



on Fundamentals of Electronics, Communications and Computer Sciences

**VOL. E97-A NO. 9
SEPTEMBER 2014**

**The usage of this PDF file must comply with the IEICE Provisions
on Copyright.**

**The author(s) can distribute this PDF file for research and
educational (nonprofit) purposes only.**

Distribution by anyone other than the author(s) is prohibited.

A PUBLICATION OF THE ENGINEERING SCIENCES SOCIETY



The Institute of Electronics, Information and Communication Engineers

Kikai-Shinko-Kaikan Bldg., 5-8, Shibakoen 3 chome, Minato-ku, TOKYO, 105-0011 JAPAN

[PAPER] *Special Section on Spatial Acoustic Signal Processing and Applications*

3D Sound-Space Sensing Method Based on Numerous Symmetrically Arranged Microphones

Shuichi SAKAMOTO^{†a)}, Satoshi HONGO^{††}, *Members*, and Yōiti SUZUKI[†], *Fellow*

SUMMARY Sensing and reproduction of precise sound-space information is important to realize highly realistic audio communications. This study was conducted to realize high-precision sensors of 3D sound-space information for transmission to distant places and for preservation of sound data for the future. Proposed method comprises a compact and spherical object with numerous microphones. Each recorded signal from multiple microphones that are uniformly distributed on the sphere is simply weighted and summed to synthesize signals to be presented to a listener's left and right ears. The calculated signals are presented binaurally via ordinary binaural systems such as headphones. Moreover, the weight can be changed according to a human's 3D head movement. A human's 3D head movement is well known to be a crucially important factor to facilitate human spatial hearing. For accurate spatial hearing, 3D sound-space information is acquired as accurately reflecting the listener's head movement. We named the proposed method SENZI (Symmetrical object with ENchased ZIllion microphones). The results of computer simulations demonstrate that our proposed SENZI outperforms a conventional method (binaural Ambisonics). It can sense 3D sound-space with high precision over a wide frequency range.

key words: *sound field recording, head-related transfer function (HRTF), spherical microphone array, tele-existence*

1. Introduction

Sensing and reproduction of accurate 3D sound-space information is an important technology to realize highly realistic audio communications. Although studies related to reproduction technologies have been conducted actively [1]–[3], reported studies of sensing technologies are far fewer. Sensing is as important as reproduction. For that reason, studies of highly realistic sensing of sound-space information should be conducted more intensively.

Regarding sensing of 3D sound space-information at a certain place by compact systems, the most conventional and the simplest technique is binaural recording using a dummy-head. In this technique, dummy-head microphones are set at a recording place. The sound is recorded via microphones located at the ear drum positions of the dummy-head. This extremely simple technique can easily record 3D sound-space information. However, the sound-space information acquired using this technique has poor accuracy because head-related transfer functions (HRTFs) are highly

individual and those of a dummy-head are quite different from those of most listeners. Moreover, movement of the listener's head cannot be reflected in the acquired sound-space information. Many researchers [4]–[8] have demonstrated that listeners' head movements are effective to enhance the localization accuracy as well as the perceived realism in human spatial hearing perception, not only in real-world environments but also in virtual environments. It is important to reflect the listener's head movements when presenting sound-space information.

In this context, a few methods have been proposed to realize sensing of 3D sound-space information [9]–[12]. Toshima et al. proposed a steerable dummy-head called *TeleHead* [9], [10]. This *TeleHead* can track a listener'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 each listener. Moreover, head movements cannot be reflected in the outputs after recording through the *TeleHead*. Algazi et al. proposed a motion-tracked binaural (MTB) recording technique [11]. 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 circumference: one pair is selected according to the movement of the listener's head when sound is recorded. This technique has been modified as introduced in their subsequent report [12]. In the modified technique, synthesized HRTFs are personalized by incorporating the shape of the listener, particularly addressing the effect of the pinna. However, synthesized HRTFs are insufficiently personalized.

As another approach to sense and/or to reproduce sound information accurately, Ambisonics, especially High-order Ambisonics (HOA), have been examined specifically [13], [14]. In this technique, 3D sound-space information is encoded and decoded on several components with specific directivities based on spherical harmonic decomposition. However, ordinary HOA reproduced with loudspeakers require large systems, typically including more than 30 loudspeakers. A binaural version of HOA, of which typical output device is headphones, has been developed [15], [16]. This technique, called binaural Ambisonics, seems to the authors to be the most promising among the conventional methods, but its performance has been little examined. For example, it remains unclear how a high order of HOA is necessary to yield directional resolution that is sufficient to satisfy perceptual resolution. As an important step to pro-

Manuscript received December 12, 2013.

Manuscript revised April 11, 2014.

[†]The authors are with Research Institute of Electrical Communication, Tohoku University, Sendai-shi, 980-8577 Japan.

^{††}The author is with the Department of Design and Computer Applications, Sendai National College of Technology, Natori-shi, 981-1239 Japan.

a) E-mail: saka@ais.rieic.tohoku.ac.jp

DOI: 10.1587/transfun.E97.A.1893

viding sensing systems matching the performance of human spatial hearing are still highly desired.

For this study, we propose a method that can sense and record accurate sound-space information. This recorded information can not only be transmitted to distant places in real time, but can also be stored in media and properly reproduced in the future. The core component of this method is a microphone array on a human-head-sized solid sphere with numerous microphones on its surface. This proposed method can sense 3D sound-space information comprehensively: information from all directions is available over different locations and over time once it is recorded, for any listener orientation and head/ear shape with correct binaural cues. The proposed method based on spherical microphone array is named SENZI (Symmetrical object with ENchased Zillion microphones), which means “thousands of ears” in Japanese.

