Prediction method for compression of spherical microphone array signals using geometric information

(Invited Paper)

Shuichi Sakamoto*, Arif Wicaksono*, Jorge Treviño*, Cesar Salvador* and Yôiti Suzuki*
*Research Institute of Electrical Communication and Graduate School of Information Sciences, Tohoku University, 17-131 Katahira, Aoba-ku, Sendai, 980-8577, Japan,
Email: {saka@ais., arifhw@ais., jorge@ais., salvador@ais., yoh@}riec.tohoku.ac.jp

***Abstract*—Previously, we proposed a method to achieve high-precision measurement systems to record 3D sound-space information. This enables the transmission of spatial sound to distant places, and enables its preservation [1], [2]. Our method, named Symmetrical object with ENchased ZIllion microphones (SENZI), was implemented using a spherical microphone array with 252 microphones. It was applied to the recording of 3D sound-space information. The microphone positions follow an icosahedral pattern. Reproducing a 3D sound space recorded with the SENZI implementation requires transmission of all the microphone signals. However, the necessarily large number of channels produce vast amounts of data. Therefore, it is important to compress these data without markedly reducing the accuracy of the reproduced 3D sound-space. In this paper, we propose a multi-channel sound signal compression technique for recordings done with a SENZI microphone array. Inter-channel correlation in the SENZI system is extremely high because the microphones are arranged densely on the sphere. Our proposed method exploits this correlation to predict some microphone signals from those recorded by microphones that are aligned with the vertices of the underlying icosahedron. The possibility of recovering some microphone signals from those of their neighbors is verified through computer simulations of a SENZI array.**

***Keywords*-sound field recording, spherical microphone array, prediction, compression, head-related transfer function (HRTF), icosahedral symmetry**

I. Introduction

Accurate 3D sound-space information is important for the design of highly realistic multimedia communications systems. Such auditory information enhances a user's sense of presence perceived from multimedia contents presented via the systems. Therefore, a technique for sensing and reproduction of accurate 3D sound-space information is a key technology to enrich various multimedia contents.

Although studies of sensing technologies are far fewer than those of reproduction technologies, some advanced 3D sound-space sensing methods have been proposed [3], [4], [5], [1], [2]. We previously proposed a method to realize high-precision sensors of 3D sound-space information for transmission to distant places and for the preservation of sound data for the future [1], [2]. The key hardware component of this method is a microphone array on a human-head-sized rigid sphere with numerous microphones on its surface. The method based on spherical microphone array is named Symmetrical object with ENchased ZIllion microphones (SENZI). The accuracy of synthesized 3D sound-space information is strongly related to the number of the microphones. In the actual SENZI system, 252 ch microphones were arranged densely on the rigid sphere.

To reproduce 3D sound-space recorded by this SENZI system to distant places, 252 ch signals must be transmitted to the listener via a network. However, because the amount of data obtained from microphone arrays is high because of the large number of channels, the processing of large amounts of data, especially data transmission for communication, entails considerable costs. Therefore, it is important to predict the recorded signals with small residual energy to reduce the redundancy of multi-channel signals.

In this paper, we proposed an efficient information prediction method that can deal with the characteristics of spherical microphone array signals. Based on the knowledge of geometrical information of the arrangement of the microphones, the relation between signals recorded by neighboring microphones was analyzed. Then, using the results of the analysis, an algorithm to predict the remaining signals solely from a subset of reference microphones was investigated.

II. Developed SENZI microphone array [1], [2]

In this section, a spherical microphone array developed for SENZI system is presented. Figure 1 shows the implemented spherical microphone array. The array's radius is 0.085 m. This size is determined according to the average human head size. On the spherical object, 252 microphones are installed. The position of each microphone is calculated based on a regular icosahedron. Each surface of a regular icosahedron is divided into 25 small equilateral triangles. All vertices of these triangles are projected to the surface of the sphere. These 252 points are used as microphone positions. Figure 2 shows the positions of the 252 microphones. The interval of each microphone is almost equal to those of the others: about 0.02 m. Consequently, the limit for the array's spatial resolution [6] appears at a frequency of 8.5 kHz.

