IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 13 September 2022, accepted 1 October 2022, date of publication 5 October 2022, date of current version 13 October 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3212146

RESEARCH ARTICLE

Hand Gesture Recognition With Acoustic Myography and Wavelet Scattering Transform

ALI H. AL-TIMEMY[1], YOUSSEF SERRESTOU[2], RAMI N. KHUSHABA[3], (Senior Member, IEEE), SLIM YACOUB[4], AND KOSAI RAOOF[4]

[1]Biomedical Engineering Department, Al-Khwarizmi College of Engineering, University of Baghdad, Baghdad 10071, Iraq
[2]LAUM, Le Mans University, 72085 Le Mans, France
[3]Australian Center for Field Robotics, The University of Sydney, Sydney, NSW 2006, Australia
[4]LTSIRS, INSAT-Carthage University, 1080 Tunis, Tunisia

Corresponding author: Ali H. Al-Timemy (ali.altimemy@kecbu.uobaghdad.edu.iq)

This work was supported in part by the RFI-Wise, in part by the University of Baghdad, and in part by Le Mans University (ENSIM and LAUM). The work of Ali H. Al-Timemy was supported in part by the Campus France; and in part by the French Embassy, Iraq.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Local Committee at the University of Baghdad, and performed in line with the Declaration of Helsinki.

ABSTRACT In the past decade, improving upper limb prostheses control methods with pattern recognition (PR) has been the focus of an extended amount of research. However, several challenges associated with the processing of the Electromyogram (EMG) signals still need to be tackled to enable widespread and clinical implementation of upper limb prostheses with PR. As a result, alternative modalities functioning as promising control signals have been proposed as source of control input rather than the surface EMG, such as Acoustic myography (AMG) and Force myography. In this paper, eight high sensitivity array microphones were utilized to acquire the AMG signals, with 8 custom-built 3D printed microphone housing developed for the purpose of this research. Twenty subjects were recruited for data collection in this paper with the hardware design developed specifically by our research team, making our database the largest open-access dataset in the AMG literature. We proposed a novel feature extraction (FE) method based on the Wavelet Scattering Transform (WST) to tackle the challenge of extracting the relevant information from AMG to classify 14 hand and finger movement classes. The WST is a translation-invariant non-linear signal representation that has a strong theoretical support as it maintains stability to time-warping deformations, while preserving a high degree of discriminability. The performance results showed that WST outperformed all state-of-the-art FE methods with an accuracy of 88% on average across 20 subjects when classified with Quadratic Discriminant Analysis (QDA) classifier for a large dataset of AMG signals. These results suggest that the AMG signals can be utilized as a reliable source of control, especially when the windows sizes and number of channels are carefully selected. The AMG dataset is available from the link https://drive.google.com/drive/folders/1r0rBnrNG5c8qegffKUhAYYHQUT8cGVHD?usp=sharing.

INDEX TERMS Acoustic myography, pattern recognition, wavelet scattering, hand gesture recognition, upper limb prostheses.

I. INTRODUCTION

For individuals with limb amputations, who lost a limb due to trauma, accidents and wars, upper limb prostheses controlled with biosignals can offer a mean to allow them to perform the activities of daily living. In an approach to help amputees to gain partial functionality of the lost hand, Pattern Recognition (PR) methods are usually applied on the surface Electromyogram (EMG) signals to decipher the movement intentions of control [1], [2], an approach which has achieved commercial success.[1] However, the high cost of upper limb prostheses is still a barrier for most amputees. Additionally,

The associate editor coordinating the review of this manuscript and approving it for publication was Ahsan Khandoker.

[1]https://coaptengineering.com/

the lack of stability overtime and performance degradation due to daily factors such as signal non-stationarity [3], contraction level change [4] and limb position change [5] will limit their widespread utilization. Furthermore, the acquired signals can be easily contaminated with variety of noises such as power line interference, ECG contamination, and environmental noise.

Since the EMG signal is electrical in its nature, other modalities such as mechanical signals have been investigated for the control of upper limb prostheses [6], including: Force myography (FMG) [7], Mechanomyography (MMG) [8], [9], [10] and Acoustic Myography (AMG) [11], [12]. These methods have shown the potential as a source of control, being less affected by noise contamination than that usually affecting surface EMG.

AMG is the recording of the mechanical vibrations of the muscle at a low frequency [13], usually below 100 Hz [14], [15], with the dominant frequency below 50 Hz. Both accelerometers [11], [16] or microphones [12] can be used to record the AMG signal. The mechanical vibrations recorded with the microphones are termed acoustic myography [6]. A variety of applications have been performed with AMG such as providing control signals for prosthesis control with PR system [16] and muscle activity assessment [17]. However, while extensive research has been made on EMG feature extraction [18], there is still a need for emerging feature extraction methods to be investigated for extracting relevant information from the acoustic muscle signals, i.e. AMG.

Early research investigating pattern recognition for gesture recognition with AMG was reported in the work in [19], where 8 microphones were utilized to record 7 gestures offline. The authors utilized a variety of power spectral density features and a Support Vector Machines (SVM) classifier to obtain a classification accuracy of 88%. However, the number of gestures was relatively small. To harvest the information provided with both EMG and MMG, a hybrid EMG and MMG system made with microphones and accelerometer for an assisted EMG-MMG control of prostheses was developed by Guo et al. [16]. Seven intact-limbed and 2 amputee subjects were recruited to perform 13 classes of hand and wrist movements, while utilizing four channels of EMG and MMG channels. It was shown that adding the MMG improved the results significantly upon using EMG only, where the AMG collected by microphones achieved better performance than those of the MMG acquired with accelerometers. In a latter study [9],a real-time gesture recognition wristband was developed based on EMG and inertial measurement unit (IMU) for recognizing 12 gestures (8 air gestures and 4 surface gestures) with 2 force levels. The system achieved an accuracy of 92.6% and 88.8 %, when tested on 10 subjects for the air and surface gestures, respectively.

A multi-channel finger pattern recognition was performed in [20] with MMG, consisting of statistical, frequency-domain, and acoustic features; and classified by artificial neural networks (ANNs). A real-time average classification accuracy for seven movements of 88.6% - 95.1% was obtained for the intact-limbed subjects where the best performers were the male athletes' subgroup, while for the amputees, the accuracy was 74.4% and 77.6% for unilateral transradial and wrist amputation subjects, respectively. However, the number of movements investigated was relatively small.

While single modality AMG systems performed well, adding other modalities such as MMG [10] or FMG seems to add more information, therefore improving the overall performance of the system. For instance, a combination of FMG and MMG was utilized in [21] to construct a hybrid control system of MMG and FMG signals. The hybrid system was evaluated with PR, where twelve movements were performed by 12 subjects including 6 amputees. In terms of feature extraction, time-domain, time-frequency, and acoustic features of MMG-FMG sensors were all investigated. An average classification accuracy of 91.44% $\pm$0.77 was obtained for amputees and 92.19% $\pm$1.12 intact limbed subjects. A hybrid system was also proposed in [22], with 10 microphones and IMUs. Thirteen daily life gestures were implemented with their protype. An accuracy of 75% accuracy was obtained when placing the sensors on the wrist, with the best-chosen features. In addition, the wrist was found to be the optimal area for the acoustic measurements, with an accuracy of 80%, using one microphone and IMU. To investigate the effect of adding more microphones, 40 microphones were placed on 4 rows near the wrist for recognizing hand gestures [12]. An accuracy of 90% for 34 American sign language gestures was obtained with time-domain features with LDA and SVM classifiers. However, challenges exist when attempting to implement this number of microphones within any real-time platform for daily usage. Furthermore, a performance of up to 84% was reported when using 10 microphones, and testing on 10 subjects, which in turn indicates the importance of having a large number of sensors to reliably identify the large number of gestures proposed.