In Sect. 2, an outline of the proposed SENZI is presented. We then investigate the performance of SENZI by evaluating the accuracies of synthesized HRTFs using computer simulations. In Sect. 3, simulation conditions for the evaluation of the accuracy of synthesized sound-space are explained in detail. Section 4 presents salient results including a comparison with the performance of binaural Ambisonics of the same system size. In this section, the results and the realizability of the proposed method are also discussed. Finally, several concluding remarks are presented in Sect. 5.

2. Outline of the Proposed Method

The proposed method, SENZI, comprises a compact human-head-sized solid spherical object with a microphone array on its surface. The microphones are distributed uniformly and symmetrically to accommodate and adapt to the listener's head rotations. Two signals are synthesized by the recorded signals and outputted to the listener. The signals yield binaural signals so that it can function as if it were the dummy-head of each listener. Individual listeners' HRTFs are synthesized using recorded signals of all microphones. The calculated binaural signals can be presented via any type of binaural system such as headphones or transaural system. In this section, the proposed signal processing method is presented.

2.1 Synthesis Method of HRTFs for Individual Listeners with a Sphere Microphone Array

In this subsection, a signal processing algorithm to synthesize HRTFs with a microphone array is proposed. To calculate and synthesize a listener's HRTFs by inputs from spatially distributed multiple microphones, recorded signals from each microphone are simply weighted and summed to synthesize a listener's HRTF in the proposed method. Moreover, the weight can be changed easily according to a human's 3D head movement as described in the next subsection. Because of that feature, 3D sound-space information

is acquired accurately irrespective of the head movement. Detailed algorithms are explained from the next paragraph.

Let \mathbf{H}_{lis} signify a specified listener's HRTFs for one ear. Although HRTFs are usually expressed as the function of the frequency, \mathbf{H}_{lis} is expressed as a function of the direction of the sound source. For a certain frequency f , $\mathbf{H}_{\text{lis},f}$ is expressed according to the following equation.

$$\begin{bmatrix} H_{\text{lis},f}(\Theta_1) \\ \vdots \\ H_{\text{lis},f}(\Theta_m) \end{bmatrix} = \begin{bmatrix} H_{1,f}(\Theta_1) & \cdots & H_{n,f}(\Theta_1) \\ \vdots & \ddots & \vdots \\ H_{1,f}(\Theta_m) & \cdots & H_{n,f}(\Theta_m) \end{bmatrix} \begin{bmatrix} z_{1,f} \\ \vdots \\ z_{n,f} \end{bmatrix}, \quad (1)$$

In that equation, $H_{i,f}(\Theta_j)$ is the object-related transfer function of the sound propagation path between i -th microphone and j -th sound source at the direction of Θ_j . $z_{i,f}$ is the weighting coefficient of the i -th microphone at the frequency f . Equation (1) means that HRTFs of a specified listener are calculable from these transfer functions using common $z_{i,f}$ to all m directions. Consequently, $z_{i,f}$ is not dependent on the sound source direction. Because Eq. (1) is overdetermined if the number of directions of sound sources (m) is greater than that of microphones (n). Optimum $z_{i,f}$ is given by solving Eq. (1) with, for example, nonlinear least mean squares (LMS) method or the pseudo-inverse matrix. The result is represented by the following equation:

$$\begin{aligned} \mathbf{H}_{\text{lis},f} &= \mathbf{H}_f \cdot \hat{\mathbf{z}}_f + \boldsymbol{\varepsilon}, \\ \mathbf{H}_f &= [\mathbf{H}_{1,f} \cdots \mathbf{H}_{n,f}], \\ \hat{\mathbf{z}}_f &= [\hat{z}_{1,f} \cdots \hat{z}_{n,f}]^T. \end{aligned} \quad (2)$$

Coefficient $\hat{z}_{i,f}$ is given for each microphone at each frequency. A calculated $\hat{z}_{i,f}$ is a constant complex that is common to all directions of the sound sources. Moreover, when this method is applied to reverberant environment, direct sound, early reflections and reverberation that come from the same direction as that of the direct sound can be regarded as one virtual sound source at the direction of a direct sound (Θ_j). For that reason, the sound source positions need not be considered at all to sense sound-space information coming from any number of sound sources including early reflections and reverberation is acquired by this method. This is an important benefit of the proposed method.

The next step is the derivation of the output signal representing recorded sound-space at frequency f information using Eq. (1). Letting $X_{i,f}$ signify a frequency (f) component of the recorded sound of the i -th microphone from all the sound sources $S(\Theta_j)$ from $j = 1$ to m , then $X_{i,f}$ is derived according to the following equation:

$$X_{i,f} = \sum_{j=1}^m S_f(\Theta_j) H_{i,f}(\Theta_j). \quad (3)$$

This $X_{i,f}$ is given as the product of sound sources S_f and object-related transfer function $H_{i,f}$, which is the transfer function of the sound propagation path from the sound sources to the i -th microphone. All the recorded sounds for n microphones X_f are expressed according to the following

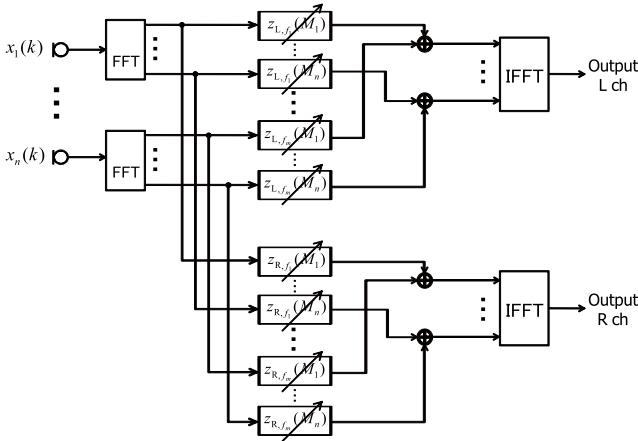


Fig. 1 Block diagram of the proposed system.

