Proceedings of 20th International Congress on Acoustics, ICA 2010

23–27 August 2010, Sydney, Australia

A 3D sound-space recording system using spherical microphone array with 252ch microphones

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-shi, Miyagi, 980-8577, Japan
(2) Sendai National Collage of Technology, 48 Nodayama, Medeshima-Shiote, Natori-shi, Miyagi 981-1239, Japan
(3) Graduate School of Engineering, Tohoku University, 6-6-05, Aza-Aoba, Aramaki, Aoba-ku, Sendai, 980-8579, Japan

PACS: 43.60.Dh, 43.60.Fg, 43.60.Jn, 43.58.Ta

ABSTRACT

We are developing a system to acquire 3D sound-space information that can transmit accurate sound-space information to a distant place using a microphone array on a human-head-sized solid sphere with numerous microphones on its surface. The system comprises an object with numerous microphones called Symmetrical object with ENchased ZIllion microphones (SENZI). Each recorded signal from spatially distributed multiple microphone is simply weighted and summed to synthesize listner's binaural signals. These are presented via headphones. Moreover, the weight is changed according to a human's 3D head movement. Therefore, 3D sound-space information is acquired accurately irrespective of the head movement. The results of the simulation of previous study indicate that 1) the microphones should be arranged densely to avoid the effect of spatial aliasing, 2) the number of *controlled directions* should be set densely at less than 5° intervals to synthesize sound sources from all directions, and 3) the radius of the microphone array should be equal to the size of listener's head to express the detail of frequency spectrum characteristics of the listener. In this study, we actually designed the system with 252 ch microphones accordingto these results and investigated the accuracy of synthesized 3D sound-space information. The results demonstrated that the developed system can synthesize 3D sound-space with high precision.

INTRODUCTION

Sensing, transmission, and reproduction of precise three-dimensional sound-space information are important to realize highly realistic communications. Although studies relating to reproduction technologies have been quite active, studies of sensing technologies are far fewer. Sensing must be at least as important as reproduction. Therefore, studies of highly realistic sensing of sound-space information should be conducted more intensively.

In this context, a few methods have been proposed to realize sensing of three-dimensional sound-space information [1, 2, 3, 4]. Toshima *et al.* proposed a mechanically steerable dummy-head called *TeleHead* [1]. This *TeleHead* can track a human's 3D head movement, as detected by the head tracker. To use this method practically, however, each listener must produce a personalized *TeleHead* to realize the sound-space precisely for the listener. Algazi *et al.* proposed a motion-tracked binaural recording technique called MTB [2, 3]. In this technique, instead of a dummy head, a sphere or a cylinder with several pairs of microphones is used. The pairs of microphones are installed at opposite positions along the diameter in a horizontal plane; one pair is selected according to the rotation of a listener's head when sound is recorded. However, the synthesized sound-space information might be insufficiently individualized.

Consequently, we have proposed a system to acquire 3D sound-space information that can transmit accurate sound-space information to a distant place using a microphone array on a human-head-sized solid sphere with numerous microphones on its surface. We designate this system as Symmetrical object with ENchased ZIllion microphones (SENZI) [5]. The system can sense 3D sound-space information comprehensively: information from all directions is available for any listener orientation 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.

A previous study examined the optimum shape of SENZI and arrangement of microphones [7]. The results of the simulation of the study indicated the specification and requirement of SENZI. In this study, we actually developed the system with 252 ch microphones according to these results and investigated the accuracy of synthesized 3D sound-space information.

SYSTEM OUTLINE

System concept

Figure 1 shows a concept of the system. The systems comprise objects with a microphone array. These objects are symmetrical to consider the listener's head motion and set at a sound field for recording the sound-space information. Recorded signals of all microphones are used for synthesizing a listener's binaural signals. These signals are presented via headphones.

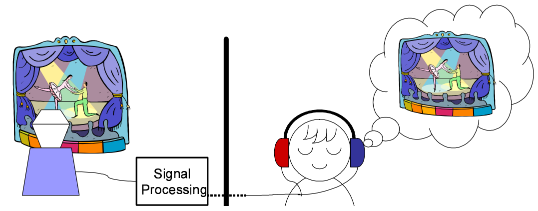

Figure 1: Concept of proposed sound-space sensing systems

Calculation method of binaural signals for individual listeners [5]

In the system, to calculate and synthesize a listener's binaural signals using input sound information from spatially distributed multiple microphones, each recorded signal from each microphone is simply weighted and summed. Moreover, the weight is changed according to a human's 3D head movement, which is an important cue to perceive a 3D sound space with high reality [6]. Therefore, 3D sound-space information is acquired accurately irrespective of the head movement.

