BEAMFORMING ARRAY TECHNIQUE WITH CLUSTERED MULTICHANNEL NOISE COVARIANCE MATRIX FOR MECHANICAL NOISE REDUCTION

Masahito Togami, Takashi Sumiyoshi, Yasunari Obuchi, Yohei Kawaguchi, and Hiroaki Kokubo

Central Research Laboratory, Hitachi Ltd.
1-280, Higashi-koigakubo Kokubunji-shi, 185-8601, Tokyo, Japan
phone: +81-42-323-1111, fax: +81-42-327-7823,
email: { masahito.togami.fe, takashi.sumiyoshi.bf, yasunari.obuchi.jx, yohei.kawaguchi.xk, hiroaki.kokubo.dz }@hitachi.com

ABSTRACT

In this paper, we propose a novel multichannel noise reduction method for a mechanical noise with a time-variant impulse response. The mechanical noise source location moves depending on the status of the actuator. In accordance with the move of the noise-source location, a most suitable multichannel beamformer is selected separately at each time-frequency bin. Each multichannel beamformer is made from a corresponding multichannel noise covariance matrix which is learned in advance. The selection criteria in the proposed method is to minimize the residual noise power in the output signal. The multichannel beamformer that minimizes the residual power after beamforming is selected. Furthermore, to reduce directional noise sources, the multichannel directional noise covariance matrix is inserted into each multichannel mechanical noise covariance matrix. Experimental results of mechanical noise reduction show that the proposed method can reduce the mechanical noise more accurately than the conventional method.

1. INTRODUCTION

Noise reduction techniques are strongly required for automatic speech recognition of communication robots or speech communication systems. Especially, the mechanical noises such as a motor noise of a communication robot or a noise source of a digital camera which happens when optical zoom is active contaminate clean speech, and degrades speech recognition performance or listenability. Conventionally, there are few works about the mechanical noise reduction. In this paper, we focus on the mechanical noise reduction. Optimized modified LSA proposed by I. Cohen [1] is the state of the art noise canceller with single microphone. This method outputs a amplitude-modified version of the microphone input signal. Noise reduction performance of these methods greatly depends on the estimation accuracy of the noise-sources amplitude at each time-frequency bin. However, the amplitude of nonstationary noise sources are difficult with single channel microphone. An alternative noise-sources amplitude estimation method is the estimation method with the indicator for the existence of the noise sources [2]. However, highly time-variant noise sources are also difficult to be reduced. Furthermore, the speech distortion of the output signal is also problematic in the single channel noise reduction techniques.

The multichannel noise reduction techniques have been actively studied. Frost' s minimum variance beamformer (MVBF) [3] is one of the major multichannel noise reduction techniques. On contrary to the single channel noise reduction techniques, when the location of the desired source is set correctly, MVBF can reduce the point noise sources theoretically without any distortion of the desired source. GSC [4] is one of the online algorithms of MVBF. The noise reduction performance is depending on the accuracy of the desired-source location. Conventionally, robust GSC algorithms have been studied [5] [6] [7]. These methods are robust against the error of the desired-source location. Blind source separation techniques for alternatives of beamforming techniques. Recently, internal noise reduction technqiues based on independent component analysis [8] have been proposed [9][10]. Conventional beamformers and ICA can reduce noise sources whose spectrum are nonstationary. When the impulse reponses of the noise sources are slowly time-variant sources, these methods can track the change by updating information about the noise sources such as a multichannel noise covariance matrix. However, when the impulse responses of the noise sources are highly nonstationary sources, conventional methods cannot track the change effectively. The mechanical noise-source location moves depending on the status of the actuator rapidly, and the impulse response of the mechanical noise-source can be approximated to be highly time-variant. Therefore, beamforming techniques which can reduce noise sources with time-variant impulse responses are required. In this paper, we propose a novel mechanical noise reduction technique. The proposed technique assumes that the existence of the mechanical noise can be detected by the external instrument, because the actuator which causes the mechanical noise is usually controlled systematically. The most important assumption in the proposed method is that the time-variant multichannnel noise covariance matrix of the mechanical noise at each time-frequency bin can be approximated by a set of the pre-learned multichannel mechanical noise covariance matrices. This assumption is often valid, because the number of the patterns of the actuator is limited in an usual case. Under this assumption, the proposed method discretizes the time-variant noise covariance matrix of the mechanical noise. The discretized multichannel noise covariance matrices are obtained in the offline learning period. Compared with single channel noise reduction techniques which estimate spectral features of the noise sources, the most discriminative point of the proposed method in the offline learning period is that the proposed method does **Not** learn power spectrum of the mechanical noise sources but also learns the impulse responses of the mechanical noise sources. In the online noise reduction period, the proposed method selects one mechanical noise covariance matrix suitable for the mechanical noise reduction at each time-frequency bin. Even when the impulse response of the noise sources change rapidly time-by-time, the proposed method can reduce the mechanical noise sources efficiently. Furthermore, the directional noise covariance matrix is estimated to reduce directional noise sources. Direction of arrival (DOA) based segregation of each time-frequency bin is performed to obtain the directional noise covariance matrix. From the segregation result, multichannel covariance matrix of the directional noise sources are updated. The multichannel directional noise covariance is inserted into each multichannel mechanical noise covariance matrix. In this paper, the proposed method was evaluated by two mechanical noise problems. The first problem is the noise reduction problem of a digital camera when optical zoom is active. The second problem is the noise reduction problem of a communication robot when it moves its arm.

