Skinput: Appropriating the Body as an Input Surface

Dhanush Shetty[1], Akash Kumar[2], Ainab[3], Megha Hegde[4]

[1,2,3]Student, Department of Computer Science Engineering, Alva's Inst. of Engg. and Tech., Moodbidri, India

[4]Assistant Professor, Dept. of Computer Science Engineering, Alva's Inst. of Engg. and Tech., Moodbidri, India

***Abstract*: We present Skinput, a technology that appropriates the human body for acoustic transmission, allowing the skin to be used as an input surface. In particular, we resolve the location of finger taps on the arm and hand by analyzing mechanical vibrations that propagate through the body. We collect these signals using a novel array of sensors worn as an armband. This approach provides an always available, naturally portable, and on-body finger input system. We assess the capabilities, accuracy and limitations of our technique through a two-part, twenty-participant user study. To further illustrate the utility of our approach, we conclude with several proof-of-concept applications we developed.**

***Keywords*: Bio-acoustics, finger input, buttons, gestures, on-body interaction, projected displays, audio interfaces.**

1. Introduction

Devices with significant computational power and capabilities can now be easily carried on our bodies. However, their small size typically leads to limited interaction space (e.g., diminutive screens, buttons, and jog wheels) and consequently diminishes their usability and functionality. Since we cannot simply make buttons and screens larger without losing the primary benefit of small size, we consider alternative approaches that enhance interactions with small mobile systems.

One option is to opportunistically appropriate surface area from the environment for interactive purposes. For example, [10] describes a technique that allows a small mobile device to turn tables on which it rests into a gestural finger input canvas. However, tables are not always present, and in a mobile context, users are unlikely to want to carry appropriated surfaces with them (at this point, one might as well just have a larger device). However, there is one surface that has been previous overlooked as an input canvas, and one that happens to always travel with us: our skin.

Appropriating the human body as an input device is appealing not only because we have roughly two square meters of external surface area, but also because much of it is easily accessible by our hands (e.g., arms, upper legs, torso). Furthermore, proprioception – our sense of how our body is configured in three-dimensional space – allows us to accurately interact with our bodies in an eyes-free manner. For example, we can readily flick each of our fingers, touch the tip of our nose, and clap our hands together without visual assistance. Few external input devices can claim this accurate, eyes-free input characteristic and provide such a large interaction area. In this paper, we present our work on Skinput – a method that allows the body to be appropriated for finger input using a novel, non-invasive, wearable bio-acoustic sensor.

The contributions of this paper are:

1. We describe the design of a novel, wearable sensor for bio-acoustic signal acquisition (Figure 1).
2. We describe an analysis approach that enables our system to resolve the location of finger taps on the body.

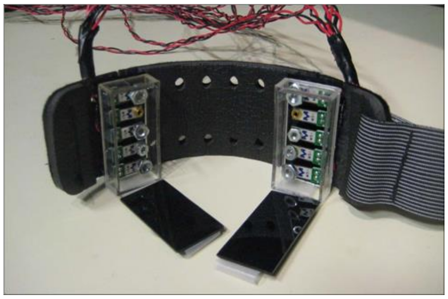

Fig. 1. A wearable, bio-acoustic sensing array built into an armband. Sensing elements detect vibrations transmitted through the body. The two sensor packages shown above each contain five, specially weighted, cantilevered piezo films, responsive to a particular frequency range.

1. We assess the robustness and limitations of this system through a user study.
2. We explore the broader space of bio-acoustic input through prototype applications and additional experimentation.

2. Related work Always-Available Input

The primary goal of Skinput is to provide an alwaysavailable mobile input system – that is, an input system that does not require a user to carry or pick up a device. A number of alternative approaches have been proposed that operate in this space. Techniques based on computer vision are popular (e.g. [3,26,27], see [7] for a recent survey). These, however, are computationally expensive and error prone in mobile scenarios (where, e.g., non-input optical flow is prevalent). Speech input (e.g. [13,15]) is a logical choice for always-available input, but is limited in its precision in unpredictable acoustic environments, and suffers from privacy and scalability issues in shared environments.

Other approaches have taken the form of wearable computing. This typically involves a physical input device built in a form considered to be part of one's clothing. For example, glove-based input systems (see [25] for a review) allow users to retain most of their natural hand movements, but are cumbersome, uncomfortable, and disruptive to tactile sensation. Post and Orth [22] present a "smart fabric" system that embeds sensors and conductors into fabric, but taking this approach to always-available input necessitates embedding technology in all clothing, which would be prohibitively complex and expensive.

The SixthSense project [19] proposes a mobile, always available input/output capability by combining projected information with a color-marker-based vision tracking system. This approach is feasible, but suffers from serious occlusion and accuracy limitations. For example, determining whether, e.g., a finger has tapped a button, or is merely hovering above it, is extraordinarily difficult. In the present work, we briefly explore the combination of on-body sensing with on-body projection.

A. *Bio-sensing*

Skinput leverages the natural acoustic conduction properties of the human body to provide an input system, and is thus related to previous work in the use of biological signals for computer input. Signals traditionally used for diagnostic medicine, such as heart rate and skin resistance, have been appropriated for assessing a user's emotional state (e.g. [16,17,20]). These features are generally subconsciously driven and cannot be controlled with sufficient precision for direct input. Similarly, brain sensing technologies such as electroencephalography (EEG) and functional near-infrared spectroscopy (fNIR) have been used by HCI researchers to assess cognitive and emotional state (e.g. [9,11,14]); this work also primarily looked at involuntary signals. In contrast, brain signals have been harnessed as a direct input for use by paralyzed patients (e.g. [8,18]), but direct brain computer interfaces (BCIs) still lack the bandwidth required for everyday computing tasks, and require levels of focus, training, and concentration that are incompatible with typical computer interaction.

