Proceedings of 20th International Congress on Acoustics, ICA 2010

23–27 August 2010, Sydney, Australia

Passive Ultrasonic Pointing System Based on Three-dimensional Position Estimation

Shinya Takayama, Toshiharu Horiuchi, and Tsuneo Kato

User Interface Laboratory, KDDI R&D Laboratories Inc.
2-1-15 Ohara, Fujimino, Saitama 356-8502, Japan
{sh-takayama, to-horiuchi, tkato}@kddilabs.jp

PACS: 43.60.Fg, 43.60.Jn, 43.60.Mn

ABSTRACT

This paper presents a novel pointing system for large displays based on three-dimensional ultrasonic positioning technology. This system consists of a display or a projector screen, a pointing device mounting two microphones on the axis of the pointing direction, and three loudspeakers set around the display. The three-dimensional position of each microphone is estimated by trilateration with three distances from the loudspeakers to the microphone, and the pointer is indicated at the intersection of a straight line connecting two three-dimensional positions of the microphones, and the display plane. This system is targeted to interact with large displays such as digital signages. In the experiment with a prototype, we use source signals of a band-limited Gaussian noise from 18 to 24 kHz, which are reproducible by normal audio-visual equipment. The estimation error of the microphone position has a standard deviation of 17 mm, which is equivalent to the error of common positioning systems. The accuracy of the pointer was measured as an angle error below 4 degrees, which is comparable to jiggle of hand, for 95% of frames when the microphones are mounted on the pointing device with an interval of 0.15 m. The pointer is displayed at 15 Hz frame rate with a latency of 31 ms.

INTRODUCTION

Laser pointers are widely used to point to areas of interest on large displays and projector screens. However, it is impossible to click on objects on the screen using a laser pointer. Ideally, users would be able to indicate and control objects on large displays as user interfaces. Thus, we developed a pointing system based on three-dimensional ultrasonic positioning technology that enables interaction with a distant large display or projector screen.

As an acoustic localization technique, three-dimensional ultrasonic positioning technology is accurate and low cost [1]. This technology is promising in diverse applications such as location-aware computing, virtual and augmented reality, and other advanced services and systems. The three-dimensional positioning technology is technically based on estimation of the time delay of arrival and/or the time difference of arrival. The three-dimensional position is determined as the intersection of multiple spherical or hyperbolic planes given by estimates of the time delay of arrival or the time difference of arrival.

There have been many studies and systems based on this technology [2, 3, 4, 5, 6, 7, 8, 9, 10]. The Active Bat [2] is an ultrasonic-based localization system that uses the time delay of arrival estimates. This system can estimate the position of a target. In this system, an ultrasonic transmitter called the Bat is attached to a target and ultrasonic receivers are located in the environment. The Cricket Compass [3] is an ultrasonic-based localization system that uses the time difference of arrival estimates. This system can estimate 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.

In this paper, we present a novel pointing system for distant large displays based on three-dimensional ultrasonic positioning technology. This system consists of a display or a projector screen, a pointing device mounting two microphones on the axis of the pointing direction, and three loudspeakers set around the display. We prototyped the system, and evaluated not only the precision of the three-dimensional position estimation but also the precision of a pointing device. The three-dimensional position of each microphone is estimated by trilateration with three distances from the loudspeakers to the microphone, and the pointer is indicated at the calculated intersection of the straight line through the estimated positions of the two microphones and the display plane. For downsizing and low power consumption of the pointing device, this system employs a passive configuration. First, we describe the details of our pointing system. Next, we perform experiments on the accuracy and the performance of the system. Finally, we summarize the paper.

PASSIVE ULTRASONIC POINTING SYSTEM

In this section, we describe the proposed passive ultrasonic pointing system. First, we introduce the configuration of our pointing system based on three-dimensional position estimation. Next, we describe the position estimation algorithm of the system.

Configuration

Figure 1 shows the configuration of our pointing system. This system consists of a display or a projector screen, a pointing device mounting two microphones and three loudspeakers set around the display. The system employs a passive configuration, where the source signal is radiated by the loudspeakers around the display, and received by the microphones on the pointing device. The position of each microphone is obtained based on trilateration using the estimated distances derived from the time delay of arrival estimates between the loudspeakers and the microphone, and the pointer is indicated at the calculated intersection of the straight line through the estimated positions of the two microphones and the display plane.

