

Wrist and Finger Gesture Recognition with Single-element Ultrasound Signals: A Comparison with Single-channel Surface Electromyogram

Jiayuan He, Henry Luo, Jie Jia, John T.W. Yeow*, *Senior Member, IEEE*, and Ning Jiang*, *Senior Member, IEEE*

Abstract—With the ability to detect volumetric changes of contracting muscles, ultrasound (US) was a potential technique in the field of human-machine interface (HMI). Compared to the US imaging (B-mode US), the signal from a static single-element US transducer, A-mode US, was a more cost-effective and convenient way towards the real-world application, particularly the wearables. This study compared the performance of the single-channel A-mode US with single-channel surface electromyogram (sEMG) signals, one of the most popular signal modalities for wrist and finger gesture recognition. We demonstrated that A-mode US outperformed sEMG in six out of nine gestures recognition, while sEMG was superior to A-mode US on the detection of the Rest state. We also demonstrated that, through feature space analysis, the advantage of A-mode US over sEMG for gesture recognition was due to its superior ability in detecting information from deep musculature. This study presented the clear complementary advantages between A-mode US and sEMG, indicating the possibility of fusing two signal modalities for the gesture recognition applications.

Index Terms—Human machine interface (HMI), pattern recognition, surface electromyogram (sEMG), ultrasound signal

I. INTRODUCTION

HUMAN-machine interface (HMI) has been a research focus for a wide range of applications in biomedicine, manufacturing and aerospace [1]. A specific type of HMI enables users to interact with mechanical or electronic systems by recognizing the intended gestures through the measurement of muscle activities. This technology is promising in the field of rehabilitation engineering for providing intuitive control of advanced powered upper-limb prostheses for amputees [2],

interpreting sign language to facilitate communication for the deaf and mute [3], assisting stroke patients with powered exoskeleton control [4], and much more [5][6]. In addition, it could be paired with virtual or augmented reality commercial devices to realize effortless and intuitive interaction between human and artificial entities, such as computers, robotics, and virtual reality objects [7].

Surface electromyogram (sEMG) is the manifestation of the electrical activities from muscle contractions and has been widely used as a source signal for hand and wrist gesture recognition. With advanced signal processing algorithms, promising results have been achieved in sEMG-based gesture recognition [8]. However, this technique is still facing many challenges in the practical application, due to its inherent limitations [2][9]. One of these limitations is that the activities of deep and small muscles are difficult to be detected accurately with the electrodes at the skin surface. Consequently, sEMG is not reliable in recognizing fine finger gestures and movements [10], such as fine pinch, thumb extension, which are mostly controlled by deep and fine muscles of the forearm.

Ultrasound (US) refers to the sound waves with the frequency over 20 kHz. When the ultrasound pulses are emitted into human tissues, it would propagate, as well as reflect at boundaries where differences of US conductance exist [11]. Such reflections contain structure information of the tissues under investigation. As such, the volumetric changes of the contracting skeletal muscles, both superficial and deep ones, can be extracted from the US echoes. US modality is a potential source signal for wrist and hand gesture recognition. There are several modes of US, and the most common one in gesture recognition studies is B-mode US (US imaging), which presents the human tissue state in two-dimensional images. US imaging has been reported in finger [12][13] and wrist [14] gesture recognition, individual finger joint angle prediction [12][15][16], and fingertip force prediction [17]. It was presented that the performance of US imaging was better than that of sEMG in finger gesture classification and metacarpophalangeal joint angle prediction [12]. However, the condition of the US imaging generation is critical. To produce US images, the acoustic beam must be moved in a prescribed direction, either mechanically or electronically, for the scan of

This work was supported by the Natural Sciences and Engineering Research Council of Canada (Discovery Grant 072169).

J He, J. T. W. Yeow and N Jiang are with Department of Systems Design Engineering, Faculty of Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada (E-mail: ning.jiang@uwaterloo.ca, jyeow@uwaterloo.ca).

H Luo is with Uniron Hearing Ltd, Kitchener, ON N2R 1S1, Canada.

J. Jia is with Department of Rehabilitation Medicine, Huashan Hospital, Fudan University, Shanghai 200040, China.

the defined image plane. This makes the corresponding probe complicated and bulky, significantly limiting its potential wearable applications where compact physical size and low-energy consumption is required.

An alternative mode of US is A-mode, where only a single static transducer is used, scanning a line through the body. Compared to B-mode US, A-mode US is one-dimensional, containing less information of human tissues. However, A-mode US is easy to generate and process, requiring much less energy to operate. It should be noted here that A-mode US is not simply one column data of B-mode US. Sophisticated algorithms are employed to integrate information from multiple US transducers into one image. Therefore, one column data of B-mode US contains information from multiple transducers, not only one [18]. Previous work reported the effectiveness of A-mode US on predicting the angle of wrist extension with one single transducer (one channel) [19], and classifying the flexions of each individual finger with four transducers (four channels) [20]. These studies indicated the potential of A-mode US in the application of HMI. However, the performances of A-mode US and sEMG in gesture recognition have not been compared systematically. Further, although it is claimed that the information from deep muscles is the main advantage of US signal over sEMG in gesture recognition, no further analysis is presented to support the point except the classification performance.

The goal of this study is to compare the performance of A-mode US with that of sEMG on the classification of the hand and wrist gestures. For both signal modalities, only one channel data were used. The sensors were placed as close as possible to ensure the comparison was as realistic as possible. Further, the role of the depth information for gesture recognition was illustrated through feature space analysis. This study presented the characteristics of A-mode US on gesture recognition, demonstrating the advantage of A-mode US over sEMG, and the importance of the fusion of two signal modalities.

