EVALUATION OF MICROPHONE ARRAY BASED ON DIFFUSED SENSING WITH VARIOUS FILTER DESIGN METHODS

Kenta Niwa[1]*, Yusuke Hioka*[2]*, Kazunori Kobayashi*[1]*, Ken'ichi Furuya*[3]*, and Yoichi Haneda*[4]

[1]: NTT Media Intelligence Laboratories, Tokyo 180-8585, JPN
[2]: Department of Electrical and Computer Engineering, University of Canterbury, Christchurch 8140, NZ
[3]: Faculty of Engineering, Oita University, Oita 870-1192, JPN
[4]: Graduate School of Informatics and Engineering, The University of Electro-Communications, Tokyo 182-8585, JPN

ABSTRACT

We investigated methods for achieving sharp directivity over a broad frequency range by using a multichannel inverse filter. We previously proposed the *diffused sensing* method for segregating sound sources by decorrelating the transfer functions of each sound source by placing an array in a diffused acoustic field. Because many transfer functions must be measured in advance in our previous method, we discuss how to achieve sharp directivity based on diffused sensing using only a few pre-measurements. As a physical phenomenon, transfer functions are expected to be decorrelated automatically in a diffused acoustic field. By exploiting this property, we found that sharp directivity could be achieved with the multichannel inverse filter, which dereverberates the target paths only. Through simulations, we show that to achieve sharp directivity, the length of the multichannel inverse filter should be the same as the impulse response length.

***Index Terms*—** Microphone array, beamforming, transfer function, diffuse acoustic field, multichannel inverse filter

1. INTRODUCTION

Beamforming techniques using microphone arrays have been studied to clearly emphasize target sources [1]. Most conventional studies on beamforming have focused on the reception of target sources 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 remote position just as a camera zooms in on objects. These situations might include, e.g., zooming in on the voices of athletes on a playing field, actors in a theater, and speakers in a teleconference. The goal of this research was to achieve sharp directive beamforming.

How to design beamforming filters (S1) and how to structure the arrays to reduce the power of interference sources in the beamforming output (output interference power) (S2) have been investigated [1]. There are various well known filter design methods for (S1), for example, delay-and-sum (DS), minimum variance distortionless response (MVDR), and maximum likelihood (ML) [1]. Blind source separation (BSS) is another method for estimating the filters without prior information, such as source positions, and frequency domain independent component analysis (FD-ICA) [2] is an effective algorithm for BSS. For the latter subject (S2), various sensor arrangements have been studied, for example, linear, spherical, and random arrangements [1]. Recently, the use of a rigid spherical microphone array to generate spherical harmonic directivity patterns has been extensively studied [3]. By rearranging the array structure in such ways, the transfer function, which is the frequency response of the impulse response from a source to a microphone, can be physically varied. However, it is difficult to achieve sharp directivity over a broad frequency range since the cross-correlation between transfer functions increases when sound sources are closely positioned.

We recently proposed the diffused sensing method [4], which makes it possible to form sharp directivity over a broad frequency range by placing arrays in a diffused acoustic field. It can be implemented in practice by, e.g., positioning an array inside a reflective enclosure. Because the transfer functions of each sound source position are mutually uncorrelated in a diffused acoustic field [5], this would enable us to segregate even closely positioned sound sources. There have been studies that were focused on the characteristics of a diffuse acoustic field [6], but not on generating such a field for decorrelating the transfer functions. In our previous work [4], we designed a beamforming filter by minimizing the output interference power while constraining the response gain to the target source. However, this filter design method requires the transfer functions of every position of interference as well as that of the target sound source. Since it is difficult to model the transfer functions in a diffused acoustic field, we measured many transfer functions in advance.

The essence of diffused sensing is to decorrelate the transfer functions by varying the array structure. Provided that the transfer functions are automatically decorrelated in a diffused acoustic field, the output interference power can still be minimized even if the filters were designed only to emphasize the target source. To confirm this idea, we discuss the performance of a representative filter for retrieving a target source, i.e., the multichannel inverse filter, in terms of reducing the output interference power. The minimum length of the filter, which dereverberates the target path, is determined using the multiple input/output inverse theory (MINT) [7]. We evaluate the effective filter design method for sharp directivity bemforming based on the diffused sensing method by investigating the relationships between the filter length and the output interference power.

