ACOUSTIC-BASED PASSIVE POINTING SYSTEM FOR DISTANT SCREENS

Toshiharu Horiuchi, Shinya Takayama, and Tsuneo Kato

KDDI R&D Laboratories, Inc.
2-1-15 Ohara, Fujimino, Saitama 356-8502, Japan

ABSTRACT

This paper presents a passive pointing system for a distant screen based on an acoustic position estimation technology in conjunction with a gravity sensor on a smartphone. The system is designed to interact with a distant large screen such as a television set at home or digital signage in public. The system consists of a screen, two loudspeakers set around the screen, and a smartphone as a pointing device having a microphone and a gravity sensor inside. The position of the pointer is theoretically determined by the position and direction. This smartphone-based system approximates the position and direction by the two-dimensional position of the microphone horizontally and the pitch angle from the gravity sensor vertically. In this paper, we report experiments to evaluate the performance of the system. The loudspeakers of the system radiate burst signal from 18 to 24 kHz. The position of the smartphone is estimated at a frame rate of 15 Hz with a latency of 0.4 second. The accuracy of the pointer was measured as an angle error below 10 degrees for 100% of all frames. We confirmed that it has enough accuracy to point to each region which is divided area in the screen for applications such as quiz or questionnaire on digital signage.

Index Terms— acoustic application, acoustic position measurement, delay estimation

1. INTRODUCTION

Laser pointers are widely used to point to regions of interest on a distant large display or a projector screen. However, it is impossible to control objects on the display using a laser pointer. Ideally, users would like to control objects on a distant display like Nintendo's Wii. However, the Wii requires special equipment, namely infrared LEDs for the display and an infrared image sensor for the pointing device. Thus, we developed a novel pointing system using a standard smartphone and two loudspeakers without special devices based on an acoustic position estimation technology.

As a three-dimensional positioning technology, an ultrasonic position estimation technology is accurate and low-cost [5]. This technology shows promise for a wide range of applications such as location-aware computing, virtual and augmented reality. Three-dimensional ultrasonic positioning technology is technically based on estimation of time delay and/or time difference of arrival. The three-dimensional position is determined as the intersection of multiple spherical and/or hyperbolic planes given by the estimates.

There have been many studies and systems based on this technology [1–12]. The GCC-PHAT [6] is the most common approach in the sound source localization community to estimate the time delay and/or time difference of arrival. Active Bat [12] is an ultrasonic-based localization system that uses the time delay of arrival estimates. This system estimates the position of a target. In this system, an ultrasonic transmitter called Bat is attached to a target and ultrasonic receivers are placed in the environment. Cricket Compass [9] is an ultrasonic-based localization system that uses the time difference of arrival estimates. This system estimates the orientation of a target as well as its position. In this system, the receiver device has five ultrasonic sensors to determine its orientation. A motion capture system with multiple ultrasonic sensors on a human body is also a typical product of this technology.

We previously presented a three-dimensional pointing system based on this technology using three loudspeakers set around the screen and two microphones on the pointing device [10]. In this system, the three-dimensional position of each microphone is estimated by three distances from the loudspeakers, and the pointer is indicated at the calculated intersection of the straight line through the estimated positions of two microphones and the screen plane. The pointer is rendered at a frame rate of over 100 Hz with linear interpolation between the frames. Users can operate the pointer smoothly on the distant large display or the projector screen. However, it is difficult to apply a system that requires two microphones and three loudspeakers to consumer products such as cell-phones, smartphones, and television sets. Thus, we proposed a novel idea of an approximate system combining the acoustic position estimation technology and a pitch angle from the gravity sensor to use only a smartphone and two loudspeakers [4].

In this paper, we present the passive pointing system that enables interaction with a distant large display or a projector screen using general-purpose equipment. First, we describe the mechanism of our pointing system based on the acoustic position estimation technology. Next, we report experiments to evaluate the performance of the system. Finally, we summarize the paper.

2. PASSIVE POINTING SYSTEM BASED ON ACOUSTIC POSITION ESTIMATION

First, we introduce our previously presented system, which is based on an acoustic position estimation only. Next, we describe the proposed pointing system using a smartphone based on the acoustic position estimation in conjunction with a gravity sensor.

2.1. Fully Acoustic System with Three Loudspeakers and Two Microphones

Figure 1 shows the configuration of the fully acoustic system with three loudspeakers set around the screen and two microphones on the pointing device [10]. The system employs a passive configuration. The source signal is radiated by three loudspeakers around the screen, and received by two microphones on the pointing device. The three-dimensional position of each microphone is obtained with three distances derived from the time delay of arrival between the loudspeakers and the microphone. The pointer is indicated at the calculated intersection of the straight line through the estimated positions of the two microphones and the screen plane.

