

## Tracking Contact and Free Gesture Across Large Interactive Surfaces

Joseph A. Paradiso  
Responsive Environments Group  
MIT Media Lab  
joep@media.mit.edu

Large projection displays and video walls are already common in public spaces such as shopping malls and airports. As the enabling technologies continue to advance and decrease in price, these devices will become even more popular. At the moment, however, such displays are mainly noninteractive, and merely play uninterrupted video streams. When they are made responsive, however, they open up entirely new types of group interactions, in contrast with video kiosks, their smaller, presently ubiquitous cousins that deal mainly with single users.

User interaction with large displays is a topic of considerable interest in the CHI and ubiquitous computing communities [1,2,3], where current research is exploring ways in which the user interface is distributed between various portals (e.g., handhelds, mobile and wearable devices, and large interactive surfaces) in responsive environments and augmented rooms. Many applications have been explored in professional niches like electronic blackboards for presentation, audiovisual portals for teleconferencing, augmented business and office environments, large electronic bulletin boards in corporate “water cooler” settings, interactive visualizations for design studios, and big-board displays for military and situation rooms. In contrast, the majority of the implementations introduced in this article are directed at public settings, where they are used for casual information browsing, interactive retail, and artistic installations or entertainment. Because their activity tends to be highly visible, participants at public interactive walls often become performers. These systems are intrinsically collaborative - crowds tend to gather around to watch, participate, and suggest choices as a user interacts with a large display; essentially all applications attain a social, gamelike quality.

Although there are several products available that identify and track objects accurately across large electronic whiteboards and tablets, in order to be usable in public settings, it is important that such interactive walls respond to bare hands and do not require the user to wear any kind of active or passive target. At the moment, there are several sensing and tracking approaches that have been used to make large surfaces bare-hand interactive, many of which are introduced in [4]. The majority of these (e.g., capacitive sensing, resistive sandwiches, light curtains, active acoustics) are derived from touch screen technology [5]; while others are based on video tracking [6]. Most do not scale well to very large surfaces, however, or involve significant complication and robustness issues, especially in unstructured public or outdoor installations.

The Responsive Environments Group at the MIT Media Lab has developed several relatively simple techniques to track activity across large surfaces [4]. All are essentially retrofits, as they do not require the installation of custom-designed material or any significant infrastructure. The first of these projects, the *Gesture Wall*, was an interactive music installation designed in 1996 for the *Brain Opera* [7], a large touring interactive media production currently installed at the *Haus der Musik* museum (<http://www.hdm.at/>) in Vienna. Here, an array of four pickup electrodes placed at the

corners of a projection screen, receives a signal capacitively coupled from the body of a participant standing atop an antenna transmitting a 50 kHz electric field. The amplitude sensed at each pickup, reflecting the proximity of the body to the corresponding receive electrode, is processed to determine a mean position, which is used by a rule base that produces interactive music and graphics. Although this system responded well to bulk gesture, its tracking accuracy varied widely with the posture of the participant, limiting its use to abstract kinetic expression of the sort exploited by the *Gesture Wall* installation.

Our next system [4] employed a scanning laser rangefinder placed at a corner of a large projection screen, where it monitored a plane just above the projection surface. As commercial rangefinders were prohibitively expensive for this project, we designed our own relatively low-cost, continuous phase-shift device that could track bare hands with roughly 1 cm accuracy out to circa 4 meters at 30 Hz. As the laser illumination was synchronously detected, this device was insensitive to background light and accurate enough for detailed, causal interaction. This system was used in interactive music applications and graphical database interfaces shown at the *Emerging Technologies* exhibitions during SIGGRAPH 98 and SIGGRAPH 2000. Despite its success in these trials, this technique requires the electromechanical scanning rangefinder to be mounted in a corner at the front of the display, potentially limiting its application, especially for outdoor settings.

