Passive Acoustic Sensing for Tracking Knocks Atop Large Interactive Displays

Joseph A. Paradiso, Che King Leo, Nisha Checka, Kaijen Hsiao

Responsive Environments Group, MIT Media Laboratory
Cambridge, MA 02142 USA
{joep,cheking,nchecka,kjhsiao}@media.mit.edu

Abstract

We describe a system that locates the position of knocks and taps atop a large sheet of glass. Our current setup uses four contact piezoelectric pickups located near the sheet's corners to record the acoustic wavefront coming from the impacts. A digital signal processor extracts relevant characteristics from these signals, such as amplitudes, frequency components, and differential timings, which are used to estimate the location of the hit and provide other parameters, including the rough position resolution, the nature of each hit (e.g., knuckle knock, metal tap, or fist bang), and the strike intensity. As this system requires only simple hardware, it needs no special adaptation of the glass pane, and allows all transducers to be mounted on the inner surface, hence it is quite easy to deploy as a retrofit to existing windows. This opens many applications, such as an interactive storefront, with content controlled by knocks on the display window.

Keywords

touch screen, interactive display, passive acoustic tracking

INTRODUCTION

Glass is now a very common construction material, often used as clear walls for room dividers or large windows bordering urban buildings. The techniques described in this paper aim at enabling these surfaces to become interactive. For example, information displayed on a projection or monitor on the inside of the glass can be determined by knocking appropriately on the outside. A straightforward application of this niche is an interactive storefront, where passers-by can navigate through information on the store's merchandise or explore special offers by knocking appropriately, even when the store is closed.

OTHER APPROACHES

Many techniques have been developed to track the position of hands above interactive surfaces [1]. In general, most techniques used in conventional touch screens [2] don't scale gracefully to large displays. Pressure-sensitive resistive sandwiches, the most common technique, are not yet produced in sizes that can cover large windows, and because of their operational principle, would need to be mounted on the window's outside (active) surface, where they would be subject to potential damage over time. Active acoustic touch screens detect the absorption of ultrasound launched into the outer surface of the glass when a finger is in contact. They require piezoceramic transducers to be mounted at a corner on the outside of the glass and need the glass' surface to be properly patterned with etched reflectors along the edges that direct the acoustic energy towards the middle of the glass plate. These systems also become error-prone as the surface scratches and deteriorates, plus attenuation of ultrasound in the glass can become an issue for large panes. Capacitive techniques are able to measure hands through the glass, but require a matrix of transparent electrodes to be patterned or adhered across the entire sensitive surface. Infrared LED curtains, featuring dense arrays of opposing IR transmitters and receivers measure the hand position when the corresponding optical beams are occluded. Although these systems can be scaled up to a larger screen, their expense increases accordingly, plus, as the IR array must line the perimeter of the screen's outer surface, they can be subject to damage, dirt, and deterioration, especially if they're used outdoors.

Other techniques, such as video tracking [3,4] have been used to make large screens interactive. Although they are steadily improving, vision-based approaches can be slow and are often sensitive to image clutter, target reflectance, and changes in background lighting. Time-of-flight laser rangefinding has also been used to scan the surface of large displays [1] and find hands, but the potentially expensive laser scanner must be mounted outside the window, leading to reliability difficulties for outdoor operation.
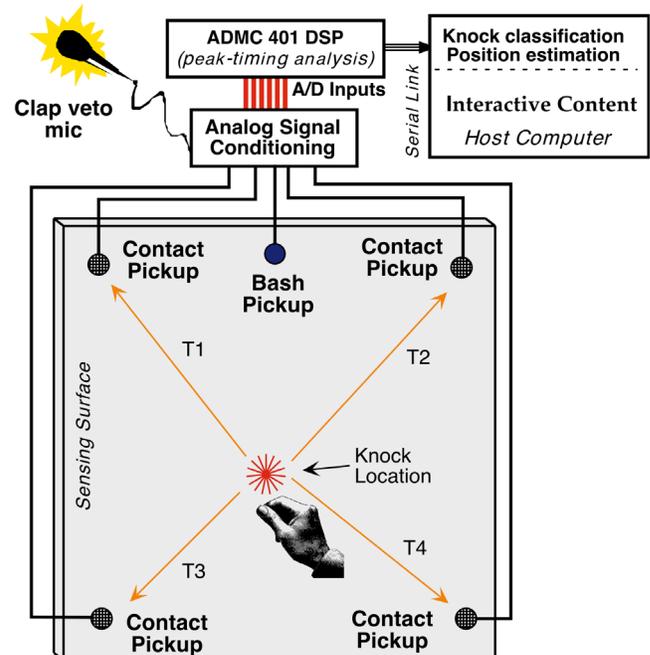
Figure 1: Hardware configuration for knock tracker system

