Acustico: Surface Tap Detection and Localization using Wrist-based Acoustic TDOA Sensing

Jun Gong[1,2], Aakar Gupta[1], Hrvoje Benko[1]
Facebook Reality Labs, Redmond, WA, USA[1], Dartmouth College, Hanover, NH, USA[2]
jun.gong.gr@dartmouth.edu, aakarg@fb.com, benko@fb.com

ABSTRACT
In this paper, we present *Acustico*, a passive acoustic sensing approach that enables tap detection and 2D tap localization on uninstrumented surfaces using a wrist-worn device. Our technique uses a novel application of acoustic time differences of arrival (TDOA) analysis. We adopt a sensor fusion approach by taking both "surface waves" (i.e., vibrations through surface) and "sound waves" (i.e., vibrations through air) into analysis to improve sensing resolution. We carefully design a sensor configuration to meet the constraints of a wristband form factor. We built a wristband prototype with four acoustic sensors, two accelerometers and two microphones. Through a 20-participant study, we evaluated the performance of our proposed sensing technique for tap detection and localization. Results show that our system reliably detects taps with an F1-score of 0.9987 across different environmental noises and yields high localization accuracies with root-mean-square-errors of 7.6mm (X-axis) and 4.6mm (Y-axis) across different surfaces and tapping techniques.

Author Keywords
Passive Acoustic; Tap Detection and Localization; Wrist

CCS Concepts
Human-centered computing - Human computer interaction (HCI);

INTRODUCTION
As computing devices become increasingly ubiquitous, there is a pressing need for input technologies that can be always available [34]. While wearable devices like smartwatches and smartglasses enable always-available touch input, it comes at the cost of small physical size, which limits user's input due to the fat finger problem [41]. One potential solution is to exploit the surfaces in the environment around us [18], which provide a large area for accurate and comfortable input.

In submission to UIST 2020

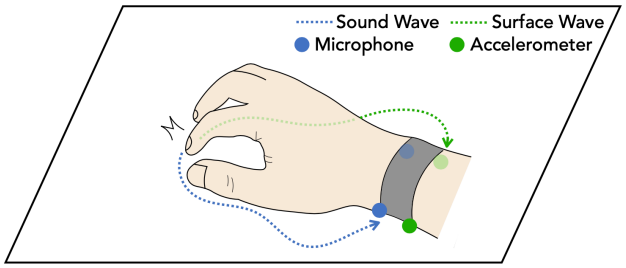

Figure 1: Acustico detects and localizes taps on unmodified surfaces using wrist-worn accelerometers and microphones on the bottom.

Prior work has investigated tracking touch input on unmodified surfaces. However, these efforts primarily use optical schemes with fixed cameras in the environment [2, 7, 30] or wearable cameras [19, 47] that are expensive, power consuming and might draw privacy concerns [3, 22]. Another thread of research focuses on finger-worn ring devices that can track fingertips on the surface using accelerometers [27], optical flow sensor [50], and infrared sensors and gyroscope [26]. However, rings have power and individual sizing constraints that limit their practicality as consumer devices. Smartwatches, on the other hand, have increasingly become popular. But wrist-worn devices that can track touch input on uninstrumented surfaces have been largely overlooked in the literature.

In this paper, we present *Acustico*, a passive acoustic sensing approach that enables tap detection and 2D tap localization on uninstrumented surfaces using a wrist-worn device. This is achieved using a novel application of acoustic time differences of arrival (TDOA) analysis. To overcome the challenges posed by the wristband form factor (e.g., sensors have to be close to each other), we adopt a sensor fusion approach by exploiting the propagation speed difference of "surface wave" (i.e., vibrations through surface) and "sound wave" (i.e., vibrations through air) to better estimate the TDOA and localize the tap (Figure 1). Through a careful design procedure, we come up with an optimal sensor configuration that meets wristband constraints and has better sensing performance. To validate the use of *Acustico*, we implemented a proof-of-concept wristband prototype with four acoustic sensors, two accelerometers and two microphones. We tested the system for tap detection and 2D tap localization. For tap detection, we evaluated the system under different types of environmental noise. For 2D tap localization, we evaluated

the system on five different surface materials (wood, steel, plastic, glass and fabric), using two tap methods (finger-pad or finger-nail) and two hand configurations (one-hand or two-hands). Results from 20 participants show that our system can reliably detect taps under different noise conditions and achieve average error distances of 7.57mm (s.e. = 0.20mm) and 4.62mm (s.e. = 0.11mm) in two axes across all different test conditions.

Our primary contributions include: 1) a sensor fusion approach for wrist-worn devices to detect and localize taps using acoustic TDOA analysis; 2) design and development of a prototype using two different types of acoustic sensors and customized software; and 3) a validation of this approach through a series of experiments.

RELATED WORK

This work builds and extends on prior work in many areas, including touch input on surfaces, wrist-based gesture sensing, active sensing, and passive acoustic localization.

Touch Input on Surfaces

Research on touch input on surfaces can be mainly divided into two categories, touch input on instrumented surfaces and touch input on uninstrumented surfaces.

Touch Input on Instrumented Surfaces. Plenty of existing work supports touch sensing by instrumenting or modifying the surface with capacitive [29, 43, 59], optical [16, 32], electrical impedance [51, 55, 58] and acoustic sensors [36, 37]. For example, Wall++ [59] uses conductive paint for patterning large electrodes onto a wall, turning ordinary walls into smart infrastructure supporting capacitive touch tracking. TouchLight [45] presents a touch screen display system on a sheet of acrylic plastic by instrumenting two video cameras behind the semi-transparent plane. Electrick [58] is a low-cost electrical impedance sensing technique enabling touch input on a surface painted with conductive coating. Since the surfaces always need to be instrumented or modified before sensing touch, these are not always practical.

Touch Input on Uninstrumented Surfaces. Cameras in the environment or worn on user's body allow sensing touch without instrumenting or modifying the surfaces. There are many existing optical schemes for touch sensing in the literature, including RGB cameras [2, 7, 25, 30], infrared cameras [1] and thermal cameras [39]. However, these approaches still require fixing the cameras in the environment or using wearable cameras [19, 47], which are expensive, power consuming and might introduce privacy issues [3, 22]. Aside from cameras, work has also been done to exploit a ring with IMU and light proximity sensor to sense touch [15] or track fingertip movements on uninstrumented surfaces [26]. None of the existing work uses wrist-based sensing. With the growing popularity of smartwatches, combined with their small touchscreen input space, enabling such surface input via wrist-based sensing holds a high potential for impact.

Gesture Sensing using a Wristband / Smartwatch

Another body of related research focuses on sensing finger gestures (e.g., pinch) [14, 31, 40, 44] and hand gestures (e.g., fist) [9, 11, 23, 33, 38, 42, 56] using the sensors on a wristband or a smartwatch. For example, GestureWrist [38] uses capacitive sensors to detect the changes in forearm shape to infer hand postures. CapBand [42] uses a similar ultra-low power capacitive sensing technique but achieves significant improvements on accuracy and gesture quantity. WristFlex [9] and Tomo [56] use force resistors or electrical impedance tomography (EIT) sensors to identify different hand postures. WristWhirl [13] supports 2D continuous input from wrist whirling using infrared proximity sensors on the wristband.

Localization using Active Sensing

Previous work has also shown the possibility to detect or localize a finger/finger-tap using active sensing methods, which use infrared [4], electrical [57, 60], magnetic [8] or acoustic signals [35, 54]. These approaches typically involve active transmission of a signal from a transmitter node and then analyzing the reception of that signal at the receiver node. FingerIO [35], for example, transforms the device into an active sonar system that transmits inaudible sound signals and tracks the echoes of the finger at its microphones. We investigated an active sensing approach, but there are fundamental constraints in the physics of this approach (detailed later in Discussion).

Gesture Input and Localization with Passive Acoustics

Many passive acoustic approaches have been employed to detect gestures [12, 20, 21, 52] and localize signals [17, 24, 36, 37, 48, 49]. For gesture inputs, SurfaceLink [12] exploits a combination of accelerometers and microphones to sense gestures and uses them to control information transfer among devices. ScratchInput [20] relies on the unique sound produced when a fingernail is dragged over different surfaces to identify six scratch gestures.

