www.s2is.org
International Journal on Smart Sensing and Intelligent Systems
ISSN 1178-5608

Sound-Separation System using Spherical Microphone Array with Three-Dimensional Directivity—KIKIWAKE 3D: Language Game for Children

Takahiro Nakadai[1], Tomohiro Nakayama[1] Tomoki Taguchi[1], Ryohei Egusa[2],
Miki Namatame[3], Masanori Sugimoto[4], Fusako Kusunoki[5], Etsuji Yamaguchi[2],
Shigenori Inagaki[2], Yoshiaki Takeda[2], and Hiroshi Mizoguchi[1]

[1]Department of Science and Technology, Mechanical Engineering,
Tokyo University of Science, 2641 Yamazaki, Noda-shi, Chiba, Japan

[2]Graduate School of Human Development and Environment,
Kobe University, Yayoigaoka 6, Sanda-shi, Hyogo, Japan

[3]Faculty of Industrial Technology,
Tsukuba University of Technology, 4-3-15, Amakubo, Tsukuba, Ibaraki, Japan

[4]Graduate School of Information Science and Technology,
Hokkaido University, Kita 15, Nishi 8, Kita-ku, Sapporo, Hokkaido, Japan

[5]Department of Information Design,
Tama Art University, 2-1723, Yarimizu, Hachioji, Tokyo, Japan

E-mail: 7514633@ed.tus.ac.jp, 7513632@ed.tus.ac.jp, 7512639@alumni.tus.ac.jp,
126d103d@stu.kobe-u.ac.jp, miki@a.tsukuba-tech.ac.jp, sugi@ist.hokudai.ac.jp, kusunoki@tamabi.ac.jp,
etuji@opal.kobe-u.ac.jp, inagakis@kobe-u.ac.jp, takedayo@kobe-u.ac.jp, hm@rs.noda.tus.ac.jp

Submitted: Oct. 15, 2014 ***Accepted: Nov. 12, 2014*** ***Published: Dec. 1, 2014***

Abstract–Mixed sounds can be separated from multiple sound sources using microphone array sensor and signal processing. We believe that promotion of interest in this technique can lead to significant future development in science and technology. To investigate this technique, we designed a language game for children called "KIKIWAKE 3D" that uses a sound-source-separation system to arouse children's interest in this technology. However, the microphone array sensor in a previous research had a limited scope in separating sounds. We developed a spherical microphone array sensor with three-dimensional directivity designed for this game. In this paper, we report the evaluation of this microphone array sensor in adapting to this game by separating the sound level and using questionnaires.

Index terms*:* Supporting learning system; signal processing; frequency-band selection, implementation, and evaluation

I. INTRODUCTION

A microphone array sensor is mainly consists of many microphones. The sound signal being input into these many microphones is processed. Consequently, an objective sound can be selectively captured from the background noise; this process is called "microphone array signal processing [1]."

Figure 1 shows the sensitivity distribution map of an omnidirectional microphone, which is plotted in a spherical form. Figure 2 shows the sensitivity distribution map of a microphone array sensor, which is plotted in a beam form. Therefore, the microphone array sensor can more selectively capture an objective sound.

In this paper, we report the development of a spherical microphone array sensor for a language game. Microphone array signal processing is implemented to separate the objective sound from mixed sounds. This technology is currently being actively researched for different applications such as in sound-source identification apparatus and environment recognition function of service robots [2, 3]. The novel technology discussed in this paper was developed to spur interest in future science and technology. Thus, we hope that children will be interested in this technology.

In previous research, a participatory language game was introduced to promote the interest of children in this technology [4]. However, only three players were able to participate in that game because the microphone array used had only a two-dimensional directivity control. We have since improved our design to a microphone array with three-dimensional directivity that allows as many as six players to participate.

We designed a game called "KIKIWAKE 3D," which is a language game, to entice children to be interested in this technology. This new design helps promote greater interest by increasing the number of players in the game. We developed a spherical microphone array sensor. In this paper, we report the evaluation of this sensor to adapt to this game by separating the sound level and using questionnaires.