PASSIVE ACOUSTIC IMPACT TRACKING

Although the original motivation for this system was to track the location of knocks on a virtual fishtank [5], the first implementation of this technique was an augmented ping pong table [6]. Here, four contact electret microphones recorded the impact of the ping pong ball at the corners of each contiguous half of the table. A simple 8-bit "PIC" microcontroller was able to adequately time the leading-edge of each microphone signal as it arrived, and the difference in arrival times (processed through a linear least-squares fit) produced the Cartesian coordinates of the impact, essentially in real time. The knock of a knuckle on glass, however, is markedly different from the impact of a ping pong ball on wood. The ball provides a consistent, high-amplitude waveform with a steep rising edge, requiring very little signal processing. The knuckle taps, however, are more challenging - each tap is different (depending on anatomy, intensity, style of knock, etc.), and generally of lower frequency, hence exhibit a much softer edge, which is much more difficult to time. The wave propagation characteristics of the glass add even more complexity. Knuckle taps launch "bending" or "flexural" waves [7] into the glass; the glass bends into "ripples" that propagate out from the point of impact. These are structural acoustic modes that travel much more slowly than the sound velocity in glass. Although this helps for resolving position, flexural waves are highly dispersive, causing the wavefront to increasingly slide apart as it travels further through the glass and high-frequency components arrive earlier. Different propagation velocities and analysis techniques must be used for knuckle taps and hard-object (e.g., "metal") taps, which travel many times faster through the glass. These complications preclude the simple constant-threshold discriminator; in contrast with the table, the signals for glass need to be digitized and processed algorithmically.

HARDWARE SETUP

Figure 1 shows the layout for our acoustic tap tracker system. Although the contact microphones that we originally used [8] were made from strips of simple PVDF piezoelectric foil laminated onto the surface of the glass, we have shifted to piezoceramic transducers made by Panasonic (the EFV-RT series), which provide considerably more sensitivity across the full bandwidth of various knocks and taps (e.g., 100 Hz - 5 kHz). Figure 2 shows photos of both transducers attached to the glass (the electronics for the Panasonic microphone are shielded from interference by a copper housing). Both systems employ a gain-of-10 voltage amplifier mounted at the transducer, enabling a long cable to be driven to the signal conditioning electronics, which filter out unwanted noise while applying additional gain. Although we generally adhere the microphones to the glass surface with standard cyanoacrylate adhesive, long-term installations may prefer a more permanent epoxy joint. An inexpensive digital signal processor (the ADMC 401 from Analog Devices, designed for motor control applications) extracts relevant parameters from the digitized waveforms and ships them across a serial link to a standard PC, which classifies the knock, estimates its coordinates, and runs appropriate interactive content.
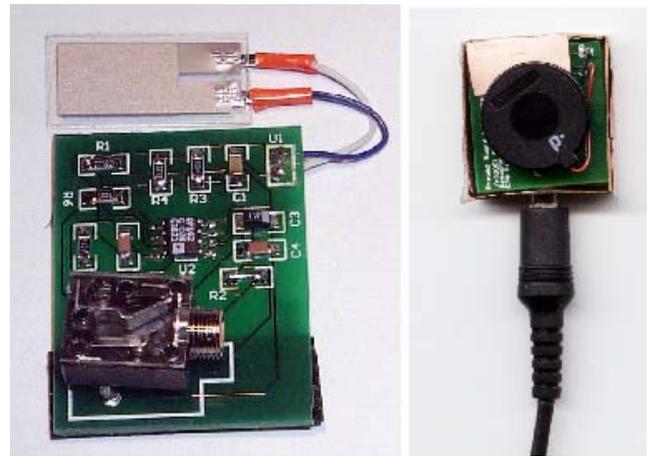
Figure 2: Contact pickups - original PVDF w. preamp (left) and current piezoceramic unit w. assembly (right)

Although the piezoceramic transducers respond well to knocks and metallic hits, they don't give much signal when the window is banged with a fist. This primarily introduces low frequency vibration, down below 50 Hz, where the pickups aren't as sensitive. Accordingly, we have attached another transducer to the window to detect these "bash" events; it is an inexpensive electrodynamic cartridge, with the diaphragm epoxied to the window glass. As such bashes are of very low frequency, they don't tend to provide good timing resolution and tracking, hence one transducer is sufficient, placed near the edge of the glass where it is out of the way. As this pickup responds very strongly to bash events (and very weakly to knuckle taps and metal hits), it detects them with essentially no ambiguity, as illustrated in the next section.

