MICROPHONE ARRAY FOR INCREASING MUTUAL INFORMATION BETWEEN SOUND SOURCES AND OBSERVATION SIGNALS

Kenta Niwa, Tatsuya Kako, and Kazunori Kobayashi

NTT Media Intelligence Laboratories, Tokyo, Japan

ABSTRACT

We investigated the basic principle of how spatial signals should be captured with a microphone array to estimate each source signal and its practical implementation. Most conventional studies on array signal processing have been focused on the design of beamforming and Wiener filters. To achieve further effective noise reduction, designing an optimum array structure to segregate a target from other noises is necessary. We found the optimum structure of the spatial correlation matrix to estimate each source signal. This is achieved by receiving signals whose eigenvalues of the spatial correlation matrix are homogenized. To homogenize the eigenvalues of the spatial correlation matrix while maintaining a short impulse response length, we propose an array structure composed of parabolic reflectors and 96 microphones. Through experiments using the proposed array structure, we confirmed that the eigenvalues of the spatial correlation matrix was asymptotically homogenized and that sharp directivity could be formed.

Index Terms— Microphone array, transfer function, parabolic reflector, multiple input multiple output (MIMO), beamforming

1. INTRODUCTION

Microphone array signal processing techniques [1, 2] have been studied to emphasize a target source in noisy environments. Most research on microphone arrays have been focused on the reception of the target sound source within a range of a few meters from the array. However, there are situations in which we would like to zoom in on a target source placed in a very remote position, such as the voice of an athlete on a stadium playing field. To distinguish between a target and other noises in remote places, similar to a camera zooming in on a target object, obtaining effective cues for segregating them from array observations is necessary. Therefore, the goal of this study is to investigate how spatial signals should be captured with a microphone array and its practical implementation to estimate a target source without both long latency in audio output and sound quality degradation.

Most conventional studies on microphone arrays have been mainly focused on how to design beamforming filters or a nonlinear Wiener filter. The delay-and-sum method and minimum variance distortionless response (MVDR) method [3] are commonly used for designing beamforming filters. Applying a Wiener filter to the beamforming output is effective in boosting noise reduction performance [4]–[10]. On the other hand, several researchers have focused on array structures such as linear and minimum redundancy [1]. A rigid spherical microphone array [11]–[14] has been studied to point the beam in an arbitrary direction. Placing microphones on a rigid spherical surface prevents instability filtering at certain frequencies. In our previous work, we found that the noise power in the beamforming output could be minimized by observing signals in a diffused acoustic field (diffused sensing [15, 16]). The effective cues for segregating a target from other noise sources could be obtained with diffused sensing. However, since the impulse response length was long, the audio output latency increased, and the output sound quality degradation, caused by instable filters, occurred.

For this study, we investigated how spatial signals should be captured with microphones to estimate sound sources and its practical implementation. We believe that the optimum property of array observations is determined independent of sound enhancement processing, e.g., beamforming and Wiener filtering. To measure how much array observations tell us about source signals, mutual information between multiple input and multiple output (MIMO) [17, 18] is used. If the mutual information of MIMO is increased over broad frequencies, effective cues for segregating source signals can be obtained. To avoid instable filtering, it is also important to shorten the impulse response length so as not to generate common zero points. As an implementation to satisfy the above requirements, we propose an array structure in which microphones are positioned in front of parabolic reflectors. If microphones are arranged optimally, the mutual information of MIMO can be increased. Then, the impulse response length would be short since the parabolic reflector is a simple shape.

This paper is organized as follows. In Sec. 2, the basic property of observation signals is defined. In Sec. 3, how spatial signals should be captured with a microphone array to estimate sound sources is discussed. In Sec. 4, its practical array structure using parabolic reflectors is proposed. After investigating the performances of the proposed array structure in Sec. 5, the paper is concluded in Sec. 6.