The TDOA approach. For localizing the acoustic signal, the most prevalent approach, which we also use in this work, is time difference of arrival (TDOA) analysis [10, 24, 36, 37, 48, 49]. For example, PingPongPlus [24] instruments four contact microphones located at the outermost corners of the desired interactive region. When a Ping-Pong ball falls inside of this region, the signal arrives to the four sensors at different times, enabling a hyperbolic intersection localization. Likewise, SurfaceVibe [36] also uses four geophones in a similar setup to enable two interaction types, tap and swipe, on multiple types of surfaces. Instead of instrumenting the surfaces, Toffee [48] augments the mobile devices and laptops with four piezo sensors and demonstrates accurate resolution of the bearings of touch events around the devices, although the evaluation considers only a single user. Besides TDOA analysis, SoundCraft [17] instead uses a target signal subspace method, by adopting the basic idea of Multiple Signal Classification technique, which can localize acoustic signals even in noisy environment.

Unlike the existing work, we are the first to embed two types of acoustic sensors in a wristband form factor to detect and localize finger-taps on uninstrumented surfaces. The wristband location and form-factor introduce two primary challenges: (1) Small distance between sensors: Due to the small form factor of the wristband, the distances between sensors are extremely small, which inherently requires a more precise estimate of TDOA; (2) Inconsistent coupling between sensors and surface: Since the accelerometers are embedded on the bottom of the wristband, the coupling between the accelerometers and surface highly depends on how users put their hands on the surface. Further, the accelerometer signals may also be affected by any hand movements right before tapping. Both introduce noises and instability in the received signals at the accelerometers. Our work overcomes these challenges and is the first work to our knowledge that detects and localizes taps on uninstrumented surfaces using wrist-based sensing.

ACUSTICO: SENSING PRINCIPLE

When a finger taps a table, the force applied to the surface causes deformation. As the contact point is relieved of the force, the surface retracts due to its elasticity, which generates vibrations propagating outward from the point of contact. On one hand, the vibration propagates through the surface, which we call the "surface wave". The speed of the "surface wave" depends on the surface medium. In solid materials, such as wood, "surface wave" typically propagates at around 600 meters per second [24]. On the other hand, the vibrations also propagate through air, which we call the "sound wave". The speed of the "sound wave" is relatively low, because air is compressible. In common indoor environments (20°C), the speed of the "sound wave" is about 343 meters per second [46].

In this work, we take both "surface wave" and "sound wave" into consideration. To capture these two waves, our wristband prototype has four sensors underneath, two accelerometers for capturing the "surface wave" and two microphones for capturing the "sound wave" (see implementation details later).

The sensor fusion approach is beneficial for both tap detection and tap localization. For tap detection, a tap is registered only when both waves are detected and pass a pre-defined threshold, which prevents many false detections if we only use a single type of sensor. To be specific, if we only use microphones, the system might not be able to work under a noisy environment. And if we only consider the data from accelerometers, multiple false positives may be introduced due to random hand motion.

For tap localization, we used the time differences of arrival (TDOA) between each pair of the sensors to interpolate the tap location. To help explain, let us first consider an example of two sensors of the same type. If a tap occurs equidistant to the two sensors, the time difference of arrival will be the same. And the system can conclude that the tap occurred along a line equidistant from the two sensors. If the tap is closer to one sensor than the other, the tap can be inferred to lie somewhere along a hyperbolic curve, a set of points having a constant difference of the distances to two fixed points (i.e., two sensor locations). With more sensors, the tap location can be determined by calculating the intersections of multiple hyperbolic curves. A similar approach is used in Toffee [48] and SurfaceVibe [36]. However, we are targeting a wristband device which has specific constraints as mentioned earlier. The distances between each pair of the sensors have to be small (in our case, we assume 4cm along the length of the wrist and 1.5cm along the width of the band). This makes the differences of TDOAs at different tapping locations extremely small (e.g., 1 - 30µs) and hard to detect. To overcome these issues, we calculate the TDOA *between* accelerometers and microphones. Since the propagation speeds in surface and air are different, this adds extra time differences by using the data from two different types of sensors. We further use a high sampling rate DAQ (data acquisition device) (1MHz) to increase the sampling rate so that we can capture such a small time difference.

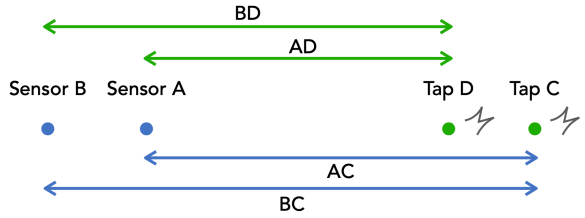

Figure 2: TDOAs from taps C and D are same for sensors A and B if they are both mics or both accelerometers, but different if one is mic and other is accelerometer.

Let us consider a simplified example shown in Figure 2. We assume there are two taps (C, D) that are aligned with two sensors (A, B) on a wooden surface. The signals that are captured by Sensor A and Sensor B have propagation speeds of V_A and V_B respectively. Here, the time difference of arrival between Sensor A and Sensor B when tapping at C should be:

$$TDOA_C = \frac{BC}{V_B} - \frac{AC}{V_A}$$

Similarly, time difference of arrival between Sensor A and Sensor B when tapping at D should be:

$$TDOA_D = \frac{BD}{V_B} - \frac{AD}{V_A}$$

Our goal is to maximize the difference between these two TDOAs so that the two taps can be easily distinguished:

$$Maximize\ (|TDOA_C - TDOA_D|)$$

If Sensor A and Sensor B are of the same type, then the propagation speeds V_A and V_B would be the same as well. The difference between the two TDOAs (both equal to AB/V) is *zero*. However, if Sensor A is an accelerometer and Sensor B is a microphone, then we have $V_A \approx 2V_B$

(600m/s vs. 343m/s). In this case, the difference between two TDOAs is:

$$|TDOA_C - TDOA_D| \approx \left(\frac{2BC}{V_A} - \frac{AC}{V_A}\right) - \left(\frac{2BD}{V_A} - \frac{AD}{V_A}\right) = \frac{CD}{V_A}$$

From this simplified example, we can see that extra time differences can be added by calculating the TDOA between accelerometer and microphone. And with larger time differences between two taps, the localization accuracy can be improved accordingly.

SENSOR CONFIGURATION

Now that we know we require the two sensor types, the next question is what should be the configuration of those sensors on the wristband. In this section, we discuss the pros and cons of different potential sensor configurations and provide the rationales of our final prototype design.

We define the X-axis to be the wrist's radial-ulnar and the Y axis to be the orthogonal direction (Figure 3). In theory, localization along X axis would always be easier than that along Y axis. To explain, let us consider another simple example about two taps along X axis (C, D, Figure 3(a)) and Y axis (E, F, Figure 3(b)) using two sensors of the same type (e.g., Sensor A and B are both microphones).

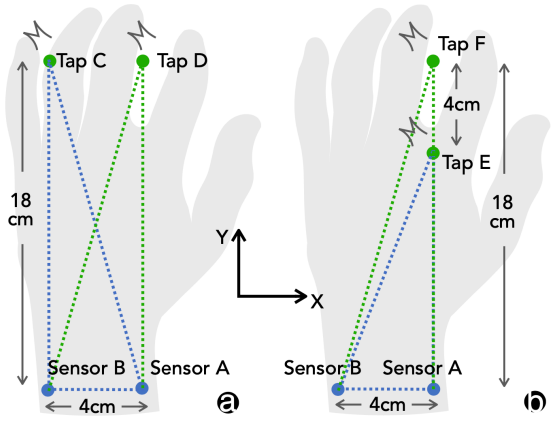

Figure 3: (a) Two taps along X axis, (b) Two taps along Y axis.

Assuming a signal propagation speed of V, the difference between the two TDOAs when tapping on C and D (X axis) and E and F (Y axis) should be respectively:

$$\text{X-axis: } |TDOA_C - TDOA_D| = \left|\left(\frac{BC}{V} - \frac{AC}{V}\right) - \left(\frac{BD}{V} - \frac{AD}{V}\right)\right|$$

$$\text{Y-axis: } |TDOA_E - TDOA_F| = \left|\left(\frac{BE}{V} - \frac{AE}{V}\right) - \left(\frac{BF}{V} - \frac{AF}{V}\right)\right|$$

We assume the distance between Sensor A and B to be 4 cm (a common wrist length), the distance between Sensor A and Tap D to be 18cm (a common hand size), the distance between the two taps (C and D, E and F) to be 4cm, and the speed of sound to be 340m/s. Populating the equations with these numbers, we get the differences of two TDOAs along X axis and Y axis as 25.8μs and 3.6μs respectively. This is why X axis localization is easier than Y axis. A similar conclusion can also be found in Toffee [48] and SoundCraft [17] since both solutions only work in angular estimation (similar to X-axis) instead of distance interpolation (similar to Y-axis).

