Toffee: Enabling Ad Hoc, Around-Device Interaction with Acoustic Time-of-Arrival Correlation

Robert Xiao[1] Greg Lew[2] James Marsanico[2] Divya Hariharan[2] Scott E. Hudson[1] Chris Harrison[1]

[1]Human-Computer Interaction Institute [2]Electrical and Computer Engineering

Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh PA 15213

[1]{brx, scott.hudson, chris.harrison}@cs.cmu.edu [2]{glew, jmarsani, dharihar}@andrew.cmu.edu

ABSTRACT

The simple fact that human fingers are large and mobile devices are small has led to the perennial issue of limited surface area for touch-based interactive tasks. In response, we have developed *Toffee*, a sensing approach that extends touch interaction beyond the small confines of a mobile device and onto ad hoc adjacent surfaces, most notably tabletops. This is achieved using a novel application of acoustic time differences of arrival (TDOA) correlation. Previous time-of-arrival based systems have required semi-permanent instrumentation of the surface and were too large for use in mobile devices. Our approach requires only a hard tabletop and gravity – the latter acoustically couples mobile devices to surfaces. We conducted an evaluation, which shows that *Toffee* can accurately resolve the bearings of touch events (mean error of 4.3° with a laptop prototype). This enables radial interactions in an area many times larger than a mobile device; for example, virtual buttons that lie above, below and to the left and right.

ACM Classification: H.5.2 [Information interfaces and presentation]: User Interfaces - Input devices and strategies, Graphical user interfaces.

Keywords: Time of flight correlation; time difference of arrival; TDOA; TOA; TOF; ADI; everyday surfaces; touch interaction; interfaces everywhere; vibro-acoustic sensors; around device interaction.

INTRODUCTION

In order for mobile devices to fit into our pockets and bags, they are generally designed with small screens and physical controls. Simultaneously, human fingers are relatively large (and are unlikely to shrink anytime soon). This has led to the recurring problem of limited surface area for touch-based interactive tasks. Thus, there is a pressing need to develop novel sensing approaches and interaction techniques that aim to mitigate this fundamental constraint.

One option is to transiently appropriate surface area from the environment around us [14]. This allows devices to remain small, but opportunistically provide large areas for accurate and comfortable input (and potentially graphical output if e.g., projectors are used). Tables, in particular, have presented an attractive target for researchers. Foremost, mobile devices often reside on tables, enabling several ad hoc sensing approaches (e.g., vibro-acoustic [13,18] and optical [7,20]). Moreover, unlike e.g., painted walls, tables are accepted areas for work, have durable surfaces, and typically have surface area available for interactive use.

We present *Toffee*, a system that allows devices to appropriate tables and other hard, flat surfaces they are placed on for ad hoc radial tap input. This is achieved using a novel application of acoustic time differences of arrival (TDOA) analysis [4,8,9,16,23,24,25] – bringing the technique, for the first time, to small devices in a way that is compatible with their inherent mobility. At a high level, *Toffee* allows users to define virtual, ad hoc buttons on a table's surface, which can then be used to trigger a variety of interactive functions, including application launching, desktop switching, music player control, and gaming. Leveraging the large surface area of the table, users can potentially trigger simple functionality eyes free.

OPERATION

Our *Toffee*-augmented devices have four vibro-acoustic piezo sensors, located on the underside of each corner (Figure 1). Their exact configuration is discussed in depth in the next section. When a finger or object taps a table, mechanical vibrations (i.e. structural acoustics) propagate outward from the point of contact [11,23]. The speed of propagation depends on the host medium and the frequency

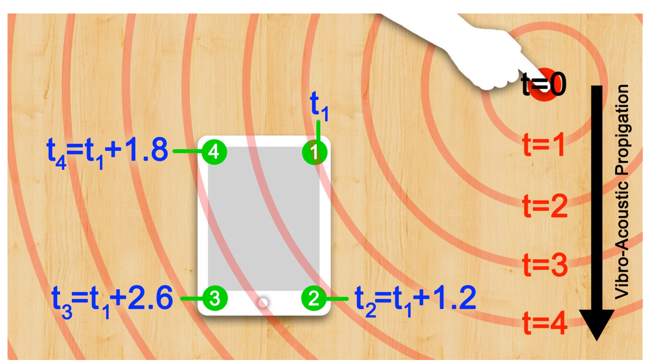

Figure 1. When a finger taps a table, mechanical vibration propagates outward from the point of contact (red concentric circles). Vibro-acoustic sensors (green) can detect when the leading edge of the wave reaches them. From these timing offsets (blue), it is possible to estimate the originating position of the finger tap.

MobileHCI 2014, September 23 - 26 2014, Toronto, ON, Canada

ACM 978-1-4503-3004-6/14/09 $15.00. http://dx.doi.org/10.1145/2628363.2628383

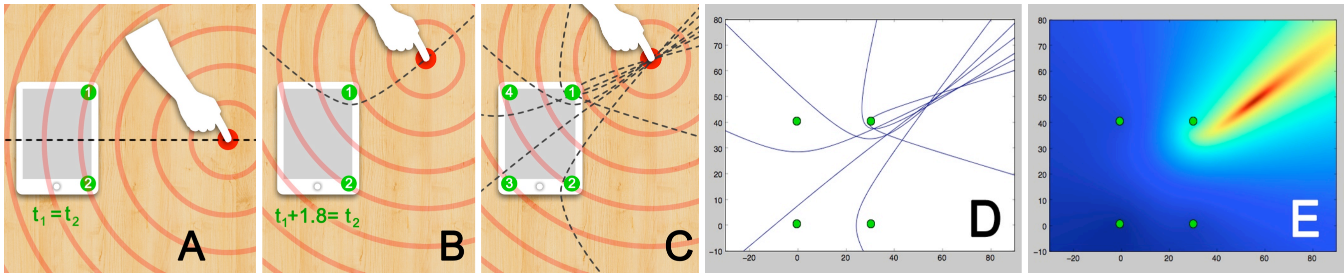

Figure 2. A: When a finger taps equidistant to two sensors, such that arrival time t_1=t_2, it is only possible to infer that the finger tapped somewhere along a line (dashed). B: If the finger lies closer to one sensor, this function becomes hyperbolic. C: In a four-sensor setup, six hyperbolas can be computed; intersections are solutions for the originating touch location. D: Real world data of the scenario in C plotted in matplotlib. E: A visualization of total squared error.

of the vibration. In solid materials, such as wood, waves typically propagate at around 600 meters per second [16]. Impulses will tend to disperse into their individual frequency components as they travel through the medium due to differences in propagation velocity between different frequencies. Importantly, because the propagation is not instantaneous, the leading edge of the fastest frequency component in the vibro-acoustic impulse reaches each of the four sensors at different times (Figure 1). Furthermore, this leading edge arrives at each sensor before any reflected wave, so that measurement of the leading edge is expected to be relatively robust to multipath propagation effects. Please see [23,24] for an extended discussion of the related physics.