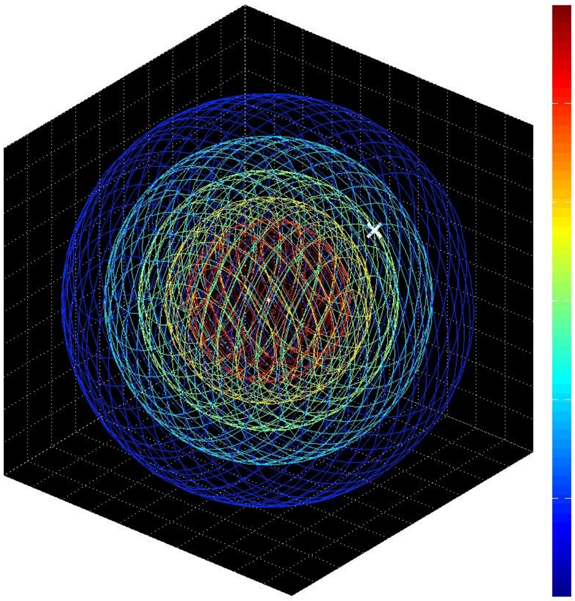


Figure 1. Sensitivity distribution map of an omnidirectional microphone

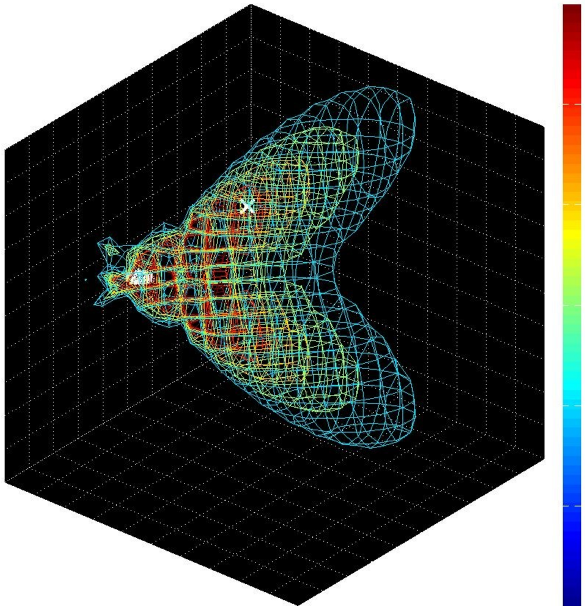


Figure 2. Sensitivity distribution map of a microphone array sensor

II. MICROPHONE ARRAY WITH THREE-DIMENSIONAL DIRECTIVITY

In this section, we describe the microphone array sensor utilized for the sound-separation system in this research.

a. Design of microphone array sensor

Several microphone array signal processing methods are available. In our research, we selected the delay-and-sum beamforming (DSBF) [5, 6, 7] method because it is robust in real environment.

The DSBF method can selectively and locally capture a sound to form a high-sensitivity beam in an objective direction. The characteristic of this method is that the performance of the microphone array sensor depends on the placement of the microphone and the distance between each microphone.

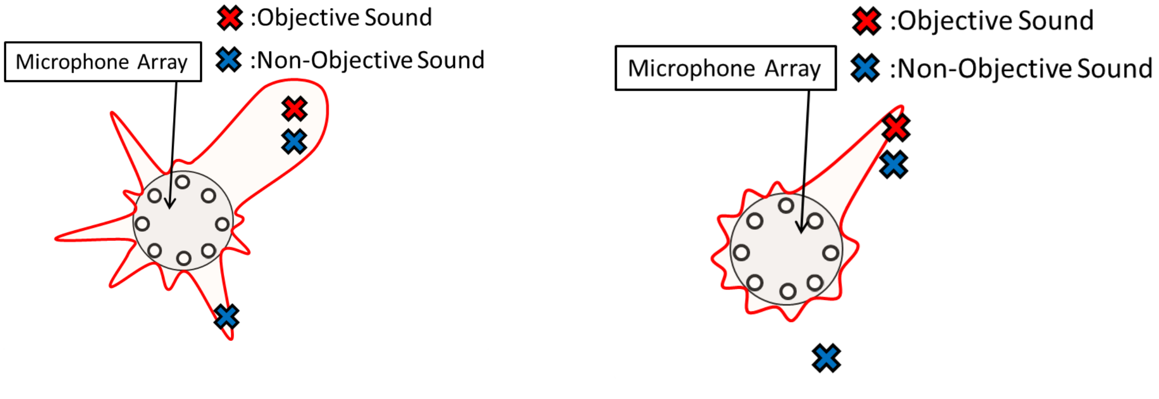
The following two points can indicate the performance of the microphone array sensor:

(1) Sharpness of the main lobe

(2) Minimization of the sidelobe gain

The main lobe shows the sensitivity distribution map of a beam shape formed in an objective direction. The sidelobe shows the sensitivity distribution map of a beam shape formed in a non-objective direction. Figure 3 shows these outlines. If the main lobe width is large and the sidelobe gain is high, a noise source can easily cover the high-sensitivity area. To effectively capture an objective sound, we recommend forming a beam that is as narrow as the main lobe width and minimizing the sidelobe gain [8].

We assumed the main lobe width and the sidelobe gain size as an index in a beamforming simulation. Analytically calculating the optimal microphone arrangement that can narrow the main lobe and minimize the sidelobe is difficult. Therefore, we devised a number of candidate microphone arrangements and evaluated the performance of each candidate to obtain the optimal arrangement [9]. Figure 4 shows the simulation results.



