24th INTERNATIONAL CONGRESS ON SOUND AND VIBRATION
23–27 July 2017, London

ACOUSTIC DISTANCE MEASUREMENT BASED ON PHASE INTERFERENCE BETWEEN TRANSMITTED AND REFLECTED WAVES AND POSITION ESTIMATION BY USING KINECT V2'S MICROPHONE ARRAY

Ryohei Oda, Toshihiro Shinohara, Tetsuji Uebo, Noboru Nakasako
Faculty of B.O.S.T., 930 Nishi-Mitani Kinokawa City, Wakayama, 649-6493 Japan
email: 1633730013t@waka.kindai.ac.jp

The distance to a target is fundamental and very important information in many engineering applications. Kinect is known to have many sensors (e.g., depth sensor, RGB camera, color camera sensor, four-elements microphone array and so on). A depth sensor of Kinect v2 determines distance from Time-of-Flight (TOF) of infrared ray. The sensing range of the depth sensor is adjustable, but it is difficult to measure the distance closer than 0.500 m. Although TOF is typically used for distance estimation, it is difficult to estimate short distance. Alternatively, an acoustic distance measurement (ADM) method has been proposed based on the phase interference between transmitted and reflected waves. The procedure of the ADM is as follows: first, to obtain a power spectrum, applying Fourier transform to the observed wave, and then, to obtain a range spectrum, applying Fourier transform again to the power spectrum and taking its absolute value. The power spectrum is a periodical function, whose period is inversely proportional to the distance. Therefore, the distance can be determined from the range spectrum. This ADM method is able to measure the distance closer than 0.500 m. In this study, we focus on performing this ADM using Kinect v2's microphone array. Then we compare measurement accuracies of Kinect's depth sensor and the ADM method. We confirmed the accuracy of the ADM method by performing a computer simulation under an assumption of using Kinect v2's microphone array and by applying it to an actual sound field. In addition, using four-channel (4ch) microphone array, we geometrically estimate the position of the target by combining distances estimated at each channel.

Keywords: Acoustic distance measurement based on phase interference, range spectrum, Kinect v2, four-channel microphone array, depth sensor

---

1. Introduction

In various engineering fields, it is important to estimate distance to a target or an azimuth direction. For example, to estimate azimuth direction, MUSIC (MUltiple SIgnal Classification), a method using eigenvalue expansion of the correlation matrix has been often used[1]. MUSIC usually requires a microphone array with a number of microphone elements. Kinect is a device that can be operated by gesture and/or speech recognition released from Microsoft and it is known to have many sensors (e.g., depth sensor, RGB camera, color camera sensor, four-element microphone array and so on). The depth sensor of Kinect v2 determines distance from Time-of-Flight (TOF) of infrared ray. The sensing range of the depth sensor is adjustable, but it is difficult to measure the distance closer than 0.500 m[2]. Alternatively, acoustic distance measurement (ADM) method has been proposed based on the phase interference between transmitted and reflected waves[3]. ADM method estimates the close range by applying the principle of microwave radar to a distance measurement method using an acoustic signal[4]. The procedure of the ADM is as follows: first, to obtain a power spectrum, applying Fourier transform to the observed wave, and then, to obtain a range spectrum, applying the Fourier transform again to the power spectrum and taking its absolute value. The power spectrum is a periodical function, whose period is inversely proportional to the distance between microphone and target. Therefore, the distance can be determined from the range spectrum. This method can estimate the distance closer than 0.500 m.

In this study, we focus on Kinect v2's microphone array and perform this ADM method. Then we compare measurement accuracies of Kinect v2's depth sensor and the ADM method. More concretely we confirmed the accuracy of the ADM method by performing a computer simulation under the assumption of using Kinect v2's microphone array and by applying it to an actual sound field. Finally, we applied the ADM method to the actual sound field using four-channel microphone array (Kinect v2) and we geometrically estimate the position of the target by combining distances estimated at each channel. As a result, we could estimate distance and position to a target by using Kinect.

2. ADM method based on the phase interference[2]

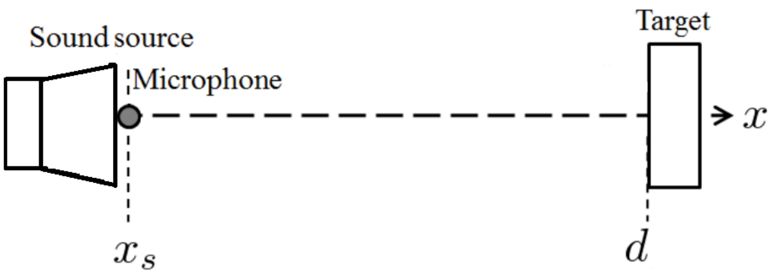


Figure 1: Geometrical configuration of sound source, microphone and target.