Because the piezoceramic transducers are strongly adhered to the glass, they are essentially contact microphones and give very little response to signals coming from the air and not generated by taps on the glass. Certain sounds, however, such as a very loud clap or "snap" produced near the window, couple enough energy into the system to trigger a false event. To prevent this from occurring, we have provided for another transducer in our system; this is a simple crystal microphone, not adhered to the window, but listening to ambient sounds made in its vicinity. The peaky high-frequency response of the crystal microphone makes it respond strongly to the sharp sounds that instigate false triggers, hence any events that exhibit significant amplitude from this channel are vetoed as external sounds, not assumed to be generated in the glass.

Although the signals from the four piezoceramic transducers of Fig. 1 appear to provide sufficient information to detect bashes and veto claps on their own, this would require considerable additional signal processing and tweaking - the additional two transducer channels make this determination much more simple and robust.

The DSP provides 26 MIPs worth of processing and samples all 6 input channels into 12 bits at 50 kHz. It continuously samples the input signals, triggering when any rise above a preset threshold. We currently retain 8 ms

worth of data from the 4 piezoceramic transducers and the clap veto transducer (5 ms before the trigger and 3 ms afterwards, nicely capturing the incoming wavefront before the glass begins steady-state modal oscillation). As its low-frequency signal comes somewhat later, we retain 20 ms from the bash transducer (5 ms before trigger and 15 ms afterwards).

Figure 3 shows data from one transducer triggered by the signal from another for a series of knuckle taps running between them atop a 1 cm-thick piece of glass; the pickups are separated by 0.9 meters. The progressive time delay, upon which this technique is based, is quite evident, as is the dispersive nature of the glass.
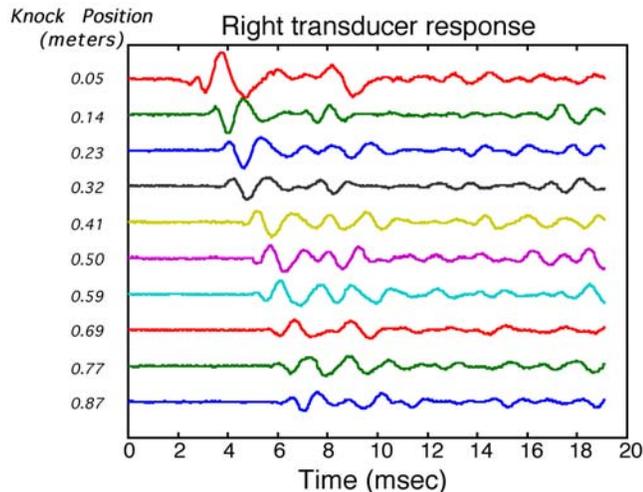
Figure 3: Increased delay as knock moves away from pickup

DATA PROCESSING

Our initial strategy [1] was centered around triggering the system when the amplitude of a transducer rose well above noise, then walking up the attack of the knock transient to determine the point at which the signal emerged from the noise floor. Although this attained a basic degree of performance, it could be quite erratic, as background noise and the low-amplitude high-frequency signals arriving earlier through dispersion could often have considerable (and variable) influence on the timing. Our next steps [8] digitized the signals with a data acquisition card, allowing us to explore more complex algorithms under MATLAB. We attained much more robust performance by cross-correlating the signals across pairs of sensors and extracting the differential time from the correlation peak. Because of the distortion encountered when propagating through the glass, however, the correlation could often become ambiguous, producing two or more significant peaks. In these cases, the redundancy in the system (provided by the additional sensor - only 3 are needed to specify position in a plane) and data from a calibration procedure were used to select the appropriate peak. Details are provided in Refs. [8,9].

