BeamBand: Hand Gesture Sensing with Ultrasonic Beamforming

Yasha Iravantchi   **Mayank Goel**   **Chris Harrison**

Carnegie Mellon University, Human-Computer Interaction Institute

5000 Forbes Avenue, Pittsburgh, PA 15213

{ysi, mayank, chris.harrison}@cs.cmu.edu
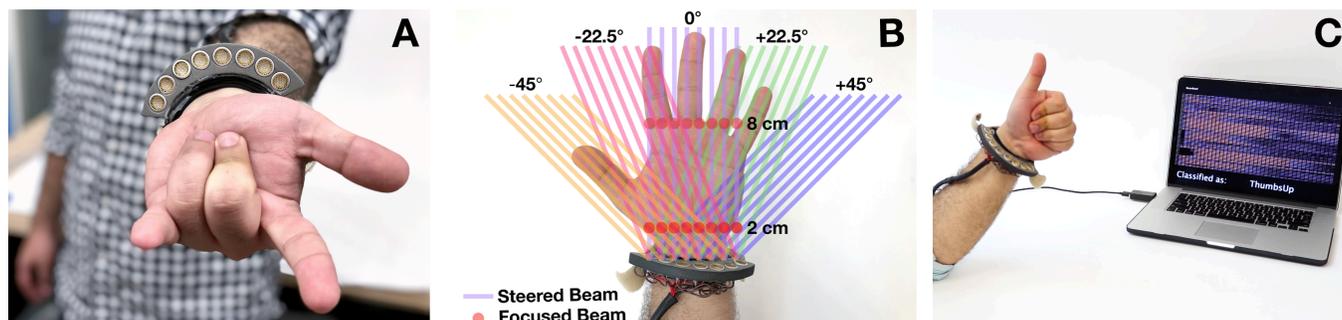
Figure 1. Beamband is a wrist worn sensor containing a transducer array (A) that uses beamforming to direct and focus ultrasound at areas of interest (B) in order to recognize a wide variety of hand gestures (C).

ABSTRACT

BeamBand is a wrist-worn system that uses ultrasonic beamforming for hand gesture sensing. Using an array of small transducers, arranged on the wrist, we can ensemble acoustic wavefronts to project acoustic energy at specified angles and focal lengths. This allows us to interrogate the surface geometry of the hand with inaudible sound in a raster-scan-like manner, from multiple viewpoints. We use the resulting, characteristic reflections to recognize hand pose at 8 FPS. In our user study, we found that BeamBand supports a six-class hand gesture set at 94.6% accuracy. Even across sessions, when the sensor is removed and reworn later, accuracy remains high: 89.4%. We describe our software and hardware, and future avenues for integration into devices such as smartwatches and VR controllers.

CCS CONCEPTS

Human-centered computing → Human computer interaction (HCI) → Interaction techniques → Gestural input

KEYWORDS

Hand Input; Hand Gesture; Acoustic Reflectometry; Acoustic Beamforming; Acoustic; Interaction Techniques; Wearables

1   INTRODUCTION

Robust hand gesture detection holds the promise to enrich user interfaces and improve immersiveness, whether it be smartwatches to AR/VR systems. Unfortunately, identifying hand gestures without instrumenting the hand (e.g., gloves, controllers) has proven to be challenging, which motivates the need to identify new methods. Prior research includes leveraging electromyography [38][39], bio-acoustics [23] [15], electrical impedance tomography [50][51], contour sensing [7], and worn cameras [20]. While each approach has its strengths and drawbacks, a common weakness is robust accuracy across users and worn sessions.

In this paper, we present our work on BeamBand, a new approach for worn hand gesture sensing, which leverages acoustic beamforming. We use small in-air ultrasonic transducers arranged along the contour of the wrist (Figure 1A), which offers a stable vantage point from which to capture hand pose. Using active beamforming, we steer and focus ultrasound towards areas of interest on the hand (Figure 1B). We also multiplex our transducers, capturing beamformed reflections from slightly different viewpoints, offering rich signals for machine-learning-driven hand gesture recognition (Figure 1C).

To assess BeamBand's recognition performance, we conducted a ten-participant study, adopting two gesture sets from the literature in order to enable direct comparison (i.e., rather than developing a custom set). The first set contained seven hand poses, while the second set has six gestures along three axes of rotation. On these two gesture sets, BeamBand demonstrates accuracies of 92.5% and 94.6% respectively. More unique is that accuracy remains high – 87.7% – in sessions after the band is removed and reworn.

2 RELATED WORK

First, we review prior work that intersects with our application area of gesture recognition. We then move to work using acoustic reflectometry, with a particular focus on the HCI literature. Finally, we discuss beamforming more specifically, as this is our main technical approach, and review the few systems that have employed it in the HCI domain.

2.1 Hand Gesture Sensing

Robust sensing of the pose and movement of the hands has been a long-standing goal in HCI. The most immediate approach is to instrument the hands directly, with, for example, gloves containing accelerometers [34][43], strain gauges [24] and capacitive sensors [37]. These methods typically place the sensors in locations well-suited for their gesture tasks. For example, Perng *et al.* [34] place the accelerometers at the fingertips for finger interactions such as pointing and which finger is raised. Whitmire *et al.* [48] use conductive fabrics as a capacitive sensor to detect finger and thumb interactions.

Slightly less conspicuous and invasive are systems that attempt to sense the hand from the wrist or arm. BeamBand falls into this category. One of the most popular approaches use optical sensors to detect hand geometrical changes that occur when a user performs a hand gesture. For example, WristWhirl uses an array of infrared proximity sensors to detect the angle of the hand with respect to the wrist [13]. Another optical approach uses a camera to observe hand gestures and reconstruct a 3D model of the hand [20]. The camera may also be mounted on a head mounted display [6]. There is also a significant body of research that leverages arm contour changes, using pressure sensors [7][18], infrared sensors [10][13][29][47], and capacitive sensors [37].

Apart from querying the external state of the hands, people have investigated using signals from inside the body to determine hand state. A very prevalent approach is Electromyography (EMG) [19][38][39][41], which passively detects electrical signals sourced from skeletal muscle movements. Active sensing has also been explored, as seen in Electrical Impedance Tomography [50][51], which has been used to sense changes in the interior structure of the arm for hand gesture sensing.

Most related to BeamBand are the approaches that use acoustic signals. For example, Amento *et al.* [1], Hambone [8], Skinput [15], and Tactile Teacher [16] place passive acoustic sensors on the skin to listen to micro-vibrations resulted from finger taps, flicks, and pinches for detection. More recently, research has shown that off-the-shelf smartwatches can also detect these signals [23][32][49][52]. Way et. [46] offers an excellent survey of wrist worn sensing approaches (including acoustic). We cover active acoustic approaches in the next section.