From the timing data, we can estimate the originating point of contact. First, let us consider an example setup with only a pair of sensors (Figure 2A). If a touch occurs equidistant to the two sensors, the time difference of arrival will be the same. From this, the system can infer that the touch must have occurred somewhere along a straight line equidistant from the two sensors (Figure 2A, dashed line). However, if the touch is closer to one sensor than the other, the touch can be inferred to lie somewhere along a hyperbolic curve – the locus of points having a constant difference between distances to two fixed points (Figure 2B, dashed line).

With three sensors, there are two non-redundant pairings and thus the touch can be inferred to lie at one of the two intersections of two hyperbolic curves. To disambiguate between these two possibilities, we need one more sensor.

With four sensors there are six combinations of sensor pairs, which allow us to compute six hyperbolic functions (Figure 2C). In a perfect system with absolute accuracy, three of these pairings would be redundant, as the curves would all intersect at the location of the touch. In a system with some error (i.e., the real world), these loosely intersect at the probable location of the touch, and the additional "redundant" functions are used to reduce the error. Figures 2D and 2E show real world data for the scenario illustrated in Figure 2C.

HARDWARE IMPLEMENTATION

We constructed three prototype platforms to demonstrate the feasibility of our approach with devices of different size. Specifically, we augmented a Samsung Galaxy S3 smartphone, an Apple iPad 2 tablet, and a 15" Apple Macbook Pro (see Figures 3, 4 and 8). An acrylic frame was attached to the bottom of each device. This provided a flat surface, to which we affixed 20mm piezo discs at each corner. These piezo discs sense acoustic vibrations by bending slightly and compressing the piezoelectric material, generating a voltage differential. For our laptop and tablet prototypes, the acrylic frame is separated into four pieces connected by soft foam tape. This dampens the transmission of mechanical vibrations through the frame (Figure 4, center).

Each piezo disc is connected to a simple analog circuit. This is configured as an operational amplifier with unity gain (i.e., no amplification). Amplification is not necessary because the piezo sensors easily produce voltages in the

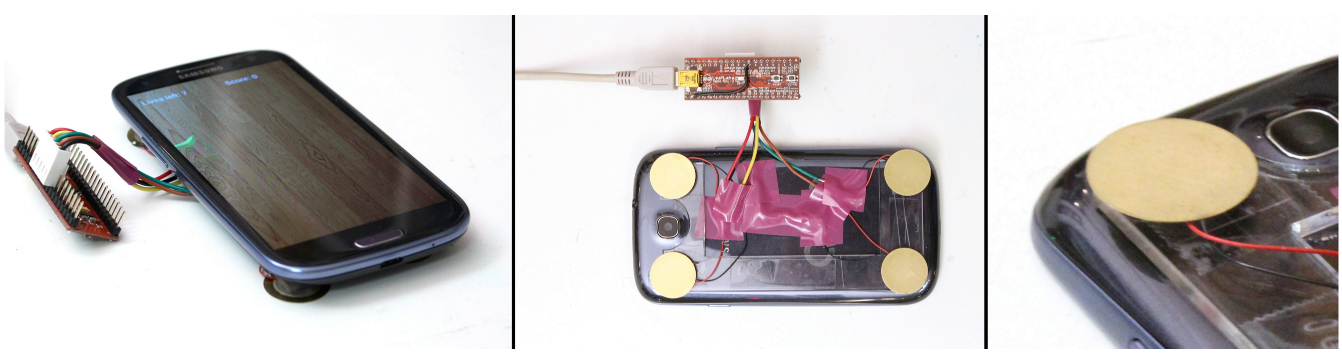

Figure 3. *Toffee*-Augmented smartphone. Piezo sensors are situated on a 107 x 49 mm acrylic frame.

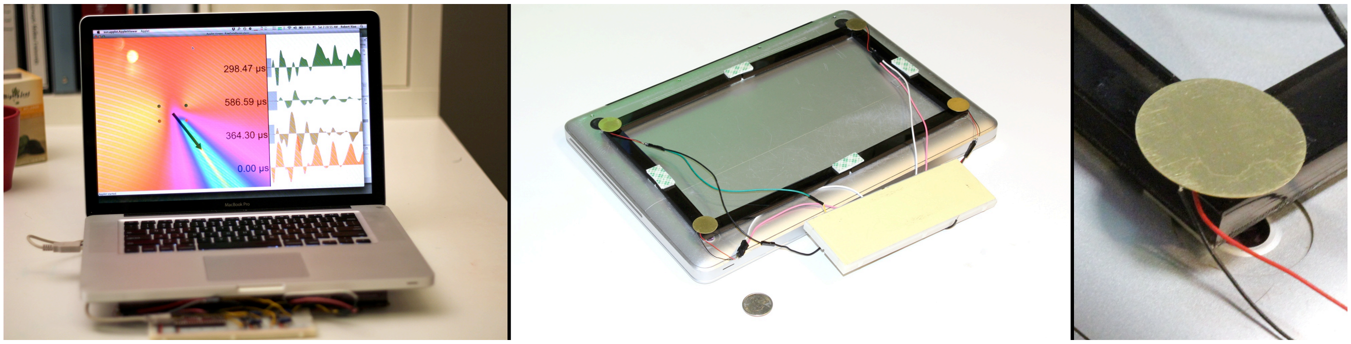

Figure 4. *Toffee*-Augmented laptop. Piezo sensors are situated on a 280 x 167 mm acrylic frame.

range of -10V to +10V. The circuit is used to provide the high input impedance (approximately 1MΩ) required by the piezo sensor, and to add a DC bias to enable our ADC to sample the full range of the signal. Due to limited space, this circuit was omitted for our smartphone prototype and the piezos were directly connected to the microcontroller.

