Effects of microphone arrangement on the accuracy of a spherical microphone array (SENZI) in acquiring high-definition 3D sound space information

Shuichi Sakamoto[1], Jun'ichi Kodama[1], Satoshi Hongo[2],
Takuma Okamoto[3], Yukio Iwaya[1] and Yôiti Suzuki[1]

[1]*Research Institute of Electrical Communication and Graduate School of Information Sciences, Tohoku University,*
2–1–1, Katahira, Aoba-ku, Sendai, 980–8577, Japan
E-mail: {saka, kodama, iwaya, yoh}@ais.riec.tohoku.ac.jp
www.ais.riec.tohoku.ac.jp/index.html

[2]*Dept. of Design and Computer Applications, Sendai National College of Technology,*
48 Nodayama, Natori, 981–1239, Japan
E-mail: hongo@sendai-nct.ac.jp

[3]*Research Institute of Electrical Communication and Graduate School of Engineering, Tohoku University,*
2–1–1, Katahira, Aoba-ku, Sendai, 980–8577, Japan
E-mail: okamoto@ais.riec.tohoku.ac.jp

We propose a three-dimensional sound space sensing system using a microphone array on a solid, human-head-sized sphere with numerous microphones, which is called SENZI (Symmetrical object with ENchased ZIllion microphones). It can acquire 3D sound space information accurately for recording and/or transmission to a distant place. Moreover, once recorded, the accurate information might be reproduced accurately for any listener at any time. This study investigated the effects of microphone arrangement and the number of controlled directions on the accuracy of the sound space information acquired by SENZI. Results of a computer simulation indicated that the microphones should be arranged at an interval that is equal to or narrower than 5.7° to avoid the effect of spatial aliasing and that the number of controlled directions should be set densely at intervals of less than 5° when the microphone array radius is 85 mm.

Keywords: Microphone array, Dummy head recording, Head-related transfer function (HRTF), Tele-existence

1. Introduction

Sensing technologies of three-dimensional (3D) sound space information are indispensable partners of 3D sound reproduction technologies. Comprehensive and accurate sensing of 3D spatial audio information is therefore a key to realization of high-definition 3D spatial audio systems. Although several research efforts have addressed sensing topics, few technologies reflect listeners' head movements in the sensing. Many authors [1–3] have described that listeners' head movements are effective to enhance the localization accuracy as well as the perceived realism in human spatial hearing perception.

In this context, a few methods have been proposed to realize sensing of three-dimensional sound space information considering the listener's movement [4–6]. All of these methods apply special objects to sense sound space information. These objects are set at the recording place. Then recorded information is transmitted to a distant place. However, these methods are insufficient to sense accurate 3D audio space information and to provide appropriate sound information to plural listeners individually and simultaneously.

As another approach to sense and/or to reproduce accurate sound information, ambisonics, especially higher-order ambisonics, have been specifically examined [7,8]. In this technology, 3D sound space information is encoded and decoded on several components with specific directivities based on spherical harmonic decomposition. However, even with higher-order ambisonics of the highest order available such as five, the directional resolution might be insufficient compared with human resolution of spatial hearing. A sensing system matching human performances is highly desired but it remains unclear how many orders are necessary to yield directional resolution that is sufficient to satisfy perceptual resolution.

Consequently, we have proposed a system that can sense accurate 3D sound space information and/or transmit it to a distant place using a microphone array on a human-head-sized solid sphere with numerous microphones on its surface. We designate this spherical microphone array as SENZI (Symmetrical object with ENchased ZIllion microphones) [9]. The system can sense 3D sound space information comprehensively: information from all directions is available, over locations and over time if once recorded, for any listener orientation and head/ear shape with correct binaural cues. However, the accuracy of the acquired 3D sound space information necessarily depends on the arrangement of microphones that are set on the solid sphere [10].

In this study, we investigated the effect of the microphone arrangement

on the accuracy of the acquired sound space information of SENZI.