2.2 Acoustic Reflectometry in HCI

BeamBand is built on the principle of ultrasonic reflectometry, which examines objects of interests by emitting structured acoustic waves and measuring reflected signals. The time of flight of sounds can be used to infer the distance of objects, which is the most basic information that can be acquired. One example is single-emitter sonar, which has been in use for roughly a century in marine applications, and also echolocation, which animals have used for considerably longer. In addition to time of flight, the amplitude of reflections (including non-linear damping of different frequencies) and multipath effects can also reveal facets of the environment (e.g., material properties, room geometry).

In the HCI literature, acoustic reflectometry is most commonly encountered in the form of low-cost sonar sensors, used for range-finding. For example, "Sound of Touch" [31] and "FingerPing" [53] both use in-body sonar to detect hand gestures. Using in-air sonar sensors, Point Upon Body [25] detects touch input on the user's arm. Measuring the Doppler shift of reflections has been used to detect the direction of hand gestures [3] and swipes on the forearm [31] (see [36] for a survey of ultrasonic doppler sensing in HCI).

2.3 Acoustic Beamforming

Beamforming can be achieved in any transmission medium. However, it is most commonly applied to radio waves (*e.g.*, radar [21], wireless communication [12]) and sound (*e.g.*, medical ultrasound [11]). When multiple wavefronts are created, signals experience constructive and destructive interference, which can be used to *form* controlled *beams* of energy, hence the technique name. See Figure 2 and Video Figure for a concise visual primer (and [4][12][21][44] for more comprehensive background). Beamforming can also be used in reverse (*i.e.*, inverse beamforming) [30], using an array of passive receivers to *e.g.*, localize voices in a room [2] or finger snaps [14].

Most similar to BeamBand in operation are multi-emitter/receiver towed sonar arrays [22]. In single-emitter sonar (regardless of the number of receivers), the first object encountered will typically reflect the largest signal. However, with multiple emitters, it is possible to have coordinated beamforming "pings" concentrate energy on an area of interest at varying distances. This is similar to medical ultrasound [11], which uses beamforming to focus acoustic energy at a particular depth in the body, and then essentially raster scans to produce an 2D interior image (which was used in EchoFlex [27][28] for hand gesture sensing). Of course, both of these examples cost many thousands of dollars and require liquid or gel to interface to the sensed medium. These methods operate using MHz-range ultrasound are confined to operate in-body. BeamBand utilizes lower frequency 40 kHz ultrasound, which can effectively propagate through air and interact with surfaces without the use

of an interfacing medium. Beamforming has also been used for haptics [5][26] and in-air levitation [17] in the HCI literature.

3   SIMULATIONS AND MEASUREMENTS

Prior to developing our system, we wanted to gain a better understanding of how beamforming operates both in computer-based simulations and physical measurements. We first built a series of simulations in software where we changed the relative phase of 7 evenly spaced emitters (i.e., 13 mm interval) outputting 40 kHz waves with 8 mm wavelength, which changes the angle and focal point of the wavefront (Figure 2, top).

To verify our theoretical model, we ran physical measurements, which also takes account for interactions such as transducer impedance mismatches, multipath interference, and environmental noise. Similar to the software simulations, we changed the relative phase of the emitters to create 5 distinct angles (-45, -22.5, 0, +22.5, +45) and 3 distinct focal lengths (infinite, 2cm, 8cm). To match the setup in our software simulations, we built a linear array of 7 evenly spaced 40 kHz transducers [35] with 13 mm spacing. These transducers were used to emit structured waveforms, which was measured by a same transducer (i.e., sensing mode) attached to a CNC gantry.

We moved this gantry along a 4mm grid (half of the wavelength in air) within a 12.4cm x 12.4cm square. At each point on the grid, the linear transducer array would generate a beam at a selected angle or focal length and the sensor would capture the acoustic interaction at the location. The sensor would move to the next point in the grid, the linear array would repeat the same beam, the sensor would again record, and move until measurements were collected from

all locations. All of the waveforms captured could then be synchronously replayed to visualize the wavefronts (Figure 2 and Video Figure). We found that our software and physical models matched, with our physical models in particular allowing us to tune parameters for optimal beamforming.

During these measurements, we also tested many different ultrasonic transducers, power ratings, physical sizes, and beam widths. Our first requirement was to select only transducers that could handle being driven by a 100 V_{pp} signal (discussed in detail in next section). Of those transducers, we selected the one with the best signal recovery. We spaced two of the same transducers 1 cm apart facing each other and measured the amplitude of the received signal. We found a 12.8 mm, 40 kHz, 30 V_{RMS}, and 70° beam width [35] offered the best SNR.

4   IMPLEMENTATION

BeamBand consists of three main components. First is our custom sensor board (Figure 3), which generates, captures, and processes ultrasonic signals. Next is a wristband (Figure 1A), which contains ultrasonic transducers in a wrist-mounted array that emits and receives signals. The total cost of our proof-of-concept hardware was $220. Finally, we have a laptop-based software that receives data from the hardware and performs further processing and machine learning. We now describe these elements in greater detail.

4.1   Sensor Board and Transducers

We used eight 40 kHz in-air ultrasonic piezoelectric transducers [35], same as the ones we used in the physical measurements. To minimize the effects of multipath and randomly scattered acoustic energy, we fire a single strong pulse using 7 transducers each with a structured phase
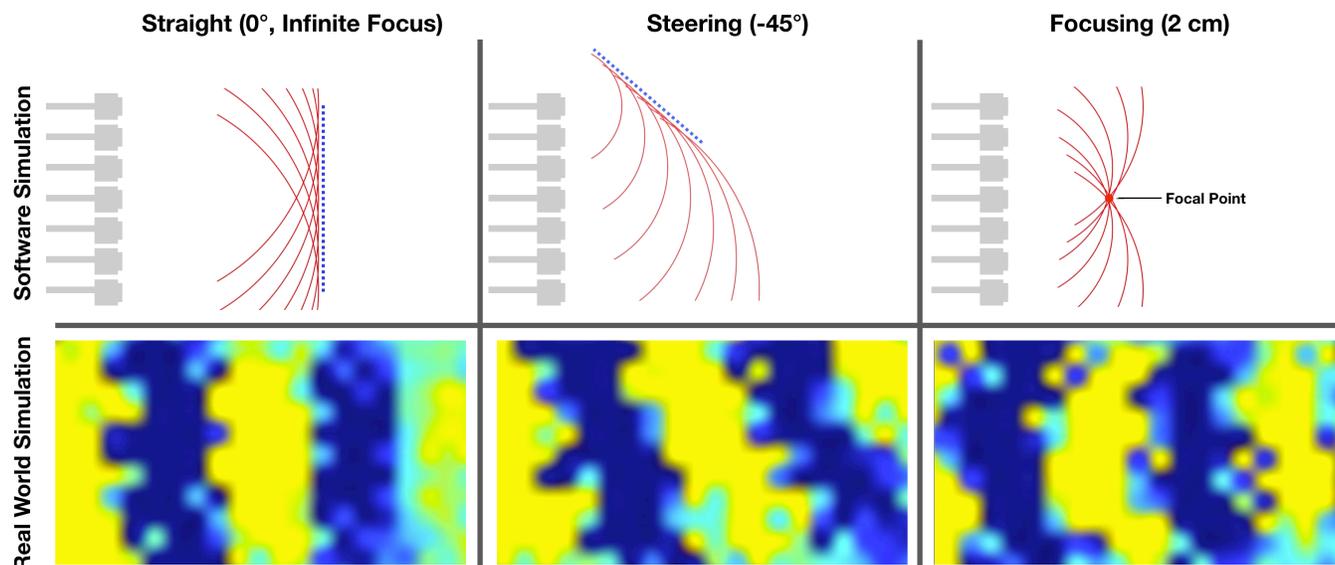

