Ubiquitous Character Input Device Using Multiple Acoustic Sensors on a Flat Surface

Akira Urashima*　　Tomoji Toriyama†

Toyama Prefectural University

ABSTRACT

We are developing a method of using existing flat surfaces such as table-tops as drawing input devices by attaching acoustic sensors. As this device employs acoustic point sensors rather than sheet type sensors and does not require special pens, it does not wrap the surface and can be carried anywhere and easily installed. In this paper, we describe the drawing acquisition method and the preliminary results of our experimental system. In our experiment with a satin-finished surface table, we estimated the tapping position with an offset error of 1.3 cm and standard deviation of 0.21 cm in a 30 cm × 30 cm region. Furthermore, we attempted automatic recognition of a few uppercase characters traced on the table: 30 of 45 input characters were correctly recognized. This is a promising result for the proposed character input device.

Index Terms: H.5.2 [Information Interfaces and presentation]: User Interfaces—Input devices and strategies

1 INTRODUCTION

In collaborative working and teaching style, quick rough sketches and scribbles are a useful communication method. Many types of input devices use continuous pen position on a flat surface (*e.g.*, a graphics tablet or interactive white board). However some of these devices adopt sheet type sensors that cover the entire input surface, thus, making it difficult to install on existing flat surfaces such as tables or walls. Other devices require a special pen, without which drawing on the surface is ignored.

We are developing a method that uses existing flat surfaces as input devices for sketching and writing by adding acoustic sensors. Drawing a line on flat surface causes acoustic vibration when the pen contacts the surface or moves. We install more than three acoustic sensors on a surface and analyze their signals to estimate the current position of the tip of the pen. The advantage of this method is that it does not wrap the surface, while keeping the merit of the needlessness of a special pen and the easiness of carrying and installation.

Some research has used acoustic sensors as pointing input devices, but studies of acoustic sensors as drawing input devices are limited, and their capabilities as character input devices have not been investigated.

In this paper, we describe a method of estimating the position of traced lines on a flat surface by acoustic sensors. Then we investigate the precision of pointing and the accuracy of alphabetic character recognition.

2 RELATED WORKS

Pham *et al*. [1, 2] described Location Template Matching (LTM) approach on the basis of the idea that the received signals carry its source location information as a result of position-dependent scattering and reflection in the medium. They adopted correlation with training signals at pre-known locations and achieved a resolution of 10 cm experimentally.

Harrison *et al*. [3] estimated the location of finger taps on the arm and hands by analyzing vibrations that propagate through the body. They extracted 186 features from the signals of 10 sensors wrapped around the arm, and passed the features to a pre-trained support vector machine (SVM) classifier that was trained to distinguish 5 to 10 different tapping locations on the arm.

The advantage of these training data based methods is that they can be applied to objects that have other than a flat surface. On the other hand, they require training data corresponding to known positions, and the spatial resolution is not so fine enough to reconstruct the drawn lines.

Paradiso *et al*. [4] installed an array of four sensors at the each corner of a glass window and located the position of a knock using differences in the times of the wavefront arrival. Pham *et al*. [5, 2] also used an array of four sensors at the each corner of thin medium density fiberboard (MDF) and estimated the signal source location from the overlaid distribution of the correlation of two wavelet-transformed signals. They also showed a screen shot of the continuous tracking of a finger scratching the MDF.

The advantage of these computation-based methods is that the spatial resolution of the signal source position is higher than in training data based methods. Thus, they can be used as not only a sketch input device but also a character input device. However, the use of the acoustic position estimation on a flat surface as a character input device has not been investigated.

Harrison *et al*. [6] described the recognition method of scratching gestures on a surface material by one microphone. Although the gestures included alphabetic characters, the number of gestures was too few to cover all characters.

3 ESTIMATION METHOD

Line sketching consists of three steps: lowering the pen to the surface (pen-down), motion of the pen (pen-move), and lifting the pen (pen-up). We describe the detection and position estimation algorithm for each step in this section. With adequate resolution, alphabetic characters are easily recognized because they consist of a few line strokes.

3.1 Pen-down

Pen-down makes a single impact on the surface, which propagates as concentric circles when the velocity is isotropic. Thus, the time difference in signal arrival at two separate sensor positions indicates the difference in the distance between each sensor position and the signal's source. With more than three sensors, there are three or more pieces of information available on the difference in distance, making it possible to estimate the position of the signal source.