2. PROBLEM STATEMENT

2.1 Input signal model

Input signal in a microphone array is modeled as the sum of the desired source convolved with a time-invariant impulse response, the directional noise source convolved with a slowly time-variant impulse response, and the mechanical noise source signal convolved

 ISSN 2076-1465

with a time-variant impulse response. The multichannel input signal is depicted as follows:

$$\boldsymbol{x}(t) = [\ x_1(t) \quad \dots \quad x_m(t) \quad \dots \quad x_M(t)\ ]^T, \tag{1}$$

where M is the number of the microphones, T is an operator of the transpose of a matrix or a vector, and $x_m(t)$ is the t-th sample of the m-th microphone input signal. The multichannel input signal is converted from the time domain to the time-frequency domain by the short-term-Fourier transform as follows:

$$\boldsymbol{x}(f,\tau) = s(f,\tau)\boldsymbol{a}(f) + n(f,\tau)\boldsymbol{b}(f,\tau) + \sum_{i=0}^{N-1} d_i(f,\tau)\boldsymbol{c}_i(f) + \boldsymbol{v}(f,\tau), \tag{2}$$

where $\boldsymbol{x}(f,\tau)$ is the multichannel input signal at (f,τ), f is the frequency index, τ is the frame index, $s(f,\tau)$ is the desired source signal at the frequency f and the frame τ, $\boldsymbol{a}(f)$ is the steering vector of the desired source signal, which depends on the spatial location of the desired source, $n(f,\tau)$ is the mechanical noise source signal, $\boldsymbol{b}(f,\tau)$ is the time-variant steering vector of the mechanical noise source, $d_i(f,\tau)$ is the i-th directional noise source, N is the number of the directional noise sources, $\boldsymbol{c}_i(f)$ is the steering vector of the i-th directional noise source, and $\boldsymbol{v}(f,\tau)$ is a back ground noise signal. $s(f,\tau)$, $n(f,\tau)$, $d_i(f,\tau)$, $\boldsymbol{v}(f,\tau)$ are defined as mutually independent signals.

2.2 Noise reduction problem for the mechanical noise with time-variant impulse response

The conventional multichannel beamforming extracts the desired source signal from the noisy multichannel input signal by using the multichannel linear filter $\boldsymbol{w}(f,\tau)$ as follows:

$$y(f,\tau) = \boldsymbol{w}(f,\tau)\boldsymbol{x}(f,\tau), \tag{3}$$

where $y(f,\tau)$ is the output signal, which is required to be the desired source signal $s(f,\tau)$. One way to obtain the suitable $\boldsymbol{w}(f,\tau)$ is to maximize the power ratio between the extracted desired signal after filtering by $\boldsymbol{w}(f,\tau)$ and the residual noise signal in the output signal [11]. The beamformer maximizing SNR, $\boldsymbol{w}_{\mathrm{SNR}}(f,\tau)$,is obtained as follows:

$$\begin{aligned} \boldsymbol{w}_{\mathrm{SNR}}(f,\tau) &= \underset{\boldsymbol{w}(f,\tau)}{\mathrm{argmax}} \frac{\boldsymbol{w}(f,\tau)^H \boldsymbol{R}_s(f,\tau)\boldsymbol{w}(f,\tau)}{\boldsymbol{w}(f,\tau)^H \boldsymbol{R}_n(f,\tau)\boldsymbol{w}(f,\tau)}, & (4)\\ &= \lambda \mathrm{max\_eig}(\boldsymbol{R}_n(f,\tau)^{-1}\boldsymbol{R}_s(f,\tau)), & (5) \end{aligned}$$

where max_eig is a function to extract the eigen vector whose eigen value is maximum, $\boldsymbol{R}_s(f,\tau)$ is the multichannel covariance matrix of the desired source signal, $\boldsymbol{R}_n(f,\tau)$ is the multichannel covariance matrix of the noise source signal, and λ is an arbitrary complex coefficient. $\boldsymbol{R}_s(f,\tau)$ and $\boldsymbol{R}_n(f,\tau)$ is expanded as follows:

$$\begin{aligned} \boldsymbol{R}_s(f,\tau) &= E[|s(f,\tau)|^2]\boldsymbol{a}(f)\boldsymbol{a}(f)^H, & (6)\\ \boldsymbol{R}_n(f,\tau) &= E[|n(f,\tau)|^2\boldsymbol{b}(f,\tau)\boldsymbol{b}(f,\tau)^H] \\ &\quad + \sum_{i=0}^{N-1} E[|d_i(f,\tau)|^2\boldsymbol{c}_i(f)\boldsymbol{c}_i(f)^H] \\ &\quad + E[\boldsymbol{v}(f,\tau)\boldsymbol{v}(f,\tau)^H], & (7) \end{aligned}$$

where $E[x]$ is an operator for the mathematical expectation, H is defined as an operator of Hermite transpose of a vector or a matrix, $*$ is the operator for the complex conjugate. $\boldsymbol{R}_s(f,\tau)$ is a product of the scalar coefficient $P_s(f,\tau) = E[|s(f,\tau)|^2]$ and the time-invariant matrix $\tilde{\boldsymbol{R}}_s(f) = \boldsymbol{a}(f)\boldsymbol{a}(f)^H$. $\boldsymbol{w}_{\mathrm{SNR}}(f,\tau)$ can be also separated into two terms as follows:

$$\boldsymbol{w}_{\mathrm{SNR}}(f,\tau) = \lambda\tilde{\boldsymbol{w}}_{\mathrm{SNR}}(f,\tau), \tag{8}$$

where $\tilde{\boldsymbol{w}}_{\mathrm{SNR}}(f,\tau) = \mathrm{max\_eig}(\boldsymbol{R}_n(f,\tau)^{-1}\tilde{\boldsymbol{R}}_s(f))$. Commonly , λ is obtained as follows [12]:

$$\begin{aligned} \lambda &\leftarrow \underset{\lambda}{\mathrm{argmin}} E[||s(f,\tau)\boldsymbol{a}(f) - \boldsymbol{w}_{\mathrm{SNR}}(f)s(f,\tau)\boldsymbol{a}(f)||^2] \\ &= \frac{E[\boldsymbol{x}(f,\tau)^H x_1(f,\tau)]\tilde{\boldsymbol{w}}_{\mathrm{SNR}}(f,\tau)}{\tilde{\boldsymbol{w}}_{\mathrm{SNR}}(f,\tau)E[\boldsymbol{x}(f,\tau)\boldsymbol{x}(f,\tau)^H]\tilde{\boldsymbol{w}}_{\mathrm{SNR}}(f,\tau)^H} & (9)\\ &= \frac{P_s(f,\tau)\tilde{\boldsymbol{R}}_s(f)[1]\tilde{\boldsymbol{w}}_{\mathrm{SNR}}(f,\tau)^H}{\tilde{\boldsymbol{w}}_{\mathrm{SNR}}(f,\tau)P_s(f,\tau)\tilde{\boldsymbol{R}}_s(f)\tilde{\boldsymbol{w}}_{\mathrm{SNR}}(f,\tau)^H}, & (10)\\ &= \frac{\tilde{\boldsymbol{R}}_s(f)[1]\tilde{\boldsymbol{w}}_{\mathrm{SNR}}(f,\tau)^H}{\tilde{\boldsymbol{w}}_{\mathrm{SNR}}(f,\tau)\tilde{\boldsymbol{R}}_s(f)\tilde{\boldsymbol{w}}_{\mathrm{SNR}}(f,\tau)^H}, & (11) \end{aligned}$$

