Interferi: Gesture Sensing using On-Body Acoustic Interferometry

Yasha Iravantchi   **Yang Zhang**   **Evi Bernitsas**   **Mayank Goel**   **Chris Harrison**
Carnegie Mellon University, Human-Computer Interaction Institute
5000 Forbes Avenue, Pittsburgh, PA 15213
{ysi, yang.zhang, mayank, chris.harrison}@cs.cmu.edu, pbernitsas@andrew.cmu.edu

ABSTRACT

Interferi is an on-body gesture sensing technique using acoustic interferometry. We use ultrasonic transducers resting on the skin to create acoustic interference patterns inside the wearer's body, which interact with anatomical features in complex, yet characteristic ways. We focus on two areas of the body with great expressive power: the hands and face. For each, we built and tested a series of worn sensor configurations, which we used to identify useful transducer arrangements and machine learning features. We created final prototypes for the hand and face, which our study results show can support eleven- and nine-class gestures sets at 93.4% and 89.0% accuracy, respectively. We also evaluated our system in four continuous tracking tasks, including smile intensity and weight estimation, which never exceed 9.5% error. We believe these results show great promise and illuminate an interesting sensing technique for HCI applications.

CCS CONCEPTS

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

KEYWORDS

Biosensing; Hand Input; Hand Gesture; Face Gesture; Acoustic Interferometry; Acoustic; Interaction Techniques; Wearables

1 INTRODUCTION

Sensing expressive hand and face gestures using a minimally-invasive, worn system has proven challenging, despite numerous immediate applications, including gesture-driven interfaces and social VR experiences. Many approaches have been explored in the Human-Computer Interaction literature, including electromyography [35][36], bio-acoustics [14][22], electrical impedance tomography [45][46], contour and pressure sensing [8], and worn computer vision systems [21].

In this work, we introduce *Interferi,* a new approach for on-body sensing that uses acoustic interferometry. We use ultrasonic transducers resting on the skin to project structured acoustic interference patterns inside a wearer's body. These patterns interact with anatomical features, such as fat, muscles and bones, in complex and characteristic ways. These interactions can be learned by classifiers to power interactive applications. We built and tested a series of arm-worn bands that can sense hand gestures (i.e., pose), which could be integrated into *e.g.*, a smartwatch to enable input beyond the touchscreen. We also built a facial expression sensing mask, which could be integrated into *e.g.*, AR/VR headsets to enable more expressive social experiences.

We iteratively developed Interferi, starting with software and physical simulations to confirm the basic operating principle. We then developed custom hardware and software to drive eight acoustic transducers (40 kHz) and connected in different arrangements for arm-band and face-mask form factors. We then ran a user study with four participants to evaluate these designs and different transducer configurations. From these results, we selected a final set of features, as well as our best performing arm band and face mask. Using these two final designs, we ran a second user study to assess discrete gesture recognition performance for the hands and face, demonstrating classification accuracies of 93.4% and 89.0%, respectively. We performed a third and final user study to assess continuous tracking performance, which included four tasks: smile intensity, lifted-weight estimation, wrist angle, and hand pose, which never exceed 9.5% error.

2 RELATED WORK

We first review prior sensing systems that enable similar hand and face sensing capabilities as Interferi. We focus on self-contained worn systems as opposed to those deployed in the environment. We then cover work that more closely aligns with our technical approach of ultrasound acoustometry - both reflectometry and interferometry.

2.1 Hand/Arm Gesture Sensing

One of the most popular approaches for detecting hand gestures is to attach sensors to the hands. For example, Perng et al. [32] developed a glove with accelerometers at the fingertips to detect static hand poses. ThumbRing's finger-borne accelerometer [41] can track fine grained movements for discrete input tasks. Beyond accelerometers, Whitmire et al. [43] developed a glove that uses conductive fabric to detect finger and thumb interactions, and thus gestures. BackHand [24] used strain gauges affixed to the back of hand to detect the movement of finger tendons for gesture sensing.

To avoid directly instrumenting the fingers or hands, researchers have considered many approaches that relocate sensing to the wrist or arm. Optical methods are popular for this goal, for example, WristWhirl [11] detects the angle of a hand using an array of infrared proximity sensors. Cameras have been used to recognize discrete hand gestures [42] and even reconstruct a 3D model of the hand [21]. There is also a body of research that leverages the fact that the arm's contour deforms when a user performs hand gestures, which can be captured using pressure [8][19], infrared proximity [30], and capacitive sensors [34].

Instead of relying on external signals, researchers have also investigated internal signals (i.e., inside the body) to reveal hand state. Most prevalent in the literature is Electromyography (EMG), which passively detects electrical signals generated by muscle movement [20][35][36][39]. Electrical Impedance Tomography is an example active approach, which has been used to sense changes in the interior anatomy of the arm for hand gesture sensing [45][46]. This research shows great promise, but has been shown to be sensitive to skin contact condition and environmental EM interference.

Most related to Interferi are approaches that use acoustic signals. One approach is to passively listen to the microvibrations generated by hand gestures. For example, Amento et al. [2], Hambone [9], and Skinput [14] placed acoustic sensors on the skin to detect finger taps, flicks and pinches. More recent research has shown that the motion sensors in smartwatches can be used for similar purposes [22][44][47]. Active acoustic approaches are also possible, for example, the "Sound of Touch" [29] uses a transmitter-receiver armband to detect grasping gestures on an arm. FingerPing [48] places a transmitter on the wrist and uses a ring receiver to detect gestures on the hand. Perhaps most technically sophisticated is EchoFlex [28], which uses an off-the-shelf medical ultrasound machine (MSRP ~$90K) to resolve a high-resolution interior image of an arm for computer-vision-driven hand gesture sensing. This method requires the use of acoustic gel, making it cumbersome for everyday use.

2.2 Facial Gesture Sensing

Digitization of facial expressions has been a long-standing research topic. The most common approach is to use cameras and computer vision operating in the environment (see [7] for a survey). More relevant to Interferi are headset-mounted camera approaches, for example, EyeSpyVR [1], which uses the front facing camera on a smartphone to capture eye blinks and coarse gaze direction by looking through the VR headset lens. In [23], a camera on the underside of a VR headset was used to capture mouth pose. Camera based methods require line of sight, which can be challenging with e.g., a worn headset that occludes much of the face. For this reason, other non-visual methods have been considered, such as measuring facial muscle movement with EMG [12][13]. It is also possible to infer face pose by measuring skin contour changes, using e.g., infrared proximity [25][26][38] and pressure sensors [23][37]; deformations in the ear canal has also been used [3][27].

2.3 Acoustic Reflectometry

Ultrasonic sensing has been used in many domains, from oceanography [5] to applied physics [4][18]. A common class of ultrasound sensing is reflectometry, where an acoustic signal is emitted, and reflections are recorded. Sonar (and echolocation in the animal world) emits pulses of ultrasound and use the time-of-flight of reflections to determine distance. More complex is medical ultrasound, which uses an array of small ultrasonic transducers to create a steerable focal point of acoustic energy (i.e., beamforming) to reveal interior structures [4][16]. To effectively beamform in a liquid medium, megahertz-scale frequencies must be used, which also necessitates the use of impedance-matching gels on the body, impractical for consumer use. Finally, we note that ultrasound in the body has been found to be safe in many comprehensive studies [15][16].

