# High Performance Wearable Ultrasound as a Human-Machine Interface for wrist and hand kinematic tracking

Bruno Grandi Sgambato, Milia H Hasbani, *Member, IEEE*, Deren Y Barsakcioglu, *Member, IEEE*, Jaime Ibáñez, Anette Jakob, Marc Fournelle, Meng-Xing Tang, *Senior Member, IEEE*, Dario Farina, *Fellow, IEEE*

***Abstract*— *Objective:* Non-invasive human machine interfaces (HMIs) have high potential in medical, entertainment, and industrial applications. Traditionally, surface electromyography (sEMG) has been used to track muscular activity and infer motor intention. Ultrasound (US) has received increasing attention as an alternative to sEMG-based HMIs. Here, we developed a portable US armband system with 24 channels and a multiple receiver approach, and compared it with existing sEMG- and US-based HMIs on movement intention decoding. *Methods:* US and motion capture data was recorded while participants performed wrist and hand movements of four degrees of freedom (DoFs) and their combinations. A linear regression model was used to offline predict hand kinematics from the US (or sEMG, for comparison) features. The method was further validated in real-time for a 3-DoF target reaching task. *Results:* In the offline analysis, the wearable US system achieved an average $R^2$ of 0.94 in the prediction of four DoFs of the wrist and hand while sEMG reached a performance of $R^2$ = 0.60. In online control, the participants achieved an average 93% completion rate of the targets. *Conclusion:* When tailored for HMIs, the proposed US A-mode system and processing pipeline can successfully regress hand kinematics both in offline and online settings with performances comparable or superior to previously published interfaces. *Significance:* Wearable US technology may provide a new generation of HMIs that use muscular deformation to estimate limb movements. The wearable US system allowed for robust proportional and simultaneous control over multiple DoFs in both offline and online settings.**

***Index Terms*— Human-Machine Interfaces, Ultrasound, A-mode, Surface Electromyography**

This work was supported by funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 899822, SOMA project.

Bruno Grandi Sgambato, Milia H Hasbani, Deren Y Barsakcioglu, Jaime Ibáñez, Meng-Xing Tang and Dario Farina are with the Department of Bioengineering, Imperial College, London, United Kingdom.

Milia H Hasbani was supported by the UKRI EPSRC.

Jaime Ibáñez works at I3A, University of Zaragoza, Spain and was supported by a Ramón y Cajal grant (RYC2021-031905-I) funded by MCIN/AEI/10.13039/501100011033 and UE's NextGenerationEU/PRTR funds.

Marc Fournelle and Anette Jakob are with the Department of Ultrasound, Fraunhofer Institute for Biomedical Engineering, Sulzbach, Germany.

(correspondence e-mail: d.farina@imperial.ac.uk).

## I. Introduction

Human-machine interfaces (HMIs) are commonly used to track the upper limb position and to recognize gestures. Ideally these interfaces should not only allow for robust and complex interactions but should also be intuitive to use. The main applications for HMI decoding of movements or motor intention are prosthetic control, remote manipulation, and virtual reality. These applications impose high portability requirements, which implies that the technologies used for a HMI should be as unobstructive as possible, preferably wearable and light, allowing natural movements. Additionally, these systems may be deployed in a variety of environmental conditions where robustness is a key issue.

Surface electromyography (sEMG) is the main technology currently used for HMI. Commercial applications mainly rely on sEMG-based estimates of muscle contraction levels to implement proportional control strategies [1]. In prosthetics, this kind of technology typically allows users to control one degree of freedom (DoF) at a time or multiple DoFs sequentially [2]. Myoelectric pattern recognition systems have also been applied for simultaneous multi-DoF control [3].

Limitations of sEMG-based HMIs are mostly due to the inherent stochastic and non-stationary nature of the sEMG signal, and to the fact that it provides superficial information with relatively low spatial resolution, which makes it difficult to compensate for crosstalk effects [4]. Furthermore, even though sEMG has a direct relation to muscle activation force, its relationship with specific hand and wrist positions is not subject-specific and may vary over time. This implies that sEMG-based HMIs typically rely on users making active efforts to generate specific activation patterns with their arms and hands for each gesture.

Ultrasound (US), on the other hand, can be used to identify the muscle morphology independent of the presence, or lack of, muscle activity. Therefore, it may provide better mapping to hand kinematics. Its application for HMIs was initially proposed by Zheng *et al*. [5], using brightness mode (B-mode) to continuously monitor muscle activity (referred to as sonomyography). Since then, this method has been validated by multiple groups [6-8] for the task of prosthetic control via computer vision-inspired pattern recognition pipelines.

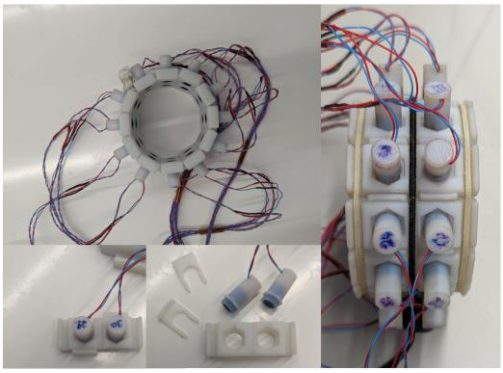

Fig. 1. Custom A-mode US bracelet. Rectangular holders were individually 3D printed to hold a pair of transducers at 1.5cm from each other. Clips hold the transducers in place. The design allows for a variable number of transducers to be placed depending on participant forearm circumference.

Although these methods showed remarkably higher performance compared to sEMG [9, 10], they are not ready for translation outside laboratory settings because of the large size of the US probes needed, and the low robustness they have against probe shifts and donning and doffing.

A more translatable design of US systems for HMIs involves independent US transducers, in amplitude mode (A-mode), positioned around the forearm in an armband-like fashion [11]. In exchange for local spatial resolution, A-mode wearable systems ensure wide coverage of the forearm muscles, redundancy, and robustness to transducer shift. Initially proposed with a single transducer tracking muscle deformation versus wrist extension [12], subsequent works proposed improvements of this technique [11, 13, 14]. Its most recent application uses 8 channels and provides reliable estimates in proportional 2-DoF target achievement control tasks [15].

In this study, we propose a novel wearable US system and its application to accurately control more than 2 simultaneous DoFs. The developed wearable US armband has 24 independent transducers and advances the state of the art because of a greater number of channels than previous systems and of the operation model that implies that each transducer acts as receiver for waves generated by all other transducers. With this system, we implemented a machine learning pipeline that achieved simultaneous position control of wrist and hand kinematics. We first validated this pipeline offline in combination with a motion capture system and compared its performance against sEMG, and then validated it online in a pointer control scenario.

## II. Methods

### A. Participants

8 healthy participants (n=8, 6 female and 2 male, 27 ± 4 *yr*) were recruited for the offline experiments and other 8 healthy participants (n=8, 3 female and 5 male, 28 ± 4 *yr*) took part on the online target achievement control (TAC) experiment. All procedures and experiments were approved in accordance with the declaration of Helsinki by the Imperial College Research Ethics Committee (refs: 18IC4685 and 22IC7602). Before data collection, the volunteers were briefed on the study, presented with a participant information form, allowed to ask any questions, and asked to sign a consent form. As the datasets acquired are large, the data was not registered in a public database but may be made available, as appropriate, upon request.