Based on this fact, we chose to place two different types of sensors along Y axis to add the extra time difference. We decided to use four sensors in total (two for each type of sensor) to reduce the error with additional information [17, 48]. With these constraints, we have three possible sensor configurations (Figure 4).

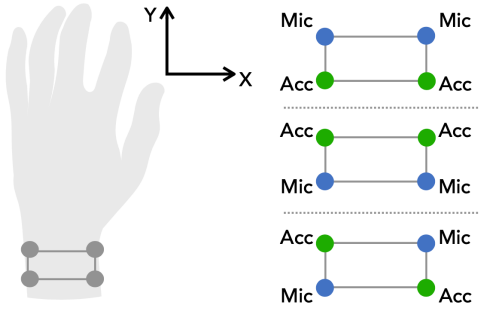

Figure 4: Three possible sensor configurations given that two different types of sensors should lie along the Y-axis. The topmost configuration was chosen to keep mics closer to taps.

Our initial tests showed that the signal to noise ratio (SNR) of the accelerometer is larger than the SNR of the microphones when tapping on the surface probably due to higher attenuation of the signals in air. We therefore chose the topmost setup as our final sensor configuration. After investigating the common wristband sizes, we used a rectangular configuration, where two microphones (or accelerometers) are separated by 4cm and distance between accelerometer and microphone is set to 1.5cm.

HARDWARE IMPLEMENTATION

Our wristband prototype consists of two highly sensitive accelerometers (Model 352A24, PCB Piezotronics) and two MEMS microphone breakouts (INMP401, Sparkfun) with built-in amplifiers (gain = 67dB). Four sensors are situated on the bottom of the wrist with a mechanical structure (Figure 5(a,b,c)).

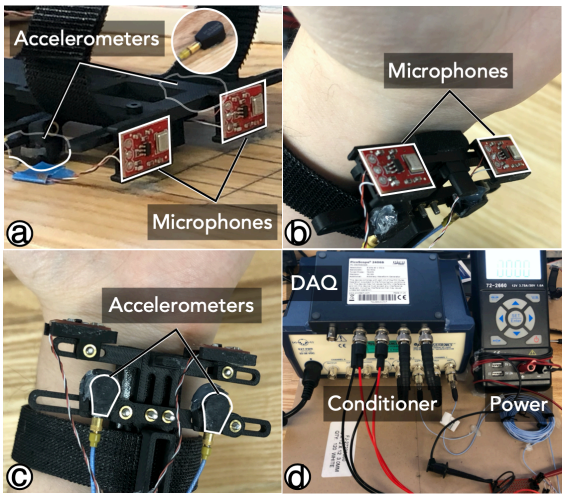

Figure 5: (a, b, c) Acustico wristband, (d) External hardware.

The accelerometers were connected to an ICP signal conditioner (Model 482C24, PCB Piezotronics) with a gain of 10. Since the TDOAs we targeted are on the order of microseconds, we needed a sampling rate of at least 1MHz. We therefore used the PicoScope 2406B for sampling the amplified signals from both accelerometers and microphones and set its sampling rate to 1MHz with the raw data streaming into a desktop via USB.

SOFTWARE IMPLEMENTATION

Our software engine that detects and localizes taps was developed in Matlab. The software pipeline has two stages: First we detect whether there is a tap on the surface, and then we localize the tap using regression models. Our implementation utilizes the machine learning toolbox in Matlab. Once trained, our system can work in real time.

Tap Detection

The basic premise here is to detect surface taps by looking at the peaks they cause in the accelerometer signal. The primary challenge here is in filtering out instances when the user moves their wrist randomly or purposely to position the finger over the target before tapping. We describe our algorithm below and test its tolerance to such instances in our evaluation.

To detect a tap, we first slice the streaming data in 0.1s windows. Within each window, the raw data from four sensors are filtered by a bandpass filter with cut-off frequencies of 10Hz and 1000Hz [36]. We then sum up the data from the two accelerometers and the two microphones separately. For accelerometer data, in order to distinguish a tap from other coarse hand movements, we search for a "pulse" having a stronger signal power (i.e., sum of the squares for each data point) than a pre-set threshold within a short time period (i.e., 0.01s). We pick this time period based on our observation of the tap signal peak length and it is also long enough for us to ignore the small time shift between each pair of sensors. To implement this, we divide the data within the 0.1s window into 10 pieces and calculate the signal power for each piece. If the power of one piece is higher than the threshold, and the neighboring pieces show a much lower power (i.e., 30% of the selected piece), the accelerometer requirements are satisfied. Then we check the microphone signal power in the exact same 0.01s time slot. If it is also higher than the pre-set threshold, we assume there is a tap on the surface. The two thresholds for accelerometers and microphones are determined by a calibration process.

Tap Localization

We localize taps relative to the location of the wristband in discrete regions. Aside from precise TDOA estimation, another challenge for tap localization is in the inconsistent coupling between the accelerometers and the surface. When the user wears the wristband, the coupling of the sensors to the surface may keep slightly varying due to the wrist motion even while the forearm stays in the same location. This inconsistency makes it difficult to reliably calculate TDOA simply using mathematical triangulation.

Therefore, we chose to use a machine learning regression model to estimate the tap's 2D coordinate. Before we extract features from the raw data, we first concatenate the data with the data from the previous window and the next window to ensure that a complete tap signal is captured, and the same signal is not captured twice. And then we trim the data into 0.1 second by localizing the tap signal through maximum detection (i.e., only includes 0.03 second before the maximum of the data and 0.07 second after the maximum, a sufficient time slot ensuring the inclusion of a complete tap signal based on our observation). We use these data in 0.1s tap windows for feature extraction.

Feature Extraction and Machine Learning

Based on the findings from previous work [36, 48] and our initial tests, we use four different methods to estimate the TDOA between two signals in 0.1s tap windows: (1) time displacement when the cross-correlation (i.e., similarity of two signals as a function of the displacement of one relative to the other) reaches maximum; (2) time difference of the first peaks; (3) time difference of the maximum peaks; (4) time difference of minimum. In total, we feed a 24-feature vector (4 estimate methods × 6 pairs of sensors) into the machine learning model for localization.

We use Random Forest in our current implementation. Random Forest has previously been found to be accurate, robust, scalable, and efficient in many different applications [6, 31]. We use two independent Random Forest regression models (nTrees = 200) which operate in parallel – one for X position and the other for Y position.

USER EVALUATION

We ran a user evaluation to characterize the robustness and accuracy of our system in tap detection and localization.

Participants

Twenty right-handed participants (11 males, 9 females; 23-59 years old, average age: 39.2) were recruited to participate in this study from our organization. The participants were compensated for their time.

We marked the tap evaluation region (Figure 7, 8), which is a three by six grid (18 squares in total). Each square was a 1cm × 1cm target for participants to tap, a reasonable target size considering the size of the finger-pad. This tap evaluation region was determined based on the comfortable area of interaction for index finger taps given the comfortable limits on flexion-extension and radial-ulnar deviation of the wrist on a surface.

Experimental Setup

The study was conducted using our wristband prototype described in implementation section. It was conducted in a quiet room with the participant and experimenter. Except for the sofa arm condition, all experiments were completed on a large table with enough space to place each of the four stools made of different materials. The experimental

interface was shown on a 27-inch monitor, placed on the same table at a comfortable distance from the participant. Prior to the evaluation, participants were asked to wear the prototype on the wrist of their left hand (non-dominant hand). In order to control for the effect of hand movements on the accelerometers (e.g., the closer to the wrist, the larger the influence), we required participants to put the wristband about 2cm away from the first knuckle of the wrist, a common position where users would wear a watch. We then asked the participants to place their wrist on the surface along the middle line of the tap region (Figure 8) such that they could use their index fingers to tap at each corner target easily and comfortably. We recorded that wrist position (i.e., 14cm - 19cm to the evaluation region) and kept it the same throughout the duration of the study. Note that this is only to ensure the distances between taps and sensors are the same across sessions (e.g., training vs. testing). When being used in real time, the system does not require wrist or elbow to be fixed in one position.