Four microphones were also utilised in [8] to construct a wrist band human-machine interfacing based on 6 wrist movements. A PR system consisting of of 213 spectro-temporal features per channel and a Neighbourhood Component Analysis classifier was used to obtain an accuracy of 75% for 7 intact-limbed and 79-90% for the 3 amputees. It is worth noting that no finger or pinch movements were investigated in their work. The previous research shows the potential of utilising PR system based on AMG for recognizing hand movements.

From the previous literature, it is obvious that challenges still need to be tackled such as gestures of daily living by the amputees and relatively low accuracy below that of usable PR system (defined as $\leq$ 10 in the literature [23]). In this study, 14 selected hand and finger movements used by the amputees in their daily living will be investigated with 8 channels AMG signals. The data set is the largest for acoustic myography as it involves recordings from 20 subjects, including 7 females. Additionally, our analysis presents a new feature extraction method based on the wavelet scattering transform (WST)

for acoustic measurements of the wrist, on 8 channel AMG. The performance of the proposed method is then compared with other FE methods utilized in the EMG literature to show the superior performance of the proposed method. Our research hypothesis is that when considering the AMG as a single input modality one can achieve significant performances by focusing on the state-of-the-art feature extraction provided by WST rather than relying on traditional feature extraction techniques. The experiments presented in the next sections investigate the applicability of our hypothesis against traditional techniques to provide the value of our proposed approach.

II. METHODOLOGY

A. THE PROPOSED FEATURE EXTRACTION BASED ON THE WAVELET SCATTERING TRANSFORM (WST)

Given the nature of the AMG signal and the fact that previous research in this direction relied mainly upon traditional time-frequency and power spectral density features, we opted to further extend upon the methods from the literature by proposing the use of the wavelet scattering transform (WST). The WST is a translation-invariant non-linear signal representation that has a strong theoretical support as it maintains a stability to time-warping deformations, while preserving a high degree of discriminability [24]. The WST is defined as a complex-valued convolutional neural network that replaces data-driven filters by wavelets and the non-linearity by a complex modulus [24], [25]. The wavelet transform convolves an input signal $x(t)$ with a filter bank of ψ_{λ_1} that is made by dilations of a mother wavelet $\psi(t)$, whose Fourier transform $\hat{\psi}(w)$ is concentrated over the dimensionless frequency interval $[1 - 2^{1/2Q}; 1 + 2^{1/2Q}]$, with Q being the quality factor (the number of wavelets filters per octave for each filter bank). By dilating the mother wavelet, a family of bandpass filters centered at $\lambda_1 = 2^{j_1 + \frac{\chi}{Q}}$ is generated, with the indices of $j_1 \in \mathbb{Z}$ and $\chi \in \{1 \dots Q\}$ respectively denoting the octave and chroma.

$$\hat{\psi}_{\lambda_1}(w) = \hat{\psi}\left(\lambda^{-1} w\right) \text{ i.e., } \psi_{\lambda_1}(w) = \lambda_1 \psi(\lambda_1 t). \quad (1)$$

A scalogram matrix is then generated by applying the complex modulus to all wavelet convolutions (denoted by $\star$) with the signal $x(t)$, that is $x_1(t, \lambda_1) = |x \star \psi_{\lambda_1}|$ with a frequential axis that is uniformly sampled by the binary logarithm. The energy of $x(t)$ is localized by the scalogram x_1 around frequencies λ_1 over durations of $2Q\lambda_1^{-1}$ [26]. The WST coefficients are obtained by averaging the wavelet modulus coefficients with a low-pass filter $\phi(t)$ of size T, which ensures local invariance to time- shifts.

$$S_1 x(t, \lambda_1) = |x \star \psi_{\lambda_1}| \star \phi_T(t). \quad (2)$$

At the zero order, a single coefficient is generated by $S_0 x(t) = x \lambda_1 \phi(t)$, as these have very low energy at low frequencies [24]. As the low-pass filtering process removes all high-frequencies, these are then recovered by a wavelet modulus transform as the time scattering transform also convolves x_1 with a second filter-bank of wavelets ψ_{λ_2} and

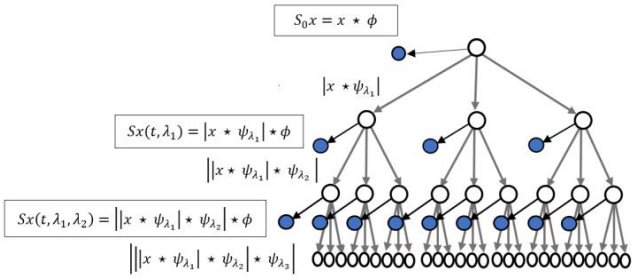

FIGURE 1. The wavelet scattering transform hierarchical representation.

applies complex modulus to obtain:

$$x_2(t, \lambda_1, \lambda_2) = |x_1 \star \psi_{\lambda_2}| = ||x \star \psi_{\lambda_1}| \star \psi_{\lambda_2}|. \quad (3)$$

In the same way as Eq.2, translation invariance in time up to T is attained by averaging. Hence, the second order coefficients, capturing the high-frequency amplitude modulations occurring at each frequency band of the first layer, are obtained by

$$S_2 x(t, \lambda_1, \lambda_2) = ||x \star \psi_{\lambda_1}| \star \psi_{\lambda_2}| \star \phi_T(t). \quad (4)$$

with an octave resolution of Q_2 for the wavelet ψ_{λ_2} having, where Q_2 is different from Q_1. By setting $Q_2 = 1$, a narrower time support is defined for wavelets that are suitable to characterize transients and attacks. As a result, a sparse representation is developed leading to concentrating the signal information over as few wavelet coefficients as possible. A low-pass filter φ then averages these coefficients, this in turn ensures local invariance to time-shifts, as with the first-order coefficients [27]. The scattering coefficients are defined by iterating the operation at any order m (see Fig.1 for an illustration of the WST representation). For any $m > 1$, iterated wavelet modulus convolutions are written as:

$$U_m x(t, \lambda_1, \dots, \lambda_m) = \left| \left| |x \star \psi_{\lambda_1}| \star \dots \right| \star \psi_{\lambda_m}(t) \right|, \quad (5)$$

where the m^{th}-order wavelets ψ_{λ_m} have an octave resolution Q_m. The application of φ on $U_m x$ gives scattering coefficients of order m

$$\begin{aligned} S_m x(t, \lambda_1, \dots, \lambda_m) &= \left| \left| |x \star \psi_{\lambda_1}| \star \dots \right| \star \psi_{\lambda_m(t)} \right| \star \phi(t) \\ &= U_m x(t, \lambda_1, \dots, \lambda_m) * \phi(t) \end{aligned} \quad (6)$$

The final scattering vector for a scattering decomposition of maximal order l collects all scattering coefficients for $0 \leq m \leq l$. The WST allows low-variance features to be extracted from real-valued time series and image data [24], [28], [29], [30], [31], [32]. It provides a signal deformation and invariance stable representation, making it suited for a wide range of signal processing and machine/deep learning applications that have already been examined in the literature. Given the nature of the WST within which the energy diminishes across the decomposition levels, a three-layers WST is usually adequate for practically many applications.