TABLE I
DESCRIPTIONS OF EACH MOVEMENT SET PERFORMED

| Movement Set | Description | Movements performed |
|---|---|---|
| Simple | 1 DoF wrist and hand movements | Wrist flexion/extension, wrist radial/ulnar deviation, pronation and supination, hand opening/closing |
| Sequential | Opening/closing hand in different wrist positions in 1 DoF | Sequential wrist movement and hand closing/opening movements. Same wrist movements as in set 1 |
| Simultaneous | Opening/closing hand while wrist moves in 1 DoF | Simultaneous wrist movement and hand closing/opening. Same wrist movements as set 1 |

### B. Equipment

#### 1) Ultrasound A-mode Armband

We designed an armband system (Fig. 1), made of multiple small 3D printed holders and held by one adjustable string and two elastic bands. It provides a flexible and wearable interface for up to 32 piezoelectric transducers. The bracelet design allowed positioning of the transducers in two parallel circles (1 cm apart). This allowed us to generate two detection planes. The bracelet modularity allows its usage with a varying number of transducers, depending on the size of the user's limb. Each of the transducers is connected via a twisted pair cable to a printed circuit board that is connected to a portable acquisition system (MoUSE, Fraunhofer IBMT, Sulzbach, DE). This system was in charge of driving the US transmission and of recording the signals. The system can send customizable sequences of transmissions for each channel and records the US echoes at up to 12-bit resolution for up to 320 $\mu s$ at a 50-MHz sampling rate [16]. The system works as a multiple receiver interface: for each transmission event is received and recorded by all transducers.

We used 1-MHz piezoelectric transducers (circular, 2.5mm radius, 15° opening angle). This frequency has been used in previous studies [17] but its lower than the frequencies used in some other A-mode studies (5-MHz [18, 19], 10-MHz [13]). Our choice was based on the working principle of the device, as a multiple-receiver system. The relatively small frequency reduced tissue attenuation, thus maximizing the signal power arriving at the receivers around the forearm.

#### 2) Motion Capture

An optical marker motion capture system (SMART DX, BTS Bioengineering, Milano, IT) with eight cameras was used to record the positions of 14 markers at 150 frames per second. The cameras were positioned to ensure that throughout the duration of the movements each marker was visible by two or more cameras. The 14 reflective markers were positioned on the hand, wrist, and elbow to measure the

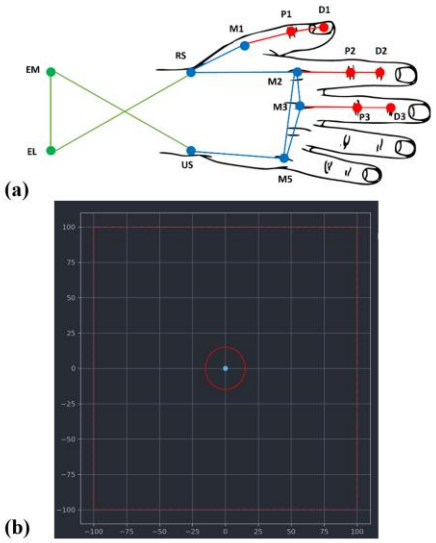


Fig. 2. (a) Placement positions of the optical markers. US: styloid process of ulna; RS: styloid process of radius; EM: medial epicondyle of elbow; EL; lateral epicondyle of elbow; Mi: head of the metacarpal bone of finger i (i = 1-3); Pi: head of proximal phalanx of finger i (i = 1-3); Di: head of distal phalanx of the thumb (i = 1) and head of middle phalanx of long fingers (i = 2-3). (b) Graphical interface developed for the online validation. The 3-DoF control was abstracted as movements of the blue point along the x and y axis as well as the changing of the red circle size.

arm and hand kinematics. The placement of the markers is described in Fig. 2a.

*3) Synchronization*

An active low output port of the US system sent 50-$\mu s$ trigger pulses at every transmission. These pulses were fed into an Arduino Due which provided an input for the motion capture system. This signal was sampled at 500 Hz and used in post-processing to align the US and motion capture recordings.

## C. Experiments

Participants were guided to an adjustable stool surrounded by motion capture cameras. The right arm was positioned on a metallic cushioned support frame resting at a neutral position (hand relaxed and ulnar styloid pointing downwards). A computer screen was positioned outside the area recorded by the video cameras, at a comfortable distance from the participant. The screen displayed videos showing the participant how to perform the required movements. The US bracelet was mounted on the forearm at around 3-cm distance from the antecubital fossa. This position was chosen as it both provided comfort and allowed coverage of the bellies of most of the wrist and extrinsic hand muscles. Water-soluble hypoallergenic US transmission gel was applied between each of the transducers and the skin to provide appropriate coupling. The transducers were designed with a slight concave curvature in the last coupling layer, so that the gel was trapped between the sensors and the skin, improving the coupling and reducing the needed amount of gel. This also had the side effect of reducing the "slipping" of the bracelet, which is often observed in similar systems due to the reduced friction caused by use of the gel. Due to the bracelet design, not all transducers could be positioned for participants with a small

TABLE II
NORMALIZED TARGETS FOR THE TARGET ACHIEVEMENT CONTROL TEST. FROM -1 TO 1 FOR MAXIMUM FLEXION TO EXTENSION AND PRONATION TO SUPINATION, AND FROM 0 TO 1 FOR HAND OPENING TO CLOSING

| | Flexion Extension | Pronation Supination | Hand Open Close |
|---|---|---|---|
| 1 DoF | 0.3 | 0.0 | 0.0 |
| | -0.3 | 0.0 | 0.0 |
| | 0.0 | 0.3 | 0.0 |
| | 0.0 | -0.3 | 0.0 |
| | 0.0 | 0.0 | 0.3 |
| | 0.0 | 0.0 | 0.6 |
| | 0.8 | 0.0 | 0.0 |
| | -0.8 | 0.0 | 0.0 |
| | 0.0 | 0.8 | 0.0 |
| | 0.0 | -0.8 | 0.0 |
| | 0.0 | 0.0 | 0.8 |
| | 0.0 | 0.0 | 1.0 |
| 2 DoFs | 0.5 | 0.5 | 0.0 |
| | 0.5 | -0.5 | 0.0 |
| | -0.5 | 0.5 | 0.0 |
| | -0.5 | -0.5 | 0.0 |
| | 0.5 | 0.0 | 0.5 |
| | -0.5 | 0.0 | 0.5 |
| | 0.0 | 0.5 | 0.8 |
| | 0.0 | -0.5 | 0.8 |
| 3 DoFs | 0.3 | 0.3 | 0.3 |
| | 0.3 | -0.3 | 0.3 |
| | -0.3 | 0.3 | 0.3 |
| | -0.3 | -0.3 | 0.3 |
| | 0.5 | 0.5 | 0.8 |
| | 0.5 | -0.5 | 0.8 |
| | -0.5 | 0.5 | 0.8 |
| | -0.5 | 0.5 | 0.8 |

forearm circumference. Therefore, for consistency across subjects, only 24 transducers (12 per circle) were positioned and used in all experiments. Finally, the 14 reflective markers for motion capture were positioned.