A combination of correlation and rising-edge results (depending on the signal quality) are used for knuckle taps, which are generally at much lower frequency (e.g., several hundred Hz) than hard taps with a metallic object, which
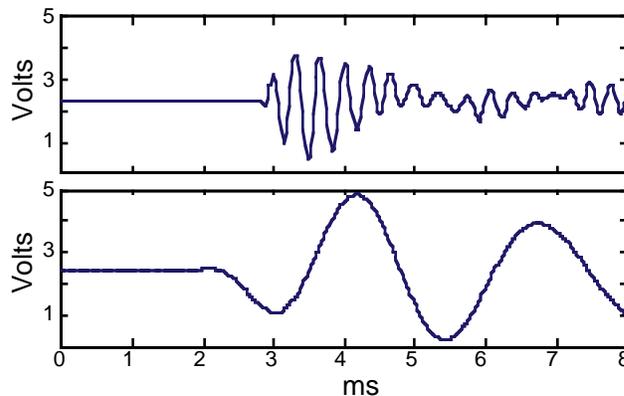
Figure 4: Knuckle (lower) and metal tap (upper) signals
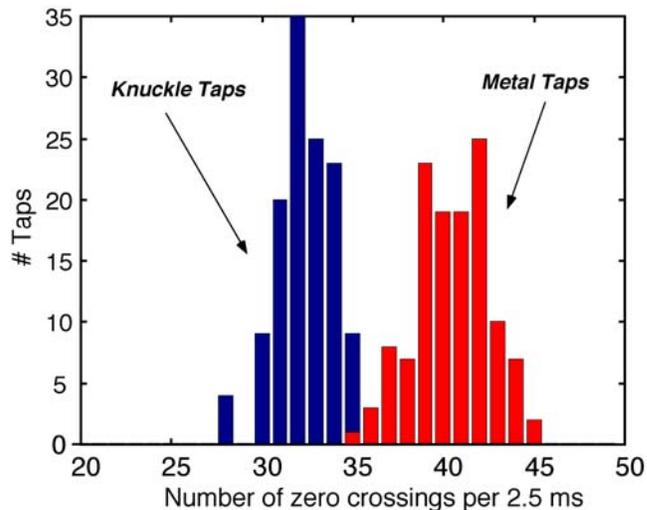
Figure 5: Histogram of knuckle and metal tap frequency

can create significant components in the vicinity of 2 kHz, as illustrated in Figure 4. As the metal taps exhibit a much sharper transient, we default to the original technique, where we walk back along the waveform after triggering to detect where the first arrival of the signal disappears below the noise floor. Likewise, because of dispersion and the different acoustical modes launched by the hard, metallic impact, the propagation velocity for metal taps is generally 2-3 times faster than for knuckle taps. Although one could discern knuckle taps from metal taps by examining the frequency distribution obtained from a FFT, we accumulate a coarse but adequate estimate of frequency by counting the number of times the sensor waveforms significantly cross zero across a fixed interval. As seen in Figure 5, which shows this quantity for several knuckle taps and metal taps, the two distributions are nicely separated and this distinction works well. In order to make this critical choice even more reliable, we also use the differential timing (e.g., the propagation velocity) between sensor pairs in this discrimination.

The operations described above are currently all executed in the DSP, and all relevant parameters are sent over a serial link to the attached PC, which then determines the impact position. In our earlier work [1,8], we derived the (x,y)