In comparison to deep convolutional neural networks, the WST can avoid the need for multiple model parameters, high computational costs, hyperparameter adjustment, and

difficulties comprehending and interpreting the extracted features. It also provides translation invariance, local deformation stability, and rich feature information storage that are all attractive properties for any feature extraction process. To further elaborate on the power of the WST based feature extraction on AMG signals, we also include comparisons with the traditional wavelet transform and wavelet-packet transform based feature extraction processes as well as some recently developed combination of time-domain based feature extraction techniques that were mainly proposed in the literature for EMG classification.

B. SUBJECTS

In this study, 20 subjects (13 males and 7 females) were recruited. The mean age of the subjects was 25.4 ± 9.3 years (Mean ± SD) (mean age: 27.7 years for males; mean age: 21.5 years for females). Before starting the experimental data collection, subjects were debriefed about the experiment, and they gave their signed consent to participate in the study. The experiments in this study were performed in accordance with the declaration of Helsinki and its updates [33].

C. AMG SIGNAL ACQUISITION

Eight high sensitivity array microphones (MPA416, BSWA technology, China) were utilised to acquire the AMG signals with a frequency range of 20Hz-20KHz, open-circuit sensitivity (50 mV/Pa)(±2dB) and inherent noise of 29dBA.

The MPA416 microphones were calibrated by the manufacturer to ensure accuracy of the measurements. The MPA416 built-in Integrated Electronics Piezo-Electric (IEPE) preamplifier helps to amplify the AMG signals, to ensure a good signal-to-noise ratio (SNR), since the amplitude of AMG signal is very low. National Instruments (NI), 24-Bit PXI module for acoustic and vibration measurements (NI PXIe-4492), mounted on an NI PXIe-1073 chasse was used to perform data acquisition of 8 channels AMG at a sampling rate of 1024 Hz. A Labview Virtual Instrument (VI) was developed to view the AMG signals in real-time and to save them for next parts of analysis.

We constructed 8 microphone housings with a 3D printer to acquire the AMG signal the shape and dimensions of the housings are inspired by those in [34]. We designed the housings so that the microphones could fit inside, and the base of the housing was shaped like a cone. The cones were then integrated with an elastic armband and placed on the forearm. This allowed us to eliminate artifacts due to movement, as proper microphone attachment can preserve the characteristics of AMG signals [35], [36].

D. SENSOR PLACEMENT

A pilot study was performed to find the best location of the AMG sensors of the forearm. Four locations were investigated by placing four microphones on the four locations and recording the hand close movement with 4 contraction levels (low, medium, high and maximum). The recording of the AMG for the 4 channels and the locations are shown in Fig. 2. We found that location 3 and 4 at the upper part of the forearm near the elbow gave a better signal quality and higher amplifier than those near the wrist. For that reason, the 8 channels of the AMG will be placed on the upper part of the forearm.

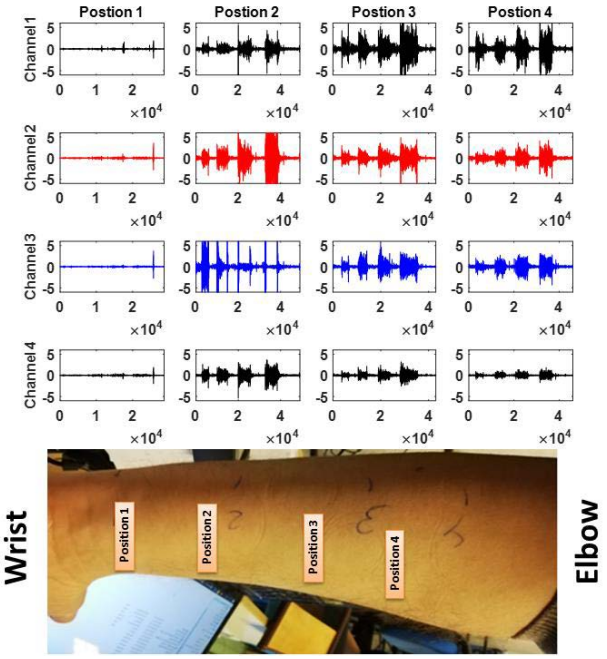

FIGURE 2. The pilot investigation of 4 sensor locations on the forearm (position 1 near the wrist and position 4 near the elbow), with the corresponding AMG signals for 4 channels while the subject performing hand close movement with 4 contraction levels.

The 8 microphone housings were combined to form 3 elastic bands, the first has 3 microphones, the second has 4 microphones and the last has only 1 microphones. The eight AMG microphones were placed on the upper part of the forearm. Fig.3 shows the AMG armband with the 8-microphone housing and the elastic bands. According to the circumference of the upper forearm of the participants, the elastic band was used to adjust and to fix the microphones on the forearm.

E. EXPERIMENTAL PROTOCOL

To acquire the 8 channel AMG, subjects sat on a chair and the forearm was resting on an armrest on the table, after putting the 8-channel on the upper forearm. We utilised the following experimental protocol to record the eight AMG signals from each participant. The subjects were asked to produce transient, moderate force contractions, that lasted 2-4 seconds. We collected AMG signals for 14 movement classes including: Pronation, Supination, Wrist Flexion, Wrist Extension, Radial Deviation, Ulnar Deviation, Hand close, Hand open, Hook grip, Fine pinch, Tripod grip, Index finger flexion, Thumb finger flexion, and No movement (Rest).

At the start of the experiment, there was a trial run to collect the data and to see how the AMG signals changes when trying different wrist/hand and finger movements. LabVIEW

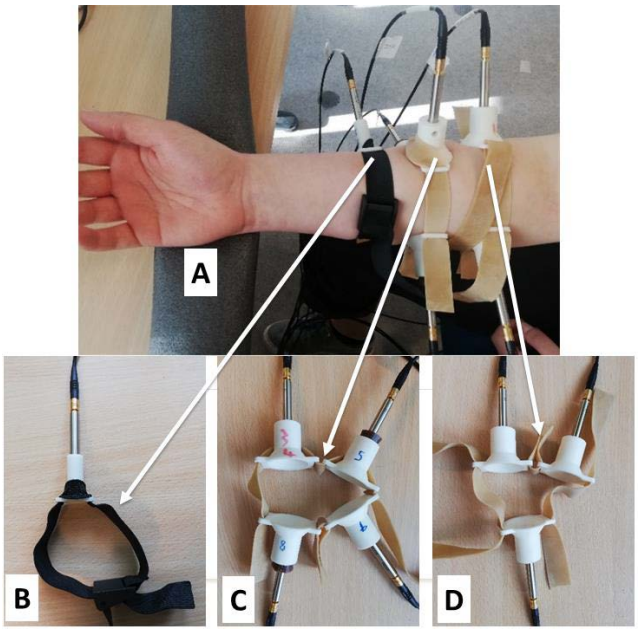

FIGURE 3. The developed AMG armband with the 8 microphones with their custom 3D-Printed housing. A. Example of the location of the 3 wrist bands on the forearm. B. Wrist band3 placed on the transradial level. C. Wrist band2 with 4 microphones and D. Wrist band1 with 3 microphones placed on the upper forearm.