When the time difference in signal arrival at (x_i, y_i) and (x_j, y_j) is τ_{ij}, the position of the signal source (x, y) obeys the relation:

$$\tau_{ij} = \sqrt{\frac{(x-x_i)^2}{v_x^2} + \frac{(y-y_i)^2}{v_y^2}} - \sqrt{\frac{(x-x_j)^2}{v_x^2} + \frac{(y-y_j)}{v_y^2}} \quad (1)$$

*e-mail: a-urasim@pu-toyama.ac.jp

†e-mail: toriyama@pu-toyama.ac.jp

20th International Conference on Artificial Reality and Telexistence (ICAT2010), 1-3 December 2010, Adelaide, Australia

ISBN: 978-4-904490-03-7 C3450

where v_x and v_y are the velocities of the signal in the x and y directions, respectively. Eq. (1) shows that the possible signal source position (x,y) draws a hyperbola on surface.

However, the time of signal arrival cannot be clearly determined because of attenuation and noise. Instead, we use the cross-correlation of two signals as a probability function of the time difference.

$$C_{ij}(\tau) = \int s'_i(t+\tau)s'_i(t)dt \tag{2}$$

$$s'_i(t) = s_i(t)w(t-t_0) \tag{3}$$
$$s'_j(t) = s_j(t)w(t-t_0) \tag{4}$$

Here, τ is the difference in arrival time, $s_1(t), s_2(t)$ are the signals at each sensor position, and $w(t)$ is the window function that limits the signal time span to the proximity of t_0.

To reduce the computational complexity, Eq. (2) is transformed as follows:

$$C_{ij}(\tau) = \mathscr{F}^{-1}[\mathscr{F}[s'_i(-t)]\mathscr{F}[s'_j(t)]] \tag{5}$$

Combining Eqs. (1) and (5) yields a scalar distribution on the surface. By summing up the distributions calculated from all pairs of signals, we can obtain the total distribution $I(x,y)$ which shows the degree of probability where the signal source exists.

$$I(x,y) = \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} C_{ij}(\tau_{ij}) \tag{6}$$

The estimated signal source position (x_e, y_e) can be calculated from $I(x,y)$ as follows:

$$(x_e, y_e) = \underset{(x,y)}{\operatorname{argmax}}\ I(x,y) \tag{7}$$

3.2 Pen-move and pen-up

The motion of a pen on a rough surface makes a scratching vibration owing to micro impacts of the pen on the surface. Thus, not only pen-down but also pen-move can be easily distinguished from the background noise by setting a threshold. However, the estimated position of a micro impact source is not very accurate because of signal weakness and overlap between signals directly arrived and reflected at the edge of the surface.

To improve the accuracy, we introduce two restrictions on position estimation. One is that the position of the pen does not jump, that is, in pen-move, estimation results that indicate a position distant from the last estimated position are eliminated.

The other restriction is that the movement of the pen changes gradually. This can be done by a simple Kalman filter that includes the equation of motion.

Here, $\mathbf{x} = (x, v_x, y, v_y)$ represents the pen's position and velocity. The prediction phase of the Kalman filter is as follows:

$$\begin{aligned}\hat{\mathbf{x}}_{k|k-1} &= \mathbf{F}\hat{\mathbf{x}}_{k-1|k-1} \\ &= \begin{pmatrix} 1 & \Delta t & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \Delta t \\ 0 & 0 & 0 & 1 \end{pmatrix} \hat{\mathbf{x}}_{k-1|k-1}\end{aligned} \tag{8}$$

$$\mathbf{P}_{k|k-1} = \mathbf{F}\mathbf{P}_{k-1|k-1}\mathbf{F}^T + \begin{pmatrix} \frac{\Delta t^4}{4} & \frac{\Delta t^3}{2} & 0 & 0 \\ \frac{\Delta t^3}{2} & \Delta t^2 & 0 & 0 \\ 0 & 0 & \frac{\Delta t^4}{4} & \frac{\Delta t^3}{2} \\ 0 & 0 & \frac{\Delta t^3}{2} & \Delta t^2 \end{pmatrix} \sigma_a^2 \tag{9}$$

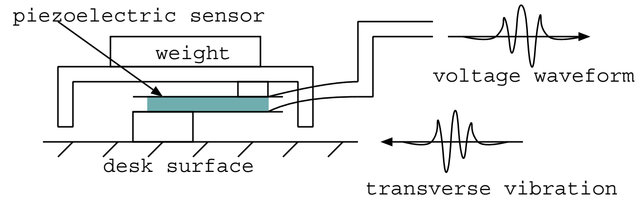

Figure 1: Structure of piezoelectric sensor for surface acoustic vibrations.