Our next system, diagrammed in Figure 1, is an extremely simple retrofit to a large display mounted inside a single-paned window. Its technical roots came from the ball impact tracker designed for the PingPongPlus [8] interactive ping-pong table, and its inspiration was the desire for a simple system that enabled taps on glass walls to annoy the nearby denizens of a virtual fish tank. Four piezoceramic contact microphones are glued to the inside corners of a large glass window. Their resultant signals are monitored by a low-cost Digital Signal Processor (DSP) that produces features analyzed by a connected PC, which generates content that can be projected onto a screen or video wall placed behind the window. When a participant knocks on the glass, flexural or bending waves [9] travel from the point of impact to the microphones. By measuring the wavefront's differential time of arrival at each location, one can infer the location of the originating impact. Although lower-accuracy coordinates can be determined when knocking outside of the square framed by the sensors, most of our applications have concentrated the interactivity and projection well within this boundary, where occlusion by the opaque, 3-cm diameter sensors and their associated cable is not an issue.

Bending waves are highly dispersive, however, hence the impulse waveform launched by the knock tends to slide apart as it propagates through the glass, making straightforward determination of its rising edge (hence time reference) difficult. In addition, there are many ways to hit glass (e.g., knock with a knuckle, fist, or metal ring for example), all producing widely varying waveforms with differing frequency content, hence different propagation velocities. Rather than try to constrain the type of knock required (essentially impossible in a public installation), we first classify the impact, then process the data with a heuristically-guided cross-correlation and edge-detection procedure [10]. The resultant system has been seen to locate knocks across 2 meter windows with resolutions of  $\sigma = 2.5$  cm in 5 mm glass and up to  $\sigma = 4$  cm in 1 cm glass [11]. Although this hardly warrants as a precision pointing device, this resolution is adequate for the relatively coarse selection needed by the applications to which it was

aimed (see below). The system was able to produce hit coordinates within 65 ms (dominated by waveform processing in the 26 MIPS DSP that was used), granting essentially real-time performance. In addition to deriving the location of the hit, the impact intensity is estimated and the type of hit is determined (e.g., knuckle knock, fist bash, or tap from a metallic object), hence the content is able to react in a more sophisticated manner than a simple touch screen response by incorporating some measure of the user's affect, especially relevant here since knocks tend to be fairly expressive gestures.

A recently announced commercial system [12] appears to operate on a similar principal; it seems, however, to exploit the ultrasound component in hard fingernail taps, hence requires a scripted "finger flick" gesture. As mentioned above, our device responds to and classifies any kind of knock, an important feature for running such a system in unattended public venues, where it is difficult to constrain user gesture.

Figure 1 shows a couple of other sensors that can be used by the system. Since the particular piezoelectric pickups and attached amplifiers that we mounted in the window's corners don't respond well to the very low frequencies present in a fist-bash, the superior low-frequency sensitivity of an attached electrodynamic pickup can be exploited to easily discriminate bash events [11]. Because of the poor impedance match to sounds not produced in the glass itself, the adhered pickups are quite insensitive to extraneous audio. Exceptions can occur, however, for loud, sharp sounds (like handclaps) produced near the pickups. In this case, a "veto" microphone can be placed in the air near the window. Signals that induce a strong response both in the glass and in the veto microphone are then rejected as background, and don't produce false hits.

An important feature of this approach is that all sensors are mounted on the inside of the glass. Nothing is attached to the outside surface. This is especially relevant for outdoor installations, where no hardware need be mounted externally on single-paned windows. No significant tracking distortion has been noticed when running this system on a window with outside conditions ranging from room temperature to below freezing – as four pickups overdetermine the position estimate, the system can self-compensate bulk changes in the wavefront propagation velocity. Depending on the glass pane's damping characteristics, multiple hits can be independently registered within a short time (e.g., 100 ms) of each other, allowing the system to be used by several people. For closer intervals, or in cases where the window has a long ringdown response, the later hit is ignored, and as the strikes approach simultaneity, the data from the four sensors becomes inconsistent and the hits are generally rejected.