2. System Outline[9]

2.1. *System Concept*

Figure 1 presents a scheme of the proposed system. The microphone array named SENZI is made from an acoustically hard sphere with plenty of microphones. The SENZI is set at the recording place and sound signals inputted to all microphones on SENZI are used for synthesizing a listener's head related transfer function (HRTF). Calculated signals are typically presented to a listener binaurally, for example, via headphones. It is noteworthy, however, that SENZI can output suitable signals to any spatial sound reproduction system. Individual HRTFs are synthesized using digital signal processing. Moreover, listener's head movement can be reflected to the output signal processing for any listener with various head and ear shape and for any time for any place if the input signals are once recorded. Therefore, it is even possible to present individualized 3D sound space information to many listeners simultaneously using this system.

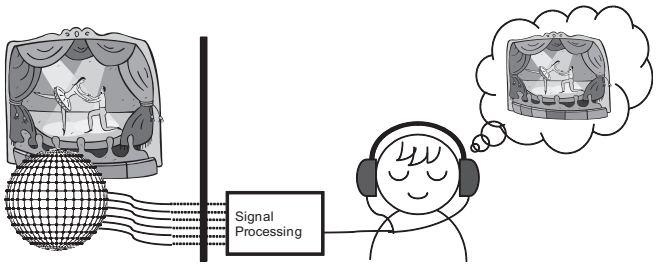

Fig. 1. Concept of the proposed system.

2.2. *Calculation method of HRTFs for individual listeners*

In the proposed system, to calculate and synthesize a listener's HRTFs using input signals from spatially distributed multiple microphones, each input signal from each microphone is simply weighted and summed to synthesize a listener's HRTF. Moreover, the weight is changed according to the 3D head movement of a human who is in a different place. Therefore, 3D sound space information is acquired accurately corresponding to any head movement. Moreover, it should be noted that this is possible for any listener for any time if input signals are once recorded.

As the simplest set of circumstances, we first assume the case in which sounds come only from the horizontal plane. Let $\boldsymbol{H}_{\text{listener}}$ signify the listener's HRTF for one ear. For a certain frequency f, $\boldsymbol{H}_{\text{listener},f}(\theta)$ is expressed according to the following equation:

$$\boldsymbol{H}_{\text{listener},f}(\theta) = \sum_{i=1}^{n} z_{i,f} \cdot \boldsymbol{H}_{i,f}(\theta) + \varepsilon. \tag{1}$$

In that equation, $\boldsymbol{H}_{i,f}(\theta)$ is the transfer function of the i-th microphone from a sound source as a function of the direction of the sound source (θ). Actually, $\boldsymbol{H}_{i,f}(\theta)$ and $z_{i,f}$ are all complex. Equation 1 reflects that the listener's HRTF is calculable from these transfer functions. Equation 1 cannot be solved; a residual ϵ remains if the number of sound source directions differs from the number of microphones n. In fact, ϵ varies according to the weighting coefficient $z_{i,f}$. Therefore, a set of optimum $z_{i,f}$ is calculated using the pseudo-inverse matrix. The coefficients $z_{i,f}$ are calculated for each microphone at each frequency in this method. Calculated $z_{i,f}$ is constant irrespective of the direction of a sound source. This feature of our method is extremely important because one important advantage of the system is that the sound source position need not be considered when sound-space information is acquired.

When $z_{i,f}$ is calculated, we must select directions (θ), that are incorporated into the calculation. The selected directions are designated as "*controlled directions*" hereinafter. However, in a real environment, sound waves come from all directions, including directions that are not incorporated into calculations. These directions are called "*uncontrolled directions*." To synthesize accurate sound information for all directions including "*uncontrolled directions*," the number of microphones, the arrangement of the microphones on the object, and the shape of the object should be optimized.

3. Accuracy of acquired sound space information for uncontrolled directions

3.1. *Experimental method*

We analyzed the accuracy of synthesized HRTFs of all the directions including at *uncontrolled directions*, when the transfer functions of the SENZI were adjusted at *controlled directions*.