2. OBSERVATION MODEL

Let us assume that K source signals are observed using M microphones. The k-th source signal is represented by $S_k(\omega, \tau)$, where ω and τ denote the index of frequency and frame, respectively. The background noise received at the m-th microphone is described by $N_m(\omega, \tau)$, where $N_m(\omega, \tau)$ is a complex Gaussian distribution. The mean and variance of $S_k(\omega, \tau)$ and $N_m(\omega, \tau)$ are given by

$$\langle S_k(\omega, \tau)\rangle = 0, \tag{1}$$

$$\langle N_m(\omega, \tau)\rangle = 0, \tag{2}$$

$$\left\langle (S_k(\omega, \tau) - \langle S_k(\omega, \tau)\rangle)^2 \right\rangle = \sigma_{\mathrm{S}}^2(\omega), \tag{3}$$

$$\left\langle (N_m(\omega, \tau) - \langle N_m(\omega, \tau)\rangle)^2 \right\rangle = \sigma_{\mathrm{N}}^2(\omega), \tag{4}$$

where $\langle\cdot\rangle$ denotes the expectation operator. The $S_k(\omega, \tau)$ and $N_m(\omega, \tau)$ are assumed to be uncorrelated as

$$\langle S_k(\omega, \tau) S_{k'}^*(\omega, \tau)\rangle = 0 \quad (k \neq k'), \tag{5}$$

$$\langle N_m(\omega, \tau) N_{m'}^*(\omega, \tau)\rangle = 0 \quad (m \neq m'), \tag{6}$$

$$\langle S_k(\omega, \tau) N_m^*(\omega, \tau)\rangle = 0, \tag{7}$$

where * denotes a complex conjugate. The variance-covariance matrix of source signals and background noise are modeled by

$$\mathbf{R}_{\mathrm{S}}(\omega) = \left\langle \mathbf{s}(\omega,\tau)\mathbf{s}^{\mathrm{H}}(\omega,\tau) \right\rangle = \sigma_{\mathrm{S}}^2(\omega)\mathbf{I}_K, \tag{8}$$

$$\mathbf{R}_{\mathrm{N}}(\omega) = \left\langle \mathbf{n}(\omega,\tau)\mathbf{n}^{\mathrm{H}}(\omega,\tau) \right\rangle = \sigma_{\mathrm{N}}^2(\omega)\mathbf{I}_M, \tag{9}$$

where $^{\mathrm{H}}$ denotes the Hermitian conjugate.

When the transfer function between the k-th source and m-th microphone is denoted by $A_{m,k}(\omega)$, the observed signals $\mathbf{x}(\omega,\tau)$ are given by

$$\mathbf{x}(\omega,\tau) = \mathbf{A}(\omega)\mathbf{s}(\omega,\tau) + \mathbf{n}(\omega,\tau), \tag{10}$$

where

$$\mathbf{x}(\omega,\tau) = [X_1(\omega,\tau),\ldots,X_M(\omega,\tau)]^{\mathrm{T}}, \tag{11}$$
$$\mathbf{A}(\omega) = [\mathbf{a}_1(\omega),\ldots,\mathbf{a}_K(\omega)], \tag{12}$$
$$\mathbf{a}_k(\omega) = [A_{1,k}(\omega),\ldots,A_{M,k}(\omega)]^{\mathrm{T}}, \tag{13}$$
$$\mathbf{s}(\omega,\tau) = [S_1(\omega,\tau),\ldots,S_K(\omega,\tau)]^{\mathrm{T}}, \tag{14}$$
$$\mathbf{n}(\omega,\tau) = [N_1(\omega,\tau),\ldots,N_M(\omega,\tau)]^{\mathrm{T}}, \tag{15}$$

and $^{\mathrm{T}}$ denotes the transposition.

To investigate the relationships between array observations, the spatial correlation matrix $\mathbf{R}_{\mathrm{X}}(\omega)$ [1, 2] is defined by

$$\begin{aligned}\mathbf{R}_{\mathrm{X}}(\omega) &= \left\langle \mathbf{x}(\omega,\tau)\mathbf{x}^{\mathrm{H}}(\omega,\tau) \right\rangle \\ &= \mathbf{A}(\omega)\left\langle \mathbf{s}(\omega,\tau)\mathbf{s}^{\mathrm{H}}(\omega,\tau) \right\rangle \mathbf{A}^{\mathrm{H}}(\omega) + \left\langle \mathbf{n}(\omega,\tau)\mathbf{n}^{\mathrm{H}}(\omega,\tau) \right\rangle \\ &= \mathbf{A}(\omega)\mathbf{R}_{\mathrm{S}}(\omega)\mathbf{A}^{\mathrm{H}}(\omega) + \mathbf{R}_{\mathrm{N}}(\omega). \end{aligned} \tag{16}$$