There has been less work relating to the intersection of finger input and biological signals. Researchers have harnessed the electrical signals generated by muscle activation during normal hand movement through electromyography (EMG) (e.g. [23,24]). At present, however, this approach typically requires expensive amplification systems and the application of conductive gel for effective signal acquisition, which would limit the acceptability of this approach for most users.

The input technology most related to our own is that of Amento et al. [2], who placed contact microphones on a user's wrist to assess finger movement. However, this work was never formally evaluated, as is constrained to finger motions in one hand. The Hambone system [6] employs a similar setup, and through an HMM, yields classification accuracies around 90% for four gestures (e.g., raise heels, snap fingers). Performance of false positive rejection remains untested in both systems at present. Moreover, both techniques required the placement of sensors near the area of interaction (e.g., the wrist), increasing the degree of invasiveness and visibility.

Finally, bone conduction microphones and headphones – now common consumer technologies - represent an additional bio-sensing technology that is relevant to the present work. These leverage the fact that sound frequencies relevant to human speech propagate well through bone. Bone conduction microphones are typically worn near the ear, where they can sense vibrations propagating from the mouth and larynx during speech. Bone conduction headphones send sound through the bones of the skull and jaw directly to the inner ear, bypassing transmission of sound through the air and outer ear, leaving an unobstructed path for environmental sounds.

B. *Acoustic Input*

Our approach is also inspired by systems that leverage acoustic transmission through (non-body) input surfaces. Paradiso et al. [21] measured the arrival time of a sound at multiple sensors to locate hand taps on a glass window. Ishii et al. [12] use a similar approach to localize a ball hitting a table, for computer augmentation of a real-world game. Both of these systems use acoustic time-of-flight for localization, which we explored, but found to be insufficiently robust on the human body, leading to the fingerprinting approach described in this paper.

C. *Skinput*

To expand the range of sensing modalities for always available input systems, we introduce Skinput, a novel input technique that allows the skin to be used as a finger input surface. In our prototype system, we choose to focus on the arm (although the technique could be applied elsewhere). This is an attractive area to appropriate as it provides considerable surface area for interaction, including a contiguous and flat area for projection (discussed subsequently). Furthermore, the forearm and hands contain a complex assemblage of bones that increases acoustic distinctiveness of different locations. To capture this acoustic information, we developed a wearable armband that is non-invasive and easily removable (Figures 1 and 5).

In this section, we discuss the mechanical phenomena that enables Skinput, with a specific focus on the mechanical properties of the arm. Then we will describe the Skinput sensor and the processing techniques we use to segment, analyze, and classify bio-acoustic signals.

D. *Bio-acoustics*

When a finger taps the skin, several distinct forms of acoustic energy are produced. Some energy is radiated into the air as sound waves; this energy is not captured by the Skinput system. Among the acoustic energy transmitted through the arm, the most readily visible are transverse waves, created by the displacement of the skin from a finger impact (Figure 2). When shot with a high-speed camera, these appear as ripples, which propagate outward from the point of contact (see video). The amplitude of these ripples is correlated to both the tapping force and to the volume and compliance of soft tissues under the impact area. In general, tapping on soft regions of the arm creates higher amplitude transverse waves than tapping on boney areas (e.g., wrist, palm, fingers), which have negligible compliance.

In addition to the energy that propagates on the surface of the arm, some energy is transmitted inward, toward the skeleton (Figure 3). These longitudinal (compressive) waves travel through the soft tissues of the arm, exciting the bone, which is much less deformable then the soft tissue but can respond to mechanical excitation by rotating and translating as a rigid body. This excitation vibrates soft tissues surrounding the entire length of the bone, resulting in new longitudinal waves that propagate outward to the skin.

We highlight these two separate forms of conduction – transverse waves moving directly along the arm surface, and longitudinal waves moving into and out of the bone through soft tissues – because these mechanisms carry energy at different frequencies and over different distances. Roughly speaking, higher frequencies propagate more readily through bone than through soft tissue, and bone conduction carries energy over larger distances than soft tissue conduction. While we do not explicitly model the specific mechanisms of conduction, or depend on these mechanisms for our analysis, we do believe the success of our technique depends on the complex acoustic patterns that result from mixtures of these modalities.

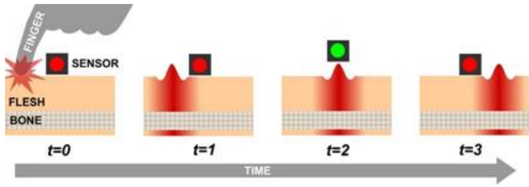

Fig. 2. Transverse wave propagation: Finger impacts displace the skin, creating transverse waves (ripples). The sensor is activated as the wave passes underneath it.

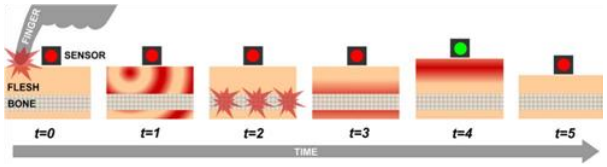

Fig. 3. Longitudinal wave propagation: Finger impacts create longitudinal (compressive) waves that cause internal skeletal structures to vibrate. This, in turn, creates longitudinal waves that emanate outwards from the bone (along its entire length) toward the skin.

Similarly, we also believe that joints play an important role in making tapped locations acoustically distinct. Bones are held together by ligaments, and joints often include additional biological structures such as fluid cavities. This makes joints behave as acoustic filters. In some cases, these may simply dampen acoustics; in other cases, these will selectively attenuate specific frequencies, creating location specific acoustic signatures.