The signal is sampled by the analog-to-digital converter (ADC) of an ARM Cortex M3-based Maple Mini micro-controller running at 72MHz [22]. Our embedded software samples the four piezo sensors at 85kHz (roughly 344,000 data points per second, the fastest possible rate on this hardware). This high sampling rate is required for capturing the tiny differences in the arrival time of the rapidly-propagating vibro-acoustic wave. However, due to the low frequency of the vibro-acoustic waveforms (Figure 5), using even higher sampling speeds will only marginally improve accuracy. When the magnitude of a converted analog value exceeds a predefined threshold, the Maple board delays for a short period to buffer the waveforms. Finally, four waveforms representing 18ms of data (one for each piezo sensor) are sent over USB to a computer for analysis, described in the next section.

SOFTWARE IMPLEMENTATION

Our *Toffee* engine (Figure 5) was developed in Java and performs the calculations necessary to localize a touch. As described above, when a touch occurs, the Maple board sends four waveforms captured by the four piezo sensors (Figure 5D). First, to reduce noise, the signal is smoothed using an exponentially-weighted moving average. The system then computes the discrete derivative of the smoothed signal. The first point (across all four waveforms) at which the derivative exceeds a predefined threshold is assumed to be the impulse resulting from the fastest frequency component of the vibro-acoustic wave. This time index anchors all further calculations (Figure 5D, top waveform).

We initially experimented with using cross-correlation to determine the propagation delay for the other three sensors. However, we discovered that the waveforms from the other sensors were often very different due to attenuation, multi-path interference, acoustic dispersion, and other effects (see e.g., [11,31] for an extended discussion of the relevant structural acoustics). This made cross-correlation a poor estimator of propagation delay – a result also found in [28]. Instead, we use the first impulse in the other waveforms as the time delta, measured in microseconds. We record the relative delay between each pair of sensors and use this information to estimate the position of the tap.

As noted previously, the time difference in the signal's arrival between a pair of sensors defines a hyperbola of possible origination points (Figure 2). Mathematically, if p is the position of the tap, s_i and s_j are the positions of the sensors, Δt_{ij} is the difference in the arrival times of the two sensors, and c is the speed of sound in the surface, then

$$\|p - s_i\| - \|p - s_j\| = c\Delta t_{ij}$$

Figure 2 illustrates an example scenario. In the ideal case (Figure 2C), there would be no error and we would simply solve the set of six equations for each pair *(i,j)* and derive the position of the tap. In the presence of measurement error, we instead aim to minimize the *squared error*

$$\sum_{i \neq j} (\|p - s_i\| - \|p - s_j\| - c\Delta t_{ij})^2$$

over all possible points p. Figures 2E and 5A show this total squared error plotted over all points near the sensors. Note that this equation depends on knowing the speed of sound in the surface, which is obtained through a calibration process described later.

The next step is to solve this optimization problem (see e.g., [10] for additional background). We employ a gradient descent approach with a small step size to avoid ill-conditioned behavior. This gradient descent takes on average about 100 iterations to converge. On our 2011 MacBook Pro (2GHz Intel Core i7), this calculation takes ~700 microseconds. The output of this optimization process is the point that minimizes the total squared error, i.e. that point which best fits the available data (Figures 2E and 5C)

Overall, the processing performed on the laptop is computationally inexpensive, and thus it would be feasible to implement the necessary calculations directly on the microprocessor (eliminating the computer entirely). We could even imagine using a specially designed ASIC that performs both sampling and calculation, and simply returns the x/y position of the touch, much like contemporary touchscreen controllers.

AUTOMATIC SURFACE CALIBRATION

The propagation speed *c* of the vibro-acoustic wave is required for reliable and accurate operation. This varies with different materials and different frequencies of sound [11,31]. Thus, before operating on a surface, it is necessary to perform a quick calibration. A manual calibration is achieved by tapping next to one of the corners of the device, and recording the time offsets to the other sensors. Because the size of the frame is known (Figures 3 and 4), the propagation speed is trivially calculated. Note that this method only examines the leading edge of the impulse, and thus we obtain only the propagation speed of the fastest frequency component in the medium.

However, relying on users for calibration is not ideal. In response, we briefly experimented with two auto-calibration approaches, which we implemented on our laptop prototype (Figure 4). One approach used a solenoid, which programmatically struck the table (much like a finger would) and allowed for the propagation speed calculation. However, this setup was bulky, required extra components, and had a power draw incompatible with mobile devices.

Our second and superior approach repurposed the piezo sensors as piezo actuators. When applying a high voltage with a transformer-based piezo driver circuit (~90V, but very low current and thus low power), the piezo discs flex, imparting a mechanical impulse to the table's surface. As a proof of concept, we modified our laptop prototype, pulsing one piezo and using the other three sensors for timing information (sensing the leading edge of the impulse using the same implementation as *Toffee).* This approach could be further extended by pulsing all four piezo elements in sequence and averaging the results.

RELATED WORK

There are two significant bodies of work relating to our approach. Conceptually, our approach aligns with other efforts to enable around-device interaction. From a technical standpoint, *Toffee* is a novel application and extension of acoustic TDOA methods, which date back to the 1930's for use in aircraft navigation (see e.g., [5] for a comprehensive history of such systems).

Around Device Interaction

The problem of fat fingers and small devices has long been recognized, and the literature on around-device interaction as a strategy to overcome this has become significant (see [17] for an excellent overview).

Magnetic sensing has been particularly popular, as magnets are unpowered, can be small, and generally avoid issues relating to occlusion and line of sight. Abracadabra [12], MagiTact [19] and similar systems have put forward a range of free-space interactive uses, including radial buttons, in-pocket gestures, and behind device interaction. Electric field [36] and capacitive sensing [32] have similar benefits without the need to wear a magnet, though typically at the cost of reduced spatial resolution.

