Proceedings of 20th International Congress on Acoustics, ICA 2010

23–27 August 2010, Sydney, Australia

Blind directivity estimation of a sound source in a room using a surrounding microphone array

Takuma Okamoto[(1)(2)], **Yukio Iwaya**[(1)(3)] **and Yôiti Suzuki**[(1)(3)]

(1) Research Institute of Electrical Communication, Tohoku University, Sendai, Japan
(2) Graduate School of Engineering, Tohoku University, Sendai, Japan
(3) Graduate School of Information Sciences, Tohoku University, Sendai, Japan

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

ABSTRACT

Sound sources in actual environments have no omni-directional feature, but they do have directivity in radiation. It is important to consider the directivity of a sound source when synthesizing a high-definition three-dimensional sound field. Therefore, we propose a simple but novel method to estimate the directivity of a sound source in a reverberant environment using a surrounding microphone array. Our method decomposes each observed signal of each microphone into an original sound signal; each impulse response is based on dereverberation technique. Furthermore, each impulse response is divided into two segments: an early response related to a directive feature and a late response related to reflections. The simulation results demonstrate the availability of our proposed method.

1. INTRODUCTION

A sound field in an actual environment comprises all sound objects, including their own characteristics (original sound source signal, sound position, and directivity) and a room's acoustic features (reflection and reverberation). We have been developing methods to decompose such characteristics of sound objects from recorded sound [1, 2]. If such a decomposition technique becomes possible, then we would be able to change the sound space characteristics artificially: editing of the sound field would become highly versatile. In such a situation, not only the original sound field but also an arbitrarily modified sound field could be synthesized. We designate such a system as an editable sound field system [3]. It might be important to realize such a system because previous sound field reproduction [4] efforts have been insufficient to realize modified sound fields such as those described above.

Previous studies have often treated sound sources as ideal point sources radiating sound waves in all directions equally, i.e., as having an omni-directional characteristic. Contrary to the circumstances in an actual environment, sound sources have directivity in radiation. Previous reports have described that humans can estimate a facing angle only by hearing a spoken voice [5]. Moreover, musical instruments have directivity [6]. For that reason, it is important to consider the directivity of a sound source according to a listening point when we synthesize a high-definition three-dimensional sound field. To that end, techniques used for estimating the directivity from recorded sounds and reproduction techniques including estimated directivity are required. Such techniques might be promising when used in combination with 3D image displays in which viewers can move around a 3D object. Using such a combination of visual and auditory displays, people can move around a virtual object as they are able to do in a real environment. For example, to reproduce an original sound source including the directivity of a sound source, several spherical loudspeaker arrays have been proposed [7–9]; the estimated directivity of a sound source could be applied in these arrays.

To record sound information including the directivity of a sound source and to realize an editable sound-space system, we constructed a test-bed room for sound acquisition in which a microphone array consisting of 157 microphones (Type 4951; Bruel and Kjær) is installed on all four walls and the ceiling of the room. We designate this as a surrounding microphone array. All microphones are installed 30 cm inside from all four walls and the ceiling using pipes. They are separated from each other by 50 cm. The microphone arrangement is portrayed in Fig. 1 and the appearance of the surrounding microphone array is shown in Fig. 2. We introduced a recording system for this microphone array to enable synchronous recording of 157 channels at the sampling frequency of 48 kHz with the linear PCM audio format [1].

In our previous study [1], we developed a method to estimate sound source positions accurately in a reverberant environment using this array. Furthermore, we improved the dereverberation method based on the linear-predictive multichannel equalization (LIME) algorithm [10] using a whitening filter so that the method can treat colored signals at high sampling frequency. We designated the proposed method as White-LIME [2]. Therefore, the remaining characteristic of a sound object is the directivity of a sound source.

Several studies have been conducted to measure the directivity of the sound source in an anechoic environment [11–14]. However, Nakadai *et al.* proposed a method to estimate the sound source position and the front direction of the sound source simultaneously [15]; other directions were not considered. No previous report has described estimation of the all-around directivity in a reverberant environment. Therefore, we propose a simple but novel method to estimate the directivity of a sound source in a room environment from signals recorded using a surrounding microphone array, along with information of the estimated source position and the original sound signal.

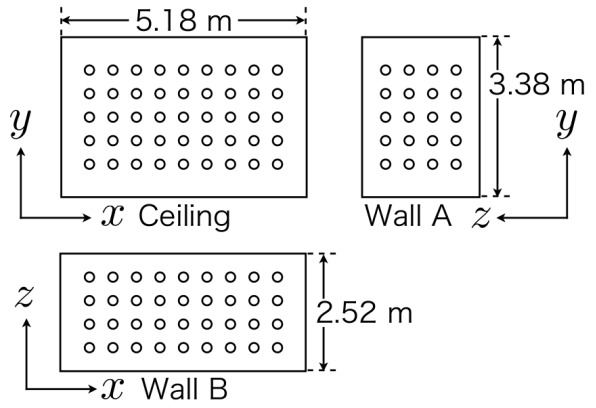


Figure 1: Arrangement of microphones in the recording room.