The update phase of Kalman filter is as follows:

$$\begin{aligned}\tilde{\mathbf{y}}_k &= \mathbf{z}_k - \mathbf{H}\hat{\mathbf{x}}_{k|k-1} \\ &= \mathbf{z}_k - \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \hat{\mathbf{x}}_{k|k-1}\end{aligned} \tag{10}$$

$$\mathbf{S}_k = \mathbf{H}\mathbf{P}_{k|k-1}\mathbf{H}^T + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \sigma_z^2 \tag{11}$$

$$\mathbf{K}_k = \mathbf{P}_{k|k-1}\mathbf{H}^T\mathbf{S}_k^{-1} \tag{12}$$

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k\tilde{\mathbf{y}}_k \tag{13}$$

$$\mathbf{P}_{k|k} = (\mathbf{I} - \mathbf{K}_k\mathbf{H})\mathbf{P}_{k|k-1} \tag{14}$$

Here, $\mathbf{z} = (x_e, y_e)$ is the observation value of the Kalman filter, which is the estimated position of the signal source in Eq. (7).

A pen-up is detected when pen-move is there is absent for a certain period. This prevents the estimated point from drifting owing to the remaining velocity.

4 EXPERIMENTS

To check the character recognition capability, we experimentally implemented a character input device using multiple acoustic sensors. We first investigated the accuracy of the pen-down position estimation. As the pen-down positions become the starting points of the stroke of the input character, their error margin is related to the recognizable size of the character. Next, we checked the character recognition capability of the experimental implementation to obtain the preliminary knowledge of the accuracy and the causes of error.

4.1 Experimental implementation

The hardware used in our implementation consists of four acoustic sensors and a processing PC. Figure 1 shows the structure of an acoustic sensor for transverse vibrations on a flat surface. The sensors are connected to the sound card on the PC via microphone amps. Data from the four sensors are synchronously sampled at a rate of 44,100 Hz and processed as shown in Figure 2. A pen position is calculated from 512 samples of each sensor using a sliding window technique with 384 sample overlap.

For the character recognition, we use Microsoft.ink.analysis.dll, which is the handwriting recognition engine for the Tablet PC in Windows SDK. Character recognition is limited to uppercase characters in the experiment.

4.2 Experimental conditions

We installed the experimental implementation system on a flat tabletop which has 180 cm $\times$ 90 cm in area (Figure 3). The tabletop is made of melamine-coated MDF and has a satin-finished surface. Four piezoelectric sensors are installed, one at each corner of a 30 cm $\times$ 30 cm square region on the table. In this region, the grid lines are drawn at intervals of 5 cm to visually confirm the pen's position.