II. METHOD

A. Subjects

Ten able-bodied subjects (nine males and one female, aged from 18 to 32 years old) took part in this study. The forearm circumference of the interest site (position of the US probe) ranged from 23 to 29 cm, and the average was approximately 26 cm. Before the participation, all subjects read and signed the

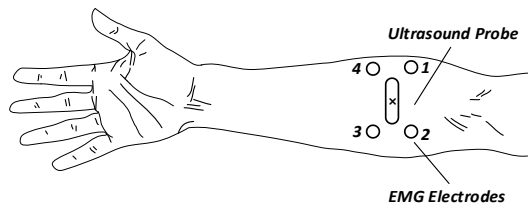


Fig. 1. The placement of the US probe and sEMG electrodes on the forearm. The US probe is around the compartments of *flexor digitorum superficialis* (FDS) and *flexor carpi ulnaris* (FCU), and the four monopolar electrodes are located on the four corners of the probe. Only the center element of the probe, labeled as the cross, is used in this study.

TABLE I
sEMG CHANNEL FORMATION FOR PERFORMANCE COMPARISON

Symbol	Description
<i>Monopolar 1 (M1)</i>	Signal from electrode 1.
<i>Monopolar 2 (M2)</i>	Signal from electrode 2.
<i>Monopolar 3 (M3)</i>	Signal from electrode 3.
<i>Monopolar 4 (M4)</i>	Signal from electrode 4
<i>Monopolar Average (MA)</i>	Signal averaged across four electrodes
<i>Bipolar 1 (B1)</i>	Signal from the differential between electrode 1 and 4
<i>Bipolar 2 (B2)</i>	Signal from the differential between electrode 2 and 3

Note: the electrode sequence is displayed in Fig. 1.

informed consent. The experiment protocol was in accordance with the Declaration of Helsinki, and approved by the local research ethics authority.

B. Experimental Protocol

A commercial US transducer with a 64-element array (P4-2v, Verasonics Inc.) was placed on the forearm, around the compartments of *flexor digitorum superficialis* (FDS) and *flexor carpi ulnaris* (FCU), which were mainly responsible for the flexion of the joints. The center frequency of the transducer was 2.976 MHz. In each echo epoch (when the transducer was receiving US echo), the echoed US signal was sampled at the frequency of four times of the central frequency by a commercial processing system (Vantage 64 LE, Verasonics Inc.). Four monopolar sEMG electrodes were placed in close vicinity to the US probe (Fig. 1), for the fairness of the comparison. As the existing of cross talk, the ability of information detection from the surrounding areas, the sEMG electrodes were not necessarily positioned on the specific muscles [21][28]. The sEMG signal was amplified and digitized by a commercial myoelectric system (g.USBamp, g.Tec Medical Engineering, Austria), with the sampling frequency of 1200 Hz. As this study aimed to compare the single-channel performance between sEMG and A-mode US,

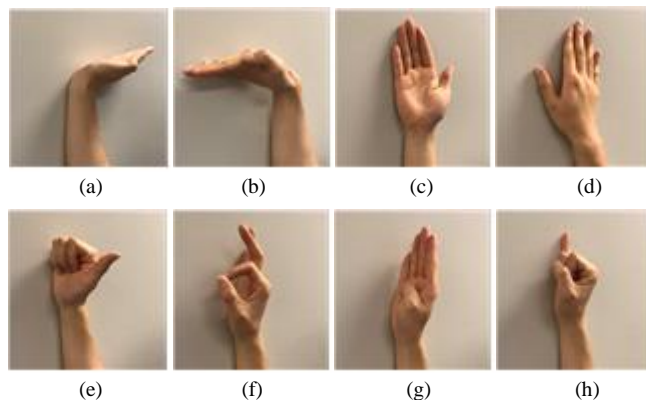


Fig. 2. The wrist and finger gesture investigated in the study: (a) wrist extension (WE), (b) wrist flexion (WF), (c) wrist supination (WS), (d) wrist pronation (WP), (e) thumb extension (TE), (f) fine pinch (FP), (g) four-finger extension (FE), (h) little-finger extension (LE).

only one element (the one in the center of the probe) of the US transducers was activated during the experiment.

Four wrist gestures, four finger gestures, and Rest were investigated in this study (Fig. 2): wrist extension (WE), wrist flexion (WF), wrist supination (WS), wrist pronation (WP), thumb extension (TE), fine pinch (FP), four-finger extension (FE) and little-finger extension (LE). In the experiment, each gesture was repeated six times, and each repetition was 12 s long. The sequence of the gestures was randomized, and there was a 5-s relax period between two consecutive gestures. A schematic diagram for sEMG and US data synchronization is illustrated in Fig. 3a. Due to a limitation of the US instrument, during each 12-s contraction, US signals were collected in five epochs. In each epoch, the transducer repeated the following procedures: sending the sound wave, receiving the echo, and pausing. One US echo frame was obtained in such a repetition. Each epoch lasting around 0.9s, containing 100 US echo frames. A representative frame is illustrated in Fig. 3b. For each contraction, the sEMG data collected during these five epochs were extracted for subsequent analysis. The data acquisition and subsequent processing procedures were performed by a customized software program running on Matlab platform (Mathworks Inc.).

C. US Processing

In each US echo, 1536 data points were recorded. Given the average sound speed in human body was 1540 m/s [18], the maximal inspecting depth was 9.94 cm, enough to cover the forearm muscles and bones of the participants in this study. The first 50 points were removed for their comparative large values, caused by the reflection at the probe/skin interface, as they were consistent across all sessions for all participants. The remaining 1486 data points, representing the echo of US when it traveled through the tissues from 0.32 to 9.94 cm away from the receiver, formed a US data frame for the subsequent analysis. The US data allocated during the five epochs in each contraction were segmented into 200-ms analysis windows, with an increment of 50 ms (150-ms overlap between windows). In each window, the root mean square (RMS) value of each data point was calculated, obtaining a feature vector with 1486 dimensions. Principal component analysis (PCA)

was employed to reduce the feature dimension to 30, avoiding the curse of dimensionality for subsequent classification. The reduced dimension was chosen empirically, mainly considering the tradeoff between the information retained (>82% variance) and the computational load of the classifier [22]. Three classifiers, linear discriminant analysis (LDA), support vector machine (SVM), and artificial neural network (ANN), were adopted to test the effect of the classifier on US signal classification. The kernel of SVM was linear for its comparable performance with other kernels [23]. The number of nodes in hidden layer of ANN was twice as the number in input layer [24]. The classification results were computed through a 5×3 cross-validation, where a 3-fold cross-validation was performed 5 times, and in each time, the training data was the four repetitions of each gesture, and the testing data was the remaining two repetitions.

