

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

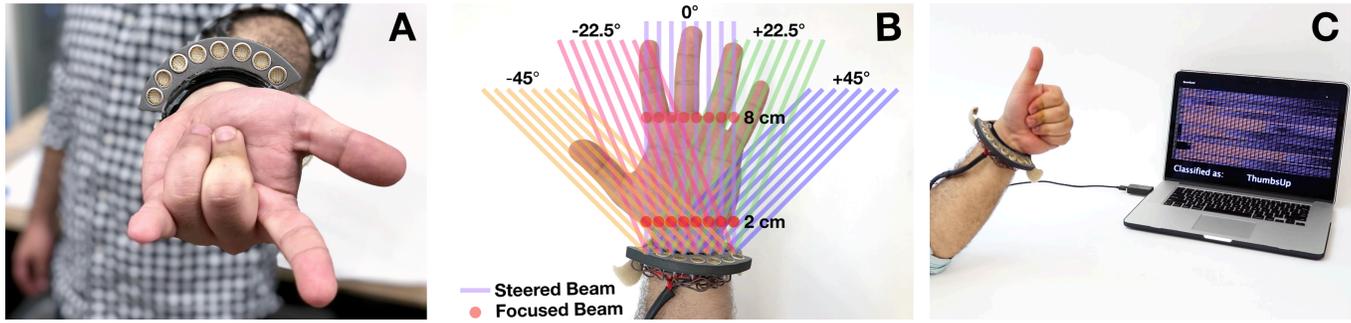


Figure 1. BeamBand is a wrist worn sensor containing eight transducers (A) that uses beamforming to direct and focus ultrasound at areas of interest (B) in order to recognize a variety of hand gestures (C).

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

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CHI 2019, May 4–9, 2019, Glasgow, Scotland, UK.

© 2019 Copyright is held by the owner/author(s). Publication rights licensed to ACM. ACM ISBN 978-1-4503-5970-2/19/05...\$15.00.

DOI: <https://doi.org/10.1145/3290605.3300245>

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 (Figure 1B), 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 – 86.0% and 89.4 respectively – 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 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 [13] uses an array of infrared proximity sensors to detect the angle of the hand with respect to the wrist. 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 [7][18], infrared [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 common approach is Electromyography (EMG) [19][38][39][41], which passively detects electrical signals from muscle contractions. Active sensing has also been explored, as seen in Electrical Impedance Tomography [50][51], which has been used to sense changes in the interior arm structure 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. More recently, research has shown that off-the-shelf smartwatches can also detect these signals [23][32][49][52]. Way *et al.* [46] offers an excellent survey of wrist worn sensing approaches (including acoustic).

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, though 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’s 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 used in maritime applications [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” focus

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 scan to produce a 2D image (which was used in EchoFlex [27][28] for hand gesture sensing). These setups cost many thousands of dollars, use MHz-range ultrasound, and require liquid or gel to couple to the sensed medium. BeamBand utilizes lower frequency 40 kHz ultrasound, which can more easily propagate through air and interact with surfaces without the use of an interfacing medium. Ultrasonic beamforming has also been used for haptics [5][26] and in-air levitation [17] in the HCI literature.

3 PILOT EXPERIMENTS

Prior to developing our system, we sought to gain a better understanding of how beamforming operates in a multi-emitter, airborne setup. We started with simulations in software, changing the relative phase of seven evenly spaced emitters outputting 40 kHz waves (~ 8.5 mm wavelength in room temperature air at 1020 mbar), allowing us to control the angle and focal point of the wavefront (Figure 2, top). To verify our theoretical model, we also ran real-world, physical experiments, which captures more complex interactions such as transducer impedance mismatch, multipath interference, and environmental noise. Our physical transducer array matched our software simulations: seven evenly spaced, 40 kHz transducers. As before, we changed the relative phase of the emitters to create different emission angles and focal lengths.

To capture and visualize ultrasound, we attached an independent transducer to a CNC gantry. We moved this gantry along a 4mm grid within a 12.4×12.4cm square. At each point on the grid, the transducer array would generate a beam at a specified angle or focal length, and the sensor would record the acoustic interaction at that location. The gantry would then move to the next point in the grid, the transducer array would repeat the same emission pattern, and the sensor would again make a recording. This procedure was repeated until all grid locations were recorded. Once complete, all waveforms could then be synchronously replayed to visualize the wavefront propagation (see Figure 2 and Video Figure). We found that our software and physical models generally matched (see examples in Figure 2).