Evaluating Tap Detection under Noise & Wrist Motion

For tap detection, we tested if users' taps can be reliably detected in noisy environments while the hand performs random wrist motion between taps. Participants tapped on a wooden stool surface under two different noise conditions. Before they started, the device needed a simple calibration process to capture appropriate thresholds for both accelerometers and microphones. For the calibration, participants were told to tap at four corners of the tap region once and we used the half amplitude of the "lightest" tap as the thresholds. Note that we did not give any instruction on how to tap in this study.

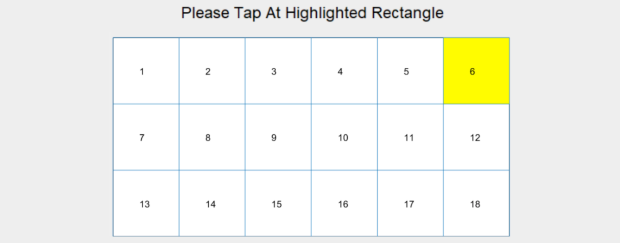

Figure 6: User interface that guided the user.

After the device was calibrated, a speaker was placed near the stool to play two different environmental noises, city traffic white noise[1] and ambient noise[2] at 80dB to simulate the acoustic noises the user may encounter in their daily activities [53]. Under each noise, participants were instructed to tap at the eighteen different target squares in a random order following the sequence shown on the experimental interface (Figure 6). To investigate the effect of hand movements on the accelerometers, we asked the participants to move their wrist randomly (left-right, up-down etc.) (without displacing their arm from its position on the surface) for about five seconds before they tapped. During the study, the experimenter manually recorded false positives and false negatives. The study took about 15 minutes to complete. In total, we had 720 taps (20 participants × 18 taps × 2 environmental noises).

Results – Tap Detection under Noise & Wrist Motion

We used F1 score to measure the accuracy of tap detection, which is defined as (2 × precision × recall) / (precision + recall). The F1 scores were analyzed using a one-way ANOVA. Violations to sphericity used Greenhouse-Geisser corrections to the degrees of freedom.

Overall, the average F1 score for tap detection under two environmental noises is 0.9987 (s.e. = 0.0003) with precision of 0.9988 (s.e. = 0.0004) and recall of 0.9986 (s.e. = 0.0004). We found no significant effect of two environmental noises on F1 scores ($F_{1, 19} = 0.998$, $p > 0.05$). The tap detection results are promising since only 17 false positives and 20 false negatives were found throughout the study, which demonstrates the robustness of tap detection algorithm under different noise conditions.

Study Design – Tap Localization

For tap localization, we conducted the investigation in two phases. The first phase investigated tap localization on a wooden surface (which is one of the most common materials used for tables) under two independent variables: hand configuration and tap method. The second phase investigated taps on different surfaces.

Phase 1 - Hand Configuration (One-Hand vs. Two-Hands): We wanted to isolate the effect of wrist motion on the accelerometers. Thus, we investigated a Two-Hands configuration where participants were asked to put their device-worn hand on the surface and use the other hand to tap (Figure 7(b)). Since the device-worn hand was kept still when the tap occurred, the accelerometers on the bottom were not affected by any hand movement but only captured the "surface wave" propagating from the tap location.

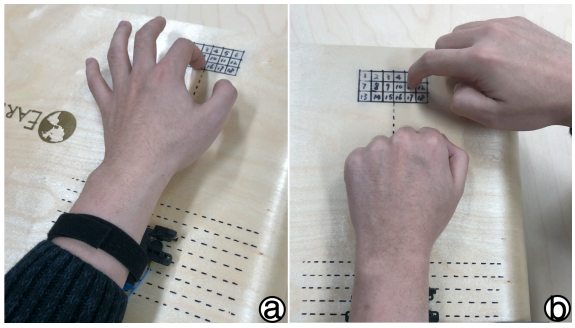

Figure 7: (a) One-Hand, and (b) Two-Hands configuration.

Phase 1 - Tap Method (Finger-pad vs. Finger-nail): We considered how users tapped on the surface when they tapped naturally (typically using the finger-pad) or when they purposely tried to incorporate the fingernail in the tap. Based on our initial observations, we found out that using fingernail to tap could create a "shorter but stronger" pulse

[1] https://www.youtube.com/watch?v=8s5H76F3SIs
[2] https://www.youtube.com/watch?v=fuwGT88P-RU

with less frequency components, which might ease the TDOA estimation between each pair of sensors.

Phase 2 - Surface Material: Since surfaces made of different materials have different properties and wave propagation velocities[3], we investigated the localization accuracy on different surfaces. We tested on four additional surfaces which constitute common table surfaces - plastic, glass and steel, and a fabric sofa arm that represented a non-rigid surface (Figure 8). For this investigation, we only looked at the Finger-pad, One-Hand scenario to ensure that the entire study did not exceed 90 minutes. Including the data for the same scenario for Wood surface from Phase 1, we had five different surfaces.

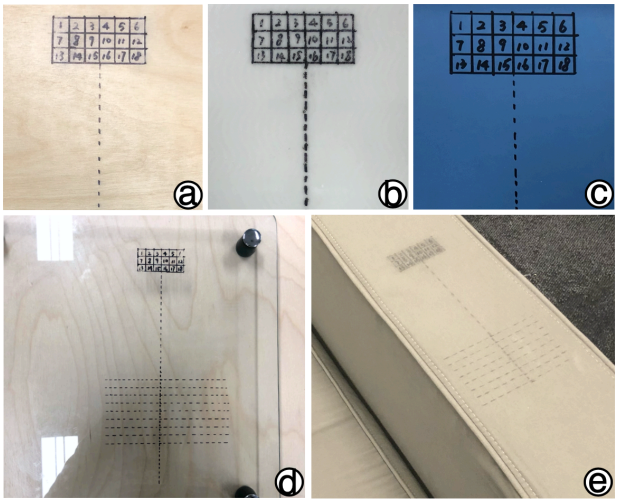

Figure 8: Five surface materials. (a) Wood, (b) Plastic, (c) Steel (painted), (d) Glass, (e) Fabric sofa arm.

In summary, we had 2 tap methods (finger-pad vs. finger-nail) × 2 hand configurations (one-hand vs. two-hands) = 4 conditions in Phase 1, and 4 surface condition (plastic, glass, steel, fabric) in Phase 2. Each condition consisted of 18 tap locations (Figure 8). Participants did four repetition sessions in each of the four conditions in Phase 1 and in each of the four conditions in Phase 2. The conditions in Phase 1 and Phase 2 were separately counterbalanced. In total, we had 8 conditions (4 Phase 1 + 4 Phase 2) × 18 taps × 4 sessions × 20 participants = 11520 taps.

Study Procedure – Tap Localization

For each test condition, the procedure was similar to the tap detection study. For each surface, participants were first asked to calibrate the device using the exact same process described in tap detection study. Unlike the detection study, no environmental noise was provided, and participants were not required to move their hands randomly before each tap. To save time, participants performed taps sequentially (i.e., from square 1 to 18) in each repetition session. When a participant's tap was registered and recorded, there was a "click" sound to notify the participant to move to the next target. The experimental interface also highlighted the next tap target accordingly. To ensure we could collect all eighteen taps in each session for localization, if a tap was not detected, participants were instructed to tap at the same location again until it was registered. The experimenter recorded all false detections manually for later analysis. A one-minute break was given between repetition sessions where participants were asked to take off the wristband, leave the desk and walk around in the room [14, 28, 56]. Note that the calibration was only performed before the first session on each surface. The whole study took about 90 minutes to complete.

Results – Tap Localization

We present experiment results to demonstrate the accuracy and reliability of our system. Data were analyzed using a one-way ANOVA with respect to the five different surface different materials and a two-way repeated measures ANOVA with respect to tap methods (i.e., finger-pad or finger-nail) and hand configurations (i.e., one-hand or two-hands) for the Wood surface. Violations to sphericity used Greenhouse-Geisser corrections to the degrees of freedom. Post-hoc tests with Bonferroni corrections were used.