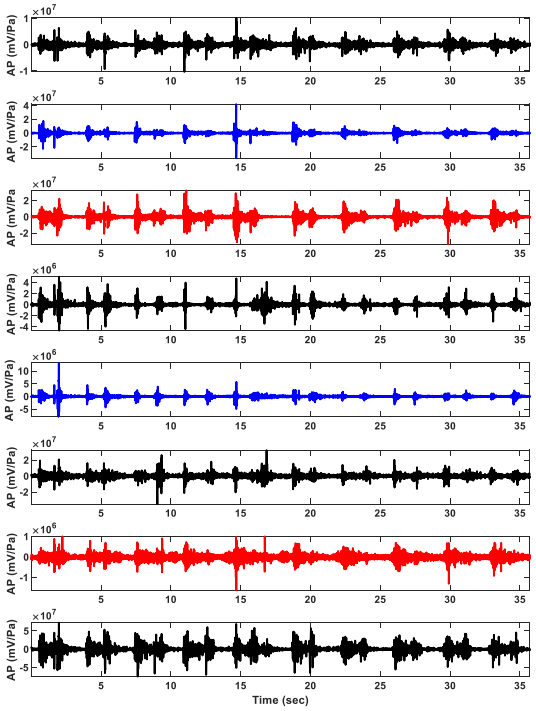

FIGURE 4. Example of the 8 channel AMG for subject 9 while performing a 10 repetitions of hand open movements. AP: Acoustic Pressure.

screen in real-time display was used by the subjects to view the 8 channels AMG to assist them to reproduce the needed movement. In Fig. 4, 8 channel AMG are shown for subject 9 while performing a sequence of ten hand open movements. For each movement, subjects produced 20 repetitions with rest periods of 2-3 seconds between the repetitions. It is noteworthy to mention that the AMG collection was done in a relatively quiet environment where the background noise was measured, and it was equal to 35 dB. The full experimental setup is illustrated in Fig. 5.

To preprocess the data for the analysis, each trial was extracted, and all 20 trials were concatenated in one file which will be used later for further analysis. The LAUM AMG datasets for all subjects are open access for research community from the third author's website[2].

Five-fold cross validation was used for testing the PR systems below with AMG signals, which will be presented in the next section. Fig. 6 shows a radar plot of the mean absolute value of the AMG signals of all AMG channels for female and male subjects, for all movements investigated in this study.

F. PATTERN RECOGNITION ANALYSIS METHODS

An overlapped windowing scheme was utilized during the WST feature extraction process. In the experiments, both the window sizes and windows increments are varied to understand the impact of the windows sizes on the achieved classification results [37]. To compare the performance of the WST feature extraction process with several existing

[2]https://www.rami-khushaba.com/

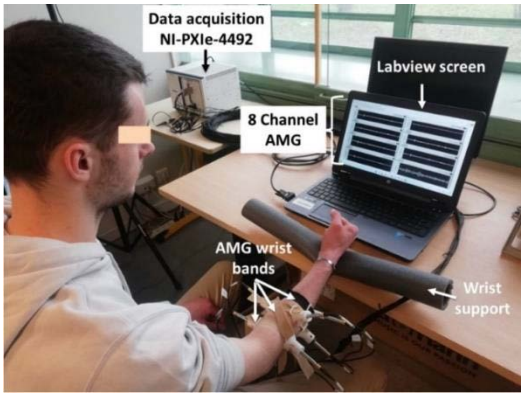

FIGURE 5. The experimental setup of the AMG signal acquisition system with a subject performing a series of hand close movements.

and notable EMG feature extraction methods reported in the literature, we decide to include the following methods in our comparison analysis to provide a fair comparison for these methods on the AMG signals, including:

- **Time Domian-Autoregressive (TDAR6)** [38]: including the combination of Mean absolute value (MAV), Waveform length (WL), number of zero crossings (ZC), number of slop sign changes (SSC), plus a 6th order AR model,
- **AR-RMS** [39]: made of a combination of the root mean square (RMS) and the 6th-order AR model parameters;
- **LSF9**: originally defined in [40], this set is made up of ZC, RMS, L-scale, Mean Value of the Square Root,

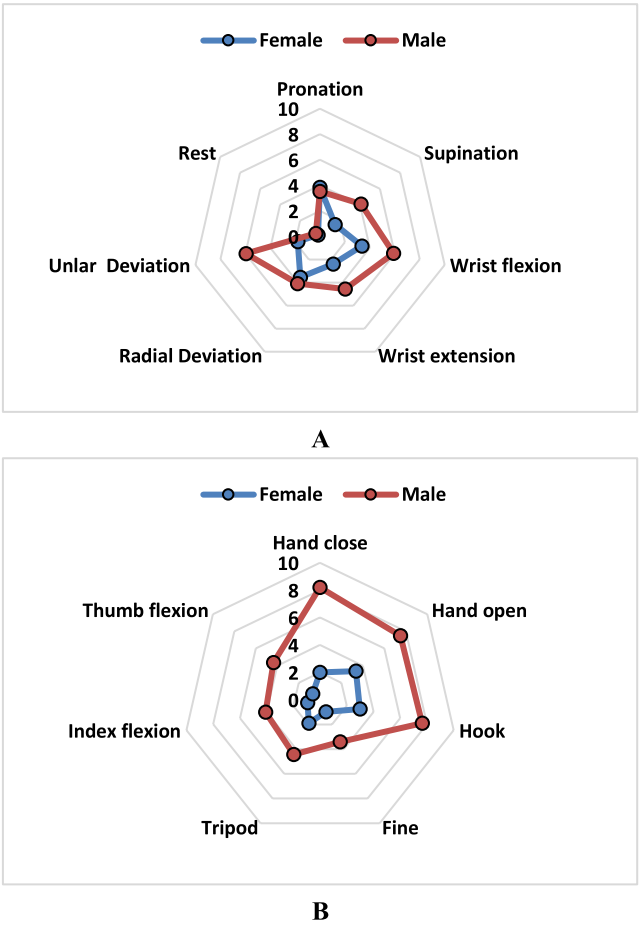

FIGURE 6. Radar plot of the mean absolute value of the AMG signals for 2 sets of movements investigated (A. wrist and B. hand movements) for male and female subjects.

Maximum Fractal Length, Willison Amplitude, Integrated Absolute Value, Variance, and Difference Absolute Standard Deviation Value;

- **ATD**: Combined TD and AR as defined in [41], made up of WL, MAV, Log-Variance (LogVar) and 4th-order AR model parameters,
- Mel-Frequency Cepstral Coefficients (**MFCC**) [42], features of the MFCCC usually utilized in audio FE.
- Spatio-temporal features **(STFS)** from [43], made up of Normalized Root-Square Coefficient of 1st and 2nd differential derivatives, Integral Square Descriptor, an estimate of Mean Derivative of the higher order moments per sliding window, Mean Log Kernel, and a measure of Spatial Muscle Information.
- Fusion of Time domain descriptors (**fTDD**) [44]: The fusion of six time-domain features extracted from each analysis window with the same features extracted from a previous window. The steps parameter of this method was empirically chosen as 5.
- Wavelet Transform (**WT**) and Wavelet-Packet Transform (**WPT**) features [45]: These included the energy of the wavelet coefficients from both transforms. The selection of the wavelet family and the decomposition levels were optimized across the AMG data showing the best performance to be achieved with Symmlet (order 8) family of wavelets across 5 decomposition levels.