During this stage of development, we also tested many different ultrasonic transducers with various power ratings, physical size, and beam widths. To assess performance, we placed two identical transducers 1cm apart, facing one another. One was driven at 100 V_{pp}, while the other was actuated by the emission. The transducer pair with the highest received signal was inferred to have the best combination of emission efficiency and air impedance match. We ultimately selected [35], a readily available transducer with a 12.8 mm diameter, 40 kHz resonant frequency, and 70° beam width.

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 sensor band,

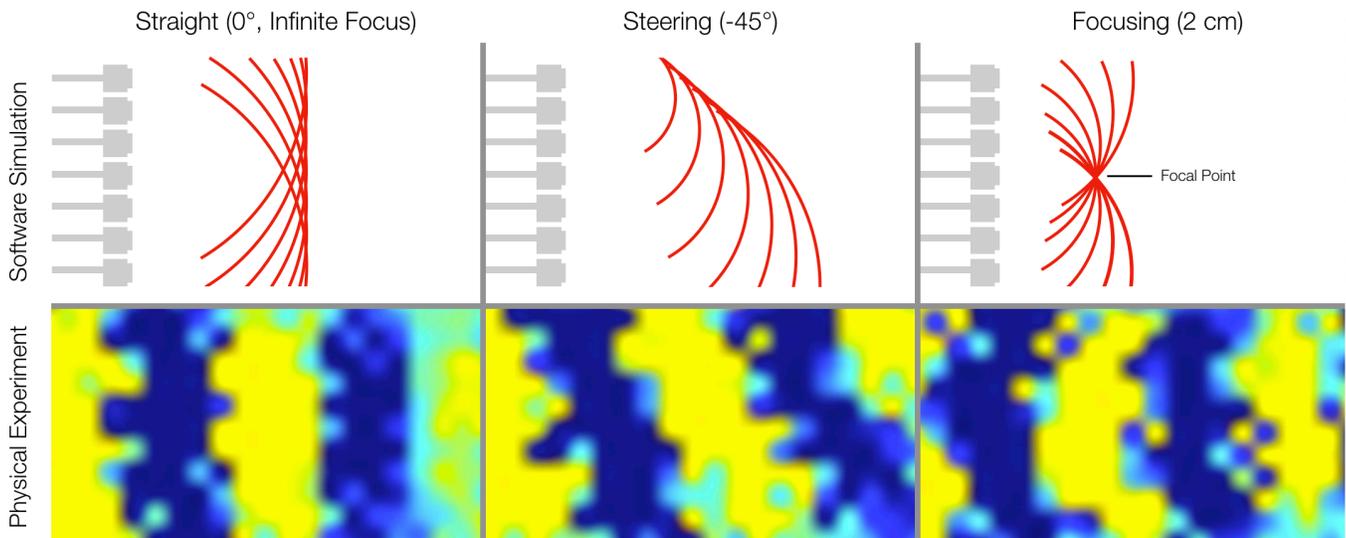


Figure 2. We performed software simulations and physical experiments to better understand ultrasonic beamforming in air (seen here from above, 40 kHz transducers with 13mm on-center spacing). Yellow denotes high acoustic energy, while blue denotes low energy. See also Video Figure.

which contains ultrasonic transducers that emit and receive signals (Figure 1A). The total cost of our proof-of-concept hardware was \$220. Finally, we have 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 use eight 40 kHz in-air ultrasonic piezoelectric transducers [35] (identified in Section 3). A single sensing instance consists of firing a single strong pulse using 7 transducers, each with a specified phase shift. To drive these transducers with software-controlled waveforms, we built a custom sensor circuit (Figure 3), which has 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, generating a 3.3 V_{pp} 40 kHz square wave. This signal is amplified to 100 V_{pp} to drive the transducers. To minimize cross-channel interference and switching overhead, each transducer has a dedicated amplifier. In order to perform accurate beamforming, we need tight control of transducer firing times. To minimize latency, we write directly into the microcontroller’s I/O map register, allowing us to toggle 8 output pins simultaneously in a single clock cycle (4.17 ns). This tight control allows us to manipulate the relative phase of our transducers at a granularity of $\sim 0.1^\circ$.