978-1-5090-0188-0/15 $31.00 © 2015 IEEE
DOI 10.1109/IIH-MSP.2015.91

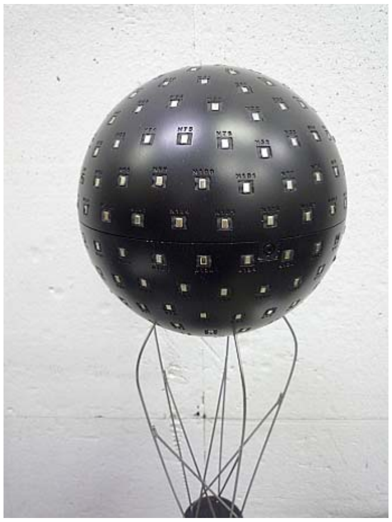


Figure 1. Photograph of actual spherical microphone array.

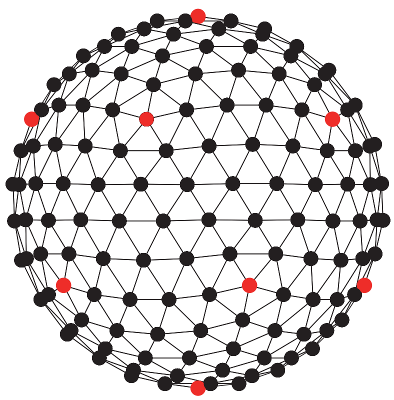


Figure 2. Microphone arrangement of spherical microphone array: red circles show vertices of a regular icosahedron.

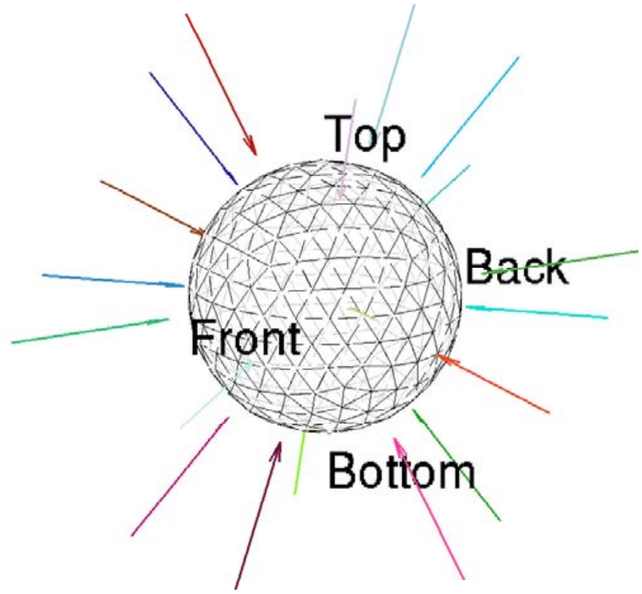


Figure 3. 20 sources located on the centroids of icosahedral faces.

III. Cross-correlation analysis of the groups of microphones based on icosahedral symmetry

As explained in the previous section, the microphones were distributed on the sphere surface by the icosahedron subdivision rule. Therefore, an analysis of the groups of microphones was conducted based on the icosahedron.

A. *Simulation condition*

As input signals, 20 sound sources were distributed around the microphone array. Each sound source was assumed as a plane wave. The positions of the sources were chosen according to the icosahedron. The angle of each sound source to the microphone array was set to the angle of incidence of the centroid of each face to the center of the sphere. Figure 3 shows the arrangement of 20 sound sources around the microphone array. From each sound position, an uncorrelated band-pass noise burst at the frequency range of 1–4 kHz was presented. The duration and interval of the burst presented from each sound source were chosen independently and randomly from 20 µs to 250 ms. Transfer functions between 20 sound sources and 252 microphones on the rigid sphere were given as a numerical calculation based on the formula of the scattering of a solid sphere [7].

To analyze the similarity between the signals recorded by each microphone in the array, cross-correlation analysis was conducted. For this study, the maximum value of the cross-correlation coefficient r_{xy} between the signals corresponding to two microphones was used as the index. This r_{xy} was calculated as follows:

$$r_{xy} = \frac{\sum_{n=0}^{L-1} x(n)y(m+n)}{\sqrt{\sum_{n=0}^{N-1} x^2(n) \sum_{n=0}^{N-1} y^2(n)}}. \quad (1)$$

In this equation, $x(n)$ and $y(n)$ respectively denote the signals recorded by two microphones. In addition, L is the length of the signals and m is the argument for the cross-correlation; a delay between the two signals being compared. These values were calculated for all microphone pairs (x, y).