E. *Sensing*

To capture the rich variety of acoustic information described in the previous section, we evaluated many sensing technologies, including bone conduction microphones, conventional microphones coupled with stethoscopes [10], piezo contact microphones [2], and accelerometers. However, these transducers were engineered for very different applications than measuring acoustics transmitted through the human body. As such, we found them to be lacking in several significant ways. Foremost, most mechanical sensors are engineered to provide relatively flat response curves over the range of frequencies that is relevant to our signal. This is a desirable property for most applications where a faithful representation of an input signal – uncolored by the properties of the transducer – is desired. However, because only a specific set of frequencies is conducted through the arm in response to tap input, a flat response curve leads to the capture of irrelevant frequencies and thus to a high signal-to-noise ratio.

While bone conduction microphones might seem a suitable choice for Skinput, these devices are typically engineered for capturing human voice, and filter out energy below the range of human speech (whose lowest frequency is around 85Hz). Thus most sensors in this category were not especially sensitive to lower-frequency signals (e.g., 25Hz), which we found in our empirical pilot studies to be vital in characterizing finger taps.

To overcome these challenges, we moved away from a single sensing element with a flat response curve, to an array of highly tuned vibration sensors. Specifically, we employ small, cantilevered piezo films (MiniSense100, Measurement Specialties, Inc.). By adding small weights to the end of the cantilever, we are able to alter the resonant frequency, allowing the sensing element to be responsive to a unique, narrow, low-frequency band of the acoustic spectrum.

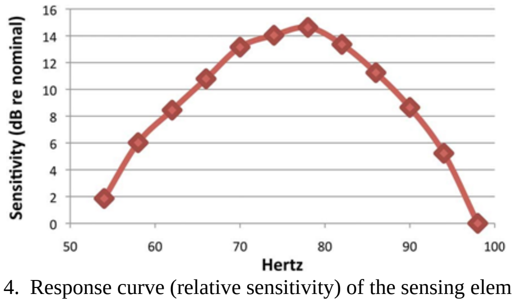

Fig. 4. Response curve (relative sensitivity) of the sensing element that resonates at 78 Hz.

Adding more mass lowers the range of excitation to which a sensor responds; we weighted each element such that it aligned with particular frequencies that pilot studies showed to be useful in characterizing bio-acoustic input. Figure 4 shows the response curve for one of our sensors, tuned to a resonant frequency of 78Hz. The curve shows a ~14dB drop-off ±20Hz away from the resonant frequency. Additionally, the cantilevered sensors were naturally insensitive to forces parallel to the skin (e.g., shearing motions caused by stretching). Thus, the skin stretch induced by many routine movements (e.g., reaching for a doorknob) tends to be attenuated. However, the sensors are highly responsive to motion perpendicular to the skin plane – perfect for capturing transverse surface waves (Figure 2) and longitudinal waves emanating from interior structures (Figure 3).

Finally, our sensor design is relatively inexpensive and can be manufactured in a very small form factor (e.g., MEMS), rendering it suitable for inclusion in future mobile devices (e.g., an arm-mounted audio player).

F. Armband Prototype

Our final prototype, shown in Figures 1 and 5, features two arrays of five sensing elements, incorporated into an armband form factor. The decision to have two sensor packages was motivated by our focus on the arm for input. In particular, when placed on the upper arm (above the elbow), we hoped to collect acoustic information from the fleshy bicep area in addition to the firmer area on the underside of the arm, with better acoustic coupling to the Humerus, the main bone that runs from shoulder to elbow. When the sensor was placed below the elbow, on the forearm, one package was located near the Radius, the bone that runs from the lateral side of the elbow to the thumb side of the wrist, and the other near the Ulna, which runs parallel to this on the medial side of the arm closest to the body. Each location thus provided slightly different acoustic coverage and information, helpful in disambiguating input location.

Based on pilot data collection, we selected a different set of resonant frequencies for each sensor package (Table 1). We tuned the upper sensor package to be more sensitive to lower frequency signals, as these were more prevalent in fleshier areas. Conversely, we tuned the lower sensor array to be sensitive to higher frequencies, in order to better capture signals transmitted though (denser) bones.

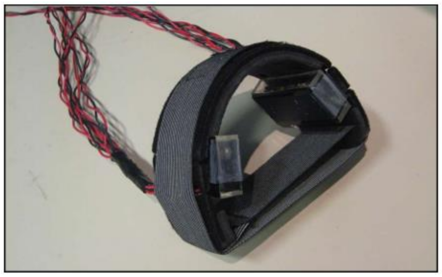

Fig. 5. Prototype armband

G. Processing

In our prototype system, we employ a Mackie Onyx 1200F audio interface to digitally capture data from the ten sensors (http://mackie.com). This was connected via Firewire to a conventional desktop computer, where a thin client written in C interfaced with the device using the Audio Stream Input/Output (ASIO) protocol.

Each channel was sampled at 5.5kHz, a sampling rate that would be considered too low for speech or environmental audio, but was able to represent the relevant spectrum of frequencies transmitted through the arm. This reduced sample rate (and consequently low processing bandwidth) makes our technique readily portable to embedded processors. For example, the ATmega168 processor employed by the Arduino platform can sample analog readings at 77kHz with no loss of precision, and could therefore provide the full sampling power required for Skinput (55kHz total).

Data was then sent from our thin client over a local socket to our primary application, written in Java. This program performed three key functions. First, it provided a live visualization of the data from our ten sensors, which was useful in identifying acoustic features (Figure 6). Second, it segmented inputs from the data stream into independent instances (taps). Third, it classified these input instances.

