Augmenting Touch Interaction Through Acoustic Sensing

Pedro Lopes *Ricardo Jota* *Joaquim A. Jorge*
INESC-ID, IST, Technical University of Lisbon
Rua Alves Redol, 9, 1000-029 Lisboa, Portugal
{pedro.lopes,jotacosta}@ist.utl.pt, jaj@inesc.pt

ABSTRACT

Recognizing how a person actually touches a surface has generated a strong interest within the interactive surfaces community. Although we agree that touch is the main source of information, unless other cues are accounted for, user intention might not be accurately recognized. We propose to expand the expressiveness of touch interfaces by augmenting touch with acoustic sensing. In our vision, users can naturally express different actions by touching the surface with different body parts, such as fingers, knuckles, fingernails, punches, and so forth - not always distinguishable by touch technologies but recognized by acoustic sensing. Our contribution is the integration of touch and sound to expand the input language of surface interaction.

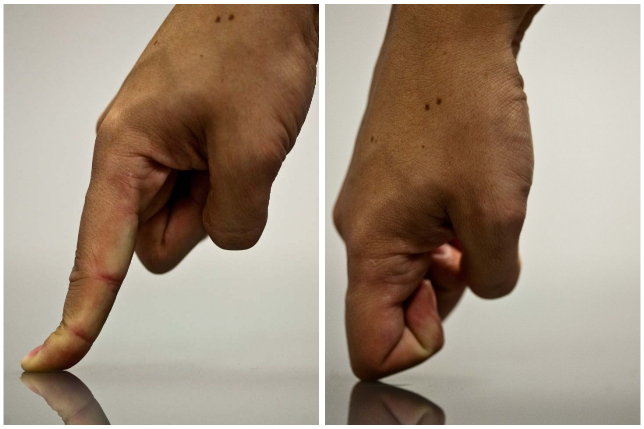
ACM Classification: H5.2 [Information interfaces and presentation]: User Interfaces - Input devices and strategies

General terms: Design, Human Factors

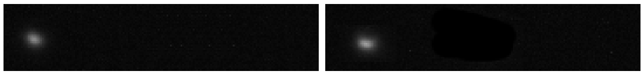
Keywords: Touch, Acoustics, Interaction

INTRODUCTION

The introduction of touch technology strongly influenced direct input interfaces. Indeed, in certain scenarios, input devices such as mouse or pens are being replaced by touch, which now supports multiple fingers [2] and gesture recognition [10]. While diversifying the input language, these advances are limited to the hand position and shape. We argue that user intention cannot be fully understood, if touch location and shape are the only cues captured. The action of touching a surface generates a contact that can be sensed with available technology, but it also generates sound, which can be sensed using acoustic analysis.



(a) Finger Touch (b) Knuckle Touch



(c) Finger Contact (d) Knuckle Contact

Figure 1: Most technologies cannot distinguish between contacts with similar signatures.

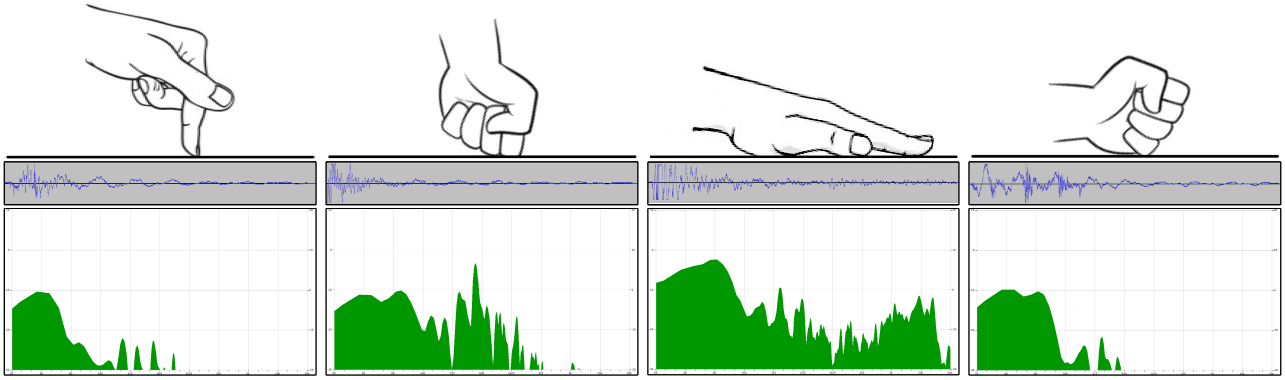
While acoustic sensing has been previously researched as an input technique [3, 6], its potential has not been fully explored in the context of surface interaction. We propose a sonically enhanced touch recognition: the integration of touch and sound. Thus, touches that have identical touch signatures (see Figure 1), can nevertheless be distinguished by their acoustic characteristics: intensity and timbre. The contribution of our research is not to replace available touch technology, but rather to provide additional cues to be integrated in current technologies. Thus, expanding the input language of multi-touch technologies with acoustic gestures, and extending the interaction space to surrounding regions, such as bezel and casing, that commonly do not have any sensing technology.

