technologies, we can use this information to form a meaningful model for researchers to study. An electromyographic (EMG) signal is one type of bioinformation that is used in many areas to help people study human muscle movement. This information can help in both clinical areas and industrial areas. EMG is a very complicated signal, so processing it is vital. The processing of EMG signals is divided into collection, denoising, decomposition, feature extraction and classification steps. In this article, the wavelet denoising step and several decomposition processes are discussed to show the usage of this technique in the final classification step. At the end of the study, we find that after the wavelet denoising step, the classification accuracy, which uses the K-nearest neighbor of the independent component analysis features, improves, but the accuracy of the wavelet coefficient features and autoregression coefficient features decreases.

Keywords: EMG signal; ICA; wavelet function; signal decomposition; KNN



Citation: Zhang, C.; Sun, T.

Discussion of the Influence of Multiscale PCA Denoising Methods with Three Different Features. *Sensors* **2022**, *22*, 1604. <https://doi.org/10.3390/s22041604>

Academic Editor: Marco Di Rienzo

Received: 17 January 2022

Accepted: 14 February 2022

Published: 18 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

An electromyographic (EMG) signal is a time-series signal that is collected mainly for clinical applications, and it has been widely used in research laboratories to meet the needs of sectors such as biomechanics and motor control [1]. After 1866, when Grace and Erlanger used an oscilloscope to show electrical signals from muscles, Cram and Steger utilized it in a clinical approach to scan a variety of muscles, and after the 1880s, when the appropriate amplifier was invented, people began to pay much more attention to EMG signals [2]. EMG signals have been investigated and widely debated over many decades. Mary described electromyographic signals as muscular activity involving muscle neurons and regulating muscle fiber inside the motor unit [3]. Based on this description and other research, the collection methods of EMG signals are divided into two types. One type is called the transcutaneous method and collects motor unit potentials (MUPs) by placing an array of electrodes on the skin. The other type, which is called the intramuscular method, requires a wire or needle electrode to be directly placed into the muscle to collect the MUPs directly [4]. The different collection methods of EMG signals mean that the processing methods are diverse. A raw EMG signal can be regarded as a time-series signal, and the features in its time and frequency domain can be obtained to perform classification. Additionally, as the number of collecting channels of the EMG signal increases, the multichannel EMG signal can be processed as a multidimensional signal process; for example, ref. [5] shows that the multichannel EMG signal is suitable for the blind source separation theory. Multidimensional signals are an important form of information. Compared with one-dimensional signals, multidimensional signals contain more energy and more information, and their accuracy can be improved through the connections between all dimensions. A common example of a two-dimensional signal is an image signal, and an example of a three-dimensional signal is a video signal [6].

A simple way to deal with a multichannel EMG signal is to deal with the average of all the signals in each channel or to use each channel's signal to perform feature extraction; however, this loses the crosstalk between different EMG signals near each other, which causes information loss. Therefore, in this article, we illustrate some methods that are used to process multichannel EMG signals, and we compare them to normal processing methods to show the differences between the features that the different methods yield.

2. Materials and Methods

2.1. EMG Signal Components

To understand how to process EMG signals, the construction of an EMG signal needs to be discussed. The fundamental component of an EMG signal is the muscle fiber action potential (MFAP), which is the action potential caused by the excitable membrane of a muscle fiber [7]. All MFAPs in a motor unit together make up the motor unit action potential (MUAP), and many motor units fire repeatedly to maintain muscle action, which is summed to produce an EMG signal [7].

Figure 1 shows the basic structure of an EMG signal and how it can be decomposed, and Figure 2 shows a raw EMG signal.

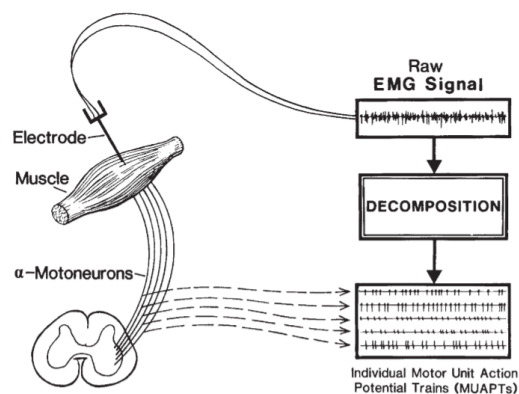


Figure 1. EMG signal composition and how it can be decomposed [7].

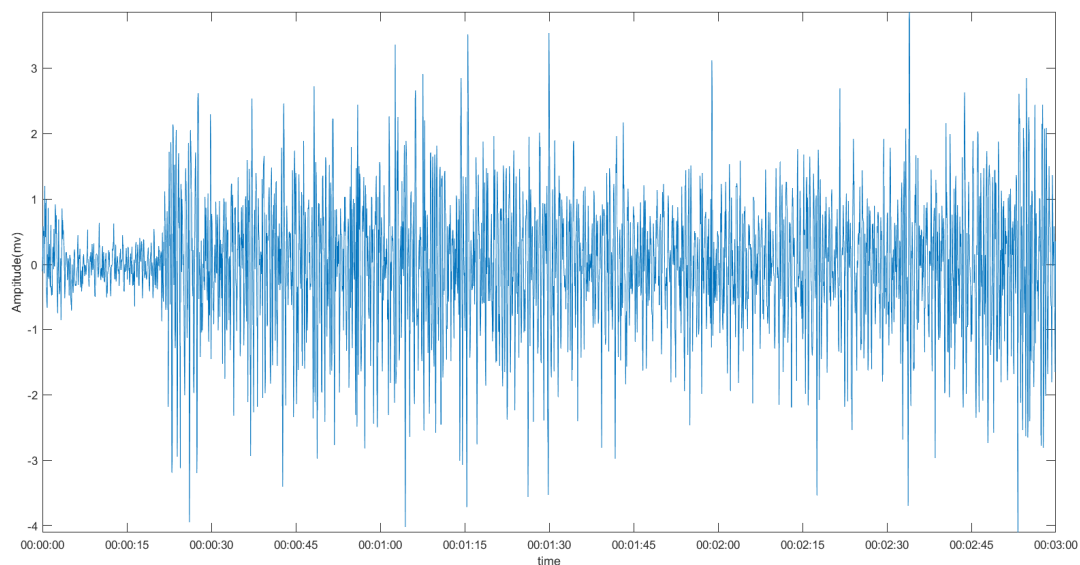


Figure 2. Example of a raw EMG signal.

The equation below shows a basic EMG signal model, where $MUAPT_j$ is the motor unit action potential train, which refers to the collection of one motor unit positioned at the time of occurrence or separated by inter-discharge intervals, and $n(t)$ denotes noise [7].

$$EMG(t) = \sum_{j=1}^{N_m} MUAPT_j(t) + n(t) \quad (1)$$

2.2. Noise and Denoising

During the detection and collection steps, there are many factors that can influence the collected signal. The denoising step is introduced in preprocessing to improve the quality of the signal. In this section, the main noise within a raw collected EMG signal is illustrated, and a denoising method is used.

2.2.1. Noising

Initially, there is noise due to the instrument used to collect the EMG signal. First, for the two main kinds of EMG signal types, that is, the intramuscular EMG signal from the needle inserted into the muscle and the surface EMG signal from the surface electrodes, the sensitivity of the collection instruments, such as the needles and the electrodes, decides the noise that is collected with the EMG signal, which ranges from zero Hz to several thousand Hz; second, for the connection cable between the muscle, the electrodes and processing circuit, a slight movement will affect the electrical forces while the signal passes through [8]. By using better equipment and an intelligent circuit design, this noise can be efficiently reduced [8].

Second, there is noise generated by the human body. One type of noise is the electromagnetic noise that is generated by the body. The human body is similar to a small magnet that constantly generates electromagnetic radiation that propagates to the surroundings. This kind of noise is also known as ambient noise. The second type of noise comes from several factors, such as anatomical noise and biochemical noise, which take place in the biological unit of the detecting muscle; another type of this kind of noise is the “crosstalk” of unwanted EMG signals, which comes from other muscle tissues. Finally, the most interference comes from the heart’s electrical activity; another name for this is “electrocardiogram artefacts” [2,8]. These two types of noise can be reduced by a high-pass filter, because they are both generated due to random factors.

The last part is due to the instability of the signal, which is caused by the unstable firing rate of the motor unit, and, usually, these noise frequency components are approximately 0–19 Hz. The active motor unit and the muscle fiber mechanical reaction both play a large part in this noise [2].