3

impact coordinates by running the timing results through a third-order polynomial that was determined via a linear least-squares fit to data collected on a regular grid. To avoid this lengthy calibration process (that was prone to overfitting), we have recently shifted to a deterministic algorithm [9]. As the position ambiguity curves for any pair of sensors form a hyperbola [10], we calculate the closest intersection between a pair of hyperbolas, each derived from the data from two different sensors. The amount of miss between the hyperbolas reflects the quality of the data, hence the resolution of the position estimate. Because of site-dependent effects (e.g., changes in the propagation velocity in different parts of the glass that are subject to different degrees of mechanical loading), we do some adjustment of the hyperbola calibration to fit specific windows.
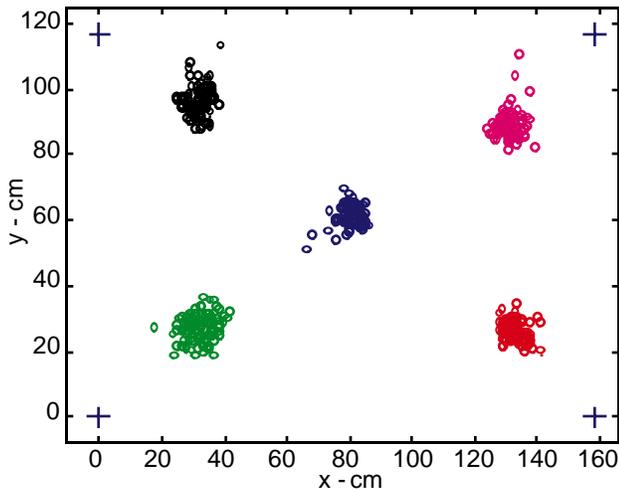
Figure 6: Estimated positions for knuckle taps at 5 locations
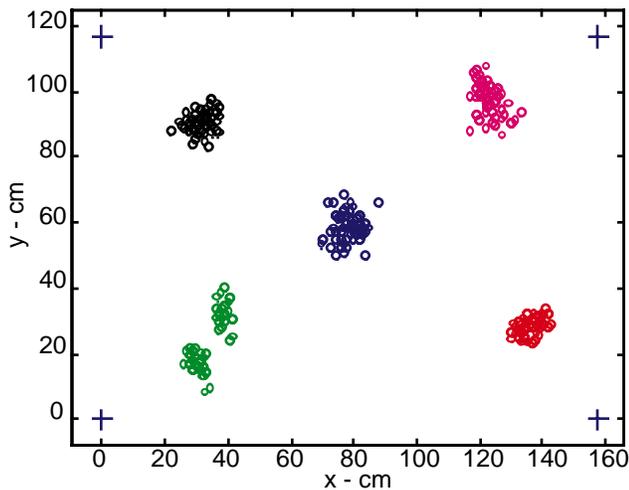
Figure 7: Estimated positions for metal taps at 5 locations

The peak amplitude is also captured for each sensor across the data acquisition interval. For the four contact pickups in the corners, these quantities reflect the hit's intensity. Although some techniques use differential amplitude to determine position of impact (e.g., [11]), as the wavefront is attenuated as it travels through the glass, we have found

that the timing data is much better behaved. The peak amplitude of the other two sensors are used to determine a "bash" event (if the bash sensor amplitude dominates), or a "clap" veto (if the clap sensor amplitude dominates).

PERFORMANCE

We have used this system thus far with primarily two different thicknesses of glass: a 1/4" thick pane of tempered window glass and a 1-cm thick pane of shatterproof room-divider or shop window glass. In the 1-cm pane, we observe a propagation velocity of 1200 m/s for the knuckle tap and 2000 m/s for the metal tap; the speeds are roughly 40% slower in the thinner glass. Our present systems have been used with an active area between the sensors of 2 by 2 meters. The results presented here in Figures 6-8 are all taken with the 1/4" glass, with sensors spaced at the corners of a 1.6 x 1.4-meter rectangle.

Figure 6 shows the (x,y) reconstructed points for knuckle knocks at 5 locations (100 knocks at each site), and Figure 7 shows the analogous data for metal taps (the "+" symbols in the corners denote the sensor positions). We see average resolutions of $\sigma = 2.5$ cm for the knuckle taps and $\sigma = 2.7$ cm for the metal taps at each position. The metal tap results tend to be somewhat less accurate because of the higher propagation velocity, while the knuckle taps can show inaccuracy due to errors in choosing the correlation peak. Because of the higher velocity, results with the thicker, 1-cm glass tend to be somewhat less precise, with sigmas approaching 3-4 cm. By using the tweaked hyperbolic fits, the accuracy was quite good, with the centroids of the distributions of Figures 6 and 7 appearing within a few cm of the knock and tap positions. Occasionally, depending on the type of glass and the style of knock, the pickup waveforms can be inconsistent, resulting in an ambiguous condition. In these cases (generally 1% or lower), the event is rejected, and no coordinate is produced.

