Initial Investigation of Position Determination of Various Sound Sources in a Room

Takeru Kadokura Yuki Hashizume Shigenori Ioroi Hiroshi Tanaka
Course of Information & Computer Science
Graduate School of Kanagawa Institute of Technology
Atsugi-Shi, Kanagawa, Japan
email: {s1885004, s1621081}@cco.kanagawa-it.ac.jp, {ioroi, h_tanaka}@ic.kanagawa-it.ac.jp

***Abstract*—Special sounds, such as ultrasonic waves and diffused sound sources used to perform high precision indoor positioning. If the positioning of various ordinary sound sources in a room becomes possible, it is thought that numerous applications are likely. This paper proposes a new method to estimate with high accuracy the position of various ordinary indoor sound sources, rather than using any special sound source. We constructed an environment where positioning experiments can be performed by knowing the size of the actual sonic environment. Positioning experiments were conducted using the sound of an operating microwave oven, the ringing tone of a telephone, and a diffused sound source. Highly accurate positioning about +/- 20 cm could be realized for the diffused sound and the microwave oven. In contrast, it was confirmed that the ringtone had a positioning error exceeding 1 m.**

Keywords-Positioning; Sound Source; TDOA; CSP Analysis; Real Environment.

I. INTRODUCTION

Many methods and technologies have already been proposed for indoor positioning, including the use of radio waves, such as Wi-Fi and BLE, and those using inertial sensors, such as acceleration sensors [1]. At present, the positioning accuracy and positioning area differ depending on the principle and method, and each system is individually selected according to the user's requirements. The most widely studied methods are those using a wireless LAN [2]. There is also a positioning method using BLE [3], which has a narrower radio propagation range. Although the devices used in this method are already in widespread use and are not limited in terms of use, the positioning accuracy is about m order, and high-accuracy position detection is difficult.

The authors have been studying systems capable of highly accurate positioning by using sound, with its low propagation speed. These systems attach sound sources to a positioning target, including ultrasonic [4] and spread spectrum sound [5], and then detect the target position. In this method, although positioning accuracy within several centimeters can be secured, attachment of a dedicated sound source is essential, so its application is restricted. There is also a system that achieves high positioning accuracy (about 10 mm) by obtaining the reception time of the sound wave with high accuracy using radio waves and sound waves [6]. However, the configuration is complicated, and the application area seems to be limited to a region where some extremely accurate positioning is required.

To examine a room, there are various sound sources, such as a device that emits a notification/warning sound like a calling buzzer, the voice of a person, and home electrical appliances. Also, a drone can be a sound source. If the positions of the various sound sources are known, the utility will be greatly expanded. For example, the location of the speaker, indoor flight control of the drone, and indirect identification of a sound source from the position of the sound source can be considered. A lot of studies have been conducted to estimate the direction of the sound sources in an indoor area [7] [8]. The purpose of these investigations is to separate each sound from multiple sound sources, and the position of the sound source is not obtained.

Our positioning method is based on the Time Difference Of Arrival (TDOA) method [9], which conducts positioning calculations using the reception time differences of the sound wave between a reference point receiver and another three or more reception points. The Cross-power Spectrum Phase (CSP) analysis method has been proposed [10] as a method to detect the reception time difference of sound waves at two reception points. There are also studies in which positioning was performed using the time difference obtained by this method. However, in these papers, the arrival direction of the sound source was obtained by time difference, the position was obtained as an intersection of the sound direction [11]. Although this is a simple calculation, the position could not be determined with high accuracy. Because it didn't use the information on the reception time differences obtained from other microphone sensors existing neighbors. The sound source is only speech sound.

This paper describes the possibility of high accuracy positioning in the same area as an actual room environment for various indoor sound sources. The presentation is as follows. Section 2 describes the positioning principle and the essential method of detecting the reception time difference including the method of selecting reference points. Section 3 shows an experimental environment where the receiving points are installed on the ceiling of the laboratory to secure the same area as an actual usage environment. Section 4 shows the results of detecting reception time differences for three types of sound sources, and the positioning results. Section 5 presents a summary.