(a) Bad (b) Good

Figure 3. Main lobe and side lobe

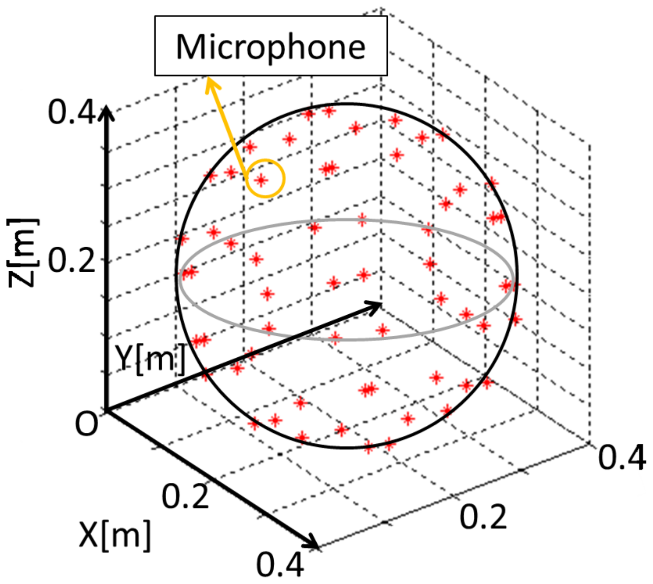


Figure 4. Simulation results to obtain the optimal microphone arrangement

b. Signal processing

A microphone array reduces the non-objective sound and accentuates the objective sound. However, a listener cannot clearly recognize any sound because the processing relatively accentuates the sound. To eliminate the non-objective sound, we implemented frequency-band selection (FBS) [10] after the microphone array signal processing. The frequency components of a human voice have individual differences; thus, this method was utilized. This signal processing method can clearly separate the voice of each speaker.

Figure 5 shows the outline of this method. We discuss the sound separation of two speakers. First, the mixed sound of two speakers is obtained by a microphone array sensor. This sound is called "mixed sound" in this study. Next, the DSBF method is implemented to this mixed sound. The voices of the two speakers are accentuated by the DSBF method. The resulting sound is called "accentuated sound" in this research. Finally, to select the desired accentuated frequency components, the frequency bands of the two voices are compared. The selected frequency components are retained, and all other components are removed. This signal processing method can more clearly separate the sound of each speaker. The resulting sound is called "separated sound" in this research.

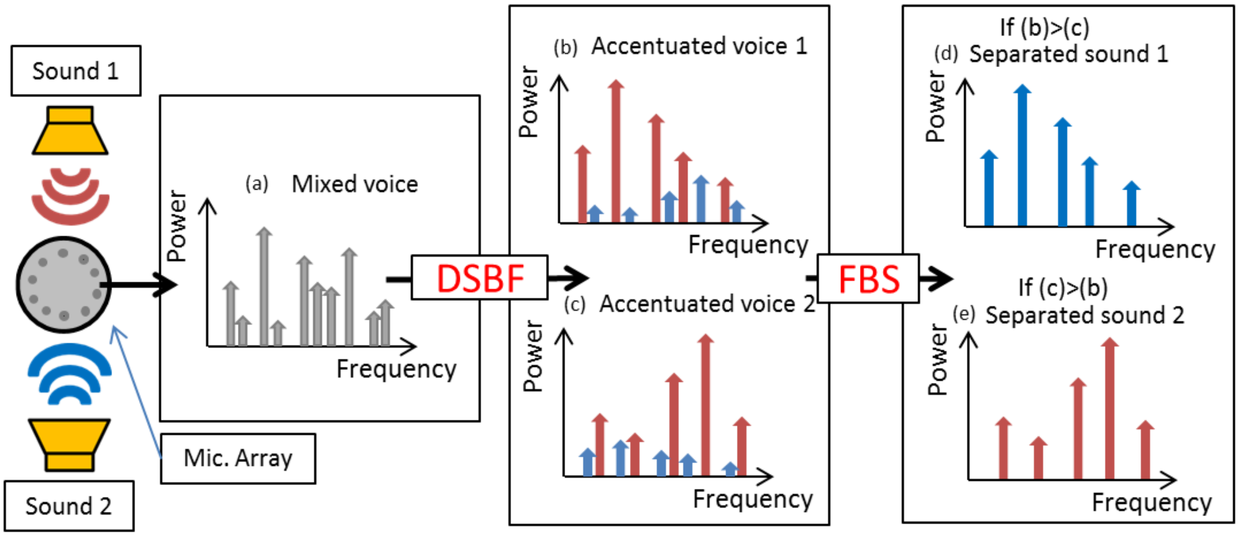


Figure 5. Outline of the FBS method

The signal of the separated sound by the FBS method is expressed by (1) and (2). Short-time Fourier transform is applied on the two accentuated sounds by the DSBF method. The two accentuated sounds are compared in terms of each frequency component after the transformation. These sounds return the signal in the time domain by applying inverse Fourier transform. Consequently, we can obtain the separated sounds.

$$X_f(\omega_i) = \begin{cases} X_d(\omega_i) & \text{if } X_d(\omega_i) \geq X_e(\omega_i) \\ 0 & \text{else} \end{cases} \qquad (1)$$

$$X_g(\omega_i) = \begin{cases} X_e(\omega_i) & \text{if } X_e(\omega_i) \geq X_d(\omega_i) \\ 0 & \text{else} \end{cases}. \qquad (2)$$

III. SOUND-SEPARATION SYSTEM

In this section, we discuss the implementation of the sound-separation system.

a. Sound-separation system using spherical microphone array sensor

Figure 4 shows our developed spherical microphone array sensor design. Figure 6 shows the implemented spherical microphone array sensor. The diameter of the spherical microphone array sensor is 0.4 m. The number of microphone elements is 64. Figure 7 shows the outline of the sound-separation system with a spherical microphone array sensor.