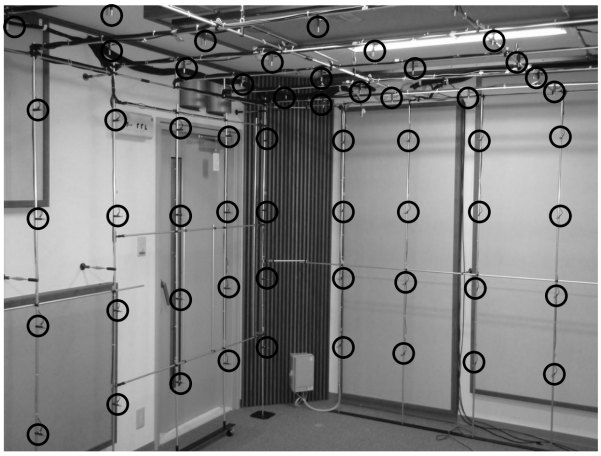


Figure 2: Appearance of the surrounding microphone array (each black circle represents a microphone).

2. PROPOSED METHOD FOR SOUND DIRECTIVITY ESTIMATION

2.1 Directivity model of a sound source in a room

The transmission of a sound wave with directivity is modeled. First, we consider the directivity of a direction θ_0 from the sound source position to a microphone of the surrounding microphone array. The original sound signal $s(n)$ is radiated with directivity $d(\theta_k,n)$, where $d(\theta_k,n)$ is the time response of directivity of a direction θ_k. The wave component radiated to direction θ_k can be described as $s(n)*d(\theta_k,n)$, where $*$ signifies the convolution. Each component is convolved with room impulse responses $h(\theta_k,n)$ and arrives at the microphone. Consequently, the output signal of the microphone $x(n)$ can be described as a summation of each component as

$$\begin{aligned} x(n) &= \sum_{k=0}^{\infty} s(n)*d(\theta_k,n)*h(\theta_k,n) \\ &= s(n)*\sum_{k=0}^{\infty}\{d(\theta_k,n)*h(\theta_k,n)\}. \end{aligned} \quad (1)$$

Here, we must estimate $d(\theta_0,n)$. The equation can be extracted as

$$\begin{aligned} x(n) &= s(n)*\{d(\theta_0,n)*h(\theta_0,n) \\ &\quad +\sum_{k=1}^{\infty} d(\theta_k,n)*h(\theta_k,n)\}. \end{aligned} \quad (2)$$

If the distance between the sound source and the microphone is $r(\theta_0)$, then Eq. 2 can be extracted as

$$\begin{aligned} x(n) &= s(n)*\left\{d(\theta_0,n)\cdot\frac{1}{r(\theta_0)}\right. \\ &\quad \left.+d(\theta_0,n)*h'(\theta_0,n)+\sum_{k=1}^{\infty} d(\theta_k,n)*h(\theta_k,n)\right\} \quad (3) \\ &= s(n)*\{h_D(n)+h_R(n)\} \quad (4) \\ &= s(n)*h(n), \quad (5) \end{aligned}$$

where $h_D(n)=d(\theta_0,n)/r(\theta_0)$, $h'(\theta_0,n)$ is the reflection component of $h(\theta_0,n)$, $h_R(n)$ is the sum of second and third term in Eq. 3, and $h(n)$ is the abbreviation of the component in Eq. 4. The system model including the directivity of a sound source is shown in Fig. 3.

2.2 Estimation of the directivity component

We can obtain the output signal $x(n)$ from each microphone directly. When the original signal $s(n)$ can be estimated, the response $h(n)$, which is usually treated as an impulse response, is obtainable with deconvolution with $s(n)$. The early response of $h(n)$ independently indicates $h_{Di}(n)$ when the length of $h_{Di}(n)$ is shorter than the time at which other reflective waves come. With the surrounding microphone array, the microphone is installed from the wall by distance d. Therefore, the minimum arrival path is not the case of oblique incidence, as in the case of head-on incidence, as depicted in Fig. 4. Therefore, the minimum arrival time interval between the direct sound and the reflected sound is $t=2d/c$, where c stands for the acoustic velocity. Then, $h_{Di}(n)$ can be extracted as the early response from the first response to time $t=2d/c$. From Eqs. 3 and 4, we must correct the amplitude of each $h_{Di}(n)$ corresponding to the difference of each distance r_i finally. Thereby, we obtain the estimated directivity $d_i(n)$ as

$$d_i(n)=r_i h_{Di}(n). \quad (6)$$

The distance between the source and each microphone can be estimated from the estimated sound source position.

2.3 Estimation of impulse responses

To apply the method described above, estimation of each impulse response between the source and each receiving point based solely on the observed signals is needed. However, such an estimation is difficult to perform because the acoustic impulse response taps were too long and the length of the response is unknown in the actual environment [10]. Therefore, we might estimate the impulse responses from the observed signals $x(n)$ using the estimated sound signal $\hat{s}(n)$ obtained using White-LIME [2].

The LIME algorithm is based on linear prediction. Therefore, if the length of the inverse filter in LIME is L and the order of the original signal $s(n)$ is U, then the estimated signal $\hat{s}(n)$ is calculated from the point at $L+1$ tap to the point at U tap of the original signal $s(n)$; the residual taps $M-1$ in $\hat{s}(n)$ are the zeros shown in Fig. 5. From this feature in LIME, the order of the longest impulse response among the observed signals was estimated.