This paper is organized as follows. The principle of diffused sensing is explained in Sec. 2. In Sec. 3, the relationships between the output interference power and a multichannel inverse filter based on diffused sensing are investigated. Numerical simulations are described in Sec. 4, and the paper is concluded in Sec. 5.

2. PRINCIPLE OF DIFFUSED SENSING

2.1. Modeling of observed signals

Let us assume that M microphones receive a target and K interference sources. Sharp directivity is achieved by emphasizing an arbitrary target source by suppressing many interference sources (K is assumed to be a large number). The impulse responses from the m-th microphone to the target and to the k-th interference source are respectively described by $\mathrm{a}_m(l)$ and $\mathrm{b}_{k,m}(l)$ whose lengths are L. When the source signal of the target and the k-th interference at time t are respectively denoted as $\mathrm{s}(t)$ and $\mathrm{n}_k(t)$, the observed signal at the m-th microphone $\mathrm{x}_m(t)$ is expressed as

$$\mathrm{x}_m(t)=\sum_{l=0}^{L-1}\mathrm{a}_m(l)\mathrm{s}(t-l)+\sum_{k=1}^{K}\sum_{l=0}^{L-1}\mathrm{b}_{k,m}(l)\mathrm{n}_k(t-l). \quad (1)$$

By applying a short-time Fourier transform to $\mathrm{x}_m(t)$, the convolved mixture in Eq. (1) is approximated as an instantaneous mixture at each frequency:

$$x_m(\omega,t)=a_m(\omega)s(\omega,t)+\sum_{k=1}^{K}b_{k,m}(\omega)n_k(\omega,t), \quad (2)$$

where ω represents frequency, $x_m(\omega,t)$, $s(\omega,t)$, and $n_k(\omega,t)$ denote the time-frequency representation of $\mathrm{x}_m(t)$, $\mathrm{s}(t)$, and $\mathrm{n}_k(t)$ respectively, and $a_m(\omega)$ and $b_{k,m}(\omega)$, which we call the transfer functions, are the frequency responses from the target and the k-th interference source to the m-th microphone, respectively. Let us rewrite Eq. (2) in matrix notation:

$$\boldsymbol{x}(\omega,t)=\boldsymbol{a}(\omega)s(\omega,t)+\sum_{k=1}^{K}\boldsymbol{b}_k(\omega)n_k(\omega,t), \quad (3)$$

where

$$\boldsymbol{x}(\omega,t)=[x_1(\omega,t),\ldots,x_M(\omega,t)]^{\mathrm{T}},$$
$$\boldsymbol{a}(\omega)=[a_1(\omega),\ldots,a_M(\omega)]^{\mathrm{T}},$$
$$\boldsymbol{b}_k(\omega)=[b_{k,1}(\omega),\ldots,b_{k,M}(\omega)]^{\mathrm{T}},$$

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

2.2. Beamforming to minimize output interference power

The output signal of beamforming $\mathrm{y}(t)$ is calculated by convolving $\mathrm{x}_m(t)$ with the beamforming filter $\mathrm{w}_m(t)$, which is designed to emphasize the target source,

$$\mathrm{y}(t)=\sum_{m=1}^{M}\sum_{j=0}^{J-1}\mathrm{w}_m(j)\mathrm{x}_m(t-j), \quad (4)$$

where J is the filter length. When the time-frequency representation of $\mathrm{y}(t)$ is described as $y(\omega,t)$, it is approximately calculated by

$$y(\omega,t)=\sum_{m=1}^{M}\boldsymbol{w}^{\mathrm{H}}(\omega)\boldsymbol{x}(\omega,t), \quad (5)$$

where $^{\mathrm{H}}$ denotes a Hermitian conjugate, and the complex conjugate of $w_m(\omega)$ corresponds to the frequency response of $\mathrm{w}_m(j)$,

$$\boldsymbol{w}(\omega)=[w_1(\omega),\ldots,w_M(\omega)]^{\mathrm{T}}.$$

The output interference power $p_{\mathrm{N}}(\omega)$ is defined as the power of $y_{\mathrm{N}}(\omega,t)$, which is the interference component included in $y(\omega,t)$ [1],

