Feasibility Experiment on Position Estimation of Various Sound Sources in Indoor Environment

Takeru Kadokura Shigenori Ioroi Hiroshi Tanaka
Course of Information & Computer Science
Graduate School of Kanagawa Institute of Technology
Atsugi-Shi, Kanagawa, Japan
email: s1885004@cco.kanagawa-it.ac.jp, {ioroi, h_tanaka}@ic.kanagawa-it.ac.jp

Abstract— **This paper describes a method for estimating the position of various sound sources in an indoor environment and presents the results of basic positioning experiments. To date, we have performed position estimation based on the Time Difference Of Arrival (TDOA) method. In this investigation, the reception time difference at each reception point is compared to the theoretical value using a diffused sound, the human voice, the sound made by an operating microwave oven, and the ringing of a telephone. The reception time difference is necessary for position estimation. The reception time difference was obtained by cross correlation processing. Then, although the sound source position was obtained using this result, satisfactory positioning accuracy was not obtained except for the diffused sound source. Therefore, the whitening cross correlation method called Cross-power Spectrum Phase (CSP) analysis was applied. As a result, we could obtain a more accurate time difference than simple cross correlation for all sound sources, and we obtained a prospect that high accuracy positioning is possible.**

Keywords-Indoor Positioning; Sound Source; TDOA; Cross Correlation; CSP Analysis.

I. INTRODUCTION

We have been studying a high-accuracy indoor positioning system using sounds. A special sound source that transmits ultrasonic waves [1] or diffused sound [2] as a sound source was used in this system. In this study, we experimentally investigated the feasibility of estimating the position of a sound source originating in an indoor environment, such as the human voice and the sounds made by electrical appliances.

If the location of these sound sources can be established, it should be possible to monitor the conversation environment in an office by combining it with speaker recognition technology, and can be used to monitor an operating situation of household appliances. By using the position information of the sound source, it should be possible to identify the position of operating household appliances with a high degree of accuracy by incorporating sound classification technology. For example, this could be applied to watching over an elderly person living alone by monitoring the usage of home appliances. The proposed technology would appear to have many potential applications. In addition to a diffused sound that is used in conventional positioning systems, the source position of the human voice, the sound a microwave oven makes while operating and the ringing sound of a telephone were investigated in this paper.

The rest of the paper is structured as follows. The Section 2 describes the differences between related work and this paper. The basic principle of our proposed method is shown in Section 3. In Section 4 and 5, we show the experimental results of detecting the reception time difference from various sound sources by the simple cross correlation, and the result of the positioning experiment using these time differences. It is shown that it is difficult to satisfy the target positioning accuracy. We apply the CSP analysis and show the result of obtained reception time difference in Section 6. By applying this method, we can detect the reception time difference with higher accuracy, and show that we got the prospect of high accuracy positioning. This may lead to privacy concerns, and it is considered necessary to discuss and examine from this viewpoint. However, this paper focuses on the technology for position detection and does not discuss the viewpoint of privacy.

II. RELATED WORK

In position estimation using sound, research is being conducted to obtain accurate position of the sound source. It uses ultrasonic waves [3] or a dedicated sound source [4] for positioning, thus a special sound source is necessary for use, which is an impediment to popularization. On the other hand, for many environmental sound sources, the estimation of arrival direction is investigated rather than the estimation of the position of the sound source. In these methods, for example MUltiple SIgnal Classification (MUSIC) method is proposed and the direction of arrival of sound is detected from the spatial spectrum called the MUSIC spectrum [5][6].

CSP analysis [7] is a method used for detecting the difference in time of arrival of acoustic waves from a sound source to two microphone sensors. Although a method of estimating the position of the sound source using this information has been proposed, the authors estimated the position by statistical processing or filter processing, and accuracy is about several tens of centimeters [8][9]. As far as we know, there are no cases where the realized high accuracy is about several cm using environmental sound source other than the dedicated sound source. This research aims to realize the accurate positioning of the sound source in indoor space without the ultrasonic wave or the diffused sound source.