matrix:

$$\mathbf{X}_f = \mathbf{S}_f \cdot \mathbf{H}_f, \quad (4)$$

$$\mathbf{X}_f = [X_{1,f} \cdots X_{n,f}],$$

$$\mathbf{S}_f = [S_f(\Theta_1) \cdots S_f(\Theta_m)].$$

The following equation is obtained by multiplying $z_{i,f}$ obtained by Eq. (1) and Eq. (4) from the right side:

$$\begin{aligned} \mathbf{X}_f \cdot \mathbf{z}_f &= \mathbf{S}_f \cdot \mathbf{H}_f \cdot \mathbf{z}_f \\ &\approx \mathbf{S}_f \cdot \mathbf{H}_{lis,f}. \end{aligned} \quad (5)$$

The right side of this equation expresses the recorded sound itself arriving at the position of the listener's one of two ears from all sound sources. The left side of this equation is the simple inner product of recorded signals of all microphones and weighting coefficients \mathbf{z}_f , which is given by Eq. (1) in the design stage.

Figure 1 presents a block diagram showing the structure of the proposed method comprehensively. Input signals are weighted by calculated $z_{i,f}$ and summed to derive the binaural signals to be presented to left/right ear as shown by the left side of Eq. (5).

2.2 Consideration of Listener's Head Rotation

In this subsection, the signal processing method to reflect the listener's head rotation is examined. SENZI can reflect the listener's head rotations without difficulty when the synthesized sound-space is reproduced once the microphone signals are recorded. This simplicity is another important benefit of the method.

The situation portrayed in Fig. 2 is described next. Here, 0° of sound source is defined as the listener's initial frontal direction. The sound source angle is defined as positive when the sound source is positioned as the direction of the target ear. Therefore, the counter-clockwise direction is positive for the left ear, as shown in Fig. 2, whereas a clockwise direction is positive for the right ear. Furthermore, 0° of head rotation is defined as the direction relative to the initial frontal direction of the listener. This angle is, as shown

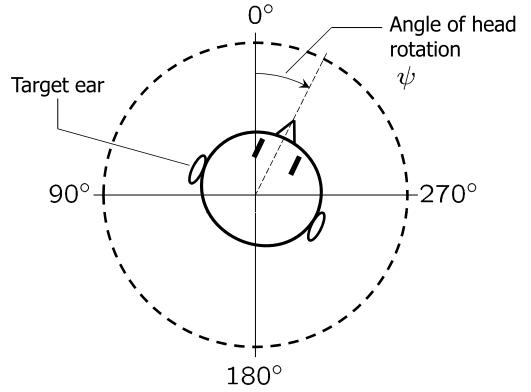


Fig. 2 Assumed scenario of the listener's head rotation.

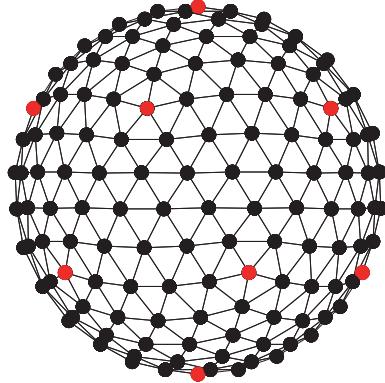


Fig. 3 Microphone arrangement of spherical microphone array evaluated in this study (17 cm diameter): red circles represent apices of a regular icosahedron.

by the arced arrow in Fig. 2, defined as positive when the head is moving opposite to the direction of the target ear. Clockwise rotation is positive for the left ear, as shown in Fig. 2, whereas the counter-clockwise direction is positive for the right ear.

The relative position of the sound source to the listener changes in the opposite direction of the listener's motion when the listener moves, although the sound source is stationary. For that reason, the target HRTFs to be synthesized are simply changed to the opposite direction of the listener's head rotation. For instance, when the listener's head rotates ψ as shown in Fig. 2, the HRTF of $\theta + \psi$ should be used instead of HRTF of θ for a sound source at the direction of θ . In the signal processing of SENZI, this conversion can be adapted simply and uniformly to any of the three-dimensional rotations of yaw, roll, and pitch because the shape is completely symmetry, i.e. a sphere.

2.3 Recording Procedure of Sound-Space by SENZI

To realize accurate sound-space reproduction with SENZI, microphones should be arranged densely to avoid spatial aliasing [17]. Figure 3 shows spherical microphone array designed for evaluation in this study. In this proposed SENZI, the number of microphones can be chosen freely

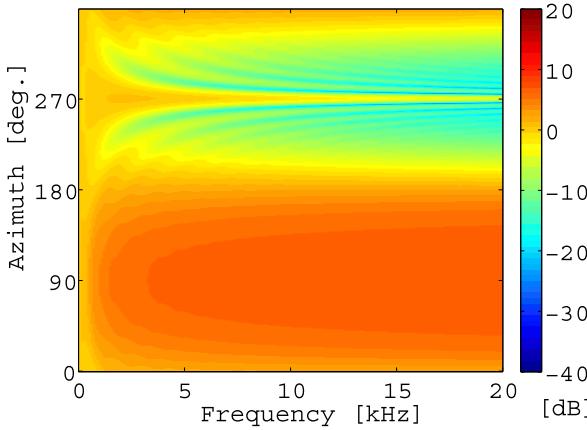


Fig. 4 Object-related transfer functions of sound propagation paths from sound sources 1.5 m above the equator of the spherical microphone array (Fig. 3) at the longitude shown by the coordinate to a microphone on its equator at longitude of either 90 or 270 deg.

according to the required accuracy of sound space or hardware limitation. In this study, the number of microphones was chosen as 252 according to the current hardware performance for development of a real-time system. The radius of the object is set to 17 cm. This size is determined according to the average human head size. Because the microphone array consisted on the SENZI is spherical, transfer functions of sound propagation paths from a sound source to the microphones, i.e. $H_{i,f}$, in Eq. (1), are very simple. The transfer functions are given as a numerical calculation based on the formula of the diffraction of a solid sphere [18]. Calculated transfer functions are depicted in Fig. 4. In this figure, a sound source is placed at 1.5 m in front of the object and signals are recorded by the microphone placed at the position corresponding to the ear (90°).