Each impulse response $\hat{h}_i(n)$ might be estimated by system identification based on least squares [16] as

$$\hat{\mathbf{h}}_i(n)=(\mathrm{E}\{\hat{\mathbf{s}}(n)\hat{\mathbf{s}}^{\mathrm{T}}(n)\})^{-1}\mathrm{E}\left[\begin{pmatrix} x_i(n)s(n) \\ x_i(n-1)s(n-1) \\ \vdots \\ x_i(n-M+1)s(n-M+1)\end{pmatrix}\right], \quad (7)$$

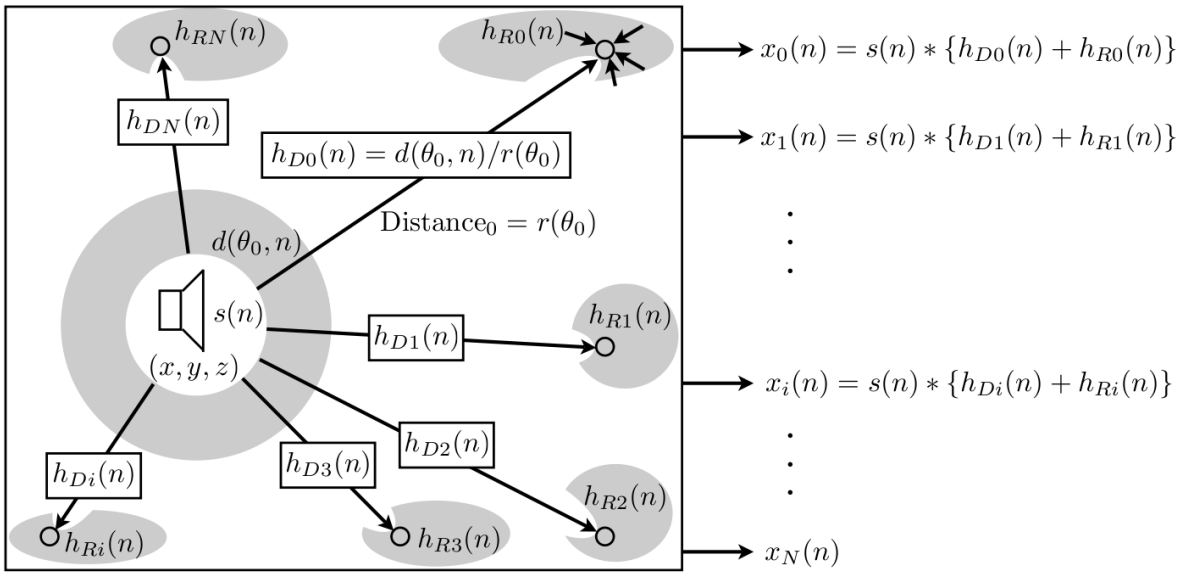


Figure 3: Single-Input Multiple-Output (SIMO) model including directivity of a sound source: $s(n)$ is the sound source signal, (x, y, z) is the sound source position, $r(\theta_i)$ is each distance between the source and each microphone i, $d(\theta_i, n) = r(\theta_i) \cdot h_{Di}(n)$ is each time response of directivity of a direction θ_i, $h_{Ri}(n)$ is each time response of reflection and $x_i(n)$ is each observed signal at microphone i.

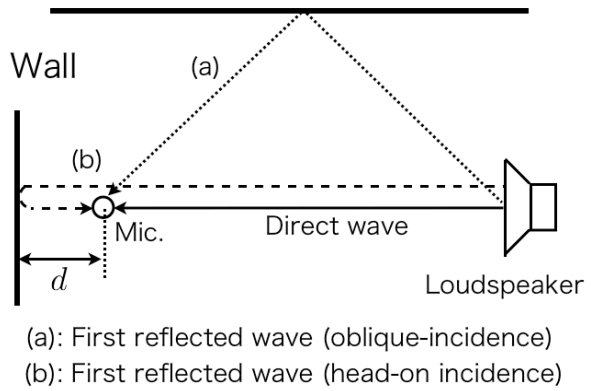


Figure 4: Propagation interval difference between a direct sound wave and the first reflected wave.

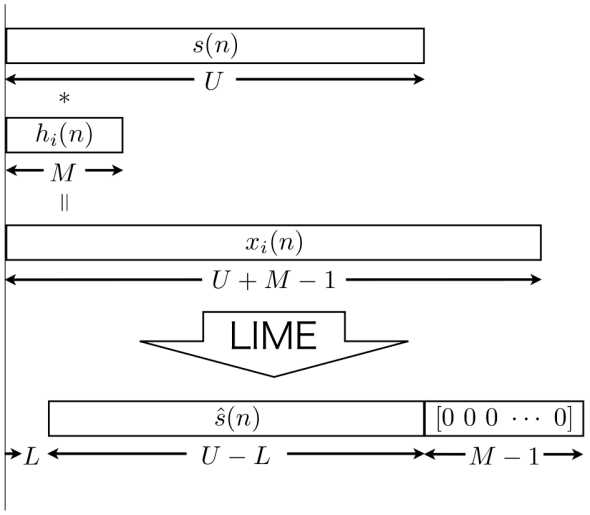


Figure 5: Signal length of estimated source signal $\hat{s}(n)$ and the observed signal $x_i(n)$ in LIME.

where $\mathrm{E}\{\cdot\}$ denotes mathematical expectation and $\hat{\mathbf{s}}(n) = [\hat{s}(n)\ \hat{s}(n-1)\ \cdots\ \hat{s}(n-M+1)]^{\mathrm{T}}\}$.