We have used this system now on many different windows. In one case, with the 1-cm glass, the window extended for over a meter above and below the region plotted here (it was grabbed along its sides by a rubber bumper to quickly damp resonances and avoid rattling). Accordingly, a user was able to tap above and below the region bounded by the sensors. The differential timings are able to determine this condition, however, and still produce coordinate estimates. As the time differences in the vertical direction show little change with vertical position in these regions, the resolution here was poor; the horizontal coordinate was still usable, however, reaching the order of $\sigma = 5.5$ cm when tapping a foot above or below the sensor-bounded perimeter.
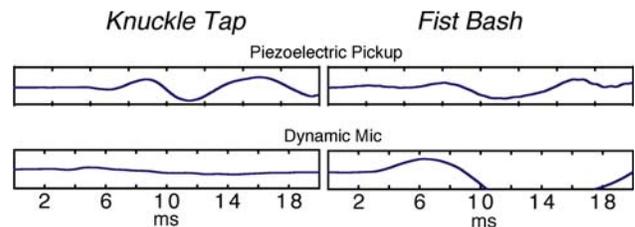
Figure 8: Piezo & dynamic pickup responses to tap and bash

The trigger thresholds were placed sufficiently low to respond to a soft knock and not activate sporadically with room noise. When sharp room sounds did occur, the "clap" veto signal successfully prevented the system from falsely responding. As poor pickup shielding in our prototype hardware made the system susceptible to interference from nearby cellphones, we implemented a simple waveshape identification algorithm in the DSP that successfully prevented this situation from triggering the system. The "bash" detector reliably discriminated between knocks and even modest fist "bangs", as illustrated in Figure 8, which shows the bash sensor saturating for a modest bang but showing very little signal for a knock.

The system responded quite quickly. The DSP generally produced parameters within 50 ms of an impact, and a PC (a 1 GHz Pentium 3) produced (x,y) coordinates circa 15 ms later, yielding a net system latency within 65 ms. Tighter coding could speed this up considerably, and the DSP could potentially be replaced with a simple 4-channel audio digitizer, with all processing happening in the PC.

Figure 9: A Knock-Activated Browser

Figure 10: The Responsive Window at Ars Electronica

APPLICATIONS

As this system has evolved, we have explored its use with increasingly sophisticated applications. All involved graphics projected onto a screen behind the glass; the user interacts with the information by knocking, tapping, and banging on the front surface. In all cases, the pickups were mounted on the rear of the glass.
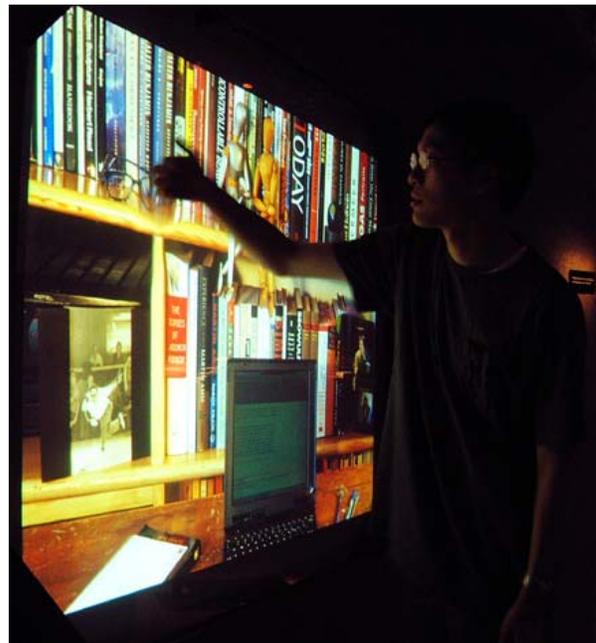
Figure 11: TelephoneStory at the Kitchen Gallery in NYC

Figure 12: Interactive window browsing at Rockefeller Ctr.

Our first application [1] was a simple diagnostic that plotted circles centered at the location of the knock, with radius proportional to the estimated coordinate accuracy. After perfecting our hardware and analysis with this tool, we built a more user-relevant application, illustrated in operation in Figure 9. This is a "knock-driven" browser [8] demonstrated in our lab, where a user navigates group projects by knocking on the corresponding picture, launching relevant web pages and/or playing associated video clips. The targets (bounded by the interpicture gaps) were sufficiently far apart to make this system quite usable, even with a coarse resolution of several centimeters.

