Touch & Activate: Adding Interactivity to Existing Objects using Active Acoustic Sensing

Makoto Ono, Buntarou Shizuki, and Jiro Tanaka
University of Tsukuba, Japan
Tennoudai 1-1-1, Tsukuba, Ibaraki, Japan 305-8571
{ono,shizuki,jiro}@iplab.cs.tsukuba.ac.jp

ABSTRACT
In this paper, we present a novel acoustic touch sensing technique called *Touch & Activate*. It recognizes a rich context of touches including grasp on existing objects by attaching only a vibration speaker and a piezo-electric microphone paired as a sensor. It provides easy hardware configuration for prototyping interactive objects that have touch input capability. We conducted a controlled experiment to measure the accuracy and trade-off between the accuracy and number of training rounds for our technique. From its results, per-user recognition accuracies with five touch gestures for a plastic toy as a simple example and six hand postures for the posture recognition as a complex example were 99.6% and 86.3%, respectively. Walk up user recognition accuracies for the two applications were 97.8% and 71.2%, respectively. Since the results of our experiment showed a promising accuracy for the recognition of touch gestures and hand postures, Touch & Activate should be feasible for prototyping interactive objects that have touch input capability.

Author Keywords
Touch; grasp; gestures; sensors; acoustic classification; tangibles; machine learning; prototyping; support vector machine; piezo-electric sensor.

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

General Terms
Human Factors.

INTRODUCTION
Interactive objects that have touch input capability are still challenging to prototype, even while touch input is increasingly used in many consumer products such as mobile devices or tablets. Even engineers spend much time on circuit design and hardware configuration. This is because prototyping usually requires the developers to redesign the objects many times, which involves replacing sensors frequently especially in the early stage of prototyping.

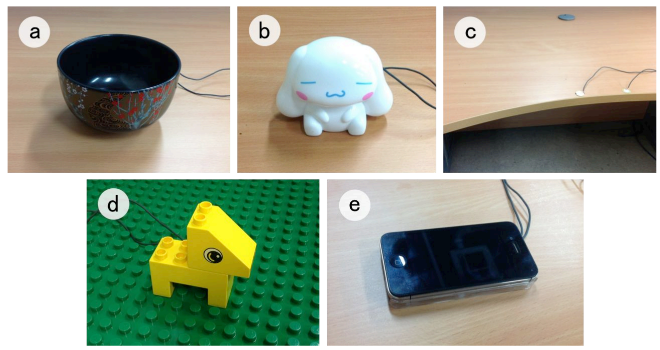

Figure 1. Examples of applicable objects: a) ceramic bowl, b) plastic toy, c) wood desk, d) Duplo block, e) mobile device (hard case).

Moreover, grasps, a kind of touch with rich context, have been explored intensively in recent years, and many grasp sensitive objects have been elaborated. For example, Kim et al. recognized hand grip patterns on mobile devices by embedding 64 capacitive sensors into mobile devices [32]. FlyEye [51] spread optical fibers across a surface, with their other ends attached to a camera, to prototype grasp sensitive surfaces. Taylor et al. developed objects with 23-72 capacitive sensors on their surfaces to explore how the way users hold and manipulate physical objects can be used as inputs [47]. HandSense [53] detected how a mobile device is held using four capacitive sensors, each of which has an antenna made of a tin sheet and is actively shielded by guard electrodes wrapping the antenna on three sides to improve sensitivity. MTPen [44] is a stylus, whose body's surface has multi-touch sensing capability using conductive ink, which detects how the user grips the stylus. FlexAura [33] detects how the user grips an object by using a flexible range finder, which is an 16×24 array of infrared LEDs and phototransistors. Note that these works all use many sensors, specialized hardware, or complicated hardware configuration; thus grasp sensitive objects are also challenging to prototype.

In this paper, we describe Touch & Activate, a novel acoustic touch sensing technique that enables prototyping interactive objects that have touch input capability rapidly and easily. It recognizes a rich context of touches, including grasps on existing objects, such as those shown in Figure 1, by attaching only a vibration speaker and a piezo-electric microphone paired as a sensor. As a result, developers can concentrate on the design of touch input rather than be overly concerned with circuit design or hardware configuration.

Our sensing technique is based on the resonant property of solid objects, which is highly sensitive to the way objects are touched. This leads to the following additional advantages: since most solid objects have this property, our approach is applicable to existing solid objects and does not require conductivity or magnetism on them and is thus safe; our sensing technique can estimate rough positions of touch and different postures of touch (e.g., pinch or press) on a single position; due to the sensitivity, developers can assign a position and/or posture of touch to an action quickly by training the system using machine learning. These advantages provide developers with possibilities to flexibly design various styles of interaction with the object.

RELATED WORK

Prototyping Physical Objects

Many tools have been developed to lower the barrier to physical prototyping. Phidgets [14] are packaged devices consisting of sensors and actuators that can be programmed with a well-defined software API. Arduino [1] and mbed [2] are electronics prototyping platforms that can be programmed via offline or online IDE. d.tools [25] is a prototyping environment that integrates prototype designing, testing, and analyzing into the visual programming interface. .Net Gadgeteer [48] provides a similar packaged module that can be programmed by using .Net Micro Framework. Prototyping with these abovementioned tools usually involves the use of off-the-shelf sensors. This means that while the prototype, once made, can work stably, the forms (i.e., shape and size) of sensors used are restricted to the ones commercially available. Therefore, developers may need to search for other forms of sensors when they want to examine prototypes of various shapes and sizes. Since this scenario occurs frequently in the early stage of prototyping, a more flexible approach is required.

Prototyping tools targeting touch sensitive prototypes have also been proposed. BOXES [28] is a tool with which developers can construct touch sensitive prototype from existing objects by attaching thumbtacks and foil to the objects. Tactile Tape [27] is a one-dimensional touch sensor that can be used to add touch sensitive areas on existing objects. Wimmer et al. explored how to extend time domain reflectometry in order to make thin, modular, and deformable surfaces and devices touch sensitive [52]. Midas [42] helps developers create their custom sensor layouts by the GUI editor. These previous works construct touch sensitive prototypes from everyday objects by attaching conductive elements to each position that will be touched. Meanwhile, our approach does not require these elements except for a pair of a speaker and a microphone. This does not only make the hardware configuration easier, but also makes the instrumentation used for sensing less obtrusive than the above approaches.