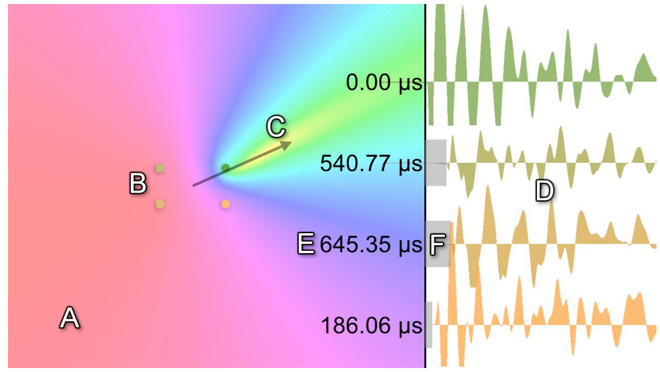

Figure 5. A screenshot of the *Toffee* engine. The multicolored area on the left is a visualization of the total squared error (A). The device with its four sensors (B) is located in the center. Note the error minima visualized as a yellow depression (C) - this is the estimated location of the touch. The waveforms captured by the four sensors are shown on the right (D). In this example, sensor 1 (top most) was triggered first. Note the computed time deltas for the other three sensors (E), also seen as grey periods prefixing the waveforms (F).

There are several optical methods that have proven successful. SideSight [7] and HoverFlow [20] use infrared emitter/receive pairs to sense proximate hands and fingers (which are reflective to infrared light). SideSight in particular is like Toffee in that it utilizes the table as an ad hoc surface for touch input. Cameras can also be used: the Portico system [2] used a pair of stilted cameras that look downward onto an interactive play space. The advent of compact, low-cost depth cameras has recently enabled researchers to consider how they might be used to support around-device interaction in mobile contexts, for example, the PalmSpace system [21].

More similar to *Toffee* are systems relying on passive acoustics. For example, Bonfire [18] used a laptop's accelerometer to detect the vibro-acoustic signal of a touch on a table surface; a separate camera-based system was used to track finger positions. BLUI [26] allows users to blow air at different areas of a laptop screen for interaction, using the built-in microphone to capture the sound of the airflow.

Finally, most similar to our ad hoc approach is Scratch Input [13]. Both systems take advantage of gravity acoustically coupling mobile devices to tables and vibro-acoustics being faithfully preserved inside of solid materials. However, instead of performing multi-sensor TDOA correlation, Scratch Input uses a single microphone and tracks gross variations in amplitude over time to infer a set of basic gestures ([33] uses a similar approach for pen-on-surface classification). Although finger taps were part of Scratch Input's gesture vocabulary, they were not localized, only recognized, offering different interactive capabilities.

Active Acoustic Localization

There are several approaches for localizing interactive events with *active* acoustics. For example, active sonar emits pulses of sound and waits for reflected signals [9]. The Cricket system [30] used a series of ultrasonic beacons fixed in the environment, from which devices could triangulate their location based on relative timing differences. Another example is surface acoustic wave (SAW) touchscreens, which emit pulses of ultrasonic sound. A finger or other object touching the screen dampens the wave locally, which appears as a gap in the signal received by separate acoustic sensors. Using this data, continuous touch tracking is possible.

Passive Acoustic Localization

In addition to the *active* approaches discussed above, there are also several *passive* approaches for localizing signals. The most popular approach, which we utilize, is *time distance of arrival* (TDOA) correlation. This is a well-known technique used in a variety of applications, including the Global Positioning System (GPS) [4] and cell phone tower triangulation [8]. Both of these examples use TDOA analysis on signals in the electromagnetic domain.

More directly related to our approach are systems that employ *acoustic* TDOA analysis [16,23,24,25,34,35]. Most notable among these are the PingPongPlus [16] and Interactive Window [23,24,25] projects. The latter two systems have similar hardware configurations: four contact microphones located at the outermost corners of the desired interactive region. When a Ping-Pong ball or finger touches inside of this region, the signal arrives to the four sensors at different times, enabling a similar all-pairs hyperbolic intersection optimization approach. PingPongPlus uses a linear approximation to the hyperbolic intersection problem to produce an approximated position within the bounds of the microphones. Interactive Window uses a binary search approach to locate the point of minimum error within a predefined range.

Another acoustic approach is time reversal (TR) analysis. When a touch occurs, TR simulates the reverse propagation of the waves received at each sensor. The simulated waves are reflected off the surface boundaries and eventually result in a peak at the original touch point. See [3] for an example system and more in-depth discussion of related work and operational principles. A downside to this approach is that a precise physical model of the surface, including surface boundaries, is required. Due to this significant setup and calibration, ad hoc uses are precluded.

Pham et al. have put forward a series of approaches for performing tap localization on instrumented surfaces. A location template matching approach is employed in [27], which uses a calibrated database of acoustic signatures to identify tap points using a single sensor (see also [6]). Later work by Pham et al. [28] demonstrated a method of decomposing and analyzing TDOA signals using a continuous wavelet transform. In [29], a Kalman filter approach is described that enables continuous tracking of touch inputs when combined with the continuous wavelet transform.

The aforementioned passive acoustic systems differ from *Toffee* in three important respects:

Foremost, the above systems required touch inputs to land *within* the perimeter of the sensors, whereas *Toffee* operates exclusively *outside* of this area (a partial exception to this is the Interactive Window, which offered degraded position sensing when taps occurred just outside the perimeter [23]).

Secondly, the above systems had sensors separated by no less than 20cm, and often on the order of meters. Importantly, the further the sensors are apart, the larger the potential difference in the arrival times, enabling considerably more accurate position sensing. *Toffee* is the first system to demonstrate the ability to scale down to mobile devices, where the technique is potentially most useful.

Third, the above systems permanently instrumented a highly controlled surface (e.g., a glass window, Ping-Pong table). This allowed for tight calibration and excellent (rigid) acoustic coupling of the sensing elements. However, this type of instrumentation and calibration is antithetical to the ad hoc settings we believe makes *Toffee* valuable (e.g., kitchen counter, desk at work, airplane tray table, and café table). *Toffee* is the first system to apply TDOA principles to surfaces in an ad hoc, non-permanent manner for interactive purposes. This means our approach is mobile, whereas previous systems were static. This is made possible in large part due to *Toffee*'s novel, surface auto-calibration method.