As the various feature extraction methods provided different numbers of features, a dimensionality reduction method was utilized to account for different feature set sizes and ensure a fairer comparison. Specifically, the Spectral Regression (SR) feature projection method [46] was used to reduce the dimensionality of all feature sets to n-1, with n being the number of classes in the corresponding dataset.

In terms of the classification models, the following traditional classifiers were evaluated, given their common use in the literature: Linear Discriminant Analysis (LDA), Extreme Learning Machine (ELM), and Quadratic Discriminant Analysis (QDA). The details of these traditional classifiers can be found in any pattern recognition reference and were hence omitted from this paper. In order to compare our classification results to those obtained by others, and verify the statistical significance of the difference, the Wilcoxon signed rank test was applied, with the results being considered significant for a p-value < 0.05.

The final part of the experiments involved investigating the effect of number of AMG channels on the classification performance and to find the optimal number of channels with the WST and the best classifier. The channel selection process was performed using the Differential Evolution Feature Selection (DEFS) methods by Khushaba et al. [47]. This is a population-based channel selection method employing a customized version of DE method (a variant of genetic algorithms). Statistical significance for the optimal number of channels, was also tested using Wilcoxon signed rank test.

III. RESULTS

In the first part of the experiments, we varied the analysis windows size and increments to verify the usefulness of using larger windows sizes for the classification of AMG signals. The analysis here included using windows sizes of 128 ms, 192 ms and 250 ms worth of data with increments of 32 ms, 64 ms and 128 ms, respectively. As expected, the results show a decreasing error trend with larger windows, with the best performances achieved at a windows size of 250 ms with error rates of 9.48%. It is important to mention here that while the literature defines the acceptable error levels for an EMG controlled system to be $\leq$10% [23], there has been no similarly defined levels of acceptable errors when using the AMG signals, as this is a relatively new area. However, to accommodate for the deep nature of the WST and its required computational resources for online tests, we fixed the remaining experiments at a windows size of 192 ms. This would allow further time for feature processing and classification for a real-time system to achieve an overall delay of less than 300 ms defined in the literature as the maximum allowed delay for prosthesis control [37]. Recent research has further demonstrated that windows sizes of up

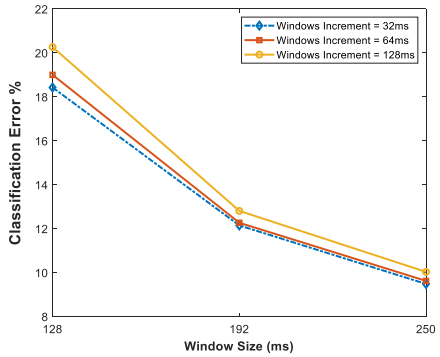

FIGURE 7. Average classification error rates across all subjects using different windows sizes/increments with the wavelet scattering transform. Sampling frequency set to 1024, invariance scale of 0.12, and quality factors of [41].

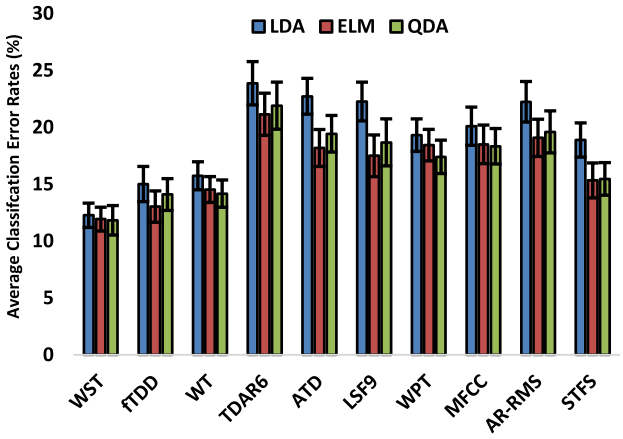

FIGURE 8. Average classification error rates of all utilized feature extraction and classification methods.

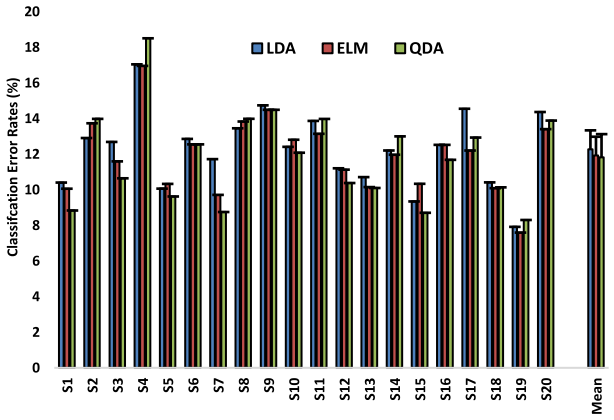

FIGURE 9. Subject performance of the 20 subjects with QDA and WST.

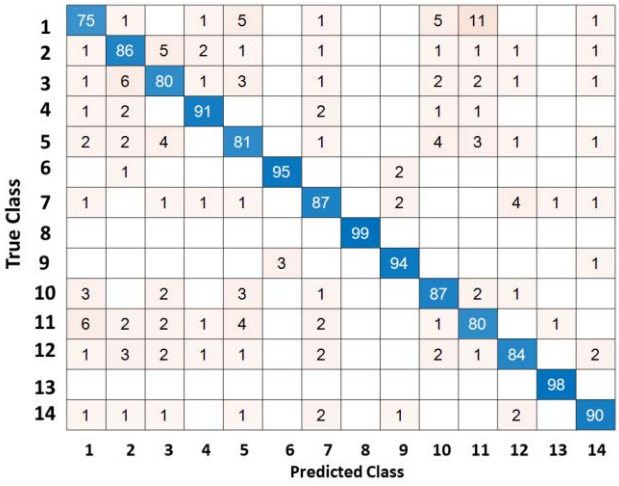

FIGURE 10. Average Confusion matrix with WST FE and QDA classifier. Movements are: 1) Pronation, 2) Supination, 3) Wrist Flexion, 4) Wrist Extension, 5) Radial Deviation, 6) Ulnar Deviation, 7) Hand close, 8) Hand open, 9) Hook grip, 10) Fine pinch, 11) Tripod grip, 12) Index finger flexion, 13) Thumb finger flexion, and 14) No movement (Rest).

to 750 ms [48], which further allows our system to produce more accurate results.