Physical prototyping tools using machine learning and pattern recognition techniques have also been proposed. Exemplar [24] is a design tool with which developers can prototype sensor based applications by demonstration for sensor-based interaction. PICL [12] is a portable in-circuit learner with which dvelopers can also prototype sensor-based applications with standalone machine learning hardware. Our approach employs machine learning to label acoustic response (i.e., frequency response in our study) as touch gestures.

Touch Sensing Technologies

Recent innovative touch sensing technologies have lowered the barrier to incorporating touch sensing into prototyping and broadened sensing capabilities. FTIR [17] enabled multi-touch surface to be made for low-cost using frustrated total internal reflection. OmniTouch [18] is a wearable depth-sensing and projection system that realizes interactive multi-touch applications on everyday surfaces. KinectFusion [30] can be used to realize multi-touch interactions on any physical surfaces, including complex ones, by using a Kinect camera. Although these vision based touch sensing technologies detect touched positions in high resolution, they require large footprints or high performance hardware for image processing. Touché [41] and Capacitive Fingerprinting [20], the works that mainly inspired us, found that the Swept Frequency Capacitive Sensing technique can detect a touch event and recognize complex configurations of human hands and body. However, Touché requires conductivity of objects. This means that, when developers apply it to non-conductive objects, they have to coat the objects with conductive ink or tape.

Acoustic Sensing

Researchers have explored many acoustic-based input techniques. These techniques are classified into passive and active approaches.

Passive approaches detect a user's input by capturing and analyzing sounds generated by the user's actions. The Sound of One Hand [3] recognizes fingertip gestures such as tapping, rubbing, and flicking by analyzing bone-conducted sound generated by the actions. Stane [36] classified sounds generated by actions such as scratching or rubbing its textured surface. Scratch Input [19] explored an acoustic based input technique using the unique sounds generated by scratching a surface of a textured material. Acoustic Barcodes [23] are tactile barcodes that, when an object like a fingernail or phone runs across the notches, identify a binary ID from the produced sound. Paradiso et al. employed multiple microphones to estimate tapped positions on a large sheet of glass based on time difference of arrival [37]. Pham et al. also employed a single microphone to localize tapped positions in non-planner objects based on time reversal [39]. Skinput [22] used bio-acoustic sensors to estimate the tapped position on the skin. TapSense [21] and Pedro et al. [34] identified the type of objects, or the part of the finger, being used for tapping from sounds generated by tapping the surface. MicPen [29] is a pressure sensitive stylus that estimates the amount of pressure applied to the stylus' tip from the scratch noise generated by dragging surfaces with its tip.

Active approaches transmit sounds using speakers, and capture and analyze the response to the sounds. Active Sonar [45] mounted on submarines and fishing boats is its typical example. It detects surrounding objects and measures distances to them by transmitting pulses of sounds and timing the response to the sound. Surface Acoustic Wave (SAW) touch screens [5] is another representative active approach. It consists of one glass sheet with a transmitter, a receiver, and reflectors. The transmitter emits SAW, which is an acoustic wave traveling on the surface of a material exhibiting elasticity, and they are reflected by reflectors that are lined up along the edges of the panel. When a finger touches the screen, the wave is absorbed. The absorption is observed by the receiver to localize the touch.

Other active approaches have also been explored in recent years. SoundWave [16] senses in-air gestures by measuring Doppler shifts. Takemura et al. proposed a wearable sensor system that estimates the angle of an elbow using a speaker and two microphones that capture bone-conducted sound in the elbow [46]. Collins estimated positions of a touch on flat surfaces by analyzing the frequency response of the surface using a single transmitter and one or two receivers [9]. It is similar to SAW touch screens in localizing touch from acoustic waves, but different in that it dose so by analyzing the frequency response with a simple instrumentation.

The advantage of active approaches for touch detection, such as those of Brenner and Fitzgibbon [5] and Collins [9], is that the system generates its own sounds and detects the changes in the sounds caused by the user's actions. This enables even the detection of gentle touches or varying pressures of touch, where almost no sound is produced. Although our approach is similar to that of Collins [9] in analyzing the frequency response of the acoustic wave, we extend it to classify grasp or farther context of touch on everyday objects and demonstrate a rapid prototyping method for interactive objects that have touch input capability with machine learning.

SENSING PRINCIPLE

All objects have their own resonant property represented by the resonant modes, natural frequency, and modal damping. The property depends on its own shape, material, and boundary condition. At present, this property has been applied to durability evaluation and vibration suppression of structures in the field of structural mechanics [11, 43].

Our study focuses on the boundary conditions among them. When an object is touched, the boundary condition changes. As a result, its resonant property also changes. The changes in the resonant property are observed as different resonant spectra. Figure 2 shows an example where a ceramic bowl touchable in different ways (Figure 2a-d) shows unique resonant spectra accordingly (Figure 2A-D). Therefore, if the object's shape and material are static, the change in the boundary condition caused by touching, grasping, and others can be observed as the change in the resonant property.

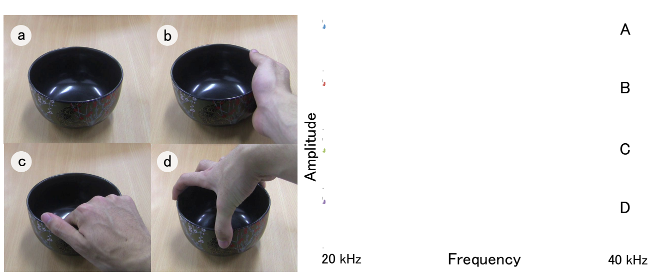

Figure 2. Resonant spectra of a ceramic bowl with different ways to touch. When the bowl is touched as shown in a-d, the resonant spectra of A-D are observed, respectively.

Our approach uses this phenomenon to estimate how an object is touched by analyzing resonant property. We use experimental frequency response testing to analyze resonant property. In general, this method involves vibrating the object at a wide frequency range using an actuator, such as an impact hammer, electrodynamic shaker, or vibration speaker, and capturing the frequency response by using a sensor, such as an accelerometer or piezo-electric microphone. In the field of structural mechanics, a theoretical modal analysis such as Finite Element Method (FEM) is then employed to calculate the details of the boundary condition. However, this requires detailed information about the shape and material of structures, and its calculation costs are high. On the other hand, we employ machine learning with frequency response labeled as touch gestures. This enables the estimation of touch gestures on an object whose structure information is unknown with lower calculation costs.

TEST SYSTEM OF TOUCH & ACTIVATE

We implemented a test system on the basis of the above sensing principle. Figure 3 shows the overview of our test system.