2.4 Acoustic Interferometry

Interferometry differs from the above techniques in several key ways. Foremost, it diverges from sonar reflectometry in that time-of-flight is not the value of interest, as well as the fact it uses multiple emitters/receivers surrounding an area of interest. Secondly, although both techniques rely on interference effects, beamforming is used to create a coherent wavefront, as opposed to a distributed pattern. Acoustic beamforming is most often used in a reflectometry-like manner, with focused "beams" reflecting signals back to a source. Indeed, while reflectometry naturally lends itself to imaging and depth sensing, the low spatial frequency and

overlapping signals of our kilohertz-scale interferometry make it a poor technique for such applications. However, the signal is perfect for projecting nodes and troughs of acoustic pressure at gross-anatomy-scale inside the body. Even small movements can cause a node to shift from muscle to bone, creating a large reflection, and also create secondary interference effects, all of which radiate back out to the skin where they can be captured by many sensors.

Despite an extensive literature search, we were unable to find any HCI work that takes advantage of active acoustic interferometry as a sensing technique. The closest we could find are systems that use a grid of ultrasonic transmitters to create acoustic interference to provide in-air haptics [6] and also levitate small objects [17]. Thus, we hope this work can draw attention to this interesting sensing technique and encourage future use.

3 PILOT MODELS

Before developing our system, we conducted a series of small-scale tests to confirm our theory-based understanding of the propagation of compressive ultrasonic waves in a liquid medium. For this, we use a 120 mm diameter column of water at 20°C (i.e., 1482 m/s propagation speed) as a rough approximation of a human limb. We started with software simulations, testing both single and multi-emitter configurations, and confirmed a wide variety of interference patterns could be generated by varying position and relative phase of transducers. Figure 1, top, shows the output of two such simulations – a single transmitter and two transmitters (in phase). Of course, the real world is more variable, with *e.g.*, transducer impedance mismatches and multipath interference. For this reason, we also ran tests using an *actual* 120 mm, water-filled, acrylic cylinder.

Using this setup, we tested different transducer arrangements, angles and relative phase offsets. In order to create a comprehensive view of ultrasonic waveforms propagating and interfering, we used a sensor attached to a CNC gantry and scanned the water bath. For any given emitter configuration (*e.g.*, two emitters, at a right angle, 0° phase offset), the sensor would be moved to 113 points inside the bath (on a 5 mm grid). At each point in the bath, the same emitter sequence was fired and the acoustic interaction at the sensor was recorded. After waveforms at all points were captured, they could be synchronously replayed (see Video Figure) to visualize the acoustics inside the bath. Figure 1, bottom, shows the physical results for a single transmitter and two transmitters (in phase).

Our software and physical models closely matched and helped ground our understanding of the operating
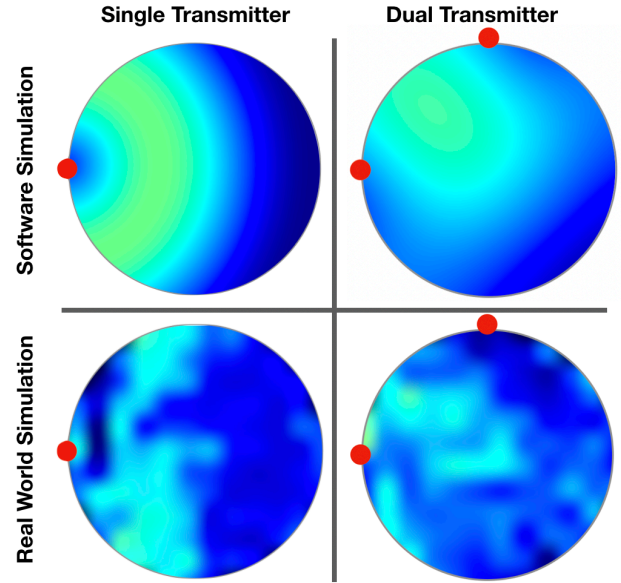

Figure 1. We ran software and physical simulations of ultrasonic propagation in a column of water (seen from above here as four circles). Bright green is high pressure and dark blue is low pressure; active 40 kHz transducers shown in red. See text for experiment setup details.

principle. More specifically, it allowed us to test and verify a wide variety of transducer arrangements and signals that would be difficult understand by collecting data on the body alone (which can only be sensed at the circumference, *i.e., on* the skin, but not *inside* the body). We also used these experiments to test over a dozen different ultrasonic transducers with different resonant frequencies (40-200 kHz), power ratings (5-140 V), physical size (10-18 mm diameter), and beam width (7-70°). We tested transducers by driving them at their maximum rated voltage and measuring V_{pp} at a matching transducer placed 1cm away in the water bath. This led us to select [33] – a 16 mm diameter, 40 kHz transducer with a 70° beam width and 140 V max rating.

4 IMPLEMENTATION

Interferi consists of three main components. First is our custom driver board (Figure 2), which generates, captures, and processes ultrasonic signals. Next are the worn sensors themselves, which contain ultrasonic transducers that emit and receive signals. Finally, we have laptop-based software that receives data from the hardware and performs additional processing and machine learning. We now describe these elements in greater detail.

4.1 Sensor Board

Our worn sensors incorporate eight piezoelectric transducers (identified in the previous study) that can emit and

receive ultrasonic acoustic signal. To overcome the acoustic impedance mismatch between transducer and skin, we selected a high-voltage transducer (140 V_{pp} max), which obviates the need for coupling agents such as acoustic gel. To drive these transducers with software-defined waveforms, we built a custom sensor board (Figure 2 & 3) with three main components: a high voltage EMCO SIP100 DC-DC power regulator [10], high voltage amplifiers, and multiplexed analog frontend. Our board can configure any number of transducers as input or output. We used a Teensy 3.6 [39] (overclocked to 240 MHz) for generating drive signals, reading transducers, and USB communication.

4.2 Waveform Generation

Since acoustic signals propagate at high speed (~1500 meters per second under water, and even faster in semi-solid mediums like the body), generating highly synchronized drive signals is crucial for controlled interferometry. In order to control the offset and phase of each transducer interpedently at a granularity of ~0.1° at 40 kHz (the equivalent of a single clock cycle on our 240 MHz microcontroller), it is necessary to pre-generate waveforms and save them to memory. We then load patterns sequentially into a I/O map register, which allows us to toggle 8 output pins simultaneously in a single clock cycle (*i.e.*, 4.17 ns).

4.3 Waveform Emission

The I/O map register toggles its digital pins to generate a 3.3 V_{pp} 40 kHz square wave signal. This signal is then amplified to 100 V_{pp} to drive the transducers. To reduce amplification lag and reduce crosstalk, each transducer has a dedicated amplifier. To ensure smooth transitions between transducer combinations, we clamp transducers to ground for 5 μs after each firing, which reduces transducer ringing and improves setting time. During this idle period is also when we perform waveform generation, described in the previous section.

