Journal of Information Processing Vol.21 No.2 362–366 (Apr. 2013)

[DOI: 10.2197/ipsjjip.21.362]

Regular Paper

Passive Pointing System for Distant Screens Using Acoustic Position Estimator and Gravity Sensor

Toshiharu Horiuchi[1,a)] Shinya Takayama[1] Tsuneo Kato[1]

Received: April 13, 2012, Accepted: November 2, 2012

Abstract: This paper presents two passive pointing systems for a distant screen based on an acoustic position estimation technology. These systems are designed to interact with a distant screen such as a television set at home or digital signage in public as an alternative to a touch screen. The first system consists of a distant screen, three loudspeakers set around the screen, and two microphones as a pointing device. The second system consists of a distant screen, two loudspeakers set around the screen, and a smartphone equipping a microphone and a gravity sensor inside as a pointing device. The position of the pointer on the screen is theoretically determined by the position and direction of the pointing device in the space. The second 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 these systems. The loudspeakers of these systems radiate burst signal from 18 to 24 kHz. The position of the microphone is estimated at a frame rate of 15 frames per second with a latency of 0.4 s. 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 one of several partitioned areas in the screen.

Keywords: acoustic application, acoustic position measurement, delay estimation

1. Introduction

Laser pointers are widely used to point to regions of interest on a distant display or a projector screen. However, it is impossible to control objects on the display with the 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 on the display side and an infrared image sensor on the pointing device side. To let users control objects without special equipment, we developed novel pointing systems based on an acoustic position estimation technology.

As a three-dimensional positioning technology, an ultrasonic position estimation technology is accurate and low-cost [6]. This technology is promising for a wide range of applications such as location-aware computing, and 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], [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13]. The GCC-PHAT [7] is the most common approach in the sound source localization technology to estimate the time delay and/or time difference of arrival. Active Bat [13] 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 [10] 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.

In this paper, we present two passive pointing systems which let users interact with a distant display or a projector screen using general-purpose equipment at the distance of about one meter as an alternative to a touch screen. First, we describe the mechanism of our pointing systems based on the acoustic position estimation technology. Next, we report experiments to evaluate the performance of these systems. Finally, we summarize the paper.

2. Passive Pointing System Based on Acoustic Position Estimation

First, we propose our basic system based only on acoustic position estimation. Next, we propose the smartphone-based pointing system 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 [11]. For downsizing and low power consumption of the pointing device, the system employs a

[1] KDDI R&D Laboratories, Inc., Fujimino, Saitama 356–8502, Japan
[a)] to-horiuchi@kddilabs.jp

© 2013 Information Processing Society of Japan

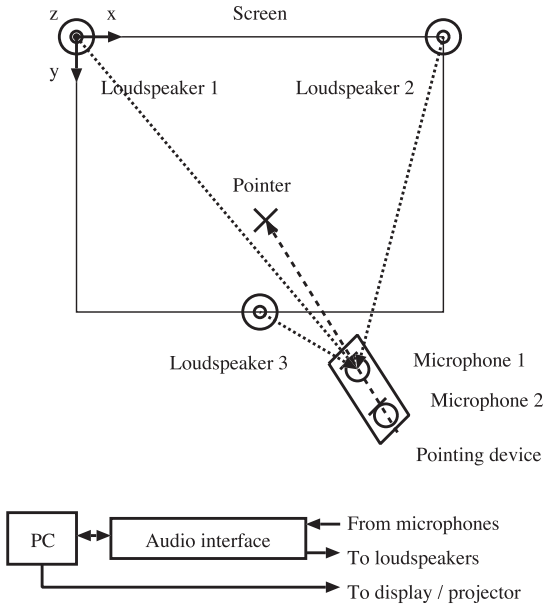

Fig. 1 The configuration of the fully acoustic system with three loudspeakers set around the screen and two microphones on the pointing device. The three loudspeakers are located at known positions.