In Fig. 1, S_i $(i = 1,2,3)$ is a loudspeaker. The position $(u_i, v_i, w_i)^T$ of the loudspeaker S_i is fixed to the coordinate originated from the display. M_j $(j = 1,2)$ is a microphone. Let us assume that the loudspeaker S_i emits the source signal $s_i(k)$ in free space. The observed signal $x_j(k)$ at the microphone M_j is given by

$$x_j(k) = \frac{1}{d_{ij}} s_i \left(k - \frac{d_{ij}}{c} \right) + n_j(k), \tag{1}$$

where k is the discrete time index, d_{ij} represents the distance between the loudspeaker S_i and the microphone M_j, c is the sound velocity, and $n_j(k)$ is the interference signal observed by the microphone M_j. Here, the distance d_{ij} is expressed as a function of the positions of loudspeaker S_i and microphone M_j on the coordinate system

$$d_{ij} = \sqrt{(x_j - u_i)^2 + (y_j - v_i)^2 + (z_j - w_i)^2}, \tag{2}$$

where $(x_j, y_j, z_j)^T$ is the position of the microphone M_j.

Our task is to find the three-dimensional position $(x_j, y_j, z_j)^T$ of each microphone M_j from the source signals and the observed signals. A three-dimensional position is calculated as the intersection of multiple spherical planes given by the distances associated with the time delay of arrival estimates. The time delay of arrival τ_{ij} is estimated by maximizing the cross-correlation between the source signal $s_i(k)$ and the observed signal $x_j(k)$ as

$$\phi_{ij}(l) = \sum_k s_i(k) x_j(k+l), \tag{3}$$

$$\tau_{ij} = \underset{l}{\operatorname{argmax}}\, \phi_{ij}(l). \tag{4}$$

Then, we can obtain the estimated distances d_{ij} from the time delay of arrival estimates τ_{ij} as

$$d_{ij} = c\tau_{ij}. \tag{5}$$

Position Estimation Algorithm

We use Newton-Raphson method for calculating the position of each microphone. $(x(n), y(n), z(n))^T$ are the position of the n-th iteration, and $(\Delta x(n), \Delta y(n), \Delta z(n))^T$ are incremental factors. The position $(x(n+1), y(n+1), z(n+1))^T$ of the $(n+1)$-th iteration can be updated by

$$\begin{pmatrix} x(n+1) \\ y(n+1) \\ z(n+1) \end{pmatrix} = \begin{pmatrix} x(n) \\ y(n) \\ z(n) \end{pmatrix} + \begin{pmatrix} \Delta x(n) \\ \Delta y(n) \\ \Delta z(n) \end{pmatrix}. \tag{6}$$

The incremental factors $(\Delta x(n), \Delta y(n), \Delta z(n))^T$ are reestimated to minimize a square error as

$$\sum_i e_i^2(n+1), \tag{7}$$

where $e_i(n+1)$ is the residual error given by

$$e_i(n+1) = \hat{d}_i(n+1) \;-\; d_i, \tag{8}$$

Here, we can approximate Eq. (8) by linearization using the relationship between Eq. (6) and the following equation

$$\hat{d}_i(n) = \sqrt{(x(n) - u_i)^2 + (y(n) - v_i)^2 + (z(n) - w_i)^2}, \tag{9}$$

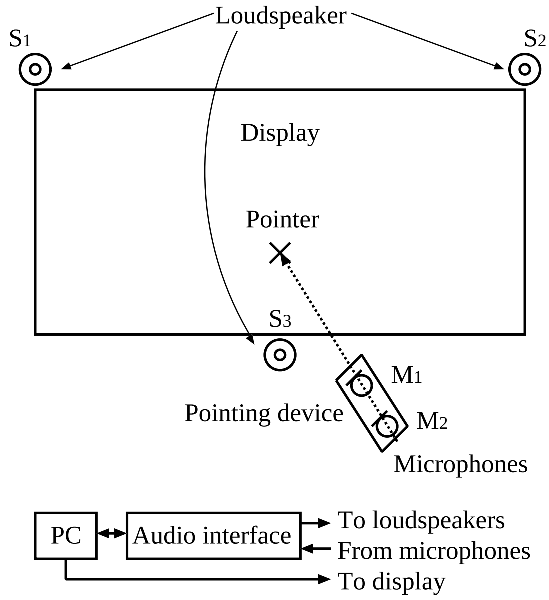


Figure 1: The configuration of our pointing system. S_i $(i = 1,2,3)$ is a loudspeaker located at known position, M_j $(j = 1,2)$ is a microphone on the pointing device. The system shows the pointer that is indicated at the calculated intersection of the straight line through the estimated positions of the two microphones and the display plane.