4.4 Analog Front End

On the receiving side, the circuit has an active high pass filter with a fixed gain (f_c=39 kHz, G= 5) and a secondary
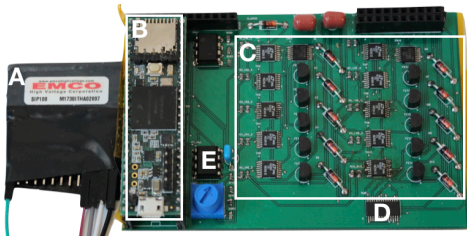

Figure 2. The custom sensor board for Interferi. A) DC-DC converter, B) Teensy 3.6, C) high voltage amplifiers, D) multiplexer, and E) filter and amplification stage.
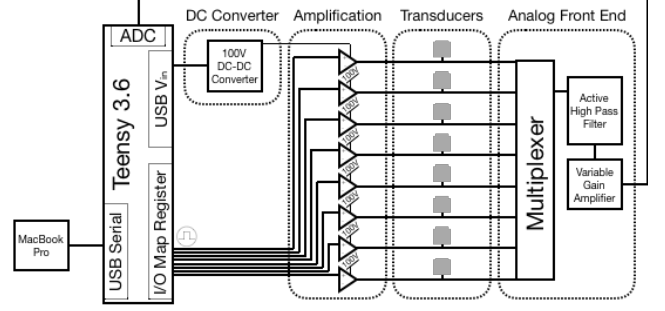

Figure 3. A schematic view of Interferi's hardware.

amplification stage with adjustable gain up to 40x. We multiplex the analog frontend to connect to the transducers, one at a time. The amplified signal is DC biased to 1.67 V (i.e., V_{ADC}/2) and sampled by the microcontroller's 12-bit ADC at 480 kHz. For each transducer configuration, we record 250 μs worth of data (i.e., 120 samples), which is transmitted to a laptop over USB for further computation.

4.5 Framerate

In total, configuring and reading 120 samples of data for a single transducer configuration takes 255 μs (including the 5 μs clamping period). Thus, our board can capture 3922 transducer configurations per second at full rate (not including *e.g.*, USB communication). As we will discuss later, our final design for hand gesture sensing utilizes 30 transducers configurations, yielding a raw sensing frame rate of 133 FPS (51 FPS with communication overhead). Our face sensor uses 48 configurations, yielding a raw sensing frame rate of 82 FPS (32 FPS in practice).

4.6 Transducer Arrangements

We generated worn designs for the hands and face using basic taxonomies, informed in part by our pilot modeling study. For the arms, one design axis was whether the transducers were arranged together (contiguous) or split into two groups (split). On the other axis, the transducers either ran down the length of the arm (linear) or wrapped around the arm (ring). For the face, we explored horizontally symmetric *vs.* asymmetric transducer arrangements, and also brow-centric *vs.* distributed. This yielded four arm designs and four face designs, illustrated in Figure 4.
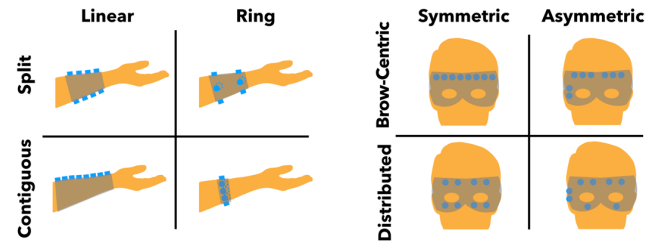

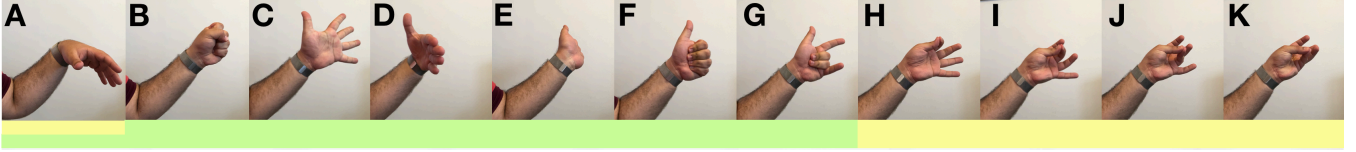Figure 4. The four geometries for the arm and face.

Figure 5. Full hand gesture set: A) Relaxed, B) Fist, C) Stretch, D) Right, E) Left, F) Thumbs Up, G) Spider-Man, H) Index Pinch, I) Middle Pinch, J) Ring Pinch, and K) Little Pinch. "Coarse hand" subset underscored in green; "pinch" subset in yellow.

Figure 6. Full facial gesture set: A) Relaxed, B) Left Eye Closed, C) Right Eye Closed, D) Both Eyes Closed, E) Eyebrows Raised, F) Smile, G) Frown, H) Mouth Open, and I) Kissy Face. "Eye" gesture subset underscored in pink; "mouth" subset in blue.

4.7 Power Consumption

We did not optimize power consumption in our proof-of-concept hardware, which is powered by 5V over USB. Nonetheless, we did measure power draw: ~400mA total, of which 250mA is from our overclocked Teeny 3.6 board. Our DC-DC converter consumes ~140mA, most of which is conversion loss. All other components, including our transducers, consume ~10mA.

4.8 Machine Learning

Our machine learning pipeline, running on a laptop, converts incoming waveform data as captured by our hardware into features. For discrete classification, we use Scikit-learn's Random Forest Classifier (default parameters, 200 trees) and for continuous classification we use Scikit-learn's Extra Trees Regressor (default parameters, 200 trees) [31]. All of these tasks were performed on a standard configuration 2017 15" MacBook Pro.

5 STUDY 1: SENSOR GEOMETRIES AND MODES

Our aforementioned software and physical simulations informed our understanding of the sensing principle, but not how well our technique worked in practice for sensing gestures on the human body. Thus, our first user study sought to quantify the value of different worn sensor geometries (Figure 4), transducer pairings, and phase offsets. Specifically, from our 8 transducers, we generated 56 single-transmitter/single-receiver combinations, 336 dual-transmitter/single-receiver combinations at each of the 4 phase offsets between transmitters (0°, 15°, 30°, 45°); resulting in (56 + 336 × 4 = 1400 possible combinations).

5.1 Gesture Set

We adopted the two hand gesture sets defined in [45], which permits direct accuracy comparison in our later

studies: seven "coarse" gestures (Figure 5, green) and five "pinch" gestures (Figure 5, yellow), which share a common gesture (*Relaxed*). We could not find a suitable face gesture set in the literature, and so we defined our own nine classes, containing both an "eye" gesture set (Figure 6, pink) and "mouth" gesture set (Figure 6, blue) with a common *Relaxed* gesture. Note these gestures are static, which are alternatively called *pose* in the HCI literature.

5.2 Procedure

We recruited 4 participants (1 female, mean age 25) who wore each of our four arm bands and four face masks (Figure 4). While wearing each design, participants performed every gesture, ten times each, in a random order. While holding the gesture, the hardware captured data for each transducer combination and phase offset. For experimentation, we extended the recording period to 1325 μs (Figure 7 offers several example waveforms). Capturing all 1400 waveforms, including transmission over USB, took our sensor board approximately 4 seconds per full frame (*i.e.*, 0.25 FPS; which we improve subsequently).
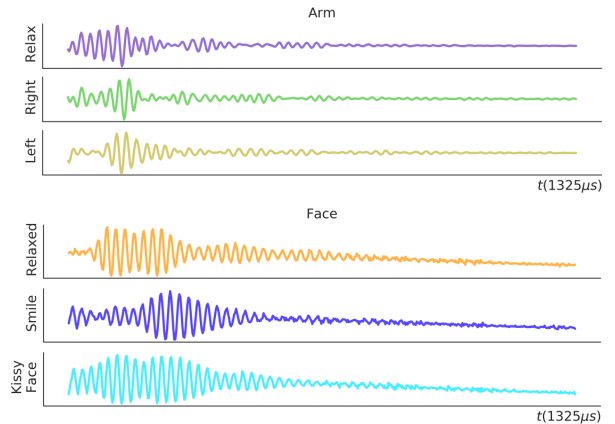