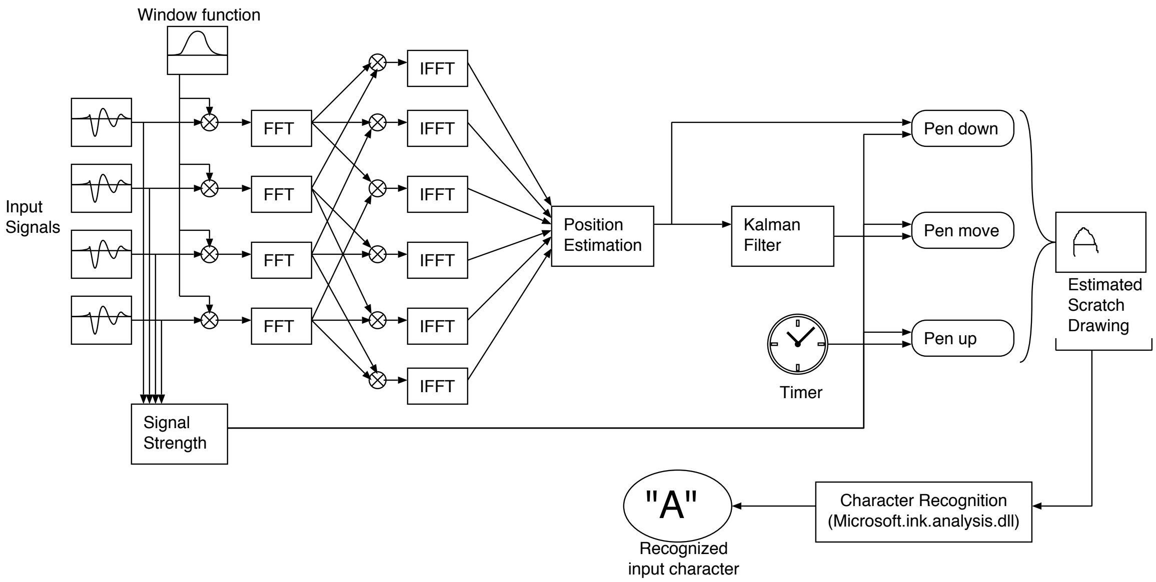

Figure 2: Functional structure of proposed character input device.

We measured velocities on the table (v_x, v_y) (377 m/s, 296 m/s) in a prior experiment.

We performed three sets of the pen-down tests. One set consisted of three pen-downs for every grid point. The estimated point of pen-down is evaluated by the following expressions.

$$E_{\text{offset}} = \frac{\sum_i \sum_j |p^i_j - p^i|}{nm} \quad (15)$$

Here, p^i_j is the estimated location of pen-down at the ith grid point for the jth try, p^i is the location of the ith grid point, m is the number of tries, and n is the number of the grid points.

$$E_{\text{stddev}} = \sqrt{\frac{\sum_i \sum_j |p^i_j - (\sum_k p^i_k)/m|^2}{nm}} \quad (16)$$

E_{offset} indicates the offset error from the physical position, and E_{stddev} indicates the inconsistency of the estimated position at each trial. Thus, larger the E_{offset} and E_{stddev} values, greater is the difficulty to stroke the pen at the expected point.

Next, we tested character recognition. Five participants wrote three characters ("T," "P," and "U") three times each with the pen shown in Figure 4. The pen is a simple pointer stick because it has a hard tip but does not mark the table. The line strokes of the character drawings are stored in the PC and processed by the character recognition program. The recognition results are judged by comparing with the intended character.

4.3 Results

Figure 5 shows the result of the pen-down test. The clusters of estimated positions are aligned with the grid points. E_{offset} is 1.3 cm, and E_{stddev} is 0.21 cm.

In the character recognition test, 30 of 45 input characters were correctly recognized. The recognition results according to the character and the participant are shown in Tables 1 and 2, respectively. Figure 6 shows a sample of correctly recognized strokes.

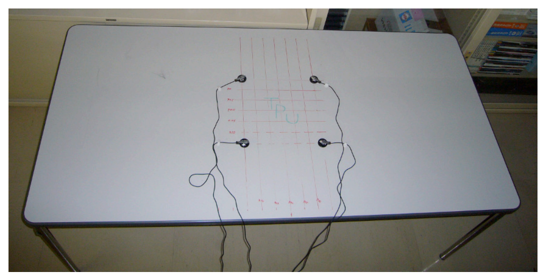

Figure 3: Table with sensors installed.

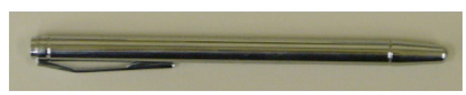

Figure 4: Pen for drawing on the table.

5 Discussion

The values of E_{offset} and E_{stddev} show that a difference in position of 1 cm or less cannot be distinguished. Considering characters composed of adjoining multiple strokes such as "M" and "W," the size of an input uppercase character should be six times or more the values of E_{offset} and E_{stddev}. So it is suggested that the size of an input uppercase character should be selected 6 cm or more.

The recognition results shown in Tables 1 and 2 have a bias

Table 1: Character recognition by character.

