Skinput: Appropriating the Skin as an Interactive Canvas

By Chris Harrison, Desney Tan, and Dan Morris

Abstract

Skinput is a technology that appropriates the skin as an input surface by analyzing mechanical vibrations that propagate through the body. Specifically, we resolve the location of finger taps on the arm and hand using a novel sensor array, worn as an armband. This approach provides an on-body finger input system that is always available, naturally portable, and minimally invasive. When coupled with a pico-projector, a fully interactive graphical interface can be rendered directly on the body. **To view video of Skinput, visit http://cacm.acm.org.**

1. INTRODUCTION

Devices with significant computational power and capability can now be easily carried with us. These devices have tremendous potential to bring the power of information, computation, creation, and communication to a wider audience and to more aspects of our lives. However, this potential raises new challenges for interaction design. For example, miniaturizing devices has simultaneously reduced their interactive surface area. This has led to diminutive screens, cramped keyboards, and tiny jog wheels, all of which impose restrictions that diminish usability and prevent us from realizing the full potential of mobile computing. Consequently, mobile devices are approaching the computational capabilities of desktop computers, but are hindered by a human–computer I/O bottleneck.

Critically, this is a problem we cannot engineer ourselves out of. While we can make computer processors faster, LCD screens thinner, and hard drives larger, we cannot add surface area without increasing size—it is a physical constraint. This has trapped us in a device size paradox: we want more usable devices, but are unwilling to sacrifice the benefits of small size and mobility. In response, designers have walked a fine line, trying to strike a balance between usability and mobility.

One promising approach to mitigate this is to appropriate surface area from the environment for interactive purposes. This can offer larger interactive surface area with no increase in device size. For example, Harrison and Hudson[7] describe a technique that allows (small) mobile devices to turn (large) tables into gestural finger input canvases. 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 previously overlooked as an input surface, 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 2 m² of 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.

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. When coupled with a pico-projector, the skin can operate as an interactive canvas supporting both input and graphical output (Figures 1 and 2).
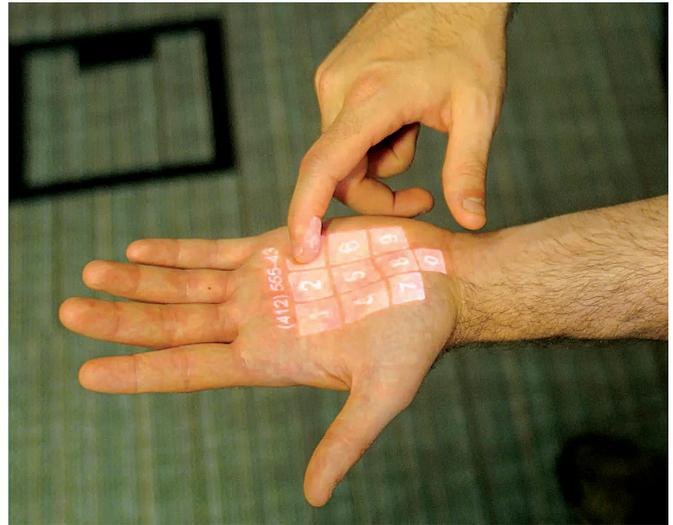
2. RELATED WORK

2.1. Always-available input

A primary goal of *Skinput* is to provide an always-available mobile input system—for example, an input system that does not require a user to carry or pick up a device. A number of alternative solutions to this problem have been proposed. Techniques based on computer vision are popular (e.g., Argyros and Lourakis,[2] Mistry et al.,[16] Wilson,[24, 25] see Erol et al.[5] for a recent survey). These, however, are computationally expensive and error prone in mobile scenarios (where, e.g., non-input optical flow is prevalent), or depend on cumbersome instrumentation of the hands to enhance performance. Speech input (e.g., Lakshmipathy et al.[9] and Lyons et al.[11]) is a logical choice for always-available input, but is limited in its precision in unpredictable acoustic environments, suffers from privacy and scalability issues in shared environments, and may interfere with cognitive tasks significantly more than manual interfaces.[22]

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 Sturman and Zeltzer[23] 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[19] present a "smart fabric" system that embeds sensors and conductors into fabric, but taking this approach to always-available input necessitates

Figure 1. Our sensing armband augmented with a pico-projector; this allows interactive elements to be rendered on the skin.

embedding technology in all clothing, which would be prohibitively complex and expensive.