Our next deployment, shown in Figure 10, was somewhat more sophisticated. It is a semipermanent installation at the Ars Electronica Center that was debuted in the 2001 Ars Electronica Festival as the "Responsive Window" [12]. A "holoscreen" [13] is mounted behind the 1-cm glass and located within the perimeter enclosed by the transducers. The holoscreen is a holographic diffuser that gives a rather "ghostly" appearance to the graphics. Nothing is projected (and the screen is mildly translucent) until the glass is knocked, at which point the images appear. The graphics that we ran were essentially a simple knock-driven drawing program that illustrates the full responsiveness of the system. One could outline and extrude an object by knocking about, with different effects and events caused by knock amplitude, bangs, and tap timbre.

Figure 11 shows another "knock-screen" application, where a desk full of objects is projected on the glass; when images of relevant objects are knocked, a corresponding video stream is launched. If the user knocks during the video, a relevant still image will briefly flash up, centered at the knock location. Bangs create a shower of images that fly away from the rough bang location. The data of Figs. 6-8 were taken with this installation, which was exhibited this winter at the Kitchen Gallery in New York City.

Figure 12 shows our device in the window of an American Greetings store near Rockefeller Center in Manhattan, where it successfully ran during the Christmas and Valentine's Day seasons this winter [14]. Passers-by could start videos or play simple games by appropriately knocking.

CONCLUSIONS

We have demonstrated a technique of easily retrofitting common windows for contact interactivity by measuring the position of a knock or tap, determining the type of impact, and estimating the impact intensity. Because of the complicated nature of the various impacts and nonideal propagation characteristics of the glass, this is not a precise pointing device, as it yields resolutions on the order of $\sigma = 2\text{-}4$ cm across roughly a meter. Its accuracy is adequate, however, for several applications, such as navigating through content at interactive storefronts. Better signal analysis could extract more reliability and accuracy from this system; some possibilities could involve locating the arrival transients with a wavelet analysis or compensating for the dispersion in the glass by using coarse estimated range information to remove the expected dispersion from the signals. For larger windows, more transducers can be added, providing increased redundancy in the measurements and keeping the minimum propagation distance small. Also, more sensitive pickup sensors could be investigated, such as wideband accelerometers, providing reliable response across larger areas. We have brought this system out to the public, where it is received with interest because of its novelty - once users recover from their initial confusion with touch screen operation (one must hit rather than touch), they generally adapt well to it. Rear-mounting of the sensors will probably not function with air-insulated double-paned windows, as the rear surface is acoustically isolated from the front, where the impact occurs.

REFERENCES
[1] 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.

[2] 2. 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.

[3] Maes, P. et al. 1997, "The ALIVE System: Full-body interaction with animated autonomous agents," Multimedia Systems, vol.5, no.2, March 1997, pp.105-112.

[4] Rekimoto, J., Matsushita, N., "Perceptual Surfaces: Towards a Human and Object Sensitive Interactive Display. In Workshop on Perceptual User Interfaces (PUI-97), Banff, Published by Microsoft, October 1997, pp. 30-32.

[5] Galyean, T., et al., "Virtual FishTank," SIGGRAPH 98, Conference Abstracts and Applications, ACM Press, New York, 1998, p. 116.

[6] 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.

[7] Cremer, L., Heckl, M., and Ungar, R.R., Structure-Borne Sound, Second Edition, Springer-Verlag, New York, 1990.

[8] Checka, N., "A System for Tracking and Characterizing Acoustic Impacts on Large Interactive Surfaces," MS Thesis, MIT Department of EECS and MIT Media Lab, May, 2001.

[9] Leo, Che King, "Contact and Free-Gesture Tracking for Large Interactive Surfaces," Meng Thesis, MIT Dept. of EECS and MIT Media Lab, May 2002.

[10] Reynolds, M., "An Immersive, Multi-User, Musical Stage Environment," ACM Computer Graphics, SIGGRAPH 2001 Proceedings, ACM Press, NY, 2001, pp. 553-560.

[11] Mathews, M.V., "The Conductor Program and Mechanical Baton," in Mathews, M.V., and Pierce, J.R., eds., Current Directions in Computer Music Research, MIT Press, Boston, MA, 1989, pp. 263-282.

[12] Paradiso, J., "The Responsive Window," in Take Over, Proc. of the 2001 Ars Electronica Festival, Springer Verlag, Linz, Austria, September 2001, pp. 261-263.

[13] See: http://www.holopro.de/

[14] Wilson, C., "If You're Game, Tap on the Window," USA Today, December 21, 2001, p. 1D.