Figure 2. We performed software and physical simulations of ultrasonic propagation in air as seen from above. Yellow denotes the high energy density, blue denotes low energy density. See also Video Figure.

shift. To drive these transducers to emit software-controlled waveforms, we built a custom sensor circuit (Figure 3), which features three main components – a high voltage EMCO SIP100 DC-DC power regulator [9], high voltage amplifiers, and a multiplexed analog frontend. A Teensy 3.6 was used to control the sensor circuit [40], which we overclocked to 240 MHz.

We configure the microcontroller to toggle its digital pins, to generate a 3.3 V_{pp} 40 kHz square wave signal. This signal is amplified to 100 V_{pp} to drive the transducers. In order to perform accurate beamforming we need to keep the timing of firings between transducers equal To minimize latency, we write directly into the I/O map register on the microcontroller. This allows us to toggle all the pins on the same register simultaneously using only a single clock cycle (*i.e.*, 4.17 ns), enabling up to 8 perfectly synchronized channels. This also allows a granularity of ~0.1° for phase-shifting the signals with respect to each other and ~0.01° granularity for beam steering. To generate more robust signals and minimize switching overhead, we isolate each transducer with its own amplifier.

To receive signals, we select the one spared transducer to act as a receiver. During the firing sequence of all other transducers, we clamp this transducer to ground, which prevents the transducer from resonating due to direct acoustic coupling and electrical noise. Once the firing sequence is complete, we disconnect the clamp and connect the transducer to our analog frontend. We then pass the signal through an active high pass filter with fixed gain (f_c=39 kHz, G=5) with an additional amplification stage with adjustable gain up to 40X. The amplified signal is then DC biased to $V_{ADC}/2$ and sampled by the microcontroller's 16-bit ADC at 667 kHz with a 1-sample on-ADC average, yielding a true sampling rate of 333 kHz. This allows us to reduce noise resulting from sampling at a high resolution at high speeds without increasing latency or compute from the microcontroller. All captured waveform data is transmitted to a laptop over USB for further computation.

4.2 Power Consumption

We did not optimize the power consumption of our proof-of-concept hardware, which is powered by 5V via its USB connection. Nonetheless, we did measure current draw: ~400mA total, which is below the 500mA limit for USB. Of the total current draw, 250mA is from our overclocked Teeny 3.6 board (100mA when not overclocked). Our DC-DC converter consumes ~140mA, most of which is conversion loss. All other components, including our transducers, consume ~10mA.

4.3 Wrist-Worn Band

As seen in Figure 1, we fabricated a band that could be worn on the arm, but primarily placed at the wrist. We placed eight transducers in a horseshoe arrangement, following
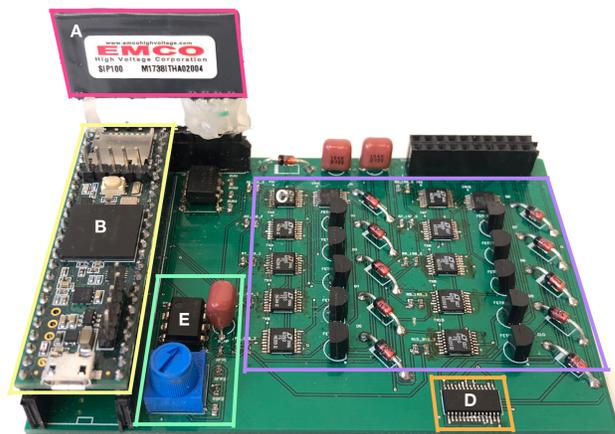
Figure 3. The custom sensor board for *BeamBand*. A) DC-DC converter, B) Teensy 3.6, C) high voltage amplifiers, D) multiplexer, and E) filter and amplification stage.

the contour of the arm. These operate 1 cm above the surface of the skin. The band is made of EVA foam [42] to allow for greater conformity and to reduce acoustic couplings between transducers. An adjustable elastic band is used to affix the sensor to the wrist. We chose not to include any transducers for the back of the hand, as fingers generally articulate inwards, and thus we found signals to not be particularly interesting (other than to capture wrist angle, which is readily captured by palm-side transducers). It is worth noting that this arrangement is slightly different than our physical simulations; we re-ran our physical simulations with the horseshoe arrangement and saw a slight degradation in the coherence/resolution of the beamforming. However, we consider the ergonomic benefits of the horseshoe arrangement to outweigh this minor effect.

4.4 Beamforming

We generate wavefronts at five angles (-45°, -22.5°, 0°, +22.5°, +45°), illustrated in Figure 1B, that cover the typical range of wrist motion. We also focus at three distances (Figure 1B). First is at 2 cm, which roughly correlates to the base of the palm. Our second focal point is at 8 cm, which is roughly at the base of the fingers. Finally, we emit a synchronous waveform (same as 0° angular beamforming), which is equivalent to infinite focus, to capture more distant features, such as finger tips. Thus, in total, each firing sequence, there are 7 unique beamformed patterns.