II. POSITIONING PRINCIPLE

The positioning method we have applied is based on the TDOA scheme. The positioning calculation is conducted by

 ISBN: 978-1-61208-739-9

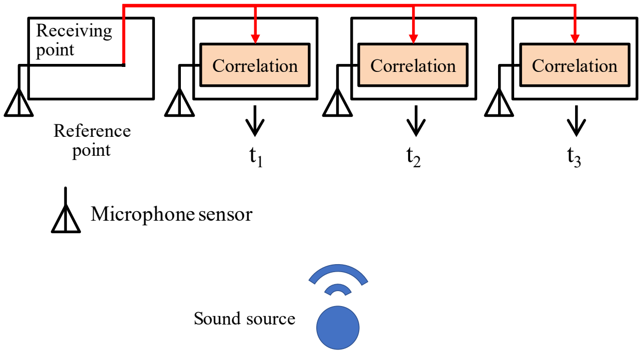

Figure 1. Principle of detection of time difference.

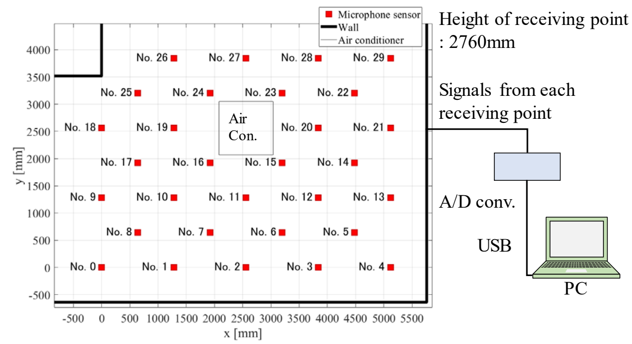

Figure 2. Microphone sensor installation position and experimental configuration.

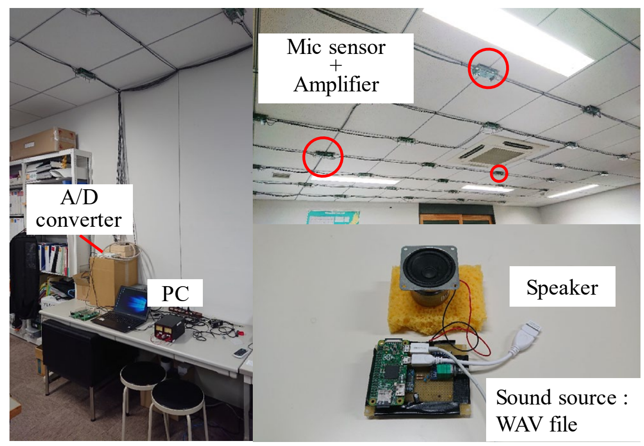

Figure 3. Experimental environment appearance.

using the reception time differences between a reference receiving point and some other points. Positioning is conducted by using (1). This equation for positioning is the same as that in a GPS/GNSS, in which radio signals are used. In previous studies, a special sound source provided a sound that had been diffused using an M sequence code. The receiving side had the same sound source data as that of the transmitting side (replica), and detected the sound reception timing by cross correlation calculation between the received signal and the replica [5].

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

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 several sounds in the room. This makes it difficult to implement a replica on the receiving side as can be done when the conventional method is used. The reception time difference is obtained by cross correlation between the signal received at the reference point and the signal at other reception points as shown in Figure 1. Based on the reception time differences obtained with this configuration, the positioning calculation is performed as it is in the conventional method. In this calculation, it is necessary to determine the reception point to use as the reference point.

III. Composition of Experiment Environment

To evaluate the reception time differences by the CSP method, a microphone sensor (Primo EM-258) was attached using a frame of about 1 m^2, and various sounds were produced by a speaker (Tang Band W2-858S) at the center of the frame, in the configuration used last year. In this investigation, the dimensions of the room were obtained to relate the sound detection time differences to the spatial separations, for positioning accuracy evaluation. The 30 receivers integrated with the microphone sensor and the amplification circuits were attached to the ceiling of the laboratory room. Figure 2 shows the installation positions and the experimental configuration. They were attached at intervals of 1.2 m, except for the area near the air conditioner. The height of the ceiling is 2.8 m. The appearance of the experimental environment is shown in Figure 3.

The sound of an operating microwave oven as a common home appliance, and the ring tone of a phone, were used as sound sources in the room. It was decided to evaluate using three types of sound sources, one of which was a diffused sound source from which high accuracy can be expected. The sounds were recorded in advance in WAV format, and each sound was sent out from a speaker, including the diffused sound source. The sampling frequency on the receiving side was 16 kHz, and the number of received samples was 10000.