where $\tilde{\boldsymbol{R}}_s(f)[1]$ is the first row of $\tilde{\boldsymbol{R}}_s(f)$. It is obvious that $P_s(f,\tau)$ is no influence on the estimation of $\tilde{\boldsymbol{w}}_{\mathrm{SNR}}(f,\tau)$ and λ. Therefore, when $\tilde{\boldsymbol{R}}_s(f)$ is estimated at the time period when there is only the desired source, the estimated value can be utilized at the noisy time period. On the other hand, $\boldsymbol{R}_n(f,\tau)$ is the time-variant matrix. Estimation of $\boldsymbol{R}_n(f,\tau)$ is required at each frame. Therefore, the problem is that estimation of the time-variant noise covariance matrix $\boldsymbol{R}_n(f,\tau)$ at each frame. Furthermore, $\boldsymbol{R}_n(f,\tau)$ is divided into 2 matrices as follows:

$$\boldsymbol{R}_n(f,\tau) = \boldsymbol{R}_{mech}(f,\tau) + \boldsymbol{R}_d(f,\tau), \tag{12}$$

where $\boldsymbol{R}_{mech}(f,\tau)$ is a multichannel mechanical noise covariance matrix, $\boldsymbol{R}_d(f,\tau)$ is a multichannel noise covariance matrix of the directional noise sources and the background noise.

3. PROPOSED METHOD

3.1 Discretization of the noise covariance matrix

The number of the patterns of the actuator is limited in an usual case. For example, the patterns of the robot's motions are limited. Therefore, even when the multichannel mechanical noise covariance are time-variant, the number of the patterns of the multichannel mechanical noise covariance is also limited. Under this assumption, the time-variant multichannnel noise covariance matrix of the mechanical noise at each time-frequency bin can be approximated by a set of the pre-learned multichannel mechanical noise covariance matrices. The time-variant noise covariance matrix, $\boldsymbol{R}_{mech}(f,\tau)$ is discretized and divided into C clusters. The multichannel mechanical noise covariance matrix at (f,τ) is depicted as follows:

$$\boldsymbol{R}_{mech}(f,\tau) \approx \boldsymbol{R}_{mech,\mathrm{index}(f,\tau)}(f), \tag{13}$$

where $index(f,\tau)$ is the segregated cluster index of (f,τ), and $\boldsymbol{R}_{mech,c}(f)$ is the c-th cluster in the discretized C clusters of the noise covariance matrix. Under the approximation of Eq. 13, when the cluster index can be obtained at each (f,τ), the beamformer maximizing SNR can be obtained by Eq. 8. The covariance matrix of the c-th noise source, $\boldsymbol{R}_{mech,c}(f)$, can be obtained by utilizing k-means clustering of the multichannel input signals in the noise-only period. k-means clustering is performed at each frequency separately for the converted multichannel input signal $\bar{\boldsymbol{x}}(f,\tau)$. The observed multichannel input signal, $\bar{\boldsymbol{x}}(f,\tau)$, is converted as follows:

$$\bar{\boldsymbol{x}}(f,\tau) = \frac{\boldsymbol{x}(f,\tau)|x_1(f,\tau)|}{|\boldsymbol{x}(f,\tau)|x_1}. \tag{14}$$

$\bar{\boldsymbol{x}}(f,\tau)$ has only the steering vector of the noise source, and this vector has no information about the power spectrum of the desired source. A distance function in k-means clustering is Euclidean distance between $\bar{\boldsymbol{x}}(f,\tau)$ and the centroid of each cluster. After k-means clustering, index(f,τ) which indicates that which cluster each time-frequency component is segregated is obtained. By using index(f,τ), the proposed method estimates the c-th multichannel mechanical noise covariance matrix as $\boldsymbol{R}_{mech,c}(f) = \sum_{\mathrm{index}(f,\tau)=c} \boldsymbol{x}(f,\tau)\boldsymbol{x}(f,\tau)^H$.