Tap Detection Performance

We looked at the numbers of false positives and false negatives during the study and evaluated the tap detection performance. Overall, the average F1 score for tap detection is 0.9995 (s.e. = 3×10^{-5}) with a precision of 0.9999 (s.e. = 4×10^{-6}) and recall of 0.9990 (s.e. = 6×10^{-5}). There was only one false positive across the whole study. And 242 false negatives occurred in total because the participant tapped much lighter than the taps in the calibration process. The ANOVA yielded a significant effect of surface on F1 scores ($F_{4, 76} = 7.793$, $p < 0.01$). Post-hoc pair-wise comparisons revealed significant differences between sofa arm and all other rigid surfaces except for the wood surface (all $p < 0.05$), which indicated that although we achieved a F1 score as high as 0.9989 (s.e. = 1.6×10^{-4}) for the sofa arm, it was still not as good as other rigid surfaces. There was also a significant effect of tap methods ($F_{1, 19} = 5.589$, $p < 0.05$) on F1-score. It showed that tapping with fingernail would be easier to detect since the received signals were "stronger and sharper". We found no significant effect of hand configuration ($F_{1, 19} = 3.416$, $p > 0.05$) which indicates that our tap detection algorithm minimized the effect of coarse hand movement on the accelerometers.

Tap Localization Performance

Since participants' hand sizes were different, the tap region in the evaluation was actually in different distances from the sensors on the wristband (i.e., 14cm - 19cm) among participants. It meant that it was not possible to create a cross-user model that worked for everyone, especially also considering the differences in how participants tapped on the surface. Further, as discussed before, different materials have different dispersion/reflection properties

[3] https://www.engineeringtoolbox.com/sound-speed-solids-d_713.html

and wave propagation velocities, making it hard for the model to be generalized across different surfaces.

Thus, we chose to evaluate the tap localization accuracy within each test condition using root-mean-square error (RMSE) measurement. The error was measured from the center of each target. We calculated the leave-one-session-out accuracy for each participant under each test condition by training a model using the data from three sessions and testing it using the remaining session. The average RMSE for each participant in each test condition was calculated by averaging all 4 possible combinations of training and test data. The overall accuracy was then averaged using the RMSEs from all participants.

Overall, the average RMSEs in X axis and Y axis across all eight tested conditions were 7.57mm (s.e. = 0.20mm) and 4.62mm (s.e. = 0.11mm) respectively. In particular, if we removed sofa arm condition, the average RMSEs decreased to 7.08mm (s.e. = 0.19mm) and 4.36mm (s.e. = 0.11mm) in X axis and Y axis (Figure 9 left).

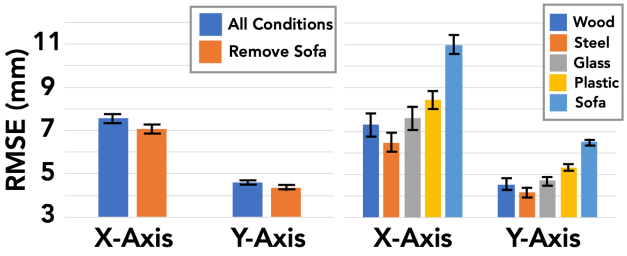

Figure 9: Tap localization RMSEs for X and Y-axis for all conditions and when the non-rigid fabric surface is excluded (left); Tap localization RMSEs across all surfaces under the Finger-pad, One-Hand scenario (right). Error bars show ± SE in all figures.

Surface. A significant effect of surface was found on the RMSEs for the X axis ($F_{4, 76} = 15.088$, $p < 0.01$) and the Y axis ($F_{4, 76} = 20.307$, $p < 0.01$) both. Post-hoc pair-wise comparisons revealed significant differences between sofa arm and other rigid surfaces (all $p < 0.05$) in both X axis and Y axis. The respective average RMSEs in X and Y axis were: Wood: 7.30mm (s.e. = 0.50mm), 4.54mm (s.e. = 0.30mm); Steel: 6.47mm (s.e. = 0.46mm), 4.17mm (s.e. = 0.25mm); Glass: 7.57mm (s.e. = 0.54mm), 4.71mm (s.e. = 0.21mm); Plastic: 8.44mm (s.e. = 0.43mm), 5.34mm (s.e. = 0.17mm); Sofa Arm: 10.98mm (s.e. = 0.44mm), 6.49mm (s.e. = 0.13mm) (Figure 9 right).

It was expected that we achieved the lowest accuracy on the sofa arm since "surface wave" propagation is more complicated in the non-rigid surface. Our system performed the best on steel surface. The reason might be two folds. Firstly, the steel surface is homogeneous, which reduces the dispersion/reflection of the "surface wave". Second, taps on the steel surface create the strongest "sound wave" among these five materials.

Tap methods and hand configurations. On the wood surface, for X axis, we found a significant effect of hand configurations ($F_{1, 19} = 9.105$, $p < 0.01$). However, no significant effect was found for tap methods ($F_{1, 19} = 2.734$, $p > 0.05$). As for Y axis, we found significant effects of tap methods ($F_{1, 19} = 5.659$, $p < 0.05$) and hand configurations ($F_{1, 19} = 7.645$, $p < 0.05$) both. We found no significant interaction effect for tap methods × hand configurations in both X axis and Y axis (both $p > 0.05$).

The average RMSEs in X axis and Y axis when tapping using finger-pad were 7.06mm (s.e. = 0.35mm) and 4.31mm (s.e. = 0.23mm) respectively while they dropped to 6.49mm (s.e. = 0.36mm) and 3.83mm (s.e. = 0.21mm) when using the fingernail to tap (Figure 10 left). Based on the results, for X axis location interpolation, using fingernail did not help but it did help Y axis location interpolation. One possible reason could be that tapping with fingernail improved the TDOA estimation between two different types of sensors (i.e., accelerometer and microphone) since the received signals from both types of sensors have clearer and stronger peaks.

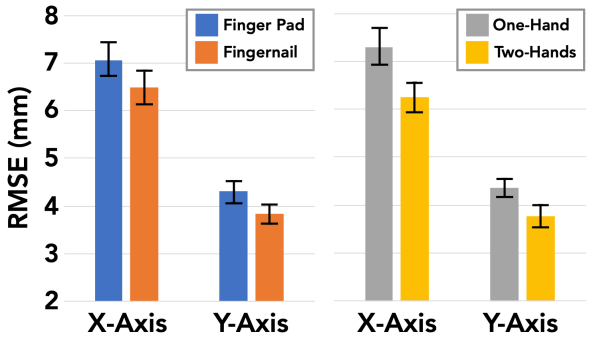

Figure 10: On the wood surface: Tap localization RMSEs for Finger-pad vs. Finger-nail (left); Tap localization RMSEs for One-Hand vs. Two-Hands configurations (right).

For hand configurations, the average RMSEs in X axis and Y axis in the one-hand condition were 7.30mm (s.e. = 0.37mm) and 4.36mm (s.e. = 0.20mm) respectively. In the two-hands condition, the RMSEs could decrease to 6.25mm (s.e. = 0.32mm) and 3.78mm (s.e. = 0.23mm) correspondingly (Figure 10 right). From the results, we demonstrated that the tap localization performance could be further improved in two-hands configuration since the effect of hand coarse movements on the accelerometers was completely removed in this situation. With this, we also envision a two-hands usage scenario, which can support applications that might require higher sensing resolution and accuracy.

Feature Importance. Aside from the accuracy, we were also interested in which features played more important roles in localization. A weighted breakdown of merit was calculated using normalized Random Forest weights. First of all, in order to see which TDOA estimation method performed better in our configuration, we summed up the

normalized weights of six TDOA features calculated by each method and averaged them from both axis (Figure 11 top). It turned out that TDOAs calculated using "the time displacement when the cross-correlation reaches maximum", contributed the most for tap localization. This might be because only this estimation method considered the complete signals from the four sensors. Second, we compared the TDOA features from each pair of the sensors in X axis and Y axis by summing up the corresponding weights from four TDOA estimation methods (Figure 11 bottom). As expected, the two TDOAs calculated from the sensors of the same type (Acc1 – Acc2, Mic1 – Mic2) were more important for X axis localization, validating our discussion earlier. For Y axis localization, the TDOAs from different types of sensors (Acc - Mic) proved more important, which again validates our idea of using the propagation speeds difference in surface and air to add extra time difference and improve the sensing resolution.