D. sEMG Processing

Seven sEMG channels were derived from four monopolar electrodes (Table I). Four of them, namely M1 to M4, were derived from each monopolar electrode, respectively. MA was the average of four monopolar electrodes, simulating the sEMG signal from the exact site of the US probe. The remaining two, B1 and B2, was the differential of two monopolar electrodes orientated parallel to the ulnar bone, i.e. electrode 1 minus electrode 4, and electrode 2 minus electrode 3 (Fig. 1). The differential direction was chosen because it had been shown that the parallel differential direction outperformed the perpendicular one [25]. For data from each of the seven derived EMG channels, the period corresponding to each US epoch was first band-pass filtered between 10 and 500 Hz using a third-order Butterworth filter. Then, the signal was windowed with the same setting as the A-mode US (200-ms window with 50-ms overlap). The classic Hudgins' time domain (TD) feature set [26], including absolute value (MAV), zero crossing (ZC), sign slope changes (SSC), wavelet length (WL), were extracted from each window (the dimension of the sEMG feature vector was four) for it was regarded as a benchmark feature in many sEMG studies [8][9][27]. As prior studies indicated that the effect of the classifier was not significant on sEMG signal classification [23][28], LDA was chosen as the classifier for its

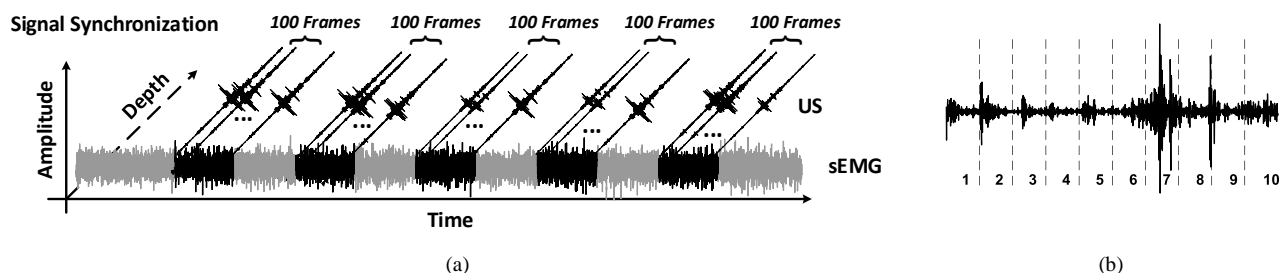


Fig. 3 Schematic diagram for sEMG and US data acquisition: (a) Relationship between the acquired sEMG signal and the corresponding US signal, illustrated with one gesture contraction. The continuously acquired sEMG signal is indicated in the grey trace. Due to the constraints of the US instrument, US data could only be collected from five separate epochs in each gesture contraction, indicated by the black segments, along the direction marked as 'Depth'. For each of such epoch, 100 frames of US echo data are acquired. The length of each epoch represents the depth of the sound reflections from the data receiver (sensor). The black overlay of sEMG is the sEMG data collected during these five US epochs. (b) A representative frame of US echo data from wrist extension. The total sample points in (b) is 1486, with the removal of the first 50 points, representing the reflections 0.32 to 9.94 cm away from the receiver (sensor). The frame is divided into ten layers for subsequent quantification analysis, and each layer is approximately 0.95 cm.