As the simplest situation, we first assume the case in which sounds come only from the horizontal plane. Let $H_{listener}$ signify the listener's HRTF (head related transfer function) of one ear. For a certain frequency f, $H_{listener,f}(\theta)$ is expressed according to the following equation:

$$H_{listener,f}(\theta) = \sum_{i=1}^{n} z_{i,f} \cdot H_{i,f}(\theta) + \varepsilon. \qquad (1)$$

In that equation, $H_{i,f}(\theta)$ is the transfer function of the i-th microphone from a sound source as a function of the direction of the sound source (θ). Equation 1 reflects that the listener's HRTF is calculable from these transfer functions. Therefore, we can synthesize individualized binaural signals by using adequate set of $z_{i,f}$ for each listener. However, equation 1 cannot be solved; the residual ε remains if the number of directions of sound sources is smaller than that of microphones n. In fact, ε varies according to the weighting coefficient. 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. Moreover, when we apply this system to reverberant condition, direct sound and reverberation which are come from same direction of direct sound are assumed as one virtual sound source at the direction of direct sound (θ) and calculated $z_{i,f}$ is applied. This feature in our methods 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.

Figure 2 shows block diagram of the processing. Imput signals are weighted by calculated $z_{i,f}$ for binaural signal. Then, these are summed and presented to left/right ear.

When $z_{i,f}$ is calculated, we must select directions (θ), which are taken into the calculation. The selected directions are called as "*controlled directions*" hereafter. 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.

Symmetrical object with a microphone array for recording sound-space

We considered objects of two types for recording sound-space information. One is with a spherical head model equipped with numerous microphones called Symmetrical object with ENchased ZIllion microphones (SENZI). Actually, SENZI has many microphones arranged symmetrically on an air sphere or an acoustically hard sphere. Because it is a sphere, transfer functions obtained using these objects are very simple. The transfer functions were calculated through computer simulation [8].

Not only SENZI, but also another type of object is considered. It is designated as A Symmetrical and Universal Recording Array (ASURA). The size and shape of this object are based on a dummy head (SAMRAI: Koken Co. Ltd.). Eight pinna-like protuberances are set symmetrically on the object and microphones are arranged on the object. It is noteworthy that there is a sufficient difference among the frequency responses of each transfer function to synthesize the HRTFs of a listener. Therefore, we considered that some protuberances set on the object might be useful to express the difference of frequency responses.

Optimization of symmetrical object with a microphone array for a recording sound-space [7]

We consider that SENZI is better than ASURA as the first step because the shape is very simple to calculate transfer functions of the object using a computer simulation. To consider the optimum size of SENZI for recording sound-space information accurately, we investigated the number of microphones that should be used and how these microphones should be arranged on SENZI. We also examined the optimum size of SENZI. The results of computer simulation indicated that the following three points should be considered to develop SENZI.

1. Microphones should be arranged densely to avoid the effects of spatial aliasing [9].
2. The number of "*control directions*" should be set densely at less than 5° intervals to synthesize sound sources from all directions.
3. The SENZI radius should be equal to the size of a listener's head to express the details of frequency and phase characteristics of the listener.

Figure 3 portrays the relation between the average of spectral distortion (SD) and the interval of *controlled directions* for each microphone arrangement. SD is defined as following equation:

$$\varepsilon_f(\theta) = \left| 20\log_{10} \left| \frac{H_{\text{listener},f}(\theta)}{H_{\text{synthesized},f}(\theta)} \right| \right| \; [\text{dB}], \qquad (2)$$

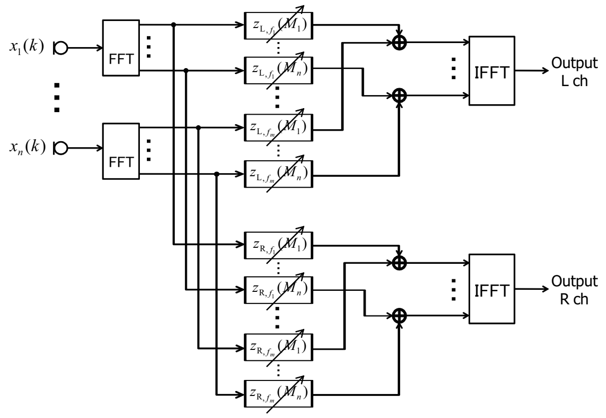

Figure 2: Block diagram of the system

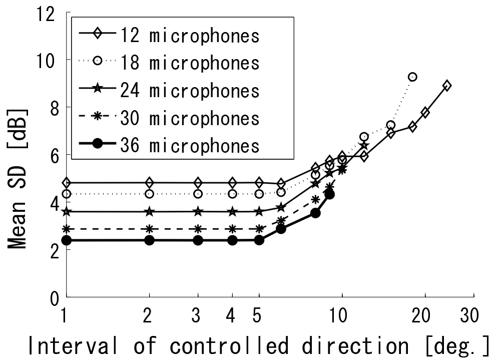

Figure 3: Relationship between mean SD and interval of *controlled directions* for each microphone arrangement