For this study, the HRTFs of a dummy head (SAMRAI; Koken Co. Ltd.) were used as the target characteristics to be realized using this system. They were calculated using the boundary element method (BEM). Calculated

HRTFs are depicted in Fig. 2. The frequency range for synthesis was set as 0–20 kHz in 93.75 Hz steps.

Table 1 presents verified conditions that were considered. In SENZI, microphones were set at steps of 20° (conditions **a**, **b** and **c**), 10° (conditions **d**, **e** and **f**), or 5° (condition **g**) from 0° (in front of the listener) to 359° in the horizontal plane and at steps of 20° from −60° to 60° in the vertical plane. The *controlled directions* were set at steps of 20° (condition **a**), 10° (conditions **b**, **d**), 5° (conditions **c**, **e**, **g**), or 2° (condition **f**) from 0° (in front of the listener) to 359° in the horizontal plane and at steps of 20° from −40° to 80° in the vertical plane. After calculating the weighting coefficient $z_{i,f}$ in each condition, the transfer functions of 2,520 (360×7) directions (including both *controlled* and *uncontrolled directions* were synthesized using $z_{i,f}$ in all conditions.

The error of the synthesized HRTFs in terms of the spectral distortion (SD) was calculated as follows [9]:

$$\varepsilon_f(\theta) = \left| 20\log_{10} \left| \frac{\boldsymbol{H}_{\text{listener},f}(\theta)}{\boldsymbol{H}_{\text{synthesized},f}(\theta)} \right| \right| \quad [\text{dB}]. \tag{2}$$

3.2. *Results and discussion*

Figures 3 to 6 show examples of the SD between the HRTFs of the dummy-head and the synthesized HRTFs. From these figures, a certain boundary

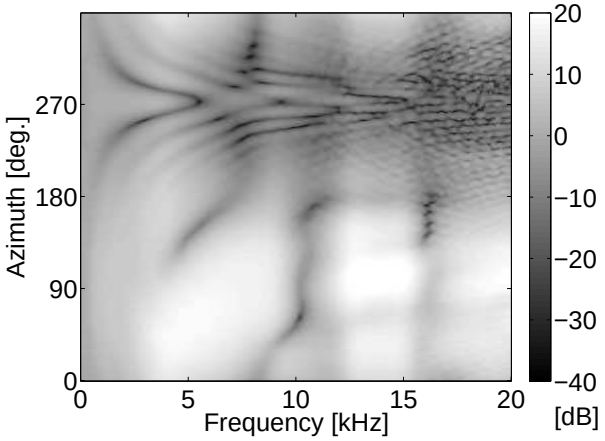

Fig. 2. HRTF of SAMRAI (0 deg elev. angle).

Table 1. Conditions considered

Condition	Number of microphones	Number of controlled directions
a	126 (18×7)	126 (18×7)
b	126 (18×7)	252 (36×7)
c	126 (18×7)	504 (72×7)
d	252 (36×7)	252 (36×7)
e	252 (36×7)	504 (72×7)
f	252 (36×7)	1260 (180×7)
g	504 (72×7)	504 (72×7)

frequency is visible between the frequency range where sound space information is synthesized accurately and that where a large synthesis error is observed. That is expected to be attributable to the effect of spatial aliasing from the intervals between microphones [11]. When the microphones are set at $\theta[rad]$ steps on the SENZI, the interval between each microphone d is calculated using the following equation:

$$d = r\theta, \tag{3}$$

where r is the radius of the SENZI.

Because the radius of the SENZI in this study was 85 mm, d was 29.5 mm in conditions **a**, **b** and **c**, and 14.8 mm in conditions **d**, **e** and **f**. Therefore, the frequency at which the half-wavelength is equal to d is 5.8 kHz in conditions **a**, **b** and **c**, and 11.6 kHz in conditions **d**, **e** and **f**. These values correspond to the boundary frequency of the error in Figs. from 3 to 6. Therefore, microphones should be set at intervals of 5.7° to avoid the effect of spatial aliasing when the radius of the SENZI is 85 mm.