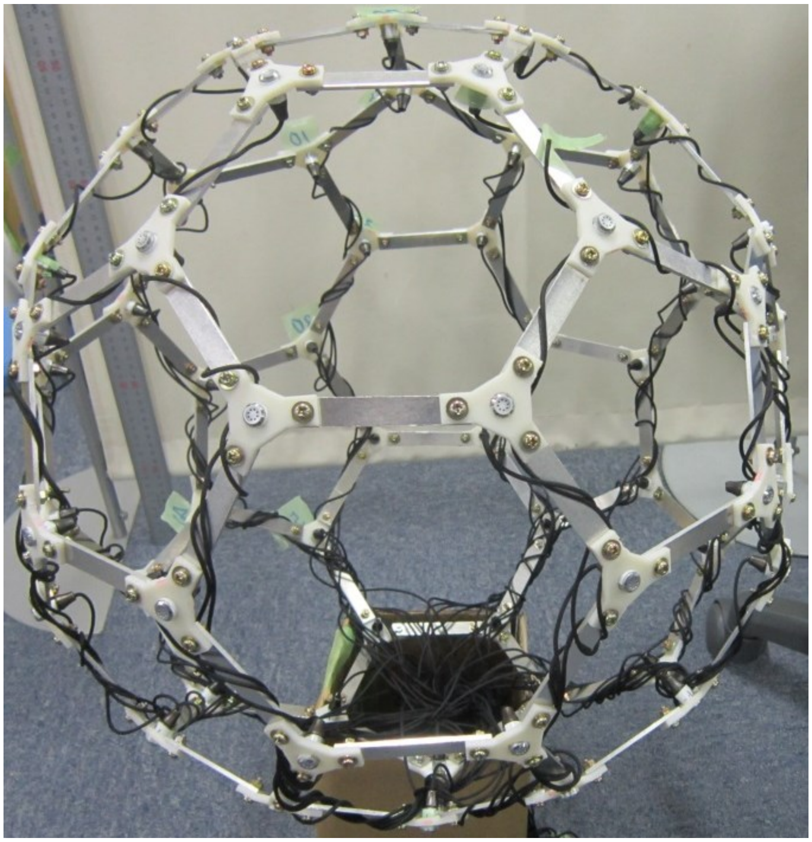


Figure 6. Spherical microphone array sensor

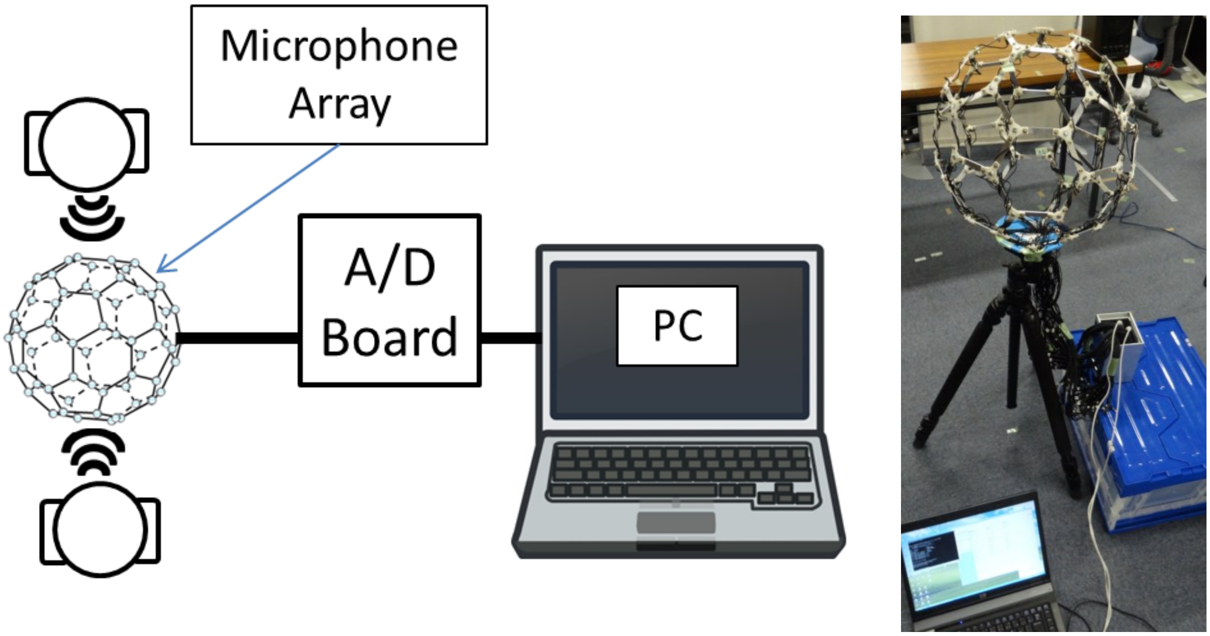


Figure 7. Sound-separation system with a spherical microphone array sensor

b. Flow of the separated-sound generation

Separated sounds can be obtained using the sound-separation system and signal processing. We discuss the flow of the separated-sound generation.