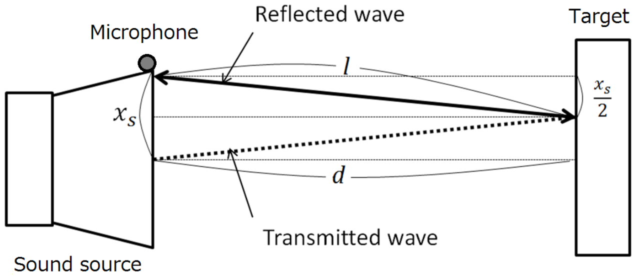


Figure 2: Geometrical configuration of sound source, microphone and target of experiment.

In conventional ADM based on phase interference, a microphone is located between a target and a loudspeaker as shown in Fig. 1,

Let $v_{\mathrm{T}}(t, x_s)$ be a transmitted wave, as follows:

$$v_{\mathrm{T}}(t, x_s) = \int_{f_1}^{f_N} A(f) e^{j(2\pi f t - \frac{2\pi f x_s}{c} + \theta)} df, \tag{1}$$

where t [s] is time, x_s [m] is the position of microphone, c [m/s] is the velocity of sound. $A(f)$ and $\theta(f)$ are the amplitude and initial phase at a frequency f [Hz]. f_1 [Hz] and f_N [Hz] represent the lowest and highest frequency, respectively.

The wave $v_{\mathrm{R}}(t, x_s)$ reflected by the target can be expressed as follows:

$$v_{\mathrm{R}}(t, x_s) = \int_{f_1}^{f_N} A(f)\gamma e^{j\left\{2\pi f t - \frac{2\pi f}{c}(2d - x_s) + \theta + \phi\right\}} df, \tag{2}$$

where γ and ϕ [rad] are the magunitude and phase of reflection coefficient of target. Let positions of the sound source(microphone) and the target be $x_s = 0$ m(origin) and $x = d$ m, respectively, supposing a single sound source and a single target. The composite wave $v_{\mathrm{C}}(t, x_s)$, which is the composition of transmitted and reflected waves, is formulated as follows:

$$v_{\mathrm{C}}(t, 0) = v_{\mathrm{T}}(t, 0) + v_{\mathrm{R}}(t, 0). \tag{3}$$

Now let $V_{\mathrm{C}}(f, 0)$ be a frequency spectrum of composite wave $v_{\mathrm{C}}(t, 0)$. The power $p(f, 0)$ at each frequency f [Hz] of the observed wave is the square of the absolute value of $V_{\mathrm{C}}(f, 0)$. When magnitudes of reflection coefficient is sufficiently small($\gamma \ll 1$), the power spectrum $p(f, 0)$ can be approximated as follows:

$$p(f, 0) \approx A(f)^2 \left\{1 + 2\gamma \cos\left(\frac{4\pi f d}{\mathrm{c}} - \phi\right)\right\}. \tag{4}$$

In Eq. (4), the first term represents the component of the transmitted wave and the second term represents the component of interference between the transmitted and reflected waves.

If $A(f)$ is constant regardless of frequency f [Hz], $p(f, 0)$ is a periodical function whose period is inversely proportional to the distance d between microphone and target. So the distance d can be obtained by applying Fourier transform to $p(f, 0)$ again. In order to extract only the fluctuation component related to the distance, subtracting the average value $\overline{p(f, 0)}$ from $p(f, 0)$ $\Big($we define $\Delta p(f, 0)$ as $\Delta p(f, 0) = p(f, 0) - \overline{p(f, 0)}\,\Big)$ and applying the Fourier transform yield to:

$$P(x) = \int_{f_1}^{f_N} \Delta p(f, 0) e^{-j2\pi \frac{2x}{c} f} df. \tag{5}$$

The absolute value $|P(x)|$ of $P(x)$ is referred to as a range spectrum. The peak position of the range spectrum corresponds to the estimated value of the distance d between the microphone and the target. Minimum measurable distance $d_{\min}$ is defined in terms of the frequency bandwidth $f_N - f_1$ [Hz] and velocity of sound c [m/s] as follows:

$$d_{\min} = \frac{c}{2(f_N - f_1)}. \tag{6}$$

In this research, the microphone is set on the loudspeaker as shown in Fig. 2. In this situation, the conventional ADM estimates the distance as the half difference between paths of direct and reflected waves from sound source. From Fig. 2, since paths of direct and reflected waves from sound source are 2ℓ [m] and x_s [m], respectively, the distance estimated by the ADM method is $d' = \ell - x_s/2$ [m]. Thus the distance $\hat{d}$ is obtained by correcting the estimated distance d' with use of the distance x_s between sound source and microphone.