Assuming that the received power at each microphone is normalized to $\sigma_{\mathrm{A}}^2(\omega)$, $\mathbf{R}_{\mathrm{X}}(\omega)$ is rewritten by

$$\mathbf{R}_{\mathrm{X}}(\omega) = \sigma_{\mathrm{S}}^2(\omega)\mathbf{R}_{\mathrm{A}}(\omega) + \sigma_{\mathrm{N}}^2(\omega)\mathbf{I}_M, \tag{17}$$

where $\mathbf{R}_{\mathrm{A}}(\omega)$ is composed of the received power and cross-correlation between microphones $\Gamma_{i,j}(\omega)$ as

$$\begin{aligned}\mathbf{R}_{\mathrm{A}}(\omega) &= \mathbf{A}(\omega)\mathbf{A}^{\mathrm{H}}(\omega) \\ &= \begin{bmatrix} \sigma_{\mathrm{A}}^2(\omega) & \Gamma_{1,2}(\omega) & \cdots & \Gamma_{1,M}(\omega) \\ \Gamma_{2,1}(\omega) & \sigma_{\mathrm{A}}^2(\omega) & \cdots & \Gamma_{2,M}(\omega) \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma_{M,1}(\omega) & \Gamma_{M,2}(\omega) & \cdots & \sigma_{\mathrm{A}}^2(\omega) \end{bmatrix}, \end{aligned} \tag{18}$$

$$\Gamma_{i,j}(\omega) = \sum_{k=1}^{K} A_{i,k}(\omega)A_{j,k}^*(\omega). \tag{19}$$

3. BASIC PRINCIPLE OF ARRAY OBSERVATIONS TO MAXIMIZE MUTUAL INFORMATION OF MIMO

We investigated the optimum structure of the spatial correlation matrix to estimate source signals. As a quantity to measure how much $\mathbf{x}(\omega,\tau)$ tells us about $\mathbf{s}(\omega,\tau)$, the mutual information of MIMO is defined as

$$I(\mathbf{s},\mathbf{x}) = \iint q_{\mathrm{xs}}(\mathbf{x},\mathbf{s})\log_2\frac{q_{\mathrm{xs}}(\mathbf{x},\mathbf{s})}{q_{\mathrm{x}}(\mathbf{x})q_{\mathrm{s}}(\mathbf{s})}dXdS, \tag{20}$$

where $q_{\mathrm{xs}}(\mathbf{x},\mathbf{s})$, $q_{\mathrm{x}}(\mathbf{x})$, and $q_{\mathrm{s}}(\mathbf{s})$ denote the joint probability density function (PDF) of $\mathbf{x}(\omega,\tau)$ and $\mathbf{s}(\omega,\tau)$, $\mathbf{x}(\omega,\tau)$, and $\mathbf{s}(\omega,\tau)$, respectively. If $I(\mathbf{s},\mathbf{x})$ is increased, effective cues to segregate sound sources can be obtained.

When K is a large number, $\mathbf{x}(\omega,\tau)$ is assumed to be a complex Gaussian distribution. Then, the maximum value of $I(\mathbf{s},\mathbf{x})$, which is called the channel capacity $C(\omega)$ [17, 18], is given by

$$\begin{aligned}C(\omega) &= \max\{I(\mathbf{s},\mathbf{x})\} \\ &= \log_2 \det\left(\mathbf{R}_{\mathrm{N}}^{-1}(\omega)\mathbf{R}_{\mathrm{X}}(\omega)\right) \\ &= \log_2 \det\left(\mathbf{R}_{\mathrm{N}}^{-1}(\omega)\mathbf{A}(\omega)\mathbf{R}_{\mathrm{S}}(\omega)\mathbf{A}^{\mathrm{H}}(\omega) + \mathbf{I}_M\right) \\ &= \log_2 \det\left(\sigma_{\mathrm{SN}}^2(\omega)\mathbf{R}_{\mathrm{A}}(\omega) + \mathbf{I}_M\right), \end{aligned} \tag{21}$$

where

$$\sigma_{\mathrm{SN}}^2(\omega) = \sigma_{\mathrm{S}}^2(\omega)/\sigma_{\mathrm{N}}^2(\omega). \tag{22}$$