IV. Experimental Results

In this section, the results of the experiment for detecting the reception time difference of each sound source, and the positioning experiment results obtained from the time difference are described by dividing into two subsections.

A. Time difference detection

The spectrogram of the sound sources is shown in Figure 4. It can be confirmed that the diffused sound source has a uniform frequency distribution and that the other sound sources have identifiable frequency characteristics. It was confirmed that all receivers could receive correctly.

The reception time difference between two receivers of the various sound sources was determined using the CSP method shown in (2).

$$\boldsymbol{CSP}_{x,y} = DFT^{-1}\left[\frac{DFT[\boldsymbol{x}]\overline{DFT[\boldsymbol{y}]}}{|DFT[\boldsymbol{x}]||DFT[\boldsymbol{y}]|}\right] \qquad (2)$$

where,

$\boldsymbol{CSP}_{x,y}$: normalized cross-power spectrum

$\boldsymbol{x}$: received signal at reference receiving point

$\boldsymbol{y}$: received signal at target receiving point

The reception time difference between two points is given as follows.

$$\tau = \arg\max_k(\boldsymbol{CSP}_{x,y}) \qquad (3)$$

The reception time difference is obtained from the data of two receiving points, that is, microphone sensors. The number of combinations of calculations for the reception time difference between two points becomes enormous, and the calculation time for that will greatly affect the positioning time interval. By determining a reference point for obtaining the reception time difference, it becomes a realistic calculation. The following two methods were compared in this investigation for selecting the reference point.

Method (1): the receiving point directly above the sound source (Since the sound source position is unknown, this method cannot be applied in real life circumstances.),

Method (2): the receiving point at which the signal strength is the highest, here, that strength is the sum of the sound energy received during the positioning time.

The reception time difference of various sound sources between two points, that is, a reference point and other points, was determined using the CSP method. An example of the cross correlation waveform by the CSP method is shown in Figure 5. The result from method (1) is in the upper column, and that from method (2) is the lower column. It was confirmed that although there is a difference in peak sharpness depending on the sound source, it is possible to detect the peak position indicating the reception time difference. The peak position was detected 100 times, and the mean and standard deviation were evaluated. The results are shown in Table I. Here, since the reference point changes every time in method (2) (reception strength fluctuates), evaluation method (2) was not used. Although the time difference matched with the theoretical value was obtained for the sound of the diffused and the microwave sound, it was confirmed that these values became unstable for the ringing tone of the telephone. The theoretical value is the reception time difference of the sound determined from the positions of the speaker and the two reception points.

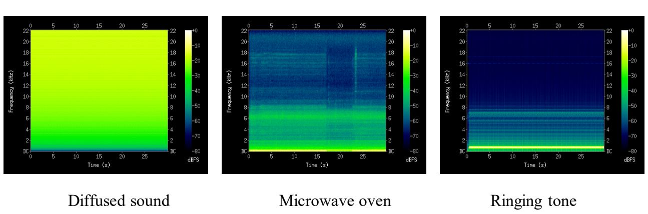

Figure 4. Spectrogram of received sound.

B. Positioning experiment result

The positioning results produced by 4 kinds of experimental conditions were examined. One sound source was set directly below reception point No.11, and another was set at an intermediate position between reception points No.14 and No.15. Two methods of reception time difference detection were applied to each case. Environmental noise was 36.9-55.6 dB at the time of the experiment.

An example of a positioning experiment result is shown in Figure 6. This is when the sound source was placed on a table 70 cm in height and below a position between reception points No.14 & No.15, and the reception time difference was detected by method (2). The positioning results of three sound sources are shown. While the positioning of the diffused sound and the operating sound of the microwave oven can be realized with high accuracy, a large error occurred in locating the ring tone of the telephone.

The positioning experiment results are summarized in Table II. Average errors and Root Mean Square (RMS) errors are shown for 100 times positioning results. Here, any solution that was out of the positioning range was excluded, that is, any that was not in the area of the receiving points, and the number of exclusions in 100 tests of positioning is also shown as a ratio in the table. The difference in positioning accuracy due to a particular sound source can be confirmed as shown in Figure 6.