Hardware

Our test system consists of a vibration speaker, a piezoelectric microphone, and a computer running software for signal processing and machine learning.

We use a bimorph piezo-electric vibration speaker (Figure 4 left; Thrive OMR20F10-BP-310, 0.3 mm thick, 21 mm in diameter) to vibrate the object. The vibration speaker can transmit sounds by vibrating the object directly. We also use a unimorph piezo-electric microphone (Figure 4 right; Murata 7BB-20-6L0, 0.2 mm thick, 20 mm in diameter) to capture the vibration response. To vibrate an object with sufficient power, piezo elements with a bimorph structure are preferred to unimorph ones. Thus, we used a bimorph piezo element as an actuator and selected a smaller, thinner one that can be easily attached to any surface. In contrast, we used a unimorph piezo element as a microphone because the power is unnecessary for capturing the vibration response and it is less expensive than a bimorph one.

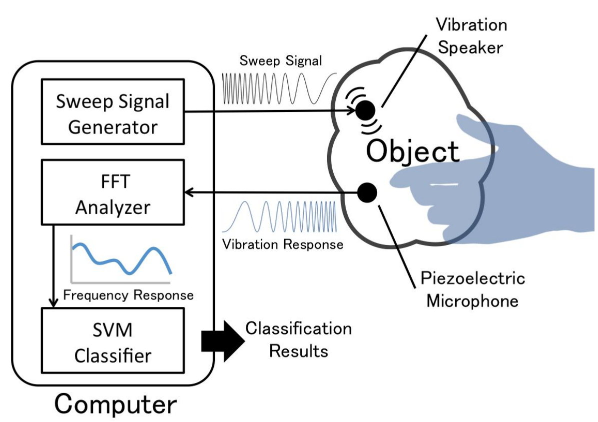

Figure 3. Overview of our test system.

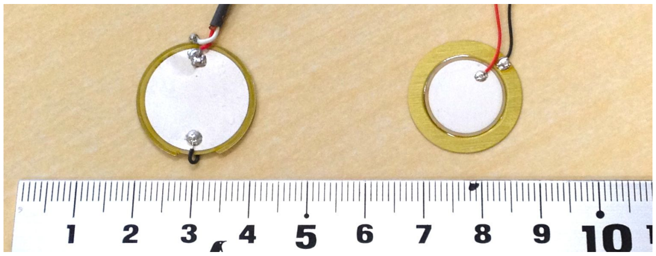

Figure 4. Vibration speaker and piezo-electric microphone.

These piezo elements are attached to various objects with double-sided tapes. Double-sided tapes are suitable as a mounting method in the early stage of prototyping using existing objects because this allows developers to remove the instrumentation from the objects easily. Instead, glues are appropriate for the final stage of prototyping because glues will provide lower mechanical impedance, making the actuation power necessary for the sensing lower. Moreover, glues connect piezo elements and objects tightly while eliminating ease to remove the instrumentation.

The signals current to/from the piezo elements are amplified and connected to a computer (Apple MacBook Air, CPU: Intel Core 2 Duo 1.4GHz, RAM: 2GB) via an USB audio interface (Native Instruments Audio Komplete 6) and processed by the software mentioned below.

Software

The software consists of three modules: Sweep Signal Generator, FFT Analyzer, and SVM Classifier. We used the BASS library[1] for audio I/O.

The Sweep Signal Generator we programmed generates sinusoid sweep signals from 20kHz to 40kHz, whose frequency increases linearly in 20 ms, at a 96kHz sampling rate. The module plays the signals repeatedly via the USB audio interface through the vibration speaker. The module also fades in/out both end of each period of the signals so as to suppress the impulse noise caused by the frequency gap. When the object is vibrated by the speaker, some sounds are emitted from the object to the air. We use 20k-40kHz sound since it is inaudible to humans [8] and robust to the contact noise produced by touching the object.

The FFT analyzer converts the vibration response captured from a piezo-electric microphone into the resonant frequency response. The module samples audio at 96kHz, which is the same rate generated by the Sweep Signal Generator, to obtain the high-frequency signal that ranges in 20k-40kHz, and uses the FFT with 8192-point hamming window vectors (85 ms long and overlapped with 20 ms) to compute the signal's 4096-point frequency vectors from 0 to 48kHz. This sampling does not need to be synchronized with the Sweep Signal Generator because the window vector of 8192 samples include at least four cycles of the sweep signal. This ensures that the frequency response we want is included in all window vectors at any moment during sampling while there are subtle differences depending on the phase of each sampling section. We then extract 400-point features representing from 20kHz to 40kHz with down-sampling. The purpose of down-sampling is to speed up recognition. To do this, we first determined the valid 400 frequency points that we use as features when the software is started. This is because there are many invalid frequency points (showing almost 0 value) within a raw frequency spectra, as the frequency resolution of the FFT is higher than that due to the sweep signal. Thus, we determined the valid 400 frequency points by applying peak detection algorithm to the raw frequency spectra.

The SVM Classifier classifies the touch gestures in real-time using a Support Vector Machine (SVM) implemented in LIBSVM [6] . We used the RBF kernel with default parameters as a kernel function for SVM. The parameters of the kernel can be tuned through a grid search to improve the accuracy. The 400-point features extracted by the FFT Analyzer are passed to the classifier every 20 ms (i.e., all data extracted by the FFT Analyzer are passed). Before the classification, the classifier must be trained for each gesture. For the informal training in the very early stage of prototyping, the module provides developers with a simple training mode where developers can train it by pressing any key (except the space bar) while performing a gesture. While a key is pressed, all features are passed to the classifier as training data labeled as its key code. After the training for all gestures is finished, the classification is started by hitting the space bar. For the formal training such as used in an evaluation and exhibition of the prototype, developers can use an auto-controlled training mode, where the training session for each gesture transits automatically at regular time intervals and the classifier can be trained by equal number of training data in each gesture.

APPLICATIONS

In this section, we demonstrate applications to show the broad range of sensing capabilities of Touch & Activate. We set up the applications using three types of objects: everyday objects, Duplo blocks, and a mobile device.

Everyday Objects

Our approach can make everyday objects touch or grasp sensitive. To show this, we applied Touch & Activate to three objects that have different shapes or material properties.