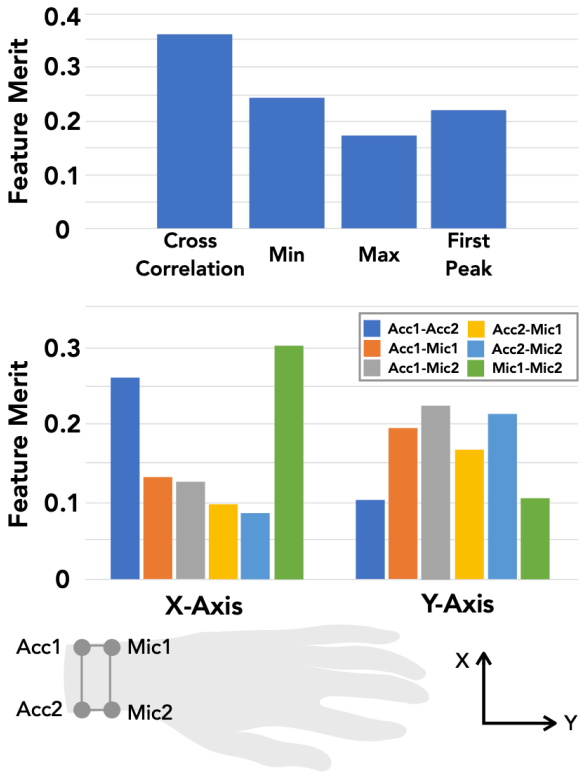

Figure 11: Feature importance for tap localization among the four TDOA estimation methods (top), and the six sensor pairs (bottom).

With Less Training Data. We also analyzed the system accuracy with less training data. We still took leave-one-session-out approach but this time we only used the taps from four corners in three sessions (i.e., 12 taps) to train the machine learning model. And we tested the model on the remaining sessions (i.e., 18 taps).

Overall, with less training data, the average RMSEs in X axis and Y axis across all eight tested conditions increased to 9.64mm (s.e. = 0.18mm) and 5.79mm (s.e. = 0.08mm) respectively. If we removed sofa arm condition, the average RMSEs could decrease to 9.16mm (s.e. = 0.16mm) and 5.63mm (s.e. = 0.08mm) in X axis and Y axis. These results showed that our system could still achieve a reasonable localization accuracy with a small set of training data. The amount of collected training data should depend on different application requirements.

EXAMPLE APPLICATIONS

We built four demo applications to showcase the potential use cases of *Acustico*. The first three applications show how *Acustico* can be used with AR devices (e.g., Microsoft HoloLens) to enrich the input expressiveness. The last application provides a coherent "mouse experience" by combining *Acustico* with an optical flow sensor.

AR Applications: The first application we built is a dial pad for AR devices. User can simply tap at different locations to enter a phone number and call the person he wants to contact (Figure 12(a)). Similarly, we also developed a calculator for AR devices. User can input the numbers on any surfaces nearby and get the calculation results (Figure 12(b)). Our third application is a whack-a-mole AR game. User can hit moles by taping on the surface, which offers an immersive gaming experience with corresponding haptic tactile feedback (Figure 12(c)).

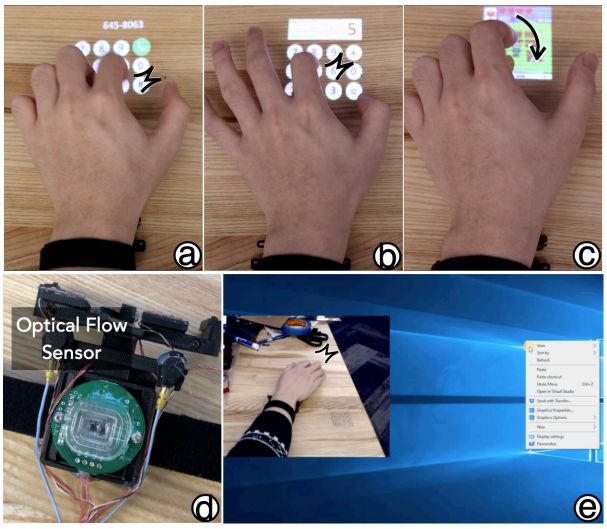

Figure 12: Demo applications. (a) Dial pad, (b) Calculator, (c) Whack-a-mole game, (d) Acustico with an optical flow sensor, (e) User taps on the right side to "right click".

Optical Flow + Acustico for Surface-wide Localization: Acustico focuses on tap localization relative to the wristband. We combine the *Acustico* with an optical flow sensor underneath to additionally track the forearm motion on the surface. This combination can be used to expand any of the applications demonstrated above to support tap localization over larger surface-wide AR interfaces. We further implement a mouse experience (Figure 12(e)) where the user simply moves his or her wrist to control the pointer position and taps on left/right region with respect to the wristband to invoke left/right click.

DISCUSSION AND LIMITATIONS

In this section, we discuss the insights gained from this work, and discuss limitations and future work.

Active Acoustic Approach. As discussed in Related Work, active acoustic approach also shows potential to localize a finger [35, 54]. Based on the existing work, we also initially investigated the idea of embedding both transducer (e.g., speaker, actuator) and receiver (e.g., microphone, accelerometer) on the wristband, and tried to localize the finger using reflected signal. However, as also mentioned in FingerIO [35], we found out that it was hard to completely isolate the transducer and receiver within a small wristband form factor. The received signal would be largely masked by the emitted signal from the transducer, making it hard to separate and extract the reflected signal. One possible way to solve the masking problem is to increase the frequency of the emitted signal. To test this, we built a prototype using an ultrasound emitter. However, this presents another problem – signal attenuation. High frequency signals attenuate very quickly and therefore the received signals do not have enough SNR to make any reasonable inferences. The set-up therefore needed more power and coupling liquids which make the approach untenable for practical use.

Tap Detection and Localization in Real-World Settings. For the tap detection, we did not evaluate the system when the hand is not on a surface, but vibration and sound signals are still detected (e.g., user's hand hits an object or user is walking). This is because we assume the system is aware of the user's hand being placed on a surface, which could be easily achieved using a proximity sensor. However, our threshold-based method still needs to be tested more intensively in the field. Another research direction would be exploring the effect of environmental noises on tap localization. While we envision *Acustico* to be most useful in indoor scenarios where the noises are limited, we are interested to quantify the lowest SNR for the system to work. Furthermore, it is also interesting to look into how surface properties (e.g., thickness/flatness) would affect system performance. We leave these for future work.

Ecological Validity. Acustico focuses on tap detection and localization under the situation that the user's wrist is being placed on a surface. Although we did not receive any negative feedback from participants, we understand that this requirement might introduce fatigue and be uncomfortable after long time use. Future research should be conducted in ecological validity of this approach.

Re-calibration/Retrain for New Surfaces. Since surfaces made of different materials have different properties and wave propagation velocities, our system needs to be re-calibrated and retrained for any new surfaces that it has not seen before. However, we show in the evaluation that we can make the training set as small as 12 taps to achieve a reasonable localization accuracy (X: 9.64mm error, Y: 5.79mm error). Another possible solution is to train on the surfaces around the user beforehand and load the specific model when the user needs to interact on that surface.

Practicality. The system we presented is an early-stage proof-of-concept research prototype. Although it is in a constrained wristband form factor similar to current wrist wearables in the market, more work is still needed to incorporate these sensors into a flexible and stretchable wristband. One direction to pursue in this regard is to only place the accelerometers under the wristband in direct contact with the surface and place the microphones on the top (in the watch-face) since they rely on in-air propagation. This may introduce certain inconsistencies in the sensor distances which could impact accuracies. Moreover, to capture the small TDOAs, we sampled the data in 1MHz, which is not supported in most of current wrist wearable devices due to their limited computational resources and batteries. But we believe this is not impossible as technology advances (e.g., ADS8330 from TI can sample at 1MHz and only consume 21mW). Plenty of engineering efforts would be necessary to fully embed this technique into commercial wrist wearable devices.

System Evaluation. We evaluated the system using a region-based approach since *Acustico* mainly focuses on discrete tap detection and localization. We plan to investigate how to use this technique to facilitate continuous position tracking or gesture-based sensing in our future work. To ensure the studies can be completed in 90 minutes, participants were asked to perform taps sequentially in each session, which might reduce the wrist movement between taps. Evaluation in random tap locations might need to be included in the future.

Finger-up Detection. The mouse demo enables clicking on targets. However, *Acustico* only detects the finger-down event which produces the acoustic waves but not the finger-up event since it does not produce any acoustic waves. The detection of this release event will enable a dragging state in the mouse [5] and is an interesting challenge for wrist-based sensing.

CONCLUSION

This paper presents a passive acoustic sensing approach for wrist-worn devices to detect and localize tap. We discuss the sensing principle and our investigation on different sensor configurations. We built a wristband prototype with four acoustic sensors including two accelerometers and two microphones. Through a 20-participant study, we demonstrate that our system can reliably detect taps with an F1-score of 0.9987 across different environmental noises and yield high localization accuracies with root-mean-square-errors of 7.6mm (X-axis) and 4.6mm (Y-axis) across different surfaces and tapping techniques. Our work presents a novel sensing methodology for always-available input on any unmodified surface. We believe it holds the potential to further enrich the input expressiveness of today's computing devices (e.g., wearables and AR devices).