Figure 7 portrays the average SD for "*controlled directions*" and "*controlled directions* and *uncontrolled directions*." In this figure, the accuracy of synthesized HRTFs calculated with *controlled* and *uncontrolled directions* is much lower than that calculated only with *controlled directions* in conditions **a**, **d**, and **g**. In these conditions, the number of microphones was equal to the number of *controlled directions*. Therefore, the calculated coefficients $z_{i,f}$ were fitted strictly at the *controlled directions*, but they were not fitted at *uncontrolled directions*.

For quantitative consideration, we simplified the situation to one in which microphones and *controlled directions* were set to 0° elevation angle, i.e. the horizontal plane. In this situation, we investigated the relation between the intervals of *controlled directions* and the accuracy of synthesized

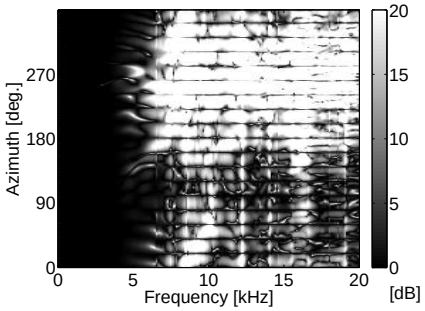

Fig. 3. Spectral distortion of condition **a** (0 deg elev. angle).

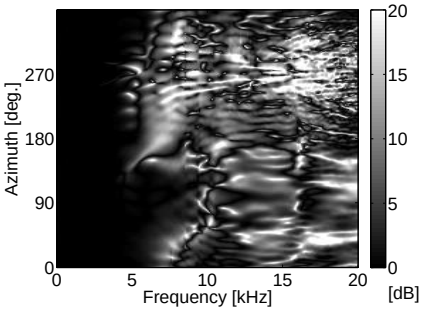

Fig. 4. Spectral distortion of condition **c** (0 deg elev. angle).

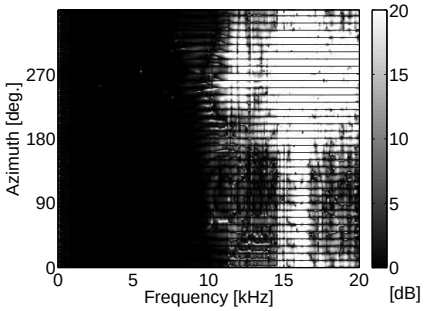

Fig. 5. Spectral distortion of condition **d** (0 deg elev. angle).

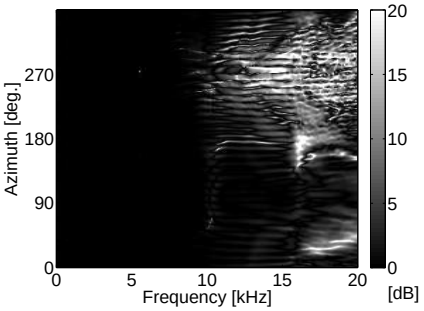

Fig. 6. Spectral distortion of condition **e** (0 deg elev. angle).

HRTFs. The microphones were set at steps of 90°, 45°, 30°, 15°, or 10° from 0° (in front of the listener) to 359° in the horizontal plane of the solid sphere. The intervals of the *controlled directions* were changed from 1° to 90°, with the intervals of *controlled directions* always narrower than that of microphones. The sampling frequency was 48 kHz. The range for synthesis was set as 0–20 kHz in steps of 93.75 Hz. Then, SD was calculated for all directions including *controlled* and *uncontrolled directins*.

Figure 8 portrays the relation between the average SD and the intervals of *controlled directions* for each microphone arrangement. Results show that the average SD decreases when the intervals of *controlled directions* becomes narrow. Moreover, the average of SD is almost constant when the *controlled directions* are set densely at intervals of less than 5°. Results show that the *controlled directions* should be set at 5° intervals.

Figures 9 to 12 show the SD of each condition with 36 microphones.