4.5 Acoustic Viewpoints & Waveforms

Of note that we have eight total transducers, and at any given time, seven act as transmitters and one as a receiver. The sensor circuit controls the transducers to cycle through all transmitter-receiver combinations which results in 8 configurations. At each configuration, we measure all 7 beamforming modes sequentially before moving to the next configuration. At each beamforming mode, we collect 500
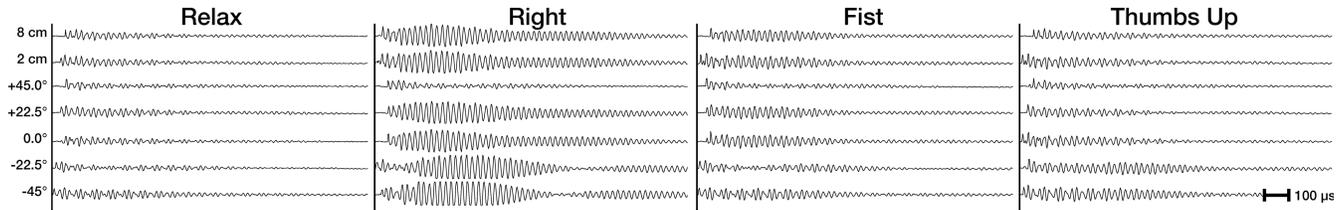
Figure 4. Raw waveforms from a subset (1st receiver, thumb side) of the 56 total waveforms across different gestures.

samples of the reflected waveform with a 333 kHz sampling rate (*i.e.*, 3 μs period). This procedure yields 56 collected waveforms (8 configurations × 7 modes) which we assemble into a single sensing frame, as illustrated in Figure 1B.

4.6 Framerate

Each pattern requires 0.5 ms to generate and emit followed by a 1.5 ms data collection period. Thus, each full cycle (all 56 waveforms) of beamforming generation, emission, and data capture takes 112 ms ($0.5 \times 56 + 1.5 \times 56$). This results in approximately 8 frames per second.

4.7 Features and Machine Learning

Our machine learning pipeline converts the 56 incoming waveforms captured by our hardware into features. We bin each waveform into 20 bins and take the standard deviation of each bin as a feature, which yields 1,120 total features forming a frame. For discrete classification and all of our evaluations, we use Scikit-learn's Random Forest Classifier (default parameters, 500 trees) [33]. All tasks were performed on a standard configuration 2013 MacBook Pro 15".

5 GESTURE SET

Rather than invent a custom gesture set, we purposely chose to adopt two gesture sets from the literature [18][50] to reduce design bias and enable direct comparison between systems. We note that most prior work creates custom gesture sets that work well with their technique. This, unfortunately, precludes direct comparison.

We adopted the hand gesture set defined in Tomo [50]. These seven gestures (relax, six "hand" gestures) are depicted in Figure 5 (green underscore). For future reference, we will refer to this gesture set as the "Tomo" set. We also adopted the hand gesture set defined in [18], which extends or flexes the hand along three different axes (two wrist axes, and one finger axes). These six gestures are depicted in Figure 5 (purple underscore). We refer to this gesture set

as "6-axis" in later text. Note these two gesture sets have four common gestures, *Right = Wrist Flexion*, *Left = Wrist Extension*, *Fist = Finger Flexion*, and *Relax = Finger Extension*.

6 EVALUATION

In this study, we evaluate the gesture classification performance of BeamBand. We recruited 10 participants (4 female, mean age 25), which had a mean wrist diameter of 5.5 cm (SD=0.8). The study took approximately one hour to complete and paid $20.

6.1 Procedure

The study consisted of participants wearing the BeamBand on their non-dominant wrist (i.e., like a watch). All of our participants were right handed, so the BeamBand was worn on the left wrist. A single round of data collection consisted of each gesture being performed once, in a random order. Each gesture took roughly one second to complete, during which time 10 sensor frames were recorded. A session consisted of ten rounds of data collection. To add variety and realism, we collected two sessions of data for each user, with the worn sensor being removed in between. This procedure yielded 18,000 sensor frames (10 sensor frames × 9 gestures × 10 rounds × 2 sessions × 10 users).

6.2 Within-Session Accuracy

To simulate the performance of gesture recognition when the system is calibrated when first worn, we performed a leave-one-round-out cross validation, where we trained on nine rounds within a session and tested on the tenth (all combinations). We repeated this for both sessions independently and averaged the results.

In the full, nine-class combined gesture set, the average within-session accuracy across all participants was 90.2% (SD=3.7). In the Tomo gesture set, the average within-session accuracy was 92.5% (SD=2.2). In the 6-axis gesture the

Figure 5. Our two gesture sets: A) Relax/Finger Extension B) Fist/Finger Flexion C) Right/Wrist Flexion D) Left/Wrist Extension E) Stretch F) Thumbs Up G) Spider Man H) Radial Deflection I) Ulnar Reflection
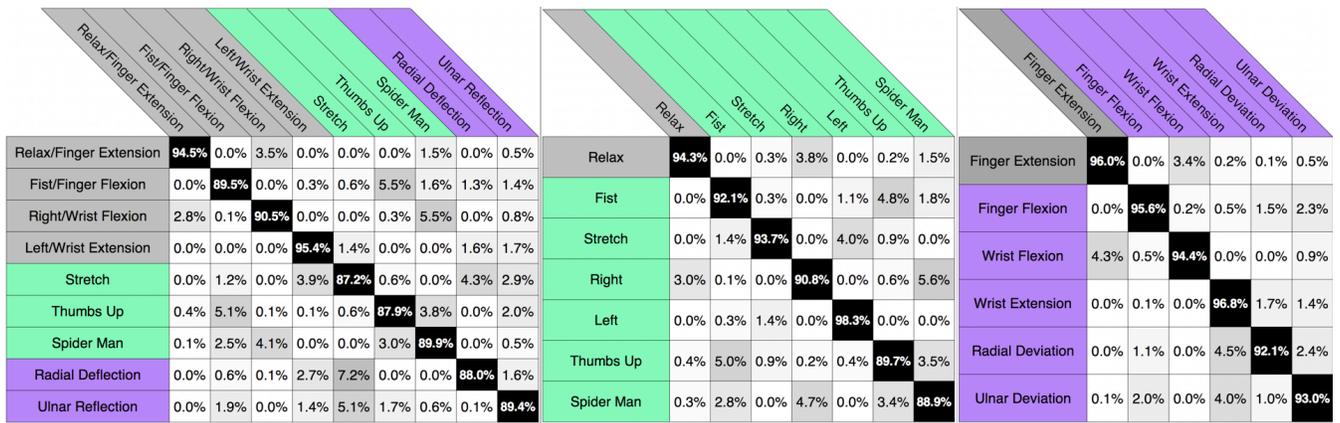
Figure 6. Confusion matrices (within-session accuracies) for the combined gesture set (mean accuracy 90.2%), Tomo gesture set (mean 92.5%), and 6-axis gesture set on the arm (mean 94.6%).

average within-session accuracy was 94.6% (SD=3.4). In each of those gesture sets, there was not a significant outlier in gesture performance. Interestingly, we noticed that there was confusion between similar gestures involving closing the hand, such as *Fist* and *Thumbs Up*, which accounted for 15.2% of the total error in the hand gesture set. The confusion matrices can be found in Figure 6.