the residual error $e_i(n+1)$ is rewritten as

$$\begin{aligned} e_i(n+1) \quad \approx \quad & \frac{x(n) - u_i}{\hat{d}_i(n)} \Delta x(n) \\ + \quad & \frac{y(n) - v_i}{\hat{d}_i(n)} \Delta y(n) \\ + \quad & \frac{z(n) - w_i}{\hat{d}_i(n)} \Delta z(n) \; + \; \hat{d}_i(n) \; - \; d_i. \end{aligned} \tag{10}$$

The estimated positions updated by the incremental factors, and this reestimation is iterated until the following convergence condition is fulfilled

$$\sum_i e_i^2(n+1) - \sum_i e_i^2(n) < \varepsilon. \tag{11}$$

EVALUATIONS

In this section, we carry out experiments to evaluate the accuracy and the performance of our pointing system.

Figure 2 shows an elevation view and a top view of the experimental setup. The loudspeakers were located triangularly on the display plane. The three loudspeaker positions were $(x,y,z)^T = (0,0,0)^T$, $(2,0,0)^T$, and $(1,1.7,0)^T$ in meters. The two microphones were embedded in the pointing device. The interval between two microphones was 0.15 m. In a soundproof room with reverberation time of 0.1 s, the pointing device was mounted on a turntable system that simulated a hand motion. The rotation axis of the turntable system is $(x,z)^T = (1,1.5)^T$. When the rotation angle $\theta = 0$ deg, the positions of microphones were $(x,y,z)^T = (1,0.7,1)^T$ and $(1,0.7,0.85)^T$. Table 1 shows the equipment for the experiments.

Three loudspeakers S_i radiated short bursts for the source signals $s_i(k)$ alternately. The short bursts were a band-limited Gaussian noise from 18 to 24 kHz. The level of the short bursts was 80 dBSPL at the front of each loudspeaker. The sampling

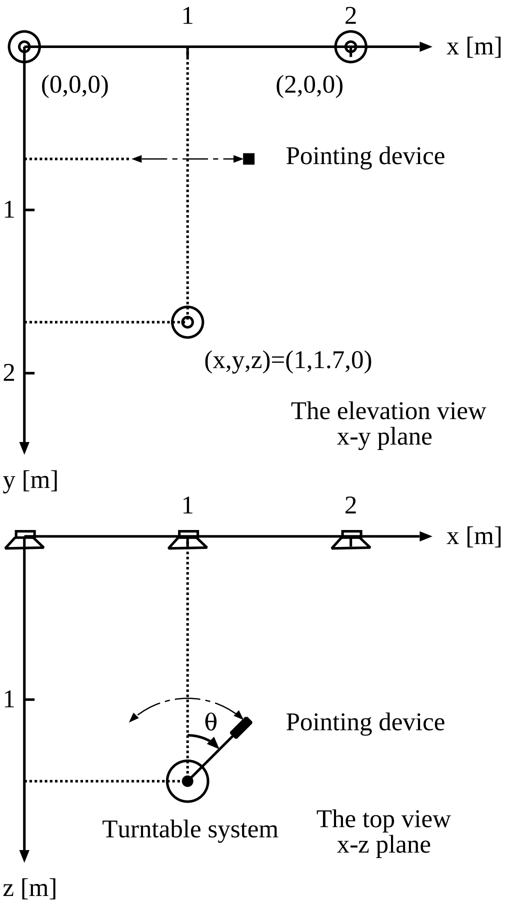


Figure 2: The experimental setup. The pointing device moves using the turntable system. The upper panel shows the elevation view and the lower panel shows the top view.

Table 1: The equipment for the experiments.