We used this system for simple, in-house demonstrations in 1999 [4], and developed it sufficiently for formal installations by 2001. Our first applications were in the realm of interactive art. Figure 2 (top) shows a semi-permanent installation running at the *Ars Electronica Center* in Linz, Austria. Called the *Responsive Window*, it is an interactive drawing program written by Ben Fry of the Media Lab's Aesthetics and Computation Group, where the user extrudes rotating objects by knocking on a 1 cm sheet of plate glass backed by a holographic projection screen. Figure 2 (middle) shows *Telephone Story*, an installation run at New York's Kitchen Gallery, where a user selects a video clip to launch (shot by Bay Area artist J.D. Beltran) by knocking on a particular region of a projected desktop. If the user knocks on the screen while the video is running, an image relevant to the current segment of the video appears at the knock

position, rotating faster with harder knocks. Bashes cause a group of images to appear at the fist position and fly off to the edges of the screen. We last ran the system on a large window (2 x 2 meters, 0.5-cm glass) at the *Emerging Technologies* exhibition during SIGGRAPH 2002. The content was based around a complex visualization called *Weather*, a behavior-driven environment written by Marc Downie of the Media Lab's Synthetic Characters Group that evolved in intricate ways with each knock, as shown in Figure 2 (bottom). A pair of very low power 2.4 GHz Doppler motion sensing radars mounted behind the screen detected people moving in front. The radars, modified versions of those introduced in [7], have an onboard processor that extracts three features corresponding to the net amount of motion, the mean speed, and the average direction of motion for the objects in their field of view. Although their spatial discrimination is quite coarse, these sensors are immune to changes in light conditions or optical characteristics of cloth – unlike video imagers, they see directly through nonconductive walls and penetrate clothing, sensing the skin directly. The radars accordingly open up a degree of noncontact interaction as people approach the wall – in this case, motion in front of the screen generated global, nonspecific behavior (e.g., rolling, scrolling, boiling effects) in the graphics in accordance with the motion characteristics. Knocking created more specific and highly localized phenomena.

After gaining experience with the system in museum installations, we collaborated with one of our Media Lab sponsor companies, American Greetings, in a retail application, installing our system on a large window in their New York City store at Rockefeller Center from December 2001 through February 2002, spanning their peak periods of Christmas and Valentine's Day. Figure 3 shows the setup in operation, with random passers-by interacting. A small speaker mounted outside provided audio prompts and narration for the interaction, otherwise all hardware (transducers, electronics, and the holographic projection screen) was inside the 1-cm thick window. The pickups were mounted at the corners of a 2-meter square, placed well away from the projection screen to avoid any user distraction. The interactive content was fairly straightforward – users could choose to watch either of two brief video clips or engage in a game of “Three Card Monty”, where, after three successful rounds of knocking on the correct card image, they would be invited to enter the store and receive a free greeting card. The game was a ploy to get people into the store; indeed, their data indicated a significant increase in store traffic when the system was running.

Large interactive surfaces in public spaces enable interesting applications where games and practicality converge. They are intrinsically communal, encouraging people to converge and collaborate. A very simple system has been described that we have used to make single-pane windows, common features in any city, into a large tracking surface, enabling large interactive displays that let passersby explore content at venues ranging from storefronts to museums. The system requires people to knock relatively lightly on the glass – a common gesture, but one that is still unusual for digital interaction. Our installations have found that once people are invited to knock (e.g., either through audio prompts, visual suggestion, or by example from observing others), they take to this interface quite easily, at least until their knuckles fatigue after scores of hits. With the inclusion of noncontact sensing and ranging away from the plane of the display, such systems can detect people approaching and vary their resolution or adapt their content with proximity of the prospective user.

Although the applications described here all involve close-up interaction with a large dynamic display, this tracking system is appropriate for other niches, e.g., for selecting objects placed behind a glass partition. This would enable, for example, interactive museum cases, where knocking near an object brings up an audio stream of related information. One could similarly use this system in a vending machine, where knocking atop a desired snack causes it to be delivered. Current implementations that use keypads are very indirect and often error-prone (think how often we've spent our last bit of change on the wrong candy bar), and nearby buttons on a museum display case often ruin the aesthetic, especially when compared to an unbroken, knock-sensitive surface.