To capture reflected ultrasound, the one unused transducer is configured to act as a receiver. During the firing

sequence of the other seven transducers, we clamp the receiver transducer to ground, which helps prevent inadvertent actuation due to acoustic coupling and electrical noise. After firing is complete, we disconnect the clamp and connect the receiver 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 333 kHz. 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: ~ 400 mA total. Of the total current draw, 250mA is from our overclocked Teeny 3.6 board (100mA when not overclocked). Our DC-DC converter consumes ~ 140 mA, most of which is conversion loss. All other components, including our transducers, consume ~ 10 mA.

4.3 Transducer Band

As seen in Figure 1, we fabricated a band that could be worn on the arm or wrist. We placed eight transducers in a horseshoe arrangement, following the contour of the arm, and roughly 1cm above the surface of the skin. The band is made of EVA foam [42] to allow for greater conformity and to reduce acoustic coupling between transducers. An adjustable elastic band is used to affix the sensor to the user. We chose not to include any transducers aimed at the back of the hand, as fingers generally articulate inwards. It is worth noting that this arrangement is slightly different than our physical simulations (where the transducers were arranged in a linear array); we re-ran our physical simulations with the horseshoe arrangement and saw a slight degradation in the coherence and resolution of the beamforming. However, we consider the compactness of the horseshoe arrangement to outweigh this minor effect.

4.4 Beamforming

We selected 5 angles for beamforming (-45° , -22.5° , 0° , $+22.5^\circ$, $+45^\circ$), illustrated in Figure 1B, that cover the typical range of finger and wrist motion. We also focus at 3 distances (Figure 1B): 2 cm, which roughly correlates to the base of the palm; 8 cm, which roughly correlates to the base of the fingers; and infinite focus to capture more distant features, such as finger tips. Infinite focus and 0° are the same beamforming pattern, so thus in total, each sensing round consists of 7 unique beamformed emissions.



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.

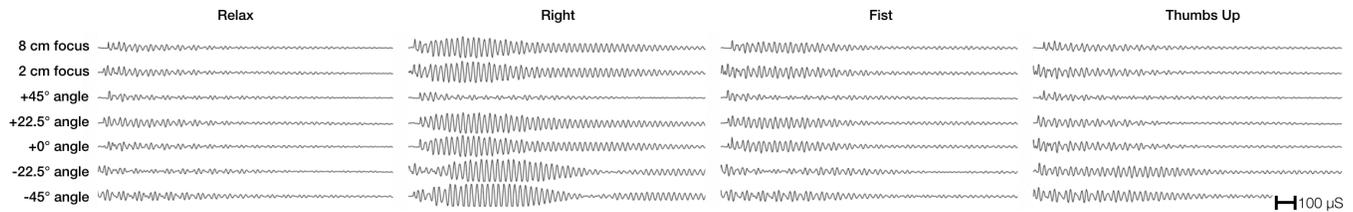


Figure 4. Example waveforms from four gestures (data received by transducer closest to thumb).

4.5 Acoustic Viewpoints & Waveforms

At any given time, seven transducers act as transmitters and one acts as a receiver. The sensor board cycles through all transmitter-receiver combinations, which results in 8 configurations. For each transducer configuration, we emit all 7 beamforming emissions, and capture 500 samples of the reflected waveform (333 kHz sampling rate), representing 1.5 ms of data. In total, this process yields 56 recorded waveforms (8 configurations \times 7 beamforming sequences) which we call a “sensing frame” (illustrated in Figure 1B).

4.6 Framerate

Each beamforming firing sequence takes 0.5 ms to generate and emit, followed by 1.5 ms of data collection. Thus, capturing a full sensor frame (56 waveforms) takes 112 ms. This results in \sim 8 full sensor frames per second.

For experimental purposes, we captured extra large buffers to see if there were interesting reflections at longer ranges. However, our study (and also seen in Figure 4) shows most signal returns within 0.8 ms, and if we contract to this smaller recording period, framerate increases to \sim 14 Hz. It is also possible to pre-generate beamforming sequences and store them in memory, which would save a further \sim 28 ms per sensor frame and increase framerate to \sim 22 Hz. Further optimizations could include time multiplexing the emissions such that one is in-flight while another is returning.