$$p_{\mathrm{N}}(\omega)=\mathrm{E}\left[|y_{\mathrm{N}}(\omega,t)|^2\right], \quad (6)$$

where $\mathrm{E}[\cdot]$ is the expectation operator, which can be replaced by the time-averaging operator with the assumption of ergodicity. When assuming that the source signals are uncorrelated to each other, $p_{\mathrm{N}}(\omega)$ is calculated using transfer functions and filters as

$$p_{\mathrm{N}}(\omega)=\sum_{k=1}^{K}|\boldsymbol{w}^{\mathrm{H}}(\omega)\boldsymbol{b}_k(\omega)|^2. \quad (7)$$

Various filter design methods have been studied to minimize $p_{\mathrm{N}}(\omega)$ [1]. When using the ML method, the filter is designed to minimize $p_{\mathrm{N}}(\omega)$ while emphasizing the target:

$$\boldsymbol{w}_{\mathrm{ML}}(\omega)=\frac{\boldsymbol{R}^{-1}(\omega)\boldsymbol{h}(\omega)}{\boldsymbol{h}^{H}(\omega)\boldsymbol{R}^{-1}(\omega)\boldsymbol{h}(\omega)}, \quad (8)$$

where $\boldsymbol{h}(\omega)=[h_1(\omega),\ldots,h_M(\omega)]^{\mathrm{T}}$ is the array manifold vector that models the direct propagations between the target source and microphones, and $\boldsymbol{R}(\omega)$ denotes the spatial correlation matrix of interferences. Since $\boldsymbol{R}(\omega)$ is composed of cross-correlations between microphones, it is calculated using transfer functions of interferences when assuming that the source signals are uncorrelated:

$$\boldsymbol{R}(\omega)=\sum_{k=1}^{K}\boldsymbol{b}_k(\omega)\boldsymbol{b}_k^{\mathrm{H}}(\omega). \quad (9)$$

However, it is difficult to minimize $p_{\mathrm{N}}(\omega)$ if the transfer functions are highly correlated.

2.3. Diffused sensing to decorrelate transfer functions

We recently introduced techniques for optimizing the transfer functions in order to achieve sharp directivity in our diffused sensing method [4]. If the transfer functions are decorrelated over a broad frequency range, which satisfies Eq. (10), sound sources can be segregated even if they were positioned closely.

$$\boldsymbol{a}(\omega) \perp \boldsymbol{b}_1(\omega) \perp, \ldots, \perp \boldsymbol{b}_K(\omega) \quad (\forall\, \omega) \tag{10}$$

One method of decorrelating transfer functions involves placing an array in a diffuse acoustic field. When the cross-correlation between microphones is described by $\gamma(\omega)$, the spatial expectation of $\gamma(\omega)$ in a diffuse acoustic field is modeled by [5],

$$\mathrm{E}\{\gamma(\omega)\} = \mathrm{sinc}\left(\frac{\omega\,||\mathbf{p}||}{c}\right), \tag{11}$$

where $\mathbf{p}$ and c denote the position vector between two microphones and the sound velocity, respectively.

The expectation of $\gamma(\omega)$ can be calculated by averaging the cross-correlation of the signals observed by two fixed microphones. The average is calculated over the signals arriving from various spatially distributed location of the sound source. By positioning microphones widely apart in a diffuse acoustic field, the transfer functions can be decorrelated as

$$\lim_{||\mathbf{p}|| \to \infty} \mathrm{E}\{\gamma(\omega)\} \to 0. \tag{12}$$

Therefore, placing an array in a diffuse acoustic field is a method of decorrelating the transfer functions.