## References

- 1) See: <http://ipsi.fraunhofer.de/ambiente/collabtablewallws/papers/index.html>
- 2) Funkhouser, T., and Li K., (eds.). Special issue of *IEEE Computer Graphics and Applications*. *Onto the Wall: Large Displays*, 20(4) (Jul/Aug 2000).
- 3) Johanson, B., Fox, A., Winograd, T., "The Interactive Workspaces Project: Experiences with Ubiquitous Computing Rooms," *IEEE Pervasive Computing Magazine*, 1(2), April-June 2002.
- 4) Paradiso, J.A., Hsiao, K., Strickon, J., Lifton, J. and Adler, A., "Sensor Systems for Interactive Surfaces," *IBM Systems Journal*, Vol. 39, No. 3&4, October 2000, pp. 892-914.
- 5) Quinnell R.A., "Touchscreen technology improves and extends its options," *EDN*, vol.40, no.23, 9 Nov. 1995, pp.52, 54-6, 58, 60, 62, 64.
- 6) Martin, D.A., Morrison, G., Sanoy, C., and McCharles, R., "Simultaneous Multiple-Input Touch Display," Presented at the UbiComp 2002 Workshop on Collaboration with Interactive Walls and Tables, Gothenburg, Sweden, September 29, 2002. See URL of Ref. 1.
- 7) Paradiso, J., "The Brain Opera Technology: New Instruments and Gestural Sensors for Musical Interaction and Performance," *Journal of New Music Research*, Vol. 28, No. 2, 1999, pp. 130-149.
- 8) Ishii, H., Wisneski, C., Orbanes, J., Chun, B., and Paradiso, J., "PingPongPlus: Design of an Athletic-Tangible Interface for Computer-Supported Cooperative Play," in *Proceedings of Conference on Human Factors in Computing Systems (CHI '99)*, Pittsburgh, Pennsylvania, May 15-20, 1999, ACM Press, NY, pp. 394-401
- 9) Cremer, L., Heckl, M., and Ungar, R.R., **Structure-Borne Sound**, Second Edition, Springer-Verlag, New York, 1990.
- 10) Leo, Che King, "Contact and Free-Gesture Tracking for Large Interactive Surfaces," M.Eng. Thesis, MIT Dept. of EECS and MIT Media Lab, May 2002.
- 11) J. A. Paradiso, C. Leo, N. Checka, K. Hsiao, "Passive Acoustic Sensing for Tracking Knocks Atop Large Interactive Displays," in the *Proceedings of the IEEE Sensors 2002 Conference*, Orlando, Florida, June 11-14, 2002, Vol. 1, pp. 521-527.
- 12) Pearson, H., "Bus Shelters to Talk Back," *Nature News Service*, September 16, 2002, See: <http://www.nature.com/nsu/020909/020909-15.html>.

*Note: video clips demonstrating the systems described in this article can be viewed at:*