Figure 8 shows the outline of the flow of the separated-sound generation. A total of 64 wave files are generated by 64 microphones using the spherical microphone array sensor. Two wave files called “accentuated sound” are generated by implementing the DSBF method on the 64 wave files. Two wave files called “separated sound” are generated by implementing the FBS method on the two wave files called “accentuated sound.”

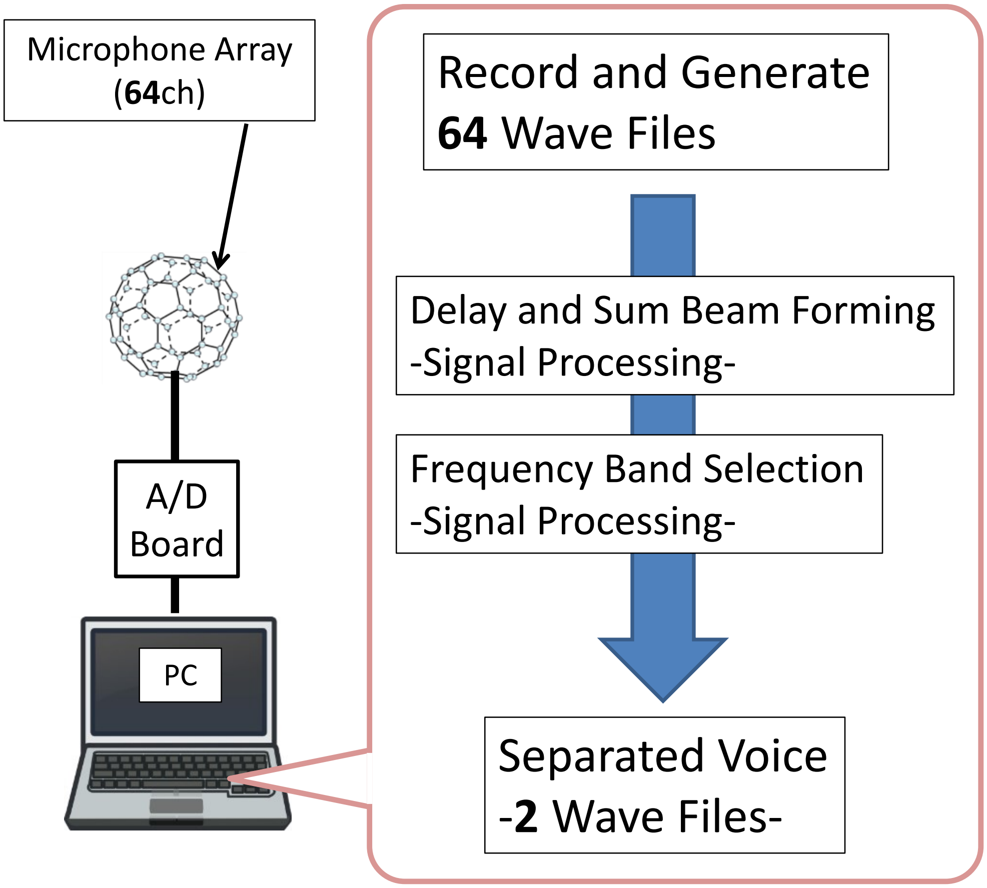


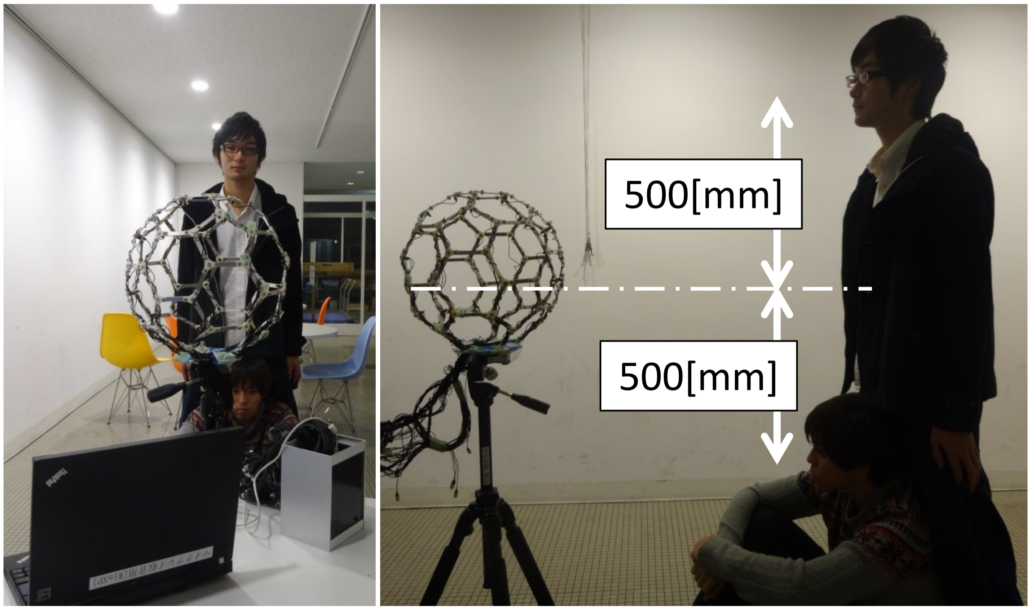
Figure 8. Generation flow of the separated sounds