6.3 Across-Session Accuracy

One significant challenge for on-body systems is their ability to retain classification performance across worn sessions. To evaluate the drop in performance after BeamBand is reworn, we ran a leave-one-session-out cross validation for each of our participants, where we train on all data from session one and test on all data from session two, and vice versa, combining the results. In the full, nine-class combined gesture set, the average across-session accuracy across all participants was 81.4% (SD=15.9). In the Tomo gesture set, the average across-session accuracy was 86.0% (SD-12.7). In the 6-axis gesture set, the average across-session accuracy was 89.4% (SD=10.9). Some gestures retained robust across-session accuracy, with *Left* and *Wrist Flexion* performing at 94.2% and 96.2%, respectively. We saw a sim-

ilar confusion between *Fist* and *Thumbs Up*, which accounted for 9.1% of the total error in the hand gesture set. The confusion matrices can be found in Figure 7.

6.4 Across-User Accuracy

Another significant challenge for on-body systems is their ability to classify gestures across participant, where there is no guarantee that the gesture will be performed similarly to another participant. In evaluating the across-user accuracy, we ran a leave-one-user-out cross validation for each of our participants, where we train on all the data across both sessions from nine participants and test on both sessions from a tenth participant, all combinations. In the full, nine-class combined gesture set, the average across-user accuracy across all participants was 44.2% (SD=8.8). In the Tomo gesture set, the average across-user accuracy was 51.7% (SD=10.4). In the 6-axis gesture set, the average across-user accuracy was 63.2% (SD=8.5). Some gestures appear to be more consistent across users, such as *Wrist Flexion* and *Radial Deviation*, which were performing at 80.1% and 79.2%, respectively.
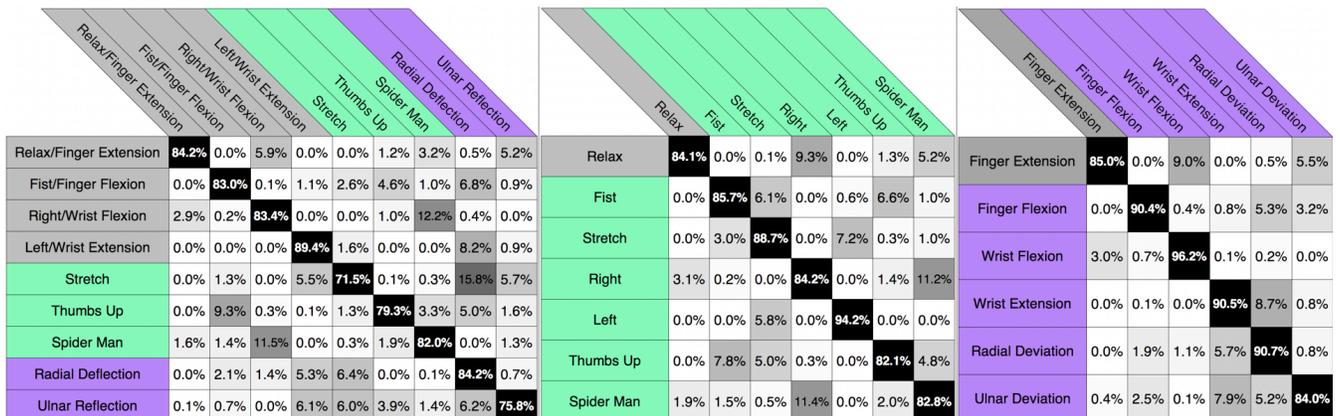

Figure 7. Confusion matrices (across-session accuracies) for the combined gestures set (mean accuracy 81.4%), Tomo gesture set (mean 86.0%), and 6-axis gesture set on the arm (mean 89.4%).

6.5 Comparison to Prior Results

Our *within-session* results are similar to the two systems from which we drew our gesture sets. Within session, Jung *et al.* [18] reports 95.4% accuracy across six gestures, while Tomo [50] on the wrist achieves accuracies of 96.6% across seven gestures. On these, BeamBand achieves 92.5% and 94.6% respectively. When the gesture sets are merged (nine classes), BeamBand is 90.2% accurate.

For further contextualize our results, our system also performs comparably when compared to other systems with their own gesture sets. Most notably, SensIR reports 93.3% accuracy across 12 gestures [29], while zSense provides 94.8% accuracy across 9 gestures [47]. Further, Mime achieves ~95% on 4 gestures [6]. Note that none of these systems evaluate across-session or across-user accuracy.

Few systems evaluate *across-session* accuracy, which is particularly challenging for on-body sensing systems. Tomo reports cross-session accuracies of 65.3% across seven gestures. On the same gesture set, BeamBand achieves 86.0%. Jung *et al.* does not report cross-session accuracy, but for reference, BeamBand achieves 89.4% accuracy on its gesture set.

Rarest are systems that evaluate *across-user* accuracy (except for worn computer vision systems, which tend to be robust). Tomo reports cross-user accuracies of 38.8% on the wrist across seven gestures, while BeamBand achieves 51.7% on the same set. We could not find any other points of comparison in the literature.

6.6 Robustness to Sleeve Occlusion

Unlike light, ultrasound can pass through thin fabrics. We found in development that we could roll our sleeves down over the sensor and train the system occluded with minimal impact on accuracy. In order to measure robustness to sleeve occlusion, we placed two identical transducers, facing each other, 8 cm apart. We drove one transducer using a function generator (40 kHz, 10 V_{pp}) while the other was connected to an oscilloscope. We then draped various fabrics over the transmitting transducer to simulate sleeve occlusion. We tested 11 different fabrics of different thicknesses and weave density (Figure 8). We found that while

Material	Signal Strength	Knit Density	Thickness (mm)
No Fabric (Uncovered)	100.0%	-	-
100% Nylon Wide Knit Lace	98.2%	Low	0.2
100% Nylon Tight Knit Lace	96.4%	Low	0.17
50% Polyester, 25% Rayon, 25% Cotton Shirt	74.5%	Medium	0.75
100% Wool Knit Sweater	73.6%	Low	1.5
90% Polyester, 10% Elastane Women UA Shirt	59.1%	High	0.93
60% Rayon, 40% Polyester Shirt	58.2%	Medium	0.66
100% Cotton Shirt	57.5%	Medium	0.83
90% Polyester, 10% Elastane Men UA Shirt	55.5%	High	1
100% Cotton Oxford Weave Shirt	54.5%	Medium	0.45
100% Cotton Flannel Shirt	49.1%	Medium	0.59
100% Polyester Dress Shirt	25.5%	High	0.41

Figure 8. Signal strength (normalized to the "No Fabric" condition) for various common sleeve materials.

thickness does seem to have a correlation to the amount of signal attenuation, a more significant factor was the knit density of the fabric (*i.e.*, while the polyester dress shirt was thinner than most of our fabrics, it performed the worst). Further, one of our better performing materials (knit wool sweater) was our thickest while being of low knit density. We believe these results present a promising starting point for future work exploring occluded sensing.