<http://www.media.mit.edu/resenv/wall.html>

# FIGURES:

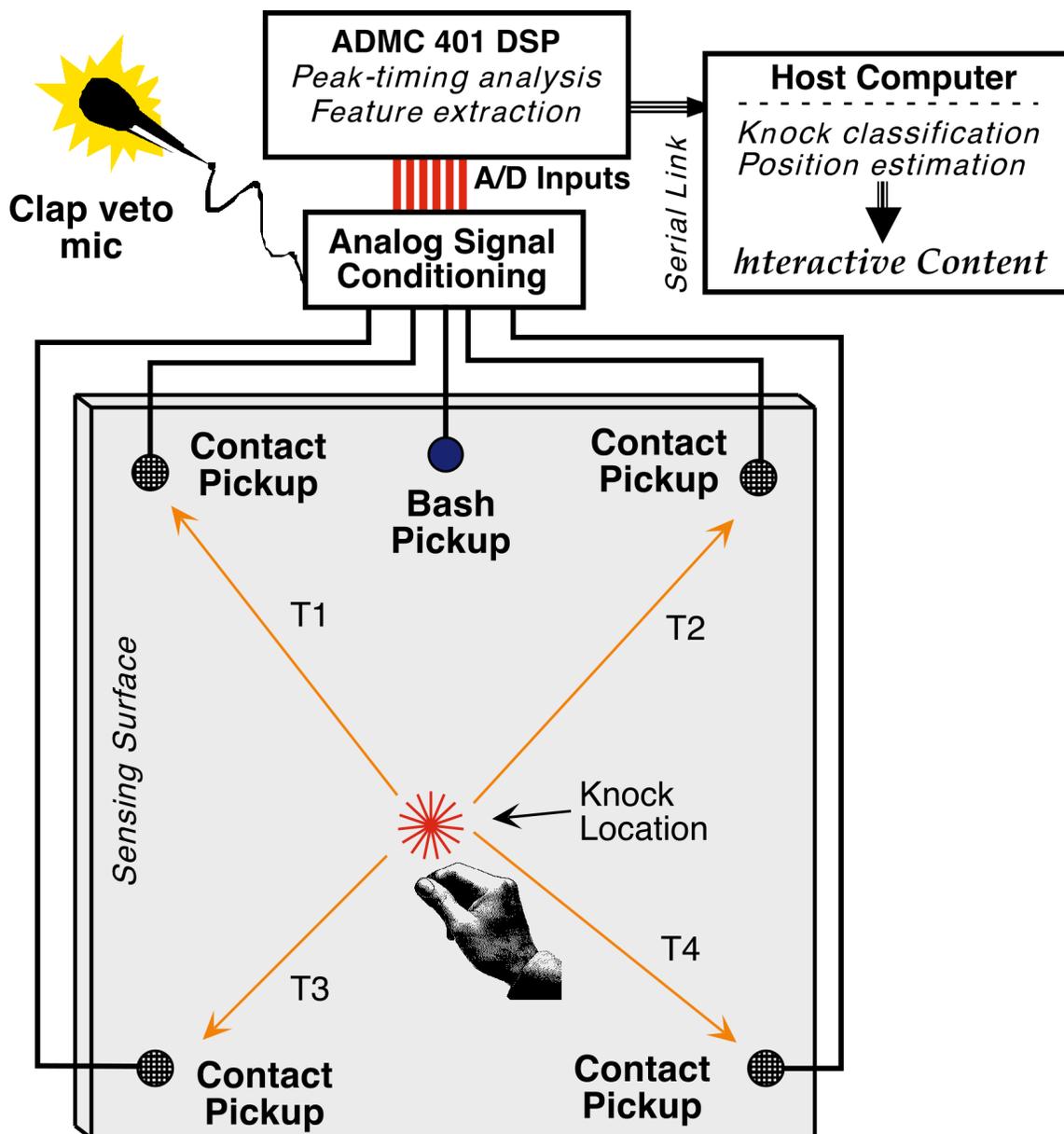


Figure 1: Essentials of the impact tracking system

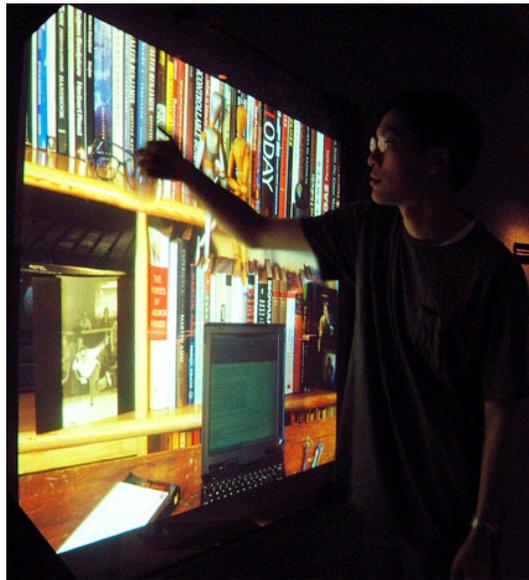
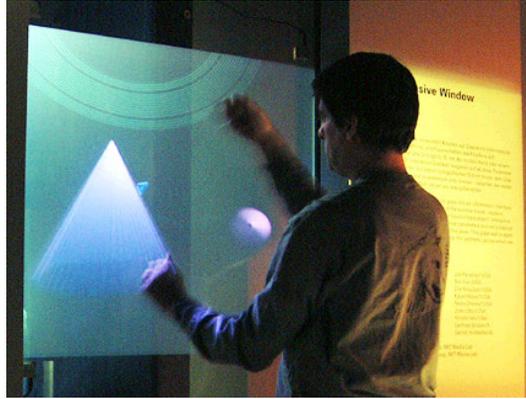


Figure 2: Interactive Art Applications – The *Responsive Window* (top), *Telephone Story* (middle), and *Weather* (bottom)



**Figure 3: Interactive window browsing at an American Greetings store in Manhattan**