REFERENCES

1. Hideki Koike, Yoichi Sato, Yoshinori Kobayashi. 2001. Integrating paper and digital information on EnhancedDesk: a method for realtime finger tracking on an augmented desk system. *ACM Trans. Comput.-Hum. Interact.*, 8 (4). 307–322. DOI=https://doi.org/10.1145/504704.504706
2. Ankur Agarwal, Shahram Izadi, Manmohan Chandraker and Andrew Blake. 2007. High Precision Multi-touch Sensing on Surfaces using Overhead Cameras. In *Second Annual IEEE International Workshop on Horizontal Interactive Human-Computer Systems* (TABLETOP'07), 197-200. DOI=https://doi.org/10.1109/TABLETOP.2007.29
3. Michael Boyle and Saul Greenberg. 2005. The language of privacy: Learning from video media space analysis and design. *ACM Trans. Comput.-Hum. Interact.*, 12 (2). 328-370. DOI=https://doi.org/10.1145/1067860.1067868
4. Alex Butler, Shahram Izadi and Steve Hodges. 2008. SideSight: multi-"touch" interaction around small devices. In *Proceedings of the 21st annual ACM symposium on User interface software and technology* (UIST '08). 201-204. DOI=https://doi.org/10.1145/1449715.1449746
5. William Buxton. 1990. A three-state model of graphical input. In *Proceedings of the IFIP TC13 Third Interational Conference on Human-Computer Interaction* (INTERACT '90), 449–456.
6. Liwei Chan, Yi-Ling Chen, Chi-Hao Hsieh, Rong-Hao Liang and Bing-Yu Chen. 2015. CyclopsRing: Enabling Whole-Hand and Context-Aware Interactions Through a Fisheye Ring. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software and Technology* (UIST'15), 549-556. DOI= https://doi.org/10.1145/2807442.2807450
7. Jae Sik Chang, Eun Yi Kim, KeeChul Jung and Hang Joon Kim. 2005. Real time hand tracking based on active contour model. In *Proceedings of the 2005 international conference on Computational Science and Its Applications.* 999. DOI=https://doi.org/10.1007/11424925_104
8. Ke-Yu Chen, Shwetak N. Patel and Sean Keller. 2016. Finexus: Tracking Precise Motions of Multiple Fingertips Using Magnetic Sensing. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (CHI'16), 1504–1514. DOI=http://dx.doi.org/10.1145/2858036.2858125
9. 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), 161-166. DOI=https://doi.org/10.1145/2642918.2647396
10. Fabrice Devige and Jean-Pierre Nikolovski. 2003. Accurate Interactive Acoustic Plate. *US Patent Application No. US2003/0066692 A1*.
11. 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), 311-314. DOI=https://doi.org/10.1145/2030112.2030154
12. Mayank Goel, Brendan Lee, Md. Tanvir Islam Aumi, Shwetak Patel, Gaetano Borriello, Stacie Hibino and Bo Begole. 2014. SurfaceLink: using inertial and acoustic sensing to enable multi-device interaction on a surface. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI'14), 1387–1396. DOI=https://doi.org/10.1145/2556288.2557120
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 ACM Symposium on User Interface Software and Technology* (UIST'16). DOI=http://dx.doi.org/10.1145/2984511.2984563
14. Jun Gong, Yang Zhang, Xia Zhou and Xing-Dong Yang. 2017. Pyro: Thumb-Tip Gesture Recognition Using Pyroelectric Infrared Sensing. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology* (UIST'17), 553-563. DOI=https://doi.org/10.1145/3126594.3126615
15. Yizheng Gu, Chun Yu, Zhipeng Li, Weiqi Li, Shuchang Xu, Xiaoying Wei and Yuanchun Shi. 2019. Accurate and Low-Latency Sensing of Touch Contact on Any Surface with Finger-Worn IMU Sensor. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology* (UIST'19), 1059–1070. DOI=https://doi.org/10.1145/3332165.3347947
16. Jefferson Y. Han. 2005. Low-cost multi-touch sensing through frustrated total internal reflection. In *Proceedings of the 18th annual ACM symposium on User interface software and technology* (UIST'05), 115–118. DOI=https://doi.org/10.1145/1095034.1095054
17. 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), 579–591. DOI=https://doi.org/10.1145/3126594.3126612
18. Chris Harrison. 2010. Appropriated Interaction Surfaces. *Computer*, 43 (6). 86-89. DOI=https://doi.org/110.1109/MC.2010.158

19. Chris Harrison, Hrvoje Benko, and Andrew D. Wilson. 2011. OmniTouch: wearable multitouch interaction everywhere. In *Proceedings of the 24th annual ACM symposium on User interface software and technology* (UIST '11), 441-450. DOI=https://doi.org/10.1145/2047196.2047255
20. Chris Harrison and Scott E. Hudson. 2008. Scratch input: creating large, inexpensive, unpowered and mobile finger input surfaces. In *Proceedings of the 21th annual ACM symposium on User interface software and technology* (UIST '08), 205-208. DOI=https://doi.org/10.1145/1449715.1449747
21. Chris Harrison, Desney S. 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), 453-462. DOI=https://doi.org/10.1145/1753326.1753394
22. Scott E. Hudson and Ian Smith. 1996. Techniques for addressing fundamental privacy and disruption tradeoffs in awareness support systems. In *Proceedings of the 1996 ACM conference on Computer supported cooperative work* (CSCW'96), 248–257. DOI=https://doi.org/10.1145/240080.240295
23. Yasha Iravantchi, Yang Zhang, Evi Bernitsas, Mayank Goel and Chris Harrison. 2019. Interferi: Gesture Sensing using On-Body Acoustic Interferometry. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (CHI'19), 276. DOI=https://doi.org/10.1145/3290605.3300506
24. Hiroshi Ishii, Craig Wisneski, Julian Orbanes, Ben Chun and Joe Paradiso. 1999. PingPongPlus: design of an athletic-tangible interface for computer-supported cooperative play. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems* (CHI'99), 394–401. DOI=https://doi.org/10.1145/302979.303115
25. Shaun K. Kane, Daniel Avrahami, Jacob O. Wobbrock, Beverly Harrison, Adam D. Rea, Matthai Philipose and Anthony LaMarca. 2009. Bonfire: a nomadic system for hybrid laptop-tabletop interaction. In *Proceedings of the 22nd annual ACM symposium on User interface software and technology* (UIST'09), 129–138. DOI=https://doi.org/10.1145/1622176.1622202
26. Wolf Kienzle and Ken Hinckley. 2014. LightRing: always-available 2D input on any surface. In *Proceedings of the 27th annual ACM symposium on User interface software and technology* (UIST'14), 157-160. DOI=https://doi.org/10.1145/2642918.2647376
27. A. H. F. Lam, W. J. Li, Liu Yunhui and Xi Ning. 2002. MIDS: micro input devices system using MEMS sensors. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 1184-1189 vol.1182. DOI=https://doi.org/10.1109/IRDS.2002.1043893
28. 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), 321-333. DOI=https://doi.org/10.1145/2984511.2984582
29. SK Lee, William Buxton and K. C. Smith. 1985. A multi-touch three dimensional touch-sensitive tablet. *SIGCHI Bull.*, 16 (4). 21–25. DOI=https://doi.org/10.1145/1165385.317461
30. Julien Letessier and François Bérard. 2004. Visual tracking of bare fingers for interactive surfaces. In *Proceedings of the 17th annual ACM symposium on User interface software and technology* (UIST'04), 119–122. DOI=https://doi.org/10.1145/1029632.1029652
31. Jaime Lien, Nicholas Gillian, M. Emre Karagozler, Patrick Amihood, Carsten Schwesig, Erik Olson, Hakim Raja and Ivan Poupyrev. 2016. Soli: Ubiquitous Gesture Sensing with Millimeter Wave Radar. In *ACM Trans. Graph* (SIGGRAPH'16), 10. DOI=https://doi.org/10.1145/2897824.2925953
32. Nobuyuki Matsushita and Jun Rekimoto. 1997. HoloWall: designing a finger, hand, body, and object sensitive wall. In *Proceedings of the 10th annual ACM symposium on User interface software and technology* (UIST'97), 209–210. DOI=https://doi.org/10.1145/263407.263549
33. 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), 593-597. DOI=https://doi.org/10.1145/3126594.3126604
34. Dan Morris, T. Scott Saponas and Desney Tan. 2011. Emerging Input Technologies for Always-Available Mobile Interaction. *Found. Trends Hum.-Comput. Interact.*, 4 (4). 245–316. DOI=https://doi.org/10.1561/1100000023
35. Rajalakshmi Nandakumar, Vikram Iyer, Desney Tan and Shyamnath Gollakota. 2016. FingerIO: Using Active Sonar for Fine-Grained Finger Tracking. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (CHI'16), 1515-1525. DOI=https://doi.org/10.1145/2858036.2858580
36. Shijia Pan, Ceferino G. Ramirez, Mostafa Mirshekari, Jonathon Fagert, Albert J. Chung, Chih C. Hu, John P. Shen, Hae Y. Noh and Pei Zhang. 2017. SurfaceVibe: Vibration-Based Tap and Swipe Tracking on Ubiquitous Surfaces. In *16th ACM/IEEE International Conference on*