7 STRENGTHS & WEAKNESSES

While BeamBand is competitive with prior systems, it is not yet sufficiently accurate for e.g., a consumer device. However, as a proof of concept, the technical approach looks promising. In order to achieve out-of-the-box classification abilities, more work is required to develop a generalizable model. Collecting more data across a wide range of participants may improve the classification robustness. There may also be merit in moving away from classical machine learning methods to a deep learning model. We also suspect the addition of a calibration stage that "homes" the orientation of the wristband could raise across-session and across-user accuracies.

Another avenue for future work is exploring different frequencies of ultrasound. Ultrasonic transducers running at 40 kHz are ubiquitous (and thus inexpensive) but are almost certainly not the optimal frequency for gesture recognition (a wavelength of ~8mm is too large). Higher frequencies could enable superior sensing of fine-grained motions and gestures, though at the cost of higher signal attenuation in air, which would have to be overcome with a higher drive voltage or more sensitive analog frontend.

While BeamBand has shown promising results, BeamBand only operates in two dimensions (*x*-direction, *y*-direction) along the plane of the palm. BeamBand currently cannot steer the beam in the axis normal to the palm. Steering the beam along *z*-axis would require building a two-dimensional transducer array. Such a "3D" BeamBand might be able to better discern similar gestures, such as *Fist* and *Thumbs Up*, which is a challenge when only scanning in 2D.

As power consumption was not optimized for our prototype, it requires our device to be physically tethered for power. In a commercial implementation, beamforming patterns would be pre-generated and specialized hardware (*e.g.*, ASICs) would drive the sensing process – dramatically more efficient. Using a general-purpose microcontroller was to facilitate research and rapid prototyping. Reducing the sensing duty cycle, running at full frame rate only when a change is detected would also improve power consumption. However, the sensing principle itself is fairly power efficient; the transducers themselves require virtually no power to drive.

In conjunction with power consumption, refresh rate could be improved as noted above by pre-generating waveforms. Each full cycle of beamforming generation/emission/data capture (56 combinations) takes 112 ms. 84 ms are for data capture, with the rest mostly waveform generation. Pre-generating waveforms would increase the frame rate to ~12 Hz (a 50% increase). Also, as can be seen in Figure 4, most signal returns within 0.8 ms. Reducing per-combination data capture from 1.5 ms results in a frame rate of ~24 Hz (a 300% increase). Further optimizations may include time multiplexing the emissions such that one combination is in-flight while another is returning.

There are some physical limitations of our current prototype. First, we need to offset the transducers in to get over the bump at the base of the palm. This offset prohibits the placement of the transducer to be at the level of the skin. Another limitation is the size of the transducer. While the transducers we used are housed in a 12 mm casing, the size of the piezo-elements inside are ~5mm in diameter – much more reasonable for integration. Also, transducers are not restricted to circular shape. For example, medical ultrasound utilizes small square elements arranged in a strip.

Finally, we believe there is no inherent reason why BeamBand could not be integrated into a standalone device. As noted earlier, there are several avenues to reduce power requirements and improve frame rate. Additionally, custom piezo-elements can be made very small. We envision Beam-Band may sit behind an acoustically-transparent plastic window on the side of smartwatches, similar to those medical ultrasound wands mentioned.

8 CONCLUSION

We have presented BeamBand, a novel worn sensing method that uses ultrasonic beamforming for on-body hand gesture recognition. BeamBand projects ultrasonic wavefronts at different angles on the user's hand, and measures waves reflected back to the band. We evaluated two gesture sets sourced from the literature and our user study results show promising accuracies, in both within-session and across-session. We hope our effort will act as a catalyst for deeper investigation into ultrasonic beamforming for enabling novel interactions.

ACKNOWELDGEMENTS

REFERENCES

[1] Brian Amento, Will Hill, and Loren Terveen. 2002. The sound of one hand: a wrist-mounted bio-acoustic fingertip gesture interface. In *CHI '02 Extended Abstracts on Human Factors in Computing Systems* (CHI EA '02). ACM, New York, NY, USA, 724-725. DOI=http://dx.doi.org/10.1145/506443.506566

[2] Shoko Araki, Hiroshi Sawada, and Shoji Makino, 2007, April. Blind speech separation in a meeting situation with maximum SNR beamformers. In *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on* (Vol. 1, pp. I-41). IEEE.

[3] Adeola Bannis, Pei Zhang, and Shijia Pan. 2014. Adding directional context to gestures using doppler effect. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication (UbiComp '14 Adjunct).* ACM, New York, NY, USA, 5-8. DOI: https://doi.org/10.1145/2638728.2638774

[4] E. H. Brandt (2001). Acoustic physics: Suspended by Sound. Nature, 413(6855), 474-475.

[5] Tom Carter, Sue Ann Seah, Benjamin Long, Bruce Drinkwater, and Sriram Subramanian. UltraHaptics: multi-point mid-air haptic feedback for touch surfaces. In *Proceedings of the 26th annual ACM symposium on User interface software and technology*, pp. 505-514. ACM, 2013.