To investigate the relationship between $C(\omega)$ and the structure of the spatial correlation matrix, eigenvalue decomposition is applied to $\mathbf{R}_{\mathrm{A}}(\omega)$ as

$$\mathbf{R}_{\mathrm{A}}(\omega) = \mathbf{V}(\omega)\mathbf{\Lambda}(\omega)\mathbf{V}^{\mathrm{H}}(\omega), \tag{23}$$

where

$$\mathbf{V}(\omega) = [\mathbf{v}_1(\omega),\ldots,\mathbf{v}_M(\omega)], \tag{24}$$
$$\mathbf{v}_m(\omega) = [V_{m,1}(\omega),\ldots,V_{m,M}(\omega)]^{\mathrm{T}}, \tag{25}$$
$$\mathbf{\Lambda}(\omega) = \mathrm{diag}\{[\Lambda_1(\omega),\ldots,\Lambda_M(\omega)]\}. \tag{26}$$

The eigenvalues are sorted in descending order as $\Lambda_1(\omega) \geq,\ldots,\geq \Lambda_M(\omega) \geq 0$. By substituting Eq. (23) into Eq. (21), $C(\omega)$ is rewritten as

$$\begin{aligned}C(\omega) &= \log_2 \det\left(\mathbf{V}(\omega)\left(\sigma_{\mathrm{SN}}^2(\omega)\mathbf{\Lambda}(\omega) + \mathbf{I}_M\right)\mathbf{V}^{\mathrm{H}}(\omega)\right) \\ &= \log_2 \det\left(\sigma_{\mathrm{SN}}^2(\omega)\mathbf{\Lambda}(\omega) + \mathbf{I}_M\right) \\ &= \log_2 \prod_{m=1}^{M}\left(\sigma_{\mathrm{SN}}^2(\omega)\Lambda_m(\omega) + 1\right). \end{aligned} \tag{27}$$

From the relationships between arithmetic and geometric means, the upper limit of the terms included in Eq. (27) is determined as

$$\begin{aligned}\sqrt[M]{\prod_{m=1}^{M}(\sigma_{\mathrm{SN}}^2(\omega)\Lambda_m(\omega)+1)} &\leq \frac{1}{M}\sum_{m=1}^{M}\left(\sigma_{\mathrm{SN}}^2(\omega)\Lambda_m(\omega)+1\right) \\ &= \sigma_{\mathrm{SN}}^2(\omega)\sigma_{\mathrm{A}}^2(\omega) + 1. \end{aligned} \tag{28}$$

In Eq. (28), the equality is satisfied if and only if all the eigenvalues are homogenized as

$$\Lambda_1(\omega) =,\ldots,= \Lambda_M(\omega). \tag{29}$$

If the signal observation to satisfy Eq. (29) is achieved, $C(\omega)$ is maximized as

$$C_{\mathrm{MAX}}(\omega) = M\log_2\left(\sigma_{\mathrm{SN}}^2(\omega)\sigma_{\mathrm{A}}^2(\omega) + 1\right). \tag{30}$$

To homogenize the eigenvalues as in Eq. (29), the array observations should be decorrelated as

$$\lim_{\Gamma_{i,j}(\omega)\to 0}\mathbf{R}_{\mathrm{A}}(\omega) \to \begin{bmatrix} \sigma_{\mathrm{A}}^2(\omega) & 0 & \cdots & 0 \\ 0 & \sigma_{\mathrm{A}}^2(\omega) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{\mathrm{A}}^2(\omega) \end{bmatrix} = \sigma_{\mathrm{A}}^2(\omega)\mathbf{I}_M. \tag{31}$$