 ISBN: 978-1-61208-679-8

III. BASIC PRINCIPLE OF PROPOSED METHOD

The location estimation method which we have applied to date is based on the TDOA scheme. In our method, a special sound source is prepared and transmits a sound that has been diffused using an M sequence code. The receiving side has the same sound source data as that of the transmitting side (replica), and detects the sound reception timing by cross correlation calculation between the received signal and the replica. Positioning calculation is conducted by using the reception time differences for each receiver. The positioning is conducted by solving the following equation (1) using numerical computation. This equation for positioning is the same as that of Global Positioning System (GPS) / Global Navigation Satellite Systems (GNSS) in which the radio signals is used.

$$\begin{aligned}\sqrt{(x-x_0)^2+(y-y_0)^2+(z-z_0)^2} &= ct \\ \sqrt{(x-x_1)^2+(y-y_1)^2+(z-z_1)^2} &= c(t+t_1) \\ \sqrt{(x-x_2)^2+(y-y_2)^2+(z-z_2)^2} &= c(t+t_2) \\ \sqrt{(x-x_3)^2+(y-y_3)^2+(z-z_3)^2} &= c(t+t_3)\end{aligned} \quad (1)$$

where,

t : propagation time [s]
x, y, z : position of transmitter [mm]
t_i : propagation time difference to each microphone sensor [s]
c : speed of sound [mm/s]
x_i, y_i, z_i : installation position of each microphone sensor [mm]

In this investigation, we considered the human voice and several other sound sources, such as home electrical appliances, in an indoor environment. Therefore, it is difficult to implement a replica on the receiving side as it can be done when the conventional method is used. In the proposed method, the reception time difference is obtained by cross correlation between the signal received at the reference point and the signal at each reception point, as shown in Figure 1. Based on the reception time differences obtained with this configuration, positioning calculation is performed in the same manner as that in the conventional method.

IV. RECEPTION TIMING EXPERIMENT

The experimental setup is shown in Figure 2. A sound source was installed on the floor (in this case, the previously

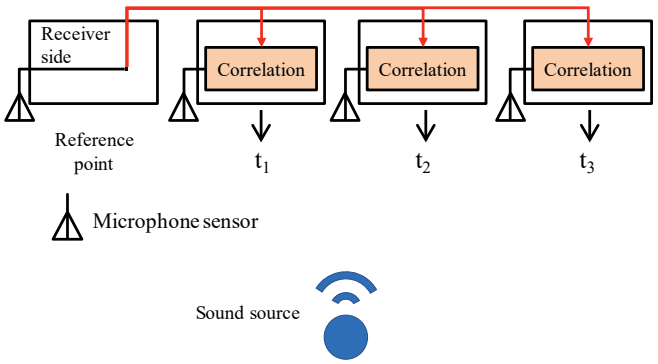
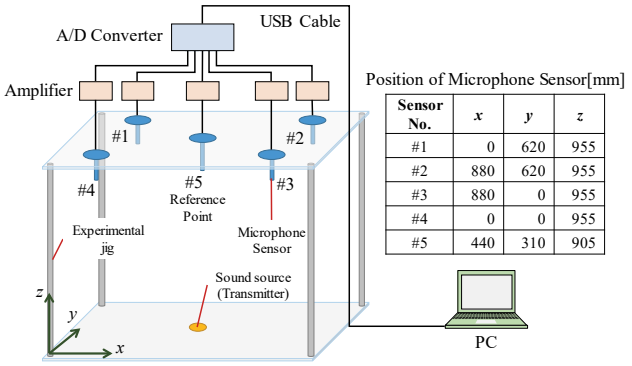


Figure 1. Positioning principle.



Position of Microphone Sensor[mm]

Sensor No.	x	y	z
#1	0	620	955
#2	880	620	955
#3	880	0	955
#4	0	0	955
#5	440	310	905

Figure 2. Experimental setup.

recorded sound source (WAV format) was reproduced by a speaker), and the time at which the sound was received (reception time) by each microphone sensor was obtained by correlation processing. The distance between the sound source and the microphone sensor was kept at about 1 m.

The received waveform (received at sensor (receiving point # 5) from the sound source) is shown in Figure 3. The sampling rate was 0.01 ms. The microphone sensors and speaker used in this experiment were the Primo EM-158 and Tang Band W2-858SB, respectively. The diffused sound source by the ninth order of the M-sequence code used in the positioning system we developed is also shown for comparison. The sounds of an operating microwave oven, a ringing phone and the human voice were examined in this investigation.