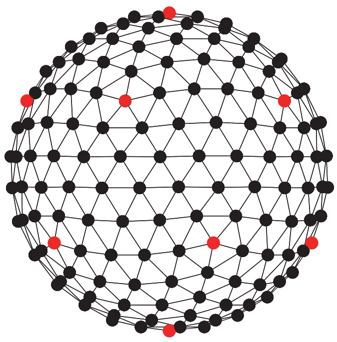

Figure 4: Microphone arrangement: Red circles indicate apices of a regular icosahedron.

where θ respectively represent the azimuth angles. The results show that the average of SD is decreasing when the interval of *controlled directions* narrows. Moreover, the average of SD is almost constant when the controlled directions are set densely at less than 5°.

ACCURACY OF SOUND-SPACE INFORMATION ACQUIRED USING SENZI

We designed SENZI, which satisfies these three requirements. To analyze the capabilities of the system, the sound-space synthesized by SENZI was compared with actual sound information and evaluated the accuracy as the index of SD. Based on the results, we actually constracted SENZI.

Specification of SENZI

Figure 4 shows SENZI designed in this study. The radius of the object is 17 cm. This size is decided according to the head size. There are 252 ch microphones on the spherical object. The position of each microphone is calculated based on a regular icosahedron [10, 11]. Each surface of a regular icosahedron is divided into 25 small equilateral triangles, and all apices of these triangles are projected to the surface of spherical object. These points on the spherical object are the microphone positions. The interval of each microphone is almost equal to those of the others: about 20 mm.

Condition of the simulation

The accuracy of synthesized sound-space was analyzed by computer simulation. The *controlled directions* were 2,562. The interval of two *controlled directions* is about 4.7°. These were set at 1.5 m distance from the center of SENZI. Positions of these directions were calculated to be identical to the positions of microphones.

The HRTFs of the left ear of a dummy head (SAMRAI; Koken Co. Ltd.) were used as the target characteristics to be realized by this system. They were calculated using the boundary element method (BEM). The frequency range for synthesis was set as 93.75 Hz steps for 0–20 kHz. Weighting coefficient $z_{i,f}$ was calculated under these conditions. The HRTFs of 1° interval were synthesized, and SD was calculated.

Results and discussions

Figure 5 shows HRTFs of SAMRAI as the target and Fig. 6 shows the transfer function synthesized by SENZI. A slight difference between both figures is observed at a part of the high-frequency region, especially from 45° to 135°. As we mentioned in the previous section, the average interval of two microphones on SENZI is 20 mm. This means that the frequency at which the effect of spatial aliasing [9] is observed more than around 8.5 kHz. Moreover, the difference observed around the range from 45° to 135° would include the influence of a shadow effect of the listener's head. Therefore, the difference is observed at this region. This tendency can be observed clearly on Fig. 7. Figure 7 shows the spectral distortion between target HRTFs and the synthesized transfer functions. This figure shows that spectral distortion is observed at a frequency of more than around 5–10 kHz. Moreover, severe distortion is observed at the part of the head shadow.

Although some differences are observed, details of peak and dip patterns below 10 kHz are almost identical between both transfer functions, indicating that the proposed SENZI can record and synthesize sound space information accurately. Improving the accuracy of synthesized sound-space is the easy way to increase the number of microphones so that microphones could be set more densely.

CONSTRACTION OF SENZI

Based on these results, we actually constracted SENZI shown in Fig. 8. A small digital EMC omnidirectional microphone (KUS5147; Hoshiden Co. Ltd.) is set at the calculated position.

We are developing proposed system with SENZI into a real-time sound space recording/synthesizing system. We cosider that it is important to evaluate accuracy of synthesized sound-space information not only from objective mesurement but also from subjective experiment as the future study.

CONCLUSION

We have proposed a system to acquire 3D sound-space information that can transmit accurate sound-space information to a distant place using a microphone array on a human-head-sized solid sphere with numerous microphones on its surface. The system comprises an object with numerous microphones called Symmetrical object with ENchased ZIllion microphones (SENZI). In this paper, we actually developed the system with 252 ch microphones according to results of a previous study and investigated the accuracy of the synthesized 3D sound-space information. The results demonstrated that the developed system can synthesize a 3D sound-space with high precision.