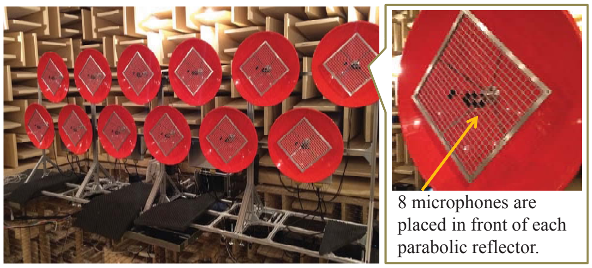

Fig. 1. Proposed array structure using parabolic reflectors (4.0 m (W) × 1.5 m (H) × 1.0 m (D))

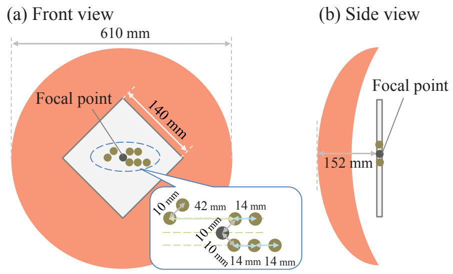

Fig. 2. Microphone arrangement of each parabolic reflector

Then, the eigenvalues are homogenized as $\Lambda_1(\omega) =, \ldots, = \Lambda_M(\omega) = \sigma_A^2(\omega)$. By receiving signals to homogenize the eigenvalues of the spatial correlation matrix, the mutual information of MIMO is increased and effective cues for segregating sound sources can be obtained.

Our previous study on diffused sensing [15, 16] involved an observation method for decorrelating between channels by placing microphones in a diffused acoustic field. However, the impulse response length becomes long and a common zero point appears at certain frequencies. Then, the audio output latency increases and the sound quality of beamforming output degrades caused by instable filters.

4. ARRAY STRUCTURE TO INCREASE MUTUAL INFORMATION USING PARABOLIC REFLECTORS

We propose an array structure to both increase the mutual information of MIMO and shorten the impulse response length. To increase the mutual information of MIMO, homogenizing the eigenvalues of the spatial correlation matrix is necessary, as in Eq. (29). It also requires the shortening of the impulse response length for low latency of audio output and stable beamforming. The basic idea to achieve the above is to place microphones around a focal point of a parabolic reflector. When a sound source is located in front of a parabolic reflector, the reflected waves pass through an area around a focal point of the reflector. Since the arrival time of each reflected wave is different without a focal point, a singular acoustic field, in which the amplitude/phase is varied drastically with the received position, will be generated around a focal point. By optimally arranging microphones around a focal point, the cross-correlation between channels can be reduced and the eigenvalues of the spatial correlation matrix can then be homogenized. Also, the impulse response length will be short since source signals arrive after reflecting several times. Since the unnecessary zero points are then removed, stable beamforming can be achieved.

Figure 1 shows the proposed array structure using parabolic reflectors and omni-directional microphones. Eight microphones were

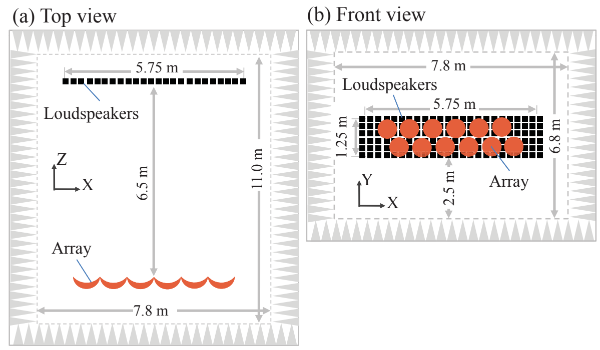

Fig. 3. Position of array and loudspeakers in anechoic chamber

Table 1. Parameter

Sampling frequency	48 kHz
Analyzed frequency band	0.3 – 16.0 kHz
Number of sound sources, K	144 (Rows: 6, Columns: 24)
Number of microphones, M	96
Impulse response length	21.3 ms

placed around a focal point, as shown in Fig. 2. Since 12 reflectors were located in parallel, a total of $M = 96$ microphones were used for capturing sounds. To determine the microphone position, the acoustic characteristics around a focal point were measured in a priori in an anechoic chamber. We placed a parabolic reflector at the center of an array located 6.5 m from loudspeakers, as shown in Fig. 3. We measured the transfer functions from 144 loudspeakers to 196 microphones placed in a square of 140 mm in front of a reflector, as shown in Fig. 2 (a). The minimum distance between the microphones was 10 mm. By calculating the mutual information of MIMO by using Eq. (27), for each microphone combination, array arrangement was determined sequentially, as shown in Fig. 2. The short impulse response length around 21.3 ms was then obtained.