[6] Andrea Colaço, Ahmed Kirmani, Hye Soo Yang, Nan-Wei Gong, Chris Schmandt, and Vivek K. Goyal. 2013. Mime: compact, low power 3D gesture sensing for interaction with head mounted displays. In Proceedings of the 26th annual ACM symposium on User interface software and technology(UIST '13). ACM, New York, NY, USA, 227-236. DOI: https://doi.org/10.1145/2501988.2502042

[7] Artem Dementyev and Joseph A. Paradiso. 2014. WristFlex: low-power gesture input with wrist-worn pressure sensors. In *Proceedings of the 27th annual ACM symposium on User interface software and technology (UIST '14).* ACM, New York, NY, USA, 161-166. DOI: https://doi.org/10.1145/2642918.2647396

[8] Travis Deyle, Szabolcs Palinko, Erika Shehan Poole, and Thad Starner. 2007. Hambone: A Bio-Acoustic Gesture Interface. In *Proceedings of the 2007 11th IEEE International Symposium on Wearable Computers (ISWC '07).* IEEE Computer Society, Washington, DC, USA, 1-8. DOI: https://doi.org/10.1109/ISWC.2007.4373768

[9] EMCO SIP100 DC-DC Converter, http://www.eie-ic.com/Images/EMCO/EMCO/sipseries.pdf

[10] Rui Fukui, Masahiko Watanabe, Tomoaki Gyota, Masamichi Shimosaka, and Tomomasa Sato. 2011. Hand shape classification with a wrist contour sensor: development of a prototype device. In *Proceedings of the 13th international conference on Ubiquitous computing (UbiComp '11).* ACM, New York, NY, USA, 311-314. DOI: https://doi.org/10.1145/2030112.2030154

[11] Reli Hershkovitz, Eyal Sheiner, and Moshe Mazor. "Ultrasound in obstetrics: a review of safety." European Journal of Obstetrics & Gynecology and Reproductive Biology101, no. 1 (2002): 15-18.

[12] Mohammad-Hossein Golbon-Haghighi, 2016. Beamforming in Wireless Networks. In Tech Open. http://cdn.intechopen.com/pdfs-wm/53332.pdf

[13] Jun Gong, Xing-Dong Yang, and Pourang Irani. 2016. WristWhirl: One-handed Continuous Smartwatch Input using Wrist Gestures. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology (UIST '16).* ACM, New York, NY, USA, 861-872. DOI: https://doi.org/10.1145/2984511.2984563

[14] Teng Han, Khalad Hasan, Keisuke Nakamura, Randy Gomez, and Pourang Irani. 2017. SoundCraft: Enabling Spatial Interactions on Smartwatches using Hand Generated Acoustics. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology (UIST '17).* ACM, New York, NY, USA, 579-591. DOI: https://doi.org/10.1145/3126594.3126612

[15] Chris Harrison, Desney Tan, and Dan Morris. Skinput: appropriating the body as an input surface. In *Proceedings of the SIGCHI conference on human factors in computing systems* (CHI 2010), pp. 453-462. ACM, 2010.

[16] Chih-Pin Hsiao, Richard Li, Xinyan Yan, and Ellen Yi-Luen Do. 2015. Tactile Teacher: Sensing Finger Tapping in Piano Playing. In *Proceedings of the Ninth International Conference on Tangible, Embedded, and Embodied Interaction (TEI '15)*. ACM, New York, NY, USA, 257-260. DOI: https://doi.org/10.1145/2677199.2680554

[17] Takeshi Ide, James Friend, Kentaro Nakamura, and Sadayuki Ueha. A non-contact linear bearing and actuator via ultrasonic levitation. *Sensors and Actuators* A: Physical 135, no. 2 (2007): 740-747.

[18] Pyeong-Gook Jung, Gukchan Lim, Seonghyok Kim, and Kyoungchul Kong. A wearable gesture recognition device for detecting muscular activities based on air-pressure sensors. *IEEE Transactions on Industrial Informatics* 11, no. 2 (2015): 485-494.

[19] Frederic Kerber, Michael Puhl, and Antonio Krüger. 2017. User-independent real-time hand gesture recognition based on surface electromyography. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '17)*. ACM, New York, NY, USA, Article 36, 7 pages. DOI: https://doi.org/10.1145/3098279.3098553

[20] David Kim, Otmar Hilliges, Shahram Izadi, Alex D. Butler, Jiawen Chen, Iason Oikonomidis, and Patrick Olivier. 2012. Digits: freehand 3D interactions anywhere using a wrist-worn gloveless sensor. In *Proceedings of the 25th annual ACM symposium on User interface software and technology (UIST '12)*. ACM, New York, NY, USA, 167-176. DOI: https://doi.org/10.1145/2380116.2380139

[21] Hamid Krim and Mats Viberg, 1996. Two decades of array signal processing research: the parametric approach. *IEEE signal processing magazine, 13*(4), pp.67-94.

[22] Carlo Kopp (December 2009). "Identification underwaterwith towed array sonar" (PDF). *Defence Today*. pp. 32–33. http://www.ausairpower.net/SP/DT-TAS-Dec-2009.pdf

[23] Gierad Laput, Robert Xiao, and Chris Harrison. 2016. ViBand: High-Fidelity Bio-Acoustic Sensing Using Commodity Smartwatch Accelerometers. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology (UIST '16)*. ACM, New York, NY, USA, 321-333. DOI: https://doi.org/10.1145/2984511.2984582

[24] Jhe-Wei Lin, Chiuan Wang, Yi Yao Huang, Kuan-Ting Chou, Hsuan-Yu Chen, Wei-Luan Tseng, and Mike Y. Chen. 2015. BackHand: Sensing Hand Gestures via Back of the Hand. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology (UIST '15)*. ACM, New York, NY, USA, 557-564. DOI: https://doi.org/10.1145/2807442.2807462

[25] Shu-Yang Lin, Chao-Huai Su, Kai-Yin Cheng, Rong-Hao Liang, Tzu-Hao Kuo, and Bing-Yu Chen. 2011. Pub - point upon body: exploring eyes-free interaction and methods on an arm. In *Proceedings of the 24th annual ACM symposium on User interface software and technology (UIST '11)*. ACM, New York, NY, USA, 481-488. DOI: https://doi.org/10.1145/2047196.2047259

[26] Asier Marzo, Richard McGeehan, Jess McIntosh, Sue Ann Seah, and Sriram Subramanian. 2015. Ghost Touch: Turning Surfaces into Interactive Tangible Canvases with Focused Ultrasound. In *Proceedings of the 2015 International Conference on Interactive Tabletops & Surfaces (ITS '15)*. ACM, New York, NY, USA, 137-140. DOI: https://doi.org/10.1145/2817721.2817727

[27] Jess McIntosh, Asier Marzo, Mike Fraser, and Carol Phillips. 2017. EchoFlex: Hand Gesture Recognition using Ultrasound Imaging. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 1923-1934. DOI: https://doi.org/10.1145/3025453.3025807

[28] Jess McIntosh And Mike Fraser, Improving the feasibility of ultrasonic hand tracking wearables, in Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces, ISS '17, New York, NY, USA, 2017, ACM, pp. 342–347.

[29] Jess McIntosh, Asier Marzo, and Mike Fraser. 2017. SensIR: Detecting Hand Gestures with a Wearable Bracelet using Infrared Transmission and Reflection. In Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology (UIST '17). ACM, New York, NY, USA, 593-597. DOI: https://doi.org/10.1145/3126594.3126604

[30] Ulif Michel. 2006. History of acoustic beamforming. In *Berlin Beamforming Conference, Berlin, Germany, Nov* (pp. 21-22).