Ceramic Bowl

First, we made an interactive media art by applying Touch & Activate to a ceramic bowl (8 cm height, 14 cm in diameter). A bowl has several ways to be held depending on the user's manner and context (e.g., if its content is very hot, we may grasp the bowl as shown in Figure 2d). Such manner and context are often utilized in interactive media arts. Touch & Activate allows artists to create an interactive media art using real-world objects rapidly and easily. We trained the system to recognize four touch gestures as shown in Figure 2a-d and associated them with corresponding sounds.

Plastic Toy

Toys are an indispensable part of children's daily life. Interactive toys that can speak or move in response to children's actions are increasingly common. However, they are generally still more expensive than static toys. By using Touch & Activate, users can add interactivity to the existing static toys

[1] http://www.un4seen.com/

Figure 5. Five touch gestures applied to the plastic toy.

Figure 6. Five touch gestures that are applied to the laboratory desk.

that children have and love. As such an example, we applied Touch & Activate to a cheap plastic toy that is normally used as a money box. We trained the system to recognize five touch gestures on the toy and associated corresponding sounds to the positions, giving a new lease of life to an otherwise static object (Figure 5).

Wood Desk

When a large object such as a desk is used, its sensitive area is limited to the small space between the speaker and microphone. This is because the vibration traveling on the object is dampened when the distance is long. We attached the piezo-electric elements to the edge of a laboratory desk (1.5 cm thick, 120 cm wide, 75 cm deep) made of wood at intervals of 15 cm and trained the system to recognize five touch gestures on the desk (Figure 6) and confirmed the gestures can be recognized locally.

Duplo Blocks

Lego or Duplo blocks can be easily assembled to create a variety of models and are often used for rapid prototyping [10]. In our study, we attached two piezo-electric elements to the inner sides of a single piece of Duplo block (2×4, arch shaped) as shown in Figure 7a. We then made four different prototypes using the same piece of Duplo block as a base to add interactivity to the prototypes.

Music Player

First, we prototyped a simple music player that utilizes six studs on the block as discrete buttons representing play, stop, previous, next, volume up, and volume down (Figure 7b). The system recognizes the touch events on the respective buttons and sends corresponding commands to the iTunes music player via Apple Script. The system supports repeat actions in the event of continuous touch.

Interactive Animal Body

DuploTrack [15] infers the assembly arrangement of the Duplo blocks and tracks the assembly process. Our approach, albeit limited, can also estimate the assembly arrangement such as position and direction, as well as the shape of the added block. This is because the added block, depending on its arrangement and shape, produces a change in the boundary condition, hence the resonant property of the object will also

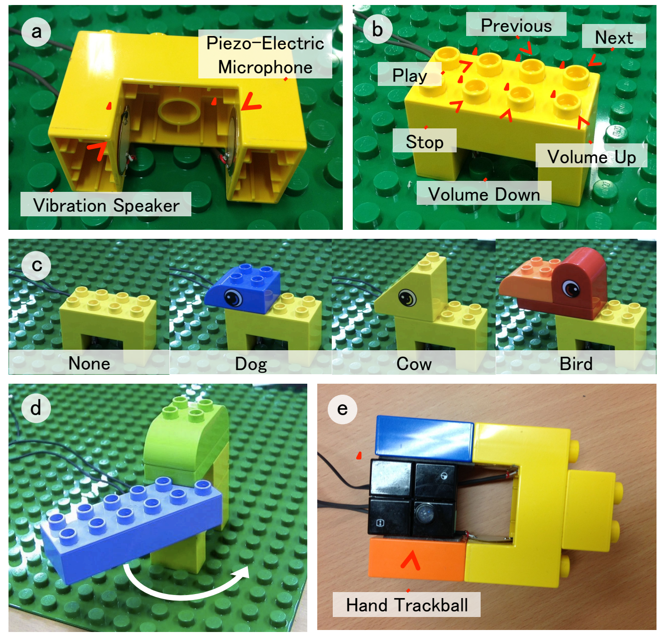

Figure 7. Examples of applications using Duplo blocks: a) a base block attached piezo-electric elements, b) music player, c) interactive animal body, d) six stage lever, and e) multi-functional input device.

change accordingly. We made a prototype of this called interactive animal body (Figure 7c) consisting of one body part, which is the base block, and three other blocks representing three animals' heads (dog, cow, bird). The system was then trained to play a corresponding sound when one of the three heads is added to the body.

Six Stage Lever

As another example of recognition of change in boundary condition without touch or grasp, we made a prototype called six stage lever (Figure 7d). In this prototype, a block which serves as a lever, is attached to the base block. By moving the lever, we can change the base block's boundary condition, thereby changing the resonant property. As the lever can be moved in a continuous manner, the resonant property can also be changed continuously. We trained the system to recognize six distinct angle of the lever with six different labels. This lever can be applied to the previous music player to control the volume.

Multi-Functional Input Device

Other studies have proposed input devices with multi-functions that can be selected by changing the way the device is held [40, 49, 44]. We developed a similar multi-functional input device (Figure 7e) designed for a paint appli-

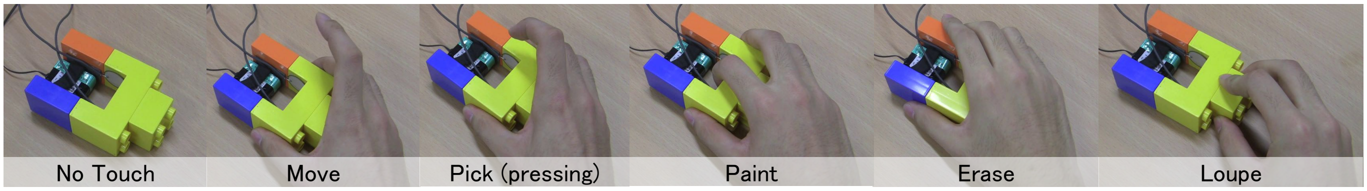

Figure 8. Tools of the multi-functional input device designed for a paint application and corresponding ways to grasp. No Touch) The device is not used. Move) This tool used to just move the mouse pointer is selected by holding out sides of the base block. Pick) By applying pressure to the base block in using move tool, pick mode is selected, with which users can create a selected area and then move the pictures in the area to another space by dragging. Paint) This tool used to paint the canvas is selected by hooking the finger in the base block. Erase) This tool is selected by holding the device deeply. Loupe) This tool can zoom into the canvas around the mouse pointer, which is selected by holding the yellow end of the device.

cation by embedding a super-mobile hand trackball (Sanwa Supply 400-MA018) upside down into the Duplo blocks. The mouse pointer movement is flipped horizontally using USB Overdrive[2]. The device has five input tools that can be selected by changing the way the device is grasped (Figure 8). For example, the user can select the loupe tool by holding the yellow end of the object.