The audio stream was segmented into individual taps using an absolute exponential average of all ten channels (Figure 6, red waveform). When an intensity threshold was exceeded (Figure 6, upper blue line), the program recorded the timestamp as a potential start of a tap. If the intensity did not fall below a second, independent "closing" threshold (Figure 6, lower purple line) between 100ms and 700ms after the onset crossing (a duration we found to be the common for finger impacts), the event was discarded. If start and end crossings were detected that satisfied these criteria, the acoustic data in that period (plus a 60ms buffer on either end) was considered an input event (Figure 6, vertical green regions). Although simple, this heuristic proved to be highly robust, mainly due to the extreme noise suppression provided by our sensing approach.

Table 1
Resonant frequencies of individual elements in the two sensor packages

Upper Array	25 Hz	27 Hz	30 Hz	38 Hz	78 Hz
Lower Array	25 Hz	27 Hz	40 Hz	44 Hz	64 Hz

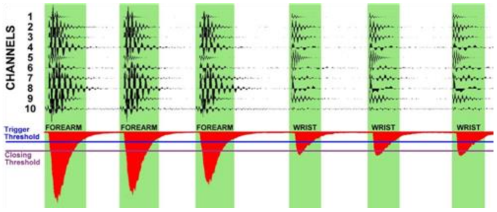

Fig. 6. Ten channels of acoustic data generated by three finger taps on the forearm, followed by three taps on the wrist. The exponential average of the channels is shown in red. Segmented input windows are highlighted in green. Note how different sensing elements are actuated by the two locations.

After an input has been segmented, the waveforms are analyzed. The highly discrete nature of taps (i.e. point impacts) meant acoustic signals were not particularly expressive over time (unlike gestures, e.g., clenching of the hand). Signals simply diminished in intensity overtime. Thus, features are computed over the entire input window and do not capture any temporal dynamics.

We employ a brute force machine learning approach, computing 186 features in total, many of which are derived combinatorially. For gross information, we include the average amplitude, standard deviation and total (absolute) energy of the waveforms in each channel (30 features). From these, we calculate all average amplitude ratios between channel pairs (45 features). We also include an average of these ratios (1 feature). We calculate a 256-point FFT for all ten channels, although only the lower ten values are used (representing the acoustic power from 0Hz to 193Hz), yielding 100 features. These are normalized by the highest-amplitude FFT value found on any channel. We also include the center of mass of the power spectrum within the same 0Hz to 193Hz range for each channel, a rough estimation of the fundamental frequency of the signal displacing each sensor (10 features). Subsequent feature selection established the all-pairs amplitude ratios and certain bands of the FFT to be the most predictive features.

These 186 features are passed to a Support Vector Machine (SVM) classifier. A full description of SVMs is beyond the scope of this paper (see [4] for a tutorial). Our software uses the implementation provided in the Weka machine learning toolkit [28]. It should be noted, however, that other, more sophisticated classification techniques and features could be employed. Thus, the results presented in this paper should be considered a baseline.

Before the SVM can classify input instances, it must first be trained to the user and the sensor position. This stage requires the collection of several examples for each input location of interest. When using Skinput to recognize live input, the same 186 acoustic features are computed on-thefly for each segmented input. These are fed into the trained SVM for classification. We use an event model in our software – once an input is classified, an event associated with that location is instantiated. Any interactive features bound to that event are fired. As can be seen in our video, we readily achieve interactive speeds.

H. *Experiment participants*

To evaluate the performance of our system, we recruited 13 participants (7 female) from the Greater Seattle area. These participants represented a diverse cross-section of potential ages and body types. Ages ranged from 20 to 56 (mean 38.3), and computed body mass indexes (BMIs) ranged from 20.5 (normal) to 31.9 (obese).

I. *Experimental conditions*

We selected three input groupings from the multitude of possible location combinations to test. We believe that these groupings, illustrated in Figure 7, are of particular interest with respect to interface design, and at the same time, push the limits of our sensing capability. From these three groupings, we derived five different experimental conditions, described below.

J. *Fingers (five locations)*

One set of gestures we tested had participants tapping on the tips of each of their five fingers (Figure 6, "Fingers"). The fingers offer interesting affordances that make them compelling to appropriate for input. Foremost, they provide clearly discrete interaction points, which are even already well-named (e.g., ring finger). In addition to five finger tips, there are 14 knuckles (five major, nine minor), which, taken together, could offer 19 readily identifiable input locations on the fingers alone. Second, we have exceptional finger-tofinger dexterity, as demonstrated when we count by tapping on our fingers. Finally, the fingers are linearly ordered, which is potentially useful for interfaces like number entry, magnitude control (e.g., volume), and menu selection.

At the same time, fingers are among the most uniform appendages on the body, with all but the thumb sharing a similar skeletal and muscular structure. This drastically reduces acoustic variation and makes differentiating among them difficult. Additionally, acoustic information must cross as many as five (finger and wrist) joints to reach the forearm, which further dampens signals. For this experimental condition, we thus decided to place the sensor arrays on the forearm, just below the elbow. Despite these difficulties, pilot experiments showed measureable acoustic differences among fingers, which we theorize is primarily related to finger length and thickness, interactions with the complex structure of the wrist bones, and variations in the acoustic transmission properties of the muscles extending from the fingers to the forearm.