The blue line is the result for a diffused sound; the variations within a short time span can be attributed to the spread spectrum by chips (chip rate: 0.04 ms) of the M sequence code. The other signals, i.e., microwave oven and phone, are shown for the same time duration. It was confirmed that sound could be received by all microphone sensors.

Based on the received signal at sensor #5, the cross correlation with the received signal at each sensor was calculated for each sound signal. The cross correlation results for each sound source between sensors #5 and # 1 are shown in Figure 4 as one example. These results are obtained by the function “xcorr” of MATLAB.

The sound source diffused by M sequence code is the clearest and shows the cyclic peak of the correlation value. Peak position can be regarded as the time difference between

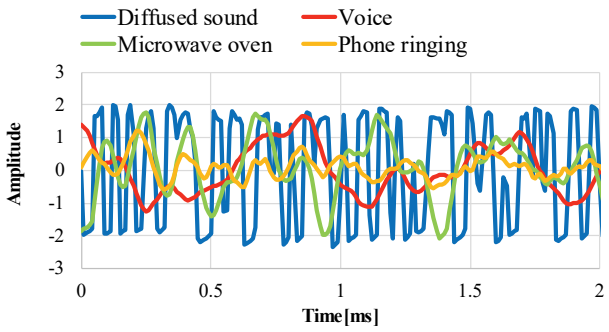


Figure 3. Received waveform.

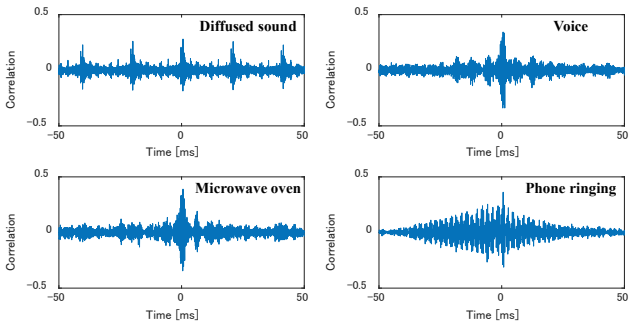


Figure 4. Example of cross correlation results.

two sound reception points. The correlation value can also be obtained from other sensors (receiving points).

We took 100 timings obtained from the peak position, and their average (μ) and standard deviation (σ) were evaluated before conducting the positioning experiment. In this experiment, the peak position, i.e., the time having the highest value, is considered to be the time difference. Table I shows the reception time differences obtained by correlation processing at each sensor as an evaluation of the reception time difference. As shown in Figure 2, the reception signals at sensors from #1 to #4 were used to calculate time difference and the signal at sensor #5 was used as reference. The theoretical value obtained from the sound velocity and distance between sensor #5 and each of the other sensors is shown in the right-hand column.

The differences between the experimental value and the theoretical values and the difference at each sensor result from the error of the peak position appearance caused by, for example, the effect of multipath signals or a low signal-to noise ratio of each of the sensors. The quality of the phone ringing was the worst in this experiment. Therefore, it seems to be difficult to get good positioning accuracy from the sound made by a phone.

V. POSITIONING RESULTS

The time difference values obtained by this method were used for the positioning calculation. Figure 5 shows the results of the positioning experiment obtained by correlation processing. The sound source was set at the center point (440, 310). As shown in Figure 2, it was confirmed that a positioning result closer to the center point can be obtained when the reception time difference is closer to the theoretical value.

As is clear from the result of the reception time difference, the sound source diffused by the M sequence, which is a

TABLE I. RECEPTION TIME DIFFERENCE IN EACH POINT

	Difference of Reception Timing[ms]								
	Diffused sound		Voice		Microwave oven		Phone ringing		Theoretical value (Temp.:25°C)
	μ	σ	μ	σ	μ	σ	μ	σ	
t_1	0.58	1.1E-18	0.99	3.1E-4	0.55	1.2E-4	0.98	9.8E-4	0.58
t_2	0.58	1.4E-6	0.66	2.9E-4	0.25	1.4E-4	0.70	1.5E-3	0.58
t_3	0.58	3.3E-6	0.61	1.8E-4	0.38	1.9E-4	1.21	2.8E-3	0.58
t_4	0.58	1.7E-6	0.62	2.3E-4	0.61	5.6E-4	0.89	1.3E-3	0.58