Mobile Device

Many researchers have developed grasp sensitive mobile devices to explore how the user interface of mobile devices can be improved using postures and pressure when holding the device [13, 32, 7, 47, 53]. Most researches used multiple touch sensors to recognize hand postures and pressure. Although GripSense [13] can detect the posture and pressure on commodity mobile devices by using built-in sensors, it relys on the swipe or touch gesture on a touch screen to recognize the postures. Our approach is applicable for recognizing the hand posture and pressure on mobile devices. We made a grasp sensitive case by attaching two piezo-electric elements to the bottom of a transparent hard case for iPhone 4S (Figure 9). As the structure is wobbly when the device is mounted on the case due to the thickness of the two elements and cables, we glued six plastic beads (white ones in Figure 9; 5 mm thick, 5 mm in diameter) on the case to achieve the uniformity in thickness. Then, we tested the prototype to recognize hand posture and pressure.

Posture Recognition

Many applications on recent mobile devices are used with the user holding the device in a unique hand posture. For example, during a call, the user's thumb is on one side while the other four fingers are on the other side of the mobile phone. When using the camera function of the device, users usually hold the four corners of the device with both hands. Therefore, if the holding postures can be recognized, the device can be taught to launch the applications associated with these postures automatically [32, 47]. Additionally, the operational performance and usage of the device differ between different hand postures due to its touch behavior and possibility [4]. If the device can recognize the hand posture of the user, it can adjust the UI accordingly such that maximum ease of use can be achieved (e.g., optimizing keyboard layout and algorithm [4], hand posture based screen rotation [7]). Touch & Activate provides easy hardware configuration to researchers who want to implement these interactions. To test this, we trained the system to recognize six hand postures (Figure 10; determined by referring to [32]) that consists of feasible combinations of three applications (call, sms, camera) and three holding-hand conditions (left handed, right handed, two-handed).

Pressure Recognition

Although a mobile device that has a touch screen as its primary input device allows several multi-touch gestures, single-touch operations using the thumb are still often used because the vast majority of users prefer single-handed interaction with the mobile device [38, 31]. In single-touch operations, however, the interaction with the touch screen is limited because only the thumb can be used. On the other hand, if the grip pressure of the hand can be recognized as a form of control, single-handed interaction with the mobile device will improve (e.g., pressure based zoom-in/out [35, 13], menu selection [50], non-verbal communication during phone call [26]). Touch & Activate can also recognize the grip pressure on the mobile device because applying the pressure changes the boundary condition of the device, and it can be used to implement these interactions. To test this, we trained the system to recognize three levels of pressure (i.e., support, hold, and grasp as shown in Figure 11) and confirmed the levels can be recognized. This example application shows that our technique also can be used to test these pressure based interactions without additional pressure sensors.

[2] http://www.usboverdrive.com/USBOverdrive/Information.html

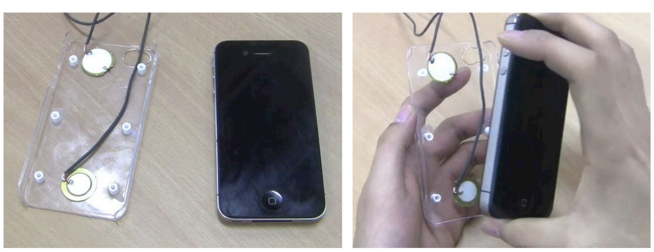

Figure 9. iPhone 4S and a grasp sensitive case with Touch & Activate.

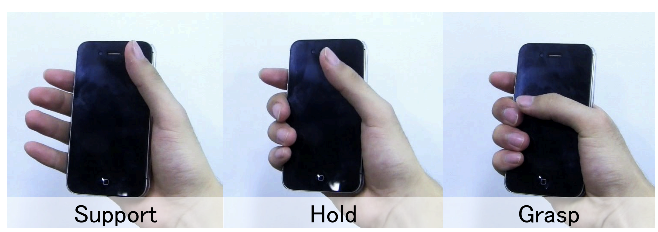

Figure 11. Three hand postures for pressure recognition. The "grasp" posture is grasping the device applying pressure strongest.

Figure 10. Six hand postures for posture recognition.

LIMITATIONS

Although Touch & Activate is applicable to various objects, our approach could not prototype fully touch-enabled objects due to the instrumentation. Moreover, it has limitations in material, size, and shape of objects; and noise will affect the recognition in some cases.

Materials

As Touch & Activate uses the vibration of objects for its sensing, the material of objects that are applied to need to conduct the vibration efficiently. In general, elastic modulus, a measure of the stiffness of materials, affects the conductivity of vibration of objects. If a material exhibits lower elastic modulus (such as gel, clay, and rubber), our approach is difficult to apply to it because this characteristic damps the vibration. In contrast, we confirmed that a material exhibiting high elastic modulus (such as wood, metal, plastic, and ceramic) generally works well.

Size

Even in the material that conducts the vibration efficiently, the vibration is dampened with the distance from the source. Thus, if the objects are large, the sensitive area is limited to a small area between the speaker and microphone as shown in the wood desk application. Extending the sensitive area requires extra sensors (i.e., speakers or microphones) or a more powerful amplifier. However, in our tests so far, most handheld-sized objects can be made touch and grasp sensitive by using a casual amplifier (e.g., consumer USB audio interfaces) and a single speaker and microphone pair.

On the other hand, if the objects are too small, there is little interaction space and developers might feel that the instrumentation is obtrusive in the prototyping. To remedy this issue, we recommend using smaller piezo elements as a sensor, because the surface area of the elements barely affects the sensitivity of the high frequency component. We confirmed even 15-mm-diameter piezo elements work well in our test.

Shape

The shape of objects also limits the sensing capability in some cases. For instance, a symmetrical object cannot recognize as many gestures as an asymmetrical object, where, due to its symmetry, the same resonant frequency is provided by the symmetric touch gestures. In such case, developers must displace the sensor from the axis of symmetry if they want to distinguish between the gestures. Moreover, if the objects that do not have clear touch points due to having too complex a shape (e.g., spiked surfaces), our approach is difficult to apply to them because our approach requires a clear change in the boundary condition.