Participants were instructed to perform 19 different tasks with the guidance of videos. Each task consisted of five repetitions of a specific type of movement resulting in 95 recordings. The movements, described in Table I, involved moving the wrist and/or fingers from a neutral position to a target position activating one (or more) DoFs (2 s), holding the position (3 s), and moving back to neutral (2 s). For the sequential tasks, participants first moved the wrist to the target position and maintained that position while opening and closing the hand (2 s for transitions), and then they moved the wrist back to a neutral position. The Simple movement set involved each of the four DoFs moving in isolation in its simplest form. The Sequential set aimed to explore having the wrist DoFs in different positions while activating the hand DoF. Lastly, the Simultaneous set involved moving each of the wrist DoFs and the hand DoF at the same time.

## D. Data Analysis

*1) Motion Capture*

From the motion capture, the positions of the markers were calculated in (x, y, z) coordinates in a 3D reference space, interpolated via $1^{st}$ degree spline, and filtered with a $3^{rd}$ order Butterworth low-pass filter with frequency determined via residual analysis [20]. Flexion-extension, radial-ulnar deviation, and supination-pronation angles of the wrist and the metacarpophalangeal and proximal interphalangeal joints of the first three fingers were calculated following methods by

Carpinela *et al.* [21] (adapted by [22]). Lastly, the nine calculated angles were filtered via a low-pass filter ($3^{rd}$ order, Butterworth, cut-off frequency 5 Hz) and linearly interpolated to the US acquisition frame rate. In this work, only the angles of the wrist flexion-extension, radial-ulnar deviation and pronation-supination and the metacarpophalangeal angle of the index finger (representing one DoF of hand opening and closing) were used. Lastly, the kinematic data was used to split each recording into five parts, one for each repetition. This was done via semi-automatic cumulative summation, followed by a manual check by an expert.

*2) US Data*

The developed processing pipeline is illustrated in Fig. 3. The processing is performed one frame of US at a time. Each frame of A-mode data is comprised of 24 sonification events (one for each transducer in order) acquired one after the other. For each sonification event, one transducer acted as a transmitter and the same transducer as well as the rest of the transducers acted as receivers. Therefore, for each transmission event, 24 lines of US echoes were recorded for 80 μs at 50-MHz, resulting in 4000 samples per line. Therefore, each frame comprised 24 events x 24 lines x 4000 samples. The time between the first and last sonification event in each frame was approximately 2 ms, and therefore we considered them all to be done instantaneously for processing purposes. The offline experiment was performed at approximately 4.5 frames per second.

Pre-processing started with channel selection and Time Gain Compensation, where half of the channels for each shot were excluded. This was done because the transducers were placed in two circles parallel to each other, meaning that only minimal information was recorded between them due to the low opening angle of the transducers. On the remaining channels, Time Gain Compensation was applied (soft tissue attenuation approximated to 0.5 dB/MHz*cm) to correct for attenuation.

Afterwards, noise and delay corrections were applied. Static noise artifacts were present at the start of each channel and therefore the initial 800 samples were removed from all channels. Furthermore, as transducers were at different positions on the forearm, the initial samples often did not contain significant physiological information due to the propagation time needed for the wave to arrive at them. As the distance between each was variable, we used a shifting thresholding method to define a delay before any significant activity was found in each line. This resulted in channels with a variable number of samples, from around 900 to 3200 samples per line. The last step of prepossessing was the smoothing and enveloping. Smoothing was performed via a Gaussian time filter (sigma 3) and the envelope was extracted via the absolute value of a Hilbert transformation.

The feature extraction step started with windowing into non-overlapping 200-samples windows (the last window with length less than 200 samples was excluded). From each window, the Root Mean Square (RMS) was computed and used as the only feature. Due to the high dimensionality of the US data this process resulted in over 3000 features per frame, many of which did not contain discriminative information because of the positioning of tissue and bone. Therefore, Principal Component Analysis (PCA) was applied to reduce the dimensionality of the feature space. Experimentally, maintaining the first 100 principal components was a good compromise between performance, robustness, and reduced potential for overfitting. The results, however, were similar across large changes in the number of kept components. The kept PCA components were used to train a group of linear regression (LR) models, using the motion capture calculated joint angles as the labels for each frame of US.

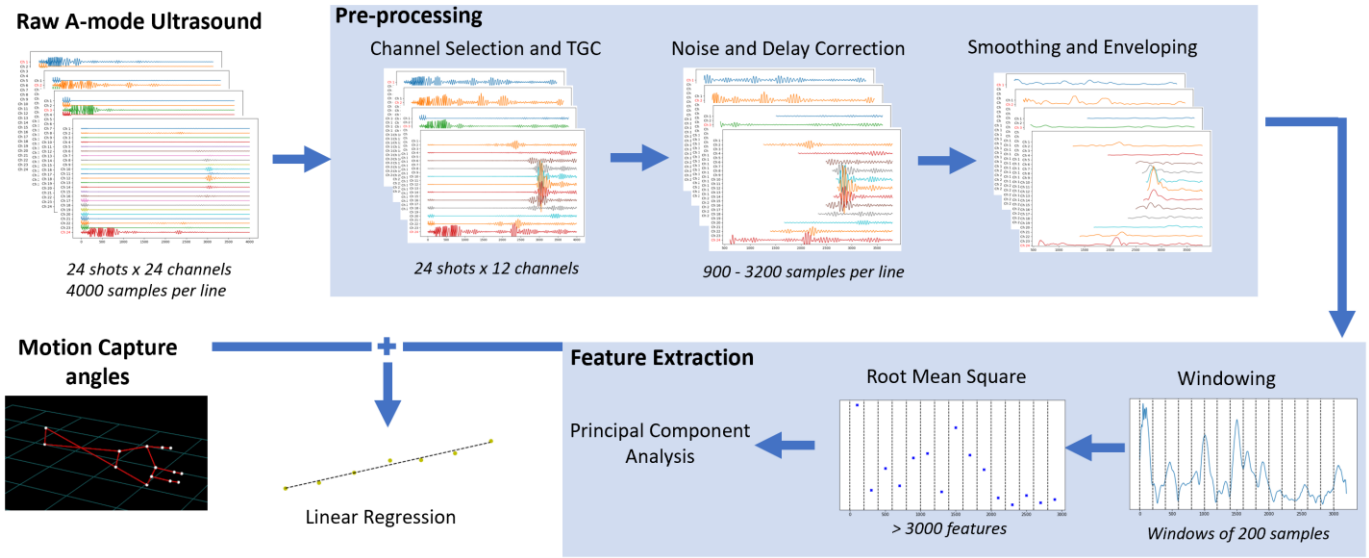


Fig. 3. Diagram of the US processing pipeline. The raw US data goes through preprocessing where the channels on plane are selected, corrected via time gain compensation, have their static noise removed, are corrected propagation delay, are smoothed, and enveloped. This follows with a feature extraction step where data is windowed, and the root mean square is extracted from each followed by a principal component analysis were only the first 100 components are kept. Lastly, together with the motion capture calculated angles one linear regressor is trained for each DoF.