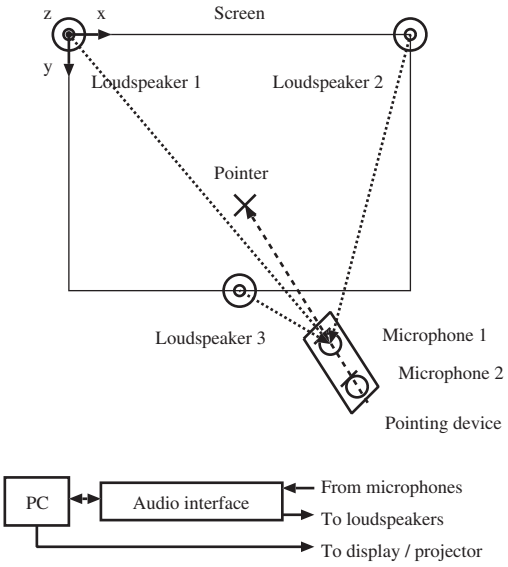

Fig. 1. The configuration of the fully acoustic system with three loudspeakers set around the screen and two microphones on the pointing device. The three loudspeakers are located at known position.

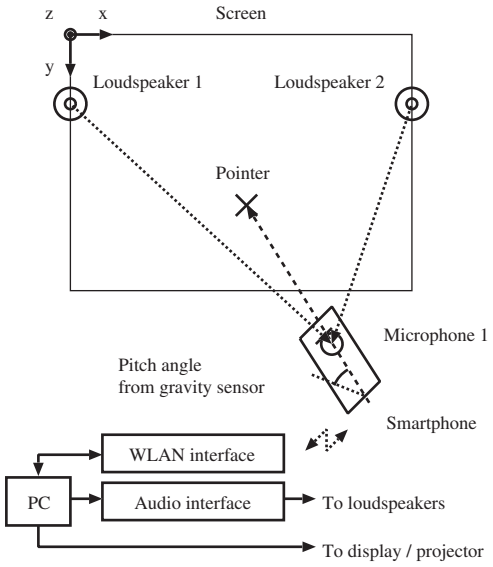

Fig. 2. The configuration of the approximate combined system with two loudspeakers set around the screen, and a smartphone as the pointing device containing a single microphone and a gravity sensor inside.

In Fig. 1, the position (u_i, v_i, w_i) of the loudspeaker i $(i = 1, 2, 3)$ is fixed to the coordinate originating from the screen. Let us assume that the loudspeaker i emits the source signal $s_i(k)$ in free space. The observed signal $x_j(k)$ at the microphone j $(j = 1, 2)$ is given by $x_j(k) = 1/d_{ij} \cdot s_i\,(k - d_{ij}/c) + n_j(k)$, where k is the discrete time index, d_{ij} represents the distance between loudspeaker i and microphone j, c is the sound velocity, and $n_j(k)$ is the interference signal observed by microphone j. Here, the distance d_{ij} is expressed as a function of the positions of loudspeaker i and microphone j as follows $d_{ij} = \sqrt{(x_j - u_i)^2 + (y_j - v_i)^2 + (z_j - w_i)^2}$, where (x_j, y_j, z_j) is the position of the microphone j.

Our task is to determine the position $\mathbf{p}_j = (x_j, y_j, z_j)^T$ of each microphone j from the source signals and the observed signals. The position is calculated as the intersection of multiple spherical planes given by the distances associated with the time delay of arrival. Here, we use the GCC-PHAT [6] for estimating the time delay of arrival τ_{ij} as $\phi_{ij}(l) = \mathrm{IDFT}\left[\{S_i(\omega)X_j^*(\omega)\}/\{|S_i(\omega)||X_j(\omega)|\}\right]$, $\tau_{ij} = \mathrm{argmax}_l\, \phi_{ij}(l)$, where $S_i(\omega)$ and $X_i(\omega)$ are the DFT transforms of $s_i(k)$ and $x_j(k)$ respectively. Then, we can obtain the estimated distances d_{ij} from τ_{ij} as $d_{ij} = c\tau_{ij}$. We use Newton-Raphson method for calculating the position $\mathbf{p}_j$ of each microphone.

Finally, the position of the pointer is calculated as the intersection of the straight line $(x, y, z)^T = (\mathbf{p}_2 - \mathbf{p}_1)t + \mathbf{p}_1$ through the estimated positions of two microphones and the screen plane, where t is the parameter as the real number, e.g. when the screen plane is $z = 0$, $t = -z_1/(z_2 - z_1)$.