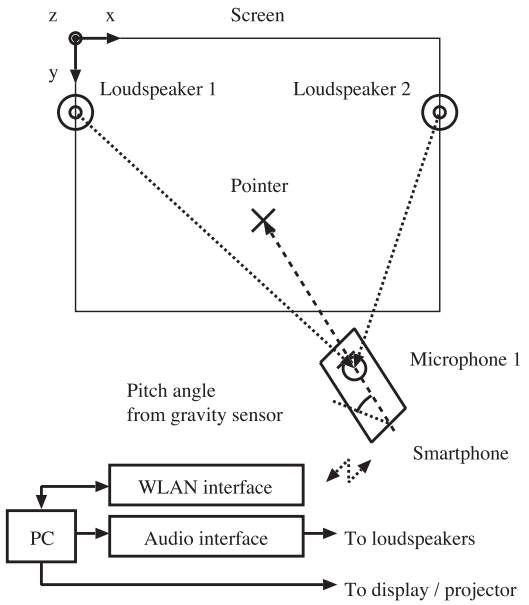

Fig. 2 The configuration of the approximate combined system with two loudspeakers set around the screen, and a smartphone as the pointing device containing a single microphone and a gravity sensor inside.

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 which correspond to 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.

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 $s_i(t)$ in free space. The observed signal $x_j(t)$ at the microphone j $(j = 1, 2)$ can be mathematically modeled as [2], [7], [9]

$$x_j(t) = \alpha_{ij} s_i(t - \tau_{ij}) + n_j(t) = \alpha_{ij} s_i(t - d_{ij}/c) + n_j(t), \qquad (1)$$

where t is the time index, α_{ij} is an attenuation factor due to propagation effects, τ_{ij} represents the time delay of arrival between loudspeaker i and microphone j, d_{ij} represents the distance, c is the sound velocity, i.e., $\tau_{ij} = d_{ij}/c$, and $n_j(t)$ 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

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

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 by trilateration as the intersection of multiple spherical planes given by the distances corresponding to the time delay of arrival. Now, Eq. (1) can be rewritten in the frequency domain as

$$X_j(f) = \alpha_{ij} S_i(f) e^{-j2\pi f \tau_{ij}} + N_j(f), \qquad (3)$$

where f is the frequency index, and $X_j(f)$, $S_i(f)$, and $N_j(f)$ are the Fourier transforms of $x_j(t)$, $s_i(t)$, and $n_j(t)$ respectively.

Here, we use the GCC-PHAT [7], which has a good performance under noise or reverberation environments, for estimating the time delay of arrival τ_{ij} as follows

$$\tau_{ij} = \operatorname*{argmax}_{\tau} \sum_{f} \frac{S_i(f) X_j^*(f)}{|S_i(f) X_j^*(f)|} e^{j2\pi f \tau}. \qquad (4)$$

Then, we can obtain the estimated distances d_{ij} from τ_{ij} as

$$d_{ij} = c\tau_{ij}. \qquad (5)$$

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 $(x, y, z)^T = (\mathbf{p}_2 - \mathbf{p}_1)\gamma + \mathbf{p}_1$ through the estimated positions of two microphones and the screen plane, where γ is a real number when the screen plane is $z = 0$, $\gamma = -z_1/(z_2 - z_1)$.

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

The fully acoustic system composed of three loudspeakers and two microphones calculates the position and direction of the pointing device exactly [11]. However, it is not practical to mount two microphones on a cellphone or a smartphone and to install three loudspeakers for a television set. Thus, we proposed a novel approximate system consisting of a smartphone and two loudspeakers by combining the acoustic position estimation technology and a pitch angle from the gravity sensor on the smartphone [4], [5].