4.7 Features and Machine Learning

Our machine learning pipeline first converts the 56 incoming waveforms captured by our hardware into features. We segment each waveform into 20 bins and take the standard deviation of each bin as a feature, yielding 1120 values. We

use Scikit-learn’s Random Forest (default parameters, 500 trees) [33] for classification. 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 to reduce design bias and enable direct comparison between systems. The first is the hand gesture set defined in Tomo [50]. These seven gestures (relax + six “hand” gestures) are depicted in Figure 5 (green underscore). We also adopted the hand gesture set defined in Jung *et al.* [18], which extends or flexes the hand along three axes (two wrist axes and one finger axis). These six gestures are depicted in Figure 5 (purple underscore). We refer to this gesture set as “six-axis” in later text. Note these 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 evaluated 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

Participants wore 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 was



Figure 5. Our two gesture sets – Tomo set underscored in green and Six-Axis set underscored in purple (note four gestures are shared). 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, and I) Ulnar Reflection.

	Relax/Finger Extension	Fist/Finger Flexion	Right/Wrist Flexion	Left/Wrist Extension	Stretch	Thumbs Up	Spider Man	Radial Deflection	Ulnar Reflection
Relax/Finger Extension	94.5%	0.0%	3.5%	0.0%	0.0%	0.0%	1.5%	0.0%	0.5%
Fist/Finger Flexion	0.0%	89.5%	0.0%	0.3%	0.6%	5.5%	1.6%	1.3%	1.4%
Right/Wrist Flexion	2.8%	0.1%	90.5%	0.0%	0.0%	0.3%	5.5%	0.0%	0.8%
Left/Wrist Extension	0.0%	0.0%	0.0%	95.4%	1.4%	0.0%	0.0%	1.6%	1.7%
Stretch	0.0%	1.2%	0.0%	3.9%	87.2%	0.6%	0.0%	4.3%	2.9%
Thumbs Up	0.4%	5.1%	0.1%	0.1%	0.6%	87.9%	3.8%	0.0%	2.0%
Spider Man	0.1%	2.5%	4.1%	0.0%	0.0%	3.0%	89.9%	0.0%	0.5%
Radial Deflection	0.0%	0.6%	0.1%	2.7%	7.2%	0.0%	0.0%	88.0%	1.6%
Ulnar Reflection	0.0%	1.9%	0.0%	1.4%	5.1%	1.7%	0.6%	0.1%	89.4%

	Relax	Fist	Stretch	Right	Left	Thumbs Up	Spider Man
Relax	94.3%	0.0%	0.3%	3.8%	0.0%	0.2%	1.5%
Fist	0.0%	92.1%	0.3%	0.0%	1.1%	4.8%	1.8%
Stretch	0.0%	1.4%	93.7%	0.0%	4.0%	0.9%	0.0%
Right	3.0%	0.1%	0.0%	90.8%	0.0%	0.6%	5.6%
Left	0.0%	0.3%	1.4%	0.0%	98.3%	0.0%	0.0%
Thumbs Up	0.4%	5.0%	0.9%	0.2%	0.4%	89.7%	3.5%
Spider Man	0.3%	2.8%	0.0%	4.7%	0.0%	3.4%	88.9%

	Finger Extension	Finger Flexion	Wrist Flexion	Wrist Extension	Radial Deviation	Ulnar Deviation
Finger Extension	96.0%	0.0%	3.4%	0.2%	0.1%	0.5%
Finger Flexion	0.0%	95.6%	0.2%	0.5%	1.5%	2.3%
Wrist Flexion	4.3%	0.5%	94.4%	0.0%	0.0%	0.9%
Wrist Extension	0.0%	0.1%	0.0%	96.8%	1.7%	1.4%
Radial Deviation	0.0%	1.1%	0.0%	4.5%	92.1%	2.4%
Ulnar Deviation	0.1%	2.0%	0.0%	4.0%	1.0%	93.0%

Figure 6. Confusion matrices (within-session accuracies) for the combined gesture set (mean accuracy 90.2%), Tomo gesture set (mean 92.5%), and Six-Axis gesture set (mean 94.6%).

held for a few seconds, 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 sensor being removed and reworn in between. This procedure yielded 18,000 sensor frames (10 sensor frames \times 9 gestures \times 10 rounds \times 2 sessions \times 10 participants).

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 specifically, the average within-session accuracy was 92.5% (SD=2.2), while the six-axis gesture set achieved 94.6% (SD=3.4) accuracy. The largest source of error was confusion between similar hand-closing gestures, such as *Fist* and *Thumbs Up*, which accounted for 15.2% of the total error in the hand gesture set. Confusion matrices can be found in Figure 6.