For the next part of the experiments, using an analysis windows size of 192 ms with 64 ms increments, we benchmarked the performances of all the feature extraction methods in comparison to the WST across three different classifiers, as shown in Fig.8. In this case, the WST outperformed all other feature extraction methods with an average classification error of 12.26%, 11.92%, and 11.82% across the LDA, ELM and QDA classifiers respectively. In comparison to the other methods, the Wilcoxon signed rank test showed that the WST performance was statistically significantly different than all other methods with $p < 0.001$, except the fTDD when using the ELM classifier with $p = 0.179$. On the other hand, the performances of the fTDD and WT methods were not statistically significantly different from each other across the LDA and QDA classifiers ($p > 0.05$), but significantly different across the ELM classifier ($p = 0.013$), and only marginally different from STFS across the QDA classifier ($p = 0.033$). It is also interesting to note here that the WT outperformed WPT across all classifiers ($p < 0.001$) despite the fact that WPT further generalizes upon the WT by decomposing the low and high frequencies. This is in turn driven by the nature of the AMG signal and the concertation of the useful energy of this signal within the low frequency side of the spectrum on which WT is focused, while WPT is further focused on decomposing the noise from the high frequency part of the spectrum.

In Fig.9, The classification error rates for 20 subjects are shown when using WST FE and three classifiers, i.e, LDA, ELM and QDA. The mean and standard deviation of all subjects are also shown. S19 was the best performing subject with an error rate of 8% while S4 was the subject with the highest error of approximately 17%. Fig.10 illustrated the average confusion matrix of 20 subjects with WSR FE and QDA classifier for 14 movement classes.

For the final part of the experiment, we fixed the feature extraction method on WST and the classification method on QDA and analyzed the required number of channels to solve the current classification problem. By varying the number

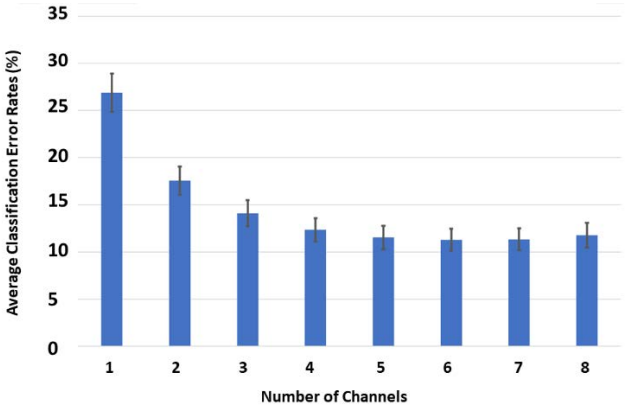

FIGURE 11. The achieved classification error rates averaged across all 20 subjects while implementing AMG channel selection.

of selected channels, the classification error rates were analyzed and the average across all subjects is shown in Fig.11. The channel selection results were achieved by using DEFS methods [47] as mentioned in the methodology section. These results indicated that the classification error rates gradually decreased by considering more channels in this problem. However, the decrease in the average classification error rates did not continue after 6 channels, i.e., the average classification errors were slightly higher when using 7 and 8 channels respectively.

IV. DISCUSSION

We proposed a novel feature extraction method based on WST while investigating AMG as a potential control signal with pattern recognition system for recognizing hand gestures. Increasing the window size of the AMG signal from 128 ms to 250 ms reduced the error rate from 18% till 10% (Fig.7); this can be attributed to the fact that increasing the amount of information gathered when increasing the windows size.

When comparing the proposed novel WST feature extraction method, it outperformed the traditional methods with all classifiers investigated, i.e. LDA, ELM and QDA (Fig.8) including time domain features: TDAR, AR-RMS, and fTDD; frequency domain features: STFT; and time frequency features: WT and WPT. In addition, it outperformed other methods MFCC and LSF9. As for the performance of the three classifiers investigated, ELM and ODA outperformed LDA classifier (Fig. 8).

The classification performance for all 20 subjects with WST and QDA (best feature extraction/classifier combination) was illustrated (Fig.9) with the range of classification error from 8% to 20% with an average of 11%. When using WST with QDA classifier, the confusion matrix in Fig.10 showed that the most errors was in movement 1 while for other movements the performance was within 80% and 90% accurate. The performance of movement 1 (Pronation) can be further improved by acquiring more AMG data from the subjects to train the classifier. It should be noted that our studies validated the PR based on WST feature extraction of AMG collected from 20 subjects, larger than previous [9], [10], [12], [16], [20] while achieving performance of 88% for 14 gestures with only single modality without the use of other modality such as MMG [16], [22] or FMG [21].

In our study, we chose the channel location in the upper part of the forearm after performing the pilot study (Fig.2). Unlike the work in [12] and [22], where the use of sensor placement locations near the wrist was recommended, we found the opposite of upper forearm part for the current set of movements considered in this research.

We originally recorded 8 AMG channel for each subject and run an experiment with DEFS [47] to investigate the effect of channel number on the classification performance. We found that no statistically significant differences were observed between the results when using 6 and 7 channels (average error of 11.28% with 6 channels and 11.34 % with 7 channels, $p = 0.390$), the analysis indicated statistically significant differences ($p < 0.001$) between the results achieved with 6 vs. 8 channels. This in turn indicated that 6 channels are sufficient to solve this problem, after which adding more channels can further increase the average error rates. This can be attributed in part to the effect of the muscles crosstalk contamination which could impact the quality of the AMG signals. On the other hand, when looking at the indices of the most frequently utilized channels when selecting a subset of 6 channels only, it was found that sensor locations 6 and 8 were the least selected, while all other channels were necessary to achieve the optimum performance in this experiment.

It should be noted that only four different positions of the upper forearm were compared in a pilot test, and the signals may be correlated to some extent, when investigating the selection of the best position of the sensors in this study. Anatomically, the distribution of the different muscles of the forearm may be analyzed, and signals with low correlation may be obtained. This matter can be further investigated in a future study, with a large number of participants.

The study has a limitation that only offline experiments have been conducted to validate the PR system with AMG signals. Also, despite the large number of 20 intact-limbed participants who were recruited, it did not include transradial amputees. Future research will include testing the proposed methods in this study on amputees and performing real-time experiments.

V. CONCLUSION

In this paper, the AMG signals from the subjects' forearm have been proposed as a source of control for hand gesture recognition with PR. A novel feature extraction for AMG based on WST was proposed and validated on 14 movement classes acquired from a large dataset of 20 subjects. The proposed novel FE outperformed all the state-of-the-art achieving an accuracy of 88% for 14 gestures with QDA classifier. In addition, reducing the channel number to 6 channels showed statically significant differences in the average classification error results than those of the full 8 channels

suggesting that 6 channels were the optimum set of channels to resolve the current problem. The outcomes of this study show the potential of using AMG for the control of upper limb prostheses with PR.

ACKNOWLEDGMENT

The author Ali H. Al-Timemy would like to thank the discretionary membership at Plymouth University, U.K. The authors are grateful for the study participants and also for Boulila Atef, Matthieu Milharo, and Julien Nicolas for their help on the microphone support.

REFERENCES

[1] R. N. Khushaba, E. Scheme, A. H. Al-Timemy, A. Phinyomark, A. Al-Taee, and A. Al-Jumaily, "A long short-term recurrent spatial–temporal fusion for myoelectric pattern recognition," *Expert Syst. Appl.*, vol. 178, Sep. 2021, Art. no. 114977.

[2] A. H. Al-Timemy, G. Bugmann, J. Escudero, and N. Outram, "Classification of finger movements for the dexterous hand prosthesis control with surface electromyography," *IEEE J. Biomed. Health Informat.*, vol. 17, no. 3, pp. 608–618, May 2013.