Each participant was analysed individually, over the 95 recordings. Tests were performed on four sets of training and test data. The first three comprised each individual movement set described in Table I, and the fourth (indicated as All) used all movements sets together. Results for each combination were obtained via 5-fold cross-validation, where the training set was the concatenation of four repetitions of each gesture in the set and the test set contained the remaining repetitions. For each of the five folds, four LR were used to estimate the three angles of the wrist and one for the hand opening and closing. For all folds, all four angles were always estimated regardless of them being those instructed to be active, i.e. the regressor always worked on four DoFs estimates. From the estimations made by the LR system, the Coefficient of Determination ($R^2$) was calculated as the performance metric. Therefore, each participant produced 20 performance metrics per movement set.

### *E. sEMG Experiment*

sEMG data were obtained from a previous study performed in the same laboratory as the US experiments (detailed description in [22]). Briefly, motion capture and sEMG were recorded simultaneously from the forearms of eight subjects (n=8, 6 male and 2 female, 27 ± 5 *yr*) while performing movements of the wrist and hand. The sEMG used for the analyses in the current study were 24 bipolar signals (2-mm electrodes; 20-mm inter-electrode distance) positioned around the forearm below the elbow. They were sampled by an amplifier (Quattrocento, OT Bioelettronica, Torino, IT) at 2048 Hz, A/D converted to 16-bits, and digitally band-pass filtered (10-500 Hz). An Arduino uno was used to generate an active low trigger pulse to be sampled both by the sEMG amplifier and by the motion capture system and used to align the recordings. To ensure consistency between the US and the sEMG experiments, participants performed the same tasks, using the same videos as guidance.

The motion capture data was filtered in the same way as in the US experiment. The sEMG data was further filtered between 20-500 Hz by a $5^{th}$ order Butterworth band pass filter, and by $3^{rd}$ order band stops at 50, 100 and 150 Hz, and rectified. Noisy or disconnected channels were excluded from further analysis. The data was windowed in non-overlapping 220-ms intervals, and the RMS feature was extracted for each channel. LR on the same data splits used for US were used for EMG with $R^2$ and the RMSE as the performance metrics. However, only for the EMG the predicted angles from the linear regressor were further low-pass filtered with cut-off frequency of 1 Hz ($5^{th}$ order Butterworth filter) to reduce the variation between consecutive predictions to a physically compatible range.

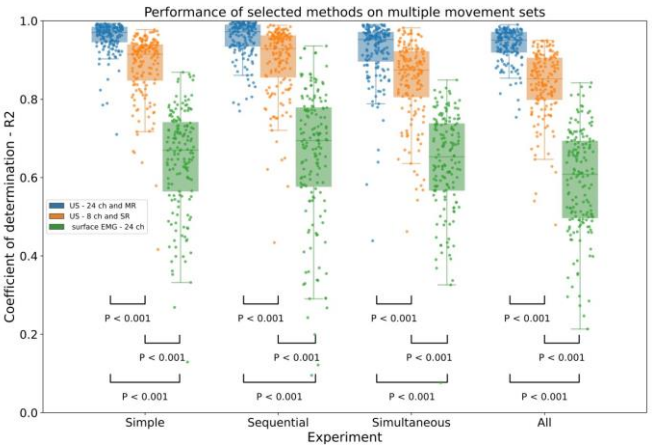


Fig. 5. $R^2$ scores for three different models in four different movement sets. Blue uses the US processing detailed above with 24 channels and the multiple receiver approach, orange uses the same processing with eight channels and only single receivers, and green uses sEMG. Jitter graph shows individual results for each cross-validation of each DoF for each participant. Jitter results below zero were not included.

### *F. Online Validation*

The US pipeline previously described was implemented in an online fashion with a higher frame rate of approximately 14 frames per second and used to perform a 3-DoF TAC test [23]. As in previous studies [24, 25], control was presented in an abstract manner (Fig. 2b). The blue point was controlled over the x-axis via the participant flexion-extension (left to right) and over the y-axis via pronation-supination (bottom to top) while the circumference radius was controlled via the opening and closing of the hand. With the system positioned, participants were guided by visual cues and by the

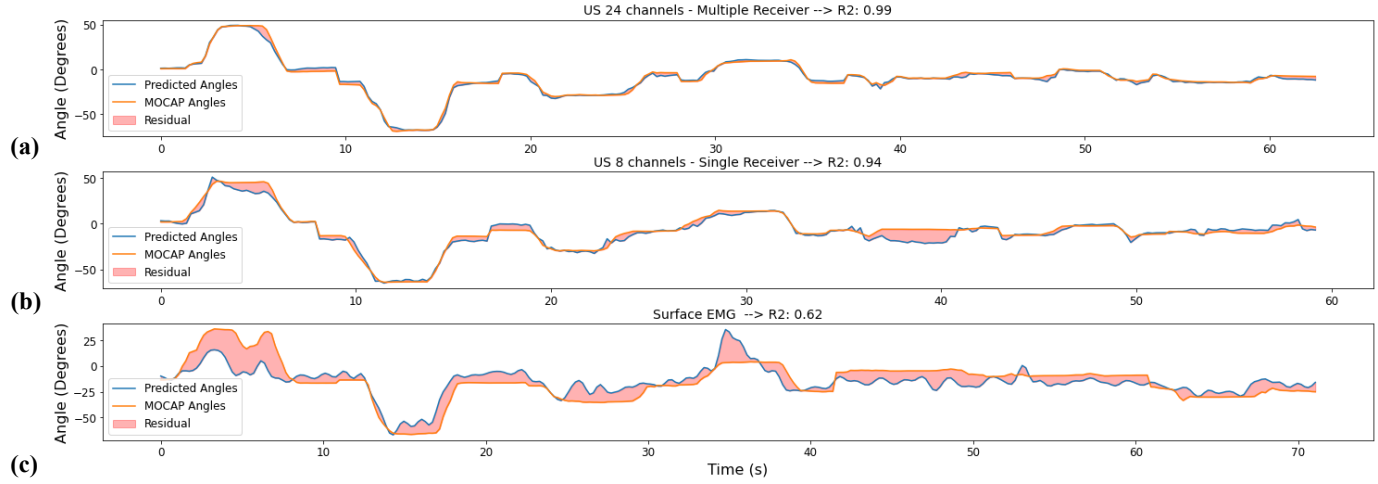


Fig. 4. Examples of target, prediction, and residual graphs for one testing group using 3 distinct models. Model trained with (a) A-mode US with 24 channels with multiple receivers, (b) A-mode US with 8 channels with single receivers, and (c) 24 channels of sEMG data followed by a low-pass 1 Hz filter.

experimenter to follow predefined training movements. Movements spanned the 3 DoFs and their combinations lasting in total less than 150 s. Participants were allowed to retake the training up to three times until they were able to follow the path correctly. After training, participants were allowed around 5 min of familiarization before they were presented and asked to reach 28 different targets. Targets (Table II) were defined in a normalized space (from -1 to 1 for maximum flexion to extension and pronation to supination, and from 0 to 1 for hand opening to closing) presented in the same order for all participants, starting from those requiring movements along only 1 DoF, then 2 simultaneous DoFs and lastly all 3 DoFs. Between targets, the participants had to return to the center neutral position. To successfully complete a target, the participant had to stay within a normalized +-10% range of the target for 7 consecutive samples (dwelling time of approximately 500ms). If the target was not reached within 30 s, the trial was considered not successful. The completion rate (percentage of successful trials) was measured, and from the successful trials the time to completion was also calculated.