In our previous work, the filters were calculated to minimize $p_{\mathrm{N}}(\omega)$ with deconvolution between the target and microphones [4]. After pre-measuring $(K+1)M$ transfer functions from both the target and interferences to the microphones, the filter was computed by substituting $\boldsymbol{a}(\omega)$ instead of $\boldsymbol{h}(\omega)$ into Eq. (8) as

$$\boldsymbol{w}_{\mathrm{CDS}}(\omega) = \frac{\boldsymbol{R}^{-1}(\omega)\boldsymbol{a}(\omega)}{\boldsymbol{a}^{\mathrm{H}}(\omega)\boldsymbol{R}^{-1}(\omega)\boldsymbol{a}(\omega)}. \tag{13}$$

Since the cross-correlation between microphones was reduced by capturing sounds in a diffuse acoustic field, $p_{\mathrm{N}}(\omega)$ was minimized over a broad frequency range. However, pre-measuring of many transfer functions was required to calculate the filter in Eq. (13).

3. MULTICHANNEL INVERSE FILTER BASED ON DIFFUSED SENSING

3.1. Basic properties of spatial correlation matrix when capturing sound in diffused acoustic field

Let us consider minimizing $p_{\mathrm{N}}(\omega)$ based on diffused sensing without pre-measuring of many transfer functions. Since the transfer functions are decorrelated automatically, which satisfies Eq. (12), when capturing sounds in a diffuse acoustic field and K is a large number, $\boldsymbol{R}(\omega)$ would be whitened as

$$\boldsymbol{R}(\omega) = \sum_{k=1}^{K} \boldsymbol{b}_k(\omega)\boldsymbol{b}_k^{\mathrm{H}}(\omega) \approx \boldsymbol{I}. \tag{14}$$

Then, the filter of Eq. (13) is rewritten by omitting $\boldsymbol{R}(\omega)$ as

$$\boldsymbol{w}(\omega) = \frac{\boldsymbol{a}(\omega)}{\boldsymbol{a}^{\mathrm{H}}(\omega)\boldsymbol{a}(\omega)}. \tag{15}$$

This means that there are possibilities to minimize $p_{\mathrm{N}}(\omega)$ by emphasizing the target source without constraints to minimize $p_{\mathrm{N}}(\omega)$. If this is true, the number of transfer functions to be pre-measured is drastically reduced from $(K+1)M$ paths to M target paths.

3.2. Methods of calculating multichannel inverse filter

We now explain the filter design methods for emphasizing the target source. To discuss the relationships between $p_{\mathrm{N}}(\omega)$ and filter length J discussed in Sec. 3.3, filter coefficients are calculated in the time-domain. If the unit impulse function $\mathbf{z}$ is output when convolving the filter with impulse responses of the target using Eq. (16), the target source is emphasized without distortion:

$$\mathbf{z} = \mathbf{A}\mathbf{w}, \tag{16}$$

where

$$\mathbf{A}_m = \left.\overbrace{\begin{bmatrix} \mathrm{a}_m(0) & & & \mathbf{O} \\ \mathrm{a}_m(1) & \mathrm{a}_m(0) & & \\ \vdots & \vdots & \ddots & \\ \mathrm{a}_m(L{-}1) & \mathrm{a}_m(L{-}2) & \ddots & \mathrm{a}_m(0) \\ & \mathrm{a}_m(L{-}1) & \ddots & \vdots \\ & & \ddots & \mathrm{a}_m(L{-}2) \\ \mathbf{O} & & & \mathrm{a}_m(L{-}1) \end{bmatrix}}^{J}\right\}{\scriptstyle J+L-1},$$

$$\begin{aligned} \mathbf{A} &= [\mathbf{A}_1, \ldots, \mathbf{A}_M], \\ \mathbf{w}_m &= [\mathrm{w}_m(0), \ldots, \mathrm{w}_m(J-1)]^{\mathrm{T}}, \\ \mathbf{w} &= [\mathbf{w}_1^{\mathrm{T}}, \ldots, \mathbf{w}_M^{\mathrm{T}}]^{\mathrm{T}}, \\ \mathbf{z} &= [\overbrace{0, \ldots, 0}^{J-1}, 1, \overbrace{0, \ldots, 0}^{L-1}]^{\mathrm{T}}, \end{aligned}$$

where we call $\mathbf{A}$ the convolution matrix of the target source.