[31] Adiyan Mujibiya, Xiang Cao, Desney S. Tan, Dan Morris, Shwetak N. Patel, and Jun Rekimoto. 2013. The sound of touch: on-body touch and gesture sensing based on transdermal ultrasound propagation. In *Proceedings of the 2013 ACM international conference on Interactive tabletops and surfaces (ITS '13)*. ACM, New York, NY, USA, 189-198. DOI: https://doi.org/10.1145/2512349.2512821

[32] Masa Ogata, and Michita Imai. SkinWatch: skin gesture interaction for smart watch. *In Proceedings of the 6th Augmented Human International Conference*, pp. 21-24. ACM, 2015.

[33] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel et al. Scikit-learn: Machine learning in Python. Journal of machine learning research 12, no. Oct (2011): 2825-2830.

[34] John Kangchun Perng, Brian Fisher, Seth Hollar, and Kristofer SJ Pister. Acceleration sensing glove (ASG). In *Wearable Computers,* 1999. Digest of Papers. The Third International Symposium on, pp. 178-180. IEEE, 1999.

[35] PUI Audio 40 kHz Ultrasonic Transducer, http://www.puiaudio.com/pdf/UT-1240K-TT-R.pdf

[36] Bhiksha Raj, Kaustubh Kalgaonkar , Chris Harrison , Paul Dietz, Ultrasonic Doppler Sensing in HCI, *IEEE Pervasive Computing*, v.11 n.2, p.24-29, April 2012. DOI, 10.1109/MPRV.2012.17.

[37] Jun Rekimoto. Gesturewrist and gesturepad: Unobtrusive wearable interaction devices. In *Wearable Computers, 2001. Proceedings. Fifth International Symposium on*, pp. 21-27. IEEE, 2001.

[38] T. Scott Saponas, Desney S. Tan, Dan Morris, and Ravin Balakrishnan. 2008. Demonstrating the feasibility of using forearm electromyography for muscle-computer interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08)*. ACM, New York, NY, USA, 515-524. DOI: https://doi.org/10.1145/1357054.1357138

[39] T. Scott Saponas, Desney S. Tan, Dan Morris, Ravin Balakrishnan, Jim Turner, and James A. Landay. 2009. Enabling always-available input with muscle-computer interfaces. In *Proceedings of the 22nd annual ACM symposium on User interface software and technology (UIST '09)*. ACM, New York, NY, USA, 167-176. DOI: https://doi.org/10.1145/1622176.1622208

[40] Teensy 3.6 Microcontroller, PJRC, https://www.pjrc.com/store/teensy36.html

[41] Thalmic Lab, Inc. http://www.thalmic.com/myo/

[42] The Thickest Foam, HilltopStudio, https://www.amazon.com/HilltopStudio-The-Thickest-Foam/dp/B01N7XDSAN

[43] Hsin-Ruey Tsai, Cheng-Yuan Wu, Lee-Ting Huang, and Yi-Ping Hung. 2016. ThumbRing: private interactions using one-handed thumb motion input on finger segments. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct (MobileHCI '16)*. ACM, New York, NY, USA, 791-798. DOI: https://doi.org/10.1145/2957265.2961859

[44] R. J. Urick *Principles of Underwater Sound*, 3rd edition. (Peninsula Publishing, Los Altos, 1983).

[45] Wei Wang, Lei Xie, and Xun Wang. 2017. Tremor detection using smartphone-based acoustic sensing. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers(UbiComp '17)*. ACM, New York, NY, USA, 309-312. DOI: https://doi.org/10.1145/3123024.3123168

[46] David Way and Joseph Paradiso. 2014. A Usability User Study Concerning Free-Hand Microgesture and Wrist-Worn Sensors. In *Proceedings of the 2014 11th International Conference on Wearable and Implantable Body Sensor Networks (BSN '14)*. IEEE Computer Society, Washington, DC, USA, 138-142. DOI=http://dx.doi.org/10.1109/BSN.2014.32

[47] Anusha Withana, Roshan Peiris, Nipuna Samarasekara, and Suranga Nanayakkara. 2015. zSense: Enabling Shallow Depth Gesture Recognition for Greater Input Expressivity on Smart Wearables. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15). ACM, New York, NY, USA, 3661-3670. DOI: https://doi.org/10.1145/2702123.2702371

[48] Eric Whitmire, Mohit Jain, Divye Jain, Greg Nelson, Ravi Karkar, Shwetak Patel, and Mayank Goel. 2017. DigiTouch: Reconfigurable

Thumb-to-Finger Input and Text Entry on Head-mounted Displays. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 3, Article 113 (September 2017), 21 pages. DOI: https://doi.org/10.1145/3130978

[49] Chao Xu, Parth H. Pathak, and Prasant Mohapatra. Finger-writing with smartwatch: A case for finger and hand gesture recognition using smartwatch. In *Proceedings of the 16th International Workshop on Mobile Computing Systems and Applications*, pp. 9-14. ACM, 2015.

[50] Yang Zhang and Chris Harrison. 2015. Tomo: Wearable, Low-Cost Electrical Impedance Tomography for Hand Gesture Recognition. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology (UIST '15)*. ACM, New York, NY, USA, 167-173. DOI: https://doi.org/10.1145/2807442.2807480

[51] Yang Zhang, Robert Xiao, and Chris Harrison. 2016. Advancing Hand Gesture Recognition with High Resolution Electrical Impedance Tomography. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology (UIST '16)*. ACM, New York, NY, USA, 843-850. DOI: https://doi.org/10.1145/2984511.2984574

[52] Cheng Zhang, AbdelKareem Bedri, Gabriel Reyes, Bailey Bercik, Omer T. Inan, Thad E. Starner, and Gregory D. Abowd. 2016. Tap-Skin: Recognizing On-Skin Input for Smartwatches. In *Proceedings of the 2016 ACM International Conference on Interactive Surfaces and Spaces (ISS '16)*. ACM, New York, NY, USA, 13-22. DOI: https://doi.org/10.1145/2992154.2992187

[53] Cheng Zhang, Qiuyue Xue, Anandghan Waghmare, Ruichen Meng, Sumeet Jain, Yizeng Han, Xinyu Li, Kenneth Cunefare, Thomas Ploetz, Thad Starner, Omer Inan, and Gregory D. Abowd. 2018. FingerPing: Recognizing Fine-grained Hand Poses using Active Acoustic On-body Sensing. *In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Paper 437, 10 pages. DOI: https://doi.org/10.1145/3173574.3174011