Figure 2 shows the configuration of the approximate combined system with two loudspeakers set at both sides of the screen, and a smartphone as a pointing device equipping a single microphone and a gravity sensor inside [4], [5]. The two-dimensional position of the microphone in the horizontal plane is calculated by trilateration based on two distances from the two loudspeakers, and the

© 2013 Information Processing Society of Japan

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 the direction by a linear function of the displacement of the microphone position from its initial position $\mathbf{p}_0 = (x_0, z_0)^T$. On start-up, press the button, and the initial position sets the estimated position at the time. This approximation is reasonable because users usually use the pointing device by moving their arm in a narrow angle with their wrist or elbow fixed as a pivot. In contrast, 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 $(x, z)^T = (\mathbf{p}_1 - \mathbf{p}_0)\gamma + \mathbf{p}_0$ through the estimated and the initial positions of the microphone, when the screen plane is $z = 0$, $\gamma = -z_0/(z_1 - z_0)$. 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 a reverberation time of 0.1 s, the pointing device was mounted at the tilt of 20 degrees from the rotation axis on a turntable system which simulated an arm motion. The turntable system is rotated 360 degrees at a rate of 30 degrees per second. The rotation axis of the turntable system is $(x, y) = (1.00, 0.70)$. Thus, the position and direction of the pointing device move in three-dimensional space. **Table 1** shows the equipment used for the experiments.

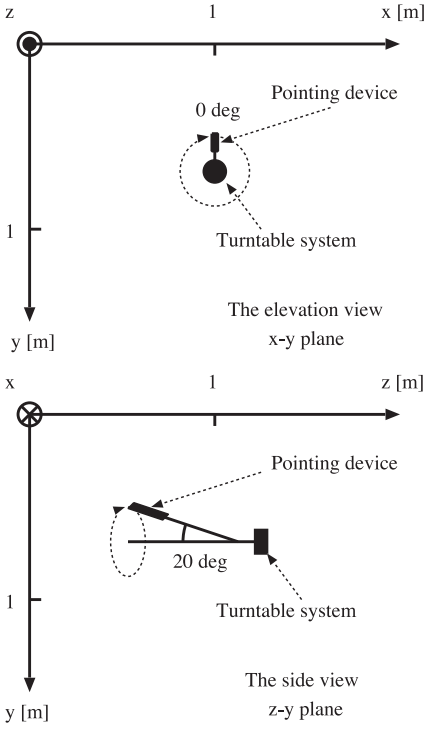

Fig. 3 The experimental setup. The pointing device rotates on a turntable system. The upper panel shows the elevation view and the lower panel shows the side view.

The loudspeaker i of the three or two radiated short burst source signal $s_i(k)$ alternately. The short burst was a band-limited Gaussian noise from 18 to 24 kHz. Generally, the short burst in this frequency range is not contaminated significantly by real environmental noise, however, noise reduction is an issue for the future. For estimating the time delay of arrival, we use the GCC-PHAT for this frequency range. 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 loudspeakers, the interval of two successive emissions of the short burst was set at 64 ms. These systems have a time resolution of approximately 15 frames per second, because the position can be estimated as every short burst radiates. The pointer is rendered at a frame rate of over 100 Hz with linear interpolation between the frames.

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 mounted on 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.

In the approximate combined system, the loudspeakers were set 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

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

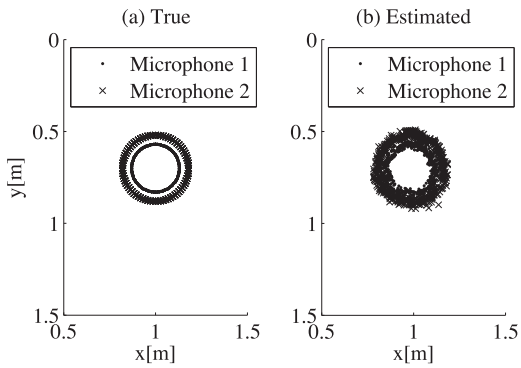

Fig. 4 The trajectories of the true and estimated positions of each microphone in the fully acoustic system.