As described above, the first L taps of $\hat{s}(n)$ could not be estimated because the LIME algorithm is based on linear prediction. However, if the number of microphones M might be large, the order of inverse filter L might be short. Moreover, if the taps of the observed signal $x_i(n)$ are large, then the first L taps of $\hat{s}(n)$ are apparently a few taps compared with the total taps and inserting the zeros in the first L taps apparently yields a small error.

3. MEASUREMENTS AND SIMULATIONS

3.1 Measurements

To evaluate the effectiveness of the proposed method, a computer simulation was performed using measured impulse responses. The measurements of the impulse responses including both the directivity of the sound source and the reflected properties were recorded using the surrounding microphone array system.

We measured the impulse responses from the loudspeaker (AP-5001: Micropure Co. Ltd.) in each direction as the directivity of this loudspeaker in the anechoic room using the Time Stretched Pulse (TSP) [17]. The sampling frequency in recording was 48 kHz. The measurement system is depicted in Fig. 6. In this measurement, the front direction is 0 deg; measurements were taken from 0 deg to 180 deg with a clockwise rotation by 15 deg; the impulse responses for 13 directions were measured.

The impulse responses in the reverberant environment were measured in the room using the surrounding microphone array system. The measurement system using 157 microphone array is shown in Fig. 7. The source signal and the sampling frequency in recording were the same as those for the anechoic measurements. Figures 8(A) and 8(B) show two patterns of the arrangements of the loudspeaker and 28 microphones ($z = 1.0$ m) in the room. The reverberation time of this room

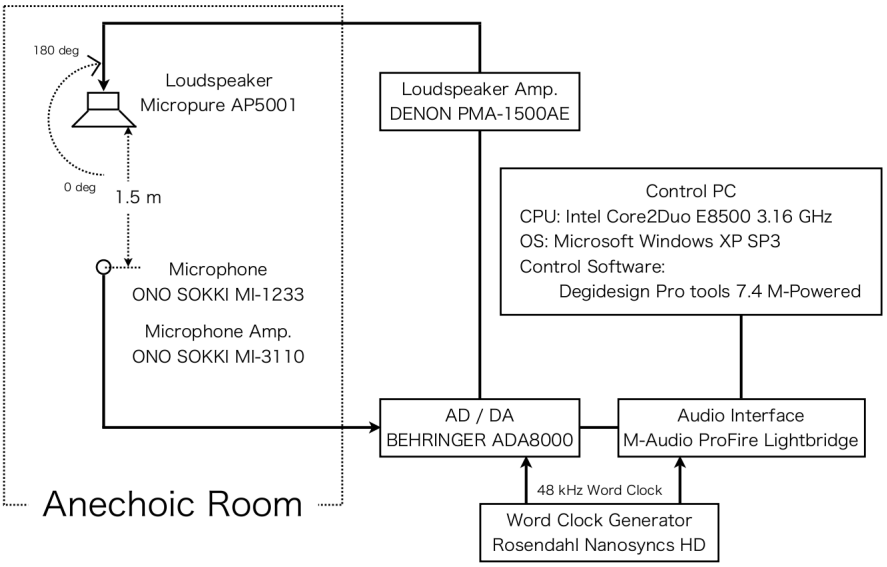


Figure 6: Measurement system in anechoic room.

was 0.15 s and the order of the room impulse responses was 7200.

3.2 Simulations

Each observed signal at each microphone $x_i(n)$ was obtained by convolving the source signal $s(n)$ to each measured impulse response $h_i(n)$. The source signal was a musical piece (2.7 s) [18]. The sampling frequency of the source signal was 44.1 kHz. Therefore, the measured impulse responses were downsampled 48 kHz to 44.1 kHz; moreover, the order of all responses was shortened from 7200 to 6615.

The estimated source signal $\hat{s}(n)$ was obtained using WhiteLIME. The score of the signal to distortion ratio (SDR) of the original signal $s(n)$ to the estimated signal $\hat{s}(n)$ was 57.3 dB. Each estimated impulse response $\hat{h}_i(n)$ was calculated from each observed signal $x_i(n)$ and the estimated source signal $\hat{s}(n)$. The average of the score of the SDR of each original response $h_i(n)$ to each estimated response $\hat{s}(n)$ was 62.6 dB. From these results, the estimated responses can be inferred accurately.

In the surrounding microphone array system, the distance between each microphone and the wall was 30 cm. Therefore, the clipping length of each estimated response was approximately $44{,}100 \times 2 \times 0.6/340 \approx 78$ taps. The clips and responses were extracted as the estimated directivity of the sound source of each direction after amplitude correction using Eq. (6). Each distance between the sound source and each microphone was estimated using an appropriate estimation method of the source position, such as RAP-MUSIC [1].

3.3 Results

The amplitude-frequency responses measured in the anechoic room, in the reverberant room, and estimated by clipping 78 taps are shown, respectively, in Figs. 10 (10-a – 10-c). Moreover, the results at 2000 Hz in the 1/3 octave-band analysis of three patterns are shown, respectively, in Figs. 11 (11-d – 11-f). The results shown in Figs. 10 and 11 demonstrate that the proposed method was able to estimate the directivity of the sound source. In particular, the dips toward 110 deg and 175 deg were estimated accurately.