Figure 7. Example waveforms from a single receiver (same emitter configuration). Top: hand gestures (Relaxed, Right, Left). Bottom: face gestures (Relaxed, Smile, Kissy Face).
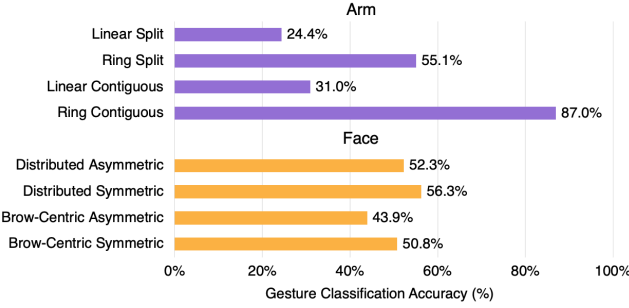
Figure 8. Performance of different sensor geometries.

Figure 9. Performance of different transducer pairings.

5.3 Results: Worn Sensor Design

Our first goal was to identify transducer geometries that captured the most useful signals on the arm and face. Due to the large number of transducer configurations we captured, it was important to limit the feature space to prevent overfitting. For this, we use mean absolute z-scored amplitude to quantify how different a received waveform is from its relaxed state. Though basic, we found this metric to be a highly descriptive feature, and we use it to produce a more manageable 1400-length feature vector (one feature per transducer configuration) for information power analysis.

We then ran a leave-one-user-out cross-validation to get a rough estimate of gesture classification accuracy (Figure 8). On the arm, Ring Contiguous clearly out-performed all other geometries across gestures. We also found a horseshoe shaped array centered on the extensor digitorum provided better contact, and so we moved to this design for future experiments. On the face, the two Brow-Centric designs were superior on tasks that involved movement of the eyes and forehead, but performed poorly on mouth movements. The two Distributed geometries performed more evenly across gestures. The Distributed Symmetric arrangement slightly outperformed its Asymmetric counterpart, and so we selected this design for our future face experiments.

5.4 Results: Transducer Configuration

Having selected a worn sensor design for the arm and face, we next sought to identify strong transducer pairings and phase offsets. For these analyses, we also use the mean absolute z-scored amplitudes for features. Instead of using classification accuracy as a measure of success, we performed a feature selection analysis (InfoGainAttributeEval with Ranker attribute selection, default parameters, Weka) on all 1400 transducer combinations, which allowed us to assess families of configurations by relative merit.

5.4.1 Transmit Mode – We first investigated the performance difference between the single-transmitter/single-
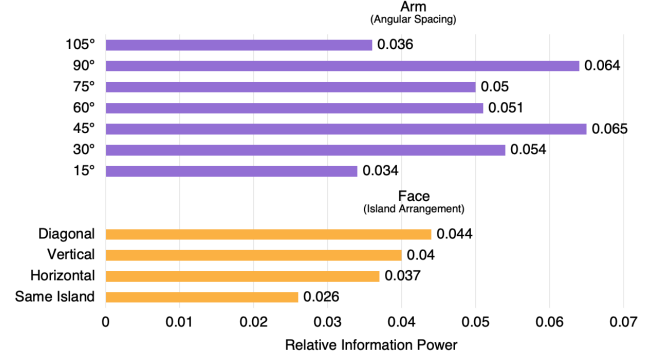
receiver and dual-transmitter/single-receiver drive modes. Single-transmitter configurations offered a mean information power of 0.030 while dual-transmitter configurations scored 0.044. Although interference patterns are also created with just a single emitter (due to multipath and scattering in the body), this result served to further underscore the value of controlled interference patterns created by multiple emitters. Thus, we selected dual-transmitter/single-receiver as our sole operation mode for the remainder of our experiments.

5.4.2 Transducer Pairings – Once we selected the dual-transmitter drive mode, the next question was which transducer pairings were most effective. On our arm band, eight transducers were spaced apart at roughly 15° intervals. This allowed us to test pairings with spacing from 15° (adjacent transducers) to 105° (transducers 1 and 8). The results are shown in Figure 9, with 45° and 90° spacing offering the most information power. Closer examination found these spacings to be largely rudiment in terms of information power, and so we selected 45° as our sole angular spacing on the arm. Figure 11, left, illustrates these five combinations.

On the face, the transducer geometry is not linear like the arm. Instead of angular spacing, we grouped transmitters into one of four "islands" (above left eye, below left eye, above right eye, below right eye). We then tested four types of pairings: whether the transmitters were located in the same island (e.g., the two transmitters above the left eye), or on islands separated horizontally, vertically, or diagonally. The results, shown in Figure 9, shows that diagonal pairings have the highest mean feature merit, followed by vertical pairings. To limit combinatoric explosion, which affects sensor framerate, we selected representative pairings from the two families of combinations, four diagonal pairings and four vertical pairings, seen in Figure 11, right.

5.4.3 Phase Offset – We next investigated phase offset, which refer to the relative pulse delay between two
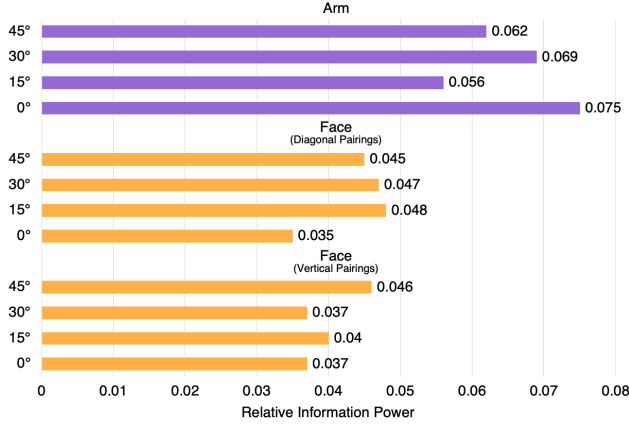
Figure 10. Performance of different phase offsets.

transducers. In evaluating phase offset, we focused on the performance within our previously identified pairing arrangements. On the arm, using 45° transducer spacing, we found 0° phase offset (i.e., synchronous firing) performed the best. On the face, with diagonal pairings, 15° had a slight edge over other phase offsets. For vertical pairings, 45° was a clear winner. See Figure 10, for a full breakdown of feature merit results.

5.4.4   Final Waveforms and Machine Learning Features – To summarize, we selected the dual-transmitter configuration for the arm and face. For the arm, we use all 5 possible transducer combinations with 45° spacing, firing with a 0° phase offset, and capture data on the 6 remaining transducers, for a total of 30 waveforms. For the face, we use the 4 diagonal transducer pairings (15° phase offset) and the 4 vertical transducer pairings (45° phase offset), receiving on the 6 unused transducers, for a total of 48 waveforms.

As noted in our procedure, our software captured 1325 μs of waveform data for this study. However, we found that the first 250 μs contained most of the information power, and so we limited future measurements to this interval, which boosted our sensor's framerate over five-fold, to 51 FPS on the arm and 32 FPS on the face (including USB transmission). Finally, we decimate these 250 μs waveforms into 20 bins of mean and standard deviation, which we use as features for machine learning.