3. Accuracy Evaluation of a Model of This Proposed Method

The HRTFs of a dummy-head (SAMRAI; Koken Co. Ltd.) were used as the target to be realized by the SENZI (Fig. 3) to evaluate the performance. A person having the same ear, head, and torso shape as the dummy-head was assumed as the listener of the SENZI. Using these target HRTFs, the synthesized HRTFs are portrayed in Fig. 5. The target HRTFs were for sound on the horizontal plane on the ear level of the dummy-head at the distance of 1.5 m and were computed numerically using the boundary element method [19].

In the SENZI, 252 ch microphones were almost uniformly distributed on the rigid surface (Fig. 3). The position of each microphone was calculated based on a regular icosahedron [20]. Each surface of a regular icosahedron was divided into 25 small equilateral triangles. All apices of these triangles were projected to the surface of the sphere. These 252 points were regarded as microphone positions. The interval of each microphone was almost equal to those of the others: about 2.0 cm.

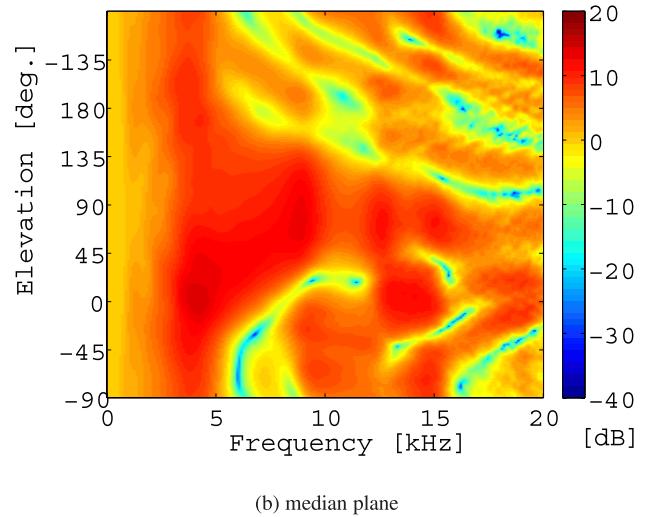
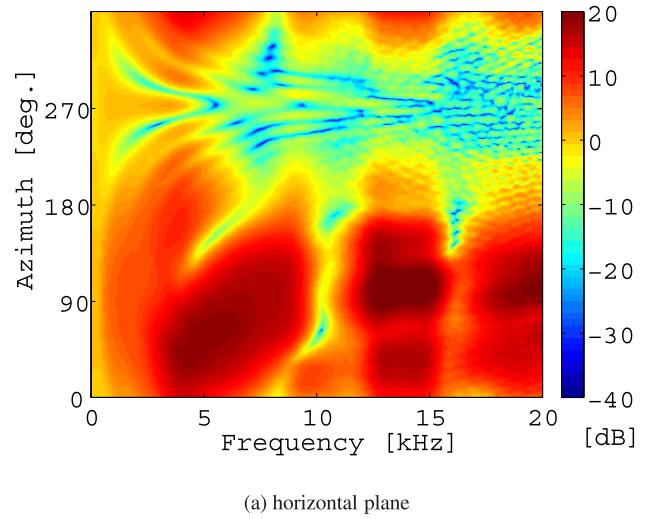


Fig. 5 HRTFs pattern of the dummy-head (SAMRAI).

The accuracy of the synthesized sound-space in terms of the synthesized HRTF for a point sound source was analyzed using a computer simulation. To do this, 2,562 HRTFs were used for calculating weighting coefficients $z_{i,f}$ expressed in Eq. (1). The 2,562 sound source positions were determined as follows: First a regular icosahedron inscribed in a sphere with radius of 1.5 m was assumed. Next, each surface of the regular icosahedron was divided into 256 small equilateral triangles. All apices of these triangles were projected to the surface of the sphere. Results show that 2,562 apices were obtained at the distance of 1.5 m from the center of the sphere and were used as sound source positions. HRTFs from obtained sound positions were calculated. The sampling frequency was 48 kHz.

As a conventional 3D sound-space reproduction technique with a compact spherical microphone array, a binaural sound reproduction method based on High-order Ambisonics [15], [16] (binaural Ambisonics) was implemented and compared with the proposed method. In this technique, a

3D sound-space recorded using a compact spherical microphone array is first analyzed using spherical harmonic analysis. A loudspeaker array is then arranged virtually. The analyzed sound-space is reproduced using the virtual speaker array by convolving HRTFs corresponding to each of the virtual loudspeaker positions. The recording setups for the conventional technique are identical to those for SENZI. Therefore, 252 ch microphones allow spherical harmonics decompositions up to the order of 14. The number and position of the virtual speaker array is the same as that of HRTFs used for calculating weighting coefficients $z_{i,f}$ of SENZI. To compensate for low-frequency distortion, the reconstruction order was changed according to the frequency, i.e. smaller orders were applied in low-frequency regions [21].

The spectral distortion (SD) of HRTFs synthesized using each method was calculated based on the following equation.