Passive Sonar

Functionally, our system most closely resembles passive range-finding sonar systems used in maritime applications [9], though obviously the context, scale and use are significantly different. Passive sonar systems are most frequently employed in military applications, as active sonar is subject to enemy detection. A typical system uses multiple microphones mounted to the vessel exterior in a known configuration (e.g., in a line along the hull – much like the piezos mounted to the underside of our devices). The system calculates time differences by correlating signals (transmitted through an uncontrolled water medium) from each sensor. Using TDOA hyperbolic estimation techniques, sound sources *external* to the submarine can be localized [9,10]. As far as we are aware, *Toffee* is the first system to use these underlying principles for interactive purposes.

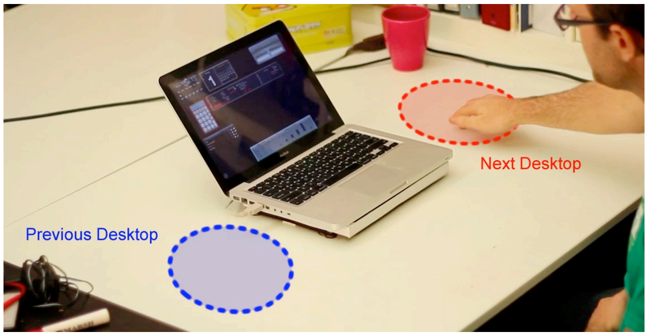

Figure 6. Users can navigate between active desktops by tapping to the left and right of the laptop.

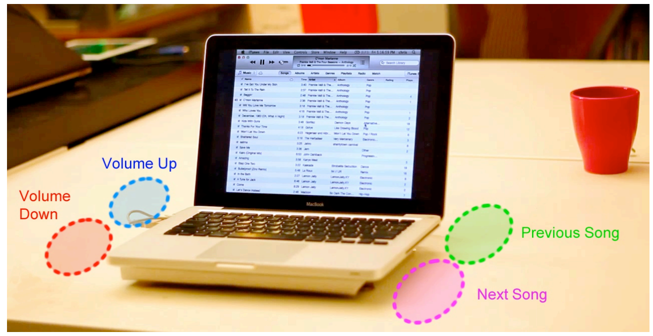

Figure 7. Music player controls can be bound to regions around the laptop, e.g., volume up and down.

EXAMPLE APPLICATIONS

To illustrate some potential uses of *Toffee* – and the surround device interactions it enables – we created proof of concept applications for our three prototype platforms. Please also see the Video Figure.

For our laptop, we created a desktop switching application. Using *Toffee*, areas to the left and right of the keyboard can be tapped to slide left and right, respectively, between active desktops (Figure 6).For a second application, we connected virtual tabletop buttons to the previous/next song and volume up/down controls in iTunes (Figure 7). Users can use these large radial buttons without needing to look at or interact with the screen, enabling inattentive or peripheral use of the interface.

For our tablet prototype, we created a game that spills interaction out onto the surrounding table. At the center of the screen is the player's tank; from all sides, enemies advance towards the player (Figure 8). Tapping on the table orients the tank's turret to the corresponding angle, and fires one projectile. This not only embeds interaction more deeply into the environment, but also opens opportunities for multiplayer experiences.

Finally, on our smartphone, we created an app launcher, where taps above, below, left and right launch the address book, map, calendar and browser respectively (Figure 9).

EVALUATION

We conducted an experimental evaluation to assess the feasibility, accuracy, and ultimately the utility of our approach. We varied a number of experimental parameters to isolate different performance characteristics of our system. In particular, we tested 1) tap angular position, 2) tap distance, 3) tap implement (finger/stylus), and 4) sensor distance (smartphone/laptop).

As an experimental compromise to mitigate a combinatorial explosion of conditions, we selected a single surface for all tests: an unmodified Ikea Galant table (particleboard with wood veneer). This decision was made in part because TDOA has already been shown to work on many surfaces, including glass [23,24,25] and whiteboard [29].

In our first experiment, we sought to understand the angular accuracy of *Toffee*. To do this, we used an overhead digital projector to render a ring around our test device, with rays at 5° intervals (72 angles). A simple interface running on a standalone laptop requested an angle (e.g., 155°), at which time the experimenter tapped the table at the requested angle. The live angular result (e.g., 158°) from *Toffee* was then saved for later analysis. Each angle was requested three times, in randomly interleaved order.

We repeated the above procedure at two different distances (45cm and 65cm), with two different tap types (finger and

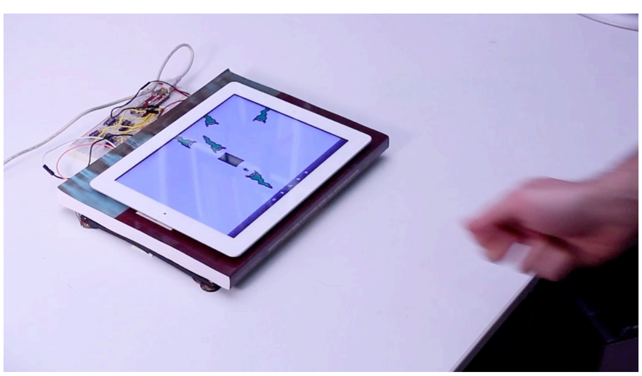

Figure 8. The Tank vs. Aliens game we created for our *Toffee*-augmented tablet. When the player taps the table, the tank fires a shell outwards in the direction of the tap.