$$\hat{d} = \sqrt{(d')^2 + x_s d'}. \tag{7}$$

3. Computer simulation

To confirm the validity of the principle of the original ADM method, we performed the simple computer simulation here.

3.1 Simulation conditions

Simulation conditions are shown in Table 1. We use a band-limited impulse signal as a transmitted wave shown in Fig. 3. The distance between the sound source(microphone) and the target was set to 0.300 m. In calculating range spectrum, we apply 0-padding to the power spectrum in frequency domain so that the number of data points of the power spectrum become 2048. So the minimum step of estimated distance is 0.013m from Eq. (6).

Table 1: Simulation conditions.

Sound source		Band-limited impluse
Sampling frequency		16 kHz
Frequency bandwidth		2.0 kHz (2.1 kHz$\sim$4.1 kHz)
Sound speed		340 m/s
Reflection coefficient		$\gamma = 0.1$,$\phi = 0$ rad
Data points (Time domain)		2048
Data points (Frequency domain)	before 0-padding	256
	after 0-padding	2048
Minimum measurable distance	before 0-padding	0.087 m
	after 0-padding	0.013 m

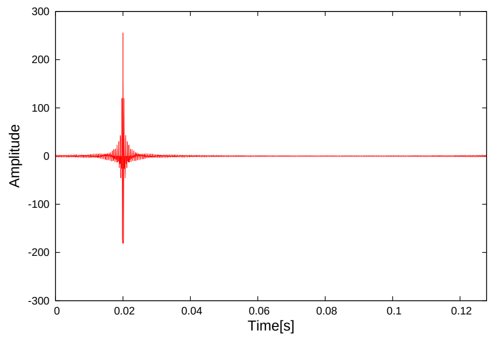


Figure 3: Transmitted wave $v_{\mathrm{T}}(t, 0)$

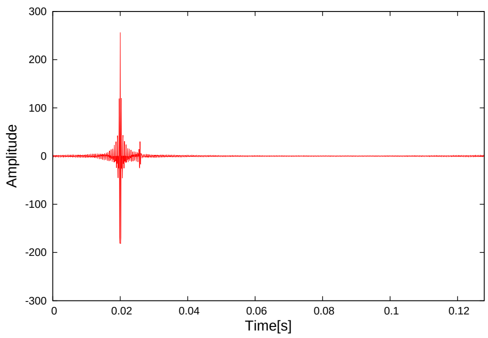


Figure 4: Composite wave $v_{\mathrm{C}}(t, 0)$

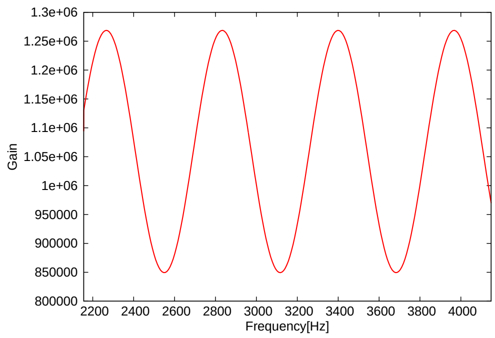


Figure 5: Power spectrum $p(f, 0)$

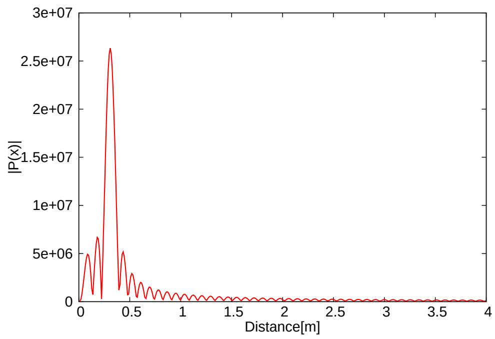


Figure 6: Range spectrum (after 0-padding)

3.2 Simulation results

Figures 4, 5 and 6 respectively show the composite wave, the power spectrum of composite wave, and the range spectrum obtained by the ADM method. The power spectrum in Fig. 5 has periodicity and includes distance information. In Fig. 6, we can see one peak of range spectrum, whose position corresponds to the distance between microphone and target. The peak position in Fig. 6 is 0.308 m. The error between the estimated distance and the true distance is smaller than the minimum measurable distance 0.013 m.