Character	"T"	"P"	"U"	total
Correct	13	9	8	30
Wrong	2	6	7	15

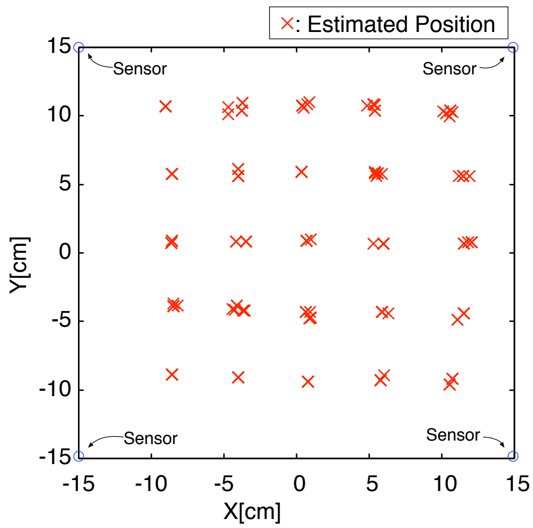

Figure 5: Estimated pen-down positions.

Table 2: Character recognition by participant.

Participant	(A)	(B)	(C)	(D)	(E)	total
Correct	8	8	6	2	6	30
Wrong	1	1	3	7	3	15

resulting from idiosyncrasies of participants and the existence of similar characters. We analyzed the incorrect results and classified them according to the following causes of error.

(a) Some or all the strokes of the pen are missing, and we cannot identify the character visually.
Five incorrectly recognized characters are included here. A sample of this type is shown in Figure 7.

(b) Some strokes are missing, but we can identify the character visually. Four incorrectly recognized characters are included here. A sample of this type is shown in Figure 8.

(c) Some strokes are distorted, and we cannot distinguish the character from a similar character visually. Four incorrectly recognized characters are included here. A sample of this type is shown in Figure 9.

(d) All the strokes exist, and we can identify the character visually. Two incorrectly recognized characters are included here. A sample of this type is shown in Figure 10.

Items (b) and (d) can be resolved by adjusting the character recognition program because there is enough information to identify the input character in the estimated stroke. For (c), recognition can be improved by introducing a time profile of the amplitude of the scratching vibration to the character recognition program. For example, "U" has one continuous vibration, whereas "V" has two separate vibration. In total, 10 of the 15 incorrect identifications can be rectified by improving the character recognition program.

In this paper, we performed the experiment in only a single environment to examine the possibility of our method. However, in the real situation, there are many kinds of elements which influence the precision of our method. For example, roughness of surfaces shall affect the strength of scratching vibration. The sensor near the edge of the table can capture the reflected vibration also. Heavy object

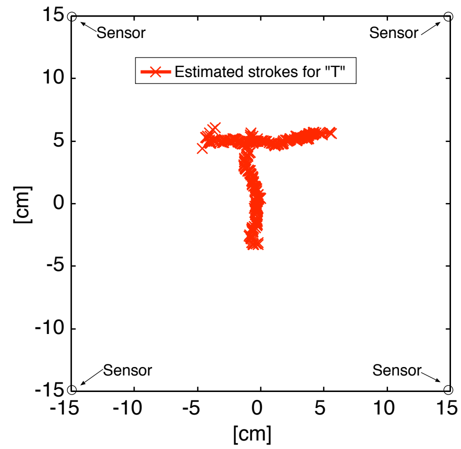

Figure 6: Estimated strokes of character "T," which is correctly recognized as "T."

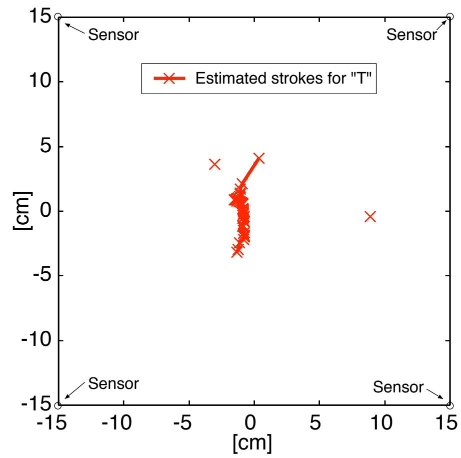

Figure 7: Estimated strokes of character "T," which is not recognized as any character.

on the surface may block vibration transmission. To investigate the effects of these elements by the experiments in several conditions is our future work.