B. *Results*

Figure 4 presents results of cross-correlation analysis of the groups based on an icosahedron. The microphones are positioned on the vertices of each triangle. Given the spherical symmetry of the microphone distribution, it is sufficient to perform the analysis for only one microphone.

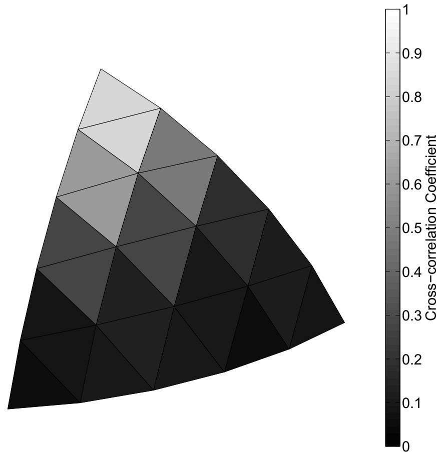


Figure 4. Cross-correlation on the group based on the icosahedron. In this plot, the reference microphone used for comparison is the microphone located at the top of the triangle.

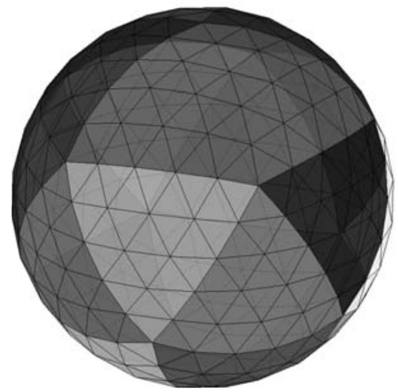


Figure 5. Grouping of microphones based on an icosahedron. One group in this figure is one triangle with the same color. Each group includes 21 microphones.

Therefore, the cross-correlation coefficients were calculated only between the signal recorded by the top microphone and those recorded by the other microphones. Averaged cross-correlation coefficients at each triangle were calculated from the coefficients of the vertices of the triangle. These averaged values were drawn on the face of the triangle in the figure.

In this figure, it might be readily apparent that in these small groups, some areas have high cross-correlation. Therefore, in these areas the redundancy is quite high as well. The microphone placement is also designed by following the icosahedron subdivision rule. It might be beneficial to group the microphones based on this arrangement.

IV. Prediction algorithm based on geometric information

The results explained in the previous section indicated that high cross-correlation can be observed around the position of the reference microphone, which means that the signals of neighboring microphones can be predicted accurately from that of the reference microphone. Selecting the reference signals to predict other signals with a small residual is important. This section introduces the method to select reference signals from the microphones on the developed spherical microphone array. Based on the results, the prediction algorithm of the recorded signals obtained using the SENZI microphone array is discussed.

A. Method of selecting the reference signals

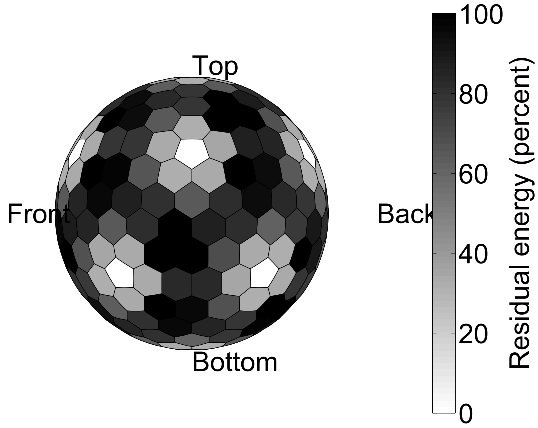
The arrangement of the microphones on SENZI microphone array is based on an icosahedron. Grouping the microphones according to the faces of the underlying icosahedron is advantageous because it results in identical groups. Therefore, the same algorithms with the same parameters are applicable to each of them. Based on this, 252 ch microphones were subdivided into 20 groups according to the face of the icosahedron. The resulting groups is presented in Fig. 5. One group in this figure is one triangle with the same color. Each group includes 21 microphones. As shown in the figure, 12 microphones on the vertices of the icosahedron belong to five groups. Therefore, these 12 microphones were selected as reference microphones and were used to predict other signals of the microphones based on the groups.