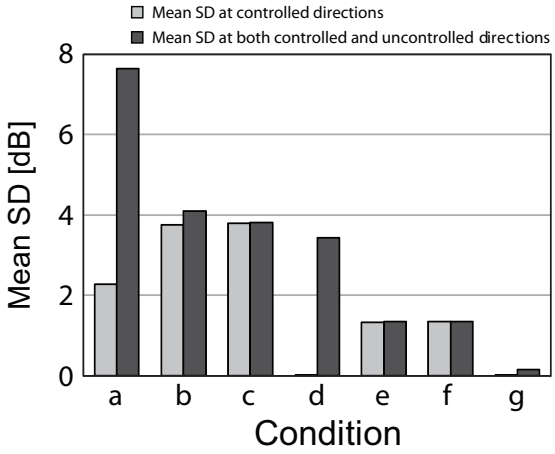

Fig. 7. Average of spectral distortion in all conditions.

In this case, the intervals of microphones are 10° (14.8 mm); thus spatial aliasing frequency is around 11.5 kHz. These figures show that the error in the head shadow region around 270° decreases as the intervals of *controlled directions* become narrow and that the error becomes constant when the *controlled directions* are set densely at intervals of less than 5°.

In summary, if the microphones and the "controlled directions" are arranged appropriately, then SENZI can accurately synthesize any HRTF up to around the spatial aliasing frequency for the shadow region and up to even several kilohertz higher frequency for the sunny-side region.

4. Summary

In this study, we investigated the effects of microphone arrangement and the number of *controlled directions* on the accuracy of the comprehensive 3D sound space information acquisition system called SENZI. The simulation results indicate that the microphones should be arranged at intervals of 5.7° or narrower to avoid the effect of spatial aliasing. Furthermore, the number of *controlled directions* should be set densely at intervals of less than 5° when the radius of the microphone array is 85 mm.

Acknowledgements

This work was supported by Strategic Information and Communications R&D Promotion Programme (SCOPE) No. 082102005 from the Ministry of

Internal Affairs and Communications (MIC) Japan and a grant for Tohoku University Global COE Program CERIES from MEXT Japan. We thank Dr. Makoto Otani (Shinshu University) for assistance in calculating the HRTFs.

References

1. H. Wallach, On sound localization, *J. Acoust. Soc. Am.* **10**, 270 (1939)
2. W. R. Thurlow and P. S. Runge, Effect of induced head movement in localization of direction of sound, *J. Acoust. Soc. Am.* **42**, 480 (1967)
3. Y. Iwaya, Y. Suzuki and Kimura D, Effects of head movement on front-back error in sound localization, *Acoust. Sci. & Tech.* **24**, 322 (2003)
4. I. Toshima, H. Uematsu and T. Hirahara, A steerable dummy head that tracks three-dimensional head movement, *Acoust. Sci. & Tech.* **24**, 327 (2003)
5. V. R. Algazi, R. O. Duda and D. M. Thompson, Motion-Tracked Binaural Sound, *J. Audio Eng. Soc.* **52**, 1142 (2004)
6. J. B. Melick, V. R. Algazi, R. O. Duda and D. M. Thompson, Customization for personalized rendering of motion-tracked binaural sound, *Proc. 117th AES Convention* **6225**, 1 (2004)
7. D.H. Cooper and T. Shige, "Discrete-Matrix Multichannel Stereo," *Journal of Audio Engineering Society*, 20(5), pp. 346–360, 1972
8. R. Nicol and M. Emerit, "3D-sound reproduction over an extensive listening area: A hybrid method derived from holophony and ambisonic," Proc. AES 16th International Conference, 16(39), pp. 436–453, 1999
9. S. Sakamoto, S. Hongo, R. Kadoi and Y. Suzuki, SENZI and ASURA: New high-precision sound-space sensing systems based on symmetrically arranged numerous microphones, *Proc. Second International Symposium on Universal Communication (ISUC2008)*, 429 (2008)
10. J. Kodama, S. Sakamoto, M. Otani, S. Hongo, Y. Iwaya and Y. Suzuki, Numerical investigation of reproduction accuracy for ASURA (A Symmetrical and Universal Recording Array): Geometry and microphone position effects, *Proc. ASJ Spring meeting* (in Japanese), 1-9-6, 1457-1456 (2008)
11. S. U. Pillai, *Array signal processing* (Springer-Verlag, New York, 1989)