Noises

In our technique, noise will affect the accuracy in some cases. When the speaker and microphone are placed close together, especially if they are on the same surface, the accuracy might be lowered because the reflected sounds with Doppler shifts are caused by moving hands or some objects passing above the sensors and affect the frequency response captured from the microphone. To avoid this, both piezo elements should be placed sufficiently apart, or the microphone should be covered with some material such as rubber to prevent it from capturing the noise caused by Doppler shifts.

Our current implementation of Touch & Activate cannot recognize touch gestures with certainty when a high impact is applied to the object. This is because a consumer audio hardware usually clips such sounds from the microphone and the high frequency in the signal, which is necessary to obtain the resonant property, is lost. In other words, if an audio hardware with a wide dynamic range is used, our approach can be used with a passive one (e.g., TapSense [21]) and further interactions combined with impacts, such as "tap and hold", can be detected, because our approach uses a very high frequency, which is not affected by tapping.

However, our approach is robust in noisy environments because little noise in air can be transformed as vibration of the object and the most of the noise is reflected. Due to this robustness, our system worked successfully when we demonstrated Touch & Activate in a noisy hall of an academic conference, where 65 demonstrations were presented simultaneously and over 400 people participated and had discussions.

EVALUATION

For our evaluation, we conducted a controlled experiment to measure the accuracy and trade-off between the accuracy and number of training rounds using the two applications: the plastic toy as a simple example and the posture recognition as a complex example.

Participants

Our experiment had 10 participants (seven males, three females) ranging in age from 22 to 32 years old. All participants were smartphone users and had used the device from between one week to three years.

Procedure

We collected the data for our evaluation by asking each participant to perform the following task for each application independently (i.e., we used the auto-controlled training mode described in the Software section).

Participants were shown a picture of one of the gestures selected from the gesture set (five touch gestures for the plastic

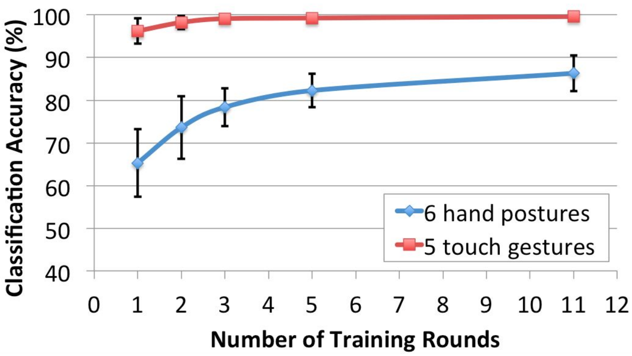

Figure 12. Accuracy using a per-user classifier.

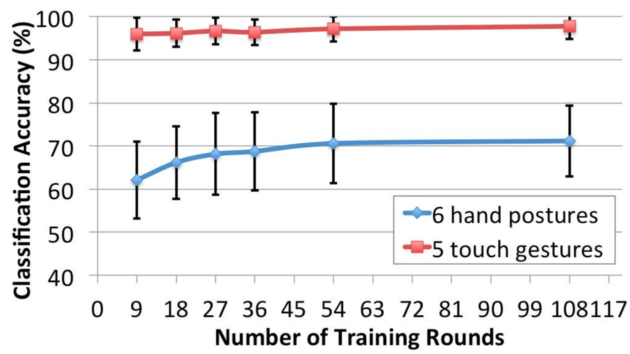

Figure 13. Accuracy using a general classifier.

toy shown in Figure 5 and six hand postures for the posture recognition shown in Figure 10) for seven seconds in a randomized order and asked to perform the gestures. While performing a gesture, the participants were also asked to continuously adjust the gesture's posture to make the system give the same label to slightly different postures of the gesture. This is necessary because our technique is so sensitive that it can detect even subtle differences in postures that the participants cannot. In the last second, because the participants were performing the gestures stably now, we collected the 40 pieces of frequency response data, which is representing the feature vectors in the last second. Seven seconds after seeing the previous picture, the participants were shown the next picture. This procedure was repeated until all gestures were performed (one round). Each participant performed 12 rounds, yielding 480 pieces of data per user per gesture.

In total, 24000 pieces of data (10 participants $\times$ 5 gestures $\times$ 12 rounds $\times$ 40 data) for the plastic toy and 28800 pieces of data (10 participants $\times$ 6 postures $\times$ 12 rounds $\times$ 40 data) for the posture recognition were collected, respectively. The entire experiment took approximately 25 minutes per participant (seven minutes for the plastic toy, nine minutes for the posture recognition, and nine minutes for the explanation and questionnaire) to complete.

RESULTS

By using the data, we evaluated two types of classifier for each application: per-user classifiers and general classifiers. We used a LIBSVM (RBF kernel, c=1.0, g=0.0025), which is the same SVM library used in our test system, in all training and classification processes.

Per-User Classifiers

To understand the feasibility of Touch & Activate for individual users, we assessed the accuracy of user-specific classifiers. We trained the classifier with 11 rounds of data collected from a single participant, and tested on the rest round data. All train/test combinations were tested and averaged per participant (i.e., 12-fold round-independent cross validation). This process was also tested for each participant's data and we averaged the accuracies. In the results, the plastic toy including five touch gestures achieved an accuracy of 99.6% (SD=0.36%). On the other hand, the posture recognition including six hand postures achieved an accuracy of 86.3% (SD=4.2%).

To understand the trade-off between the accuracy and number of training rounds for individual users, we also tested the accuracy when the system is trained with different amounts of data. To do this, we first divided the per-participant data into two, three, four, and six segments (e.g., we divided the per-participant data consisting of 12 rounds data into two segments: first-sixth rounds of data and seventh-twelfth rounds of data). Then, six-, four-, three-, and two-fold cross validations were performed on both segments. The results were averaged and are plotted in Figure 12. The plastic toy with only one round of data achieved an accuracy of 96.2% (SD=3.0%). The posture recognition with one round of data maintained an accuracy of 65.3% (SD=7.9%). By using five rounds of data, the accuracy jumped to 82.3% (SD=3.9%).

General Classifier

To understand the feasibility of Touch & Activate for "walk up" users, we assessed the accuracy of a general classifier. We trained the SVM with nine participants' data and tested on the other participant's data (all-combinations, i.e., 10-fold participant-independent cross validation). In the results, for the plastic toy achieved 97.8% accuracy (SD=2.9%) that is about 2% lower than per-user classifiers. On the other hand, the posture recognition maintained 71.2% accuracy (SD=8.2%).