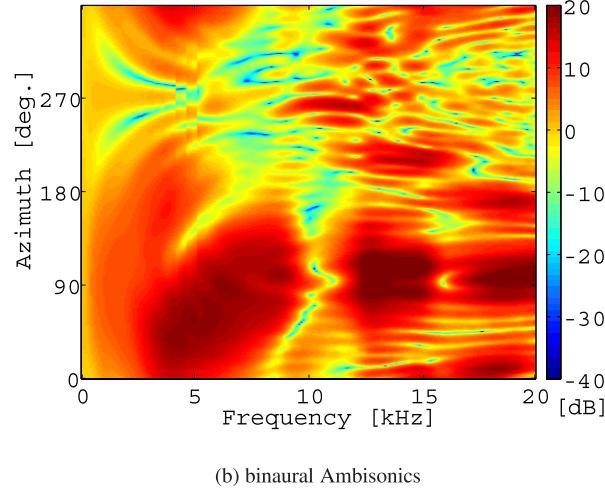
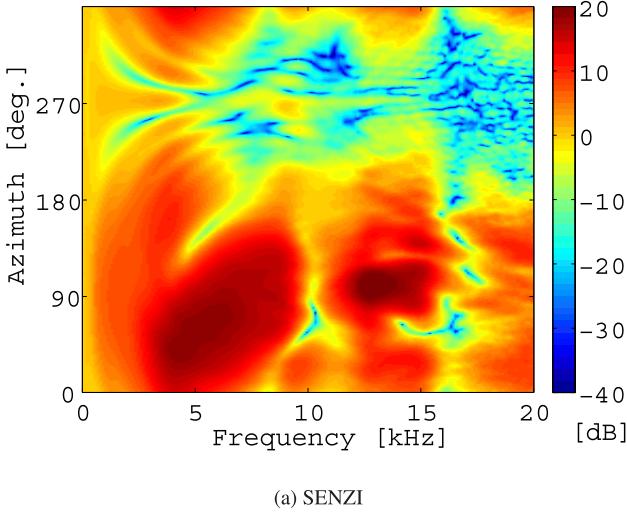


Fig. 6 Azimuthal patterns of synthesized HRTFs using the proposed SENZI and binaural Ambisonics.

$$\varepsilon_{SD}(f, \theta, \phi) = \left| 20 \log_{10} \left| \frac{HRTF_{target}(f, \theta, \phi)}{HRTF_{synthesized}(f, \theta, \phi)} \right| \right| [\text{dB}] \quad (6)$$

Therein, θ and ϕ respectively represent the azimuth and elevation angles. The synthesis accuracy along frequency f was also measured using the normalized spherical correlation (SC) based on the following equation [22]:

$$\varepsilon_{SC}(f) = \frac{\sum_{\theta} \sum_{\phi} HRTF_{target}(f, \theta, \phi) HRTF_{synthesized}(f, \theta, \phi)}{\sqrt{\sum_{\theta} \sum_{\phi} HRTF_{target}(f, \theta, \phi)^2} \sqrt{\sum_{\theta} \sum_{\phi} HRTF_{synthesized}(f, \theta, \phi)^2}}. \quad (7)$$

4. Results and Discussion

Directional patterns of HRTFs synthesized for all azimuth angles at 0° of the elevation angle are presented in Fig. 6

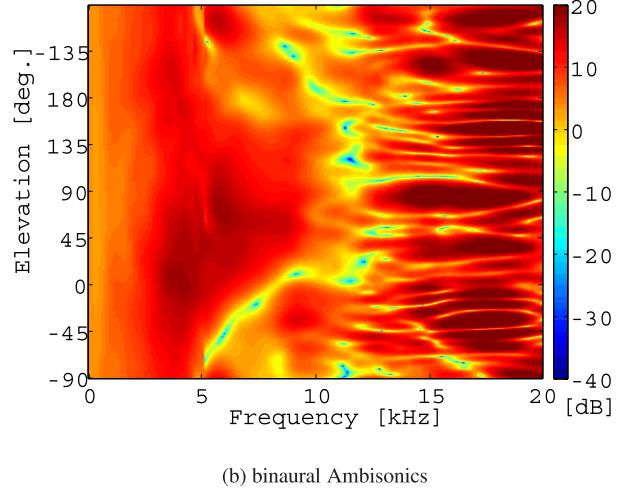
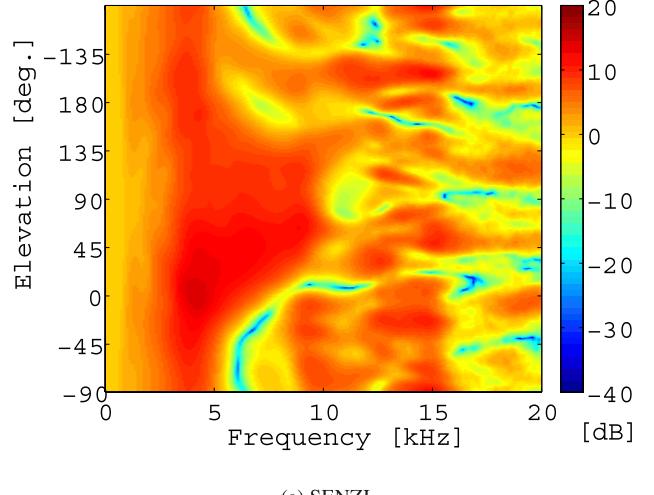


Fig. 7 Elevational patterns of synthesized HRTFs using the proposed SENZI and binaural Ambisonics.