In [7], the raw EMG signal was detected through an electrode and passed through a bandpass filter and a low-pass filter to reduce the temporal overlap, the number of superimposed waveforms and the amplitude of meaningless motor units. Additionally, nondiscriminative, low-frequency information is removed, which makes it easier to distinguish different motor units’ MUAPs, which are some of the features that can be extracted by reducing the variability of the shape of the signal.

2.2.2. MPCA Denoising

The main denoising method that is used for EMG signals is the multiscale principal component method. Multiscale principal component analysis (MPCA) combines the features and capabilities of PCA and enhances the robustness of PCA in revealing hidden structures [9,10]. The main idea of PCA is to project N data points that have the mean value \mathbf{m} and covariance matrix \mathbf{C} onto the direction \mathbf{u} to determine the direction that can maximize the project variance $\mathbf{u}^t \mathbf{C} \mathbf{u}$. To determine the maximum variance, we can set up an optimization problem that finds $\max_{\mathbf{u}} \mathbf{u}^t \mathbf{C} \mathbf{u}$ under the condition $\mathbf{u}^t \mathbf{u} = 1$ [9]. Hence, we can use the Lagrangian multiplier method, which is

$$\mathcal{L} = \mathbf{u}^t \mathbf{C} \mathbf{u} - \lambda(\mathbf{u}^t \mathbf{u} - 1) \quad (2)$$

Then, we differentiate \mathcal{L} , and when $\frac{d\mathcal{L}}{du} = 0$, we can obtain $Cu = \lambda u$.

Another important part of multiscale PCA is the wavelet transform process. The general idea is that noise is a high-frequency signal that is normally distributed with a Gaussian distribution. As in the general model mentioned in the previous section, the EMG signal can also be described as

$$F(t) = s(t) + e(t) \quad (3)$$

where $s(t)$ and $e(t)$ are the EMG signal and white noise, respectively, which means that the mean value of the noise is zero and the covariance is σ^2 . Assuming the wavelet function is $\varphi(t)$, $W_e(\alpha, t)$ is the wavelet transform of $f(t)$; hence, we have

$$W_e(\alpha, t) = \int_{-\infty}^{+\infty} e(u)\varphi_e(t-u)du \quad (4)$$

From this assumption, we can obtain the average power of $W_e(\alpha, t)$,

$$\begin{aligned} E\{|W_e(\alpha, t)|^2\} &= \iint_{-\infty}^{+\infty} E\{e(u)e(v)\}\varphi_a(t-u)\varphi_a(t-v)dudv \\ &= \sigma^2 \iint_{-\infty}^{+\infty} \delta(u-v)\varphi_a(t-u)\varphi_a(t-v)dudv \\ &= \sigma^2 \int_{-\infty}^{+\infty} |\varphi_a(t-u)|^2 du \\ &= \frac{\sigma^2 \|\varphi\|^2}{a} \end{aligned} \quad (5)$$

We can determine that the average power of $W_e(\alpha, t)$ is inversely proportional to the scale a [11].

The main process of MPCA is as follows [12,13]:

1. Use wavelet transformation to transform each channel of the EMG signal to the L of the decomposition level, and collect all the wavelet details in each channel into matrix D_L and the wavelet approximations into matrix A_L .
2. Eigen-decompose the covariance matrix of each wavelet matrix, determine the eigenvector and eigenvalue of each, and then arrange them in descending order. The eigenvalues that select the threshold will determine the number of principal components that form a new matrix.
3. The relevance factor μ_i (when two statistics, namely, T^2 statistic and SPE statistic, exceed the control limits, $\mu_i = 1$ otherwise, 0) will select the nonsignificant scale and form a new EMG signal:

$$EMG'_i = A_L + \sum_i^L \mu_i D_L \quad (6)$$

The two statistics and their control limit, which are used in Equation (6) to decide the value of μ_i , are calculated as follows:

The T^2 statistic reflects the degree to which each principal component deviates from the model in terms of trend and amplitude; it is a measure of the internalization of the model, and it can be used to monitor multiple principal components simultaneously:

$$T_i^2 = X_i P_k \lambda^{-1} P_k X_i \quad (7)$$

$$T_\alpha = \frac{k(m-1)(m+1)}{m(m-k)} F_{k,m-k,\alpha} \quad (8)$$

T_α is the control limit of the T^2 statistic, where X is the data matrix and P is the loading matrix; λ is the eigenvalue-diagnosed matrix, in which the eigenvalue belongs to the first k principal component; m represents the dimension of the data matrix X ; and $F_{k,m-k,\alpha}$ is the F distribution at the significance level of α .

The SPE statistic describes the degree to which the measured value of the input variable deviates from the principal component model, and it is a measure of external variation in the model.

$$SPE_i = X_i(I - P_k P_k^T) X_i^T \quad (9)$$

$$SPE_\alpha = \theta_1 \left[\frac{C_\alpha \sqrt{2\theta_2 h^2}}{\theta_1} + 1 + \frac{\theta_2 h(h-1)}{\theta_1^2} \right]^{\frac{1}{h}} \quad (10)$$

$$\theta_n = \sum_{j=k+1}^m \lambda_j^n \quad n = 1, 2, 3 \quad (11)$$

$$h = 1 - \frac{2\theta_1 \theta_3}{3\theta_2^2} \quad (12)$$

SPE_α is the control limit of the SPE_i statistic, where C_α is the standard normal deviate corresponding to the upper $(1 - \alpha)$ percentile.

2.3. Raw EMG Signal Processing

For a multichannel EMG signal, a necessary step is decomposition. By obscuring the composition of the EMG signal, we can easily determine that the EMG signal is composed of many MUAPs, so the features of classical techniques used for analysis, which are extracted from s-EMG signals, cannot guarantee effectiveness [14]. Therefore, if we can decompose the EMG signal, it will be easier to perform the next feature extraction step and make the features more representative.

2.3.1. Independent Component Decomposition

In recent years, the independent component analysis (ICA) algorithm has appeared in many multi-dimension signal processing processes and mostly plays an important role, which is a special case of the blind source separation technique [8]. The main purpose of ICA is to transform an experimental multivariate random vector into independent sources [8]. The basis of the ICA is to normalize, whiten and iterate [5]. The whitening process uses a linear transformation to transform the components of the observed vector x into uncorrelated components with unit variance. After whitening, the parameters that need to be estimated in the ICA process will be reduced [5]. The iteration process is an algorithm used to calculate the independent components, and different algorithms have been used in this process, such as joint approximate diagonalization of eigenmatrices (JADE), the infomax estimation or maximum likelihood algorithm and the fast ICA algorithm [8]. We first discuss the standard ICA process. The standard process considers that the signal matrix x can be represented as an unknown, invertible, square matrix A with unknown pure signal vector s .

$$x = As \quad (13)$$

We want to find an unmixing matrix w to obtain the following equation:

$$s = wx = wAs \quad (14)$$

where $wA = I$ [15]. The fast ICA algorithm is used to estimate matrix w . The fast ICA algorithm is based on a fixed-point iterative structure algorithm, and the goal is to make wx have the most non-Gaussian nature.

The first step of the fast ICA algorithm is to initialize the vector matrix w and then use $g(u) = \tanh(au)$ or other suitable functions as the derivative of the contraction function. Under the constraint condition $E\{(wx)^2\} = \|w\|^2 = 1$, the optimal condition of $E\{G(wx)\}$ can be obtained by the following formula:

$$E\{xg(wx)\} - \beta w = 0 \quad (15)$$

To use Newton iteration to solve this problem, we can obtain the vector matrix w' , which is

$$w' = E\{xg(wx)\} - E\{g'(wx)\}w \quad (16)$$

and then use

$$w = \frac{w'}{\|w'\|}. \quad (17)$$

To calculate the new matrix, if the old w is in the same direction as the new w , then we can consider it convergent; otherwise, we calculate w' iteratively. The obtained vector matrix is the matrix that we want, and based on this matrix, we can obtain the decomposed signal that we want [16].

2.3.2. Wavelet Transform

The wavelet transform (WT) method is also a very common way to decompose a raw EMG signal. As a type of Fourier transform (FT), the WT is also a tool for time-frequency analysis; the difference between the WT and FT is that the WT possesses characteristics of multiresolution, which solves problems of the FT, such as the lack of localized analysis capabilities and the inability to analyze nonstationary signals [11].