Whole Arm (Five Locations): Another gesture set investigated the use of five input locations on the forearm and hand: arm, wrist, palm, thumb and middle finger (Figure 7, "Whole Arm"). We selected these

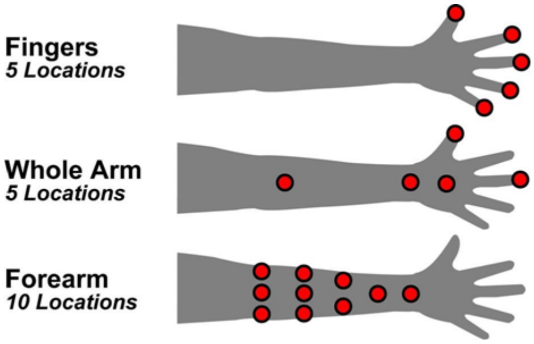

Fig. 7. The three input location sets evaluated in the study

We used these locations in three different conditions. One condition placed the sensor above the elbow, while another placed it below. This was incorporated into the experiment to measure the accuracy loss across this significant articulation point (the elbow). Additionally, participants repeated the lower placement condition in an eyes-free context: participants were told to close their eyes and face forward, both for training and testing. This condition was included to gauge how well users could target on-body input locations in an eyes-free context (e.g., driving).

K. *Forearm (Ten Locations)*

In an effort to assess the upper bound of our approach's sensing resolution, our fifth and final experimental condition used ten locations on just the forearm (Figure 6, "Forearm"). Not only was this a very high density of input locations (unlike the whole-arm condition), but it also relied on an input surface (the forearm) with a high degree of physical uniformity (unlike, e.g., the hand). We expected that these factors would make acoustic sensing difficult. Moreover, this location was compelling due to its large and flat surface area, as well as its immediate accessibility, both visually and for finger input. Simultaneously, this makes for an ideal projection surface for dynamic interfaces.

To maximize the surface area for input, we placed the sensor above the elbow, leaving the entire forearm free. Rather than naming the input locations, as was done in the previously described conditions, we employed small, colored stickers to mark input targets. This was both to reduce confusion (since locations on the forearm do not have common names) and to increase input consistency. As mentioned previously, we believe the forearm is ideal for projected interface elements; the stickers served as low-tech placeholders for projected buttons.

L. *Design and setup*

We employed a within-subjects design, with each participant performing tasks in each of the five conditions in randomized order: five fingers with sensors below elbow; five points on the whole arm with the sensors above the elbow; the same points with sensors below the elbow, both sighted and blind; and ten marked points on the forearm with the sensors above the elbow.

Participants were seated in a conventional office chair, in front of a desktop computer that presented stimuli. For conditions with sensors below the elbow, we placed the armband ~3cm away from the elbow, with one sensor package near the radius and the other near the ulna. For conditions with the sensors above the elbow, we placed the armband ~7cm above the elbow, such that one sensor package rested on the biceps. Right-handed participants had the armband placed on the left arm, which allowed them to use their dominant hand for finger input. For the one left-handed participant, we flipped the setup, which had no apparent effect on the operation of the system. Tightness of the armband was adjusted to be firm, but comfortable. While performing tasks, participants could place their elbow on the desk, tucked against their body, or on the chair's adjustable armrest; most chose the latter.

M. *Procedure*

For each condition, the experimenter walked through the input locations to be tested and demonstrated finger taps on each. Participants practiced duplicating these motions for approximately one minute with each gesture set. This allowed participants to familiarize themselves with our naming conventions (e.g. "pinky", "wrist"), and to practice tapping their arm and hands with a finger on the opposite hand. It also allowed us to convey the appropriate tap force to participants, who often initially tapped unnecessarily hard. To train the system, participants were instructed to comfortably tap each location ten times, with a finger of their choosing. This constituted one training round. In total, three rounds of training data were collected per input location set (30 examples per location, 150 data points total). An exception to this procedure was in the case of the ten forearm locations, where only two rounds were collected to save time (20 examples per location, 200 data points total). Total training time for each experimental condition was approximately three minutes.

We used the training data to build an SVM classifier. During the subsequent testing phase, we presented participants with simple text stimuli (e.g. "tap your wrist"), which instructed them where to tap. The order of stimuli was randomized, with each location appearing ten times in total. The system performed real-time segmentation and classification, and provided immediate feedback to the participant (e.g. "you tapped your wrist"). We provided feedback so that participants could see where the system was making errors (as they would if using a real application). If an input was not segmented (i.e. the tap was too quiet), participants could see this and would simply tap again. Overall, segments not included in further analysis.

3. Results

In this section, we report on the classification accuracies for the test phases in the five different conditions. Overall, classification rates were high, with an average accuracy across conditions of 87.6%. Additionally, we present preliminary results exploring the correlation between classification accuracy and factors such as BMI, age, and sex.

A. *Five fingers*

Despite multiple joint crossings and ~40cm of separation between the input targets and sensors, classification accuracy remained high for the five-finger condition, averaging 87.7% (SD=10.0%, chance=20%) across participants. Segmentation, as in other conditions, was essentially perfect.

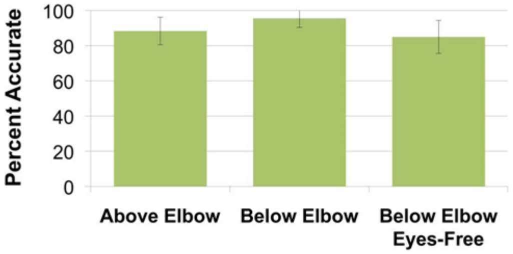

Fig. 8. Accuracy of the three whole-arm-centric conditions. Error bars represent standard deviation. mentation error rates were negligible in all conditions, and

Inspection of the confusion matrices showed no systematic errors in the classification, with errors tending to be evenly distributed over the other digits. When classification was incorrect, the system believed the input to be an adjacent finger

60.5% of the time; only marginally above prior probability (40%). This suggests there are only limited acoustic continuities between the fingers. The only potential exception to this was in the case of the pinky, where the ring finger constituted 63.3% percent of the misclassifications.