To confirm the effectiveness of the proposed method, similarity based on the nearest neighbor method [19] was calculated from results of the 1/3 octave-band analysis. We defined $\boldsymbol{P}_1(\tau)$ as the pattern vector of the acoustic pressure distribution for each direction measured in the anechoic room, $\boldsymbol{P}_2(\tau)$ as that measured in the reverberant room, and $\boldsymbol{P}_3(\tau)$ as that estimated using our proposed method. Each similarity was calculated using the following equation.

$$S_{1,i} = 1 - \frac{|\boldsymbol{P}_1(\tau) - \boldsymbol{P}_i(\tau)|}{|\boldsymbol{P}_1(\tau)|} \quad (i = 2,3) \tag{8}$$

In that equation, $S_{1,2}$ signifies the similarity between the result measured in the anechoic room and the result measured in the reverberant room. In addition, $S_{1,3}$ denotes that between the result measured in the anechoic room and the estimated result. These results are presented in Figs. 9 (9-A – 9-B). From these results, it might be inferred that $S_{1,3}$ at all frequency bands were high scores, although $S_{1,3}$ at high-frequency bands were low scores. Therefore, our proposed method can estimate directivity patterns that closely resemble real ones up to around 16 kHz.

For these measurements, the sampling frequency was 44.1 kHz and the frequency resolution of the impulse response clipped 78 taps was approximately 44,100/78, or 565 Hz. Therefore, the estimation accuracy at a frequency less than 565 Hz was inferior to that at higher frequencies. The interval between each microphone and the wall might be set adequately to solve this problem.

4. FUTURE WORKS

The estimation method of directivity of a sound source was proposed and its validity was confirmed. Our method, however, presents a problem in estimating the sound source directivity when the walls and microphones are close to each other. Therefore, the part of an impulse response corresponding to the sound source and that to the reflections overlap because, in this proposed method, the early part of an impulse response corresponding to the sound source directivity is derived using very simple rectangular time-windowing. Therefore, an important subject of future work to widen the applicability of our proposed method is development of a method extracting the early part from an impulse response with overlapped reflections.

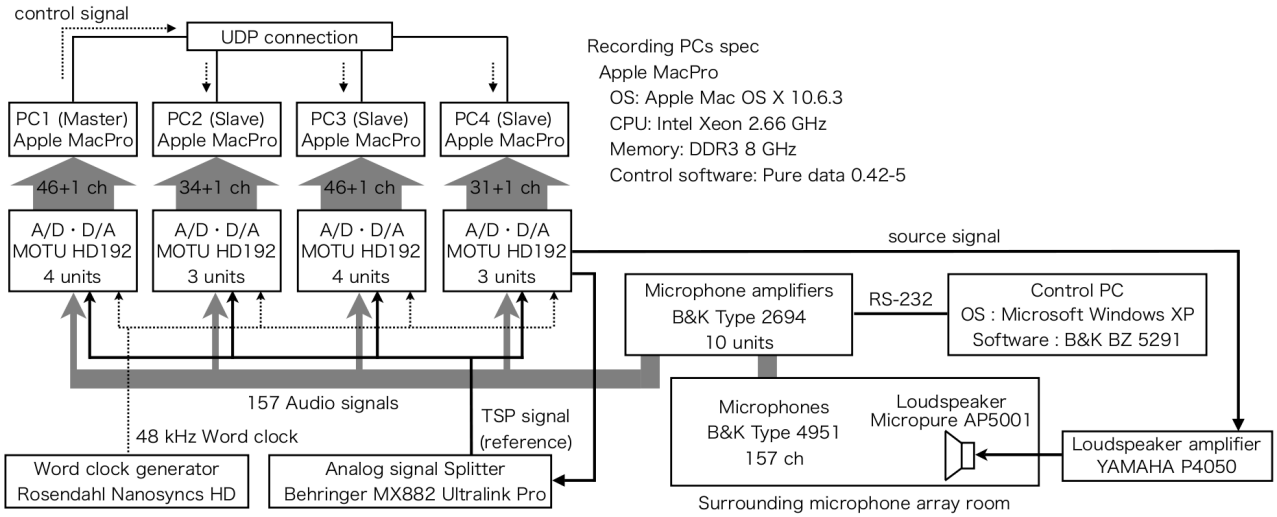


Figure 7: Measurement system in the room using the surrounding microphone array.

ACKNOWLEDGMENT

This study was partly supported by the GCOE program (CERIES) of the Graduate School of Engineering, Tohoku University and cooperative research with NTT Corp.

REFERENCES

[1] T. Okamoto, R. Nishimura and Y. Iwaya, "Estimation of sound source positions using a surrounding microphone array," *Acoust. Sci. & Tech.*, vol. 28, no. 3, pp. 181–189, May 2007.

[2] T. Okamoto, Y. Iwaya and Y. Suzuki, "New blind dereverberation method based on multichannel linear prediction using pre-whitening filter," *2009 Spring Meet. Acoust. Soc. Jpn.*, pp. 675–676, Mar. 2009. (in Japanese)

[3] T. Okamoto, Y. Iwaya and Y. Suzuki, "Toward an editable sound-space system using high-resolution sound properties," *Proc. IWPASH 2009 (eProceedings)*, no. 10.1142-9789814299312-0048, Nov. 2009.