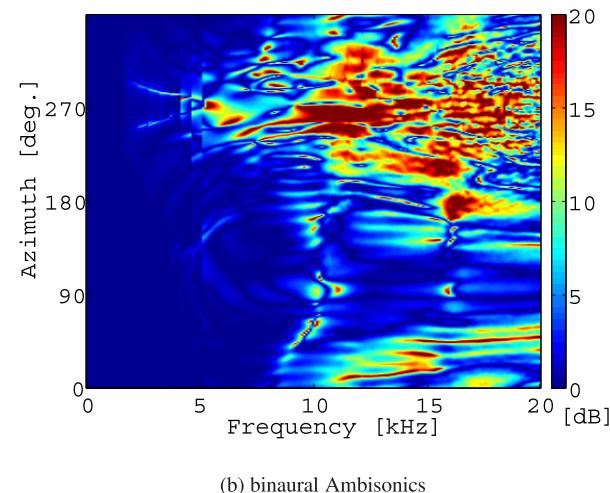
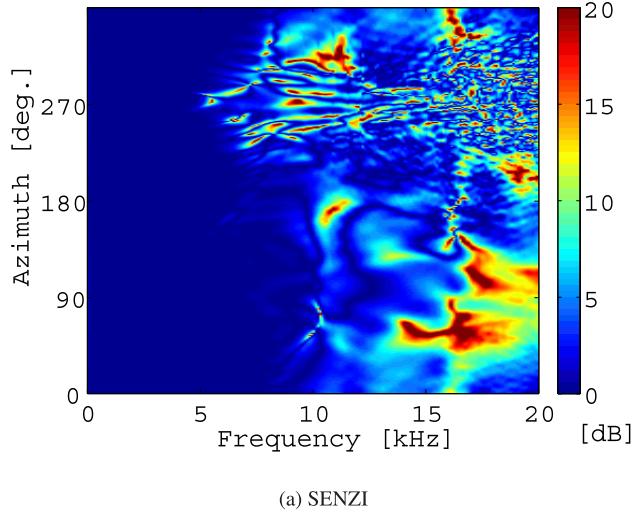


Fig.8 Reproduction error of synthesized azimuthal HRTFs using the proposed SENZI and binaural Ambisonics.

and those synthesized for all elevation angles at 0° of the azimuth angle are presented in Fig. 7. The frequency characteristics up to around 10 kHz are well expressed using both methods. In contrast, because of the spatial aliasing, the accuracy of synthesized HRTFs is degraded over 10 kHz. Nevertheless, this degradation is apparently small in the proposed method compared to that within conventional binaural Ambisonics. Figures 8 and 9 present spectral distortions at 0° of the elevation angle and 0° of the azimuth angle calculated using both methods. As shown in Figs. 6 and 7, they are shown clearly that both methods present the listener's sound-space information at almost all frequencies up to around 10 kHz for all azimuths and elevations except for the head shadow region around 270° . It is extremely interesting that much smaller SDs are observed in the high frequency region where spatial aliasing occurs using the proposed method compared with the conventional binaural Ambisonics because the weighting coefficients are

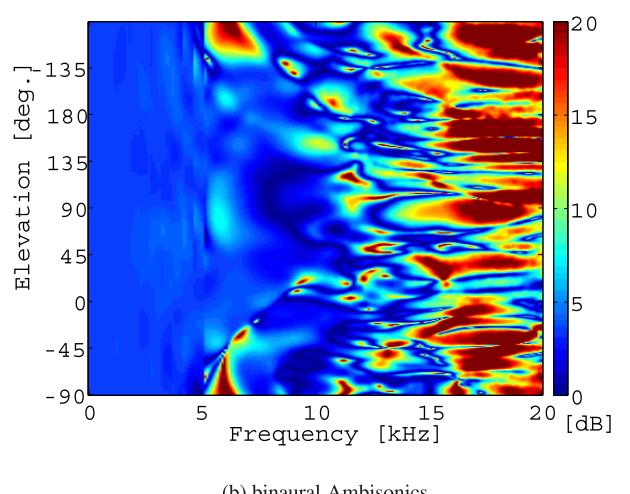
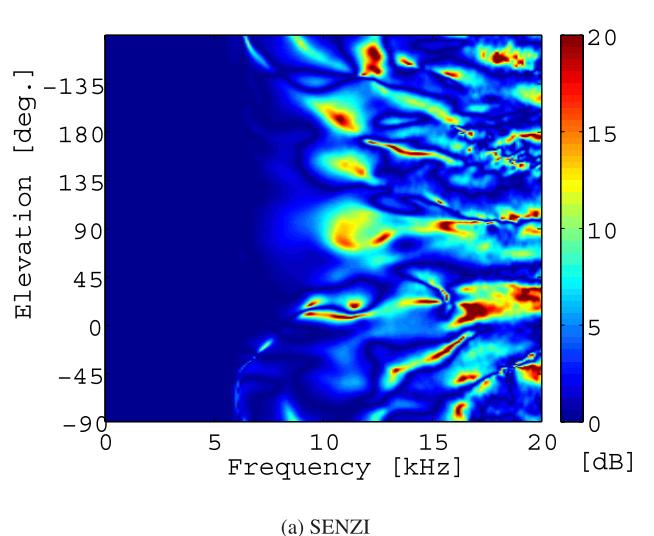


Fig.9 Reproduction error of synthesized elevational HRTFs using the proposed SENZI and binaural Ambisonics.