© 2013 Information Processing Society of Japan

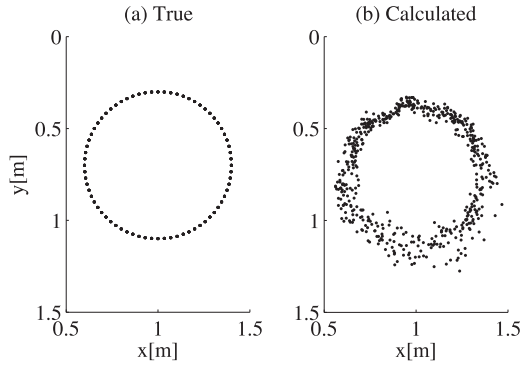

Fig. 5 The trajectory of the true and calculated positions of the pointer in the fully acoustic system.

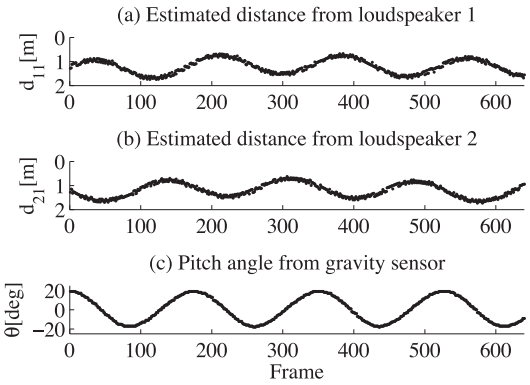

Fig. 6 The trajectories of the estimated distances and pitch angle in the approximate combined system.

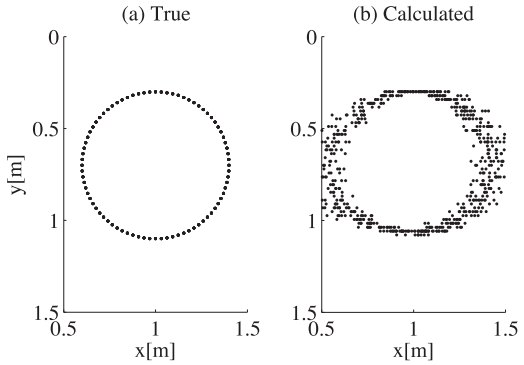

Fig. 7 The trajectory of the true and calculated position of the pointer in the approximate combined system.

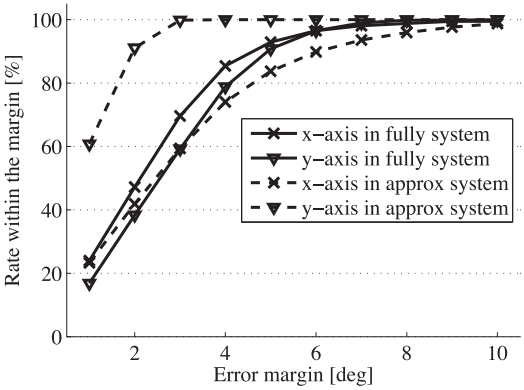

Fig. 8 The accuracy of the pointer. The relationship between the error margin and the rate within the margin.

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 the same 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 s. The latency is mainly caused by the smartphone's large audio buffer of 0.3 s, and the emission interval of 64 ms 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 due to estimation error which is both independent and dependent between frames. When 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. The estimation error of the distances from the loudspeakers to the microphones was 7 mm at maximum, which is equivalent to the sampling period. In contrast, the microphones move approximately 20 mm during the emission and its frame interval of 64 ms. The standard deviation was increased to 38 mm in the moving case. Therefore, the estimation error is mainly caused by the movement of the microphones. In addition, the fluctuation as shown in the lower half of Fig. 5 is caused by the reflection on the floor in the experimental setup.

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 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 corresponding to 0.10 m on the screen in the experimental setup. This means that it is difficult to draw a precise picture by using these systems, however it has enough accuracy to point to one of several partitioned areas in the screen.

4. Conclusions

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 equipment. The position of the smartphone is estimated at a frame rate of 15 frames per second with a latency of 0.4 s. The accuracy of the pointer was measured as an angle error below 10 degrees for 100% of all frames. This means that it is difficult to draw a precise picture by using this system, however it has enough accuracy to point to one of several partitioned areas in the screen.