### G. Statistical Analysis

Statistical analysis was performed to compare the $R^2$ results. In all factors tested at least one group did not follow a normal distribution (Shapiro-Wilk test with $p<0.05$). Therefore, the non-parametric Kruskal Wallis H test with factors being the movement combinations was used. When each factor comprised more than two groups, Mann-Whitney U-tests for post-hoc multiple comparison, with Bonferroni correction, were used. The n and p values are reported on each graph with the significance level set as $p < 0.05$. The p values are reported on the graphs up to 3 decimal points with values lower than 0.001 written as $p < 0.001$.

## III. Results

Fig. 4 shows three examples of the motion capture measured angles and the predicted ones by the regressors. The first, (Fig. 4a) showcases the results obtained using all 24 channels with our multiple receiver approach. The second, (Fig. 4b) showcases an emulation of a system closer to what has been previously presented in the literature [15] using only 8 channels and single receivers. The last, (Fig. 4c) shows the results using sEMG. These examples highlight that the system is able to predict the actual act of flexion and extension (initial portion until around 15s) and also follow the smaller flexion movements caused when abducting the wrist (from around 15s to 25s) as well as keeping the angle steady while the hand opens and close (end portion from around 50s to 60s).

### A. Comparison US vs sEMG

In Fig. 5, we compare the overall results between our US A-mode approach, a simulation of the performance of a US system based on the literature, and sEMG. Overall, higher $R^2$ values were obtained with both US approaches when compared to the sEMG. This difference was observed for all the movement sets tested, which indicates that US could outperform sEMG both on single DoF as well as simultaneous multi-DoF tasks. Between both US approaches, our proposed method had higher $R^2$ values when compared to the lower channel count and with the single receiver method. All factors showed a significant ($p < 0.001$) difference between each group.

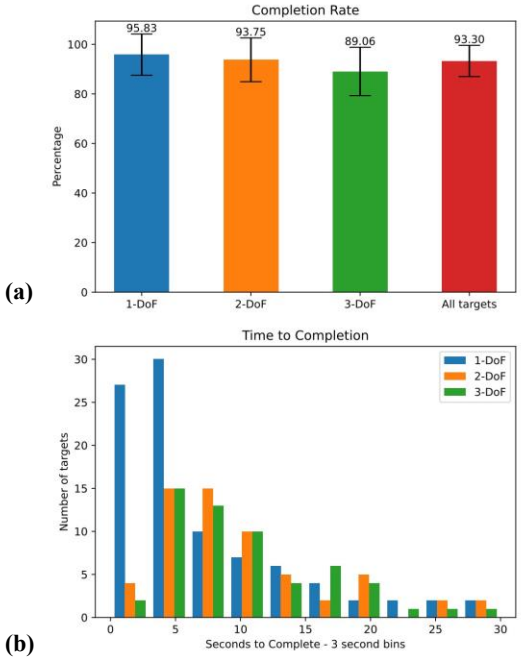


Fig. 7. Performance metrics for the online validation. (a) Completion rate for 1-DoF, 2-DoF, 3-DoF, and all targets. (b) Time to completion for all successful trials.

### B. Comparison of US acquisition methods

The impact of the proposed multiple receiver approach against the conventional single receiver method on performance is shown in Fig. 6a. The multiple receiver approach led to a $R^2$ distribution with significantly higher values ($p < 0.001$ for Simple and All, and $p = 0.007$ and 0.002 for Sequential and Simultaneous, respectively) than with the single receiver method.

Fig. 6b evaluates the change in performance on both approaches by artificially reducing the number of available channels. For each $N$, from one to 23 channels, 10 random combinations of $N$ channels were tested. For $N = 24$ all channels were used. Increasing the channel count had an impact on performance up to $N$=18, after which the results were not significantly different anymore ($p > 0.05$).

Lastly, we evaluated whether having two independent circles of transducers improved the performance. This was done by comparing the $R^2$ distributions obtained with the 2 different models being trained on each circle and one being trained on both. The distributions for each of these were similar with statistical significance difference only in the "All" movement set ($p = 0.01$ and 0.0004).

### C. Online US Validation

Fig. 7 displays the metrics calculated for the online validation. Fig. 7a shows and compares the average completion rate for all participants over 1-DoF ($95.8 \pm 8.3\%$), 2-DoF ($93.7 \pm 8.8\%$), 3-DoF ($89. \pm 9.7\%$), and all targets

($93.3 \pm 6.2\%$). Fig. 7b shows the histogram of the time to completion for all successful trials.

## IV. Discussion

In this study we built and validated an HMI hardware interface and software pipeline that uses A-mode US signals to estimate the kinematics of the hand and wrist. Leveraging on a larger number of transducers than previous US-based HMIs and a novel recording methodology that uses information from all transducers on each transmission, we achieved improvements in the offline decoding of movements compared to previously proposed US- and EMG-based approaches. Furthermore, we presented a full validation of this novel approach in real-time for a 3-DoF pointer movement task.

Current work on HMIs using upper limb muscle activation mainly uses EMG and shows important limitations. Nonetheless, sEMG has been used in successful product implementations for prostheses (e.g., Michelangelo hand by Ottobock [26]) and remote control (e.g., MyoArmband by Thalmic Labs [27]). However, EMG signal quality is influenced by the low-pass filtering and attenuation effect of the tissues and its intrinsic variability [4], while also having low spatial resolution.

US on the other hand naturally avoids some of the pitfalls of the sEMG signal, while still being non-invasive. Being artificially elicited, US signals are highly stable over time and provide deep penetration in the tissues with good spatial resolution. Many previous works demonstrated the use of B-mode imaging US for HMI [6, 7, 10, 28, 29]. B-mode allows for higher lateral resolution and for the generation of 2D images while A-mode works on individual lines of lower lateral resolution where echoes between each line do not have a clear spatial relation to each other. A recent review [30] found similar results for A- and B-mode US used for interfacing but pointed that the possibility to detect the movement on neighboring muscles could be an advantage of B-mode. Using B-mode, recent studies even show the feasibility of identifying single motor unit activity from US [31], as a new approach to movement intention decoding.

While B-mode US may have some advantages over A-mode, the path to implement B-mode as a wearable interface is still unclear. Currently, B-mode probes are bulky and hard to secure to the arm [30, 32]. Recent systems include flexible arrays for long term monitoring [33, 34], but the miniaturization of the electronics required for image acquisition and processing is still a challenge. Lastly, the B-mode results have sometimes showed issues due to their limited field of view [6].