RELATED WORK

To expand the input language, baseline technologies [2] have been complemented with alternatives cues, including: hand shape [10], pose estimation [4], and muscle sensing [1].

Sound has been previously proposed as the sole interaction modality with surfaces. Robinson [8] provides an additional input to mobile devices by acoustically detecting finger taps in the back of the device. Paradiso proposes a large-scale surface that infers touch location [6] and distinguishes gestures by their acoustic signature (timbre), such as knocks, metal taps, or fist bashes (which we denote as punch). However, the system solely relies on acoustic sensing to detect touch location, not integrating capacitive or optical sensing. Murray-Smith and Harrison focus on recognizing continuous gestures, by analyzing the amplitude envelope (intensity) of

Conference ITS'11, November 13-16, Kobe, Japan.



(a) Finger (b) Knock (c) Slap (d) Punch

Figure 2: Gestures amplitude envelope (blue graphic) and frequency distribution (green graphic).

surface gestures. In Stane [5], a hand-held device is controlled by rubbing motions on a rugged surface; whereas in Scratch input [3] existing surfaces are augmented with acoustic sensing to support finger interaction. In TouchLight [9], Wilson captures the impact intensity of the surface contact to distinguish a tap from a knock, the latter is recognized by a peak above a certain amplitude threshold. However, this does not address the possibility of using timbre, as a cue to distinguish different gestures.

We distinguish from previous work by, not only looking at how the integration of touch and acoustic cues (timbre and intensity) can help identify finger parts and tangibles, but also by further expanding interaction to regions where touch is not available, such as bezels or the side of the table, and proposing interactions that use different impact amplitudes to convey separate meanings to the same gesture.

SONICALLY ENHANCED TOUCH

Most touch capture systems only provide two dimensions: position and area of contact. As illustrated in Figure 1, two touches, with distinct body parts such as knuckles and fingertips, can exhibit equivalent contact areas and, therefore, become indistinguishable. However, they have different acoustic signatures; in this case, the acoustic difference is audible to the naked ear because bone is significantly harder than the fingertip. From our understanding, touch produces two main sonic characteristics: attack amplitude (the intensity of impact) and spectral quality (the sound timbre). Thus, current touch cues may be sonically enhanced with two new dimensions to improve upon the range of surface gestures.

We characterize gestures based on their acoustic cues, as depicted in Figure 2. For example: the finger tap in Figure 2(a) exhibits a soft amplitude envelope and a fundamental frequency around 38Hz with partials in 120Hz and 940Hz, while a knock gesture (Figure 2(b)) portraits a higher peak, a fundamental frequency around 40Hz but with strong partials in 110Hz and 500Hz up to 5KHz that distinguish it from a finger tap. This evidence supports our vision to provide a richer input language, by augmenting gesture classification with impact intensity and timbre plus location and shape.