5. EXPERIMENTS

5.1. Experimental conditions

We conducted experiments to evaluate the proposed array structure. As shown in Fig, 3, the proposed array structure and sound sources (loudspeakers) were placed in an anechoic chamber. The loudspeakers were arranged in a grid pattern of 6 rows by 24 columns whose interval was 0.25 m. The shortest distance between the microphones and loudspeakers was 6.5 m. By using the measured impulse responses between the microphones (M=96) and loudspeakers (K= 144), $\mathbf{A}(\omega)$ and $\mathbf{R}_A(\omega)$ were calculated. Other experimental parameters are listed in Table 1.

We placed omni-directional microphones in midair as a conventional array structure. The position of the microphones was the same as with the proposed array structure, i.e., the difference between the proposed and conventional microphone arrays was whether there are parabolic reflectors. The transfer function between the m-th microphone and k-th loudspeaker when using the conventional array structure was calculated based on the following wave equation:

$$H_{m,k}(\omega) = \exp\left(j\omega||\mathbf{p}_m - \mathbf{q}_k||/c\right), \tag{32}$$

where $\mathbf{p}_m$, $\mathbf{q}_k$, and c denote the position of the m-th microphone, that of the k-th sound source, and sound velocity, respectively. The

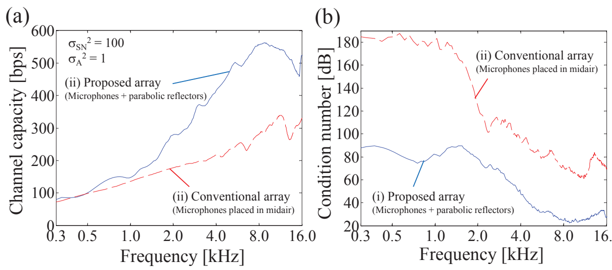

Fig. 4. Homogeneity of eigenvalues of spatial correlation matrix evaluated by (a) channel capacity and (b) condition number

spatial correlation matrix when using the conventional microphone array structure was modeled as

$$\mathbf{R}_{\mathrm{H}}(\omega) = \mathbf{H}(\omega)\mathbf{H}^{\mathrm{H}}(\omega), \tag{33}$$

where

$$\mathbf{H}(\omega) = [\mathbf{h}_1(\omega), \ldots, \mathbf{h}_K(\omega)], \tag{34}$$

$$\mathbf{h}_k(\omega) = [H_{1,k}(\omega), \ldots, H_{M,k}(\omega)]^{\mathrm{T}}. \tag{35}$$

5.2. Experimental results

We investigated the homogeneity of the eigenvalues of the spatial correlation matrix when using the proposed array $\mathbf{R}_{\mathrm{A}}(\omega)$ and conventional array $\mathbf{R}_{\mathrm{H}}(\omega)$. To measure the homogeneity of the eigenvalues, (a) the channel capacity $C(\omega)$ defined in Eq. (21) when $\sigma^2_{\mathrm{SN}}(\omega) = 100$ and $\sigma^2_{\mathrm{A}}(\omega) = 1$ and (b) the matrix condition number $\Phi(\omega)$ were used.

$$\Phi(\omega) = 10\log_{10}\left(\Lambda_1(\omega)/\Lambda_M(\omega)\right) \quad [\mathrm{dB}] \tag{36}$$

When the eigenvalues of the spatial correlation matrix were homogenized perfectly, (a) $C(\omega)$ was maximized to 639.2 bps and (b) $\Phi(\omega)$ was minimized to 0 dB. Figure (4) shows the experimental results when using the proposed/conventional array structure. The $C(\omega)$ increased with the proposed array and $\Phi(\omega)$ reduced, especially in a high frequency band of more than 1.5 kHz. Thus, it was confirmed that the eigenvalues of the spatial correlation matrix were homogenized using the proposed array structure.