As a method for emphasizing the target source without distortion, the multichannel inverse filter, which is the solution to the inverse problem of Eq. (16), is derived. Depending on the relationships between L, J, and M, several cases can be considered to calculate the multichannel inverse filter [7]. When $\mathbf{A}$ is a square matrix, i.e., $J_{\mathrm{I}} = (L-1)/(M-1)$, the multichannel inverse filter is solved by

$$\mathbf{w}_{\mathrm{I}} = \mathbf{A}^{-1}\mathbf{z}. \tag{17}$$

We call this solution the MIF-I method.

On the other hand, the multichannel inverse filter is calculated using the overdetermined least squares method when $J_{\mathrm{II}} > (L-1)/(M-1)$ as

$$\mathbf{w}_{\mathrm{II}} = \mathbf{A}^{+}\mathbf{z} = \mathbf{A}^{\mathrm{T}}(\mathbf{A}\mathbf{A}^{\mathrm{T}})^{-1}\mathbf{z}, \tag{18}$$

where $\mathbf{A}^{+}$ denotes the pseudo-inverse matrix of $\mathbf{A}$. We call this solution the MIF-II method. Even though $J_{\mathrm{I}} < J_{\mathrm{II}}$, the target source can still be emphasized without distortion by using either $\mathbf{w}_{\mathrm{I}}$ or $\mathbf{w}_{\mathrm{II}}$ if no common zero point exists [7].

3.3. Relationships between output interference power and length of multichannel inverse filter

To determine the filter length of $\mathbf{w}_{\mathrm{II}}$ to form sharp directivity, we investigated the relationships between $p_{\mathrm{N}}(\omega)$ and J_{II}. Because $p_{\mathrm{N}}(\omega)$ is the total sum of the output power of K interference sources, we calculate the response to the k-th interference source $\mathbf{r}_k = [\mathrm{r}_k(0), \ldots, \mathrm{r}_k(J_{\mathrm{II}}+L-2)]^{\mathrm{T}}$ by convolving $\mathbf{w}_{\mathrm{II}}$ with impulse responses as

$$\mathbf{r}_k = \mathbf{B}_k\mathbf{w}_{\mathrm{II}} = \mathbf{B}_k\mathbf{A}^{\mathrm{T}}(\mathbf{A}\mathbf{A}^{\mathrm{T}})^{-1}\mathbf{z}, \tag{19}$$

where $\mathbf{B}_k$ is the convolution matrix of the k-th interference source and is composed of $\mathrm{b}_{k,m}(l)$, in the same way that $\mathbf{A}$ in Eq. (16) is composed of $\mathrm{a}_m(l)$. The covariance matrix of $\mathbf{A}$ in Eq. (19) is expanded as

$$\mathbf{A}\mathbf{A}^{\mathrm{T}} = \begin{bmatrix} \sigma_0 & \varphi_{0,1} & \cdots & \varphi_{0,J_{\mathrm{II}}+L-2} \\ \varphi_{1,0} & \sigma_1 & \cdots & \varphi_{1,J_{\mathrm{II}}+L-2} \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_{J_{\mathrm{II}}+L-2,0} & \varphi_{J_{\mathrm{II}}+L-2,1} & \cdots & \sigma_{J_{\mathrm{II}}+L-2} \end{bmatrix}, \tag{20}$$

where

$$\sigma_i = \sum_{m=1}^{M}\sum_{l=0}^{i} \mathrm{a}_m^2(i-l), \tag{21}$$

$$\varphi_{i,j} = \sum_{m=1}^{M}\sum_{l=0}^{\min(i,j)} \mathrm{a}_m(i-l)\mathrm{a}_m(j-l), \tag{22}$$

and $\mathrm{a}_m(i) = 0$ when $i > L-1$.

When capturing sounds in a diffused acoustic field, $\mathrm{a}_m(l)$ becomes a long uncorrelated series such as white noise. Since then the autocorrelation function of $\mathrm{a}_m(l)$ has a sharp peak, such as a unit impulse, $\varphi_{i,j}$ will decrease automatically. If all $\varphi_{i,j}$ are approximated to zero, Eq. (20) is rewritten as

$$\mathbf{A}\mathbf{A}^{\mathrm{T}} \approx \mathrm{diag}\left[\sigma_0, \ldots, \sigma_{J_{\mathrm{II}}+L-2}\right]. \tag{23}$$