The SixthSense project[16] proposes a mobile, always-available I/O system by combining projected information with a color-marker-based vision tracking system. This approach is feasible, but suffers from the limitations of vision-based systems discussed above and requires instrumentation of the fingertips. Like SixthSense, we explore the combination of on-body sensing with on-body projection.

2.2. 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., Mandryk and Atkins,[12] Mandryk et al.,[13] Moore and Dua[17]). 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., Grimes et al.,[6] Lee and Tan[10]); 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., McFarland et al.[15]), 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., Rosenberg[20] and Saponas et al.[21]). 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.,[1] who placed contact microphones on a user's wrist to assess finger movement. However, this work was never formally evaluated and is constrained to finger motions in one hand. The Hambone system[4] employs a similar approach using piezoelectric sensors, yielding classification accuracies around 90% for four gestures (e.g., raise heels, snap fingers). Performance of false positive rejection remains untested in both systems.

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 lossy transmission of sound through the air and outer ear. The mechanically conductive properties of human bones are also employed by Zhong et al.[27] for transmitting information through the body, such as from an implanted device to an external receiver.

2.3. Acoustic input

Our approach is also inspired by systems that leverage acoustic transmission through (non-body) input surfaces. Paradiso et al.[18] measured the arrival time of a sound at multiple sensors to locate hand taps on a glass window. Ishii et al.[8] 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.

3. SKINPUT

To expand the range of sensing modalities for always-available input systems, we developed *Skinput*, a novel input technique that allows the skin to be used as a finger input surface, much like a touchscreen. 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 (Figure 2).

3.1. 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 3). When shot with a high-speed camera, these appear as ripples, which propagate outward from the point of contact (like a pebble into a pond). The amplitude of these ripples is correlated to the tapping force and 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 4). These longitudinal (compressive) waves

Figure 2. Skinput rendering a list interface rendered on the arm. Pico-projector and sensing armband highlighted.
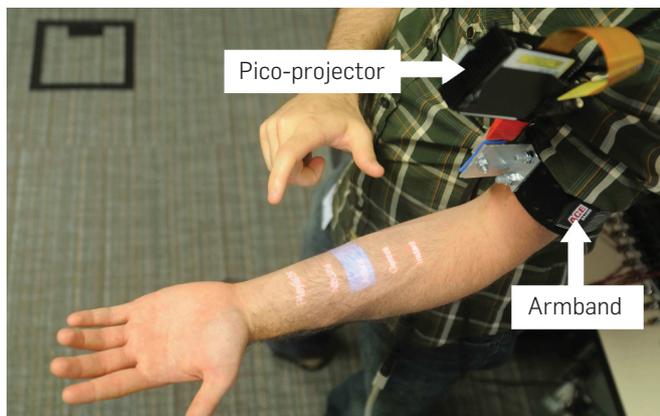

Pico-projector

Armband

Figure 3. *Transverse wave propagation*: Finger impacts displace the skin, creating transverse waves (ripples). The sensor is activated as the wave passes underneath it.
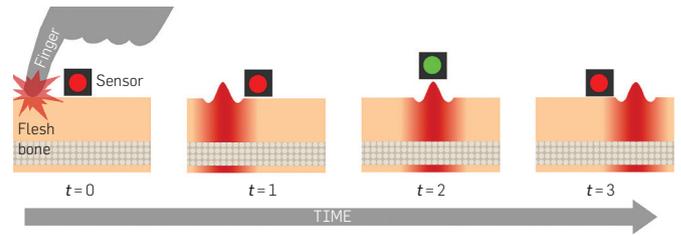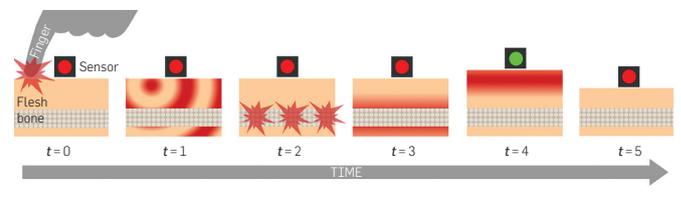

Figure 4. Finger impacts create longitudinal (compressive) waves that cause internal skeletal structures to vibrate. This, in turn, creates longitudinal waves that emanate outward from the bone (along its entire length) toward the skin.

travel through the soft tissues of the arm, exciting the bone, which is much less deformable than 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.