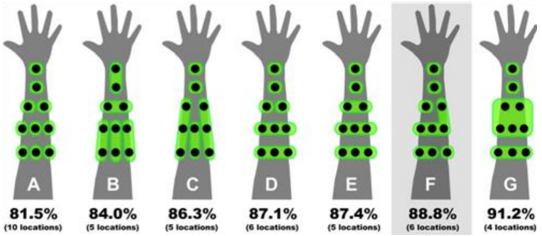

Fig. 9. Higher accuracies can be achieved by collapsing the ten input locations into groups. A-E and G were created using a design-centric strategy. F was created following analysis of per-location accuracy data.

B. *Whole arm*

Participants performed three conditions with the whole-arm location configuration. The below-elbow placement performed the best, posting a 95.5% (SD=5.1%, chance=20%) average accuracy. This is not surprising, as this condition placed the sensors closer to the input targets than the other conditions. Moving the sensor above the elbow reduced accuracy to 88.3% (SD=7.8%, chance=20%), a drop of 7.2%. This is almost certainly related to the acoustic loss at the elbow joint and the additional 10cm of distance between the sensor and input targets. Figure 8 shows these results. The eyes-free input condition yielded lower accuracies than other conditions, averaging 85.0% (SD=9.4%, chance=20%). This represents a 10.5% drop from its vision assisted, but otherwise identical counterpart condition. It was apparent from watching participants complete this condition that targeting precision was reduced. In sighted conditions, participants appeared to be able to tap locations with perhaps a 2cm radius of error. Although not formally captured, this margin of error appeared to double or triple when the eyes were closed. We believe that additional training data, which better covers the increased input variability, would remove much of this deficit. We would also caution designers developing eyes-free, on-body interfaces to carefully consider the locations participants can tap accurately.

C. *Forearm*

Classification accuracy for the ten-location forearm condition stood at 81.5% (SD=10.5%, chance=10%), a surprisingly strong result for an input set we devised to push our system's sensing limit (K=0.72, considered very strong). Following the experiment, we considered different ways to improve accuracy by collapsing the ten locations into larger input groupings. The goal of this exercise was to explore the tradeoff between classification accuracy and number of input locations on the forearm, which represents a particularly valuable input surface for application designers. We grouped targets into sets based on what we believed to be logical spatial groupings (Figure 9, A-E and G). In addition to exploring classification accuracies for layouts that we considered to be intuitive, we also performed an exhaustive search (programmatically) over all possible groupings. For most location counts, this search confirmed that our intuitive groupings were optimal; however, this search revealed one plausible, although irregular, layout with high accuracy at six input locations (Figure 9, F).

Unlike in the five-fingers condition, there appeared to be shared acoustic traits that led to a higher likelihood of confusion with adjacent targets than distant ones. This effect was more prominent laterally than longitudinally. Figure 9 illustrates this with lateral groupings consistently outperforming similarly arranged, longitudinal groupings (B and C vs. D and E). This is unsurprising given the morphology of the arm, with a high degree of bilateral symmetry along the long axis.

D. *BMI Effects*

Early on, we suspected that our acoustic approach was susceptible to variations in body composition. This included, most notably, the prevalence of fatty tissues and the density/mass of bones. These, respectively, tend to dampen or facilitate the transmission of acoustic energy in the body. To assess how these variations affected our sensing accuracy, we calculated each participant's body mass index (BMI) from self-reported weight and height. Data and observations from the experiment suggest that high BMI is correlated with decreased accuracies. The participants with the three highest BMIs (29.2, 29.6, and 31.9 – representing borderline obese to obese) produced the three lowest average accuracies. Figure 10 illustrates this significant disparity - here participants are separated into two groups, those with BMI greater and less than the US national median, age and sex adjusted [5] (F1,12=8.65, p=.013).

Other factors such as age and sex, which may be correlated to BMI in specific populations, might also exhibit a correlation with classification accuracy. For example, in our participant pool, males yielded higher classification accuracies than females, but we expect that this is an artifact of BMI correlation in our sample, and probably not an effect of sex directly.

E. *Supplemental experiments*

We conducted a series of smaller, targeted experiments to explore the feasibility of our approach for other applications. In the first additional experiment, which tested performance of the system while users walked and jogged, we recruited one male (age 23) and one female (age 26) for a single-purpose experiment. For the rest of the experiments, we recruited seven new participants (3 female, mean age 26.9) from within our institution. In all cases, the sensor armband was placed just below the elbow. Similar to the previous experiment, each additional experiment consisted of a training phase, where participants provided between 10 and 20 examples for each input type, and a testing phase, in which participants were prompted to provide a particular input (ten times per input type). As before, input order was randomized; segmentation and classification were performed in real-time.

F. *Walking and jogging*

As discussed previously, acoustically-driven input techniques are often sensitive to environmental noise. In regard to bio-acoustic sensing, with sensors coupled to the body, noise created during other motions is particularly troublesome, and walking and jogging represent perhaps the most common types of whole-body motion. This experiment explored the accuracy of our system in these scenarios.

Each participant trained and tested the system while walking and jogging on a treadmill. Three input locations were used to evaluate accuracy: arm, wrist, and palm. Additionally, the rate of false positives (i.e., the system believed there was input when in fact there was not) and true positives (i.e., the system was able to correctly segment an intended input) was captured. The testing phase took roughly three minutes to complete (four trials total: two participants, two conditions). The male walked at 2.3 mph and jogged at 4.3 mph; the female at 1.9 and 3.1 mph, respectively.

In both walking trials, the system never produced a falsepositive input. Meanwhile, true positive accuracy was 100%. Classification accuracy for the inputs (e.g., a wrist tap was recognized as a wrist tap) was 100% for the male and 86.7% for the female (chance=33%).