To evaluate the effectiveness of the proposed array structure, we calculated the sensitivity to each source position. As a basic source enhancement method, MVDR beamforming was applied. The beamforming filters to emphasize the i-th sound source when using the proposed and conventional array structures are respectively given by

$$\mathbf{w}_{\mathrm{A},i}(\omega) = \frac{\mathbf{R}_{\mathrm{A}}^{-1}(\omega)\mathbf{a}_i(\omega)}{\mathbf{a}_i^{\mathrm{H}}(\omega)\mathbf{R}_{\mathrm{A}}^{-1}(\omega)\mathbf{a}_i(\omega)}, \tag{37}$$

$$\mathbf{w}_{\mathrm{H},i}(\omega) = \frac{\mathbf{R}_{\mathrm{H}}^{-1}(\omega)\mathbf{h}_i(\omega)}{\mathbf{h}_i^{\mathrm{H}}(\omega)\mathbf{R}_{\mathrm{H}}^{-1}(\omega)\mathbf{h}_i(\omega)}, \tag{38}$$

where the sensitivity to the target source was constrained to 0 dB no matter which filter was used. Sensitivity to the k-th sound source was calculated by

$$Z_{\mathrm{A},i,k}(\omega) = 10\log_{10}\left|\mathbf{w}^{\mathrm{H}}_{\mathrm{A},i}(\omega)\mathbf{a}_k(\omega)\right|^2 \quad [\mathrm{dB}], \tag{39}$$

$$Z_{\mathrm{H},i,k}(\omega) = 10\log_{10}\left|\mathbf{w}^{\mathrm{H}}_{\mathrm{H},i}(\omega)\mathbf{h}_k(\omega)\right|^2 \quad [\mathrm{dB}]. \tag{40}$$

Figure 5 shows the sensitivities at 144 loudspeakers placed in a grid, as shown in Fig. 3 (b), when using the conventional array structure.

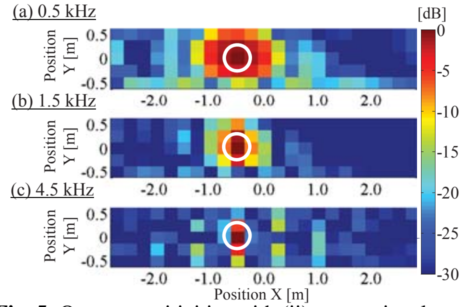

Fig. 5. Output sensitivities with (ii) conventional array

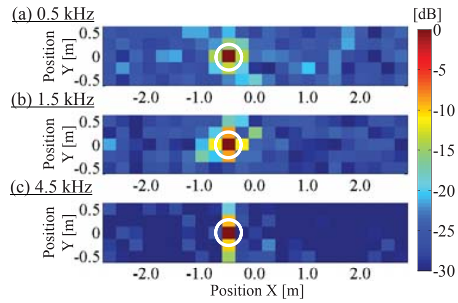

Fig. 6. Output sensitivities with (i) proposed array

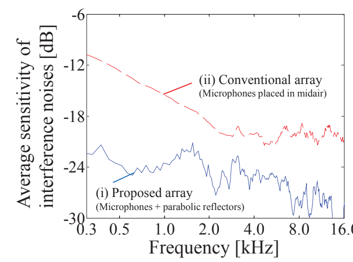

Fig. 7. Average sensitivity of interference noise sources

The position of the target source is denoted with a circle. With the conventional array structure, the beam-width of the mainlobe was broadened to less than 1.5 kHz. Figure 6 shows the sensitivities with the proposed array structure. The beam-width narrowed over broad frequencies. Figure 7 shows the average sensitivities to the interference noise sources. Compared with the conventional array, the noise power in the beamforming output reduced with the proposed array structure. Thus, it was confirmed that the eigenvalues of the spatial correlation matrix was homogenized and sharp directivity over broad frequency range was formed with the proposed array structure.