The multiresolution theory means that when a scaling function $\{V_j\}$, $j \in Z$ satisfies several conditions, it can be approximated by finite subspace $\{V_{j+1}\}$, which represents the low-frequency large-scale approximation part, and $\{W_{j+1}\}$, which represents the high-frequency detail part [11]:

$$V_0 = V_1 \oplus W_1 = V_2 \oplus W_2 \oplus W_1 = \dots = V_j \oplus W_j \oplus W_{j-1} \oplus \dots \oplus W_2 \oplus W_1 \quad (18)$$

The conditions that the scaling function need to meet are as follows:

1. $V_{j-1} \subset V_j$, $j \in Z$;
2. $\cup_{j \in Z} V_j = L^2$, $\cap_{j \in Z} \{0\}$;
3. $f(t) \in V_j \Leftrightarrow f(2t) \in V_{j+1}$;
4. $\forall k \in Z$, $\phi(2^{-\frac{j}{2}}t) \in V_j \Leftrightarrow \phi(2^{-\frac{j}{2}}t - k) \in V_j$;
5. $\exists \phi(t) \in V_0$, which make $\{\phi(2^{-\frac{j}{2}}t - k) - k | k \in Z\}$ become orthogonal.

The basic FT is expressed as

$$F(t) = \int_{-\infty}^{+\infty} f(t)e^{-i\omega t} dt \quad (19)$$

The WT uses a basic wavelet $\varphi(t)$ as the mother wave to replace the sinusoidal signal base, which is expressed as

$$W(\alpha, t) = \frac{1}{\sqrt{\alpha}} \int_{-\infty}^{+\infty} f(u)\varphi_{\alpha,t}\left(\frac{u-t}{\alpha}\right) du \quad (20)$$

in which α and t are the scale and translation factors, respectively [11]. In practice, the discrete WT (DWT) is often used in the decomposition steps of the signal to discretize a continuous wavelet [1].

2.4. Feature Extraction Method

Feature extraction is a very useful way to transform a raw EMG signal into a set of features, and it is used by researchers to extract useful information in a signal and eliminate unnecessary information [17]. It is also very helpful for later classification. There are three main kinds of features: time-domain (TD) features, frequency-domain (FD) features and time-frequency-domain (TDF) features [17].

2.4.1. Autoregression Coefficient

The autoregression (AR) coefficient is a time-domain feature that involves a linear combination of the previous signal $x(n-1)$ to $x(n-p)$. In this model, the EMG signal is described as a combination of the previous signal and noise:

$$x(n) = \sum_{k=1}^p a_k x(n-k) + w_n \quad (21)$$

where a_k is the autoregression coefficient and w_n is white noise [18,19].

There are many mathematical methods to derive the autoregressive coefficients that define the autoregressive model, the most famous of which is the Yule–Walker model. This model uses the estimated values of the correlation function, which can be calculated as [20]

$$r_{xx}(n, n-k) = E[x(n)x(n-k)] \quad (22)$$

Based on the above equation, the Yule–Walker matrix can be obtained:

$$\begin{bmatrix} r_{xx}(0) & r_{xx}(-1) & \cdots & r_{xx}(-p+1) \\ r_{xx}(1) & r_{xx}(0) & \cdots & r_{xx}(-p+2) \\ \vdots & \vdots & \cdots & \vdots \\ r_{xx}(p-1) & r_{xx}(p-2) & \cdots & r_{xx}(0) \end{bmatrix} \begin{bmatrix} 1 \\ a_1 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} \sigma_w^2 \\ r_{xx}(1) \\ \vdots \\ r_{xx}(p) \end{bmatrix} \quad (23)$$

where the value σ_w^2 is the variance of the model stochastic process. Hence, from the matrix, we can obtain

$$r_{xx}(p) = r_{xx}(p-1) + a_1 r_{xx}(p-2) + \cdots + a_p r_{xx}(0) \quad (24)$$

Since this matrix describes an equation system and fulfils the Toeplitz definition, the recursive Levinson–Durbin algorithm can be used to obtain the autoregressive coefficients a_p .

In the Toeplitz definition, the autoregressive coefficients a_p can be obtained using the recursive Levinson–Durbin algorithm [20]. By observing the above matrix, we can see that this matrix fulfils the definition. Hence, we can use the m th-order Levinson–Durbin algorithm as follows to calculate the coefficients [21]:

$$k_m = -\frac{[r_x(m) + \sum_{k=1}^m a_k r_x(m-k)]}{\rho_{m-1}} \quad (25)$$

$$a_m(k) = a_{m-1}(k) + k_m(a_{m-1}) * (m-k), \quad 1 \leq k \leq m-1, 1 \leq m \leq p \quad (26)$$

$$\rho_m = \rho_{m-1} (1 - k_m^2) \quad (27)$$

2.4.2. The Decomposition Coefficient Features

In addition to the time-domain feature, the frequency-domain feature is an available component that can represent the information of the signal. Two types of features can be extracted based on the decomposition process: the first is the wavelet coefficient, and the other is the ICA coefficient. The wavelet coefficient is the feature vector of each decomposition scale, and the ICA coefficient is the weight of the decomposition signals.

2.5. Classification Method

2.5.1. The K-Nearest Neighbor

There are many ways to classify EMG features, such as Bayesian techniques [22], the neural network method, the K-nearest neighbor (KNN) method [19] and the decision tree [23]. These methods all perform well in the classification process. Comparing several papers that use the matching learning tools implemented in calculating EMG features, KNN seems to be the simplest way to classify EMG features, and it can achieve relatively high accuracy. In reference [24], the author compared the artificial neural network (ANN) and KNN, and when the size of the feature set was the same, the operation time and accuracy

of KNN were much better. Additionally, in reference [25], KNN and other classification methods were compared to classify the features in the time domain. Hence, in this paper, the KNN method is applied.

The KNN algorithm is a suboptimal process and is only useful when the dataset is relatively small. This algorithm assigns labels based on the k-nearest values of the data for classification. This distance is usually determined by Euclidean distance, and then iteratively updates by comparing new records with training records [23,26]. However, there are two issues to consider when using this algorithm. The first issue is that the algorithm assigns equal weights to each sample, which makes the labels difficult to determine when the sample set overlaps with the class labels. The second issue is that the strength of each classification sample cannot be determined [26].

2.5.2. The Naïve Bayesian

The Bayesian aggregation has been used in many fields, such as medical and clinical applications. Generally, the probability of event A under the condition of event B is different from the probability of event B under the condition of event A; however, there is a definite relationship between the two, and Bayes' rule is a statement of this relationship. The main function of the Bayesian aggregation is to use a priori numbers of occurrence and then combine it with specific sampled data values. As such, the Bayesian aggregation algorithm can produce an overall characterization [22].

3. Experiment

The experiment was conducted to determine how the denoising step influences the classification accuracy of the three features. The most important aspects of the experiment were to determine how denoising influences the signal in each processing step, what can be seen by illustrating every step and how each denoising process influences the feature extraction process.

3.1. The Dataset Used

This research used an EMG signal database downloaded from the UCL machine learning website. This dataset was collected using LabVIEW with national directives as the core, and it has been widely applied in many research papers that study EMG signals. This dataset has two different data groups with same six grasping movements see as Figure 3, and the EMG signals of each data group were collected differently. The sampling rate of the EMG signal was 500 Hz, and a Butterworth bandpass filter with a range of 15–500 Hz and a notch filter with a bandstop of 50 Hz were applied to eliminate the artefacts of signals caused by interference, such as those from wires.

The EMG signal was taken from two forearm surface EMG sensors, meaning that the signal is a two-channel EMG system. There are two kinds of datasets in the UCL machine learning repository: the one used in this experiment was collected from five people, three of whom are female and the other two of whom are male. This dataset was formed by collecting the data of the people performing six full grasping movements thirty times each, and the signal collected from each person had its own mat file. These six grasping movements are shown below.

3.2. Comparison of The Denoising Signal and Original Signal

The figure shows the MPCA process that was performed on the two different channels, which calculates the average of all 30 signals in one channel. The MPCA algorithm is based on MATLAB and can be accessed worldwide. The basic wavelet used in this process is the Daubechies least asymmetric wavelet, and a heuristic rule is used to determine how many principal components are retained. The heuristic rule ensures that the component associated with the eigenvalues is greater than 0.05 times the sum of all eigenvalues.