6   STUDY 2: DISCRETE GESTURES

In Study 1, we identified strongly performing worn sensor designs for the arm and face (Figure 12), and selected the strongest transducer configurations. In Study 2, we evaluate the gesture classification performance of our significantly faster prototype on a larger set of users. Specifically, we recruited 10 participants (3 female, mean age 30), which

Figure 11. Final transducer firing pairings. On the armband, each color represents a transducer firing pair. On the facemask, the orange dotted line refers to transducers within the same "island", blue and red lines refer to vertical and diagonal firing pairs, respectively.

had a mean arm diameter of 9.7 cm (SD=1.8) and a mean face width (measured at the eyes) of 11.5 cm (SD=1.6). The study took approximately one hour and participants were paid $10 USD.

6.1   Procedure

The study was broken into two phases: arm and face gestures. For the arm, participants wore the armband on their non-dominant hand – as all ten of our participants were right handed, this meant the armband was always worn on the left arm. We used the same arm and face gesture sets as Study 1 (Figures 5 and 6). A single round of data collection consisted of each gesture being performed once, in a random order. Each gesture was held for approximately two seconds, during which time 50 sensor frames were recorded. A session consisted of ten rounds of data collection. To add realism, we collected two sessions of data for each user, with the worn sensor being removed in between.

In total, this procedure yielded 110,000 sensor frames (50 sensor frames × 11 gestures × 10 rounds × 2 sessions × 10 participants) for the arm gesture set and 90,000 sensor frames (50 sensor frames × 9 gestures × 10 rounds × 2 sessions × 10 participants) for the face gesture set.

6.2   Within-Session Accuracy

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

Figure 12. Our final prototypes for the arm and face.

Figure 13. Confusion matrices (within-session) for the complete set, hand set and pinch set on the arm.

Complete set (arm):

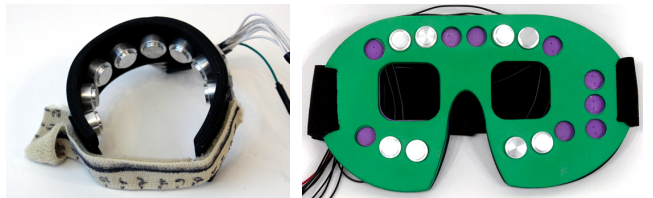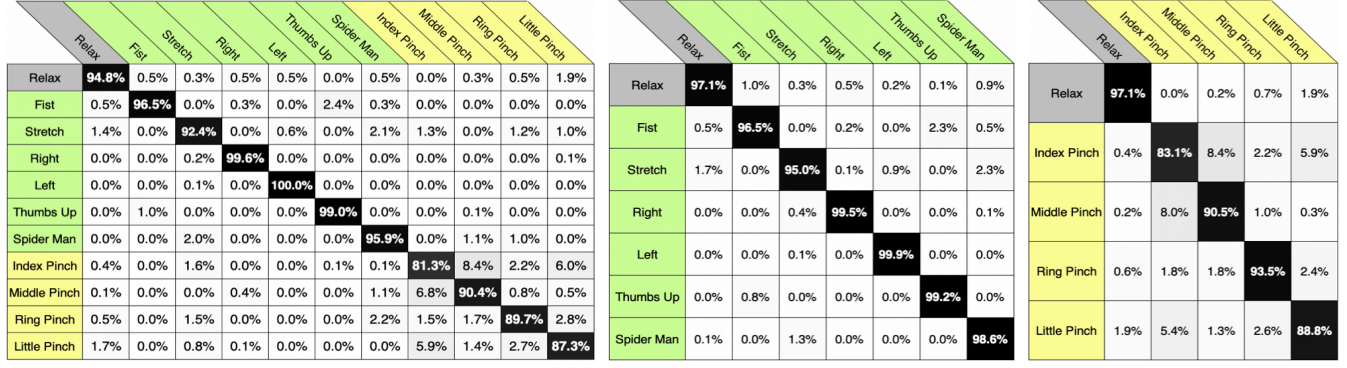|	Relax	Fist	Stretch	Right	Left	Thumbs Up	Spider Man	Index Pinch	Middle Pinch	Ring Pinch	Little Pinch
Relax	94.8%	0.5%	0.3%	0.5%	0.5%	0.0%	0.5%	0.0%	0.3%	0.5%	1.9%
Fist	0.5%	96.5%	0.0%	0.3%	0.0%	2.4%	0.3%	0.0%	0.0%	0.0%	0.0%
Stretch	1.4%	0.0%	92.4%	0.0%	0.6%	0.0%	2.1%	1.3%	0.0%	1.2%	1.0%
Right	0.0%	0.0%	0.2%	99.6%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%
Left	0.0%	0.0%	0.1%	0.0%	100.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Thumbs Up	0.0%	1.0%	0.0%	0.0%	0.0%	99.0%	0.0%	0.0%	0.1%	0.0%	0.0%
Spider Man	0.0%	0.0%	2.0%	0.0%	0.0%	0.0%	95.9%	0.0%	1.1%	1.0%	0.0%
Index Pinch	0.4%	0.0%	1.6%	0.0%	0.0%	0.1%	0.1%	81.3%	8.4%	2.2%	6.0%
Middle Pinch	0.1%	0.0%	0.0%	0.4%	0.0%	0.0%	1.1%	6.8%	90.4%	0.8%	0.5%
Ring Pinch	0.5%	0.0%	1.5%	0.0%	0.0%	0.0%	2.2%	1.5%	1.7%	89.7%	2.8%
Little Pinch	1.7%	0.0%	0.8%	0.1%	0.0%	0.0%	0.0%	5.9%	1.4%	2.7%	87.3%

Hand set (arm):

	Relax	Fist	Stretch	Right	Left	Thumbs Up	Spider Man
Relax	97.1%	1.0%	0.3%	0.5%	0.2%	0.1%	0.9%
Fist	0.5%	96.5%	0.0%	0.2%	0.0%	2.3%	0.5%
Stretch	1.7%	0.0%	95.0%	0.1%	0.9%	0.0%	2.3%
Right	0.0%	0.0%	0.4%	99.5%	0.0%	0.0%	0.1%
Left	0.0%	0.0%	0.1%	0.0%	99.9%	0.0%	0.0%
Thumbs Up	0.0%	0.8%	0.0%	0.0%	0.0%	99.2%	0.0%
Spider Man	0.1%	0.0%	1.3%	0.0%	0.0%	0.0%	98.6%

Pinch set (arm):

	Relax	Index Pinch	Middle Pinch	Ring Pinch	Little Pinch
Relax	97.1%	0.0%	0.2%	0.7%	1.9%
Index Pinch	0.4%	83.1%	8.4%	2.2%	5.9%
Middle Pinch	0.2%	8.0%	90.5%	1.0%	0.3%
Ring Pinch	0.6%	1.8%	1.8%	93.5%	2.4%
Little Pinch	1.9%	5.4%	1.3%	2.6%	88.8%

For the arm, the average within-session accuracy across all participants on the full, eleven-class hand gesture set was 93.4% (SD = 4.0%). A significant source of error was confusion in the *Index Pinch* gesture, contributing 25.5% of misclassifications. In the "coarse" hand gesture subset (7 classes) the average within-session accuracy was 98.0% (SD=1.8), and in the "pinch" gesture subset (5 classes) the average within-session accuracy was 90.6% (SD = 7.5%). The confusion matrices can be found in Figure 13.