2.2. Approximate Combined System with Two Loudspeakers, Single Microphone, and Gravity Sensor

Figure 2 shows the configuration of the approximate combined system with two loudspeakers set around the screen horizontally, and a smartphone as a pointing device equipping a single microphone and a gravity sensor inside [4]. The two-dimensional position of the microphone in the horizontal plane is obtained with two distances from the two loudspeakers, and the vertical direction is obtained by the pitch angle from the gravity sensor.

In Fig. 2, the acoustic position estimation is applied to estimation of the two-dimensional position $\mathbf{p}_1 = (x_1, z_1)^T$ of the microphone 1 in the horizontal plane. Though the position of the pointer requires not only the position but also the direction of the pointing device in principle, we approximate it gives the direction with the displacement of the microphone position from its initial position $\mathbf{p}_0 = (x_0, z_0)^T$. This approximation is reasonable because users move their arm holding the pointing device in a narrow angle with their wrist or elbow fixed as a pivot. Next, in the vertical plane, we use the pitch angle from the gravity sensor instead of the acoustic position estimation because the vertical position and direction cannot be obtained by the two loudspeakers setup. Here, we set a pitch angle θ of 0 degree at the vertical center, $+45$ degrees at the top, and -45 degrees at the bottom of the screen.

Finally, the x-axis position of the pointer is calculated as the intersection of the straight line $(x, z)^T = (\mathbf{p}_1 - \mathbf{p}_0)t + \mathbf{p}_0$ through the estimated and the initial positions of the microphone, where e.g. when the screen plane is $z = 0$, $t = -z_0/(z_1 - z_0)$. The y-axis position of the pointer is calculated as $y = h(1 - \tan\theta)/2$, where h is the height of the screen.

3. EVALUATION EXPERIMENTS

We conducted experiments to evaluate the performance of both systems. Figure 3 shows an elevation view and a side view of the experimental setup. In a soundproof room with reverberation time of 0.1 s, the pointing device was mounted on a turntable system that simulated an arm motion. The turntable system is rotated 360 degrees of 30 degrees per second. The rotation axis of the turntable system is $(x, y) = (1.00, 0.70)$. Table 1 shows the equipment used for the experiments.

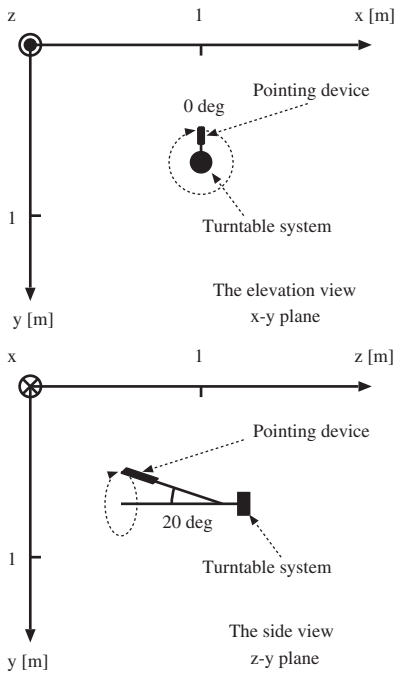

Fig. 3. The experimental setup. The pointing device moves using the turntable system. The upper panel shows the elevation view and the lower panel shows the side view.

Table 1. The equipment used for the experiments.

Equipment	Manufacture	Model
Microphone	DPA	4060
Smartphone	HTC	Desire
Loudspeaker	YAMAHA	NS-pf7
Power amplifier	BOSE	1200VI
Audio interface	M-AUDIO	FireWire1814
Turntable system	B&K	9640

The loudspeaker i of the three or two radiated short burst for the source signal $s_i(k)$ alternately. The short burst was a band-limited Gaussian noise from 18 to 24 kHz. The level of the short burst was 80 dBSPL at the front of each loudspeaker. The sampling condition was 48 kHz/16 bit. To prevent the interference of the direct signals from other loudspeaker(s), the interval of two successive emissions of the short burst was set at 64 ms. These systems have a time resolution of approximately 15 Hz, if it executes the position estimation every short burst radiates.

In the fully acoustic system, the loudspeakers were located triangularly on the screen plane. The three loudspeaker positions were $(x, y, z) = (0.00, 0.00, 0.00)$, $(2.00, 0.00, 0.00)$, and $(1.00, 1.70, 0.00)$ in meters. The two microphones were embedded in the pointing device. The interval between two microphones was 0.15 m. When the rotation angle was 0 degree, the positions of microphones were $(x, y, z) = (1.00, 0.57, 0.76)$ and $(1.00, 0.52, 0.62)$. Figure 4 shows the trajectories of the position of each microphone. Figure 5 shows the trajectory of the position of the pointer. In Fig. 4 and Fig. 5, the left half shows the true positions and the right half shows the estimated and calculated positions.