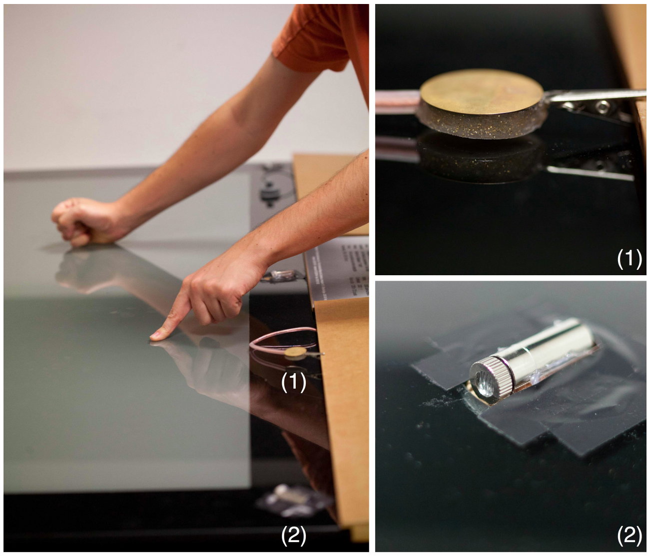


Figure 3: Tabletop with LLP technology (bottom-right) and contact microphone (top-right).

IMPLEMENTATION DETAILS

Acoustic cues require the use of microphones to capture the sound signature of gestures. To capture vibrations, caused by users' interaction, we use a contact microphone placed against the surface. These small components can be placed above, below, or on the corners of the surface, with no visual interference. Alternatively, they can be positioned either on the side, or on the bezels of touch-screens. Remarkably, some touch devices such as tablets or mobile phones, already have a built-in microphone that can be used to recognize gestures with acceptable quality.

Figure 3 depicts our tabletop prototype, including the sound recognition setup. The top is a 1.5 x 0.75m, 15mm thick, glass surface with a 150mm border. Multi-touch is sensed with an optical Laser Light Plane (LLP), shown in Figure 3 as the grey region. The contact microphone was installed outside the multi-touch enabled region, but still touching the glass surface (Figure 3, top-right).

Touch Recognition

The gesture recognition module is implemented in Pure Data, which is cross-platform and available on mobile devices. The workflow of audio processing is as follows: (1) the contact microphone captures the audio signal, sampled at 44.1kHz; (2) a noise reduction notch filter is applied to preserve signal quality, i.e., most optical multi-touch setups include noisy components such as projectors or computers that can be easily filtered out; (3) the signal is analyzed through an envelope follower, which measures peak amplitude (in dBs) - which we denote as the **intensity cue**; (4) to determine the spectral signature, chunks of 256 samples (or 5.8 ms of audio at 44.1kHz) are passed through 11 narrow band filters (a FIR filter bank, with two filters per octave, as suggested by [7]). Then, a Discrete Fourier Transformation (DFT) calculates the magnitude for each filter. The spectral signature, that we denote as the **timbre cue**, is stored as a vector (11 elements, one per filter) of those magnitudes; finally, (5) the spectral signature is compared against a database of trained gestures using the method described in [7]. If a match is found, an event with the gesture type (e.g., "fingernail"), intensity (dBs), and timestamp is issued. The timestamp is used to match touch data (location) with the aforementioned acoustic cues. If no touch events are reported by the optical tracker within a certain timeframe, the event is considered a touch in the non-sensitive areas (bezel, casing). Moreover, touch data with no associated acoustic signature is considered a false positive and discarded.

This method allows for fast recognition of attacks and spectral signature comparisons; providing, on our prototype, latencies on par with the optical touch recognition. For the recognizer, our databases were comprised of 10 templates per gesture, e.g., a total of 40 if we intended to simultaneously recognize: finger taps, fingernail taps, punches, and slaps.

DISCUSSION

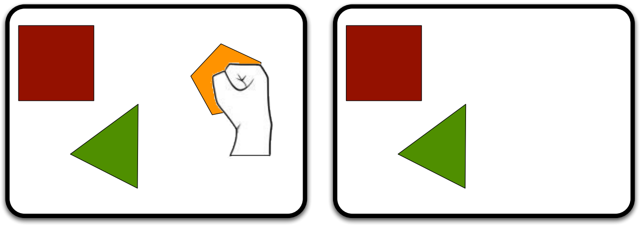
In this section, we present several interaction design opportunities that benefit from the inclusion of acoustic sensing.