6. CONCLUSION

We found that the mutual information of MIMO is maximized by homogenizing the eigenvalues of the spatial correlation matrix. We proposed an array structure to homogenize the eigenvalues of the spatial correlation matrix using parabolic reflectors and microphones. Through experiments conducted in an anechoic chamber, we confirmed that the proposed array structure was effective in asymptotically homogenizing the eigenvalues of the spatial correlation matrix and forming sharp directivity over broad frequencies.

Other topics requiring further study include investigation of how far away sound sources can be captured with the proposed microphone array structure and robustness against noises and reverberations to capture the sound sources in practical environments.

7. REFERENCES

[1] H. L. V. Trees, *Optimum array processing*, Wiley-Interscience (Part IV ed.), 2002.

[2] D. H. Johnson and D. E. Dudgeon, *Array processing: concepts and techniques*, Englewood Cliffs, NJ: Prentice-Hall, 1993.

[3] O. L. Frost III, "An algorithm for linearly constrained adaptive array processing," *in Proc. IEEE*, vol. 60, pp. 926–935, 1972.

[4] C. Marro, Y. Mahieux, K. U. Simmer, "Analysis of noise reduction and dereverberation techniques based on microphone arrays with postfiltering," *IEEE Trans. on Speech and Audio Proc.*, no. 6, pp. 240–259, 1998.

[5] T. Wolff and M. Buck, "A generalized view on microphone array postfilters," *in Proc. IWAENC 2010*, 2010.

[6] R. Zelinski, "A microphone array with adaptive post–filtering for noise reduction in reverberant rooms," *in Proc. ICASSP 1988*, 5, 2578–2581, 1988.

[7] I. A. McCowan, H. Bourlard, "Microphone array post-filter based on noise field coherence," *IEEE Trans. on Audio, Speech, and Language Proc.*, no. 11, pp. 709–716, 2003.

[8] K. U. Simmer, J. Bitzer, and C. Marro, *Microphone arrays: signal processing techniques and applications*, chapter 3, pp. 39–60, Springer, 1 edition, 2001.

[9] Y. Hioka, K. Furuya, K. Kobayashi, K. Niwa, Y. Haneda, "Underdetermined sound source separation using power spectrum density estimated by combination of directivity gain," *IEEE Trans. on Audio, Speech, and Language Proc.*, vol. 21, pp. 1240–1250, 2013.

[10] K. Niwa, Y. Hioka, and K. Kobayashi, "Post-filter design for speech enhancement in various noisy environments," *in Proc. IWAENC 2014*, pp. 36–40, 2014.

[11] J. Meyer, and G. W. Elko, "A spherical microphone array for spatial sound recordings," *J. Acoust. Soc. Am.* , vol. 111, Issue 5, 2346, 2002.

[12] B. Rafaely, "Open–sphere designs for spherical microphone arrays," *IEEE Trans. Audio, Speech, and Language Proc.*, vol. 15, no. 2, pp. 727–732, 2007.

[13] H. Morgenstern and B. Rafaely, "Analysis of acoustics MIMO systems in enclosed sound fields," *in Proc. ICASSP 2012*, pp. 209–212, 2012.

[14] T. D. Abhayapala, and D. B. Ward, "Theory and design of high order sound field microphones using spherical microphone array," *in Proc. ICASSP2002*, vol. II, pp. 1949–1952, 2002.

[15] K. Niwa, Y. Hioka, K. Furuya, and Y. Haneda, "Diffused sensing for sharp directive beamforming," *IEEE Trans. on Audio, Speech and Language Proc.*, vol. 21, pp. 2346–2355, 2013.

[16] K. Niwa, Y. Hioka, K. Kobayashi, K. Furuya, and Y. Haneda, "Evaluation of microphone array based on diffused sensing with various filter design methods", *in Proc. EUSIPCO 2013*(156974157), 2013.

[17] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Technical Journal*, vol. 1, no. 2, pp. 41–59, 1996.

[18] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multi-element antennas," *Wireless Personal Communications*, vol. 6, no. 3, pp. 311–335, 1998.