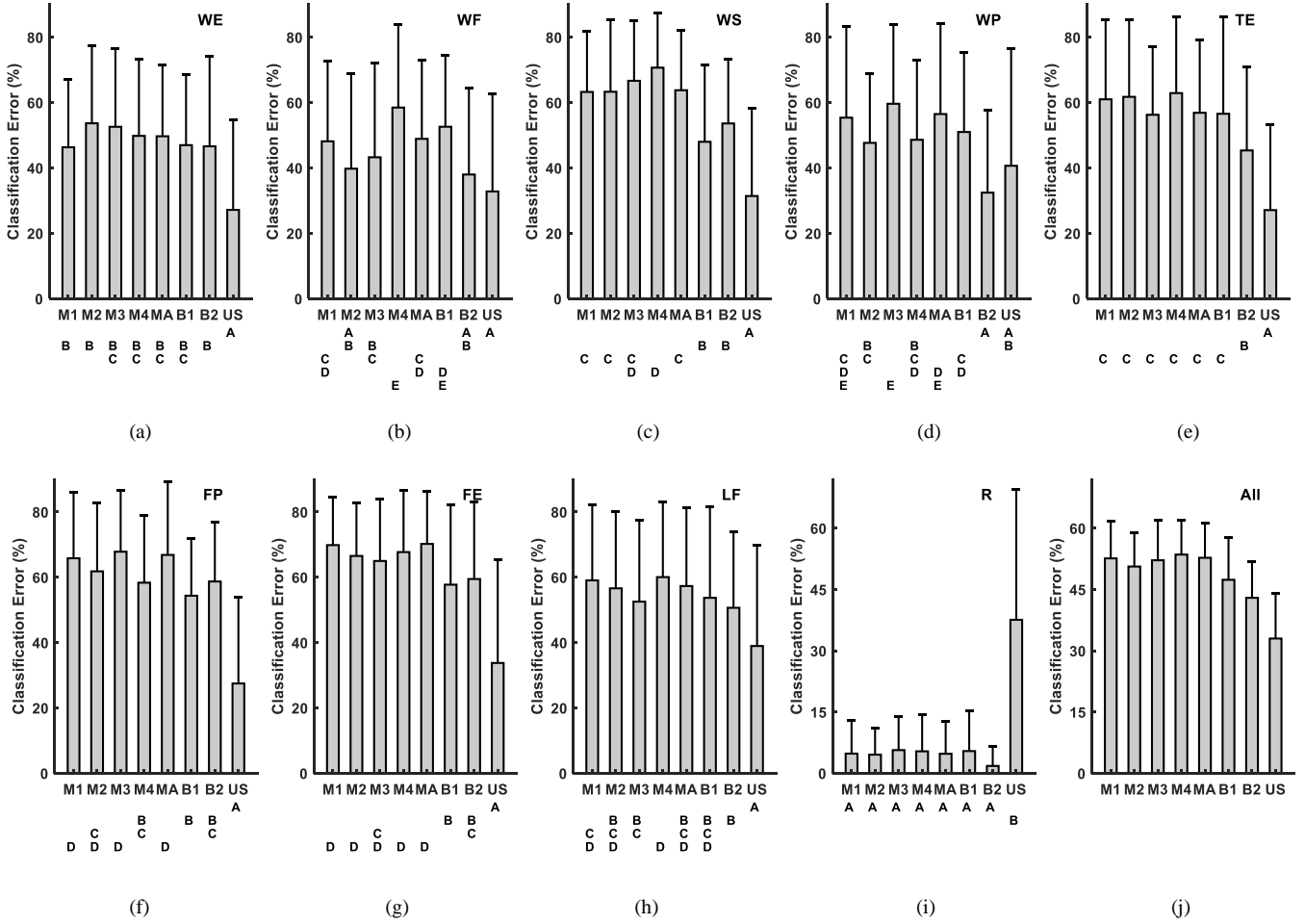


Fig. 5. Comparison of the two signal modalities, sEMG, including M1, M2, M3, M4, MA, B1, B2, and A-mode US, on wrist and finger gesture classification: (a) wrist extension, (b) wrist flexion, (c) wrist supination, (d) wrist pronation, (e) thumb extension, (f) fine pinch, (g) four-finger extension, (h) little-finger extension, (i) rest, and (j) all the movements. The values are averaged across all the subjects. The letters below each plot indicate the results of the post-hoc comparison. There is no significant difference ($p > 0.05$) if two methods share the same letter. For example, in (a), the difference between M1 and M2 is significant for labeled with different letters, B and C, respectively. The difference between M3 and M2 is not significant for sharing the same label B. The error rate is in ascending order alphabetically. The results varied against gesture classes.

simplicity. The following cross-validation procedure was the same as that of US data.

E. Quantification of US Feature Space

As US signal could detect volumetric change of deep muscles, whose electrical activities were usually masked by those of superficial muscles at the sEMG recordings, to systematically investigate the importance of such depth information in A-mode US, we introduced separability index (SI) [27] to quantify the effect of depth information on gesture classification

$$SI = \frac{1}{N} \sum_{j=1}^N \min_{1 \leq i \leq N; i \neq j} \frac{1}{2} \sqrt{(\mu_j - \mu_i)^T S_j^{-1} (\mu_j - \mu_i)}$$

where μ_j and μ_i is the centroid of class i and class j ; S_j is the covariance matrix of class j ; N is the number of gesture classes. SI measures the half Mahalanobis distance from one class to its nearest class, and averages the values across the classes in the same group.

Each US feature vector was segmented into ten layers from the most superficial to most deep. The dimension of first six

layers was 149 data points, and the dimension of last four layers was 148 data points (approximately 0.95 cm per layer). There were no overlap between two adjacent layers. The SI values of each layer for three groups, wrist gestures (WE, WF, WS, WP), finger gestures (TE, FP, FE, LE) and all the gestures were separately calculated.

F. Statistical Analysis

To compare the performance of three classifiers, LDA, SVM, ANN, on A-mode US data classification, a one-way repeated-measure analysis of variance (ANOVA) was conducted on the classification error rates. Further, for the comparison of two signal modalities, A-mode US and sEMG, a two-way repeated-measure ANOVA was performed. The two factors were gesture class (nine gestures) and data source (M1, M2, M3, M4, MA, B1, B2, US). If a significant interaction between two factors was detected, focused ANOVA would be performed by fixing the level of the gesture class. In addition, a one-way repeated-measure ANOVA was conducted on SI to compare the separability among the layers of the US feature vector for the three gesture groups: finger gestures, wrist