Implicit Semantics

To perform complex actions (e.g., copy-paste, group, ungroup, delete), most touch-based applications rely on complex GUI elements, such as contextual menus, that are far from ideal for touch interaction. By increasing the number of available gestures, we can refrain from using GUI elements. Each sonic enhanced gesture can be associated with a meaning (semantic) that corresponds to the gesture affordance. For instance, punch can delete an object, as depicted in Figure 4(a). This allows the already familiar multi-touch lexicon to remain available for direct manipulation of objects (move, scale, rotate, and so forth). Another example is the cloning objects that can be performed bimanually: one finger points at the target object and a knock gesture clones the object onto the desired location, as depicted in Figure 4(c).

Gesture Intention

Distinct intentions may exhibit different attacks, i.e., a strong and soft punches may convey unique meanings. Intention can be valuable to interaction design, because this dimension gains more resolution that just two states (on/off). This technique can be used to create situations where the same gesture results in different actions, accordingly to the impact amplitude, e.g., a soft knuckle tap cracks an object, while a strong knuckle tap destroys it.


(a) Delete (b) Object Deleted

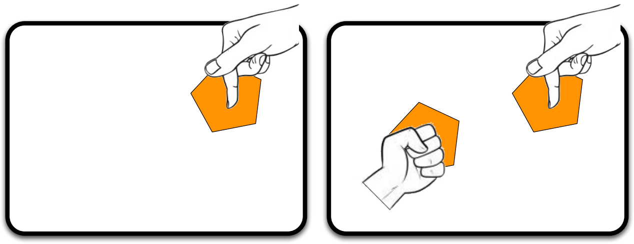

(c) Copy (d) Paste

Figure 4: Example of Implicit Semantics.

Beyond Touch: outside the touch area and tangibles

As sound propagates quite easily through most surfaces, including those that are used in tabletops casing or bezels, interactions can be made available outside the touch area. Therefore, one can devise scenarios where gesture recognition is strictly based on sonic cues. For example, a user could rotate the whole tabletop visualization towards his position by hitting the side of the case; another example is a punch on one of the bezels to trigger all objects to move towards that side of the table.

Timbre recognition enables tangible objects to come into play without the need for optical tracking. This widens the possibilities of interaction, allowing us to capture the acoustic signature of everyday objects, such as mobile phones, wooden blocks, taps with metal rings or other non-tracked artifacts.

Improving Touch Technologies

Acoustics can filter out false positive touch positions, typically generated by natural or artificial lighting that interferes with IR setups. In these cases, all touch data not accompanied by a peak in the incoming audio, can be classified as noise and discarded.

Furthermore, the combination of touch data with acoustic cues can improve accidental activation avoidance, e.g., resting hands and arms on the touch sensitive area without raising touch events (often denoted as palm rejection). For this, the system designer must provide a empirical threshold that separates accidental activations from intentional touches.

Limitations of Acoustic Sensing

Interactions that are solely based on acoustic cues, such as touching the bezels or tabletop casing, are more prone to accidental activation, e.g., users might bump the side of the tabletop with their knees. We advise designers to only consider gestures that portray highly expressive and intentional acoustic cues (e.g., strong knocks), to reduce false positives in these secondary regions.

Currently, our prototype does not account for the recognition of simultaneous acoustic gestures, e.g., if a user knocks and punches at the same time. When two simultaneous gestures occur (by one or multiple users) the microphone captures the sum of both acoustic signatures, which cannot be trivially separated in order to recognize the different gestures.

Throughout our experiments (surface size of 1.5 x 0.75 m) we did not perceive variations associated to the distance between touch and microphone locations. However, for large scale interactive displays, the position of the microphone can cause discrepancies in the recognizer results, since sound energy dissipates as it propagates through the surface. Also, the surface material may affect the quality of the signal. We recommend the adoption of rigid interactive surfaces, preferably glass (which propagates sound easily [6]).