Therefore, A-mode is currently a better candidate solution for a human interface, as it has more potential for wearability [32, 35] and, using sufficient channels can record, with redundancy, information from all around the forearm.

### A. Offline Validation

Our results showed that the A-mode approach for HMIs is capable of outperforming sEMG-based decoding. Fig. 4 shows that even without a low-pass filter after the regressor, the predictions from the US model are more stable than with the sEMG model. Even with a 1-Hz low-pass filter the sEMG predictions are very variable, as seen in previous studies [36, 37]. This result supports the argument that the US-extracted morphological information is directly suited for HMI applications.

Our sEMG results (Fig. 5) are in agreement with previous results in the literature. Using sEMG, Bao *et al*. [37] reported an average $R^2$ of 0.66 for two DoFs using linear regression while Zhao *et al.* [38] reported values over 0.9 using a musculoskeletal model. Other works reported values from 0.6 to 0.8 using intramuscular EMG [39] and high-density sEMG [24].

Using A-mode US recordings, most previous works focused on classification [13, 19, 40]. Regression has been less used, with Guo *et al.* [12] achieving $R^2$ values of approximately 0.9 for one DoF wrist extension, and, more recently, Yang et al. reporting $R^2$ values over 0.95 for wrist rotation and hand closing [15, 41]. Our results indicate similar $R^2$ values but for up to four DoFs, thus expanding previous studies on US-based HMIs.

Improved performance of our approach was achieved by a higher channel count and the proposed multiple-receiver approach (Fig. 5). By simulating a system closer to those used in previous papers [15], we verified that both the channel count and the multiple-receiver methods significantly contributed to the improvement in performance. Therefore, a lower transducer count and single receiver approach cannot properly extract all the morphological information from the forearm muscle activation. This could be due to a lack of coverage of all muscle tissues.

We evaluated the performance change by using all channels with either the single and multiple receiver approach (Fig. 6a) and identified superior performance for the multiple receiving method. This implies that the US lines acquired by transducers other than the one transmitting contain information that is not present in the direct backscatters. Thus, the implementation of a multiple receiver strategy is suggested as a method to improve system performance without much additional complexity.

The performance impact of the number of transducers was also evaluated by artificially limiting the number of transducers (Fig. 6b). As expected, the increase in performance from adding an extra transducer became progressively smaller as the number of transducers increased [42, 43]. However, significant increases in accuracy were still found by using up to 18 channels. While this result is likely to be variable between systems with different transducers, it points that work using A-mode US would benefit from using more transducers than what was implemented in previous studies.

Lastly, we hope that our validation of simultaneous recording of US and motion capture can inspire future studies to proceed in this direction. While our application of motion capture required manual post-processing of the data for joint angle acquisition it may be possible to perform this process automatically and in real time. This would open the possibility

of acquiring extensive labeled data in a simplified process.

### B. Online Validation

Proportional simultaneous position control of multiple DoFs in a pointer interface is a notably hard task. Using sEMG, online control can be achieved via velocity control using classification of gestures and a proportional component resulting in completions rates of 96% [25] and 98% [44]. Simultaneous position control provides more natural usage but is considerably harder as it requires constant correct predictions of all DoFs and does not allow for control via sequential activation of the DoFs. With position control using sEMG, Hahne et al. [45] used a linear regressor to perform functional tasks, and Nowak et al. [24] used high-density sEMG and a ridge regressor to perform a 3-DoF TAC test with reported success rates of 92%, 62% and 37% for targets requiring simultaneous activation of 1-DoF, 2-DoF and 3-DoF. Our US results (Fig. 7) of 95%, 93% and 89% for 1-,2-, and 3-DoFs show that US can achieve similar high performances for 1-DoF targets and suffer little performance degradation with increase in the required simultaneous activation of DoFs.

Online control has been preliminarily evaluated over 1-DoF in B-mode US works [8, 46]. In A-mode works, classification based works have been reported [18] with a classification-based velocity control approach for a 3-DoF TAC test conducted with perfect completion rate [14]. Tests have also been done in online classification to control a prosthesis [47], and in integration between US and sEMG [48]. On online position control, Yang. et al. [15] achieved a 97% completion rate on a 2-DoF TAC test. Our work expands these results with a completion rate of 93% for 3-DoF tasks.

Completion time for all participants was overall low, with approximately 83% of the targets being completed in less than half of the available time (15 s). While not easily comparable due to different experimental setups, dwelling time, range of motion and selected targets our average completion time of $8.9 \pm 6.5$ s is close to previously reported values (e.g., $15.61 \pm 4.12$ s [14] and $4.66 \pm 0.91$ s [15]).

### C. Limitations and future work

While the work developed marks a novel combination of hardware and software with promising applications in HMI, there are some limitations of note. The bracelet prototype design, while flexible, did not physically allow the usage of all available channels in all participants. It also requires application of US gel that reduces the usability for long periods of time as the gel dries and needs reapplication. US gel also reduces the friction between the bracelet and skin and sometimes may cause the sensors to slip during usage. Future work should investigate more stable designs that can adapt to different forearm circumferences as well as substitute the US gel for a more stable interface on the long term [44]. Also, the current version of the US acquisition is considerably portable but not wearable and has multiple cables that can obstruct natural movements of the user. Design work should also be dedicated on the optimization of the transducer themselves with special considerations on optimal center frequency and opening angle.

The offline performance comparison between US and sEMG was performed on two distinct groups of participants. Ideally, both should have been recorded simultaneously, however the two systems do not physically fit in the forearm. On a similar note, our simulation of the results of other published applications of A-mode US is limited as systems are inherently unique and use different transducers, sampling rates and processing pipelines.

On the online side, further work should use the setup developed for the TAC as part of actual workflows. Use cases such as control of mechanical or virtual limbs are a clear extension of the concept and would help validating its performance in realistic scenarios.

## V. Conclusion

In conclusion, we have developed a portable A-mode US bracelet setup that, coupled with a processing pipeline, can simultaneously predict the wrist and hand kinematics for up to 4 DoFs. In doing so, we validated the applicability of US as a solution for upper limb HMIs. We also evaluated the performance of the system against sEMG and proved that the inclusion of higher US channel count and a multiple receiver approach improves the robustness of the system. Lastly, we validated the system in real-time control where participants showed remarkable performance on a 3-DoF TAC test, expanding previous results.

## References


[1] A. D. Roche, B. Lakey, I. Mendez, I. Vujaklija, D. Farina, and O. C. Aszmann, "Clinical Perspectives in Upper Limb Prostheses: An Update," *Current Surgery Reports*, vol. 7, no. 3, p. 5, 2019/02/27 2019, doi: 10.1007/s40137-019-0227-z.

[2] I. Vujaklija, D. Farina, and O. C. Aszmann, "New developments in prosthetic arm systems," (in eng), no. 1179-1462 (Electronic).