4. Experiment in a real sound field

In order to confirm the effectiveness of this method, we perform distance measurement using the Kinect v2's microphone array in the actual sound field, and try to estimate the position of the target by using the result.

4.1 Experimental conditions

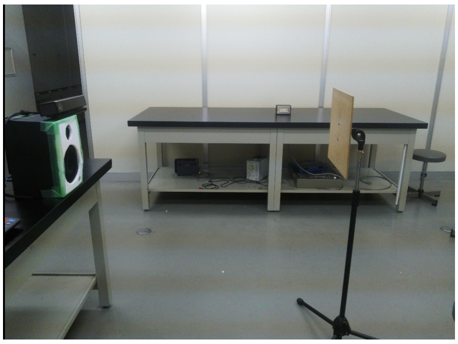
Table 2 shows the experimental apparatus. We adopt a plywood square as a target. The measurement is performed in two cases where the true distances between sound source and target are 1.000 m and 0.300 m. We experimented distance measurement using Kinect v2's microphone array in a real sound field. Figures 7(a) and (b) show the actual situation of measurement, and Kinect v2's microphone array and loudspeaker used here. Experimental conditions are shown in Table 3, which are basically same as the simulation conditions.

Table 2: Experimental apparatus.

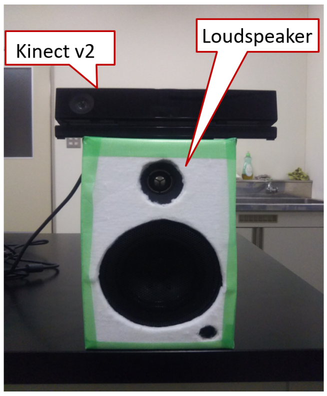
Target	Plywood square (H:30cm × W:30cm × D:0.5cm)
Audio interface	M-AUDIO, Fast Track Ultra 8R
Loudspeaker	YAMAHA, MSP5 STUDIO
Microphone	Microsoft, Kinect v2

Table 3: Experimental conditions.

Room temperature		24 °C
Sound speed		345.9 m/s
Between the sound source and microphones	ch 1	0.105 m
	ch 2	0.070 m
	ch 3	0.060 m
	ch 4	0.135 m



(a) Real sound field.



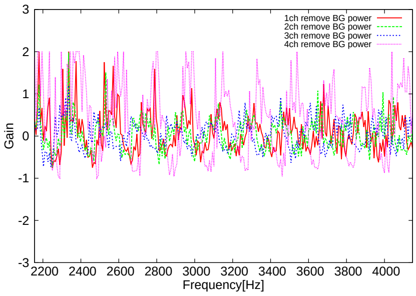
(b) Kinect v2's microphone array and loudspeaker

Figure 7: Experimental environment.

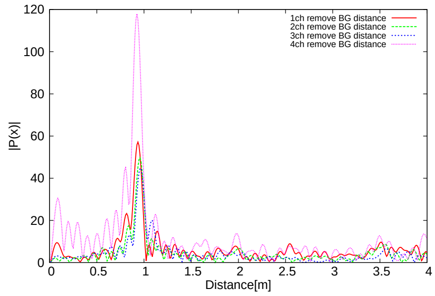
4.2 Experimental results

4.2.1 Distance estimation

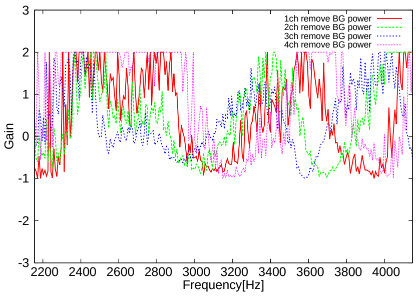
Figures 8(a) and (b) show the power spectra and the range spectra for true distance 1.000 m respectively. The 4 curves in Figs. 8 (a) and (b) show the power spectra and range spectra obtained by each microphone respectively for true distance 1.000 m. Similarly Figs. 8 (c) and (d) show the power spectra and range spectra obtained by each microphone respectively for true distance 0.300 m. All of these power spectra have periodicity, although the power spectra are saturated. From Eq.(4), there is doubled cosine term in the power spectrum, so we introduce the threshold range of the power spectrum $\Delta p(f, 0)$ as $-2 \leq \Delta p(f, 0) \leq 2$. If the $\Delta p(f, 0)$ exceeds threshold, 2 is substituted into $\Delta p(f, 0)$ when $\Delta p(f, 0)$ becomes bigger than 2. Similarly -2 is substituted into $\Delta p(f, 0)$ when $\Delta p(f, 0)$ becomes smaller than -2.