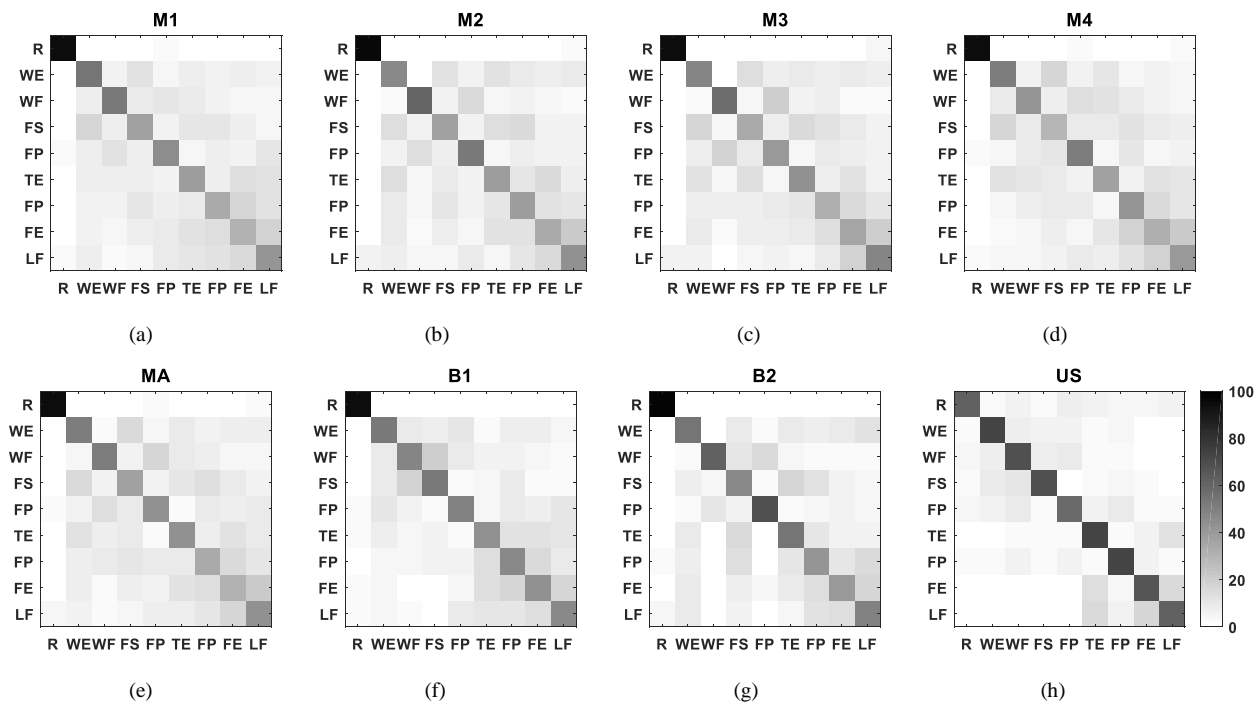


Fig. 6. Confusion matrices for the gesture classification of two signal modalities, sEMG (a-g) and A-mode US (h). The values are averaged across all the subjects. The column label represents the actual class, and the row label represents the predicted class. For a perfect classification of one movement with 0% error rate, each row should be white except that the entry on the diagonal, which should be black.

gestures and all the gestures. In all the tests, Turkey comparison with Bonferroni correction was performed when significant difference was detected for the factor of interest. Participant was a random factor in all the tests, and the significance level was set to 0.05.

III. RESULTS

A. Gesture Classification Comparison

Fig. 4 displays the classification errors of three classifiers, LDA, SVM, and ANN, on A-mode US data classification. There was no significant difference between the values of LDA and SVM, and they were significantly lower than the value of ANN. However, the abstract difference among the three average classification errors was less than 4%, indicating the limited effect of the classifier on A-mode US data classification. The results of LDA was selected for the subsequent comparison with sEMG.

For the comparison of two modalities, A-mode US and

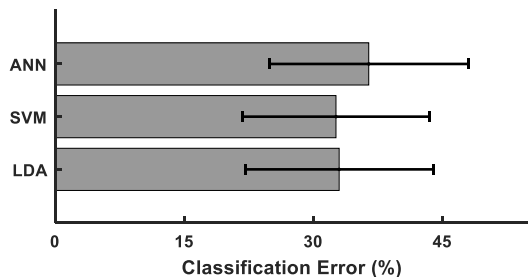


Fig. 4. Comparison of three classifiers on A-mode US data classification. The values are averaged across all the subjects. The abstract difference among the three average errors is <4%, indicating the effect of the classifier is limited.

sEMG (Fig. 5), the averaged error rate of A-mode US, $32.99 \pm 10.94\%$, was much lower than the value of sEMG ($50.29 \pm 9.69\%$ averaged across all seven sEMG methods). For the specific gestures, the results varied. A-mode US outperformed sEMG on FP, FE, WE and TE with a large margin ($>18.26\%$), as well as on LF and WF with the difference $>11.69\%$ and $>5.18\%$, respectively. On the other hand, the performance of A-mode US was inferior to sEMG on two gestures, WP and Rest class. In particular, for the Rest class, the performance of A-mode US was much worse to that of sEMG with the difference $>32.14\%$.

The statistical analysis confirmed the observation of the comparison between A-mode US and sEMG. A significant interaction was detected between the two factors, gesture class and data source, indicating the performance of the methods depended on the type of gesture class. As such, the focused ANOVAs were performed on each gesture. The effect of methods (data source) was significant in all the tests. Post-hoc comparison revealed that the performance of A-mode US was significantly better than that of the seven sEMG methods on the classification of the following six gestures, WE, WS, TE, FP, FE and LF. For gesture WF, there was no significant difference among the error rates of A-mode US, B2, M2, and the error rates of A-mode US was significantly lower than the error rates of the other methods. For gesture WP, the error rates of A-mode US were significantly higher than the error rates of M1, M3, MA, and B1, and comparable to the error rates of M2, M4 and B2. For the Rest class, the performance of A-mode US was significantly worse than all the sEMG methods.

For the misclassification displayed in confusion matrices (Fig. 6), the gestures from wrist (WE, WF, FS, FP) and finger (TE, FP, FE, LF) group of A-mode US were mostly