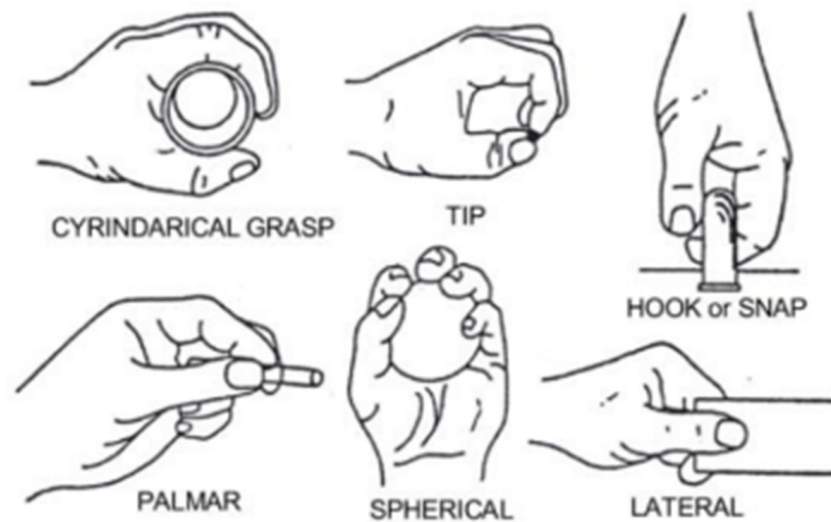


Figure 3. The grasping movements [27].

Because the signal gathers in two different channels, we want to see if there are any differences in denoising for these two channels' signals. From Figure 4, we can see that, despite the slight differences, these two channels' signals do not show many differences; they all seem to process well. By comparing the right column and left column, which means the signal with or without denoising, we can see that there are two improvements in the MPCA denoising method. The first improvement is that the density of the signal is decreased, which is because when using denoising steps, less relative information is removed. The second improvement is that the MUAPTs, which formed as a result of the hand movement, are more easily to distinguish by the amplitude change.

The correlation between the two channels can be calculated with the following equation:

$$R_{xy}(m) = E\{x_{n+m}y_n^*\} = E\{x_n y_{n-m}^*\} \quad (28)$$

where x and y are the signals, and the correlation vector of the raw signal and the denoising signal is shown in Figure 5. As an observation, the correlation between the two channels improved.

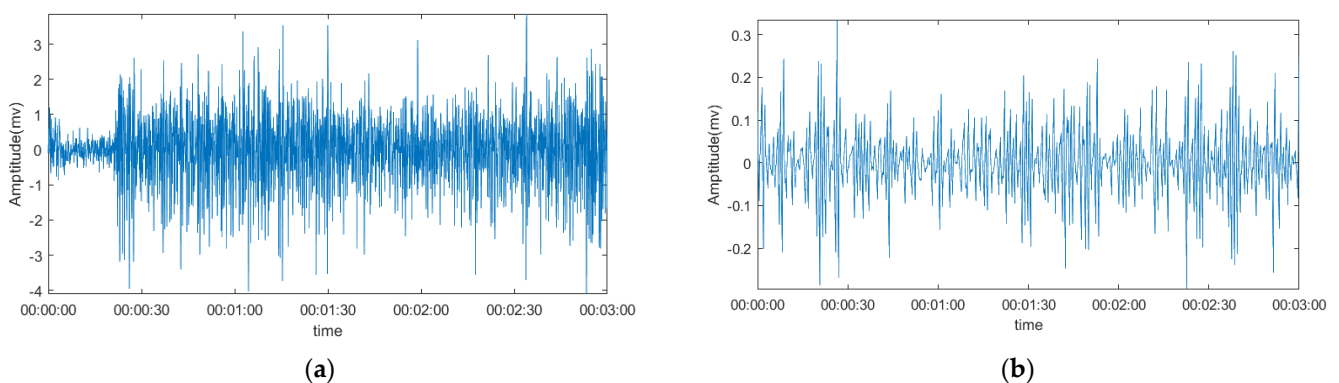


Figure 4. Cont.

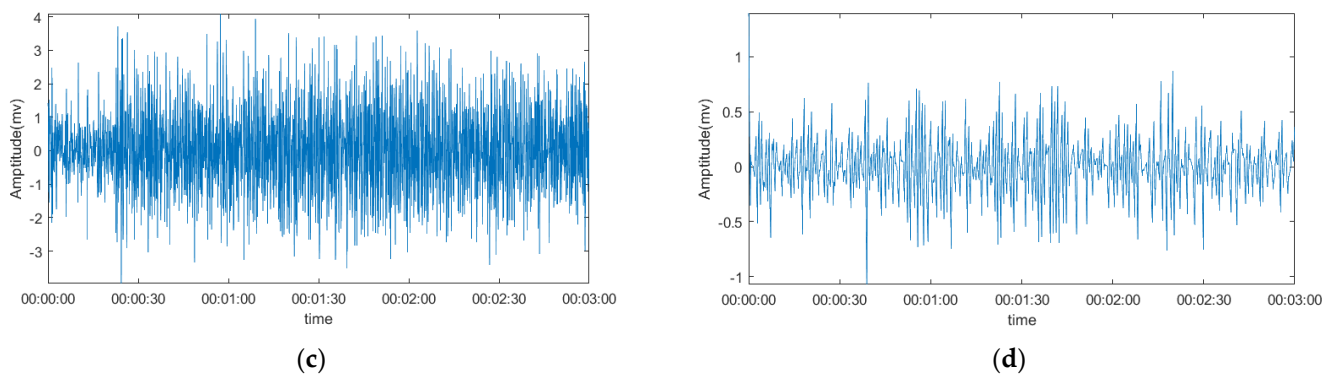


Figure 4. The outcome of the MPCA denoising process. (a) The original signal of channel 1. (b) The denoising signal of channel 1. (c) The original signal of channel 2. (d) The denoising signal of channel 2.

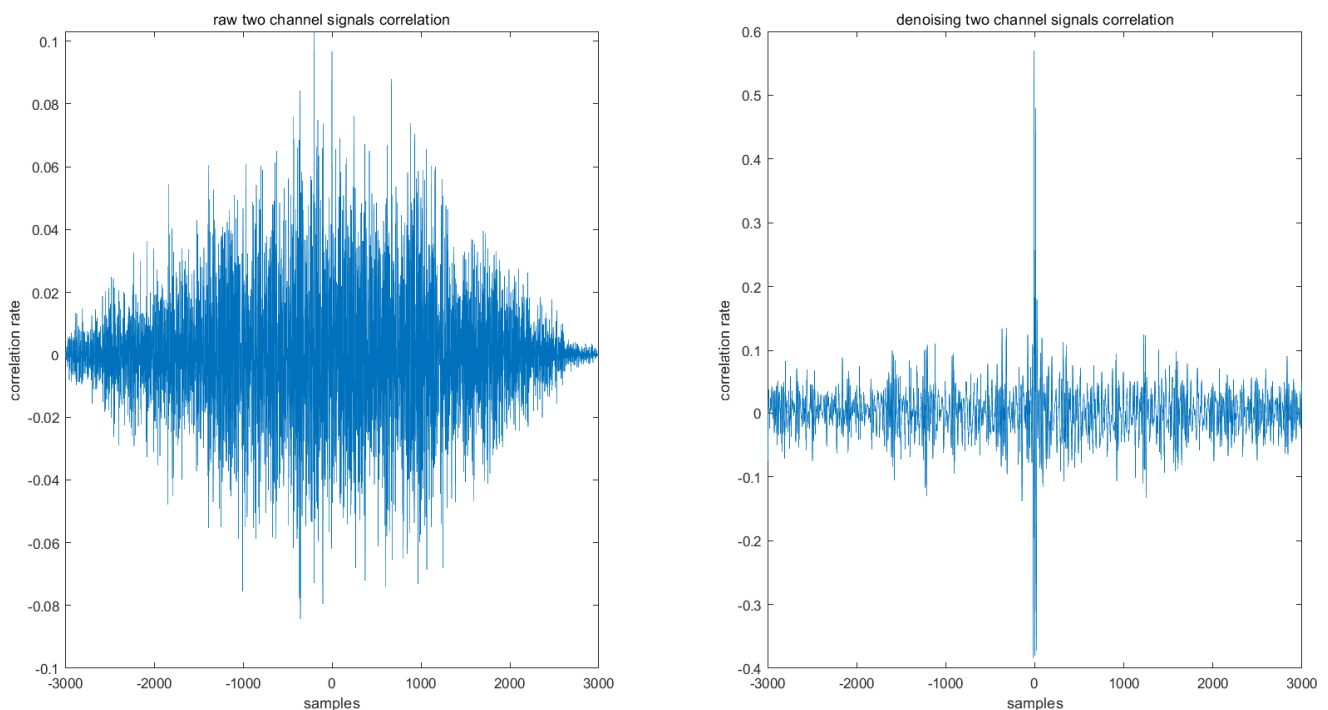


Figure 5. The correlation between the two channels.