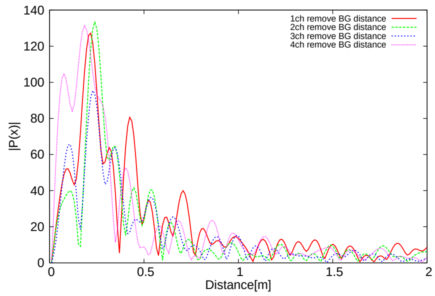
(a) Power spectra for 1.0 m.



(b) Range spectra for 1.0 m.



(c) Power spectra for 0.3 m.



(d) Range spectra for 0.3 m.

Figure 8: Experimental results.

For d =1.000 m and 0.300 m, comparison between estimated and true values are shown in Tables 4 and 5. Estimated distances in Tables 4 and 5 are the distances simply estimated by the original ADM method and corrected distances geometrically by Eq. (7). The errors of the corrected distances for 1.000 m and 0.300 m obtained by Kinect v2's microphone array are smaller than the minimum measurable distance 0.013 m. As a result, ADM method can estimate the distance closer than 0.500 m. Therefore, we could also estimate distance with a high degree of accuracy even in an actual sound field.

Table 4: Comparison between estimated and true values for $d = 1.000$ m.

Microphone	True value	Estimated distance	Corrected distance	error
ch1	1.00m	0.936 m	0.990 m	0.010 m
ch2	1.00m	0.958 m	0.995 m	0.005 m
ch3	1.00m	0.958 m	0.990 m	0.010 m
ch4	1.00m	0.922 m	0.993 m	0.007 m

Table 5: Comparison between estimated and true values for $d = 0.300$ m.

Microphone	True value	Estimated distance	Corrected distance	error
ch1	0.300m	0.258 m	0.305 m	0.005 m
ch2	0.300m	0.279 m	0.312m	0.012 m
ch3	0.300m	0.267 m	0.297m	0.003 m
ch4	0.300m	0.225 m	0.293m	0.007 m

4.2.2 Direction estimation

As shown in Fig. 9 we put a target in front of sound source and estimate direction. Similarly we put a target in the direction of 30 degrees from sound source as shown in Fig. 10. The position and direction of the target can be estimated from distance information by the array microphone. Figures 11 and 12 show the results of direction estimation. Although ADM method can only estimate distance, it is possible to estimate direction by using multiple microphones. In this research, we draw the circle around each microphone and find the direction using intersection of the circle. The radius of the circle is the estimated distances at each microphone, and all combination of intersection of the circle ${}_4\mathrm{C}_2 = 6$. The intersection used this time takes the average of 6 points, ignoring the point far from the average. As a result, we could estimate the approximate position of the target. As a result in Fig. 12, the range of target direction is 34.2 to 35.0 degrees. So we could estimate the approximate position of target other than front.

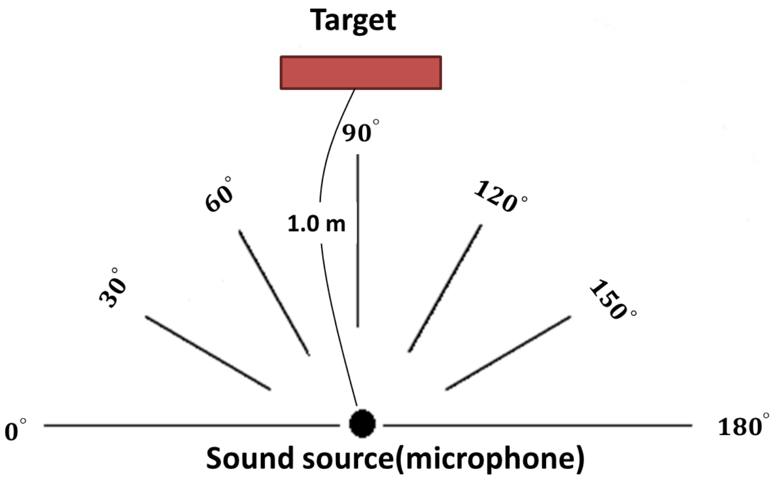


Figure 9: Target position (in front).