6.3 Across-Session Accuracy

One significant challenge for on-body sensing systems is the ability to perform well 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, averaging the results. In the full, nine-class combined gesture set, the average across-session accuracy for all participants

was 81.4% (SD=15.9). In the Tomo gesture set, the average across-session accuracy was 86.0% (SD=12.7), and in the six-axis gesture set, the average across-session accuracy was 89.4% (SD=10.9). We saw a similar confusion between *Fist* and *Thumbs Up*, which accounted for 9.1% of the total error in the hand gesture set. However, other gestures appeared unaffected after rewearing the sensor (e.g., *Left* and *Wrist Flexion* performed at 94.2% and 96.2%, respectively). See Figure 7 for this experiment’s confusion matrices.

6.4 Across-User Accuracy

Another significant challenge for on-body systems is the ability to be trained once and work for all users (i.e., without per-user training or calibration). To investigate this potential, we ran a leave-one-user-out cross validation for each of our participants, where we train on all of 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 was 44.2% (SD=8.8). In the Tomo gesture set, the average across-user accuracy was 51.7% (SD=10.4), and in the six-axis gesture set, the average across-user accuracy was 63.2% (SD=8.5). This low performance suggests that users’ hands are different and perform gestures differently. Nonetheless, some gestures appear to be more consistent across users, such as *Wrist Flexion* and *Radial Deviation*, which performed at 80.1% and 79.2%, respectively.

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

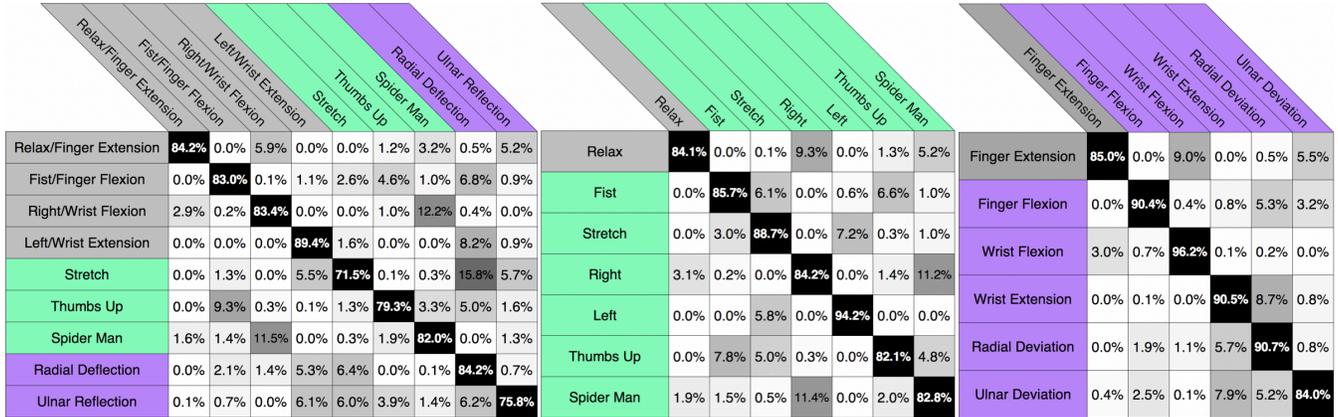


Figure 7. Confusion matrices (across-session accuracies) for the combined gesture set (mean accuracy 81.4%), Tomo gesture set (mean 86.0%), and Six-Axis gesture set (mean 89.4%).

94.6% respectively. When the gesture sets are merged (nine classes), BeamBand is 90.2% accurate.

Our system also performs comparably to other systems with custom gesture sets. Most notably, SensIR [29] reports 93.3% accuracy across twelve gestures, zSense [47] is 94.8% accurate across nine gestures, Skinput [15] is 96.8% accurate on four finger flicking gestures, and Mime [6] achieves ~95% accuracy on four gestures. 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 across-session accuracies of 65.3% for 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 across-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 more formally 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 a single layer of various fabrics over the transmitting transducer to simulate sleeve occlusion. We tested ten fabrics of different material, thicknesses and knit density (Figure 8).

While thickness does appear to correlate with signal attenuation, a more significant factor is knit density. For example, the polyester dress shirt was among the thinnest of our tested fabrics, and yet hurt performance the most. Conversely, the wool sweater (low knit density) was one of our better performing materials, despite being our thickest.