In the jogging trials, the system had four false-positive input events (two per participant) over six minutes of continuous jogging. True-positive accuracy, as with walking, was 100%. Considering that jogging is perhaps the hardest input filtering and segmentation test, we view this result as extremely positive. Classification accuracy, however, decreased to 83.3% and 60.0% for the male and female participants respectively (chance=33%).

Although the noise generated from the jogging almost certainly degraded the signal (and in turn, lowered classification accuracy), we believe the chief cause for this decrease was the quality of the training data. Participants only provided ten examples for each of three tested input locations. Furthermore, the training examples were collected while participants were jogging. Thus, the resulting training data was not only highly variable, but also sparse – neither of which is conducive to accurate machine learning classification. We believe that more rigorous collection of training data could yield even stronger results.

G. *Single-handed gestures*

In the experiments discussed thus far, we considered only bimanual gestures, where the sensor-free arm, and in particular the fingers, are used to provide input. However, there are a range of gestures that can be performed with just the fingers of one hand. This was the focus of [2], although this work did not evaluate classification accuracy.

We conducted three independent tests to explore onehanded gestures. The first had participants tap their index, middle, ring and pinky fingers against their thumb (akin to a pinching gesture) ten times each. Our system was able to identify the four input types with an overall accuracy of 89.6% (SD=5.1%, chance=25%). We ran an identical experiment using flicks instead of taps (i.e., using the thumb as a catch, then rapidly flicking the fingers forward). This yielded an impressive 96.8% (SD=3.1%, chance=25%) accuracy in the testing phase.

This motivated us to run a third and independent experiment that combined taps and flicks into a single gesture set. Participants re-trained the system, and completed an independent testing round. Even with eight input classes in very close spatial proximity, the system was able to achieve a remarkable 87.3% (SD=4.8%, chance=12.5%) accuracy. This result is comparable to the aforementioned ten location forearm experiment (which achieved 81.5% accuracy), lending credence to the possibility of having ten or more functions on the hand alone. Furthermore, proprioception of our fingers on a single hand is quite accurate, suggesting a mechanism for high-accuracy, eyes-free input.

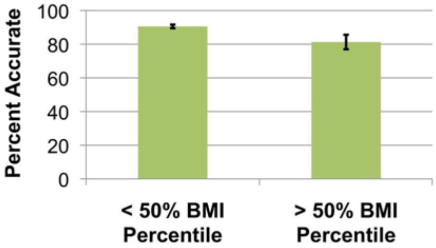

Fig. 10. Accuracy was significantly lower for participants with BMIs above the 50th percentile.

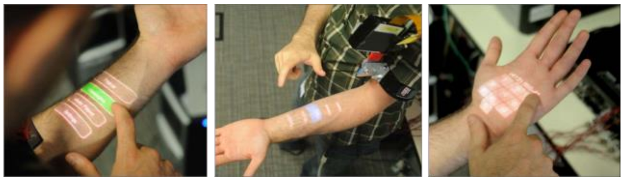

Fig. 11. Our sensing armband augmented with a pico-projector; this allows interactive elements to be rendered on the skin

H. *Surface and object recognition*

During piloting, it became apparent that our system had some ability to identify the type of material on which the user was operating. Using a similar setup to the main experiment, we asked participants to tap their index finger against 1) a finger on their other hand, 2) a paper pad approximately 80 pages thick, and 3) an LCD screen. Results show that we can identify the contacted object with about 87.1% (SD=8.3%, chance=33%) accuracy. This capability was never considered when designing the system, so superior acoustic features may exist. Even as accuracy stands now, there are several interesting applications that could take advantage of this functionality, including workstations or devices composed of different interactive surfaces, or recognition of different objects grasped in the environment.

I. *Identification of finger tap type*

Users can “tap” surfaces with their fingers in several distinct ways. For example, one can use the tip of their finger (potentially even their finger nail) or the pad (flat, bottom) of their finger. The former tends to be quite boney, while the latter more fleshy. It is also possible to use the knuckles (both major and minor metacarpophalangeal joints).

To evaluate our approach's ability to distinguish these input types, we had participants tap on a table situated in front of them in three ways (ten times each): fingertip, finger pad, and major knuckle. A classifier trained on this data yielded an average accuracy of 89.5% (SD=4.7%, chance=33%) during the testing period.

This ability has several potential uses. Perhaps the most notable is the ability for interactive touch surfaces to distinguish different types of finger contacts (which are indistinguishable in e.g., capacitive and vision-based systems). One example interaction could be that "double-knocking" on an item opens it, while a "pad-tap" activates an options menu.

J. Segmenting finger input

A pragmatic concern regarding the appropriation of fingertips for input was that other routine tasks would generate false positives. For example, typing on a keyboard strikes the finger tips in a very similar manner to the finger-tipinput we proposed previously. Thus, we set out to explore whether finger-to-finger input sounded sufficiently distinct such that other actions could be disregarded.

As an initial assessment, we asked participants to tap their index finger 20 times with a finger on their other hand, and 20 times on the surface of a table in front of them. This data was used to train our classifier. This training phase was followed by a testing phase, which yielded a participant wide average accuracy of 94.3% (SD=4.5%, chance=50%).

K. Example interfaces and interactions

We conceived and built several prototype interfaces that demonstrate our ability to appropriate the human body, in this case the arm, and use it as an interactive surface. These interfaces can be seen in Figure 11, as well as in the accompanying video.

While the bio-acoustic input modality is not strictly tethered to a particular output modality, we believe the sensor form factors we explored could be readily coupled with visual output provided by an integrated pico-projector. There are two nice properties of wearing such a projection device on the arm that permit us to sidestep many calibration issues. First, the arm is a relatively rigid structure - the projector, when attached appropriately, will naturally track with the arm (see video). Second, since we have fine-grained control of the arm, making minute adjustments to align the projected image with the arm is trivial (e.g., projected horizontal stripes for alignment with the wrist and elbow).