For the face, the average within-session accuracy across all participants on the full, 9-class gesture set was 89.0% (SD = 5.7%). The *Relax* gesture contributed 22.3% of misclassification. In the "eye" gesture subset, the average within-session accuracy was 92.7% (SD = 4.7%). In the "mouth" subset, the average within-session accuracy was 91.1% (SD = 5.1%). The confusion matrices can be found in Figure 14.

6.3 Across-Session Accuracy

Maintaining performance across worn sessions. (i.e., reworn at a late time) is a challenge for almost all bio-sensing systems, as even slight misalignments can result in significant signal change. To quantify Interferi's drop in performance when reworn, we ran a leave-one-session-out cross validation for each of our participants. Specifically, we train on all data from session one and test on all data from session two, and vice versa.

In the arm condition (11 classes), our participants had a mean across-session accuracy of 65.7% (SD=19.2). It is worth noting that Interferi's performance is not uniform across all gestures, as some gestures performed well across-session, such as *Left* and *Right*, with 96.4% and 94.2% accuracy, respectively. Broken out into the "coarse" and "pinch" subsets, participants had a mean across-session accuracy of 80.0% (SD=16.4) and 60.5% (SD=21.6), respectively. For the face (9 classes), mean across-session accuracy across all ten participants was 64.2% (SD=17.2). In the "eyes" and "mouth" subsets, participants had a mean across-session accuracy of 74.4% (SD=17.5) and 68.4% (SD=18.3), respectively.

7 STUDY 3: CONTINUOUS TRACKING

In addition to investigating Interferi's performance in detecting discrete gestures, we also wished to explore our system's ability to provide continuous hand and face tracking. For this evaluation, we recruited 10 participants (2 female, mean age 22), with a mean arm diameter of 7.6 cm (SD=0.7), and mean face width of 10.3 cm (SD=1.3). The study took approximately 30 minutes and participants were paid $10. Participants completed four continuous tasks, three for the arms and one for the face, using the same worn sensors as Study 2. All tasks followed the same general procedure, collecting ten rounds of data (task-specific details below). We
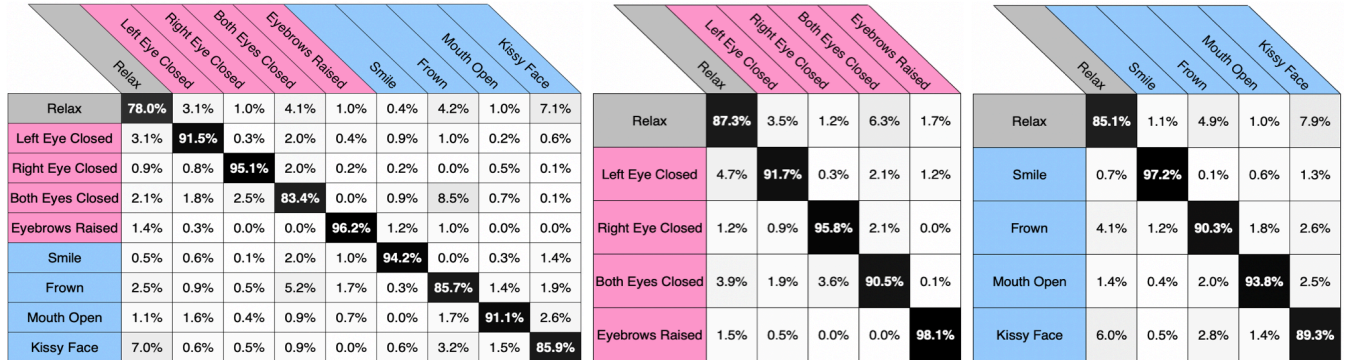

Complete set (face):

	Relax	Left Eye Closed	Right Eye Closed	Both Eyes Closed	Eyebrows Raised	Smile	Frown	Mouth Open	Kissy Face
Relax	78.0%	3.1%	1.0%	4.1%	1.0%	0.4%	4.2%	1.0%	7.1%
Left Eye Closed	3.1%	91.5%	0.3%	2.0%	0.4%	0.9%	1.0%	0.2%	0.6%
Right Eye Closed	0.9%	0.8%	95.1%	2.0%	0.2%	0.2%	0.0%	0.5%	0.1%
Both Eyes Closed	2.1%	1.8%	2.5%	83.4%	0.0%	0.9%	8.5%	0.7%	0.1%
Eyebrows Raised	1.4%	0.3%	0.0%	0.0%	96.2%	1.2%	1.0%	0.0%	0.0%
Smile	0.5%	0.6%	0.1%	2.0%	1.0%	94.2%	0.0%	0.3%	1.4%
Frown	2.5%	0.9%	0.5%	5.2%	1.7%	0.3%	85.7%	1.4%	1.9%
Mouth Open	1.1%	1.6%	0.4%	0.9%	0.7%	0.0%	1.7%	91.1%	2.6%
Kissy Face	7.0%	0.6%	0.5%	0.9%	0.0%	0.6%	3.2%	1.5%	85.9%

Eyes set (face):

	Relax	Left Eye Closed	Right Eye Closed	Both Eyes Closed	Eyebrows Raised
Relax	87.3%	3.5%	1.2%	6.3%	1.7%
Left Eye Closed	4.7%	91.7%	0.3%	2.1%	1.2%
Right Eye Closed	1.2%	0.9%	95.8%	2.1%	0.0%
Both Eyes Closed	3.9%	1.9%	3.6%	90.5%	0.1%
Eyebrows Raised	1.5%	0.5%	0.0%	0.0%	98.1%

Mouth set (face):

	Relax	Smile	Frown	Mouth Open	Kissy Face
Relax	85.1%	1.1%	4.9%	1.0%	7.9%
Smile	0.7%	97.2%	0.1%	0.6%	1.3%
Frown	4.1%	1.2%	90.3%	1.8%	2.6%
Mouth Open	1.4%	0.4%	2.0%	93.8%	2.5%
Kissy Face	6.0%	0.5%	2.8%	1.4%	89.3%

Figure 14. Confusion matrices (within-session) for the complete set, eyes set and mouth set on the face.
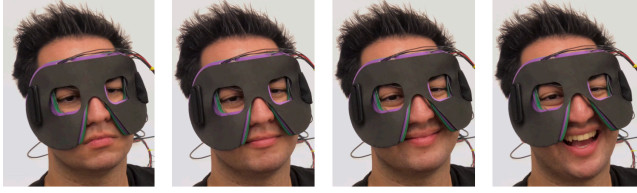
Figure 15. Our smile intensity task included four levels of smile. Examples shown here, left to right: mouth neutral, small smile, medium smile, and large smile.

then performed a leave-one-round-out cross validation, where we trained on nine rounds and test on the tenth (all combinations, results averaged).

7.1 Smile Intensity

In this task, we defined four smile intensities on a linear scale: mouth neutral (which we defined as 0%), small smile (33%), medium smile (lips closed; 66%), and large smile (with teeth showing; 100%). As there is no universal smile ground truth, we let our participants practice and define four smile levels they found comfortable and repeatable. Figure 15 shows an example of these smile intensities. A single round of data collection captured each smile level once, in a random order, during which time 50 frames of sensor data was recorded. In total, participants completed ten rounds. Using our leave-one-round-out cross validation procedure, we found a mean smile intensity error of 8.6% (SD=2.6).