[3] F. Mereu, F. Leone, C. Gentile, F. Cordella, E. Gruppioni, and L. Zollo, "Control Strategies and Performance Assessment of Upper-Limb TMR Prostheses: A Review," (in eng), *Sensors (Basel)*, vol. 21, no. 6, Mar 10 2021, doi: 10.3390/s21061953.

[4] E. Scheme and K. Englehart, "Electromyogram pattern recognition for control of powered upper-limb prostheses: state of the art and challenges for clinical use," (in eng), *J Rehabil Res Dev*, vol. 48, no. 6, pp. 643-59, 2011, doi: 10.1682/jrrd.2010.09.0177.

[5] Y. P. Zheng, J. Chan Mm Fau - Shi, X. Shi J Fau - Chen, Q. H. Chen X Fau - Huang, and Q. H. Huang, "Sonomyography: monitoring morphological changes of forearm muscles in actions with the feasibility for the control of powered prosthesis," (in eng), *Med Eng Phys*, no. 1350-4533 (Print), 2006.

[6] J. McIntosh, A. Marzo, M. Fraser, and C. Phillips, "EchoFlex: Hand Gesture Recognition using Ultrasound Imaging," presented at the Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, Denver, Colorado, USA, 2017. [Online]. Available: https://doi.org/10.1145/3025453.3025807.

[7] C. Castellini and D. S. Gonzalez, "Ultrasound imaging as a human-machine interface in a realistic scenario," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 3-7 Nov. 2013 2013, pp. 1486-1492, doi: 10.1109/IROS.2013.6696545.

[8] A. S. Dhawan *et al.*, "Proprioceptive Sonomyographic Control: A novel method for intuitive and proportional control of multiple degrees-of-freedom for individuals with upper extremity limb loss," *Scientific Reports*, vol. 9, no. 1, p. 9499, 2019/07/01 2019, doi: 10.1038/s41598-019-45459-7.

[9] S. Engdahl, A. Dhawan, G. Lévay, A. Bashatah, R. Kaliki, and S. Sikdar, "Motion prediction using electromyography and sonomyography for an individual with transhumeral limb loss," *medRxiv*, p. 2020.12.23.20248489, 2020, doi: 10.1101/2020.12.23.20248489.

[10] Y. Huang, X. Yang, Y. Li, D. Zhou, K. He, and H. Liu, "Ultrasound-Based Sensing Models for Finger Motion Classification," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 5, pp. 1395-1405, 2018, doi: 10.1109/JBHI.2017.2766249.

[11] N. Hettiarachchi, Z. Ju, and H. Liu, "A New Wearable Ultrasound Muscle Activity Sensing System for Dexterous Prosthetic Control," in *2015 IEEE International Conference on Systems, Man, and Cybernetics*, 9-12 Oct. 2015 2015, pp. 1415-1420, doi: 10.1109/SMC.2015.251.

[12] J. Y. Guo, Y. P. Zheng, Q. H. Huang, and X. Chen, "Dynamic monitoring of forearm muscles using one-dimensional sonomyography system," (in eng), *J Rehabil Res Dev*, vol. 45, no. 1, pp. 187-95, 2008, doi: 10.1682/jrrd.2007.02.0026.

[13] 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," (in eng), *Prosthet Orthot Int*, vol. 37, no. 1, pp. 43-9, Feb 2013, doi: 10.1177/0309364612446652.

[14] X. Yang, Z. Chen, N. Hettiarachchi, J. Yan, and H. Liu, "A Wearable Ultrasound System for Sensing Muscular Morphological Deformations," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 51, no. 6, pp. 3370-3379, 2021, doi: 10.1109/TSMC.2019.2924984.

[15] X. Yang, J. Yan, Z. Yin, and H. Liu, "Sonomyographic Prosthetic Interaction: Online Simultaneous and Proportional Control of Wrist and Hand Motions Using Semisupervised Learning," *IEEE/ASME Transactions on Mechatronics*, pp. 1-10, 2022, doi: 10.1109/TMECH.2022.3207359.

[16] M. Fournelle *et al.*, "Portable Ultrasound Research System for Use in Automated Bladder Monitoring with Machine-Learning-Based Segmentation," *Sensors*, vol. 21, p. 6481, 09/28 2021, doi: 10.3390/s21196481.

[17] X. Sun, X. Yang, X. Zhu, and H. Liu, "Dual-Frequency Ultrasound Transducers for the Detection of Morphological Changes of Deep-Layered Muscles," *IEEE Sensors Journal*, vol. 18, no. 4, pp. 1373-1383, 2018, doi: 10.1109/JSEN.2017.2778243.

[18] X. Yang, X. Sun, D. Zhou, Y. Li, and H. Liu, "Towards Wearable A-Mode Ultrasound Sensing for Real-Time Finger Motion Recognition," (in eng), *IEEE Trans Neural Syst Rehabil Eng*, vol. 26, no. 6, pp. 1199-1208, Jun 2018, doi: 10.1109/tnsre.2018.2829913.

[19] Y. Li, K. He, X. Sun, and H. Liu, "Human-machine interface based on multi-channel single-element ultrasound transducers: A preliminary study," in *2016 IEEE 18th International Conference on e-Health Networking, Applications and Services (Healthcom)*, 14-16 Sept. 2016 2016, pp. 1-6, doi: 10.1109/HealthCom.2016.7749483.

[20] D. Winter, "Biomechanics and Motor Control of Human Movement, Fourth Edition," 09/01 2009, doi: 10.1002/9780470549148.ch5.

[21] I. Carpinella, J. Jonsdottir, and M. Ferrarin, "Multi-finger coordination in healthy subjects and stroke patients: a mathematical modelling approach," *Journal of NeuroEngineering and Rehabilitation*, vol. 8, no. 1, p. 19, 2011/04/20 2011, doi: 10.1186/1743-0003-8-19.

[22] D. Y. Hasbani Mh Fau - Barsakcioglu, M. K. Barsakcioglu Dy Fau - Jung, D. Jung Mk Fau - Farina, and D. Farina, "Simultaneous and proportional control of wrist and hand degrees of freedom with kinematic prediction models from high-density EMG," (in eng), no. 2694-0604 (Electronic).

[23] A. M. Simon, L. J. Hargrove, B. A. Lock, and T. A. Kuiken, "Target Achievement Control Test: evaluating real-time myoelectric pattern-recognition control of multifunctional upper-limb prostheses," (in eng), *J Rehabil Res Dev*, vol. 48, no. 6, pp. 619-27, 2011, doi: 10.1682/jrrd.2010.08.0149.

[24] M. Nowak, I. Vujaklija, A. Sturma, C. Castellini, and D. Farina, "Simultaneous and Proportional Real-Time Myocontrol of Up to Three Degrees of Freedom of the Wrist and Hand," (in eng), *IEEE Trans Biomed Eng*, vol. 70, no. 2, pp. 459-469, Feb 2023, doi: 10.1109/tbme.2022.3194104.

[25] E. J. Scheme and K. B. Englehart, "Validation of a Selective Ensemble-Based Classification Scheme for Myoelectric Control Using a Three-Dimensional Fitts' Law Test," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 21, no. 4, pp. 616-623, 2013, doi: 10.1109/TNSRE.2012.2226189.