IV. EXPERIMENT

a. Experimental setup

In this section, we discuss the experimental setup. We conducted an experiment using the sound-separation system. Figure 9 shows the experimental setup. The mixed sound was confusing. A separated sound was generated by combining the microphone array sensor with signal processing. The generated separated sounds were evaluated in terms of their applicability in the "KIKIWAKE 3D," as presented in this and in the next sections.

Two speaking persons were positioned up and down. One was standing, and the other was sitting. In this research, the standing and sitting speakers were designated as "A" and "B," respectively. The origin of the coordinates was the center of the spherical microphone array sensor installed, as shown in Figure 9. The center of the spherical microphone array sensor was installed 1200 mm from the ground. The head position of A was located at (1000,0,500) mm. The head position of B was located at (1000,0,-500) mm.



(a) Front view (b) Side view

Figure 9. Experimental setup

b. Experimental procedure

The speakers simultaneously spoke different words, and their voices were recorded by the spherical microphone array. Table 1 lists the Japanese words spoken by the two speakers. "*Hashigo*," "*Ichigo*," "*Tora*," and "*Uma*" mean ladder, strawberry, tiger, and horse in Japanese, respectively.

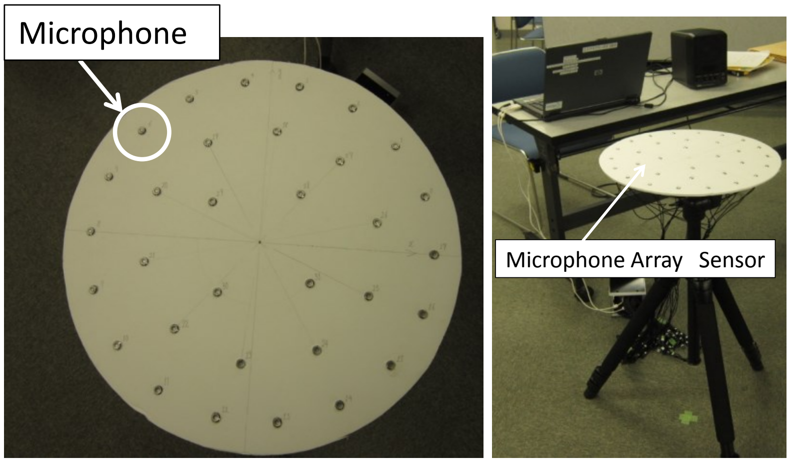


Figure 10. Existing microphone array sensor

c. Existing microphone array sensor used for comparison

We conducted a comparative evaluation of the spherical microphone array sensor developed in this research. The spherical microphone array was compared with the existing microphone array sensor shown in Figure 10. The diameter of this existing microphone array sensor was 0.4 m. The number of microphone elements was 32.

Table 1: Traditional methods of fixing fault current (IEEE Spectrum July 1997 issue)

	A	B
Question 1	*Hashigo* (means ladder)	*Ichigo* (means strawberry)
Question 2	*Tora* (means tiger)	*Uma* (means horse)
Question 3	Cat	Rat

V. EVALUATION

We discuss the evaluation of the mixed and separated sounds obtained by the experiment from two perspectives. One was from the engineering perspective. The separation level of the two voices of the speakers was evaluated using the correlation coefficient. The other was from the psychological perspective. To evaluate the clarity of the separated sounds, we conducted investigation using questionnaires. The effectiveness of the sound-separation system developed in this research for "KIKIWAKE 3D" was confirmed using the two evaluation perspectives.

a. Correlation coefficient

The separated sounds were evaluated from the engineering perspective. We calculated the correlation coefficient of the two separated sounds obtained by the sound-separation system. Correlation coefficient is a measure of the correlation between two variables; it gives an inclusive value between +1 and zero. A correlation coefficient close to zero signifies that the correlation between the two variables is poor, and vice versa. The lower the correlation coefficient value, the better is the separation.

In this research, two correlation coefficient values were compared. One was the correlation coefficient of the two separated sounds obtained by the spherical microphone array sensor, which was called "separated sound by spherical microphone array sensor (SSS)." The other was the correlation coefficient of the two separated sounds obtained by the existing microphone array sensor, which was called "separated sound by existing microphone array sensor (SSE)." Figure 11 shows the result of this evaluation.