6 Conclusion

We developed a character input device that uses only four acoustic sensors on a flat surface. We evaluated the precision of pen-down point estimation and the accuracy of character recognition. 30 of 45 input characters are correctly recognized, and most of the incorrect identifications can be rectified by improving the recognition program. The result of this preliminary experiment can be said to be promising for recognizing all other uppercase characters.

References

[1] D. T. Pham, M. Al-Kutubi, Z. Ji, M. Yang, Z. Wang, and S. Catheline. Tangible acoustic inteface approches. Proceedings of IPROMS 2005 Virtual Conference, pp. 497–502, 2005.

[2] D. T. Pham, Z. Ji, M. Yang, Z. Wang, and M. Al-Kutubi. A novel human-computer interface based on passive acoustic localisation. Human-Computer Interaction, Part II, HCII 2007, LNCS 4551, pp. 901–909, 2007.

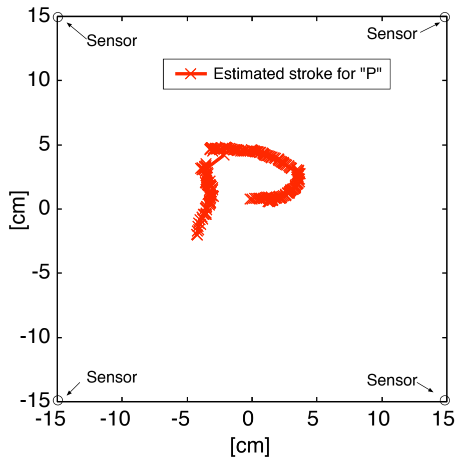

Figure 8: Estimated strokes of character “P,” which is recognized as “I” and “?”(unknown character).

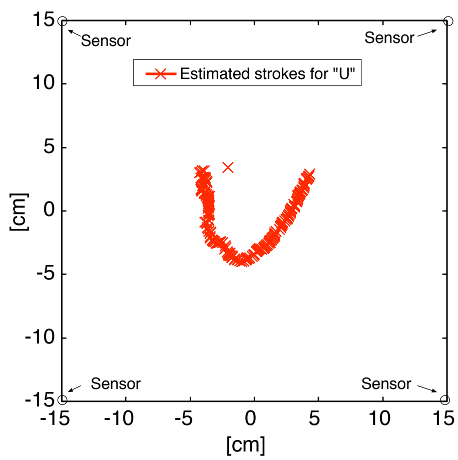

Figure 9: Estimated strokes of character “U,” which is recognized as “V.”

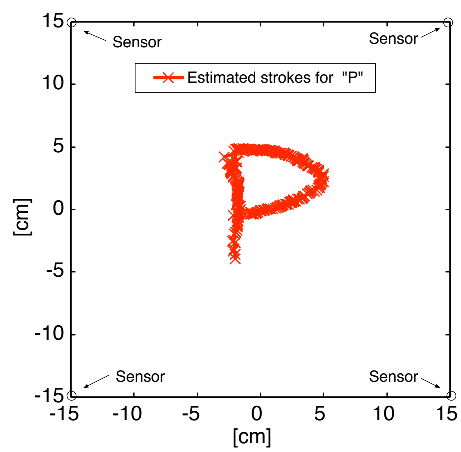

Figure 10: Estimated strokes of character “P,” which is recognized as “D.”

[3] C. Harrison, D. Tan, and D. Morris. Skinput: appropriating the body as an input surface. CHI ’10: Proceedings of the 28th international conference on Human factors in computing systems, pp. 453–462, 2010.
[4] J. A. Paradiso, C. K. Leo, N. Checka, and K. Hsiao, Passive acoustic knock tracking for interactive windows. CHI ’02: CHI ’02 extended abstracts on Human factors in computing systems, pp. 732–733, 2002.
[5] D. T. Pham, Z. Ji, O. Peyroutet, M. Yang, Z. Wang, and M. Al-Kutubi. Localisation of impacts on solid objects using the wavelet transform and maximum likelihood estimation. Proceedings of IPROMS 2006 Virtual Conference, pp. 541–547, 2006.
[6] C. Harrison and and S. E. Hudson. Scratch input: creating large, inexpensive, unpowered and mobile finger input surfaces. UIST ’08: Proceedings of the 21st annual ACM symposium on User interface software and technology, pp. 205–208, 2008.