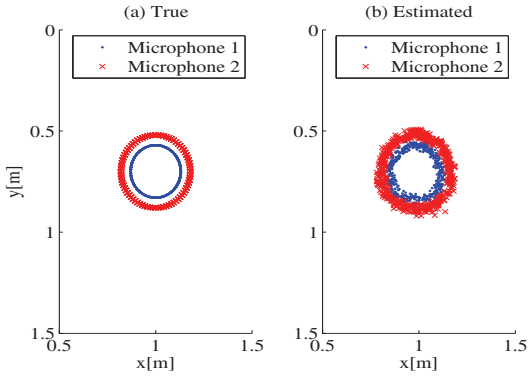

Fig. 4. The trajectories of the true and estimated positions of each microphone in the fully acoustic system.

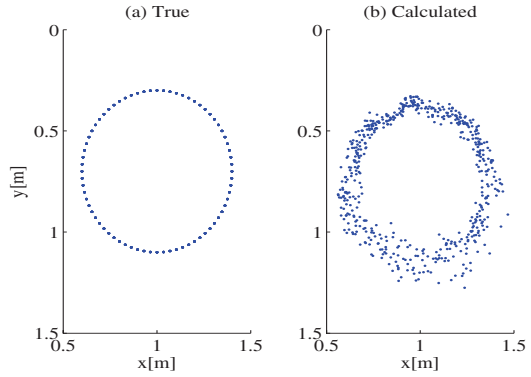

Fig. 5. The trajectory of the true and calculated positions of the pointer in the fully acoustic system.

In the approximate combined system, the loudspeakers were located horizontally on the screen plane. The two loudspeaker positions were $(x, y, z) = (0.00, 0.40, 0.00)$ and $(2.00, 0.40, 0.00)$ in meters. The microphone and gravity sensor were embedded in the smartphone. When the rotation angle was 0 degree, the position of microphone was $(x, y, z) = (1.00, 0.57, 0.76)$ as in the case of the fully acoustic system. The initial position was $(x, z) = (1.00, 1.13)$. The position estimation of the smartphone has a latency of 0.4 second. Most of the latency is caused by the smartphone's large audio buffer of 0.3 second in this implementation. Figure 6 shows the trajectories of the estimated distances and pitch angle. Figure 7 shows the trajectory of the true and calculated positions of the pointer.

We found that the estimated and calculated positions fluctuate around the true positions in both systems. This is mainly because the estimation error exists and the microphones move approximately 20 mm during the emission and its interval of 64 ms. The microphones were in static case, the standard deviation representing the estimation error of the positions of microphones was 17 mm, which is equivalent to the estimation error of common positioning systems. However, the standard deviation was increased to 38 mm in the moving case.

Figure 8 shows the accuracy of the pointer. We define the error margin as the acceptable range of the absolute value of the difference angles between the true and calculated positions of the pointer. The rate within the margin was calculated by counting the number of samples within the acceptable range from all frames. We found

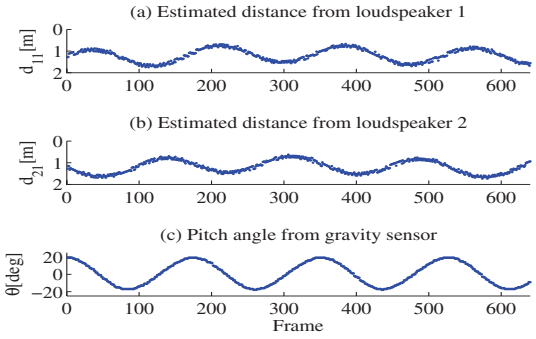

Fig. 6. The trajectories of the estimated distances and pitch angle in the approximate combined system.

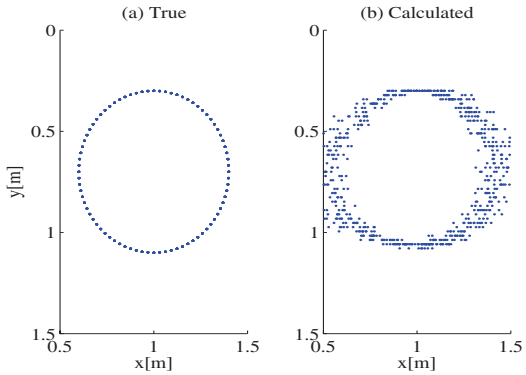

Fig. 7. The trajectory of the true and calculated position of the pointer in the approximate combined system.

from Fig. 8 that the rate within the margin of more than 90% was within 5 degrees in the fully acoustic system. On the other hand, in the approximate combined system, the rate within the margin was decreased to 80% for the x-axis and increased to 100% for the y-axis. These are because the approximate combined system doesn't determine the vertical position for the x-axis, and uses the gravity sensor for the y-axis. We also found that the rate within the margin of 100% for both systems and both axes was within 10 degrees. It means that it is difficult to draw a detailed picture by using these systems, however it has enough accuracy to point to each region which is divided area in the screen.