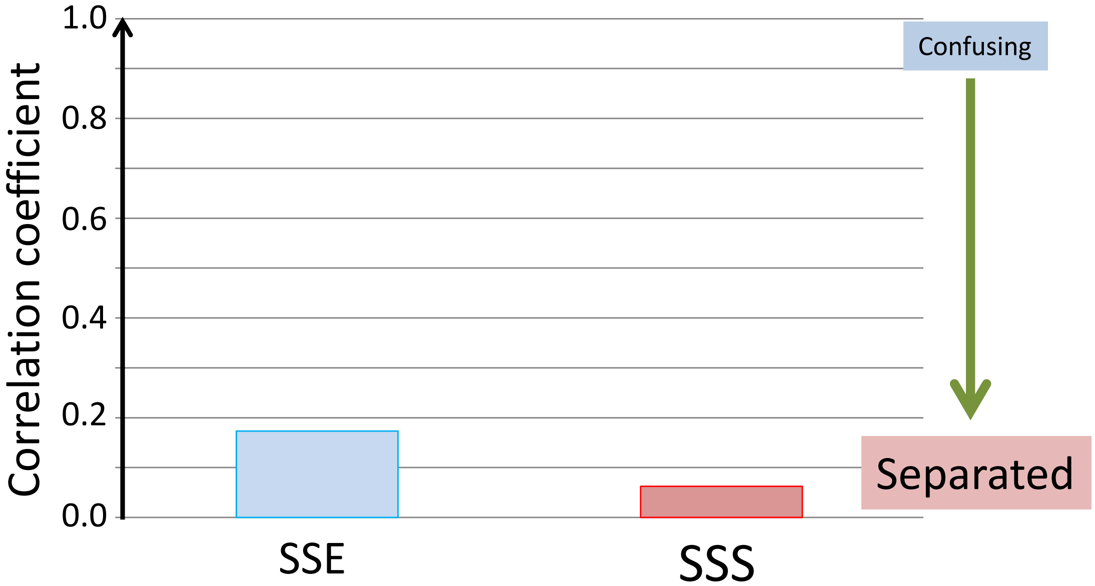


Figure 11. Correlation coefficients of the separated sounds obtained by each microphone array sensor

The SSS correlation coefficient is smaller than the SSE correlation coefficient. In other words, we confirmed that the separated sound obtained by the spherical microphone array sensor was better. However, we were not able to confirm a clear difference in the correlation coefficient values. These separated sounds were evaluated from the other perspective.

b. Evaluation using questionnaires

The SSS was evaluated from the psychological perspective. In this research, we conducted our investigation using questionnaires.

Participants: The participants were composed of 15 students (aging from 22 to 24 years old) from a private university.

Investigation: The subjects listened to six SSSs listed in Table 1. Each filled up a questionnaire on what they recognized from the SSSs. To compare the spherical microphone array sensor with the existing microphone array sensor, the subjects listened to six SSEs listed in Table 1. Similarly, each filled up the questionnaire. Therefore, we investigated 12 separated sounds.

The clarity of the SSSs and SSEs was evaluated by the number of correct answers in the questionnaires. The separated sounds obtained by each microphone array sensor were six. Thus, a full mark has a total score of six. Figure 12 shows the questionnaire results.

We focused on the average of correct answers for the SSEs. It was less than half of the full mark. We confirmed that the subjects were not comfortable with the SSEs. In contrast, when the average of the correct answers for the SSSs was focused, we found more than five correct answers. We confirmed that the subjects were comfortable with the SSSs.