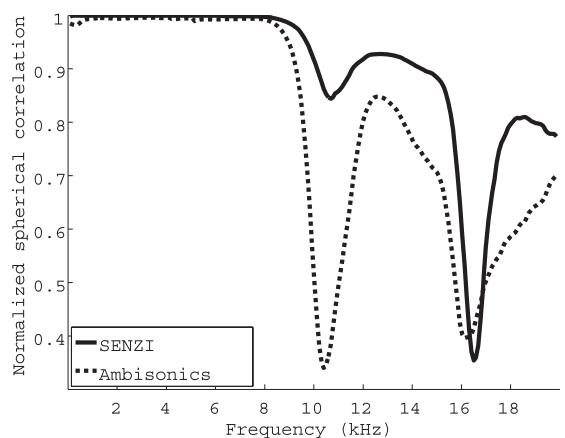
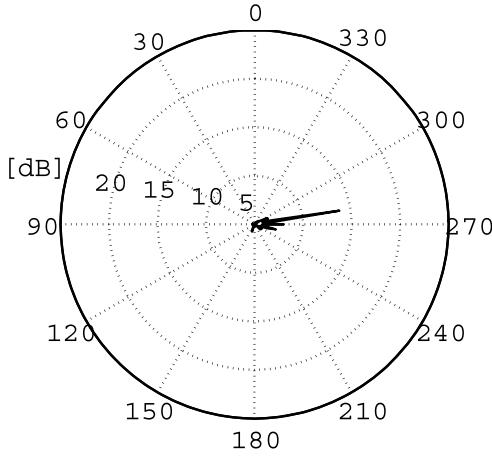
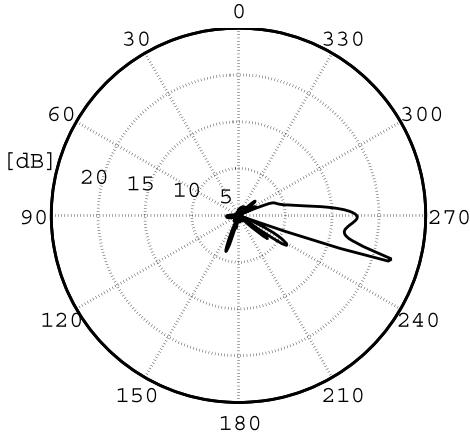


Fig.10 Normalized spherical correlation of synthesized HRTFs on the horizontal plane.



(a) SENZI



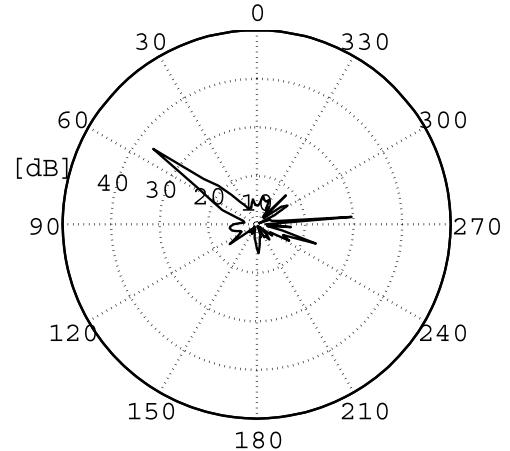
(b) binaural Ambisonics

Fig. 11 Numerical accuracy of the synthesized sound-space (directivity pattern: $f=6$ kHz).

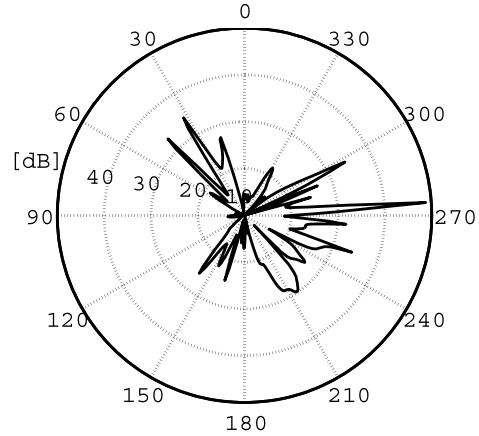
chosen to minimize the residual ε defined at Eq. (2) not only at the low frequency region but also at the high frequency region in the proposed method.

Figure 10 portrays a normalized spherical correlation of synthesized HRTFs on the horizontal plane ($\phi = 0$) calculated using Eq. (7). Two dips are observed at frequencies of around 10 kHz and 16 kHz in both methods. These frequencies correspond to multiple frequencies of spatial aliasing calculated using the interval of neighboring microphones. Even though such effects are observed here, higher spherical correlations are observed at almost all frequencies in the proposed method than in binaural Ambisonics.

Figures 11 and 12 show directivity patterns of numerical accuracy of the synthesized sound-space at the elevation angle of 0 degrees calculated using Eq. (6). To analyze the effects of spatial aliasing, two frequencies higher and lower than the spatial aliasing frequency (around 8 kHz) were selected. In both methods, the accuracy of the synthesized



(a) SENZI



(b) binaural Ambisonics

Fig. 12 Numerical accuracy of the synthesized sound-space (directivity pattern: $f=15$ kHz).

sound space is degraded in several directions. The degree of degradation is much smaller in the proposed method than in the conventional binaural Ambisonics. These results suggest that our proposed method can synthesize sound-space more accurately than the conventional method at all frequencies and in all directions.

Although a rigid sphere is used as a compact human-head-size solid object in this study, the proposed SENZI is applicable to an object of any shape or arrangement of the microphones. Actually, the accuracy of the synthesized sound-space depends on the object shape and the arrangement of microphones. It is therefore important to discuss what object requirements should be satisfied to realize an actual system. The purpose of this paper is to propose the basic idea of our method, SENZI. For that reason, this matter is beyond the scope of this study. We intend to study this topic in the near future.

5. Conclusion

As described in this paper, we presented a method for a three-dimensional comprehensive sound-space information acquisition based on a spherical microphone array. We called this method SENZI. Proposed SENZI is an effective but simple signal processing method based on a set of weighted summations to convert object-related transfer functions into a listener's head-related transfer functions (HRTFs). Moreover, we analyzed the accuracy of the proposed method and compared it with that of a conventional binaural Ambisonics technique. The simulation results showed clearly that our proposed SENZI outperforms a conventional method (binaural Ambisonics) and that it can sense 3D sound-space with high precision over a wide frequency range.

Acknowledgments

Portions of this report were presented at the Second International Symposium on Universal Communication (ISUC2008) and 20th International Congress on Acoustics (ICA2010). A part of 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. We wish to thank Prof. Yukio Iwaya and Dr. Takuma Okamoto for their intensive discussion and invaluable advice related to this study. We also wish to thank Mr. Ryo Kadoi, Mr. Jun'ichi Kodama, Mr. Cesar Salvador, and Mr. Yoshiki Satou for their help with analysis of the results.