[3] D. Farina, "The extraction of neural information from the surface EMG for the control of upper-limb prostheses: Emerging avenues and challenges," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 22, no. 4, pp. 797–809, Jul. 2014.

[4] A. H. Al-Timemy, R. N. Khushaba, G. Bugmann, and J. Escudero, "Improving the performance against force variation of EMG controlled multifunctional upper-limb prostheses for transradial amputees," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 6, pp. 650–661, Jun. 2016.

[5] R. N. Khushaba, A. Al-Timemy, S. Kodagoda, and K. Nazarpour, "Combined influence of forearm orientation and muscular contraction on EMG pattern recognition," *Expert Syst. Appl.*, vol. 61, pp. 154–161, Nov. 2016.

[6] T. Bao, S. Q. Xie, P. Yang, P. Zhou, and Z.-Q. Zhang, "Toward robust, adaptiveand reliable upper-limb motion estimation using machine learning and deep learning—A survey in myoelectric control," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 8, pp. 3822–3835, Aug. 2022.

[7] R. S. Chegani and C. Menon, "Pilot study on fine finger movement regression, using FMG," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Oct. 2017, pp. 1069–1074.

[8] C. S. M. Castillo, S. Wilson, R. Vaidyanathan, and S. F. Atashzar, "Wearable MMG-plus-one armband: Evaluation of normal force on mechanomyography (MMG) to enhance human-machine interfacing," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 196–205, 2021.

[9] S. Jiang, B. Lv, W. Guo, C. Zhang, H. Wang, X. Sheng, and P. B. Shull, "Feasibility of wrist-worn, real-time hand, and surface gesture recognition via sEMG and IMU sensing," *IEEE Trans. Ind. Informat.*, vol. 14, no. 8, pp. 3376–3385, Aug. 2018.

[10] A. A. Y. Buk, M. K. Wali, A. H. Al-Timemy, and K. Raoof, "Hand gesture recognition using mechanomyography signal based on LDA classifier," *IOP Conf. Ser., Mater. Sci. Eng.*, vol. 881, no. 1, 2020, Art. no. 12125.

[11] D. T. Barry, J. A. Leonard, and A. J. Gitter, "Acoustic myography as a control signal for an externally powered prosthesis," *Arch. Phys. Med. Rehabil.*, vol. 67, pp. 267–269, Apr. 1986.

[12] N. Siddiqui and R. H. M. Chan, "Hand gesture recognition using multiple acoustic measurements at wrist," *IEEE Trans. Human-Mach. Syst.*, vol. 51, no. 1, pp. 56–62, Feb. 2021.

[13] A. K. Tanim, K. M. T. Nahiyan, and M. A. R. Ahad, "Suitability of single-channel acoustic myography for classification of individual finger movements," in *Proc. Joint 9th Int. Conf. Informat., Electron. Vis. (ICIEV) 4th Int. Conf. Imag., Vis. Pattern Recognit. (icIVPR)*, Aug. 2020, pp. 1–8.

[14] C. Orizio, R. Perini, and A. Veicsteinas, "Muscular sound and force relationship during isometric contraction in man," *Eur. J. Appl. Physiol. Occupational Physiol.*, vol. 58, no. 5, pp. 528–533, Mar. 1989.

[15] T. W. Beck, T. J. Housh, J. T. Cramer, J. P. Weir, G. O. Johnson, J. W. Coburn, M. H. Malek, and M. Mielke, "Mechanomyographic amplitude and frequency responses during dynamic muscle actions: A comprehensive review," *Biomed. Eng. OnLine*, vol. 4, no. 1, pp. 1–27, Dec. 2005.

[16] W. Guo, X. Sheng, H. Liu, and X. Zhu, "Mechanomyography assisted myoeletric sensing for upper-extremity prostheses: A hybrid approach," *IEEE Sensors J.*, vol. 17, no. 10, pp. 3100–3108, May 2017.

[17] A. P. Harrison, B. Danneskiold-Samsøe, and E. M. Bartels, "Portable acoustic myography—A realistic noninvasive method for assessment of muscle activity and coordination in human subjects in most home and sports settings," Physiol. Rep., 2013, vol. 1, no. 2.

[18] A. Phinyomark, F. Quaine, S. Charbonnier, C. Serviere, F. Tarpin-Bernard, and Y. Laurillau, "EMG feature evaluation for improving myoelectric pattern recognition robustness," *Expert Syst. Appl.*, vol. 40, no. 12, pp. 4832–4840, Sep. 2013.

[19] H. Kato and K. Takemura, "Hand pose estimation based on active bone-conducted sound sensing," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput., Adjunct*, Sep. 2016, pp. 109–112.

[20] A. S. Asheghabadi, S. B. Moqadam, and J. Xu, "Multichannel finger pattern recognition using single-site mechanomyography," *IEEE Sensors J.*, vol. 21, no. 6, pp. 8184–8193, Mar. 2021.

[21] S. B. Moqadam, A. S. Asheghabadi, and J. Xu, "A novel hybrid approach to pattern recognition of finger movements and grasping gestures in upper limb amputees," *IEEE Sensors J.*, vol. 22, no. 3, pp. 2591–2602, Feb. 2022.

[22] N. Siddiqui and R. H. M. Chan, "Multimodal hand gesture recognition using single IMU and acoustic measurements at wrist," *PLoS ONE*, vol. 15, no. 1, Jan. 2020, Art. no. e0227039.

[23] E. J. Scheme and K. Englehart, "Electromyogram pattern recognition for control of powered upper-limb prostheses: State of the art and challenges for clinical use," *J. Rehabil. Res. Develop.*, vol. 48, no. 6, pp. 643–659, 2011.

[24] J. Andén and S. Mallat, "Deep scattering spectrum," *IEEE Trans. Signal Process.*, vol. 62, no. 16, pp. 4114–4128, Aug. 2014.

[25] S. Mallat, "Group invariant scattering," *Commun. Pure Appl. Math.*, vol. 65, no. 10, pp. 1331–1398, 2012.

[26] V. Lostanlen and S. Mallat, "Wavelet scattering on the pitch spiral," 2016, *arXiv:1601.00287*.

[27] W. Ghezaiel, L. Brun, and O. Lezoray, "Wavelet scattering transform and CNN for closed set speaker identification," in *Proc. IEEE 22nd Int. Workshop Multimedia Signal Process. (MMSP)*, Sep. 2020, pp. 1–6.

[28] B. Soro and C. Lee, "A wavelet scattering feature extraction approach for deep neural network based indoor fingerprinting localization," *Sensors*, vol. 19, no. 8, p. 1790, Apr. 2019.

[29] J. Bruna and S. Mallat, "Invariant scattering convolution networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1872–1886, Aug. 2013.

[30] J. Bruna and S. Mallat, "Classification with invariant scattering representations," in *Proc. IEEE 10th IVMSP Workshop, Perception Vis. Signal Anal.*, Jun. 2011, pp. 99–104.

[31] J. Andén and S. Mallat, "Multiscale scattering for audio classification," in *Proc. ISMIR*, 2011, pp. 657–662.

[32] L. Sifre and S. Mallat, "Rotation, scaling and deformation invariant scattering for texture discrimination," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 1233–1240.