[4] S. Ise, "A principle of sound field control based on the Kirchhoff–Helmholtz integral equation and the theory of inverse systems," *Acustica – Acta Acustica*, vol. 85, no. 1, pp. 78–87, Jan. 1999.

[5] H. Kato, H. Takemoto and R. Nishimura, "Perception of speaker's facing angle," *J. Acoust. Soc. Am.*, vol. 123, no. 5, pp. 3294, May 2008.

[6] F. Saito, M. Kasuya, T. Harima, and Y. Suzuki, "Sound power levels of musical instruments and the estimation of the influence on players' hearing ability," *Proc. Inter-Noise 2003*, pp. 2259–2266, 2003.

[7] M. Katsumoto, Y. Yamakata and T. Kimura, "Development of 3D audio display for ultra-realistic communication," *Proc. IWPASH 2009 (eProceedings)*, no. 10.1142-9789814299312-0046, Nov. 2009.

[8] F. Zotter and R. Holdrich, "Modeling a spherical loudspeaker system as a multiple source," *Proc. the 33rd German Annual Conf. Acoust.*, 2007.

[9] Boaz Rafaely, "Spherical loudspeaker array for local active control of sound, " *J. Acoust. Soc. Am.*, vol. 125, no. 5, pp. 3006–3017, May 2009.

[10] M. Delacroix, T. Hikichi and M. Miyoshi, "Precise dereverberation using multichannel linear prediction," *IEEE Trans. Audio Speech Lang. Process.*, vol. 15, no. 2, pp. 430–440, Feb. 2007.

[11] R. Jacques, B. Albrecht and H.-P. Schade, "Multichannel source directivity recording in an anechoic chamber and in a studio," *Proc. Forum Acusticum*, Aug. 2005.

[12] B. F. G. Katz and C. d'Alessandro, "Directivity measurements of the singing voice," *Proc. ICA 2007*, Sep. 2007.

[13] M. Noisternig and B. F. G. Katz, "Reconstructing sound source directivity in virtual acoustic environments," *Proc. IWPASH 2009 (eProceedings)*, no. 10.1142-9789814299312-0020, Nov. 2009.

[14] D. Devoy and F. Zotter, "Acoustic center and orientation analysis of sound-radiation recording with a surrounding spherical microphone array," *Proc. the 2nd Int. Symp. Ambisonics and Spherical Acoust.*, May 2010.

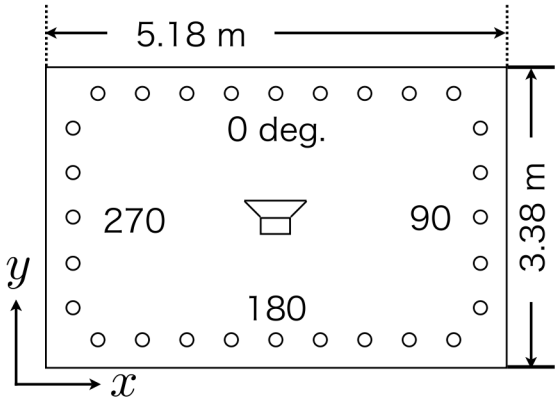
[15] K. Nakadai, H. Nakajima, K. Yamada, Y. Hasegawa, T. Nakamura, and H. Tsujino, "Sound source tracking with directivity pattern estimation using a 64 ch microphone array," *IROS 2005*, pp. 1690–1696, Aug. 2005.

[16] L. Ljung, *System Identification: Theory for the User*, Prentice Hall, Englewood Cliffs, MJ, 1987.

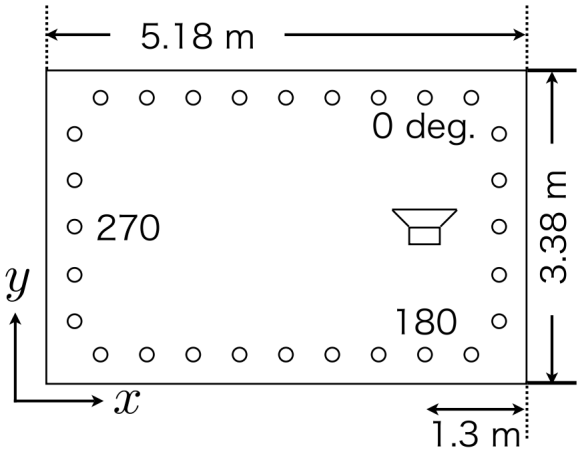
[17] Y. Suzuki, F. Asano, H.-Y. Kim, and T. Sone, "An optimum computer-generated pulse signal suitable for the measurement of very long impulse responses," *J. Acoust. Soc. Am.*, vol. 97, no. 2, pp. 1119–1123, Feb. 1995.

[18] http://staff.aist.go.jp/m.goto/RWC-MDB/index.html

[19] P. J. Clark and F. C. Evans, "Distance to nearest neighbor as a measure of spatial relationships in populations," *Ecology*, vol. 35, no. 4, pp. 445–453, Oct. 1954.