B. Prediction of the recorded signals from the references

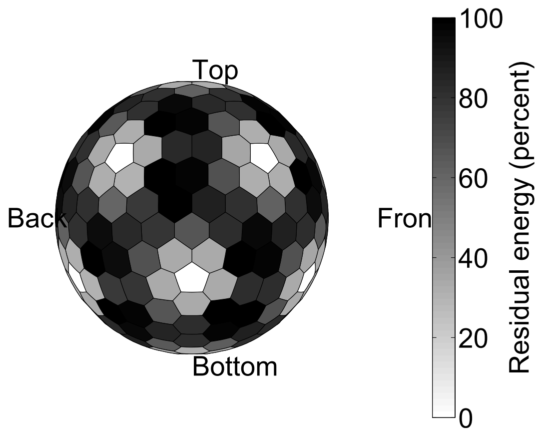
From the selected 12 reference signals, the other signals were predicted based on the signal averaging method. This technique is used widely starting from joint stereo (M/S stereo) until the recently standardized spatial audio coding [8].

The 12 reference microphone signals corresponding to the original icosahedron vertices were used as they were. Afterwards, the signals of the microphones belonging to the groups based on the icosahedron were calculated using the signals of the three closest reference microphones. These three signals were weighted and averaged to calculate the signals. The weighting coefficients for each microphone were calculated based on the angle formed by a line connecting the target microphone with the center of the array and that of three reference microphones. Therefore, the microphones' positions in the array were necessary during reconstruction.

The accuracy of the predicted signals was evaluated through computer simulation. The same sound sources in the previous section were used as stimuli. Based on the 12 reference signals, the other signals were calculated. By subtracting the original signal to the predicted signal, residual signals were calculated. Then, the energy of the residual signals



(a) Left side



(b) Right side

Figure 6. Reconstruction error attributable to application of weighted signal averaging, as described on the 20-sources bandpass-filtered noise with 1 kHz to 4 kHz frequency range.

related to the original signal was obtained. Figure 6 presents the reconstruction error attributable to applying weighted signal averaging as described on the 20-source bandpass-filtered noise with 1 kHz to 4 kHz frequency range. As the figure shows, the signals of the microphones around the reference microphone can be predicted accurately, although a high reconstruction error can be observed at the position of the center of the face, which is distant from the reference microphones.

As for future research investigating the extension of this approach, it might be worth investigating the effective subdivision of the group. According to the result of signal averaging reconstruction, the microphones that were located close to the reference microphone had low reconstruction error. Therefore, some other subdivision technique might work.

V. Conclusion

In this paper, we proposed a multi-channel sound signal compression technique recorded by SENZI microphone array. Results of cross-correlation analysis revealed that the signals of the neighboring microphones were strongly correlated. Based on this knowledge, an algorithm for prediction of all the microphone signals from some reference signals was proposed. Using geometric information, the signals of the microphones of the SENZI microphone were compressible with small residual energy from the signals of 12 reference microphones on the icosahedron vertices.

Acknowledgment

This research was partly supported by a Grant-in-Aid for Scientific Research (A) (No. 24240016) from the Ministry of Education, Culture, Sports, Science and Technology, and the A3 Foresight Program for "Ultra-realistic acoustic interactive communication on next-generation internet."

References

[1] S. Sakamoto, S. Hongo, R. Kadoi, and Y. Suzuki, "3d sound-space sensing system based on numerous symmetrically arranged microphones," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E97-A, no. 9, 2014.

[2] S. Sakamoto, S. Hongo, T. Okamoto, Y. Iwaya, and Y. Suzuki, "Improvement of accuracy of 3d sound space synthesized by real-time "SENZI", a sound space information acquisition system using spherical array with numerous microphones," in *Proc. of the 21th International Congress on Acoustics*, no. 055051, 2013, pp. 1–9.

[3] J. B. Melick, V. R. Algazi, R. O. Duda, and D. M. Thompson, "Customization for personalized rendering of motion-tracked binaural sound," in *Proc. 117th AES Convention*, no. 6225, 2004.

[4] 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," in *Audio Engineering Society Convention 119*, Oct 2005. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=13369

[5] E. Rasumow, M. Blau, M. Hansen, S. Doclo, S. van de Par, and V. Mellert, "Robustness of virtual artificial head topologies with respect to microphone positioning errors," in *Proc. Forum Acusticum 2011*, 2011, pp. 2251–2256.

[6] S. U. Pillai, *Array signal processing*. New York: Springer-Verlag, 1989.

[7] 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.

[8] J. Herre, "From joint stereo to spatial audio coding –recent progress and standardization." in *Proc. of 7th International Conference on Digital Audio Effects (DAFX04)*, 2004, pp. 157–162.