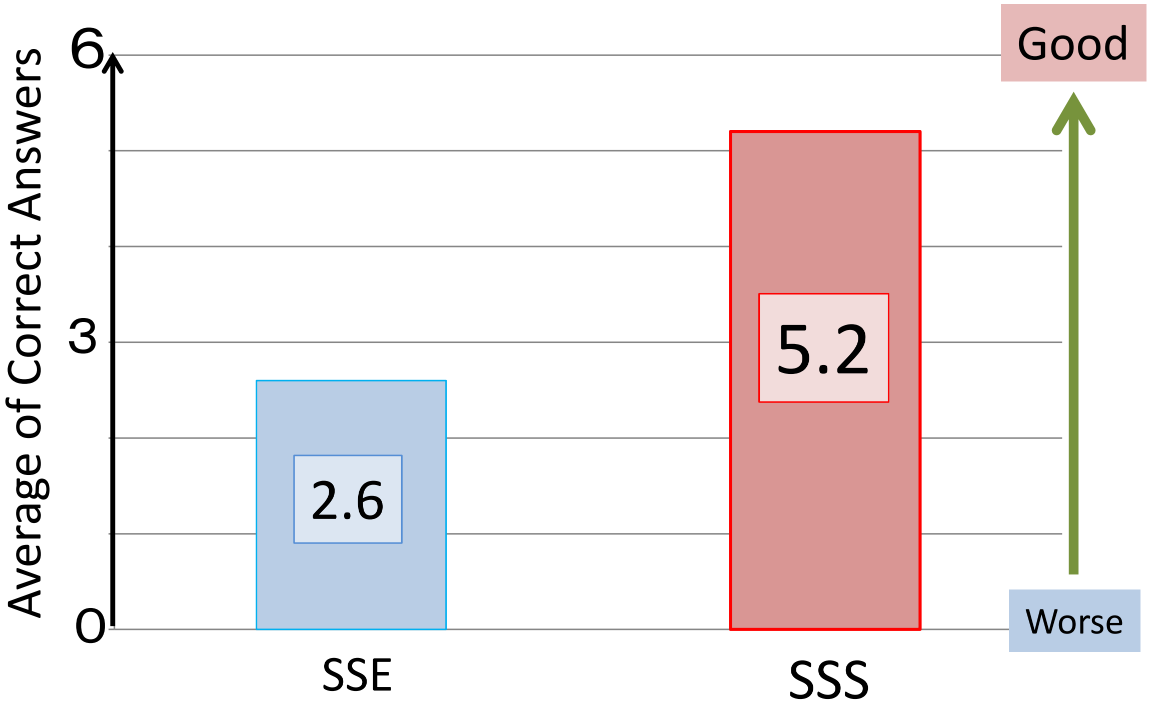


Figure 12. Number of correct answers from the questionnaires

VI. CONCLUSIONS

This paper has reported the development of a spherical microphone array sensor for use in a language game. In our study, to capture the interest of elementary school children on sound-source-separation technology, which is considered very important for future development in science and technology, we designed a language game called "KIKIWAKE 3D" in which children can experience the sound-source-separation technology using their own voices.

We developed a microphone array with three-dimensional directivity to make the game more interesting and to increase the number of participants. In this game, it is essential that the same directivity be formed in every direction in space. Therefore, a spherical microphone array was designed and developed for the game using simulation. The arrangement of the microphone array was designed for optimal performance so that it can effectively capture different sounds. The evaluation results by separating the sound level and using questionnaires indicated that the spherical microphone array sensor is applicable and effective for the game.

In the future, we plan to let children play KIKIWAKE 3D and evaluate their feelings on the technology.

ACKNOWLEDGEMENT

This work was supported in part by the Grants-in-Aid for Scientific Research (B) (24300290 and 23300303).

REFERENCES

[1] B. D. Van Veen and K. M. Buckley, "Beamforming: a versatile approach to spatial filtering," IEEE Acoust., Speech, Signal Processing Mag. , vol. 5, pp., April 4–24, 1988.

[2] T. Yoshida, K. Nakadai, H. Okuno, "Two-Layered Audio-Visual Speech Recognition for Robots in Noisy Environment", The 2010 IEEE\RSJ International Conference on Intelligent Robots and Systems (IROS-2010), pp.988-993, October 18-22, 2010.

[3] M. Goseki, M. Ding, H. Takemura, and H. Mizoguchi, "Combination of microphone array and camera image processing for visualizing sound pressure distribution." SMC2011, pp.139-143, 2011.

[4] T. Taguchi, M. Goseki, R. Egusa, M. Namatame, F. Kusunoki, M. Sugimoto, E. Yamaguchi, S. Inagaki, Y. Takeda, and H. Mizoguchi, “KIKIWAKE: Sound source separation system for children-computer interaction.” CHI 2013 Extended Abstracts, pp. 757-762, April 27–May 2, 2013.

[5] N.-V. Vu, H. Ye, J. Whittington, J. Devlin, and M. Mason, “Small footprint implementation of dual-microphone delay-and-sum beamforming for in-car speech enhancement.” International Conference on Acoustics, Speech, and Signal Processing. USA, pp. 1482-1485, March 2010.

[6] M. Fuchs, T. Haulick, and G. Schmidt, “Noise suppression for automotive applications based on directional information.” International Conference on Acoustics, Speech, and Signal Processing. Canada, vol. 1, pp. I–237-40, May 2004.

[7] E. Weinstein, k. Steele, A. Agarwal, and J.Glass.Loud: A 1020-node modular microphone array and beamformer for intelligent computing spaces. Technical Report MIT-LCS-TM-642, MIT/LCS Technical Memo, April 2004.

[8] H.Sun, S. Yan, P. Svensson, “Robust Minimum Sidelobe Beamforming for Spherical Microphone Array”, IEEE TRANSCATIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, vol. 19, No. 4, pp.1045-1051, May 2011.

[9] T. Fujihara, S. Kagami, Y. Sasaki, and H. Mizoguchi, “Arrangement optimization for narrow directivity and high S/N ratio beam forming microphone array.” IEEE SENSORS, Italia, pp. 450-453, October 2008.

[10] T. Taguchi, T. Nakadai, R. Egusa, M. Namatame, F. Kusunoki, M. Sugimoto, E. Yamaguchi, S. Inagaki, Y. Takeda, and H. Mizoguchi, “Investigation on optimal microphone arrangement of spherical microphone array to achieve shape beamforming.” Intelligent Systems, Modelling, and Simulation, pp. 330-333, January 27–29, 2014.