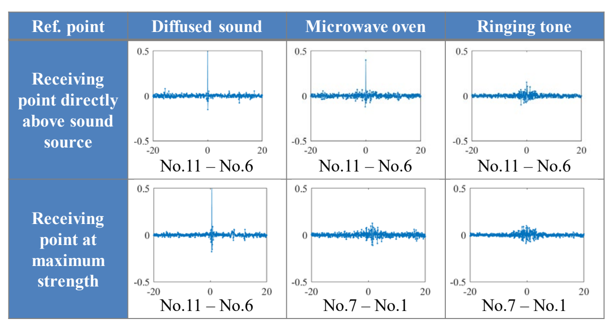

Figure 5. An example of cross correlation waveform by CSP method.

TABLE I. DETECTED RECEPTION TIME DIFFERENCES

	Reception time difference [ms]						
	Diffused sound		Microwave oven		Phone ringing		Theory
	μ	σ	μ	σ	μ	σ	μ
t1	0.56	1.1E-18	0.56	1.1E-18	0.24	7.9E-04	0.55
t2	0.56	1.1E-18	0.56	1.1E-18	0.32	7.1E-04	0.55
t3	0.56	1.1E-18	0.56	1.1E-18	0.23	5.0E-04	0.55
t4	0.56	1.1E-18	0.56	1.1E-18	0.17	7.0E-04	0.55

 ISBN: 978-1-61208-739-9

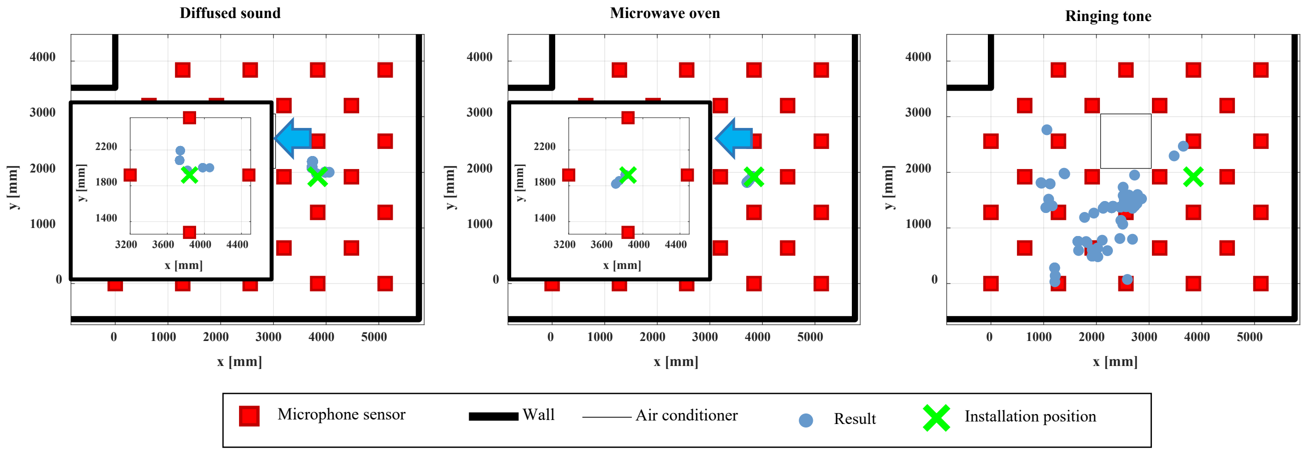

Figure 6. Example of positioning experiment result.

TABLE II. EVALUATION OF POSITIONING ACCURACY

Reference point	Fix (upper position) [mm]				Maximum receiving level [mm]			
Sound source position (2560,1280,700)								
Sound source	***X_average error***	***Y_average error***	***RMS error***	***Ratio***	***X_average error***	***Y_average error***	***RMS error***	***Ratio***
Diffused sound	0.0	0.0	0.0	100	0.0	0.0	0.0	100
Microwave oven	0.0	0.0	0.0	100	19.6	19.6	27.8	100
Phone ringing	-110.2	315.8	1118.5	87	-496.8	-143.7	1167.8	69
Sound source position (3840,1920,700)								
Sound source	***X_average error***	***Y_average error***	***RMS error***	***Ratio***	***X_average error***	***Y_average error***	***RMS error***	***Ratio***
Diffused sound	-105.5	181.7	213.9	100	-84.0	174.1	211.6	100
Microwave oven	-13.3	156.5	157.0	100	-39.1	-6.5	56.6	98
Phone ringing	-2.8	-666.6	667.2	100	-1717.4	-710.1	2006.2	98