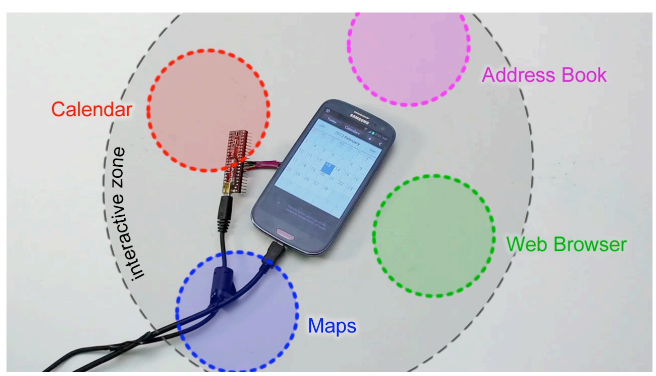

Figure 9. Common applications could be bound to regions immediately surrounding a smartphone (i.e., areas lying within a proximate interactive zone; taps outside this area are attenuated).

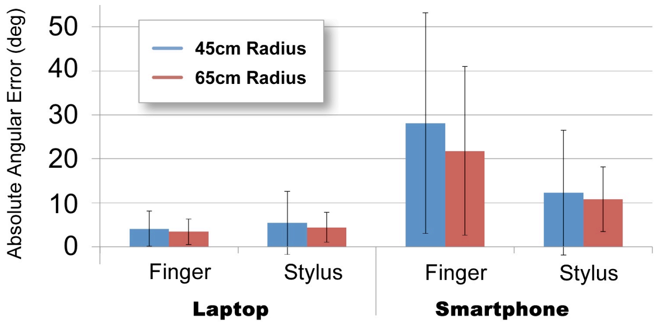

Figure 10. A summary of results in Figure 12. Mean absolute error of estimated angle across eight conditions. Error bars are standard deviation.

stylus), and with two different device sizes (laptop and smartphone), for a total of eight conditions. This yielded 1728 tap trials (eight conditions x 72 unique angles x repeated three times).

The inclusion of two devices of radically different sizes was purposeful. As the distances between the sensors shrink, the timing deltas also shrink. Because our sampling rate is constant, this means our precision decreases as devices get smaller. We also decided to include both finger and stylus taps. These produce significantly different waveforms (fundamental frequencies of around 380Hz and 1400Hz respectively). As described previously, our sensors trigger in response to an exponentially-averaged derivative exceeding a threshold. A higher frequency signal (larger derivative) is more likely to trigger the system accurately.

Our second experiment was designed to understand *Toffee*'s ability to estimate distance. Seven concentric rings were projected around our laptop prototype, starting with a diameter of 36cm, and increasing in 6cm intervals, up to 72cm. To avoid combinatorial explosion, we tested eight angles (0°, 45°, 90°, … 315°). As before, each trial (e.g., 155° @ 42cm) was requested three times to reduce noise in the data, yielding 168 data points. This experiment was run twice, once with finger taps and once with stylus taps, for a total of 336 trials.

RESULTS AND DISCUSSION

Overall, *Toffee* performs well at estimating the angle of touch events. However, distance estimation is too poor for practical use at this time. When compared against existing systems, *Toffee* offers competitive performance, despite using an ad hoc surface.

Angular Accuracy

Combining near and far distances, and finger and stylus conditions, the mean angular error (Figures 10 and 12) for our laptop prototype was 4.3° (mean SD = 4.4°). To put this in context, if we were to build a system with eighteen 20° radial buttons (which allows for a ±10° tolerance in the angle estimation) the system would trigger the correct button 91.8% of the time. If we provided eight 45° buttons (e.g., above / below / left / right and the four diagonals), the system would be 99.1% accurate; four 90° buttons would be 99.9% accurate. Figure 11 illustrates this this accuracy curve based on post hoc analysis of the collected data.

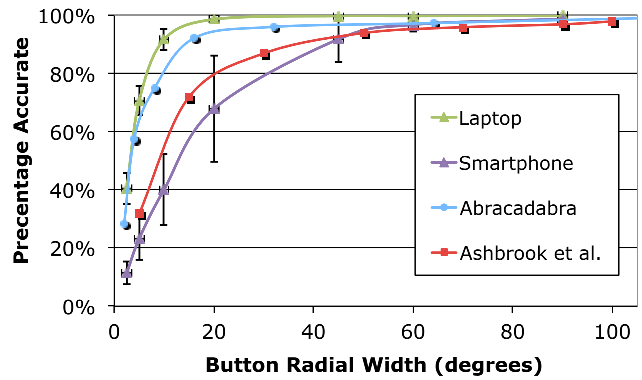

Figure 11. Post hoc simulation of mean input accuracy for different sized angular buttons with our *Toffee*-augmented laptop and smartphone. For comparison, results from two other radial input systems [1,12] are plotted. Error bars illustrate the standard deviation of accuracy across the four sub-conditions for laptop and smartphone trials.

The smartphone performed worse, primarily due to loss of precision resulting from its tightly packed sensors. Combining near/far and finger/stylus conditions, the mean angular error for our smartphone prototype was 18.3° (mean SD = 16.5°; see Figures 10 and 12). Stylus taps performed better than finger taps (mean of 11.6° vs. 25.0° respectively), though not significantly so, primarily due to high variance (SDs of 10.9° and 22.1° respectively).

Framing this again in terms of radial interactors: Eight 45° radial buttons would perform at 72.7% accuracy on the smartphone (Figure 11); four 90° buttons would be 91.9% accurate. To approach 100% accuracy, it would be necessary to use just two buttons, perhaps the areas to the left and right of the device. Although this provides less input bandwidth and interactive options than the laptop, this still opens several interesting opportunities.

The results indicate that *Toffee* is viable out to at least 65cm from the device, which enables an interaction area 1m^2 in size centered over the device. This allows for operation across a substantial portion of a typical desk surface, an area most users would consider "near" the device.

Comparison to Prior Work

We designed our radial input experiment so that our results could be plotted against two notable pieces of prior work, which offer a benchmark (Figure 11). The first system is a capacitive-sensing, direct-touch watch bezel developed by Ashbrook et al. [1]. The second system, Abracadabra [12], is a magnetically driven free-space finger-tracking watch.

The laptop condition of *Toffee* exceeds the radial accuracy of both of these systems. This is particularly notable given that *Toffee* relies on the uncontrolled properties of an ad hoc surface – compared to direct physical contact with a finger [1] and electro-magnetism [12], both of which suffer little interference. The *Toffee*-augmented Smartphone, however, performs worse overall; only when the angular buttons exceed 45 degrees in size are accuracies competitive with the aforementioned systems.