[33] W. M. Association, "World medical association declaration of Helsinki: Ethical principles for medical research involving human subjects," *J. Amer. Med. Assoc.*, vol. 310, no. 20, pp. 2191–2194, Nov. 2013.

[34] M. Y.-E. Ma, *MMG Sensor for Muscle Activity Detection–Low Cost Design, Implementation and Experimentation Masters of Engineering Mechatronics*. Palmerston North, New Zealand: Massey Univ., 2009.

[35] I. Masmoudi, A. H. Al-Timemy, S. E. Bouzid, S. Yacoub, and K. Raoof, "AMG signals acquisition method," in *Proc. 4th Int. Conf. Adv. Syst. Emergent Technol. (IC_ASET)*, Dec. 2020, pp. 284–288.

[36] S. Yacoub, A. H. Al-Timemy, Y. Serrestou, and K. Raoof, "Hand movements analysis with acoustic myography signals," in *Proc. 5th Int. Conf. Adv. Syst. Emergent Technol. (IC_ASET)*, Mar. 2022, pp. 228–232.

[37] L. H. Smith, L. J. Hargrove, B. A. Lock, and T. A. Kuiken, "Determining the optimal window length for pattern recognition-based myoelectric control: Balancing the competing effects of classification error and controller delay," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 19, no. 2, pp. 186–192, Apr. 2011.

[38] M. G. Asogbon, O. W. Samuel, Y. Jiang, L. Wang, Y. Geng, A. K. Sangaiah, S. Chen, P. Fang, and G. Li, "Appropriate feature set and window parameters selection for efficient motion intent characterization towards intelligently smart EMG-PR system," *Symmetry*, vol. 12, no. 10, p. 1710, Oct. 2020.

[39] R. N. Khushaba, A. Phinyomark, A. H. Al-Timemy, and E. Scheme, "Recursive multi-signal temporal fusions with attention mechanism improves EMG feature extraction," *IEEE Trans. Artif. Intell.*, vol. 1, no. 2, pp. 139–150, Oct. 2020.

[40] A. Phinyomark, P. Phukpattaranont, and C. Limsakul, "Feature reduction and selection for EMG signal classification," *Expert Syst. Appl.*, vol. 39, pp. 7420–7431, Jun. 2012.
[41] A. Krasoulis, I. Kyranou, M. S. Erden, K. Nazarpour, and S. Vijayakumar, "Improved prosthetic hand control with concurrent use of myoelectric and inertial measurements," *J. Neuroeng. Rehabil.*, vol. 14, no. 1, pp. 1–14, 2017.
[42] S. Wilson and R. Vaidyanathan, "Gesture recognition through classification of acoustic muscle sensing for prosthetic control," in *Proc. Conf. Biomimetic Biohybrid Syst.*, M. Mangan, M. Cutkosky, A. Mura, P. F. M. J. Verschure, T. Prescott, N. Lepora, Eds. Stanford, CA, USA: Springer, Jul. 2017, pp. 637–642.
[43] O. W. Samuel, M. G. Asogbon, Y. Geng, X. Li, S. Pirbhulal, S. Chen, N. Ganesh, P. Feng, and G. Li, "Spatio-temporal based descriptor for limb movement-intent characterization in EMG-pattern recognition system," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2019, pp. 2637–2640.
[44] R. N. Khushaba, A. Al-Ani, A. Al-Timemy, and A. Al-Jumaily, "A fusion of time-domain descriptors for improved myoelectric hand control," in *Proc. IEEE Symp. Ser. Comput. Intell. (SSCI)*, Dec. 2016, pp. 1–6.
[45] R. N. Khushaba, S. Kodagoda, S. Lal, and G. Dissanayake, "Driver drowsiness classification using fuzzy wavelet-packet-based feature-extraction algorithm," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 1, pp. 121–131, Jan. 2011.
[46] D. Cai, X. He, and J. Han, "SRDA: An efficient algorithm for large-scale discriminant analysis," *IEEE Trans. Knowl. Data Eng.*, vol. 20, no. 1, pp. 1–12, Jan. 2007.
[47] R. N. Khushaba, A. Al-Ani, and A. Al-Jumaily, "Feature subset selection using differential evolution and a statistical repair mechanism," *Expert Syst. Appl.*, vol. 38, no. 9, pp. 11515–11526, 2011.
[48] M. Dyson, S. Dupan, H. Jones, and K. Nazarpour, "Learning, generalization, and scalability of abstract myoelectric control," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 7, pp. 1539–1547, Jul. 2020.

ALI H. AL-TIMEMY received the B.Sc. and M.Sc. degrees in biomedical engineering from Nahrain University, Iraq, and the Ph.D. degree from the Centre for Robotics and Neural Systems (CRNS), University of Plymouth, U.K.

He is currently an Associate Professor at the Biomedical Engineering Department, Al-Khwarizmi College of Engineering, University of Baghdad, Iraq. His research interests include biomedical signal and image processing looking at improving the EMG-based pattern recognition control of upper limb prostheses and also using of artificial intelligence to detect ophthalmic diseases. He was a recipient of the Fulbright Visiting Scholarship at the University of Delaware in 2018 and an ARVO Collaborative Research Award in 2022.

YOUSSEF SERRESTOU received the Engineering, M.Sc., Aggregation in mathematics, and Ph.D. degrees from the Grenoble-INP, in 2003, 2006, and 2008, respectively. Since 2015, he has been with the Group of Research on Signal Processing and Instrumentation, LAUM Laboratory, Le Mans, France. He is currently an Associate Professor of mathematics with the ENSIM College of Engineering, Le Mans University (LAUM), and an Associate Researcher at the Acoustic Laboratory, LAUM.

RAMI N. KHUSHABA (Senior Member, IEEE) received the Ph.D. degree from the University of Technology, Sydney (UTS), in 2010, with a focus on myoelectric signal processing. He is currently acting as a Senior Planning and Modeling Scientist with Transport for NSW. His research interests include myoelectric control, machine and deep learning theory and applications, and vision, LiDAR, and radar signal processing.

SLIM YACOUB is currently a Professor at the INSAT, Tunisia, and a Researcher at the LTSIRS Laboratory. His research interests include signal processing in the case of biomedical signals, particularly ECG, EMG, EEG, and AMG signals.

KOSAI RAOOF received the M.Sc. and Ph.D. degrees from Grenoble University, in 1990 and 1993, respectively, and the Habilitation à Diriger des Recherches (HDR) degree in 1998. He was invited to join the Laboratoire des Images et Signaux (LIS), in 1999, to participate in the founding of the Telecommunication Research Group. He is currently the Head of Real Time Embedded Systems with the ASTRE Department, ENSIM College of Engineering, Le Mans University, Le Mans, France. His research interest was first focalized on advanced MIMO systems and joint CDMA synchronization; he studied and introduced polarized diversity MIMO systems in the research group. In 2007, he joined the GIPSA-Laboratory to continue his research on MIMO antenna selection systems. Since 2011, he has been a Full Professor at the University of Maine, and established a new group of research on signal processing and instrumentation at the LAUM Laboratory. He supervised more than 20 Ph.D. and M.Sc. students in different fields of applied signal processing and telecommunications. His research interests include localization for smart sensor networks and distributed agents for intelligent spaces. He is an editor of two books and the Chief Editor of *Wireless Sensor Networks* journal first published, in 2008.