3.3. Signal Decomposition

3.3.1. Wavelet Decomposition

In this experiment, a four-level wavelet decomposition of the EMG signal using the fourth-order Daubechies wavelet was implemented. The coarse-scale approximation coefficients and the detailed coefficients from the decomposition are shown in the upper figure.

Figure 6 shows the detail coefficients and the approximate coefficient of the denoising signal and the raw EMG signal. The difference between these two kinds of signal is obvious. The denoising signal shows a more clear change in aptitude, which is associated with the hand movement and is similar to the denoising signal in the first denoising step.

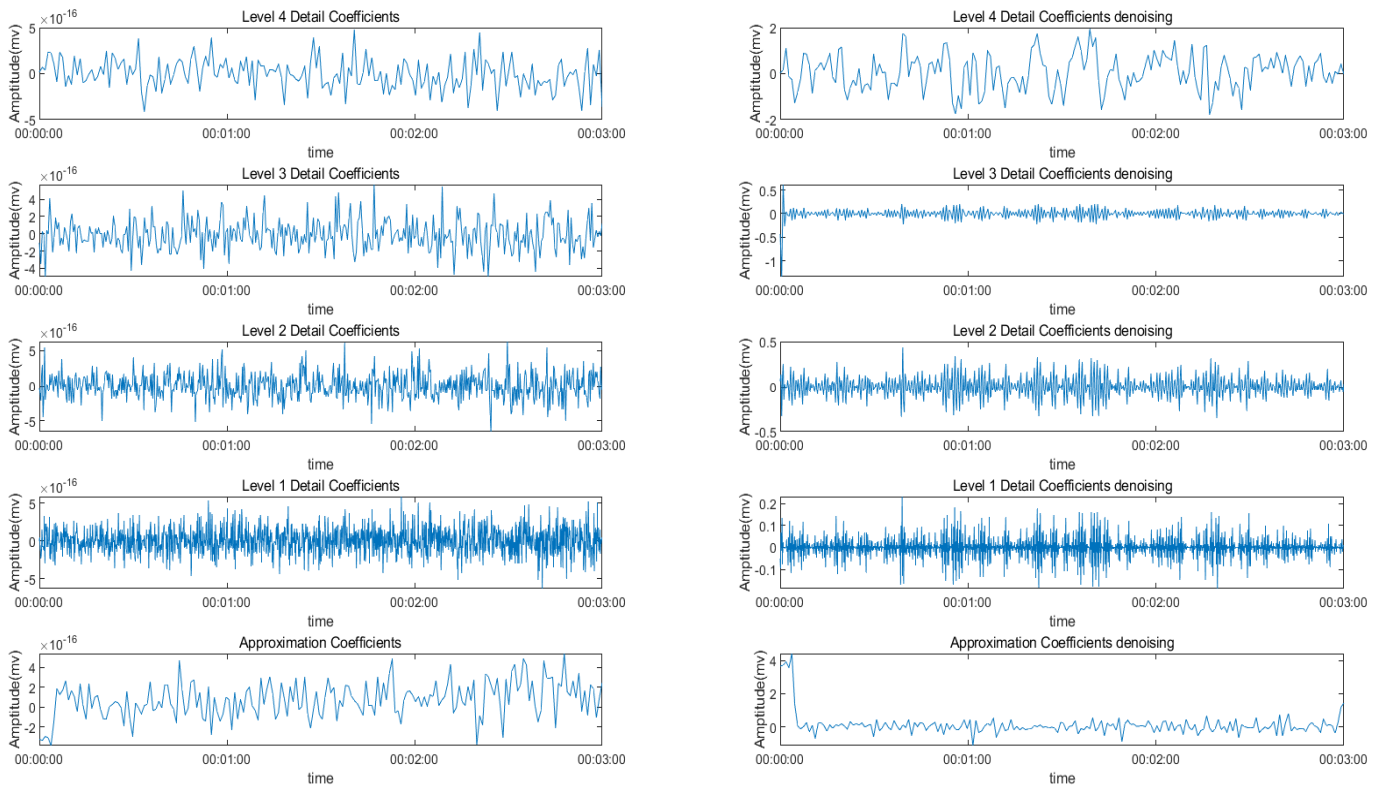


Figure 6. The wavelet decomposition coefficient feature vectors at different levels.

As can be seen in Figure 7, by calculating the correlation rate of the different levels of detail coefficients, the correlation between the different levels of wavelet detail coefficients increases after denoising.

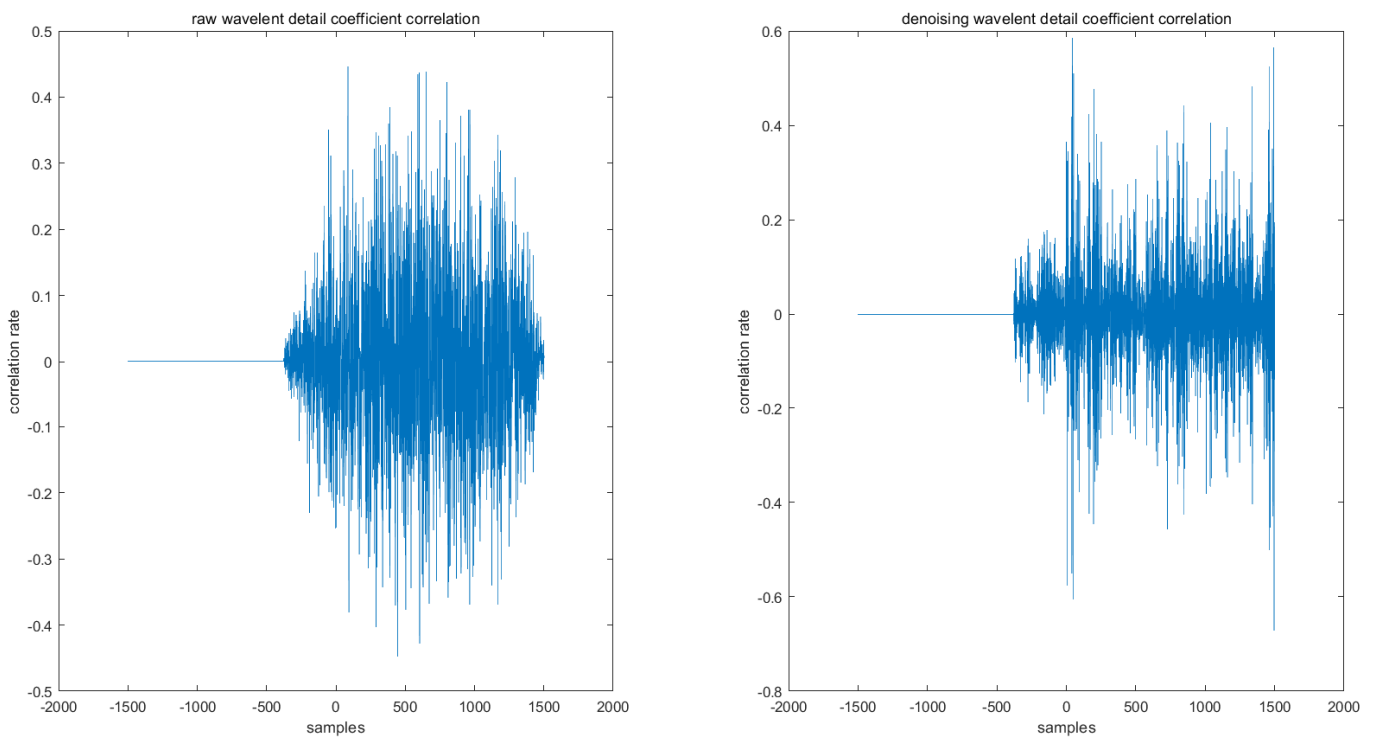


Figure 7. The correlation between the different levels of wavelet detail coefficients.

3.3.2. ICA Decomposition

Here, in the ICA decomposition steps which shows an example in the Figure 8, the construction function is used as the smoothing function, which is similar to the absolute value function and has four decomposition signals.

As well as the decomposition incorporation of the two kinds of EMG signals, these signals, when going through the ICA decomposition steps, also show that the denoised EMG signal has a clearer boundary between each MUAPT. However, different from the wavelet decomposition, the feature vector that the ICA method decomposed, the raw EMG signal seems to correlate more to each other and the original signals.