7.2 Lifted Weights

In our next task, participants were asked to hold an exercise weight in their non-dominant arm, with their elbow placed on a table. Figure 16 shows an example of this task. Participants were handed one of five weights – 0, 6, 9, 12 and 15 lbs. – once per round (ten rounds collected total). Once the arm was stable, 50 frames of sensor data were recorded. Using our leave-one-round-out cross validation procedure, we found the mean lifted weight error was 0.04 lbs. (SD=0.04).

Error was so low that we suspected these results might be the product of overfitting to the four weights we tested, and did not operate like a true continuous regression. To overcome this experimental limitation, we ran a second,

Figure 16. Our lifted weights task included five weights.

Figure 17. Our wrist angle task used a Leap Motion controller to provide ground truth. Here a participant sweeps their hand from left to right (with all intermediate angles recorded).

leave-one-weight-out cross validation, where the model was trained on all-but-one weight and tested on the holdout weight (all combinations). Scikit-learn's Extra Trees Regressor only provides interpolation and not extrapolation, which we need when e.g., training on 6, 9 and 12 lbs. and testing on 15 lbs. Thus, for this one experiment, we use Scikit-learn's MLPRegressor (default parameters, identity activation, "LBFGS" solver) to produce estimates. Using this evaluation scheme, trying to predict a weight the classifier had never seen before, we found an average error of 1.59 lbs. (SD=0.44), which we believe is a more realistic estimation of performance.

7.3 Wrist Angle

In Study 2, we evaluated discrete left- and right-bent wrist gestures. In this task, we investigated continuous tracking of the wrist angle. To provide a high-quality ground truth, and also facilitate rapid collection of participant data, we used a Leap Motion controller for wireless tracking of the hand (see setup in Figure 17). Each round of data collection consisted of a slow ~3 second sweep the wrist from roughly -60 to +60° (a range found to be comfortable in pilots), pausing at far left, center, and far right. While participants performed this motion, we recorded 50 sensor frames per second, labeled with the wrist angle reported by the Leap. This procedure was repeated for ten rounds in total.

Using the same leave-one-round-out cross validation procedure, we found a mean wrist angular error of 5.3° (SD=1.0). Figure 18 offers an example plot of Interferi's predicted values tracking against the Leap Motion's output.
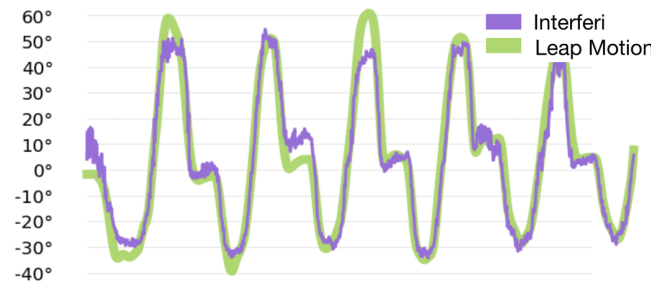

Figure 18. Example wrist-angle tracking from one participant, five rounds of data collection. Leap Motion prediction shown in green; Interferi prediction in purple.

Figure 19. Our coarse hand pose task evaluated tracking of the thumb (A → B) and four other fingers (A → C). We used a Leap Motion controller for ground truth.

7.4 Coarse Hand Pose

In our final task, we focus on Interferi's ability to estimate hand pose. Individual digit tracking is not possible, and so we track two values: the position of the thumb, and the position of the four other fingers as a single unit. Participants were asked to place their hand above a Leap Motion controller, palm facing down, thumb extended and other four fingers held together (see Figure 19). One round of data collection consisted of bringing the thumb to the center of the palm and returning it to full extension, followed by the other four fingers dropping to form a 90° angle with the palm and then returning to full extension. We briefly trained participants to perform this motion slowly, taking roughly 5 seconds to complete, with our system capturing 50 frames per second of sensor data. As in the pervious task, a Leap Motion was used to provide the ground truth labels. In total, ten rounds of data were captured per participant.

Using the same leave-one-round-out cross validation procedure, we found Interferi had a mean tracking error of 1.7° (SD = 0.8°) for the thumb and 6.5° (SD = 1.2°) for the four fingers. Figure 20 offers example plots of Interferi's predicted values tracking against the Leap Motion's output.

8 COMPARISON TO PRIOR RESULTS

Our within-session results compare favorably to other biosensing gesture recognition systems with respect to accuracy and number of gestures supported. We now briefly review prior results to help contextualize Interferi's accuracy.

Starting first with static hand gesture classification, Saponas *et al.* [36] and WristFlex [8] supported a four-class hand gesture set at 77.0% and 69.0% accuracy respectively.
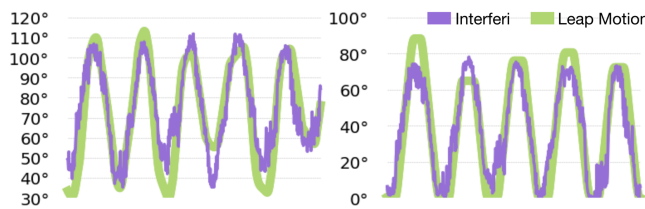

Figure 20. Example thumb (left) and four-finger (right) tracking from one participant, five rounds of data collection. Leap Motion predictions shown in green; Interferi in purple.

On the high end, Jung *et al.* [19] reports 95.4% accuracy across six gestures. We specifically adopted Tomo's [45] two gesture sets so as to directly compare results. The latter system achieved accuracies of 93.1% and 80.9% on the hand and pinch gesture sets, respectively. Later work by Zhang *et al.* [46] achieved 94.3% on the combined set, 94.7% on the hand gesture set, and 94.3% on the pinch gesture set. For reference, Interferi achieves 93.4%, 98.0%, and 90.6% on the same gesture sets. We also note that Tomo does not investigate continuous tracking, nor sensing on the face.

Few systems report across-session accuracy, as this is particularly challenging in worn sensing systems, where even slight alignment variations can lead to large signal changes. The best we found in the literature is Tomo [45], which reports across-session accuracies of 83.4% and 62.6% on its hand and pinch gesture sets, respectively. Interferi is comparable, achieving 80.0% and 60.5% on the hand and pinch gesture sets, respectively (and 65.7% across all eleven hand gestures, which [45][46] do not report).

With regards to continuous hand pose tracking, the most comparable wearable system we found is Digits [21], which can track individual fingers with a mean error of < 9° (6.7° for the thumb specifically). Interferi cannot perform individual finger tracking, except for the thumb, which had a mean angular error of 1.7°. Interferi can track the other four fingers (moving together as a unit) with a mean angular error of 6.5°.

For discrete facial gestures, the closest worn system with a comparable facial gesture set is EarFieldSensing [27], which demonstrates 5 head and face related gestures at 85.3%. Using EMG, Gruebler *et al.* [13] supported smile/frown detection at 88.3%. Interferi achieves 89.0% across 9 face gestures.

Unfortunately, we could not find worn systems that reported accuracies with which to compare our smile intensity, lifted weight and wrist angle study results.

9 LIMITATIONS & FUTURE WORK

Like most biosensing systems, Interferi needs per-user calibration to perform well. Moreover, although our reworn (i.e., across-session) accuracies compare favorably to prior work, it is not yet sufficiently robust for end user applications. Thus, it is likely a rapid calibration will be needed each time the sensor is worn, similar to commercial products such as the Myo EMG armband [39]. It may also be possible to avoid calibration gestures, and instead virtually rotate the sensors to compensate for placement variation, allowing for an existing machine learning model to be used.