The reason why the positioning accuracy by the ringing tone is worse than the other is that the reception time difference cannot be detected accurately as shown in Table I. The sound reception time is obtained from the peak position. The peak of ringing tone is not clear compared to those of the diffused sound and the microwave oven. The correlation between the two received points is obtained in the CSP calculation. The peak of CSP calculation becomes sharp when the sound source has a uniform frequency component over a wide frequency range. As can be seen from Figure 4, the ringing tone does not cover a wide frequency range compared to the diffused sound source and the microwave oven sound, so it is difficult to obtain a clear peak in the CSP calculation value. The error in reception time difference obtained from the peak is considered to be increased due to the unclearness of the peak because another peak might be selected as reception time difference. For this reason, the positioning accuracy is deteriorated for the ringing tone.

V. CONCLUSION

An experimental system was constructed in the space of an actual usage environment in order to clarify a positioning method for various sound sources in an indoor environment. A method of selecting a reference point to detect the reception time difference of sound waves was investigated and the time difference was obtained by using the CSP method. The positioning experiment was carried out using the TDOA method.

Together with a diffused sound source, which is a dedicated sound source expected to allow high positioning accuracy, position estimation was performed using the sound of an operating microwave oven and the ringing of a telephone as indoor sound sources. High-precision positioning resulted with an RMS error of about 20 cm for the diffused sound and the microwave oven, but a positioning error of more than 1 meter occurred for the ringing tone. Considering the size of the sound source and the ability to discriminate between sound sources, it may be possible to use this technique if the acceptable positioning error is a few tens of centimeters. It remains as a next step to determine the cause of the lower location accuracy of the ringing tone.

ACKNOWLEDGMENT

This work was supported by JSPS KAKENHI Grant Numbers JP17K00140.

REFERENCES

[1] A. Yassin, et al., “Recent Advances in Indoor Localization: A Surveyon Theoretical Approaches and Applications,” IEEE Communications Surveys & Tutorials, vol. 19, no. 2, Second Quarter, pp. 1327-1346, 2017.

[2] Z. Hailong, H. Baoqi, and J. Bing, “Applying kriging interpolation for WiFi fingerprinting based indoor positioning

systems," 2016 IEEE Wireless Communications and Networking Conference, pp. 1822-1827, 2016.

[3] M. Ji, J. Kim, J. Jeon, and Y. Cho, "Analysis of positioning accuracy corresponding to the number of BLE beacons in indoor positioning system," 2015 17th International Conference on Advanced Communication Technology, pp. 92-95, 2015.

[4] M. Akiyama, H. Sunaga, S. Ioroi, and H. Tanaka, "Composition and Verification Experiment for Indoor Positioning System using Ultrasonic Sensors," Journal of the Institute of Positioning, Navigation and Timing of Japan, vol. 3, no. 1, pp. 1-8, 2012 (in Japanese).

[5] S. Murata, C. Yara, K. Kaneta, S. Ioroi, and H. Tanaka, "Accurate Indoor Positioning System using Near-Ultrasonic Sound from a Smartphone," Int. Conference on Next Generation Mobile Apps, Service and Technologies, pp. 13-18, 2014.

[6] J. Qi and G.P. Liu, "A Robust High-Accuracy Ultrasound Indoor Positioning System Based on a Wireless Sensor Network," Sensors, 17(11), 17 pages, 2017.

[7] Z. Chen and J. Wang, "ES-DPR: A DOA-Based Method for Passive Localization in Indoor Environments," Sensors 19(11), 17 pages, 2019.

[8] B. Suksiri and M. Fukumoto, "An Efficient Framework for Estimating the Direction of Multiple Sound Sources Using Higher-Order Generalized Singular Value Decomposition," Sensors, 19(13), 27 pages, 2019.

[9] T. H. Do and M. Yoo, "TDOA-based indoor positioning using visible light," Photonic Network Communications, vol. 27, Issue 2, pp. 80-88, 2014.

[10] 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.

[11] T. Nishiura, T. Yamada, S. Nakamura, and K. Shikano, "Localization of multiple sound sources based on a CSP analysis with a microphone array," 2000 IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 1053-1056, 2000.

 ISBN: 978-1-61208-739-9