We calculated the trade-off curves between the accuracy and number of training rounds (Figure 13). Note that 10-fold participant-independent cross validation was performed on all results with different numbers of rounds of data while six-, four-, three-, and two-fold round-independent cross validations were performed on per-user classifier's results. In the results, the plastic toy keeps a high accuracy of 96.0% (SD=3.8%) even with nine rounds of data (one round of data per participant). On the other hand, the posture recognition maintained 62.1% accuracy (SD=8.9%) with nine rounds of data and accuracy mostly stopped improving at around 54 rounds of data (five rounds of data per participant).

DISCUSSION & FUTURE WORK

As the results of our experiment showed a promising accuracy for the recognition of touch gestures and hand postures, Touch & Activate should be feasible for prototyping interactive objects that have touch input capability. Especially for a simple application such as the plastic toy, Touch & Activate provides a very rapid prototyping experience using existing objects. Note that, as Figure 12 shows, training with only one round of data resulted in 96.2% accuracy in the plastic toy, which is high enough for developers to test their ideas by themselves, and collecting one round of data requires only 35 seconds (7 seconds $\times$ 5 gestures). Even for a complex application such as the posture recognition, the touch or grasp input should be available with a promising accuracy of 86.3% after a few minutes of training; collecting 11 rounds of data takes 7.7 minutes (7 seconds $\times$ 6 postures $\times$ 11 rounds), which is much shorter than time spent designing circuits and configuring hardware using multiple sensors. Note that in our experiment, we presented a gesture for seven seconds to ensure that the participants could understand the gesture and perform it without fail. In practice, less time is necessary to train the system to recognize a gesture (it was approximately three seconds for each gesture in our development), supporting rapid prototyping.

Resonant property of solid objects, which is used for touch sensing in Touch & Activate, is highly sensitive to changes in the boundary condition. To stabilize the recognition, developers have to continuously adjust the touches while training the system to recognize the gestures. Moreover, when the system is trained once, the piezo elements positions must be fixed to recognize the gestures using those training data because the resonant property also differs even if the relative positions of sensors remain unchanged.

While we still cannot find any general rule for the sensor's placement with which Touch & Activate provides the best performance for the recognition, there are some rules to avoid limitations: 1) avoid the axis of symmetry in the symmetric objects, 2) place them sufficiently apart when Doppler shifts occur, and 3) place them in an area that do not needed to be touched. In future work, we plan to explore the general rules of placement.

Although our primary focus is on providing easy hardware configuration for adding touch input capability to existing objects, prototyping these objects still requires some programming for implementing actions, whose processes are difficult for non-engineers. In future work, we plan to remove this difficulty by adopting easy authoring methods proposed in other works (e.g., record-and-replay [24, 28, 42] and visual programming [25]) to help developers implement actions.

In our current implementation, the portability of the objects with Touch & Activate is restricted because of the wired connection to the computer. Since this might discourage developers from stretching their imagination in designing the interaction, we plan to implement the wireless sensor module of Touch & Activate. The module will be the one similar to the wireless Swept Frequency Capacitive Sensing module of Touché [41] with an additional amplifier to vibrate the objects; hence, its running time will be limited due to the power consumption of the amplifier. Therefore, we also plan to measure the trade-off between the degree of the amplification (i.e., power consumption) and the accuracy in sensing. However, we believe that power consumption would not impose any problem even in the wireless setup because our technique requires short time to train a gesture for rapid prototyping.

CONCLUSIONS

We presented a novel acoustic touch sensing technique called *Touch & Activate*. It can detect a rich context of touches including grasp and changes in other boundary conditions on existing objects, by attaching only a vibration speaker and a piezo-electric microphone paired as a sensor. This provides easy hardware configuration for prototyping interactive objects that have touch input capability. We demonstrated with a number of applications that our approach's broad range of sensing capability can be used in prototyping a variety of interactive objects. We conducted a controlled experiment to measure the accuracy and trade-off between the accuracy and number of training rounds of our technique. From its results, per-user recognition accuracies with five touch gestures for the plastic toy and six hand postures for the posture recognition were 99.6% and 86.3%, respectively. Walk up user recognition accuracies were 97.8% and 71.2%, respectively. This suggests Touch & Activate should be feasible and accessible for prototyping interactive objects using existing objects.

REFERENCES

1. Arduino - HomePage. http://www.arduino.cc/, Accessed: June 2013.
2. mbed - Rapid Prototyping for Microcontrollers. http://mbed.org/, Accessed: June 2013.
3. Amento, B., Hill, W., and Terveen, L. The sound of one hand: a wrist-mounted bio-acoustic fingertip gesture interface. In *CHI EA '02*, 724–725.
4. Azenkot, S., and Zhai, S. Touch behavior with different postures on soft smartphone keyboards. In *MobileHCI '12*, 251–260.
5. Brenner, M.C., and Fitzgibbon, J.J. Surface acoustic wave touch panel system, Feb.17 1987. US Patent 4,644,100.
6. Chang, C.-C., Lin, C.-J. LIBSVM: a library for support vector machines., http://www.csie.ntu.edu.tw/cjlin/livsvm
7. Cheng, L.-P., Hsiao, F.-I., Liu, Y.-T., and Chen, M.Y. iRotate grasp: automatic screen rotation based on grasp of mobile devices. In *UIST '12 Adjunct*, 15–16.
8. Christopher, D. Frequency Range of Human Hearing. In *The Physics Factbook*. 2003.
9. Collins, T. Active acoustic touch interface. *Electronics Letters 45*, 20 (24 2009), 1055 –1056.
10. Cralg, L. Right–rapid–rough. *Ask Magazine*, 13 (2003), 12–15.
11. Ewins, D.J. *Modal testing: theory, practice and application*, vol.2. Research studies press Baldock, 2000.
12. Fourney, A., and Terry, M. PICL: portable in-circuit learner. In *UIST '12*, 569–578.
13. Goel, M., Wobbrock, J., and Patel, S. GripSense: using built-in sensors to detect hand posture and pressure on commodity mobile phones. In *UIST '12*, 545–554.
14. Greenberg, S., and Fitchett, C. Phidgets: easy development of physical interfaces through physical widgets. In *UIST '01*, 209–218.
15. Gupta, A., Fox, D., Curless, B., and Cohen, M. DuploTrack: a real-time system for authoring and guiding duplo block assembly. In *UIST '12*, 389–402.
16. Gupta, S., Morris, D., Patel, S., and Tan, D. SoundWave: using the Doppler effect to sense gestures. In *CHI '12*, 1911–1914.