In these two different types of decomposition functions, we can see that the signal decomposed into several separate signals, which can extract features that are more representative than the original signal, and the denoising step clarified the features of the recomposited signal.

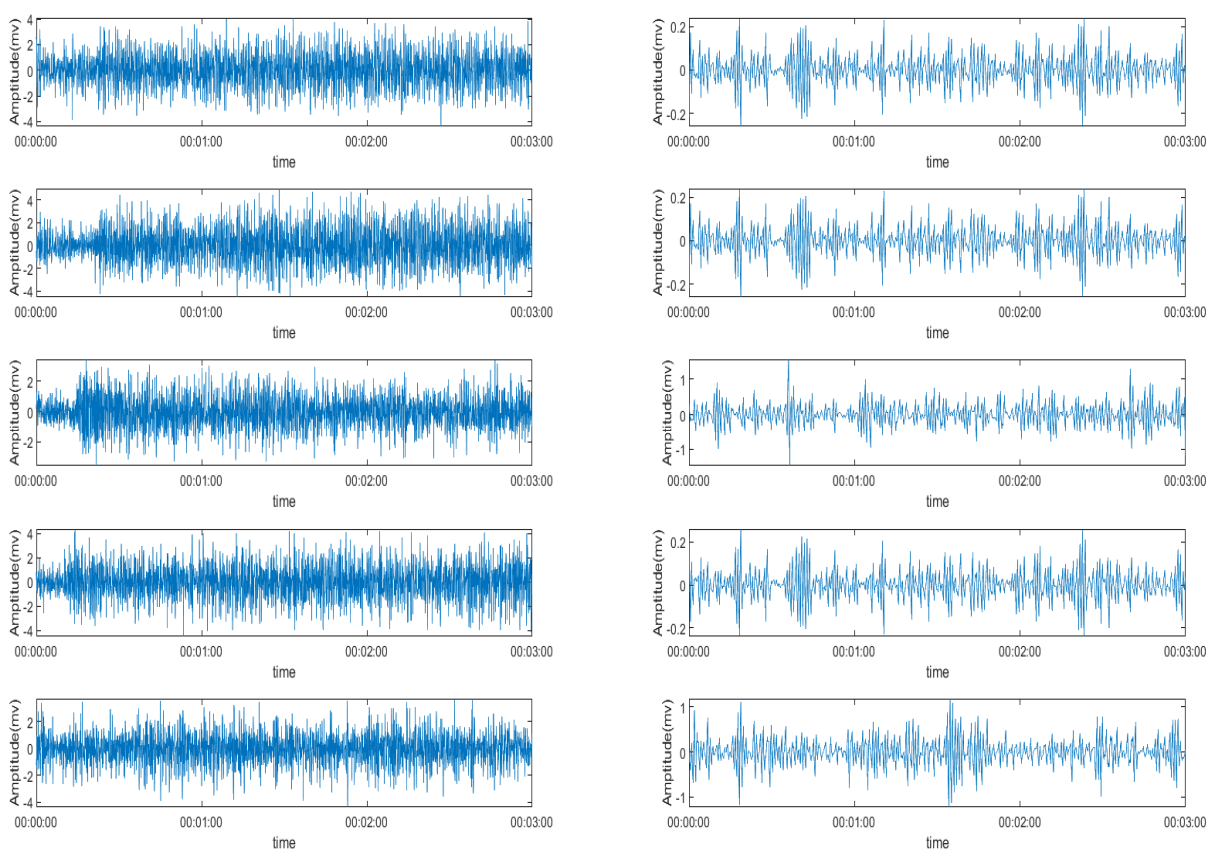


Figure 8. The ICA decomposition signals.

3.4. Features

There are three kinds of features that are extracted from the EMG signal: the first is the autoregression coefficient, which is calculated by the Yule–Walker method, and the other two are the coefficients generated by the decomposition steps as illustrated above. Here, the wavelet decomposition decomposes the EMG signal at level 9, and the detailed coefficient extracted by each level scale occupies a frequency band. The energy of each scale, which is calculated by the quadratic sum of all the coefficients in the scale, is selected and forms the feature vector DWTF.

$$ar_{coeff} = [ar_1, ar_2, ar_3, ar_4, ar_5, ar_6, ar_7, ar_8, ar_9, ar_{10}]$$

$$ICAF = [ica_1, ica_2, ica_3, ica_4, ica_5, ica_6, ica_7, ica_8, ica_9, ica_{10}]$$

$$DWTF = [ca_9, cd_9, cd_8, cd_7, cd_6, cd_5, cd_4, cd_3, cd_2, cd_1]$$

3.5. Classification

In the classification step, the features are separated into ten folds, and the cross-validation method is used to calculate the correction rate of the classification model. The correction rate is calculated by the following formula:

$$R = \frac{N_{tp} + N_{tf}}{N_t} \times 100\% \quad (29)$$

where N_{tp} is the number of features that are classified as the correct class, N_{tf} is the number of features that are classified as the wrong class, and N_t is the total number of features that participate in this classification process.

The content and meaning of Tables 1–5 are shown next. Tables 1 and 4 show the two different classification results of the three different classification methods. In Table 1, by using the K-nearest neighbor, the denoising only shows some improvement in the ICA coefficient and makes the classification correct rate of the autoregression coefficient and the WT coefficient, the confusion matrix of each classification result are illustrated in Tables 2, 3 and 5. However, in Table 4, the naïve Bayesian classification method shows a different result of the wavelet transform coefficient; the denoising step increased a lot of the classification accuracy of this feature.

Table 1. The KNN classification results.

Features	Multiscale PCA Denoising				Original Signal			
	k = 1	k = 4	k = 7	k = 10	k = 1	k = 4	k = 7	k = 10
WT coefficient	100%	80.27%	46.77%	53.33%	100%	85.05%	65%	74.5%
ICA coefficient	53.66%	47.88%	48.16%	46.61%	42.27%	38.16%	36.33%	33.05%
Autoregression coefficient	33.72%	33.27%	34.44%	31.88%	69.11%	72.38%	73.88%	73%

The difference between the KNN and the naïve Bayesian is that the classification method is different. The basic concept of the KNN is based on the distance between each sample, which is based on the Euclidean distance, but the naïve Bayesian is based on the probability of each category appearing under the condition that this item appears, whichever is the largest, considering which category the item to be classified belongs to. The differences in the classification methods also influenced the denoising method. However, in both classification methods, the MPCA has a negative influence on the classification accuracy rate of the time-domain feature—the autoregression coefficient.

Table 2. The confusion matrix of the KNN method—denoising part.

Features	Multiscale PCA Denoising																											
	k = 1						k = 4						k = 7						k = 10									
	a	b	c	d	e	f	<-- classified as	a	b	c	d	e	f	<-- classified as	a	b	c	d	e	f	<-- classified as	a	b	c	d	e	f	<-- classified as
<i>WT coefficient</i>	300	0	0	0	0	0	a = a	247	9	4	21	13	6	a = a	177	22	13	47	21	20	a = a	181	32	22	35	21	9	a = a
	0	300	0	0	0	0	b = b	9	225	22	21	11	12	b = b	25	129	47	42	20	37	b = b	32	135	47	36	28	22	b = b
	0	0	300	0	0	0	c = c	8	14	256	8	14	0	c = c	23	39	165	40	23	10	c = c	26	26	166	55	14	13	c = c
	0	0	0	300	0	0	d = d	19	22	22	213	16	8	d = d	42	50	70	88	31	19	d = d	44	53	45	121	29	8	d = d
	0	0	0	0	300	0	e = e	14	6	20	16	244	0	e = e	57	28	45	50	112	8	e = e	61	27	39	33	132	8	e = e
	0	0	0	0	0	300	f = f	6	6	11	9	8	260	f = f	24	21	28	22	34	171	f = f	16	14	19	12	14	225	f = f
<i>ICA coefficient</i>	145	37	24	27	33	34	a = a	148	42	29	30	34	17	a = a	145	43	29	33	37	13	a = a	136	43	31	30	39	21	a = a
	26	167	26	29	33	19	b = b	55	161	30	20	18	16	b = b	38	171	31	23	22	15	b = b	29	169	38	27	15	22	b = b
	26	21	194	23	18	18	c = c	32	32	188	19	15	14	c = c	28	31	193	19	12	17	c = c	30	26	189	21	17	17	c = c
	21	29	37	148	41	24	d = d	54	39	47	120	30	10	d = d	35	45	49	120	37	14	d = d	41	39	52	109	45	14	d = d
	29	29	29	27	153	33	e = e	51	41	29	41	118	20	e = e	40	42	33	47	123	15	e = e	42	42	34	44	121	17	e = e
	34	21	22	25	39	159	f = f	58	41	20	20	34	127	f = f	49	47	28	28	33	115	f = f	41	51	30	26	37	115	f = f
<i>Autoregression coefficient</i>	80	59	33	37	54	37	a = a	131	56	27	26	44	16	a = a	128	63	31	19	52	7	a = a	123	60	31	23	51	12	a = a
	57	85	39	35	45	39	b = b	85	106	33	24	37	15	b = b	72	110	36	28	42	12	b = b	82	86	38	32	40	22	b = b
	31	33	110	36	39	51	c = c	52	34	106	30	42	36	c = c	48	46	108	28	33	37	c = c	58	44	98	26	35	39	c = c
	39	38	34	115	28	46	d = d	66	48	39	89	26	32	d = d	63	49	43	104	20	21	d = d	51	55	43	101	27	23	d = d
	48	55	40	37	96	24	e = e	90	52	35	22	80	21	e = e	77	58	48	24	80	13	e = e	81	58	49	25	77	10	e = e
	28	36	56	37	22	121	f = f	50	41	58	41	23	87	f = f	43	44	56	39	28	90	f = f	50	46	51	42	22	89	f = f