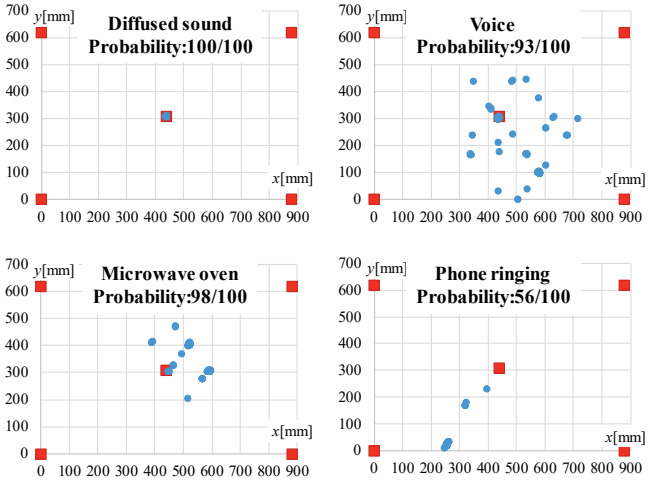


Figure 5. Positioning experiment results.

dedicated sound source used for positioning, gave the most accurate results. The numerical values in each graph, that is, the denominators and numerators in the graph, are the number of positioning times, and the number of times that the positioning result is within the range surrounded by the four reception points, respectively.

After the diffused sound, the sound generated by the operating microwave oven has the highest accuracy. Although the human voice can also be used to determine positioning, the error is significant. In this experiment, a microphone sensor was installed using a relatively small frame in order to construct an easy to use experimental system. The microphone sensors have been spaced at relatively narrow intervals of 620 mm and 880 mm. In practice, it is considered preferable for the microphone sensors to be placed at wider intervals from the viewpoint of installation load and cost. The time at peak position obtained from the correlation calculation, which governs the positioning accuracy, under a decrease in the reception intensity and in a multipath environment, is unclear. It would appear that the detection accuracy needs to be confirmed when widening the sensor interval.

The positioning error results for each sound source are summarized in Table II. In this table, the average value and the Root Mean Square (RMS) value in the x-y plane of positioning error of each sound source are shown for all 100 positioning times. Here, the average value and the RMS value, both results including and excluding the points outside the range of the four reception points, are shown. The RMS result

TABLE II. POSITIONING ERROR

	All measurement[mm]			Region within 4 points[mm]		
	x average	y average	RMS	x average	y average	RMS
Diffused sound	0.23	-0.39	1.52	0.23	-0.39	1.52
Voice	63.13	-122.1	224.36	68.06	-104.16	197.57
Microwave oven	127.15	35.34	380.88	100.30	36.40	130.75
Phone ringing	44.68	-42.84	2300.13	-182.50	-279.46	347.64

 ISBN: 978-1-61208-679-8

of phone ringing was 2300; this was because some solutions have extremely large error.

For products with large dimensions, such as microwave ovens, an accuracy of about several centimeters is not necessarily required. However, in the case of positioning for small objects, such as call buzzers or alarms and when controlling the moving objects with high accuracy, it is necessary to suppress the positioning error to less than several tens of centimeters. It is of the utmost importance to devise a method that is able to reduce the positioning error.

The positioning error seems to be caused by the error of reception time deference obtained cross correlation calculation shown in Table I. A more accurate detection of reception time difference is required to keep positioning accuracy.

VI. RESULTS BY CSP ANALYSIS

There is a method of detecting the reception time difference by CSP analysis [7]. This method is also called a whitening correlation method, and it is said that the time difference can be accurately detected even for a sound source which is not whitened different from a diffused sound source. Figure 6 shows an example of the result by the cross correlation method and CSP analysis. This result shows the possibility that the CSP analysis can detect the reception time difference more accurately.

Table III shows the result of obtaining the reception time difference by the CSP analysis for the same sound sources shown in Table II. Significant improvements in detection accuracy can be confirmed. Although a large error (2.36 ms) occurs in the detection of a part of phone ringing, this error can be removed as an abnormal value from other conditions, such as the reception range of sound waves. By using this analysis, improvement of the positioning accuracy can be expected. We will use this method for the detection of the reception time difference and plan to perform the positioning experiment.

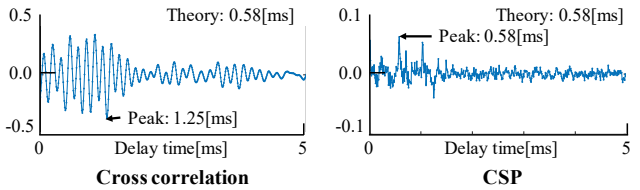


Figure 6. Difference of results by cross correlation and CSP analysis.