Figure 21. Interferi could be integrated into future smart-watch bands and AR/VR headset liners, as seen in these mock-ups.

Another challenge is achieving good acoustic coupling to the body, which is vital for robust performance. We designed our sensors as one-size-fits-most wearables, which worked well enough to complete our user studies. While participant arms were reasonably easy to accommodate, different face morphologies required extra time to ensure a proper fit (by adjusting the head band and resting position). In a real product, it seems likely many sizes would have to be produced. It might also be possible to use more transducers in a denser arrangement, and have software automatically select a subset based on fit.

Another drawback is that our implementation uses relatively large ultrasonic transducers (16 mm diameter, 10 mm thick), produced chiefly for the automotive market (backup parking sensors). These transducers would be challenging to integrate into a thin, smartwatch band. However, the face liner of AR/VR headsets should be feasible. Mockups of these two form factors appear in Figure 21. That said, thinner and smaller transducers are available at other resonant frequencies, which could enable integration into smaller form factors.

Finally, there are immediate avenues to improve our machine learning pipeline, potentially enabling fully continuous pose tracking. We envision using techniques like Hidden Markov Models (HMMs) or Recurrent Neural Networks (RNNs), which take advantage of temporality.

10 CONCLUSION

We have presented Interferi, a novel worn sensing technique that uses acoustic interferometry for on-body gesture recognition. It generates and measures ultrasonic interference patterns using an array of transducers placed on the arm or face. We investigated many worn sensor designs, transducer pairings and phase offsets to identify successful configurations. We evaluated our best designs for the arm and face in two studies: discrete and continuous sensing. Results are promising and competitive with prior worn systems, which we hope spurs interest in acoustic interferometry in HCI.

REFERENCES

[1] Karan Ahuja, Rahul Islam, Varun Parashar, Kuntal Dey, Chris Harrison, and Mayank Goel. 2018. EyeSpyVR: Interactive Eye Sensing Using Off-the-Shelf, Smartphone-Based VR Headsets. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 2, 2, Article 57 (July 2018), 10 pages. DOI: https://doi.org/10.1145/3214260

[2] 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

[3] Toshiyuki Ando, Yuki Kubo, Buntarou Shizuki, and Shin Takahashi. 2017. CanalSense: Face-Related Movement Recognition System based on Sensing Air Pressure in Ear Canals. In Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology (UIST '17). ACM, New York, NY, USA, 679-689. DOI: https://doi.org/10.1145/3126594.3126649

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

[5] Laura A. Brooks and Peter Gerstoft. Ocean acoustic interferometry. The Journal of the Acoustical Society of America 121, no. 6 (2007): 3377-3385.

[6] 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.

[7] Ciprian Adrian Corneanu, Marc Oliu Simón, Jeffrey F. Cohn, and Sergio Escalera Guerrero. Survey on RGB, 3D, thermal, and multimodal approaches for facial expression recognition: History, trends, and affect-related applications. IEEE transactions on pattern analysis and machine intelligence 38, no. 8 (2016): 1548-1568.

[8] 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

[9] 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

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

[11] 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

[12] Anna Gruebler, and Kenji Suzuki. Measurement of distal EMG signals using a wearable device for reading facial expressions. In Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE, pp. 4594-4597. IEEE, 2010.

[13] Anna Gruebler, and Kenji Suzuki. Design of a wearable device for reading positive expressions from facial emg signals. IEEE Transactions on affective computing 5, no. 3 (2014): 227-237.

[14] 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, pp. 453-462. ACM, 2010.

[15] Claude F. Harbarger, Paul M. Weinberger, Jack C. Borders, and Charles A. Hughes. "Prenatal ultrasound exposure and association

with postnatal hearing outcomes." *Journal of Otolaryngology-Head & Neck Surgery* 42, no. 1 (2013): 3.

[16] 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.

[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] N. Inoue, et al. A new ultrasonic interferometer for velocity measurement in liquids using optical diffraction 1986 J. Phys. D: Appl. Phys. 1

[19] 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.

[20] 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

[21] 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

[22] 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

[23] Hao Li, Laura Trutoiu, Kyle Olszewski, Lingyu Wei, Tristan Trutna, Pei-Lun Hsieh, Aaron Nicholls, and Chongyang Ma. 2015. Facial performance sensing head-mounted display. ACM Trans. Graph. 34, 4, Article 47 (July 2015), 9 pages. DOI: https://doi.org/10.1145/2766939

[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] Katsutoshi Masai, Yuta Sugiura, Masa Ogata, Kai Kunze, Masahiko Inami, and Maki Sugimoto. 2016. Facial Expression Recognition in Daily Life by Embedded Photo Reflective Sensors on Smart Eyewear. In Proceedings of the 21st International Conference on Intelligent User Interfaces (IUI '16). ACM, New York, NY, USA, 317-326. DOI: https://doi.org/10.1145/2856767.2856770

[26] Katsutoshi Masai, Kai Kunze, Yuta Sugiura, Masa Ogata, Masahiko Inami, and Maki Sugimoto. 2017. Evaluation of Facial Expression Recognition by a Smart Eyewear for Facial Direction Changes, Repeatability, and Positional Drift. ACM Trans. Interact. Intell. Syst. 7, 4, Article 15 (December 2017), 23 pages. DOI: https://doi.org/10.1145/3012941

[27] Denys J. C. Matthies, Bernhard A. Strecker, and Bodo Urban. 2017. EarFieldSensing: A Novel In-Ear Electric Field Sensing to Enrich Wearable Gesture Input through Facial Expressions. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17). ACM, New York, NY, USA, 1911-1922. DOI: https://doi.org/10.1145/3025453.3025692

[28] 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

[29] 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

[30] 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.

[31] 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.

[32] 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.

[33] PUI Audio 40 kHz Ultrasonic Transducer, http://www.puiaudio.com/pdf/UTR-1440K-TT-R.pdf

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

[35] 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

[36] 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

[37] Jocelyn Scheirer, Raul Fernandez, and Rosalind W. Picard. 1999. Expression glasses: a wearable device for facial expression recognition. In CHI '99 Extended Abstracts on Human Factors in Computing Systems (CHI EA '99). ACM, New York, NY, USA, 262-263. DOI=http://dx.doi.org/10.1145/632716.632878

[38] Katsuhiro Suzuki, Fumihiko Nakamura, Jiu Otsuka, Katsutoshi Masai, Yuta Itoh, Yuta Sugiura, and Maki Sugimoto. 2016. Facial Expression Mapping inside Head Mounted Display by Embedded Optical Sensors. In Proceedings of the 29th Annual Symposium on User Interface Software and Technology (UIST '16 Adjunct). ACM, New York, NY, USA, 91-92. DOI: https://doi.org/10.1145/2984751.2985714

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

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

[41] 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

[42] 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

[43] 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

[44] 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.

[45] 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

Interactive tabletops and surfaces (ITS '13). ACM, New York, NY, USA, 189-198. DOI: https://doi.org/10.1145/2512349.2512821

Software & Technology (UIST '15). ACM, New York, NY, USA, 167-173. DOI: https://doi.org/10.1145/2807442.2807480

[46] 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

[47] 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

[48] 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