17. Han, J.Y. Low-cost multi-touch sensing through frustrated total internal reflection. In *UIST '05*, 115–118.
18. Harrison, C., Benko, H., and Wilson, A.D. OmniTouch: wearable multitouch interaction everywhere. In *UIST '11*, 441–450.
19. Harrison, C., and Hudson, S.E. Scratch input: creating large, inexpensive, unpowered and mobile finger input surfaces. In *UIST '08*, 205–208.
20. Harrison, C., Sato, M., and Poupyrev, I. Capacitive fingerprinting: exploring user differentiation by sensing electrical properties of the human body. In *UIST '12*, 537–544.
21. Harrison, C., Schwarz, J., and Hudson, S.E. TapSense: enhancing finger interaction on touch surfaces. In *UIST '11*, 627–636.
22. Harrison, C., Tan, D., and Morris, D. Skinput: appropriating the body as an input surface. In *CHI '10*, 453–462.
23. Harrison, C., Xiao, R., and Hudson, S. Acoustic barcodes: passive, durable and inexpensive notched identification tags. In *UIST '12*, 563–568.
24. Hartmann, B., Abdulla, L., Mittal, M., and Klemmer, S.R. Authoring sensor-based interactions by demonstration with direct manipulation and pattern recognition. In *CHI '07*, 145–154.
25. Hartmann, B., Klemmer, S.R., Bernstein, M., Abdulla, L., Burr, B., Robinson-Mosher, A., and Gee, J. Reflective physical prototyping through integrated design, test, and analysis. In *UIST '06*, 299–308.
26. Hoggan, E., Stewart, C., Haverinen, L., Jacucci, G., and Lantz, V. Pressages: augmenting phone calls with non-verbal messages. In *UIST '12*, 555–562.
27. Holman, D., and Vertegaal, R. TactileTape: low-cost touch sensing on curved surfaces. In *UIST '11 Adjunct*, 17–18.
28. Hudson, S.E., and Mankoff, J. Rapid construction of functioning physical interfaces from cardboard, thumbtacks, tin foil and masking tape. In *UIST '06*, 289–298.
29. Hwang, S., Bianchi, A., and Wohn, K. MicPen: pressure-sensitive pen interaction using microphone with standard touchscreen. In *CHI EA '12*, 1847–1852.
30. Izadi, S., Kim, D., Hilliges, O., Molyneaux, D., Newcombe, R., Kohli, P., Shotton, J., Hodges, S., Freeman, D., Davison, A., and Fitzgibbon, A. KinectFusion: real-time 3D reconstruction and interaction using a moving depth camera. In *UIST '11*, 559–568.
31. Karlson, A.K., and Bederson, B.B. Understanding single-handed mobile device interaction. Tech. rep., Dep. of Computer Science, Univ. of Maryland, 2006.
32. Kim, K.-E., Chang, W., Cho, S.-J., Shim, J., Lee, H., Park, J., Lee, Y., and Kim, S. Hand grip pattern recognition for mobile user interfaces. In *IAAI '06*, 1789–1794.
33. Liu, S., and Guimbretière, F. FlexAura: a flexible near-surface range sensor. In *UIST '12*, 327–330.
34. Lopes, P., Jota, R., and Jorge, J.A. Augmenting touch interaction through acoustic sensing. In *ITS '11*, 53–56.
35. Miyaki, T., and Rekimoto, J. Graspzoom: zooming and scrolling control model for single-handed mobile interaction. In *MobileHCI '09*, 11.
36. Murray-Smith, R., Williamson, J., Hughes, S., and Quaade, T. Stane: synthesized surfaces for tactile input. In *CHI '08*, 1299–1302.
37. Paradiso, J.A., Leo, C.K., Checka, N., and Hsiao, K. Passive acoustic knock tracking for interactive windows. In *CHI EA '02*, 732–733.
38. Parhi, P., Karlson, A.K., and Bederson, B.B. Target size study for one-handed thumb use on small touchscreen devices. In *MobileHCI '06*, 203–210.
39. Pham, D., Ji, Z., Yang, M., Wang, Z., and Al-Kutubi, M. A novel human-computer interface based on passive acoustic localisation. In *HCII '07*, 901–909.
40. Rekimoto, J., and Sciammarella, E. ToolStone: effective use of the physical manipulation vocabularies of input devices. In *UIST '00*, 109–117.
41. Sato, M., Poupyrev, I., and Harrison, C. Touché: enhancing touch interaction on humans, screens, liquids, and everyday objects. In *CHI '12*, 483–492.
42. Savage, V., Zhang, X., and Hartmann, B. Midas: fabricating custom capacitive touch sensors to prototype interactive objects. In *UIST '12*, 579–588.
43. Schwarz, B.J., and Richardson, M.H. Experimental modal analysis. *CSI Reliability week* (1999).
44. Song, H., Benko, H., Guimbretière, F., Izadi, S., Cao, X., and Hinckley, K. Grips and gestures on a multi-touch pen. In *CHI '11*, 1323–1332.
45. Stewart, J., and Westerfield, E. A theory of active sonar detection. *Proceedings of the IRE 47*, 5 (1959), 872–881.
46. Takemura, K., Ito, A., Takamatsu, J., and Ogasawara, T. Active bone-conducted sound sensing for wearable interfaces. In *UIST '11 Adjunct*, 53–54.
47. Taylor, B.T., and Bove, Jr., V.M. Graspables: grasp-recognition as a user interface. In *CHI '09*, 917–926.
48. Villar, N., Scott, J., Hodges, S., Hammil, K., and Miller, C. . NET Gadgeteer: a platform for custom devices. *Pervasive Computing* (2012), 216–233.
49. Vogel, D., and Casiez, G. Conté: multimodal input inspired by an artist's crayon. In *UIST '11*, 357–366.
50. Wilson, G., Stewart, C., and Brewster, S.A. Pressure-based menu selection for mobile devices. In *MobileHCI '10*, 181–190.
51. Wimmer, R. FlyEye: grasp-sensitive surfaces using optical fiber. In *TEI '10*, 245–248.
52. Wimmer, R., and Baudisch, P. Modular and deformable touch-sensitive surfaces based on time domain reflectometry. In *UIST '11*, 517–526.
53. Wimmer, R., and Boring, S. HandSense: discriminating different ways of grasping and holding a tangible user interface. In *TEI '09*, 359–362.