To illustrate the utility of coupling projection and finger input on the body (as researchers have proposed to do with projection and computer vision-based techniques [19]), we developed three proof-of-concept projected interfaces built on top of our system's live input classification. In the first interface, we project a series of buttons onto the forearm, on which a user can finger tap to navigate a hierarchical menu (Figure 11, left). In the second interface, we project a scrolling menu (center), which a user can navigate by tapping at the top or bottom to scroll up and down one item respectively. Tapping on the selected item activates it. In a third interface, we project a numeric keypad on a user's palm and allow them to tap on the palm to, e.g., dial a phone number (right). To emphasize the output flexibility of approach, we also coupled our bio-acoustic input to audio output. In this case, the user taps on preset locations on their forearm and hand to navigate and interact with an audio interface.

4. Conclusion

In this paper, we have presented our approach to appropriating the human body as an input surface. We have described a novel, wearable bio-acoustic sensing array that we built into an armband in order to detect and localize finger taps on the forearm and hand. Results from our experiments have shown that our system performs very well for a series of gestures, even when the body is in motion. Additionally, we have presented initial results demonstrating other potential uses of our approach, which we hope to further explore in future work. These include single-handed gestures, taps with different parts of the finger, and differentiating between materials and objects. We conclude with descriptions of several prototype applications that demonstrate the rich design space we believe Skinput enables.

Acknowledgement

We gratefully acknowledge Equipoise Technologies, American International Enterprises and Measurement Specialties for supplying sensors. We also thank LOUD Technologies for graciously providing a Mackie Onyx 1200F. Finally, we are in indebted to John Lewis and Johnny Lee for their thoughts and pico projectors.

References

[1] Ahmad, F., and Musilek, P. A Keystroke and Pointer Control Input Interface for Wearable Computers. In *Proc. IEEE PERCOM '06*, 2-11.

[2] Amento, B., Hill, W., and Terveen, L. The Sound of One Hand: A Wrist-mounted Bio-acoustic Fingertip Gesture Interface. In *CHI '02 Ext. Abstracts*, 724-725.

[3] Argyros, A.A., and Lourakis, M.I.A. Vision-based Interpretation of Hand Gestures for Remote Control of a Computer Mouse. In *Proc. ECCV 2006 Workshop on Computer Vision in HCI, LNCS 3979*, 40-51.

[4] Burges, C.J. A Tutorial on Support Vector Machines for Pattern Recognition. *Data Mining and Knowledge Discovery*, 2.2, June 1998, 121-167.

[5] Clinical Guidelines on the Identification, Evaluation, and Treatment of Overweight and Obesity in Adults. National Heart, Lung and Blood Institute. Jun. 17, 1998.

[6] Deyle, T., Palinko, S., Poole, E.S., and Starner, T. Hambone: A Bio-Acoustic Gesture Interface. In *Proc. ISWC '07*. 1-8.

[7] Erol, A., Bebis, G., Nicolescu, M., Boyle, R.D., and Twombly, X. Vision-based hand pose estimation: A review. *Computer Vision and Image Understanding*. 108, Oct., 2007.

[8] Fabiani, G.E. McFarland, D.J. Wolpaw, J.R. and Pfurtscheller, G. Conversion of EEG activity into cursor movement by a brain-computer interface (BCI). *IEEE Trans. on Neural Systems and Rehabilitation Engineering*, 12.3, 331-8. Sept. 2004.

[9] Grimes, D., Tan, D., Hudson, S.E., Shenoy, P., and Rao, R. Feasibility and pragmatics of classifying working memory load with an electroencephalograph. *Proc. CHI '08*, 835-844.

[10] Harrison, C., and Hudson, S.E. Scratch Input: Creating Large, Inexpensive, Unpowered and Mobile Finger Input Surfaces. In *Proc. UIST '08*, 205-208.

[11] Hirshfield, L.M., Solovey, E.T., Girouard, A., Kebinger, J., Jacob, R.J., Sassaroli, A., and Fantini, S. Brain measurement for usability testing and adaptive interfaces: an example of uncovering syntactic workload with functional near infrared spectroscopy. In *Proc. CHI '09, 2185-2194.*

[12] Ishii, H., Wisneski, C., Orbanes, J., and Chun, B., Paradiso, J. PingPongPlus: design of an athletic-tangible interface for computer-supported cooperative play. *Proc. CHI '99*, 394-401.

[13] Lakshmipathy, V., Schmandt, C., and Marmasse, N. TalkBack: a conversational answering machine. In *Proc. UIST '03*, 41-50.

[14] Lee, J.C., and Tan, D.S. Using a low-cost electroencephalograph for task classification in HCI research. In *Proc. CHI '06*, 81-90.

[15] Lyons, K., Skeels, C., Starner, T., Snoeck, C. M., Wong, B.A., andAshbrook, D. Augmenting conversations using dualpurpose speech. In *Proc. UIST '04*. 237-246.

[16] Mandryk, R.L., and Atkins, M.S. A Fuzzy Physiological Approach for Continuously Modeling Emotion During Interaction with Play Environments. *Intl Journal of Human-Computer Studies*, 6(4), 329-347, 2007.

[17] Mandryk, R.L., Inkpen, K.M., and Calvert, T.W. Using Psychophysiological Techniques to Measure User Experience with Entertainment Technologies. *Behaviour and Information Technology*, 25(2), 141-58, March 2006.

[18] McFarland, D.J., Sarnacki, W.A., and Wolpaw, J.R. Brain– computer interface (BCI) operation: optimizing information transfer rates. *Biological Psychology*, 63(3), 237-51. Jul 2003.