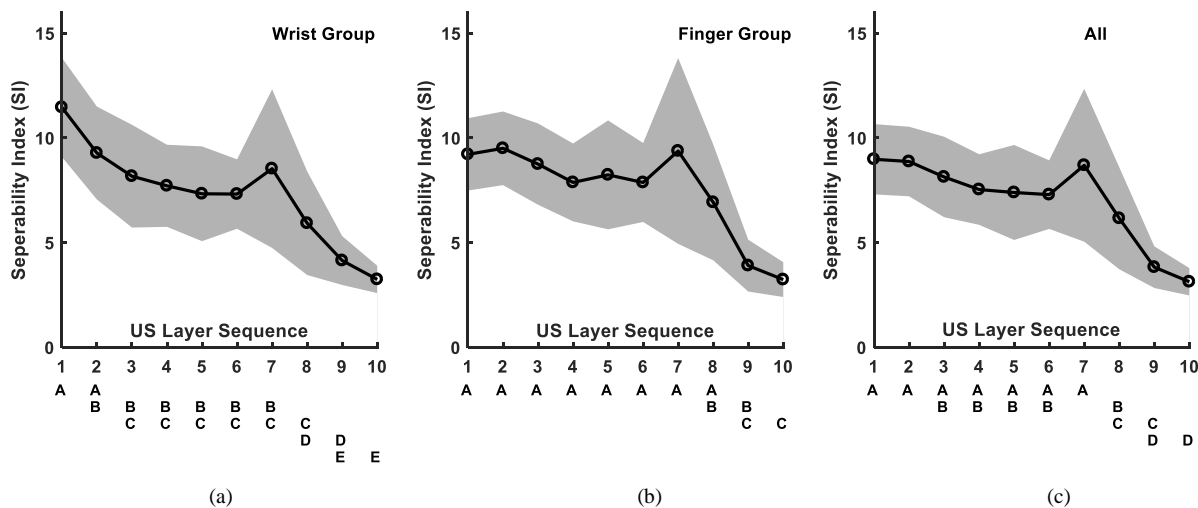


Fig. 7. Separability quantification of ten US layers for three gesture groups, (a) wrist gestures, (b) finger gestures, (c) all the gestures. The grey shade represents the standard deviation. The values are averaged across all the subjects. The letters indicate the results of post-hoc comparison, same as Fig. 5. There was no significant difference ($p > 0.05$) between two layers if they are labeled with the same letter. The SI value decreases with the sequence of the letter. The high SI value of the middle layer indicates the importance of the information from deep muscles.

misclassified as the one within the same group. Evidently, the diagonal boxes in the lower-right corner of A-mode US were darker than those of sEMG methods, indicating a significant higher accuracy achieved by A-mode US over sEMG for fine finger gestures, which was in part explained by the feature space analysis presented below.

B. Separability of US Data

SI value was adopted to quantify the separability of each US layer for three groups, wrist, finger and all the gestures (Fig. 7). For the wrist gesture group, the SI value decreased as the depth increased (away from the US transducer), and dropped sharply after the seventh layer. For the finger gesture group, the SI values did not change significantly for the first seven layers, and then it decreased sharply, similar to the scenario in wrist gestures. For the group of all the gestures, the trend of SI changes were between those of wrist and finger gestures, which was expected. It was interesting to note that the peak of SI for all cases was at the seventh layer, indicating crucial information contribution was obtained from deep muscles (the distance from this layer to the US transducer was between 6.11 and 7.06 cm). The statistical analysis indicated that US layer had a significant effect on the SI value for all the three groups. Post-hoc comparison confirmed the above qualitative observation of the three curves, displaying the significant difference among the SI value of each layer.

IV. DISCUSSION

A. Effectiveness of Depth Information

The muscle activities captured by sEMG and A-mode US were located in different areas. For sEMG, the activities were mainly from the superficial muscles, around the pickup area of the surface electrodes. For A-mode US, the activities were mainly from the muscles located on the scanning line of the transducer. The scanning area of A-mode US was not as broad as that of B-mode US, which was a plane. However, compared

to sEMG, A-mode US still had advantages in capturing information from the deep muscles (depth information), such as flexor digitorum and profundus, which controlled the fine finger gestures. This advantage was illustrated in Fig. 5, where A-mode US outperformed sEMG in six out of nine gesture recognition, including all the four finger gestures, with the US probe on flexor muscles. High classification accuracy ($>90\%$) was reported with the US sensor on flexors [12][14], extensors [13] and around the forearm [20]. As the ability of US modality in detecting activities away from the sensor, the sensor location would not affect the conclusion of this study.

The depth US signal detected was subject to the data length in one US echo frame. For simplicity, the length of the frame was set to be constant among the participants. The maximum depth US signal could reach was 9.94 cm in this study, longer than the forearm diameters of all the participants (from 7.3 to 9.2 cm). As few sound waves penetrated the forearm skin for high energy loss in the forearm, little information in the US echo was obtained from the outside of the forearm, and it would not contribute to the gesture classification for the status of the object under the forearm didn't change during the experiment.

The SI value was proposed to quantify the effectiveness of the depth information on the gesture classification for A-mode US. The difference between the curves of finger and wrist gesture group indicated that the role of each muscle layer in classification was different between two groups: the classification of wrist gestures relied more on the information from the superficial muscles, while the classification of finger gestures was more on the information from deep muscles. The peak at the seventh layer, approximately 6.11 to 7.06 cm away from the transducer, was observed on both curves. The value of US echo was positively related with the acoustic impedance difference of two adjacent substances. The biggest impedance difference was the muscle/bone interface. Considering the peak value of the raw A-mode US data in this segment (Fig. 3b), we inferred that the muscle/bone interface might be contained in

the seventh layer. It was possible in physics for the layer was located around three fourths to four fifths of the forearm diameter with the average diameter of 8.28 cm. This type of depth information was difficult to be extracted from sEMG. The layers at the two ends of the SI curve, represented the scenarios of the superficial muscles. The values close to the transducer, such as Layer 1, Layer 2 and Layer 3, were much bigger than the values away from the transducer, such as Layer 8, Layer 9, and Layer 10. It was because that the energy of the sound wave was reduced when penetrating the tissues, due to reflection, scattering, absorption and dispersion [29]. As such, the signal to noise ratio was reduced, inducing the decrease of the SI value with the increase of the depth.