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” gesture recognition, more work is required to develop a generalizable model. Collecting more data across a wide range of participants may improve robustness. There may also be merit in moving away from classical machine learning methods towards deep learning. 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. Transducers running at 40 kHz are ubiquitous (and thus inexpensive) but are almost certainly not the optimal frequency for gesture recognition (a

Fabric Composition	Relative Signal Strength	Knit Density	Thickness (mm)
No Fabric (uncovered)	100.0%	-	-
100% Nylon Wide Knit Lace	98.2%	Low	0.20
100% Nylon Tight Knit Lace	96.4%	Low	0.17
50% Poly., 25% Rayon, 25% Cotton Shirt	74.5%	Medium	0.75
100% Wool Knit Sweater	73.6%	Low	1.50
60% Rayon, 40% Polyester Shirt	58.2%	Medium	0.66
100% Cotton Shirt	57.5%	Medium	0.83
90% Polyester, 10% Elastane Shirt	55.5%	High	1.00
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 “no fabric” condition) for ten clothing fabrics.

wavelength of ~8.5mm is likely 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.

Although a horseshoe arrangement should permit some degree of beamforming in the axis normal to the palm, we treat our array as though it was linear, which permits beamforming along the plane parallel to the palm. More advanced beamforming patterns, or certainly a 2D transducer array, would enable 3D raster-scan-like capabilities, which could offer much richer signals. Without doubt, it would facilitate recognition of gestures like *Fist* and *Thumbs Up*, which in cross-section look fairly similar.

We used a general-purpose microcontroller to facilitate research and rapid prototyping. In a commercial implementation, beamforming patterns would be saved in memory and specialized, energy-efficient hardware (e.g., ASICs) would drive the entire sensing process. Reducing the sensing duty cycle and running at full frame rate only when a change is detected would also improve power consumption. The sensing principle itself is fairly power efficient; the transducers themselves require virtually no power to drive. Thus, we believe a tether-less, self-contained version of BeamBand is possible with proper engineering.

There are also some important physical constraints. For example, we needed to raise the transducers off the skin in order to project acoustics over the bump at the base of the palm, which increases the minimum thickness of the band. Another limitation was the size of the transducers we selected – almost 13 mm in diameter. However, the piezo-elements inside are ~5 mm in diameter, which suggests tighter integration is possible. Also, ultrasonic transducers are not restricted to cylindrical housings; medical ultrasound utilizes small square elements arranged in a strip. Indeed, BeamBand could sit behind an acoustically-transparent plastic window on the side of smartwatches, very similar to medical ultrasound wands.

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 and focal points 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 reveals promising accuracies, 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.

ACKNOWLEDGEMENTS

This work was generously supported with funds from the Packard Foundation and Sloan Foundation. We are also indebted to Prof. Robert Xiao for his help with embedded development, and to Evi Bernitsas who sparked our interest in beamforming for human input sensing.

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. 1-41). IEEE. DOI: 10.1109/ICASSP.2007.366611
- [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. 2013. UltraHaptics: multi-point mid-air haptic feedback for touch surfaces. In *Proceedings of the 26th annual ACM symposium on User interface software and technology (UIST '13)*. ACM, New York, NY, USA, 505-514. DOI: <https://doi.org/10.1145/2501988.2502018>
- [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 Biology*101, 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. 2010. Skinput: appropriating the body as an input surface. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. ACM, New York, NY, USA, 453-462. DOI: <https://doi.org/10.1145/1753326.1753394>
- [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, Guckhan 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. DOI: 10.1109/TII.2015.2405413
- [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 underwater with towed array sonar". *Defense 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, Chiuang 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. 2017. Improving the Feasibility of Ultrasonic Hand Tracking Wearables. In *Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces (ISS '17)*. ACM, New York, NY, USA, 342-347. DOI: <https://doi.org/10.1145/3132272.3135075>
- [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] Adiyani Mujibiyah, 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. 2015. SkinWatch: skin gesture interaction for smart watch. In *Proceedings of the 6th Augmented Human International Conference (AH '15)*. ACM, New York, NY, USA, 21-24. DOI: <http://dx.doi.org/10.1145/2735711.2735830>
- [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. DOI: 10.1109/ISWC.1999.806717
- [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. DOI: 10.1109/ISWC.2001.962092
- [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. 2015. 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 (HotMobile '15)*. ACM, New York, NY, USA, 9-14. DOI: <https://doi.org/10.1145/2699343.2699350>
- [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>