ACKNOWLEDGEMENT

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

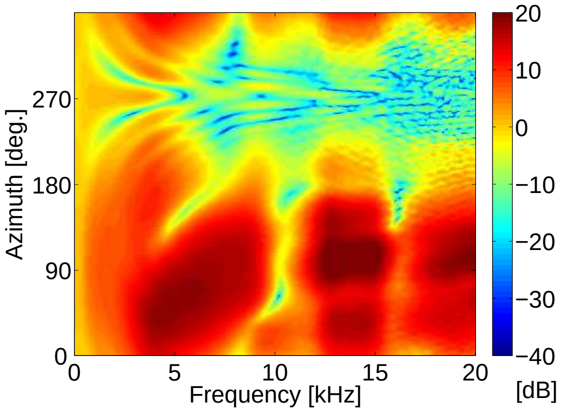

Figure 5: HRTFs of SAMRAI (target HRTF)

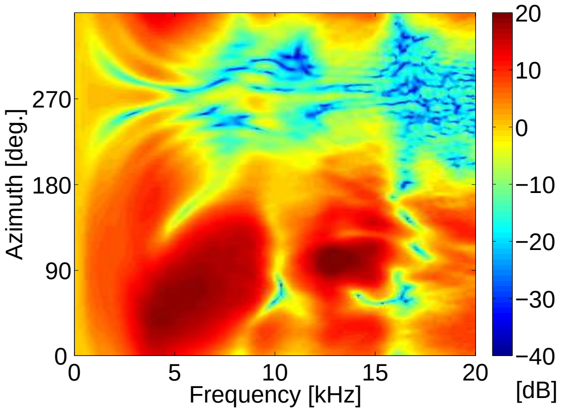

Figure 6: Transfer functions synthesized by SENZI

(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 HRTFs.

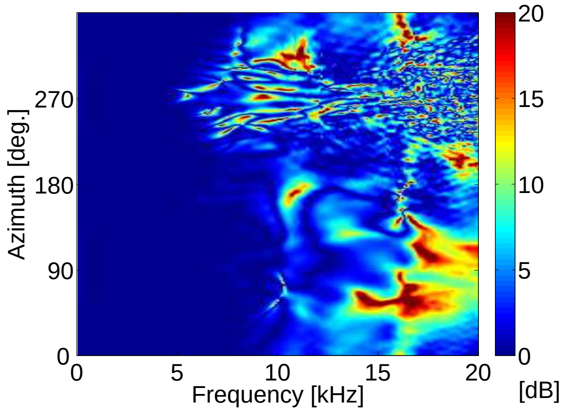

Figure 7: Spectral distortion between target HRTFs and synthesized HRTFs

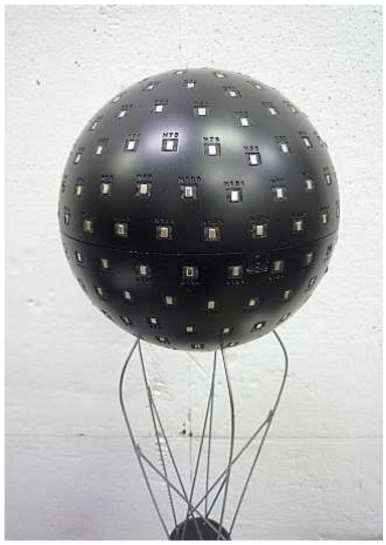

Figure 8: Photograph of constracted SENZI

REFERENCES

[1] I. Toshima, H. Uematsu and T. Hirahara, “A steerable dummy head that tracks three-dimensional head movement: *TeleHead*,” *Acoustical Science and Technology*, 24(5), pp. 327–329, 2003

[2] V. R. Algazi, R. O. Duda and D. M. Thompson, “Motion-Tracked Binaural Sound,” *Journal of Audio Engineering Society*, 52(11), pp. 1142–1156, 2004

[3] 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, pp. 1–20, 2004

[4] R. Nicol, “Sound spatialization by higher order Ambisonics: Encoding and decoding a sound scene in practice from a theoretical point of view...,” Proc. 2nd International Symposium on Ambisonics and Sperical Acoust., 2010

[5] 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. 2nd International Symposium on Universal Communication (ISUC2008) pp. 429–434, 2008

[6] Y. Iwaya, Y. Suzuki and D. Kimura, “Effects of head movement on front-back error in sound localization,” *Acoustical Science and Technology*, 24(5), pp. 322–324, 2003

[7] J. Kodama, S. Sakamoto, S. Hongo, T. Okamoto, Y. Iwaya and Y. Suzuki, “Effects of microphone arrangements on the accuracy of a spherical microphone array (SENZI) in acquiring high-definition 3D sound space information,” Proc. International Workshop on the Principal and Applications of Spatial Hearing (IWPASH2009), P28, 2009

[8] R. O. Duda, W. L. Martens, “Range dependence of the response of a spherical head model,” Journal of Acoustical Society of America 104(5), pp.3048–3058, 1998

[9] S. U. Pillai, “Array signal processing,” Springer-Verlag, New York, 1989.

[10] http://www.emit.jp/prog/prog d1.html

[11] http://kabuki.spring8.or.jp/Members/matusita/forums/reconstraction/tips/sph