Similarly, we also hypothesize 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. Finally, muscle contraction may also contribute to the vibration patterns recorded by our sensors,[14] including both contraction related to posture maintenance and reflexive muscle movements in response to input taps.

3.2. Armband prototype

Our initial hardware prototype employed an array of tuned

mechanical vibration sensors; specifically small, cantilevered piezoelectric films (MiniSense100, Measurement Specialties, Inc.). By adding small weights to the end of the cantilever, we were able to alter the resonant frequency, allowing each sensing element to be responsive to a unique, narrow, low-frequency band of the acoustic spectrum. Each element was aligned with a particular frequency pilot study shown to be useful in characterizing bio-acoustic input. These sensing elements were packaged into 2 groups of 5–10 sensors in total.
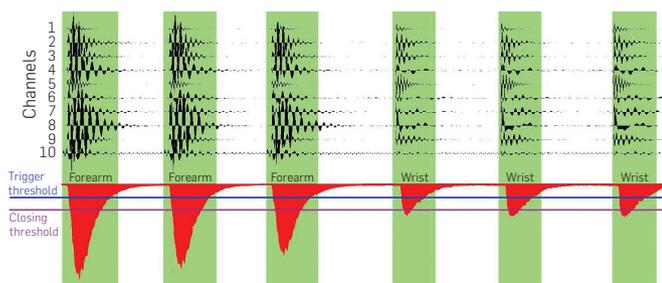
A Mackie Onyx 1200F audio interface was used to digitally capture data from the 10 sensors. Each channel was sampled at 5.5 kHz, a sampling rate that would be considered too low for speech or environmental audio, but was able to represent the relevant spectrum of frequencies transmitted through the arm. 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 77 kHz with no loss of precision, and could therefore provide the full sampling power required for *Skinput* (55 kHz in total).

3.3. Processing
The audio stream was segmented into individual taps using an absolute exponential average of all sensor channels (Figure 5, red waveform). When an intensity threshold was exceeded (Figure 5, 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 5, lower purple line) between 100 and 700 ms 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 60 ms buffer on either end) was considered an input event (Figure 5, vertical green regions). Although simple, this heuristic proved to be robust.

After an input has been segmented, the waveforms are analyzed. 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 10 channels, although only the lower 10 values are used (representing the acoustic power from 0 to 193 Hz), 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 0–193 Hz 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 Burges[3] for a tutorial). Our software uses the implementation provided in the Weka machine learning toolkit.[26] 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-the-fly 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.

4. EXPERIMENT

4.1. Participants
To evaluate the performance of our system, we recruited 13 participants (7 female) from the 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).
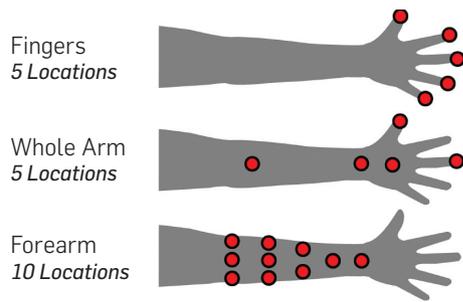
4.2. Experimental conditions
We selected three input groupings from the multitude of possible location combinations to test. We believe that these groupings, illustrated in Figure 6, are of particular interest with respect to interface design, and at the same time, push the limits of our sensing capability. From these three groupings, we derived five different experimental conditions, described below.

4.2.1. 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 clear, discrete interaction points, which are even

Figure 5. 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 activated by the two locations.

Figure 6. The three input location sets evaluated in the study.

Fingers
5 Locations

Whole Arm
5 Locations

Forearm
10 Locations

well-named (e.g., "ring finger"). In addition to 5 finger tips, there are 14 knuckles (5 major, 9 minor), which, taken together, could offer 19 readily identifiable input locations on the fingers alone. Second, we have exceptional finger-to-finger 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 measurable 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.