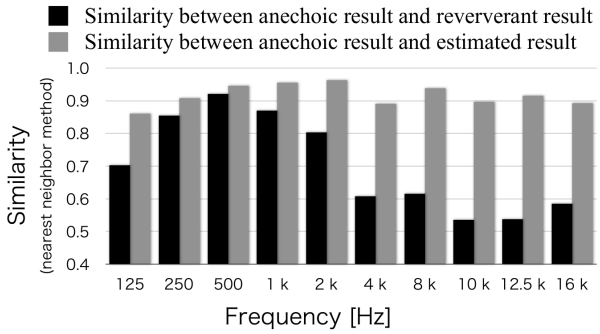


(A) loudspeaker position: ($x = 2.59$, $y = 1.69$, $z = 1.10$)

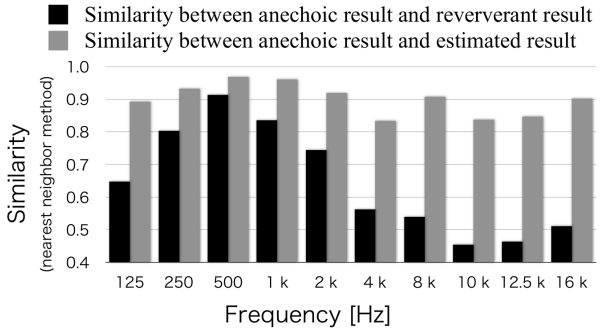


(B) loudspeaker position: ($x = 3.88$, $y = 1.69$, $z = 1.10$)

Figure 8: Two patterns of arrangement of the loudspeaker and 28 microphones.

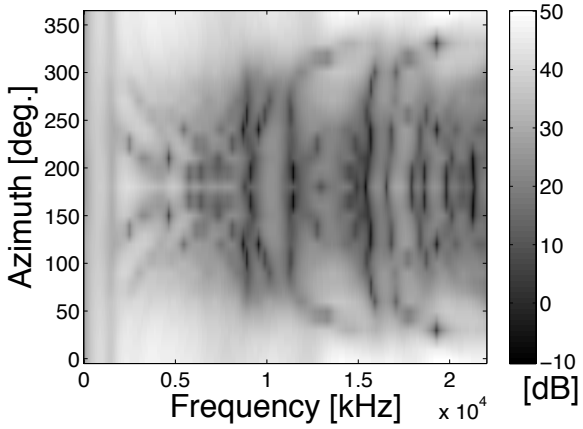


(A) loudspeaker position: ($x = 2.59$, $y = 1.69$, $z = 1.10$)

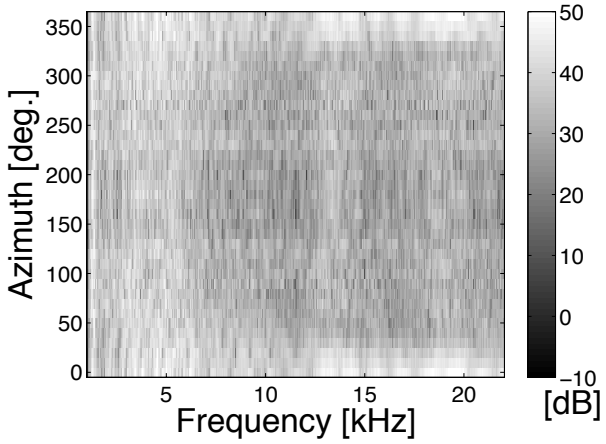


(B) loudspeaker position: ($x = 3.88$, $y = 1.69$, $z = 1.10$)

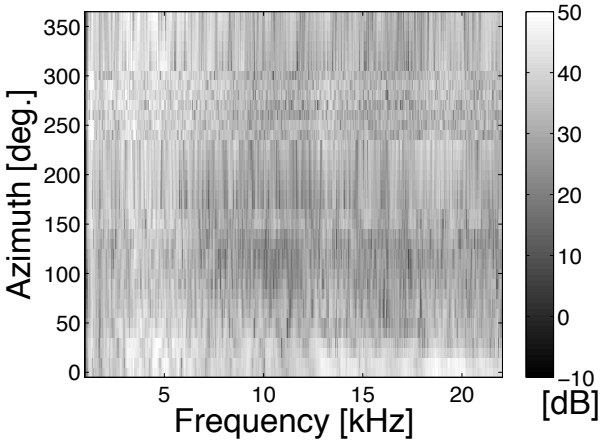
Figure 9: Results of similarity analysis (nearest neighbor method) of the reverberant response and the response extracted using the proposed method.