References

- [1] A.J. Berkhouit, D. de Vries, and P. Vogel, "Acoustic control by wave field synthesis," *J. Acoust. Soc. Am.*, vol.93, pp.2764–2778, 1993.
- [2] S. Ise, "A principle of sound field control based on the Kirchhoff-Helmholtz integral equation and the theory of inverse systems," *Acta Acustica United with Acustica*, vol.85, pp.78–87, 1999.
- [3] S. Takane, Y. Suzuki, T. Miyajima, and T. Sone, "A new theory for high definition virtual acoustic display named ADVISE," *Acoust. Sci. & Tech.*, vol.24, pp.276–283, 2003.
- [4] H. Wallach, "On sound localization," *J. Acoust. Soc. Am.*, vol.10, pp.270–274, 1939.
- [5] W.R. Thurlow and P.S. Runge, "Effect of induced head movement in localization of direction of sound," *J. Acoust. Soc. Am.*, vol.42, pp.480–488, 1967.
- [6] J. Kawaura, Y. Suzuki, F. Asano, and T. Sone, "Sound localization in headphone reproduction by simulating transfer functions from the sound source to the external ear," *J. Acoust. Soc. Jpn.*, vol.45, pp.756–766, 1989.
- [7] J. Kawaura, Y. Suzuki, F. Asano, and T. Sone, "Sound localization in headphone reproduction by simulating transfer functions from the sound source to the external ear," *J. Acoust. Soc. Jpn. (E)*, vol.12, pp.203–216, 1991.
- [8] Y. Iwaya, Y. Suzuki, and D. Kimura, "Effects of head movement on front-back error in sound localization," *Acoust. Sci. & Tech.*, vol.24, pp.322–324, 2003.
- [9] I. Toshima, H. Uematsu, and T. Hirahara, "A steerable dummy head that tracks three-dimensional head movement," *Acoust. Sci. & Tech.*, vol.24, pp.327–329, 2003.
- [10] I. Toshima, S. Aoki, and T. Hirahara, "Sound localization using an auditory telepresence robot: TeleHead II," *Presence*, vol.17, pp.392–404, 2008.
- [11] V.R. Algazi, R.O. Duda, and D.M. Thompson, "Motion-tracked binaural sound," *J. Audio Eng. Soc.*, vol.52, pp.1142–1156, 2004.
- [12] 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*, vol.6225, 2004.
- [13] D.H. Cooper and T. Shige, "Discrete-matrix multichannel stereo," *J. Audio Eng. Soc.*, vol.20, pp.346–360, 1972.
- [14] 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*, vol.16, pp.436–453, 1999.
- [15] M. Noisternig, A. Sontacchi, T. Musil, and R. Höldrich, "A 3D ambisonic based binaural sound reproduction system," *AES 24th International Conference: Multichannel Audio, The New Reality*, <http://www.aes.org/e-lib/browse.cfm?elib=12314>, 2003.
- [16] L.S. Davis, R. Duraiswami, E. Grassi, N.A. Gumerov, Z. Li, and D.N. Zotkin, "High order spatial audio capture and its binaural head-tracked playback over headphones with HRTF cues," *119th Convention of the Audio Engineering Society*, <http://www.aes.org/e-lib/browse.cfm?elib=13369>, 2005.
- [17] S.U. Pillai, *Array signal processing*, Springer-Verlag, New York, 1989.
- [18] R.O. Duda and W.L. Martens, "Range dependence of the response of a spherical head model," *J. Acoust. Soc. Am.*, vol.104, pp.3048–3058, 1998.
- [19] M. Otani and S. Ise, "Fast calculation system specialized for head-related transfer function based on boundary element method," *J. Acoust. Soc. Am.*, vol.119, pp.2589–2598, 2006.
- [20] R. Sadourny, A. Arakawa, and Y. Mintz, "Integration of the nondivergent barotropic vorticity equation with an icosahedral-hexagonal grid for the sphere," *Monthly Weather Review*, vol.96, pp.351–356, 1968.
- [21] E.G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography*, London, Academic Press, 1999.
- [22] M. Pollow, K.V. Nguyen, O. Warusfel, T. Carpentier, M. Müller-Trapet, M. Vorländer, and M. Noisternig, "Calculation of head-related transfer functions for arbitrary field points using spherical harmonics," *Acta Acustica united with Acustica*, vol.98, pp.72–82, 2012.



Shuichi Sakamoto received B.S., M.S. and Ph.D. degrees from Tohoku University, in 1995, 1997, and 2004, respectively. He is currently an Associate Professor at the Research Institute of Electrical Communication, Tohoku University. He was a Visiting Researcher at McGill University, Montreal, Canada during 2007–2008. His research interests include human multi-sensory information processing including hearing, speech perception, and development of high-definition 3D audio recording systems. He is a member of ASJ, IEICE, VRSJ, and others.



Satoshi Hongo received B.E., M.E. and Dr. Eng. degrees from Tohoku University in 1990, 1992, and 1995, respectively. From 1995 to 1999, he was a Research Associate with the Faculty of Engineering, Tohoku University. From 1999 to 2008 he was an Associate Professor of Department of Design and Computer Applications, Miyagi National College of Technology (SNCT), Japan. He has been a Professor of the Advanced Course, Sendai National College of Technology and Chairman of the Department of Design and Computer Applications, SNCT.



Yôiti Suzuki graduated from Tohoku University in 1976 and received his Ph.D. degree in electrical and communication engineering in 1981. He is currently a Professor of Research Institute of Electrical Communication, Tohoku University. His research interests include psychoacoustics and digital signal processing of acoustic signals. He served as President of the Acoustical Society of Japan from '05 to '07. He is a Fellow of the Acoustical Society of America.