B. Rest State Detection

Rest state detection was important for practical applications, because false activation was most detrimental to overall system performance [30]. One important finding of this study was the poor performance of A-mode US on the Rest state detection, comparing with sEMG. All sEMG methods achieved the low error rates on the classification of the Rest, much better than that of A-mode US. The difference between the performances of two signal modalities stemmed from their inherent properties. sEMG was related to the electrical activities from muscle contractions. When there was no contraction (the Rest), no signal could be detected. The abstract value of sEMG amplitude was the smallest, close to zero, while the values from the active gestures were all positive. As such, the expression of the Rest state in sEMG was special, easily distinguishable from the other classes, inducing its high classification accuracy. The good performance of sEMG on the Rest state detection was also reported in [31][32]. On the other hand, US signal was related to the volumetric changes in muscle. The expression of the Rest state in A-mode US was not special in values compared to that of the active gestures. As such, its detection accuracy was close to that of the other classes, not as good as the results from sEMG. This point was first reported in US-based gesture recognition studies.

C. Future work

This study presented the comparison of the gesture recognition between the single-channel A-mode US and sEMG. Advantages and limitations were illustrated for A-mode US gesture classification. Some issues needed to be addressed before its practical application. In this study, only static contraction, or discrete gesture recognition, was investigated. In daily lives, dynamic contraction, or continuous joint angle prediction, would be more useful. For dynamic contractions, the advantage of US modality over sEMG, the ability to detect deep muscle activities, was still important. Promising results were achieved with B-mode US on finger joint angle prediction [12][15]. As the similar characteristics between A-mode and B-mode US, it was expected that A-mode US could also benefit the scenario of joint angle prediction. Systematic studies were needed to shed more light on this.

As the dimension of A-mode US data was much larger than that of sEMG, the processing time could be increased. In this

study, with the PC setting of Intel Core i7-6700 CUP, 3.4 GHz clock speed, and 16 G RAM, the time to generate one decision from raw data was less than 0.3 ms for sEMG, and around 7 ms for A-mode US in matlab R2016b platform, both less than the window step length (50 ms). The time could be reduced with the low-level programming languages, such as C, C++. As online control achieved with B-mode US [12][14], it was achievable for real-time A-mode US gesture recognition. In addition, power consumption of A-mode US was also different from sEMG. A-mode US needed external power to generate sound waves, while sEMG didn't. As such, it would raise a challenge for battery capacity if it was applied in wearable devices. Methods to decrease power consumption, such as using less elements, would be considered for future application.

For sEMG-based gesture recognition, lack of robustness was one main obstacles for its practical application. As for A-mode US, the main challenge in robustness was the relative movement between the US transducer and the forearm. As the plane of the current transducer was not complied with the skin surface, it was inevitable that the transducer would move during contractions. Such change of the transducer/skin interface would change the echo waveform received and reduce the repeatability of the signal, resulting in a degradation of classification accuracy in various degree. Consequently, we observed a larger standard deviation in such as WS, FP in Fig. 5. Robust feature extraction, such as using effective spatial filtering algorithms [33][34], and/or applying deep learning for feature representation learning [35], would provide a solution for this problem.

Finally, sEMG and US have been investigated for gesture recognition separately. However, as far as we knew, there was no prior studies reporting the possibility of using both signal modalities synergistically. Clear evidences were showed in this study that the two signal modalities, sEMG and A-mode US, had complementary advantages in detecting hand and wrist gestures. Therefore, it was necessary to develop an effective sensor fusion algorithm to exploit the respective strengths of two modalities, enhancing the performance of the gesture recognition.

V. CONCLUSION

This study compared the single-channel performance of two signal modalities, sEMG and A-mode US, on the recognition of nine hand and wrist gestures. The classification performance of A-mode US was better than that of sEMG on six gestures (two wrist gestures, WE and WS; four finger gestures, TE, FP, FE and LF), but significantly worse on the Rest class. For A-mode US, the effectiveness of its depth information on gesture recognition was analyzed and confirmed by the quantification of the signal separability in each layer. This study presented the potential of A-mode US from the single-element transducer in the application of HMI.

REFERENCES

- [1] D. Tkach, H. Huang, and T. a Kuiken, "Study of stability of time-domain features for electromyographic pattern recognition.," *J.*