Information Processing in Sensor Networks (IPSN'17), 197-208.
37. J. A. Paradiso, Leo Che King, N. Checka and Hsiao Kaijen. 2002. Passive acoustic sensing for tracking knocks atop large interactive displays. In *SENSORS, 2002 IEEE*, 521-527 vol.521. DOI=https://doi.org/10.1109/ICSENS.2002.1037150
38. Jun Rekimoto. 2001. GestureWrist and GesturePad: Unobtrusive Wearable Interaction Devices. In *Proceedings of the 5th IEEE International Symposium on Wearable Computers* (ISWC'01), 21. DOI=https://doi.org/10.1109/ISWC.2001.962092
39. Elliot N. Saba, Eric C. Larson and Shwetak N. Patel. 2012. Dante vision: In-air and touch gesture sensing for natural surface interaction with combined depth and thermal cameras. In *2012 IEEE International Conference on Emerging Signal Processing Applications*, 167-170. DOI=https://doi.org/10.1109/ESPA.2012.6152472
40. 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), 167-176. DOI=https://doi.org/10.1145/1622176.1622208
41. Katie A. Siek, Yvonne Rogers and Kay H. Connelly. 2005. Fat finger worries: how older and younger users physically interact with PDAs. In *Proceedings of the 2005 IFIP TC13 international conference on Human-Computer Interaction*, 267-280. DOI=https://doi.org/10.1007/11555261_24
42. Hoang Truong, Shuo Zhang, Ufuk Muncuk, Phuc Nguyen, Nam Bui, Anh Nguyen, Qin Lv, Kaushik Chowdhury, Thang Dinh and Tam Vu. 2018. CapBand: Battery-free Successive Capacitance Sensing Wristband for Hand Gesture Recognition. In *Proceedings of the 16th ACM Conference on Embedded Networked Sensor Systems* (SenSys'18), 54–67. DOI=https://doi.org/10.1145/3274783.3274854
43. Dong Wei, Steven Z. Zhou and Du Xie. 2010. MTMR: A conceptual interior design framework integrating Mixed Reality with the Multi-Touch tabletop interface. In *2010 IEEE International Symposium on Mixed and Augmented Reality*, 279-280. DOI=https://doi.org/10.1109/ISMAR.2010.5643606
44. Hongyi Wen, Julian Ramos Rojas and Anind K. Dey. 2016. Serendipity: Finger Gesture Recognition using an Off-the-Shelf Smartwatch. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (CHI'16), 3847–3851. DOI=https://doi.org/10.1145/2858036.2858466
45. Andrew D. Wilson. 2004. TouchLight: an imaging touch screen and display for gesture-based interaction. In *Proceedings of the 6th international conference on Multimodal interfaces* (ICMI'04), 69–76. DOI=https://doi.org/10.1145/1027933.1027946
46. George S. K. Wong. 1986. Speed of sound in standard air. *The Journal of the Acoustical Society of America*, 79 (5). 1359-1366. DOI=https://doi.org/10.1121/1.393664
47. Robert Xiao, Julia Schwarz, Nick Throm, Andrew D. Wilson and Hrvoje Benko. 2018. MRTouch: Adding Touch Input to Head-Mounted Mixed Reality. *IEEE Transactions on Visualization and Computer Graphics*, 24 (4). 1653-1660. DOI=https://doi.org/10.1109/TVCG.2018.2794222
48. Robert Xiao, Greg Lew, James Marsanico, Divya Hariharan, Scott Hudson and Chris Harrison. 2014. Toffee: enabling ad hoc, around-device interaction with acoustic time-of-arrival correlation. In *Proceedings of the 16th international conference on Human-computer interaction with mobile devices and services* (MobileHCI'14), 67-76. DOI=https://doi.org/10.1145/2628363.2628383
49. Ming Yang. 2011. In-Solid Acoustic Source Localization Using Likelihood Mapping Algorithm. *Open Journal of Acoustics*, 1. 34-40. DOI=https://doi.org/10.4236/oja.2011.12005
50. Xing-Dong Yang, Tovi Grossman, Daniel Wigdor and George Fitzmaurice. 2012. Magic Finger: Always-Available Input through Finger Instrumentation. In *Proceedings of the 25nd annual ACM symposium on User interface software and technology* (UIST'12), 147 - 156. DOI=https://doi.org/10.1145/2380116.2380137
51. Sang Ho Yoon, Ke Huo, Yunbo Zhang, Guiming Chen, Luis Paredes, Subramanian Chidambaram and Karthik Ramani. 2017. iSoft: A Customizable Soft Sensor with Real-time Continuous Contact and Stretching Sensing. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology* (UIST'17), 665-678. DOI=https://doi.org/10.1145/3126594.3126654
52. Cheng Zhang, AbdelKareem Bedri, Gabriel Reyes, Bailey Bercik, Omer T. Inan, Thad E. Starner and Gregory D. Abowd. 2016. TapSkin: Recognizing On-Skin Input for Smartwatches. In *Proceedings of the 2016 ACM International Conference on Interactive Surfaces and Spaces*, 13–22. DOI=https://doi.org/10.1145/2992154.2992187
53. Cheng Zhang, Anandghan Waghmare, Pranav Kundra, Yiming Pu, Scott Gilliland, Thomas Ploetz, Thad E. Starner, Omer T. Inan and Gregory D. Abowd. 2017. FingerSound: Recognizing unistroke thumb gestures using a ring. In *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 1 (3). Article DOI=https://doi.org/120. 10.1145/3130985

54. Cheng Zhang, Qiuyue Xue, Anandghan Waghmare, Sumeet Jain, Yiming Pu, Sinan Hersek, Kent Lyons, Kenneth A. Cunefare, Omer T. Inan and Gregory D. Abowd. 2017. SoundTrak: Continuous 3D Tracking of a Finger Using Active Acoustics. In *Proc. ACM Interact. Mob. Wearable Ubiquitous* DOI=https://doi.org/10.1145/3090095
55. Yang Zhang and Chris Harrison. 2018. Pulp Nonfiction: Low-Cost Touch Tracking for Paper. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (CHI'18), 1-11. DOI=https://doi.org/10.1145/3173574.3173691
56. 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 and Technology* (UIST'15), 167-173. DOI=https://doi.org/10.1145/2807442.2807480
57. Yang Zhang, Wolf Kienzle, Yanjun Ma, Shiu S. Ng, Hrvoje Benko and Chris Harrison. 2019. ActiTouch: Robust Touch Detection for On-Skin AR/VR Interfaces. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology* (UIST'19), 1151–1159. DOI=https://doi.org/10.1145/3332165.3347869
58. Yang Zhang, Gierad Laput and Chris Harrison. 2017. Electrick: Low-Cost Touch Sensing Using Electric Field Tomography. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (CHI'17), 1-14. DOI=https://doi.org/10.1145/3025453.3025842
59. Yang Zhang, Chouchang Yang, Scott E. Hudson, Chris Harrison and Alanson Sample. 2018. Wall++: Room-Scale Interactive and Context-Aware Sensing. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 273. DOI=https://doi.org/10.1145/3173574.3173847
60. Yang Zhang, Junhan Zhou, Gierad Laput and Chris Harrison. 2016. SkinTrack: Using the Body as an Electrical Waveguide for Continuous Finger Tracking on the Skin. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (CHI'16), 1491-1503. DOI=https://doi.org/10.1145/2858036.2858082