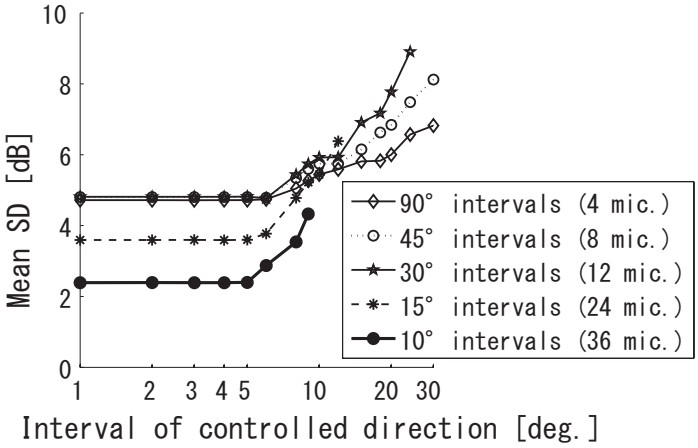

Fig. 8. Relation between the average of spectral distortion and the number of *controlled directions* for each microphone arrangement.

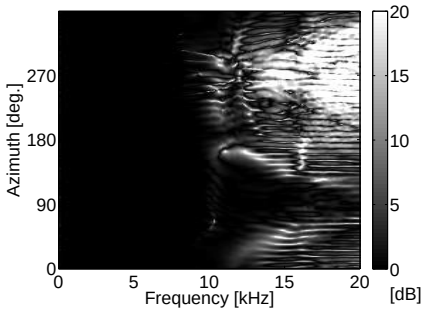

Fig. 9. Spectral distortion of 36 microphones and 8 deg intervals of *controlled directions* (45 directions).

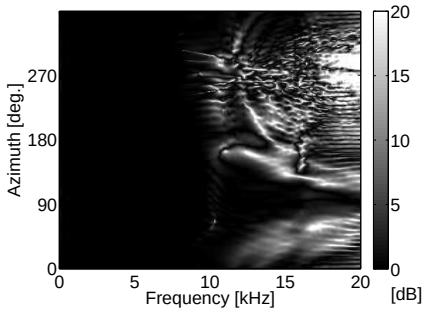

Fig. 10. Spectral distortion of 36 microphones and 6 deg intervals of *controlled directions* (60 directions).

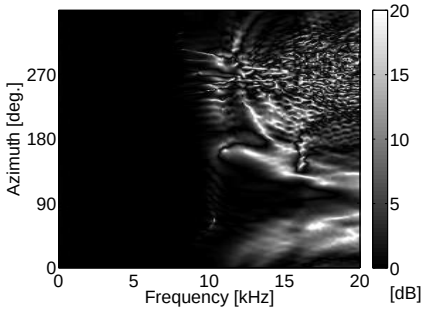

Fig. 11. Spectral distortion of 36 microphones and 5 deg intervals of *controlled directions* (72 directions).

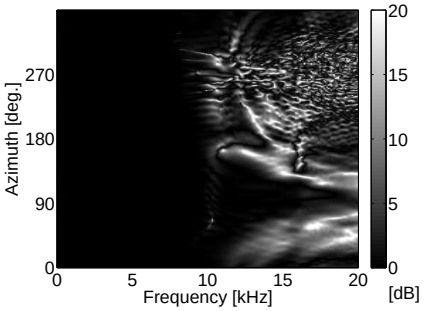

Fig. 12. Spectral distortion of 36 microphones and 4 deg intervals of *controlled directions* (90 directions).