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

condition was 48 kHz/16 bit. The emission pattern is shown in Fig. 3. To prevent the interference of the direct signals from other two loudspeakers and the reflected signals, the interval of two emissions of the short burst was set at 64 ms. The system has a time resolution of approximately 15 frames per second, if it executes the position estimation every short burst radiates.

We carried out experiments under two conditions. First, the pointing device was in static case. Second, the pointing device was in moving case. In static case, the turntable system stops at $\theta = -60, -30, 0, 30, 60$ deg. In moving case, the turntable system rotates from $\theta = -60$ deg to 60 deg of 30 degrees per second on $x - z$ plane. Two hundred trials were performed in both cases.

Figure 4 shows the estimation error of the positions of microphones. In Fig. 4, the left half shows the static case and the right half shows the moving case. The error bars show the standard deviation. We found from Fig. 4 that the standard deviation in the moving case was larger than that in the static case. This is because the microphones move approximately 17 mm during the emission interval of 64 ms in the moving case. Standard deviation representing the estimation error of the positions of microphones was 17 mm, which is equivalent to the error of common positioning systems.

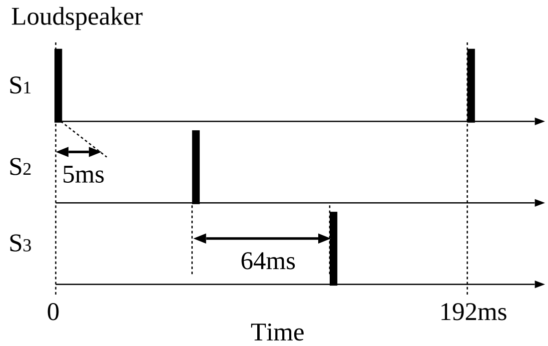


Figure 3: The emission pattern of the source signals for loudspeakers S_i. The source signals are a band-limited Gaussian noise from 18 to 24 kHz.

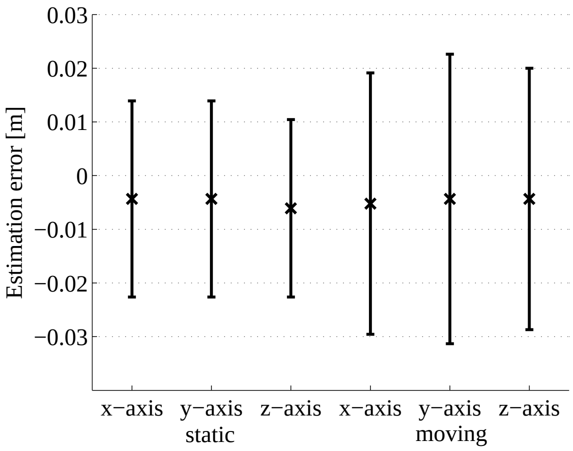


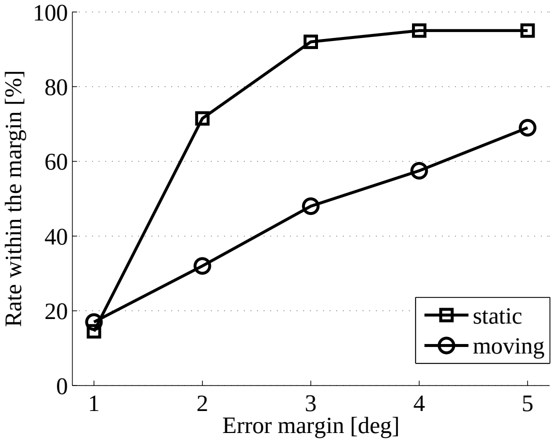
Figure 4: The estimation error of the microphones' positions. The left half shows the static case. The right half shows the moving case. The error bars show the standard deviation.

Figure 5 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 point and estimated point. The rate within the margin was calculated by counting the number of samples within the acceptable range from all frames and all trials. In Fig. 5, the upper panel shows angle errors for the horizontal direction whereas the lower panel shows angle errors for the vertical direction. We found from Fig. 5 that the rate within the margin of more than 95% of the frames and trials was within 4 degrees in the static case. On the other hand, in the moving case, the rate within the margin was decreased to 60% for the horizontal direction and 70% for the vertical direction.