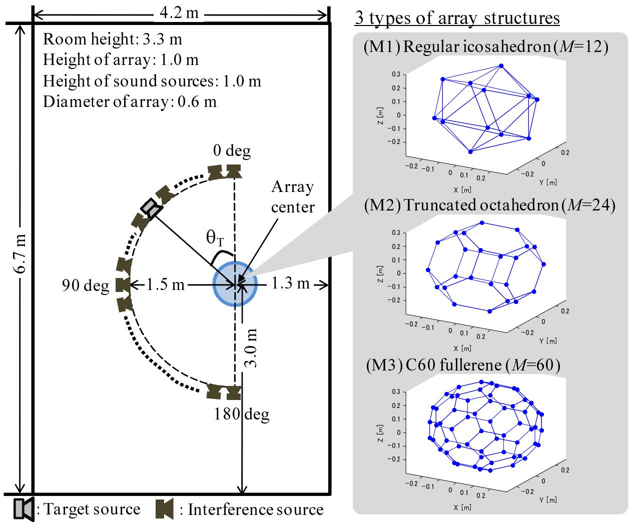

Fig. 1. Room conditions and array structure

By substituting Eq. (23) into Eq. (19), $\mathbf{r}_k$ is calculated by

$$\mathbf{r}_k \approx \mathbf{B}_k\mathbf{A}^{\mathrm{T}}[\overbrace{0,\ldots,0}^{J_{\mathrm{II}}-1}, 1/\sigma_{J_{\mathrm{II}}-1}, \overbrace{0,\ldots,0}^{L-1}]^{\mathrm{T}}, \tag{24}$$

where the i-th component of $\mathbf{r}_k$ is

$$\mathrm{r}_k(i) = \frac{\sum_{m=1}^{M}\sum_{l=0}^{\min(i,J_{\mathrm{II}}-1)} \mathrm{b}_{k,m}(i-l)\mathrm{a}_m(J_{\mathrm{II}}-l-1)}{\sum_{m=1}^{M}\sum_{l=0}^{J_{\mathrm{II}}-1} \mathrm{a}_m^2(J_{\mathrm{II}}-l-1)}, \tag{25}$$

and $\mathrm{b}_{k,m}(i) = 0$ when $i > L-1$. Since $\mathrm{a}_m(l)$ and $\mathrm{b}_{k,m}(l)$ are decorrelated in a diffused acoustic field, the numerator of Eq. (25) can be reduced automatically. The denominator of Eq. (25) is maximized by increasing J_{II} so that it is greater than L. Thus, we use $J_{\mathrm{II}} = L$ to minimize the norm of $\mathbf{r}_k$ then $p_{\mathrm{N}}(\omega)$ can be minimized.

4. NUMERICAL SIMULATIONS

4.1. Simulation conditions

Numerical simulations were conducted to evaluate the reduction in $p_{\mathrm{N}}(\omega)$ in a diffused acoustic field. Figure 1 shows the room conditions used to generate diffusely reflected impulse responses using the image method [8]. We used three types of arrays whose microphones were respectively positioned at the vertexes of (M1) a regular icosahedron ($M = 12$), (M2) a truncated octahedron ($M = 24$), and (M3) a C60 fullerene ($M = 60$). A target source arriving from $\theta_{\mathrm{T}} = 45$ degrees and interference sources (K=180) were placed 1.5 meters from the array center. Since the impulse response had $L = 2048$ taps, and the reflection coefficient of the walls was 0.85, many reflected sounds were included in the impulse response. Other parameters are listed in Table 1. The beamforming filters were calculated using the following four methods: (F1) the

Table 1. Simulation parameters

Sampling frequency	8.0 kHz
Analyzed frequency range	0.5 kHz – 3.5 kHz
Number of microphones, M	12, 24, 60
Diameter of array	0.6 m
Arrival direction of target source, θ_T	45 deg
Number of interference sources, K	180 (Angular interval: 1 deg)
Impulse response length, L	2048 taps
Room size	4.2 m(W)×6.7 m(D)×3.3 m(H)
Reflection coefficient of walls	0.85

ML method in Eq. (8) with $J = 16384$ taps, (F2) the conventional diffused sensing (CDS) method in Eq. (13) with $J = 16384$ taps, (F3) the MIF-I method in Eq. (17) with $J_\mathrm{I} = (L-1)/(M-1)$ taps, and (F4) the MIF-II method in Eq. (18) with $J_\mathrm{II} = L$ taps.