Distance Experiment

Results from our distance estimation experiment were less encouraging. The mean difference between the estimated distance and the true distance (i.e., absolute distance error) for finger trials was 51.9cm (SD=44.1cm). Given that our distances ranged from 36cm to 72cm, this should be considered a failure. Stylus taps are more promising: the mean absolute distance error was 10.2cm (SD=8.6cm). From post hoc analysis, it is clear that the higher frequency waveforms generated by the hard tip of the stylus more reliably trigger the sensors. This leads to tighter time delta estimates, which produces a tighter intersection of the hyperbolas, and ultimately yields a superior estimate of location.

Importantly, moving outward along a ray results in very small changes to the arrival time differences, which is inherently less accurate. This stands in contrast to changing the touch angle, which dramatically alters the relative timing differences among the sensors. Further, this problem is exacerbated at larger distances. The only way to overcome this issue is faster and more precise sensing to obtain more accurate timing estimates.

Robustness

Although we tested accuracy, we did not specifically test robustness outside the lab. However, we did note a few properties of the system during piloting, the study, and application development.

Foremost, the system requires a fairly "sharp" finger or knuckle tap in order to work reliably. Soft taps will fail to trigger the system, particularly as our ADC precision was reduced (to 8 bits) to allow for the faster read rate. In practice, the tap force required is no harder than one would normally knock on a door, i.e. excessive force is not required to operate *Toffee*. However, the system works even with relatively soft stylus taps.

Additionally, because *Toffee* uses only the leading edge of the tap impulse, objects placed on the table do not appear to affect accuracy, provided they are not between the device and the tap location. Indeed, our initial test table had a number of objects resting on it (e.g., books, cups) that were occasionally repositioned without affecting the behavior of *Toffee.* Also, as the leading edge of the acoustic impulse arrives before any reflections, *Toffee* appears fairly robust to multipath acoustic interference.

However, we do note that *Toffee* is subject to false positive input, as with any touch interface. Dropping objects on the table near the device, moving the device or table, or striking the device itself can potentially all produce false positives. In the latter two cases, the effect can be mitigated by rejecting input taps that appear to emanate from within the boundaries of the device. In the former case, the device can be equipped with a lock or sleep feature to ignore inputs when not in use (much like a smartphone is usually locked when in a pocket to avoid accidental input). A classification approach may also be used to distinguish the acoustic signature of a knuckle or finger tap from the sound of a larger object being dropped or placed on the surface [15]. Presently, our only implemented approach is to reject acoustic events that result in excessive total squared error (i.e., a preset threshold).

CONCLUSION

We have presented our work on *Toffee*, a system that allows (small) mobile devices to appropriate (large) tabletops for interaction. This was achieved through time-difference-of-arrival correlation of structural vibro-acoustic signals captured by sensors located at the four corners of a device. Our study results suggest that resolving a true, 2D position (angle and distance) is not sufficiently robust to enable accurate interactive use. However, angle estimation is

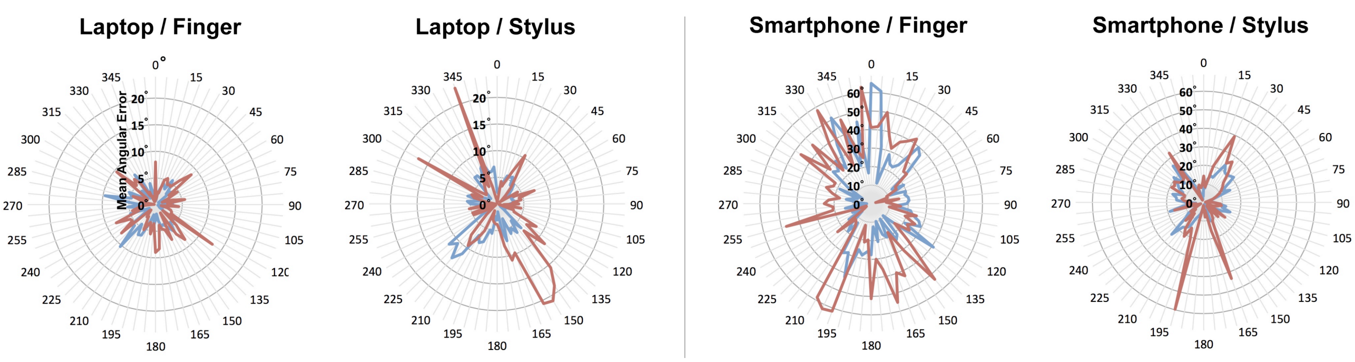

Figure 12. The mean angular error (radial axis) for taps performed in a 360° circumference at two different distances (45 and 65cm). Note the radial axis scale is different for the laptop (20°) and smartphone (60°).

robust, with an average error of 4.3° on a laptop sized setup. Thus, we suggest that interactions should be built around radial interaction, similar to bezel interactions [1] and peripheral free-space gesturing [12]. Our example applications are built using this interaction paradigm, and allow users to make use of the expanded envelope of interactive space surrounding laptops, tablets, and smartphones.

ACKNOWLEDGEMENTS

We would like to thank Bhiksha Raj for encouraging this research as a class project for his Machine Learning for Signal Processing class at Carnegie Mellon University. This work was also supported in part by a Qualcomm Innovation Fellowship, Google Ph.D. Fellowship, Natural Sciences and Engineering Research Council of Canada (NSERC) and NSF grant IIS-1217929.