4. CONCLUSION

This paper showed two pointing systems using loudspeakers and microphones based on an acoustic position estimation technology. In the experiments, we used the burst signal from 18 to 24 kHz, which is reproducible by normal audio-visual equipment. The position of the smartphone is estimated at a frame rate of 15 Hz with a latency of 0.4 second. The accuracy of the pointer was measured as an angle error below 10 degrees for 100% of all frames. It means that it is difficult to draw a detailed picture by using this system, however it has enough accuracy to point to each region which is divided area in the screen for applications such as quiz or questionnaire on digital signage.

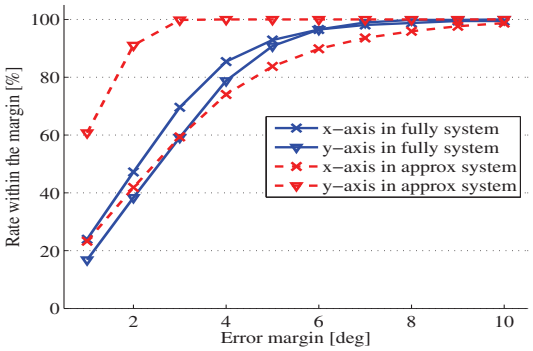

Fig. 8. The accuracy of the pointer. The relationship between the error margin and the rate within the margin.

5. REFERENCES

[1] M. S. Brandstein, J. E. Adcock, and H. F. Silverman. A closed-form location estimator for use with room environment microphone arrays. *IEEE Trans. on Speech and Audio Processing*, 5(1):45–50, Jan. 1997.

[2] Y. T. Chan and K. C. Ho. A simple and efficient estimator for hyperbolic location. *IEEE Trans. on Signal Processing*, 42(8):1905–1915, Aug. 1994.

[3] E. O. Dijk, C. H. van Berkel, R. M. Aarts, and E. J. van Loenen. 3-d indoor positioning method using a single compact base station. In *2nd IEEE Annual Conf. on Pervasive Computing and Communications (PerCom)*, pages 101–110, Mar. 2004.

[4] T. Horiuchi, S. Takayama, and T. Kato. A pointing system based on acoustic position estimation and gravity sensing. In *6th IEEE Symposium on 3D User Interfaces (3DUI)*, pages 105–106, Mar. 2011.

[5] T. Ito, T. Sato, K. Tulathimutte, M. Sugimoto, and H. Hashizume. A scalable tracking system using ultrasonic communication. *IEICE Trans. on Fundamentals*, E92-A(6):1408–1416, June 2009.

[6] C. H. Knapp and G. C. Carter. The generalized correlation method for estimation of time delay. *IEEE Trans. on Acoust. Speech and Signal Processing*, 24(4):320–327, Aug. 1996.

[7] M. R. McCarthy and H. L. Muller. Rf free ultrasonic positioning. In *7th IEEE Intl. Symposium on Wearable Computers (ISWC)*, pages 79–85, Oct. 2003.

[8] M. Omologo and P. Svaizer. Use of the crosspower-spectrum phase in acoustic event location. *IEEE Trans. on Speech and Audio Processing*, 5(3):288–292, May 1997.

[9] N. B. Priyantha, A. K. L. Miu, H. Balakrishnan, and S. Teller. The cricket compass for context-aware mobile applications. In *7th ACM Annual Intl. Conf. Mobile Computing and Networking (MobiCom)*, pages 1–14, July 2001.

[10] S. Takayama, T. Horiuchi, and T. Kato. Passive ultrasonic pointing system based on three-dimensional position estimation. In *20th Intl. Cong. on Acoustics (ICA)*, Aug. 2010.

[11] R. Want, A. Hopper, V. Falcao, and J. Gibbons. The active badge location system. *ACM Trans. on Information Systems*, 10(1):91–102, Jan. 1992.

[12] A. Ward, A. Jones, and A. Hopper. A new location technique for the active office. *IEEE Personal Communications*, 4(5):42–47, Oct. 1997.