© 2013 Information Processing Society of Japan

References

[1] Brandstein, M.S., Adcock, J.E. and Silverman, H.F.: A closed-form location estimator for use with room environment microphone arrays, *IEEE Trans. Speech and Audio Processing*, Vol.5, No.1, pp.45–50 (1997).

[2] Chan, Y.T. and Ho, K.C.: A simple and efficient estimator for hyperbolic location, *IEEE Trans. Signal Processing*, Vol.42, No.8, pp.1905–1915 (1994).

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

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

[5] Horiuchi, T., Takayama, S. and Kato, T.: Acoustic-based passive pointing system for distant screens, *Proc. 37th IEEE Intl. Conf. Acoustics, Speech and Signal Processing* (*ICASSP*), pp.1689–1692 (2012).

[6] Ito, T., Sato, T., Tulathimutte, K., Sugimoto, M. and Hashizume, H.: A scalable tracking system using ultrasonic communication, *IEICE Trans. Fundamentals*, Vol.E92-A, No.6, pp.1408–1416 (2009).

[7] Knapp, C.H. and Carter, G.C.: The generalized correlation method for estimation of time delay, *IEEE Trans. Acoust. Speech and Signal Processing*, Vol.24, No.4, pp.320–327 (1996).

[8] McCarthy, M.R. and Muller, H.L.: RF free ultrasonic positioning, *Proc. 7th IEEE Intl. Symposium on Wearable Computers* (*ISWC*), pp.79–85 (2003).

[9] Omologo, M. and Svaizer, P.: Use of the crosspower-spectrum phase in acoustic event location, *IEEE Trans. Speech and Audio Processing*, Vol.5, No.3, pp.288–292 (1997).

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

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

[12] Want, R., Hopper, A., Falcao, V. and Gibbons, J.: The active badge location system, *ACM Trans. Inf. Syst.*, Vol.10, No.1, pp.91–102 (1992).

[13] Ward, A., Jones, A. and Hopper, A.: A new location technique for the active office, *IEEE Personal Communications*, Vol.4, No.5, pp.42–47 (1997).

Toshiharu Horiuchi received his Bachelor's, Master's degrees and Ph.D. from Nagaoka University of Technology in 1999, 2001 and 2004, respectively. From 2002 to 2006, he was a researcher at the Spoken Language Communication Research Laboratories, Advanced Telecommunications Research Institute International (ATR). From 2006 to 2009, he was a research engineer at KDDI R&D Laboratories, Inc. In 2009, having joined the KDDI Corporation, he has worked as a research engineer at KDDI R&D Laboratories, Inc. His research interests include acoustic signal processing and adaptive signal processing, and in particular spatial signal processing. He received the 17th TELECOM System Technology Award for Students from the Telecommunications Advancement Foundation in 2002, the 20th Fujio Frontier Award from the Institute of Image Information and Television Engineers in 2012, and the Best Technical Demonstration Runner-up at ACM Multimedia 2012. He is a member of ASJ, IEEE and IEICE.

Shinya Takayama received his B.S. and M.E. degrees from Waseda University in 2006 and 2008, respectively. In 2008, having joined the KDDI Corporation, he has worked as an associate research engineer at KDDI R&D Laboratories, Inc. He was engaged in the research of three-dimentional position estimation using an acoustic wave. His research interests currently include multimedia processing with photo images. He is a member of IEICE and ITE.

Tsuneo Kato received his B.E., M.E. degrees and Ph.D. from the University of Tokyo in 1994, 1996 and 2011, respectively. He joined the Kokusai Denshin Denwa Co. Ltd. in 1996. He is currently with KDDI R&D Laboratories, Inc. He has been engaged in the research and development of automatic speech recognition and interactive user interface. He received the IPSJ Kiyasu Special Industrial Achievement Award in 2011 and the Best Technical Demonstration Runner-up at ACM Multimedia 2012. He is a member of ASJ, IEEE, IEICE and IPSJ.

© 2013 Information Processing Society of Japan