TABLE III. RECEPTION TIME DIFFERENCE BY CSP ANALYSIS

	Difference of Reception Timing[ms]								
	Diffused sound		Voice		Microwave oven		Phone ringing		Theoretical value (Temp.:25°C)
	μ	σ	μ	σ	μ	σ	μ	σ	
t_1	0.58	1.1E-18	0.46	6.6E-4	0.57	9.4E-5	0.59	1.0E-6	0.58
t_2	0.58	3.1E-6	0.57	1.3E-4	0.47	2.5E-4	0.57	1.0E-6	0.58
t_3	0.58	1.1E-18	0.57	1.4E-4	0.49	2.0E-4	0.56	3.9E-6	0.58
t_4	0.58	4.0E-6	0.55	1.8E-4	0.56	1.2E-4	2.36	1.3E-2	0.58

VII. CONCLUSION

We have experimentally examined a highly accurate positioning method for human voice, the sound generated by an operating microwave oven and a ringing phone. To date, sound source diffused by M sequence code is used for accurate indoor positioning in conventional systems. There would be numerous applications if other sound sources commonly present in indoor environments could be used for the positioning of a sound source.

Although the sound received timing used for positioning can be obtained by simple cross correlation calculation for each sound source, it was not possible to secure adequate positioning accuracy except for the diffusion sound source. For this reason, we conducted an experiment to find the accurate difference in reception time by applying the CSP analysis. As a result, the error of the reception time difference can be greatly reduced, and the prospect that the positioning with higher accuracy can be obtained. Future plans are to evaluate the positioning accuracy by using the time difference obtained by this method. Furthermore, it is necessary to evaluate the positioning accuracy with a more realistic configuration, that is, widening the installation intervals of the microphone sensors.

ACKNOWLEDGMENT

This work was supported by JSPS KAKENHI Grant Numbers JP17K00140.

REFERENCES

[1] C. Yara, M. Akiyama, S. Ioroi, and H. Tanaka, “Indoor Positioning System using Ultrasonic Sensors as a Position Information Infrastructure for a Wide Area,” Asian Conference on Information System 2013 (ACIS2013), pp. 386-389, 2013.

[2] K. Kaneta, M. Naruoka, S. Ioroi, and H. Tanaka, “Accurate Indoor Positioning using Sound with Distributed System Configuration,” The Fifth Asian Conference on Information Systems (ACIS2016), pp. 114-120, 2016.

[3] A. Smith, H. Balakrishnan, M. Goraczko, and N. Priyantha, "Tracking Moving Devices with the Cricket Location System," ACM MobiSYS 2004, 13 pages, 2004.

[4] A. Mandal, C. V. Lopes, T. Givargis, A. Haghighat, R. Jurdak and P. Baldi, “Beep: 3D Indoor Positioning Using Audible Sound,” Consumer Communications and Networking Conference, pp. 348-353, 2005.

[5] F. Asano, M. Goto, K. Itou and H. Asoh, “Real-time Sound Source Localization and Separation System and Its Application to Automatic Speech Recognition,” Proc. of Eurospeech, pp. 1013-1016, 2001.

[6] P. Danes and J. Bonnal, “Information-Theoretic Detection of Broadband Sources in a Coherent Beamspace MUSIC Scheme,” Proc. of IROS, pp. 1976-1981, 2010.

[7] D. Rabinkin, R. Renomeron, J. French and J. Flanagan, “Estimation of Wavefront Arraival Delay Using the Cross-Power Spectrum Phase Technique,” 132nd Meeting of the Acoustical Socienty of America, pp. 1-10, 1996.

[8] M. Omologo and P. Svaizer, “Use of the Crosspower-spectrum Phase in Acoustic Event Location,” IEEE Trans. on Speech and Audio Processing, vol. SAP-5, no.3, pp. 288–292, 1997.

[9] H. Okumura, K. Cho, T. Nishiura, Y. Yamashita, “Sound Source Localization Using a Distributed Microphones System,” IEICE Technical Report, SP2006-139, pp. 61-66, 2007 (in Japanese).

 ISBN: 978-1-61208-679-8