Table 3. The confusion matrix of the KNN method—raw signal part.

Features	Original Signal																											
	k = 1							k = 4							k = 7							k = 10						
	a	b	c	d	e	f	<-- classified as	a	b	c	d	e	f	<-- classified as	a	b	c	d	e	f	<-- classified as	a	b	c	d	e	f	<-- classified as
<i>WT coefficient</i>	300	0	0	0	0	0	a = a	230	17	6	20	27	0	a = a	151	38	11	33	65	2	a = a	180	49	14	28	28	1	a = a
	0	300	0	0	0	0	b = b	12	256	8	7	17	0	b = b	40	210	11	16	23	0	b = b	23	238	2	15	22	0	b = b
	0	0	300	0	0	0	c = c	5	2	264	12	13	4	c = c	20	10	221	18	21	10	c = c	8	10	265	6	11	0	c = c
	0	0	0	300	0	0	d = d	8	8	8	262	9	5	d = d	33	14	20	207	18	8	d = d	31	7	16	220	23	3	d = d
	0	0	0	0	300	0	e = e	18	17	20	2	243	0	e = e	56	39	44	19	139	3	e = e	60	17	22	17	183	1	e = e
	0	0	0	0	0	300	f = f	6	0	4	10	4	276	f = f	15	0	14	21	8	242	f = f	12	0	5	17	11	255	f = f
<i>ICA coefficient</i>	118	40	40	33	38	31	a = a	149	60	28	26	25	12	a = a	140	63	31	30	25	11	a = a	123	71	32	22	36	16	a = a
	34	129	31	29	47	30	b = b	71	134	36	14	32	13	b = b	52	146	35	20	36	11	b = b	52	128	30	26	52	12	b = b
	41	29	132	29	39	30	c = c	63	55	114	17	31	20	c = c	60	69	91	24	36	20	c = c	71	70	77	27	36	19	c = c
	36	35	29	134	35	31	d = d	77	41	31	106	21	24	d = d	59	55	33	105	25	23	d = d	63	60	34	92	30	21	d = d
	35	48	42	30	121	24	e = e	71	68	35	15	100	11	e = e	62	83	40	14	85	16	e = e	51	80	41	21	92	15	e = e
	41	32	27	33	40	127	f = f	82	49	22	35	28	84	f = f	71	43	35	32	32	87	f = f	70	47	33	30	37	83	f = f
<i>Autoregression coefficient</i>	194	14	7	25	41	19	a = a	229	14	10	17	22	8	a = a	232	17	11	11	21	8	a = a	229	17	9	13	25	7	a = a
	25	223	10	14	20	8	b = b	34	234	7	6	15	4	b = b	25	237	8	5	20	5	b = b	28	222	11	11	22	6	b = b
	11	17	214	25	11	22	c = c	15	21	230	17	9	8	c = c	14	22	234	14	9	7	c = c	13	17	240	14	10	6	c = c
	23	11	21	205	15	25	d = d	32	11	28	200	7	22	d = d	27	15	24	203	9	22	d = d	23	13	32	205	10	17	d = d
	47	27	17	17	190	2	e = e	58	22	11	14	195	0	e = e	46	21	12	18	203	0	e = e	43	24	13	13	207	0	e = e
	18	15	27	16	6	218	f = f	21	11	32	17	4	215	f = f	24	11	25	16	3	221	f = f	20	13	30	23	3	211	f = f

Table 4. Classification with the naïve Bayesian method.

Features	Multiscale PCA Denoising	Original Signal
WT coefficient	22.72%	46%
ICA coefficient	15.27%	8.27%
Autoregression coefficient	22.38%	43.5%

Table 5. The confusion matrix of naïve Bayesian method.

Features	Multiscale PCA Denoising							Original Signal						
	a	b	c	d	e	f	<-- classified as	a	b	c	d	e	f	<-- classified as
WT coefficient	253	5	6	9	7	20	a = a	83	105	11	31	50	20	a = a
	237	10	15	0	8	30	b = b	65	205	1	5	9	15	b = b
	191	5	37	26	21	20	c = c	20	75	127	20	36	22	c = c
	215	10	15	28	22	10	d = d	84	23	11	133	24	25	d = d
	229	5	22	18	1	25	e = e	73	97	20	12	75	23	e = e
	161	18	11	5	25	80	f = f	0	25	50	20	0	205	f = f
ICA coefficient	39	77	35	46	53	50	a = a	41	59	53	53	30	64	a = a
	53	69	41	64	33	40	b = b	85	18	58	43	33	63	b = b
	69	62	43	44	35	47	c = c	76	60	24	50	33	57	c = c
	54	75	36	49	42	44	d = d	83	56	51	19	43	48	d = d
	59	66	34	63	25	53	e = e	75	59	49	35	14	68	e = e
	42	88	30	53	37	50	f = f	79	47	50	43	48	33	f = f
Autoregression coefficient	6	139	9	38	108	0	a = a	215	4	37	2	23	19	a = a
	8	158	18	30	85	1	b = b	75	52	63	5	61	44	b = b
	5	107	20	108	60	0	c = c	34	5	211	35	3	12	c = c
	7	87	14	121	67	4	d = d	83	3	80	69	36	29	d = d
	9	116	12	65	98	0	e = e	119	10	54	30	63	24	e = e
	2	111	25	92	70	0	f = f	49	8	47	23	0	173	f = f

4. Discussion

In this experiment, several features were extracted from a raw EMG signal. The autoregression coefficient is a time-domain feature that represents the time connection of one signal. Independent component analysis and wavelet transform analysis describe a signal in the frequency domain; they decompose the signal into different scales, and ICA decomposition separates the signal into different signals that show the components of the original signal. Wavelet transform uses different basic wavelet functions to separate the signal to simulate the original one. There are some similarities between the two different decomposition methods, such as the fact that they are both frequency-domain feature extraction methods and they both divide the original signal into different levels of decomposition signals. By judging the classification accuracy and confusion matrix of the two, we can see that the features formed by the wavelet decomposition method show better results in both classification methods. The effect of wavelet denoising for ICA decomposition is much better than that of wavelet decomposition. In both classification methods, the features of the ICA decomposition method significantly improved after wavelet denoising.

In the KNN classification algorithm method, different neighbor numbers affect the correction rate, so different neighbor numbers were tested in this paper from $K = 1$ to $K = 10$ by 3 to test which value of k is suitable for classification. Moreover, we can see that when the value is 7, the classification accuracy of the ICA coefficient feature and the

autocorrelation feature reach the maximum point. However, the WT coefficient feature reaches the maximum point when $k = 1$.

The results are shown in the Classification Section. By judging the information in Tables 1–5 in Section 3.4, we can see that the classification accuracy rate of the autoregression coefficient feature shows a huge decrease, which only successfully classifies half of the features compared to the one that uses the original signal to carry out the feature extraction. The Bayesian method showed a poor accuracy rate, and none of the features expected in the naïve Bayesian classification method exceeded 50%. The denoising step improved the classification rate of the ICA features but showed a worse result in the classification of the other two features. The reason, we believe, is that multiscale PCA uses wavelet transform for denoising, so the denoising process is more efficient in the frequency domain, and this process can erase some of the important relationships in the time series, so the autoregression coefficient shows worse behavior after the denoising process. For the wavelet decomposition coefficient, we believe the reason why the classification behaves worse is that the wavelet transform was already used in the denoising step.