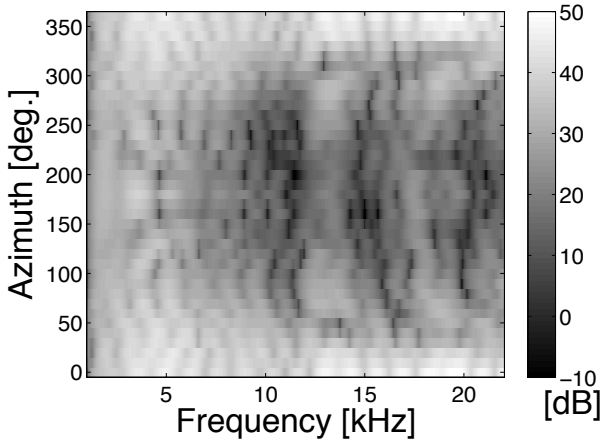
(a) Measured in an anechoic room



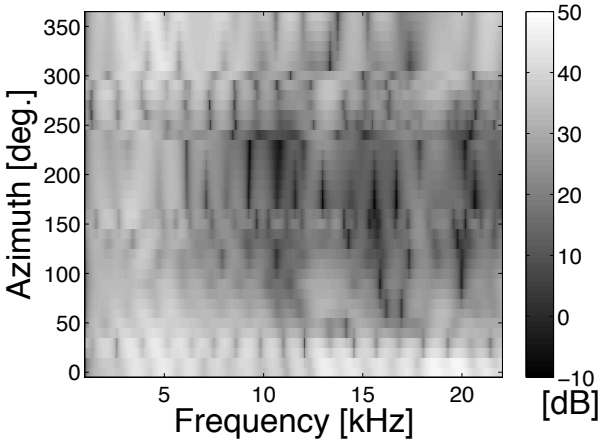
(b-A) Measured in an actual room of pattern A



(b-B) Measured in an actual room of pattern B

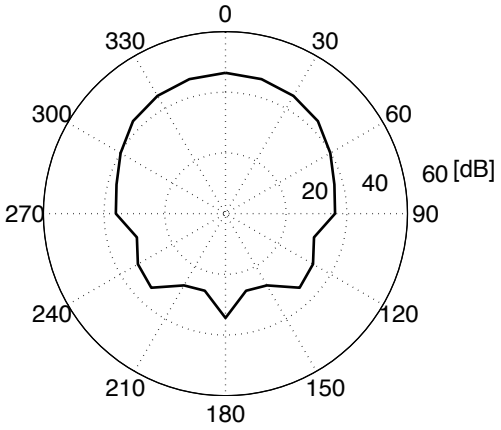


(c-A) Proposed method of pattern A

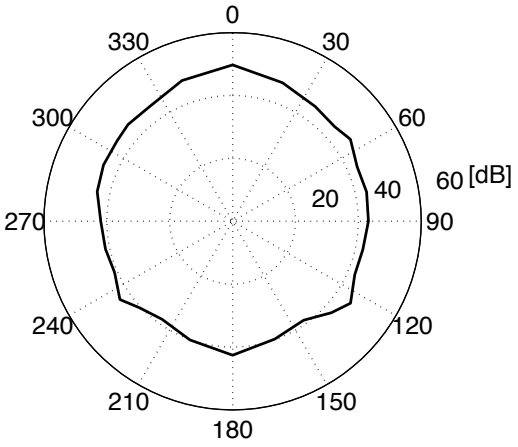


(c-B) Proposed method of pattern B

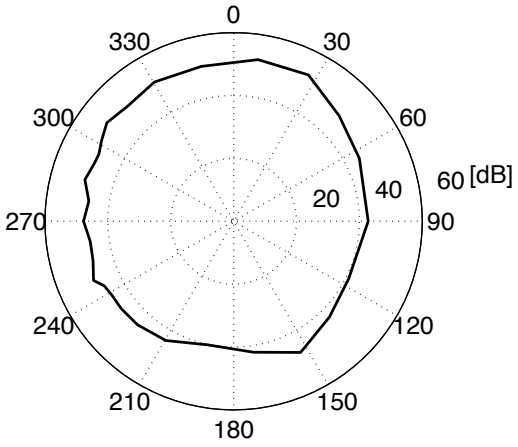
Figure 10: Amplitude—frequency characteristics.



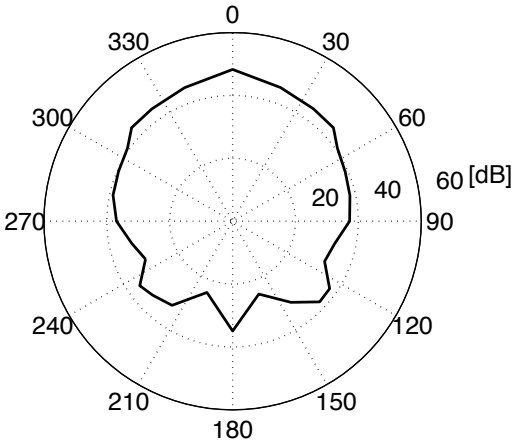
(a) Measured in an anechoic room



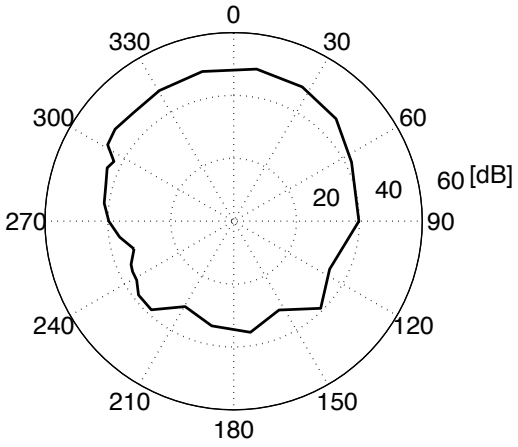
(b-A) Measured in an actual room of pattern A



(b-B) Measured in an actual room of pattern B



(c-A) Proposed method of pattern A



(c-B) Proposed method of pattern B

Figure 11: Measured and estimated directivity at 2000 Hz in the 1/3 octave-band.