4.2.2. Whole arm (five locations)

Another task investigated the use of five input locations on the forearm and hand: arm, wrist, palm, thumb, and middle finger (Figure 6, "Whole Arm"). We selected these locations for two important reasons. First, they are distinct and named parts of the body (e.g., "wrist"). This allowed participants to accurately tap these locations without training or markings. Additionally, these locations proved to be acoustically distinct during piloting, with the large spatial spread of input points offering further variation.

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).

4.2.3. Forearm (10 locations)

In an effort to assess the upper bound of our approach's sensing resolution, our fifth and final experimental condition used 10 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.

4.3. 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 10 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 ~3 cm away from the elbow, with one sensor package on the "thumb" side of the forearm and one on the "pinky" side. For conditions with the sensors above the elbow, we placed the armband ~7 cm 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.

4.4. 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 1 min 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 10 times, with a finger of their choosing. This constituted 1 training round. In total, 3 rounds of training data were collected per input location set (30 examples per location, 150 data points in total). An exception to this procedure was in the case of the 10 forearm locations, where only 2 rounds were collected to save time (20 examples per location, 200 data points in total). Total training time for each experimental condition was approximately 3 min.

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 10 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, segmentation error rates were negligible in all conditions, and not included in further analysis.

5. 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%.

5.1. Five fingers

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

5.2. 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%) 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%), a drop of 7.2%. This is almost certainly related to the acoustic loss at the elbow joint and the additional 10 cm between the sensor and input targets.

The eyes-free input condition yielded lower accuracies than other conditions, averaging 85.0% (SD = 9.4%). 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 2 cm 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 captures the increased input variability, would remove much of this deficit. However, we also caution designers developing eyes-free, on-body interfaces to carefully consider the locations participants can tap accurately.

5.3. Forearm

Classification accuracy for the 10-location forearm condition stood at 81.5% (SD = 10.5%), a surprisingly strong result for an input set we purposely devised to tax our system's sensing accuracy.
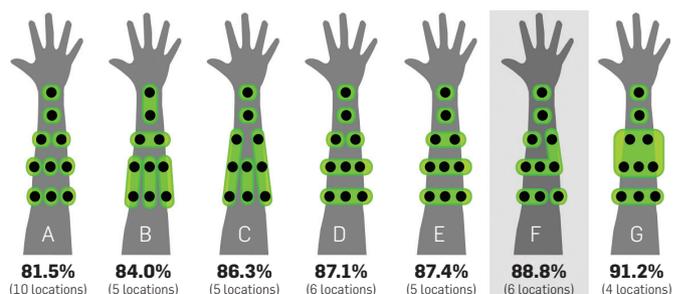
Using our experimental data, we considered different ways to improve accuracy by post hoc collapsing the 10 locations into input groupings. The goal of this exercise was to explore the tradeoff between classification accuracy and number of input locations on the forearm, which represents a particularly valuable input surface for application designers. We grouped targets into sets based on what we believed to be logical spatial groupings (Figure 7A–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 7F).

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 7 illustrates this with lateral groupings consistently out-performing 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.

6. SUPPLEMENTAL EXPERIMENTS

We conducted a series of smaller, targeted experiments

Figure 7. Higher accuracies can be achieved by collapsing the 10 input locations into groups. A–E and G were designed to be spatially intuitive. F was created following analysis of per-location accuracy data.

| 81.5% (10 locations) | 84.0% (5 locations) | 86.3% (5 locations) | 87.1% (6 locations) | 87.4% (5 locations) | 88.8% (6 locations) | 91.2% (4 locations) |

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 1 male (age 23) and 1 female (age 26) for a single-purpose experiment. For the rest of the experiments, we recruited 7 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 (10 times per input type). As before, input order was randomized; segmentation and classification were performed in real time.

6.1. Walking and jogging

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 3 min to complete (four trials in 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 false-positive 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.

In the jogging trials, the system had four false-positive input events (two per participant) over 6 min 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.

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 10 examples for each of 3 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.

6.2. 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 Amento et al.,[1] although this work did not evaluate classification accuracy.