- Neuroeng. Rehabil.*, vol. 7, p. 21, Jan. 2010.
- [2] J. Ning, S. Dosen, K.-R. Müller, and D. Farina, "Myoelectric Control of Artificial Limbs—Is There a Need to Change Focus? [In the Spotlight]," *IEEE Signal Process. Mag.*, vol. 29, no. 5, pp. 152–150, Sep. 2012.
- [3] Y. Li, X. Chen, X. Zhang, K. Wang, and Z. J. Wang, "A sign-component-based framework for Chinese sign language recognition using accelerometer and sEMG data," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 10, pp. 2695–2704, 2012.
- [4] X. Zhang and P. Zhou, "High-density myoelectric pattern recognition toward improved stroke rehabilitation," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 6, pp. 1649–1657, 2012.
- [5] W. S. Harwin, T. Rahman, and R. a. Foulds, "A review of design issues in rehabilitation robotics with reference to North American research," *IEEE Trans. Rehabil. Eng.*, vol. 3, no. 1, pp. 3–13, 1995.
- [6] O. Fukuda, T. Tsuji, M. Kaneko, and A. Otsuka, "A human-assisting manipulator teleoperated by EMG signals and arm motions," *IEEE Trans. Robot. Autom.*, vol. 19, no. 2, pp. 210–222, 2003.
- [7] H. Liang, J. Yuan, D. Thalmann, and N. Magnenat-thalmann, "AR in Hand: Egocentric Palm Pose Tracking and Gesture Recognition for Augmented Reality Applications," *ACM Multimed. Conf.*, pp. 6–7, 2015.
- [8] M. Asghari Oskoei and H. Hu, "Myoelectric control systems—A survey," *Biomed. Signal Process. Control*, vol. 2, no. 4, pp. 275–294, Oct. 2007.
- [9] J. He, D. Zhang, X. Sheng, S. Li, and X. Zhu, "Invariant Surface EMG Feature Against Varying Contraction Level for Myoelectric Control Based on Muscle Coordination," *IEEE J. Biomed. Heal. Informatics*, vol. 2194, no. c, pp. 1–1, Jun. 2014.
- [10] G. Li, A. E. Schultz, and T. a. Kuiken, "Quantifying pattern recognition-based myoelectric control of multifunctional transradial prostheses," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 18, no. 2, pp. 185–192, Apr. 2010.
- [11] R. S. C. Cobbold, "Foundations of Biomedical Ultrasound," *Oxford Univ. Press*, pp. 45–51, 2007.
- [12] Y. Huang, X. Yang, Y. Li, D. Zhou, K. He, and H. Liu, "Ultrasound-based Sensing Models for Finger Motion Classification," *IEEE J. Biomed. Heal. Informatics*, vol. 2194, no. c, pp. 1–11, 2017.
- [13] J. Shi, J. Y. Guo, S. X. Hu, and Y. P. Zheng, "Recognition of Finger Flexion Motion from Ultrasound Image: A Feasibility Study," *Ultrasound Med. Biol.*, vol. 38, no. 10, pp. 1695–1704, 2012.
- [14] N. Akhlaghi *et al.*, "Real-Time Classification of Hand Motions Using Ultrasound Imaging of Forearm Muscles," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 8, pp. 1687–1698, Aug. 2016.
- [15] C. Castellini, G. Passig, and E. Zarka, "Using ultrasound images of the forearm to predict finger positions," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 20, no. 6, pp. 788–797, 2012.
- [16] S. Sikdar *et al.*, "Novel method for predicting dexterous individual finger movements by imaging muscle activity using a wearable ultrasonic system," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 22, no. 1, pp. 69–76, 2014.
- [17] D. Sierra González and C. Castellini, "A realistic implementation of ultrasound imaging as a human-machine interface for upper-limb amputees," *Front. Neurobot.*, vol. 7, no. OCT, pp. 1–11, 2013.
- [18] T. L. Szabo, "Therapeutic Ultrasound," in *Diagnostic Ultrasound Imaging: Inside Out*, vol. 787, Elsevier, 2014, pp. 735–763.
- [19] J.-Y. Guo, Y.-P. Zheng, H.-B. Xie, and T. K. Koo, "Towards the application of one-dimensional sonomyography for powered upper-limb prosthetic control using machine learning models," *Prosthet. Orthot. Int.*, vol. 37, no. 1, pp. 43–49, Feb. 2013.
- [20] Y. Li, K. He, X. Sun, and H. Liu, "Human-machine interface based on multi-channel single-element ultrasound transducers: A preliminary study," *2016 IEEE 18th Int. Conf. e-Health Networking, Appl. Serv. Heal. 2016*, pp. 1–6, 2016.
- [21] T. R. Farrell and R. F. F. Weir, "A comparison of the effects of electrode implantation and targeting on pattern classification accuracy for prosthesis control," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 9, pp. 2198–211, Sep. 2008.
- [22] P. Parker, K. Englehart, and B. Hudgins, "Myoelectric signal processing for control of powered limb prostheses," *J. Electromyogr. Kinesiol.*, vol. 16, no. 6, pp. 541–8, Dec. 2006.
- [23] M. A. Oskoei and H. Hu, "Support vector machine-based classification scheme for myoelectric control applied to upper limb," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 8, pp. 1956–65, Aug. 2008.
- [24] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. 2000.
- [25] A. J. Young, L. J. Hargrove, and T. a. Kuiken, "The effects of electrode size and orientation on the sensitivity of myoelectric pattern recognition systems to electrode shift," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 9, pp. 2537–44, Sep. 2011.
- [26] B. Hudgins, P. Parker, and R. N. Scott, "A new strategy for multifunction myoelectric control," *IEEE Trans. Biomed. Eng.*, vol. 40, no. 1, pp. 82–94, Jan. 1993.
- [27] N. E. Bunderson and T. A. Kuiken, "Quantification of feature space changes with experience during electromyogram pattern recognition control," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 20, no. 3, pp. 239–46, May 2012.
- [28] L. J. Hargrove, K. Englehart, and B. Hudgins, "A comparison of surface and intramuscular myoelectric signal classification," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 5, pp. 847–53, May 2007.
- [29] P. N. T. Wells, "Absorption and dispersion of ultrasound in biological tissue," *Ultrasound in Medicine and Biology*, vol. 1, no. 4, pp. 369–376, 1975.
- [30] L. J. Hargrove, E. J. Scheme, K. B. Englehart, and B. S. Hudgins, "Multiple binary classifications via linear discriminant analysis for improved controllability of a powered prosthesis," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 18, no. 1, pp. 49–57, Feb. 2010.
- [31] L. Pan, D. Zhang, X. Sheng, and X. Zhu, "Improving myoelectric control for amputees through transcranial direct current stimulation," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 8, pp. 1927–1936, 2015.
- [32] W. Guo, X. Sheng, H. Liu, and X. Zhu, "Toward an Enhanced Human-Machine Interface for Upper-Limb Prosthesis Control With Combined EMG and NIRS Signals," *IEEE Trans. Human-Machine Syst.*, pp. 1–12, 2017.
- [33] J. He and X. Zhu, "Combining Improved Gray-Level Co-Occurrence Matrix with High Density Grid for Myoelectric Control Robustness to Electrode Shift," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11, no. 4, pp. 1–1, 2016.
- [34] J. M. Hahne, B. Graimann, and K.-R. Müller, "Spatial filtering for robust myoelectric control," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 5, pp. 1436–43, May 2012.
- [35] J. Shi, S. Zhou, X. Liu, Q. Zhang, M. Lu, and T. Wang, "Stacked deep polynomial network based representation learning for tumor classification with small ultrasound image dataset," *Neurocomputing*, vol. 194, pp. 87–94, 2016.