Even with noise reduction, noisy environments can still affect the quality of the input signal. To minimize this issue, we recommend the use of contact microphones as opposed to common condensers. Contact microphones are more adequate for this purpose because they react to surface vibration rather than to outside interference.

CONCLUSIONS AND FUTURE WORK

We added a new dimension, sound, to interactive surfaces and presented acoustically enhanced touch sensing. This dimension can provide more expression to user defined gestures. Our solution allows more interactions with touch surfaces, such as detecting gestures on the bezel or side, distinguishing more hand parts (e.g., knuckles) or tapping with objects (that also produce a specific audio signature). Furthermore, combining sound and touch data is an inexpensive approach to eliminate false positives in optical multi-touch setups.

As future work, we foresee three promising directions: explore fine-grained details of sound gestures, by capturing small vibrations caused by friction motion; separation of concurrent gestures through complex spectral analysis; and clarify which hardware setups (DI, FTIR, capacitive, and so forth) yield better characteristics for combining touch with acoustic cues.

Finally, we observe that nowadays, many devices (such as tablet computers or mobile phones) include built-in microphones, thus benefiting from our framework of sonically enhanced gestures to provide better interactions at no extra cost.

ACKNOWLEDGMENTS

This research was partially supported by FCT grant MIVis (PTDC/EIA-EIA/104031/2008), and through the PIDDAC via INESC-ID multi-annual funding. We acknowledge the early review of Yoichi Sato.

References

1. Hrvoje Benko, T. Scott Saponas, Dan Morris, and Desney Tan. Enhancing input on and above the interactive surface with muscle sensing. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*, ITS '09, pages 93–100, New York, NY, USA, 2009. ACM.
2. Jefferson Y. Han. Low-cost multi-touch sensing through frustrated total internal reflection. In *Proceedings of the 18th annual ACM symposium on User interface software and technology*, UIST '05, pages 115–118. ACM, 2005.
3. Chris Harrison and Scott E. Hudson. Scratch input: creating large, inexpensive, unpowered and mobile finger input surfaces. In *UIST '08: Proceedings of the 21st annual ACM symposium on User interface software and technology*, pages 205–208. ACM, 2008.
4. Nicolai Marquardt, Johannes Kiemer, and Saul Greenberg. What caused that touch?: expressive interaction with a surface through fiduciary-tagged gloves. In *ACM International Conference on Interactive Tabletops and Surfaces*, ITS '10, pages 139–142. ACM, 2010.
5. Roderick Murray-Smith, John Williamson, Stephen Hughes, and Torben Quaade. Stane: synthesized surfaces for tactile input. In *Proceeding of the 26th annual SIGCHI conference on Human factors in computing systems*, CHI '08, pages 1299–1302. ACM, 2008.
6. Joseph A. Paradiso, Che King Leo, Nisha Checka, and Kaijen Hsiao. Passive acoustic sensing for tracking knocks atop large interactive displays. In *Proceedings of the 2002 IEEE International Conference on Sensors*, pages 11–14, 2002.
7. M. Puckette and T. Apel. Real-time audio analysis tools for pd and msp. *Proceedings of the 24th ICMC, International Computer Music Conference, Univ. of Michigan, Ann Arbor, USA*, pages 109–112, 1998.
8. Simon Robinson, Nitendra Rajput, Matt Jones, Anupam Jain, Shrey Sahay, and Amit Nanavati. Tapback: towards richer mobile interfaces in impoverished contexts. In *Proceedings of the 2011 annual conference on Human factors in computing systems*, CHI '11, pages 2733–2736, New York, NY, USA, 2011. ACM.
9. Andrew D. Wilson. Touchlight: an imaging touch screen and display for gesture-based interaction. In *Proceedings of the 6th international conference on Multimodal interfaces*, ICMI '04, pages 69–76, New York, NY, USA, 2004. ACM.
10. Mike Wu and Ravin Balakrishnan. Multi-finger and whole hand gestural interaction techniques for multi-user tabletop displays. In *Proceedings of the 16th annual ACM symposium on User interface software and technology*, UIST '03, pages 193–202. ACM, 2003.