[26] "Michelangelo hand | The Michelangelo hand helps you regain your freedom."

https://www.ottobock.com/en-gb/product/8E500 (accessed 13/03/2023.

[27] "Thalmic Labs – Official Open Repository." https://github.com/thalmiclabs (accessed 13/03/2023.

[28] S. Engdahl *et al.*, "Classification Performance and Feature Space Characteristics in Individuals With Upper Limb Loss Using Sonomyography," (in eng), *IEEE J Transl Eng Health Med*, vol. 10, p. 2100311, 2022, doi: 10.1109/jtehm.2022.3140973.

[29] J. Shi, J.-Y. Guo, S.-X. Hu, and Y.-P. Zheng, "Recognition of Finger Flexion Motion from Ultrasound Image: A Feasibility Study," *Ultrasound in Medicine & Biology*, vol. 38, no. 10, pp. 1695-1704, 2012/10/01/ 2012, doi: https://doi.org/10.1016/j.ultrasmedbio.2012.04.021.

[30] V. Nazari and Y.-P. Zheng, "Controlling Upper Limb Prostheses Using Sonomyography (SMG): A Review," *Sensors*, vol. 23, no. 4, doi: 10.3390/s23041885.

[31] E. Lubel, B. Grandi Sgambato, D. Y. Barsakcioglu, J. Ibáñez, M. X. Tang, and D. Farina, "Kinematics of individual muscle units in natural contractions measuredin vivousing ultrafast ultrasound," (in eng), *J Neural Eng*, vol. 19, no. 5, Sep 6 2022, doi: 10.1088/1741-2552/ac8c6c.

[32] C. Castellini, G. Passig, and E. Zarka, "Using ultrasound images of the forearm to predict finger positions," (in eng), *IEEE Trans Neural Syst Rehabil Eng*, vol. 20, no. 6, pp. 788-97, Nov 2012, doi: 10.1109/tnsre.2012.2207916.

[33] H. Hu *et al.*, "A wearable cardiac ultrasound imager," *Nature*, vol. 613, no. 7945, pp. 667-675, 2023/01/01 2023, doi: 10.1038/s41586-022-05498-z.

[34] C. Wang *et al.*, "Bioadhesive ultrasound for long-term continuous imaging of diverse organs," *Science*, vol. 377, no. 6605, pp. 517-523, 2022, doi: doi:10.1126/science.abo2542.

[35] S. Frey, S. Vostrikov, L. Benini, and A. Cossettini, "WULPUS: a Wearable Ultra Low-Power Ultrasound probe for multi-day monitoring of carotid artery and muscle activity," in *2022 IEEE International Ultrasonics Symposium (IUS)*, 10-13 Oct. 2022 2022, pp. 1-4, doi: 10.1109/IUS54386.2022.9958156.

[36] S. Starpornchaisit, Y. Kim, A. Takagi, N. Yoshimura, and Y. Koike, "Finger Angle Estimation From Array EMG System Using Linear Regression Model With Independent Component Analysis," (in English), *Frontiers in Neurorobotics*, Original Research vol. 13, 2019-September-26 2019, doi: 10.3389/fnbot.2019.00075.

[37] T. Bao, A. Zaidi, S. Xie, and Z. Zhang, "Surface-EMG based Wrist Kinematics Estimation using Convolutional Neural Network," in *2019 IEEE 16th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, 19-22 May 2019 2019, pp. 1-4, doi: 10.1109/BSN.2019.8771100.

[38] Y. Zhao, Z. Zhang, Z. Li, Z. Yang, A. A. Dehghani-Sanij, and S. Xie, "An EMG-Driven Musculoskeletal Model for Estimating Continuous Wrist Motion," (in eng), *IEEE Trans Neural Syst Rehabil Eng*, vol. 28, no. 12, pp. 3113-3120, Dec 2020, doi: 10.1109/tnsre.2020.3038051.

[39] L. H. Smith, T. A. Kuiken, and L. J. Hargrove, "Evaluation of Linear Regression Simultaneous Myoelectric Control Using Intramuscular EMG," (in eng), *IEEE Trans Biomed Eng*, vol. 63, no. 4, pp. 737-46, Apr 2016, doi: 10.1109/tbme.2015.2469741.

[40] J. He, H. Luo, J. Jia, J. T. W. Yeow, and N. Jiang, "Wrist and Finger Gesture Recognition With Single-Element Ultrasound Signals: A Comparison With Single-Channel Surface Electromyogram," (in eng), *IEEE Trans Biomed Eng*, vol. 66, no. 5, pp. 1277-1284, May 2019, doi: 10.1109/tbme.2018.2872593.

[41] X. Yang, Y. Zhou, and H. Liu, "Wearable Ultrasound-Based Decoding of Simultaneous Wrist/Hand Kinematics," *IEEE Transactions on Industrial Electronics*, vol. 68, no. 9, pp. 8667-8675, 2021, doi: 10.1109/TIE.2020.3020037.

[42] A. Fernandes, Y. Ono, and E. Ukwatta, "Evaluation of Finger Flexion Classification at Reduced Lateral Spatial Resolutions of Ultrasound," *IEEE Access*, vol. PP, pp. 1-1, 02/02 2021, doi: 10.1109/ACCESS.2021.3056353.

[43] A. T. Kamatham, M. Alzamani, A. Dockum, S. Sikdar, and B. Mukherjee, "Sparse Sonomyography-Based Estimation of Isometric Force: A Comparison of Methods and Features," *IEEE Transactions on Medical Robotics and Bionics*, vol. 4, no. 3, pp. 821-829, 2022, doi: 10.1109/TMRB.2022.3172680.

[44] L. H. Smith, T. A. Kuiken, and L. J. Hargrove, "Use of probabilistic weights to enhance linear regression myoelectric control," (in eng), *J Neural Eng*, vol. 12, no. 6, p. 066030, Dec 2015, doi: 10.1088/1741-2560/12/6/066030.

[45] J. M. Hahne, M. A. Schweisfurth, M. Koppe, and D. Farina, "Simultaneous control of multiple functions of bionic hand prostheses: Performance and robustness in end users," (in eng), *Sci Robot*, vol. 3, no. 19, Jun 20 2018, doi: 10.1126/scirobotics.aat3630.

[46] N. Akhlaghi *et al.*, "Real-Time Classification of Hand Motions Using Ultrasound Imaging of Forearm Muscles," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 8, pp. 1687-1698, 2016, doi: 10.1109/TBME.2015.2498124.

[47] Z. Yin *et al.*, "A Wearable Ultrasound Interface for Prosthetic Hand Control," (in eng), *IEEE J Biomed Health Inform*, vol. 26, no. 11, pp. 5384-5393, Nov 2022, doi: 10.1109/jbhi.2022.3203084.

[48] X. Yang, J. Yan, and H. Liu, "Comparative Analysis of Wearable A-Mode Ultrasound and sEMG for Muscle-Computer Interface," *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 9, pp. 2434-2442, 2020, doi: 10.1109/TBME.2019.2962499.