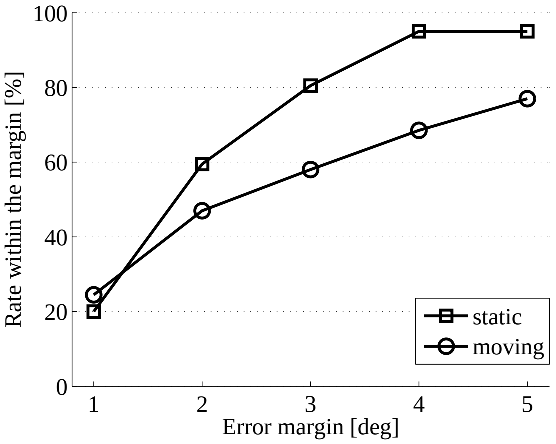
The experimental result shows the basic functionality of tracking the motion of the device. The accuracy still needs to be improved in future work.

CONCLUSION

This paper presents a novel pointing system for a large display based on three-dimensional ultrasonic positioning technology. This system is targeted to interact with large displays such as digital signages. In the experiments, we used a band-limited



(a) Horizontal direction



(b) Vertical direction

Figure 5: The accuracy of the pointer. The relationship between the error margin and the rate within the margin in the static and the moving case. The upper panel shows the difference angle of the horizontal direction and the lower panel shows the difference angle of the vertical direction.

Gaussian noise from 18 to 24 kHz for source signal, which are reproducible by normal audio-visual equipment. Standard deviation representing the estimation error of the microphones' positions is 17 mm, which is equivalent to the error of common positioning systems. The accuracy of the pointer was measured as an angle error below 4 degrees, which is comparable to jiggle of hand, for 95% of frames when the microphones are placed at 0.15 m interval. The pointer is displayed at 15 Hz frame rate with a latency of 31 ms. To make the pointer track the user operation correctly and smoothly, the accuracy of position estimation and its frame rate should be increased in future work.

REFERENCES

[1] T. Ito, T. Sato, K. Tulathimutte, M. Sugimoto, and H. Hashizume, "A Scalable Tracking System Using Ultrasonic Communication," IEICE Trans. Fundamentals, vol. E92-A, no. 6, pp. 1408–1416, 2009.

[2] A. Ward, A, Jones, and A. Hopper, "A new location technique for the active office," IEEE Pers. Commun., vol. 4, no. 5, pp. 42–47, 1997.

[3] N. B. Priyantha, A. K. Miu, H. Balakrishnan, and S. Teller, "The cricket compass for context-aware mobile applications," Proc. ACM MobiCom, pp. 1–14, 2001.

[4] M. Omologo and P. Svaizer, "Use of the Crosspower-Spectrum Phase in Acoustic Event Localization," IEEE Trans. Speech & Audio Process., vol. 5, no. 3, pp. 288–292, May 1997.

[5] Y. T. Chan and K. C. Ho, "A simple and efficient estimator for hyperbolic location," IEEE Trans. Signal Process., vol. 42, no. 8, pp. 1905–1915, Aug. 1994.

[6] M. S. Brandstein, J. E. Adcock, and H. F. Silverman, "A closed-form location estimator for use with room environment microphone arrays," IEEE Trans. Speech & Audio Process., vol. 5, no. 1, pp. 45–50, Jan. 1997.

[7] M. Minami, Y. Fukuju, K. Hirosawa, S. Yokoyama, M. Mizumachi, H. Morikawa, and T. Aoyama, "DOLPHIN: A practical approach for implementing a fully distributed indoor ultrasonic positioning system," Proc. UbiComp, pp. 347–365, Sep. 2004.

[8] E. O. Dijk, C. H. (Kees) van Berkel, R. M. Aarts, and E. J. van Loenen, "A 3-D indoor positioning method using a single compact base station," Proc. IEEE PerCom, pp. 101, 2004.

[9] M. R. McCarthy and H. L. Muller, "RF free ultrasonic positioning," Proc. IEEE ISWC, pp. 79–85, Oct. 2003.

[10] R. Want, A. Hopper, V. Flcão, and J. Gibbons, "The active badge location system," ACM Trans. Information systems, vol. 10, no. 5, pp. 42–47, 1992.