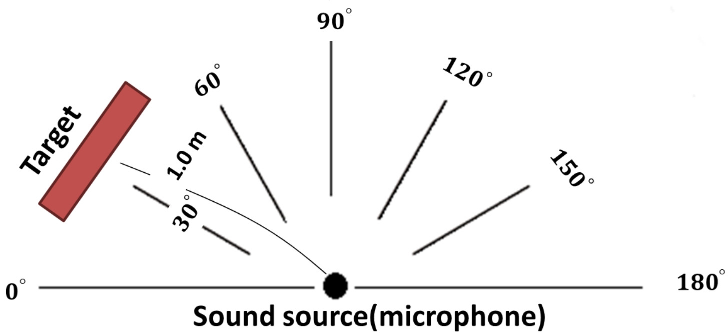


Figure 10: Target position (30 degree).

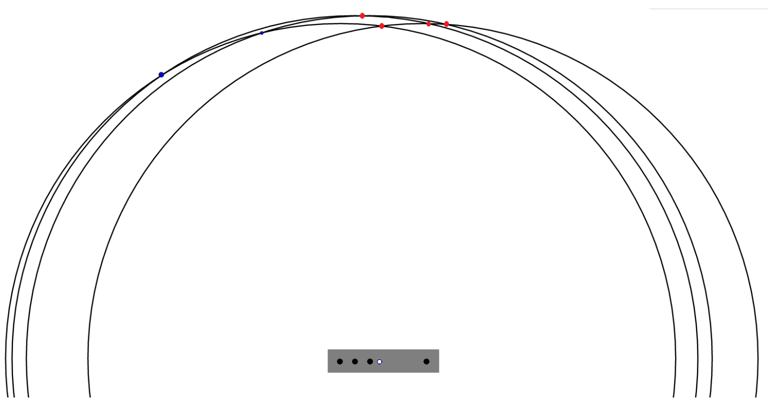


Figure 11: Direction estimation of front.

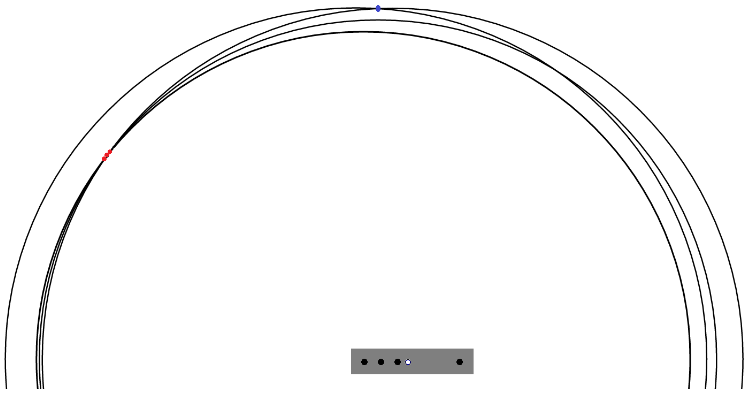


Figure 12: Direction estimation of 30 degree.

5. Conclusion

In this study, we used a Kinect v2's microphone array as 4ch microphones in estimating the distance using ADM method. As a result, the ADM method could estimate the distance closer than 0.500 m. In addition, the position of the target could be approximately estimated using intersection of the circles drawn around each microphone. The radii of these circles are the estimated distances at each microphone. This method can estimate approximately target's position but it cannot estimate accurate position of the target. So we will consider methods to improve accuracy.

Acknowledgments

The present study was supported in part by a grant from the Strategic Research Foundation Grant-aided Project for Private Universities from the Ministry of Education, Culture, Sport, Science, and Technology, Japan (MEXT), 2013-2017 (S1311045).

REFERENCES

1. S.U.Pillai, *Array Signal Processing*, **Springer-Verlag New York Inc.**, 1986.

2. Microsoft, *Kinect hardware.* [Online.] available: `https://developer.microsoft.com/en-us/windows/kinect/hardware`

3. Nakasako, N., Uebo, T., Shinohara, T. *Distance Estimation Using Audible Sound : Principle of Distance Estimation Based on Standing Wave*, System/control/information, **54** (8), pp.314∼319, 2010.

4. Uebo, T., Okubo, Y. and Iritani, T. *Standing Wave Radar Capable of Measuring Distances Down To Zero Meters.*, IEICE Trans. Commun., **E88-B** (6), 2609-2615 (2005).