REFERENCES

1. Ashbrook, D., Lyons, K., and Starner, T. An investigation into round touchscreen wristwatch interaction. In Proc. MobileHCI '08. 311-314.
2. Avrahami, D., Wobbrock, J., and Izadi, S. Portico: tangible interaction on and around a tablet. In *Proc. UIST '11*. 347-356.
3. Bai, M. R. and Tsai Y. K. Impact localization combined with haptic feedback for touch panel applications based on the time-reversal approach. *Journal of the Acoustical Society of America,* vol. 129, no. 3, pp. 1297-1305, March 2011.
4. Bancroft, S. An algebraic solution of the GPS equations. *IEEE Trans. Aerospace and Electronic Systems*, vol. AES-21, pp. 56-59, Jan. 1985.
5. Blanchard, W. F. Hyperbolic Airborne Radio Navigation Aids – A Navigator's View of their History and Development. *The Journal Of Navigation*, vol. 44, no. 3, pp. 285-315, Sept 1991.
6. Bornand, C., Camurri, A., Castellano, G., Catheline, S., Crevoisier, A., Roesch, E. B., Scherer, K. R., and Volpe, G. Usability evaluation and comparison of prototypes of tangible acoustic interfaces. In P*roc. ENACTIVE '05*.
7. Butler, A., Izadi, S., and Hodges, S. SideSight: multi-"touch" interaction around small devices. In *Proc. UIST '08*. 201-204.
8. Caffery, J. and Stuber, G. L. Overview of radiolocation in CDMA cellular systems. *IEEE Commun. Mag.*, 36(4), pp. 38-45, Apr. 1998.
9. Carter, G. C. Time delay estimation for passive sonar signal processing. *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. ASSP-29, pp. 463-470, June 1981.
10. Chan, Y. T. A Simple and Efficient Estimator for Hyperbolic Location. *IEEE Trans. on Signal Processing*, vol. 42, no. 8, pp. 1905-1915, August 1994.
11. Fahy, F. and Gardonio, P. (2007). Sound and Structural Vibration, 2nd Ed. Academic Press. Oxford, UK.
12. Harrison, C. and Hudson, S. Abracadabra: wireless, high-precision, and unpowered finger input for very small mobile devices. In *Proc. UIST '09*. 121-124.
13. Harrison, C. and Hudson, S. Scratch input: creating large, inexpensive, unpowered and mobile finger input surfaces. In *Proc. UIST '08*. 205-208.
14. Harrison, C. Appropriated Interaction Surfaces. *IEEE Computer Magazine*, June 2010, 43(6). 86-89
15. Harrison, C., Schwarz, J., and Hudson, S.E. TapSense: enhancing finger interaction on touch surfaces. In Proc. UIST '11, 627-636.
16. Ishii, H., Wisneski, C., Orbanes, J., Chun, C., and Paradiso, J. PingPongPlus: design of an athletic-tangible interface for computer-supported cooperative play. In *Proc. CHI '99*. 394-401.
17. Jones, B., Sodhi, R., Forsyth, D., Bailey, B. and Maciocci, G. Around device interaction for multiscale navigation. In *Proc. MobileHCI '12*. 83-92.
18. Kane, S., Avrahami, D., Wobbrock, J., Harrison, B., Rea, A., Philipose, M., and LaMarca, A. Bonfire: a nomadic system for hybrid laptop-tabletop interaction. In *Proc. UIST '09*. 129-138.
19. Ketabdar, H., Roshandel, M. and Yüksel, K. Towards using embedded magnetic field sensor for around mobile device 3D interaction. In *Proc. MobileHCI '10*. 153-156.
20. Kratz, S. and Rohs, M. HoverFlow: expanding the design space of around-device interaction. In *Proc. MobileHCI '09*. Article 4, 8 pages.
21. Kratz, S., Rohs, M., Guse, D., Müller, J., Bailly, G. and Nischt, M. PalmSpace: continuous around-device gestures vs. multitouch for 3D rotation tasks on mobile devices. In *Proc. AVI '12*. 181-188.
22. LeafLabs. Maple Mini. http://leaflabs.com/devices
23. Leo, C. K. Contact and Free-Gesture Tracking for Large Interactive Surfaces. MEng Thesis, MIT Dept. of EECS and MIT Media Lab, May 2002.
24. Paradiso, J. and Leo, C. Tracking and Characterizing Knocks Atop Large Interactive Displays. *Sensor Review*, vol. 25, no. 2, pp. 134-143, 2005.
25. Paradiso, J., Leo, C., Checka, N., and Hsiao, K. Passive acoustic sensing for tracking knocks atop large interactive displays. In *Proc. IEEE Sensors '02*. 521-527.

26. Patel, S. N. and Abowd, G. D. Blui: low-cost localized blowable user interfaces. In *Proc. UIST '07*. 217-220.

27. Pham, D.T., Al-Kutubi, M., Ji, Z., Yang, M., Wang, Z. and Catheline, S. Tangible acoustic interface approaches. In *Proc. IPROMS '05*. 497–502.

28. Pham, D.T., Ji, Z., Peyroutet, O., Yang, M., Wang, Z. and Al-Kutubi, M. Localisation of impacts on solid objects using the wavelet transform and maximum likelihood estimation. In *Proc. IPROMS '06*. 541–547.

29. Pham, D. T., Ji, Z., Yang, M., Wang, Z., and Al-Kutubi, M. A novel human-computer interface based on passive acoustic localisation. In *Proc. HCI '07*. 901-909.

30. Priyantha, N. B. Chakraborty, A., and Balakrishnan, H. The Cricket Location-Support system. In *Proc. MOBICOM '00*. 32-43.

31. Prosser, W. H., Gorman, M. R., and Humes, D. H. Acoustic emission signals in thin plates produced by impact damage. *Journal of Acoustic Emission*, vol. 17, no. 1–2, pp. 29–36, 1999.

32. Sato, M., Poupyrev, I., and Harrison, C. Touché: enhancing touch interaction on humans, screens, liquids, and everyday objects. In *Proc. CHI '12*. 483-492.

33. Seniuk, A. and Blostein, D. Pen Acoustic Emissions for Text and Gesture Recognition. In *Proc. ICDAR '09*. 872-876.

34. Urashima, A. and Toriyama, T. Ubiquitous Character Input Device Using Multiple Acoustic Sensors on a Flat Surface. In *Proc. ICAT '10*. 39-43.

35. Yang, M. Al-Kutubi, M., and Pham, D. T. In-Solid Acoustic Source Localization Using Likelihood Mapping Algorithm. *Open Journal of Acoustics*, vol. 1, no. 2, pp. 34-40, Sept. 2011.

36. Zimmerman, T., Smith, J., Paradiso, J., Allport, D., and Gershenfeld, N. Applying electric field sensing to human-computer interfaces. In *Proc. CHI '95*. 280-287.