4.2. Simulation results

Figure 2 shows the frequency averaged $p_\mathrm{N}(\omega)$, which was normalized by the response power to the target transfer functions when using the four filter design methods. Since the transfer functions of both the target and interferences sources were known when using (F2) the CDS method, $p_\mathrm{N}(\omega)$ was minimized in all four filter design methods. Although $p_\mathrm{N}(\omega)$ was not minimized with (F3) the MIF-I method, (F4) the MIF-II method was effective in decreasing $p_\mathrm{N}(\omega)$ even though it was calculated using the transfer functions of the target source only. Figure 3 shows the directivity patterns when using (M2) the truncated octahedron array. With (F4) the MIF-II method, sharp directivity was formed over a broad frequency range, as it was using (F2) the CDS method. These results shows that sharp directive beamforming based on diffused sensing was achieved even when the transfer functions of interference sources were unknown.

5. CONCLUSION

We investigated the performances of various filter design methods in reducing the output interference power when capturing sounds based on diffused sensing. Through theoretical analysis and numerical simulations, we showed that the output interference power was reduced sufficiently by using the multichannel inverse filter, which was designed to dereverberate the target paths only. Therefore, the length of filter should be the same as that of impulse response.

Other issues require further study, including the use of blind deconvolution techniques to achieve sharp directivity without pre-measuring the transfer functions of the target source and determining how to optimize the array structure to decorrelate transfer functions.

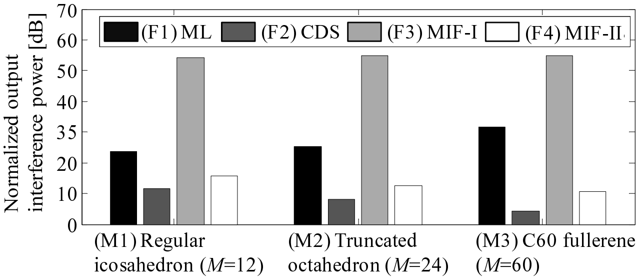

Fig. 2. Normalized output interference power

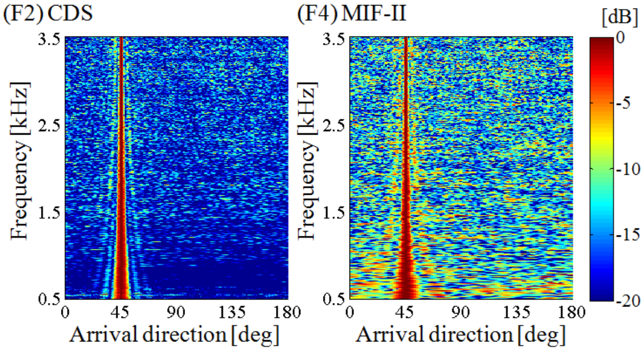

Fig. 3. Directivity patterns using (M2) truncated octahedron array, (left) with (F2) CDS method, and (right) with (F4) MIF-II method

6. REFERENCES

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

[2] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent component analysis*, Wiley, 2001.

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

[4] K. Niwa, S. Sakauchi, K. Furuya, M. Okamoto, and Y. Haneda, "Diffused sensing for sharp directivity microphone array," *ICASSP 2012*, pp. 225–228, 2012.

[5] M. Tohyama, H. Suzuki, and Y. Ando, *The nature and technology of acoustic space*, Academic Press, 1995.

[6] I. McCowan, M. Lincoln, and I. Himawan, "Microphone array shape calibration in diffuse noise fields," *IEEE Trans. Audio, Speech, and Language Process.*, vol. 16, no. 3, pp. 666–670, 2008.

[7] M. Miyoshi and Y. Kaneda, "Inverse filtering of room acoustics," *IEEE Trans. Speech Audio Process.*, vol. 36, no. 2, pp. 145–152, 1988.

[8] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.* , vol. 65, pp. 943–950, 1979.