This study has several limitations, including using too-few EMG signal data. Although our findings indicate that MPCA can improve the classification rate of ICA features with the KNN method and naïve Bayesian classification method, more empirical studies on the denoising steps need to be conducted in future tests, such as using the naïve Bayesian classification method.

In the research process, we reviewed many papers, and much of the research relates to electroencephalograms (EEGs). EEG signal processing is very popular and is similar to EMG signal processing. Additionally, the brain–computer interface, a concept that has attracted our attention, is highly relevant to these two signals. EEG signals have low accuracy, and some patients suffer from neuromuscular disease, which means that the EMG signals are limited [28]. By combining these two signals, we can acquire more accurate muscle movement for applications such as typing without hands. In future studies, we would like to study the difference between these two signals and find a way to make use of both of them.

Author Contributions: Conceptualization, C.Z.; Data curation, C.Z.; Formal analysis, C.Z.; Funding acquisition, C.Z.; Investigation, C.Z.; Methodology, C.Z.; Project administration, C.Z.; Resources, C.Z.; Software, C.Z.; Supervision, T.S.; Validation, C.Z.; Visualization, C.Z.; Writing—original draft, C.Z.; Writing—review & editing, C.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are openly available in [UCI Machine Learning Repository] at [<https://archive.ics.uci.edu/ml/datasets/sEMG+for+Basic+Hand+movements>], reference number [27].

Acknowledgments: Thanks to Christopher Freeman for helping me with my postgraduation study and giving me guidance for the graduation paper that helped me discover the beauty of the bioinformatics field. I am very grateful for my parents who help me and support me in whatever I decide, and this support made it possible to finish this work.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Pauk, J. Different techniques for EMG signal processing. *J. Vibroeng.* **2008**, *10*, 419.
2. Reaz, M.B.I.; Hussain, M.S.; Mohd-Yasin, F. Techniques of EMG signal analysis: Detection, processing, classification and applications. *Biol. Proced. Online* **2006**, *8*, 11–35. [[CrossRef](#)] [[PubMed](#)]
3. Yousefi, J.; Andrew, H.-W. Characterizing EMG data using machine-learning tools. *Comput. Biol. Med.* **2014**, *51*, 1–13. [[CrossRef](#)] [[PubMed](#)]
4. Hutten, G.J.; van Thuijl, H.F.; van Bellegem, A.C.M.; van Eykern, L.A.; van Aalderen, W.M.C. A literature review of the methodology of EMG recordings of the diaphragm. *J. Electromyogr. Kinesiol.* **2010**, *20*, 185–190. [[CrossRef](#)]

5. Negro, F.; Muceli, S.; Castronovo, A.M.; Holobar, A.; Farina, D. Multi-channel intramuscular and surface EMG decomposition by convolutive blind source separation. *J. Neural Eng.* **2016**, *13*, 026027. [CrossRef]
6. Chang, W.X. LMS Algorithm Based on C-expansion of Quaternion. *Signal Process.* **2011**, *27*, 1515–1519.
7. Dan, S. EMG signal decomposition: How can it be accomplished and used? *J. Electromyogr. Kinesiol.* **2001**, *11*, 151–173.
8. Chowdhury, R.H.; Reaz, M.B.I.; Ali, M.A.B.M.; Bakar, A.A.A.; Chellappan, K.; Chang, T.G. Surface Electromyography Signal Processing and Classification Techniques. *Sensors* **2013**, *13*, 12431–12466. [CrossRef]
9. Akinwumi, A.A.; Gorban, A.N. Multiscale principal component analysis. *J. Phys.* **2014**, *490*, 12081.
10. Gokgoz, E.; Subasi, A. Effect of multiscale PCA de-noising on EMG signal classification for diagnosis of neuromuscular disorders. *J. Med. Syst.* **2014**, *38*, 31. [CrossRef]
11. Xu, Z.; Yu, W.; Han, R. Wavelet transform theory and its application in EMG signal processing. In Proceedings of the IEEE Seventh International Conference on Fuzzy Systems & Knowledge Discovery, Yantai, China, 10–12 August 2010.
12. Wang, G.; Zhang, Y.; Liu, C.; Xie, Q.; Xu, Y. A new tool wear monitoring method based on multi-scale PCA. *J. Intell. Manuf.* **2019**, *30*, 113–122. [CrossRef]
13. Sharma, L.N.; Danadapat, S.; Mahanta, A. Abstract-multiscale multiscale PCA based quality controlled denoising of multichannel ECG signals. *Int. J. Inf. Electron. Eng.* **2017**, *2*, 107–111.
14. García, G.A.; Maekawa, K.; Akazawa, K. Decomposition of Synthetic Multi-channel Surface-Electromyogram Using Independent Component Analysis. In Proceedings of the International Conference on Independent Component Analysis and Signal Separation, Granada, Spain, 22–24 September 2004; pp. 985–992.
15. Chawla, M.P.S. A comparative analysis of principal component and independent component techniques for electrocardiograms. *Neural Comput. Appl.* **2009**, *18*, 539–556. [CrossRef]
16. Shi, Z.; Tang, H.; Tang, Y. A fast fixed-point algorithm for complexity pursuit. *Neurocomputing* **2005**, *64*, 529–536. [CrossRef]
17. Phinyomark, A.; Quaine, F.; Charbonnier, S.; Serviere, C.; Tarpin-Bernard, F.; Laurillau, Y. EMG feature evaluation for improving myoelectric pattern recognition robustness. *Expert Syst. Appl.* **2013**, *40*, 4832–4840. [CrossRef]
18. Phinyomark, A.; Pornchai, P.; Chusak, L. Feature reduction and selection for EMG signal classification. *Expert Syst. Appl.* **2012**, *39*, 7420–7431. [CrossRef]
19. Paiss, O.; Inbar, G.F. Autoregressive modeling of surface EMG and its spectrum with application to fatigue. *Math. Comput. Model.* **1987**, *10*, 793–794. [CrossRef]
20. Ramírez-Martínez, D.; Alfaro-Ponce, M.; Pogrebnyak, O.; Aldape-Pérez, M.; Argüelles-Cruz, A.-J. Hand Movement Classification Using Burg Reflection Coefficients. *Sensors* **2019**, *19*, 475. [CrossRef]
21. Brockwell, P.J.; Dahlhaus, R. Generalized Levinson–Durbin and Burg algorithms. *J. Econom.* **2004**, *118*, 129–149. [CrossRef]
22. Hamilton-Wright, A.; McLean, L.; Stashuk, D.W.; Calder, K.M. Bayesian aggregation versus majority vote in the characterization of non-specific arm pain based on quantitative needle electromyography. *J. Neuroeng. Rehabil.* **2010**, *7*, 8. [CrossRef]
23. Gokgoz, E.; Abdulhamit, S. Comparison of decision tree algorithms for EMG signal classification using DWT. *Biomed. Signal Process. Control* **2015**, *18*, 138–144. [CrossRef]
24. Al-Faiz, M.Z.; Ali, A.A.; Miry, A.H. A k-nearest neighbor based algorithm for human arm movements recognition using EMG signals. In Proceedings of the IEEE International Conference on Energy, Manama, Bahrain, 18–21 December 2010.
25. Rasheed, S.; Stashuk, D.; Kamel, M. Adaptive fuzzy k-NN classifier for EMG signal decomposition. *Med. Eng. Phys.* **2006**, *28*, 694–709. [CrossRef] [PubMed]
26. Sapsanis, C.; Georgoulas, G.; Tzes, A.; Lymberopoulos, D. Improving EMG based classification of basic hand movements using EMD. In Proceedings of the 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 13), Osaka, Japan, 3–7 July 2013; pp. 5754–5757.
27. UCL, sEMG for Basic Hand Movements Data Set. Available online: <http://archive.ics.uci.edu/ml/datasets/sEMG+for+Basic+Hand+movements#> (accessed on 14 September 2020).
28. Chowdhury, A.; Raza, H.; Meena, Y.K.; Dutta, A.; Prasad, G. EEG-EMG correlation-based brain-computer interface for hand orthosis supported neuro-rehabilitation. *J. Neurosci. Methods* **2019**, *312*, 1–11. [CrossRef] [PubMed]