We conducted three independent tests to explore one-handed gestures. The first had participants tap their index, middle, ring and pinky fingers against their thumb (akin to a pinching gesture) 10 times each. Our system was able to identify the four input types with an overall accuracy of 89.6% (SD = 5.1%). 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%) 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 retrained the system, and completed an independent testing round. Even with eight input classes in very close spatial proximity, the system was able to achieve 87.3% (SD = 4.8%) accuracy. This result is comparable to the aforementioned 10-location forearm experiment (which achieved 81.5% accuracy), lending credence to the possibility of having 10 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.

6.3. 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-tip input 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%).

7. 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.

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. 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[16]), we developed four proof-of-concept projected interfaces built on our system's live input classification. In the first interface, we project a series of buttons onto the forearm, on which a user can tap to navigate a hierarchical menu (Figure 1). In the second interface, we project a scrolling menu (Figure 2), which a user can navigate by tapping at the top or bottom to scroll up and down one item. 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, for example, dial a phone number (Figure 1). Finally, as a true test of real-time control, we ported Tetris to the hand, with controls bound to different fingertips.

8. FUTURE WORK

In order to assess the real-world practicality of *Skinput*, we are currently building a successor to our prototype that will incorporate several additional sensors, particularly electrical sensors (allowing us to sense the muscle activity associated with finger movement, as per[21]) and inertial sensors (accelerometers and gyroscopes). In addition to expanding the gesture vocabulary beyond taps, we expect this sensor fusion to allow considerably more accuracy—and more robustness to false positives—than each sensor alone. This revision of our prototype will also allow us to benefit from anecdotal lessons learned since building our first prototype: in particular, early experiments with subsequent prototypes suggest that the hardware filtering we describe above (weighting our cantilevered sensors to create a mechanical band-pass filter) can be effectively replicated in software, allowing us to replace our relatively large piezoelectric sensors with micro-machined accelerometers. This considerably reduces the size and electrical complexity of our armband. Furthermore, anecdotal evidence has also suggested that vibration frequency ranges as high as several kilohertz may contribute to tap classification, further motivating the use of broadband accelerometers. Finally, our multi-sensor armband will be wireless, allowing us to explore a wide variety of usage scenarios, as well as our general assertion that always-available input will inspire radically new computing paradigms.

9. CONCLUSION

In this paper, we presented our approach to appropriating the human body as an interactive surface. We have described a novel, wearable, bio-acoustic sensing approach that can detect and localize finger taps on the forearm and hand. Results from our experiments have shown that our system performs well for a series of gestures, even when the body is in motion. We conclude with descriptions of several prototype applications that graze the tip of the rich design space we believe *Skinput* enables. C

References

1. Amento, B., Hill, W., Terveen, L. The sound of one hand: A wrist-mounted bio-acoustic fingertip gesture interface. In *Proceedings of the CHI '02 Extended Abstracts* (2002), 724–725.
2. Argyros, A.A., Lourakis, M.I.A. Vision-based interpretation of hand gestures for remote control of a computer mouse. In *Proceedings of the ECCV 2006 Workshop on Computer Vision in HCI, LNCS 3979* (2006), 40–51.
3. Burges, C.J. A tutorial on support vector machines for pattern recognition. *Data Min. Knowl. Discov. 2*, 2 (June 1998), 121–167.
4. Deyle, T., Palinko, S., Poole, E.S., Starner, T. Hambone: A bio-acoustic gesture interface. In *Proceedings of the ISWC'07* (2007), 1–8.
5. Erol, A., Bebis, G., Nicolescu, M., Boyle, R.D., Twombly, X. Vision-based hand pose estimation: A review. *Comput. Vision Image Underst. 108* (Oct. 2007), 52–73.
6. Grimes, D., Tan, D., Hudson, S.E., Shenoy, P., Rao, R. Feasibility and pragmatics of classifying working memory load with an electroencephalograph. In *Proceedings of the CHI'08* (2008), 835–844.
7. Harrison, C., Hudson, S.E. Scratch input: Creating large, inexpensive, unpowered and mobile finger input surfaces. In *Proceedings of the UIST'08* (2008), 205–208.
8. Ishii, H., Wisneski, C., Orbanes, J., Chun, B., Paradiso, J. PingPongPlus: Design of an athletic-tangible interface for computer-supported cooperative play. In *Proceedings of the CHI'99* (1999), 394–401.
9. Lakshmipathy, V., Schmandt, C., Marmasse, N. TalkBack: A conversational answering machine. In *Proceedings of the UIST'03* (2003), 41–50.
10. Lee, J.C., Tan, D.S. Using a low-cost electroencephalograph for task classification in HCI research. In *Proceedings of the CHI'06* (2006), 81–90.
11. Lyons, K., Skeels, C., Starner, T., Snoeck, C.M., Wong, B.A., Ashbrook, D. Augmenting conversations using dual-purpose speech. In *Proceedings of the UIST'04* (2004), 237–246.
12. Mandryk, R.L., Atkins, M.S. A fuzzy physiological approach for continuously modeling emotion during interaction with play environments. *Int. J. Hum. Comput. Stud. 6*, 4 (2007), 329–347.
13. Mandryk, R.L., Inkpen, K.M., Calvert, T.W. Using psychophysiological techniques to measure user

14. experience with entertainment technologies. *Behav. Inf. Technol. 25*, 2 (Mar. 2006), 141–158.
14. Matheson, G.O., Maffey-Ward, L., Mooney, M., Ladly, K., Fung, T., Zhang, Y.T. Vibromyography as a quantitative measure of muscle force production. *Scand. J. Rehabil. Med. 29*, 1 (Mar. 1997), 29–35.
15. McFarland, D.J., Sarnacki, W.A., Wolpaw, J.R. Brain–computer interface (BCI) operation: Optimizing information transfer rates. *Biol. Psychol. 63*, 3 (Jul. 2003), 237–251.
16. Mistry, P., Maes, P., Chang, L. WUW—Wear ur world: A wearable gestural interface. In *Proceedings of the CHI '09 Extended Abstracts* (2009), 4111–4116.
17. Moore, M., Dua, U. A galvanic skin response interface for people with severe motor disabilities. In *Proceedings of the ACM SIGACCESS Accessibility and Computers '04* (2004), 48–54.
18. Paradiso, J.A., Leo, C.K., Checka, N., Hsiao, K. Passive acoustic knock tracking for interactive windows. In *Proceedings of the CHI '02 Extended Abstracts* (2002), 732–733.
19. Post, E.R., Orth, M. Smart fabric, or "wearable clothing." In *Proceedings of the ISWC'97* (1997), 167.
20. Rosenberg, R. The biofeedback pointer: EMG control of a two dimensional pointer. In *Proceedings of the ISWC'98* (1998), 4–7.
21. Saponas, T.S., Tan, D.S., Morris, D., Balakrishnan, R. Demonstrating the feasibility of using forearm electromyography for muscle–computer interfaces. In *Proceedings of the CHI'09* (2009), 515–524.
22. Starner, T. The role of speech input in wearable computing. *IEEE Perv. Comput. 1*, 3 (July 2002), 89–93.
23. Sturman, D.J., Zeltzer, D. A survey of glove-based input. *IEEE Comput. Graph. Appl. 14*, 1 (Jan. 1994).
24. Wilson, A. PlayAnywhere: A compact interactive tabletop projection-vision system. In *Proceedings of the UIST '05* (2005), 83–92.
25. Wilson, A.D. Robust computer vision-based detection of pinching for one and two-handed gesture input. In *Proceedings of the UIST'06* (2006), 255–258.
26. Witten, I.H., Frank, E. *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd edn. Morgan Kaufmann, San Francisco, CA, 2005.
27. Zhong, L., El-Daye, D., Kaufman, B., Tobaoda, N., Mohamed, T., Liebschner, M. OsteoConduct: Wireless body-area communication based on bone conduction. In *Proceedings of the ICST'07* (2007), 1–8.

Chris Harrison (chris.harrison@cs.cmu.edu), Human-Computer Interaction Institute, Carnegie Mellon University, Pittsburgh, PA.

Desney Tan (desney@microsoft.com), Microsoft Research, Redmond, WA.

Dan Morris (dan@microsoft.com), Microsoft Research, Redmond, WA.