3.2 Estimation of the multichannel directional noise covariance matrix and the multichannel desired source covariance matrix

The existence of the mechanical noise can be detected by the external instrument. Updating of the multichannel covariance matrices of the directional noise sources and the desired source is controlled by information about the existence of the mechanical noise. These two matrices are updated under the condition that the background noise is small and there are both the desired source and the directional noise sources. To update two matrices, sparseness-based updating technique [13] is utilized. When the directional noise sources and the desired source are sparse enough, these sources are rarely overlapped at the same time-frequency point [14]. The proposed method detects which source is active at each time-frequency point by using direction of arrival (DOA) estimation from multichannel input signal. When there is no mechanical noise, the covariance matrix of the desired source is updated by using the sparseness assumption as follows:

$$\boldsymbol{R}_s(f,\tau) \leftarrow \alpha_{f,\tau}\boldsymbol{R}_s(f,\tau-1) + (1-\alpha_{f,\tau})\boldsymbol{x}(f,\tau)\boldsymbol{x}(f,\tau)^H, \quad (15)$$

where $\alpha_{f,\tau}$ controls speed of updating $\boldsymbol{R}_s(f,\tau)$. When estimated DOA at (f,τ) is beyond the pre-defined desired speech area, $\alpha_{f,\tau}$ is set to be 0, otherwise $\alpha_{f,\tau}$ is set to be a constant value α. Similarly, the covariance matrix of the directional noise sources, $\boldsymbol{R}_d(f,\tau)$, is updated as follows:

$$\boldsymbol{R}_d(f,\tau) \leftarrow \beta_{f,\tau}\boldsymbol{R}_d(f,\tau-1) + (1-\beta_{f,\tau})\boldsymbol{x}(f,\tau)\boldsymbol{x}(f,\tau)^H. \quad (16)$$

When estimated DOA at (f,τ) is beyond the predefined desired speech area, $\beta_{f,\tau}$ is set to be a constant value β, otherwise $\beta_{f,\tau}$ is set to be 0. And when there is mechanical noise, the covariance matrix of the desired source and that of the directional noise sources are not updated. Therefore, $\boldsymbol{R}_s(f,\tau) = \boldsymbol{R}_s(f,\tau-1)$ and $\boldsymbol{R}_d(f,\tau) = \boldsymbol{R}_d(f,\tau-1)$ in this case.

3.3 Selection criterion of noise covariance matrix at each time-frequency point

The proposed selection criterion of the noise covariance matrix at each time-frequency point is shown. The multiple hypothesis of the multichannel noise covariance matrix $\boldsymbol{R}_n(f,\tau)$ appears depending on the number of the multichannel mechanical noise covariance matrices. The c-th multichannel noise covariance matrix $\boldsymbol{R}_{n,c}(f,\tau)$ can be obtained as follows:

$$\boldsymbol{R}_{n,c}(f,\tau) = \gamma\boldsymbol{R}_{mech,c}(f,\tau) + (1-\gamma)\boldsymbol{R}_d(f,\tau), \quad (17)$$

where γ is a parameter which controls the balance between the mechanical noise reduction performance and the directional noise reduction performance. The multichannel noise covariance matrix of the directional-noise source, $\boldsymbol{R}_d(f,\tau)$, is common in Eq. 17. The c-th noise reduction filter $\boldsymbol{w}_{\mathrm{SNR},c}(f,\tau)$ which is obtained by using the desired covariance matrix $\boldsymbol{R}_s(f,\tau)$ and $\boldsymbol{R}_{n,c}(f,\tau)$ is obtained by substituting $\boldsymbol{R}_{n,c}(f,\tau)$ for $\boldsymbol{R}_n(f,\tau)$ in Eq. 8 as follows:

$$\tilde{\boldsymbol{w}}_{\mathrm{SNR},c}(f,\tau) = \max\_\mathrm{eig}(\boldsymbol{R}_{n,c}(f,\tau)^{-1}\boldsymbol{R}_s(f,\tau)), \quad (18)$$

$$\lambda_c \leftarrow \frac{\boldsymbol{R}_s(f,\tau)[1]\tilde{\boldsymbol{w}}_{\mathrm{SNR},c}(f,\tau)^H}{\tilde{\boldsymbol{w}}_{\mathrm{SNR},c}(f,\tau)\boldsymbol{R}_s(f,\tau)\tilde{\boldsymbol{w}}_{\mathrm{SNR}}(f,\tau)^H}, \quad (19)$$

$$\boldsymbol{w}_{\mathrm{SNR},c}(f,\tau) = \lambda_c\tilde{\boldsymbol{w}}_{\mathrm{SNR}}(f,\tau). \quad (20)$$

The output signal after filtering by the c-th filter, $y_c(f,\tau)$ can be expanded as follows:

$$y_c(f,\tau) = \boldsymbol{w}_{\mathrm{SNR},c}(f)^H\boldsymbol{x}(f,\tau), \quad (21)$$

$$= o_{s,c} + o_{n,c} + o_{d,c} + o_{v,c}, \quad (22)$$

where $o_{s,c}$ is defined as $s(f,\tau)\boldsymbol{w}_{\mathrm{SNR},c}(f)^H\boldsymbol{a}(f)$, $o_{n,c}$ is defined as $n(f,\tau)\boldsymbol{w}_{\mathrm{SNR},c}(f)^H\boldsymbol{b}(f,\tau)$, $o_{d,c}$ is defined as $\boldsymbol{w}_{\mathrm{SNR},c}(f)^H\sum_{i=0}^{N-1}d_i(f,\tau)\boldsymbol{c}_i(f)$, and $o_{v,c}$ is defined as $\boldsymbol{w}_{\mathrm{SNR},c}(f)^H\boldsymbol{v}(f,\tau)$. when the λ_c is correctly estimated, $o_{s,c}$ is approximately independent of the noise cluster index c. The common directional noise covariance matrix is inserted in $\boldsymbol{R}_{n,c}(f,\tau)$, so the residual directional noise in the output signal is independent of the noise cluster index. Assuming that distribution of back ground noise is i. i. d, $\mathrm{E}[||o_{v,c}||^2] = \sigma_v||\boldsymbol{w}_{\mathrm{SNR},c}(f)||^2$, and σ_v is the average power of back ground noise at each microphone.

When the background noise level is low or the l_2-norm of each filter is constant, $||o_{v,c}||^2$ is approximately independent of the noise cluster index c. Therefore, the expectation of the spectral power of the $y_c(f,\tau)$ is obtained as follows:

$$E[||y_c(f,\tau)||^2] = E[||o_s + o_{n,c} + o_d, o_{v,c}||^2]$$

$$\approx E[||o_s||^2] + E[||o_{n,c}||^2] + E[||o_d||^2] + E[||o_v||^2]. \quad (23)$$

From Eq. 23, the noise reduction filter which minimizes the residual noise is defined as follows:

$$\boldsymbol{w}_{\mathrm{SNR}}(f,\tau) = \underset{\boldsymbol{w}_{\mathrm{SNR},c}(f)\in\Omega_C(f)}{\mathrm{argmin}} E[||\boldsymbol{w}_{\mathrm{SNR},c}(f)^H\boldsymbol{x}(f,\tau)||^2],$$

$$\approx \underset{\boldsymbol{w}_{\mathrm{SNR},c}(f)\in\Omega_C(f)}{\mathrm{argmin}} ||\boldsymbol{w}_{\mathrm{SNR},c}(f)^H\boldsymbol{x}(f,\tau)||^2, \quad (24)$$

where $\Omega_C(f)$ is composed of the C noise reduction filters (the c-th element is $\boldsymbol{w}_{\mathrm{SNR},c}(f)$).

3.4 Block diagram of proposed method

The block diagram of the proposed method is summarized in Fig. 1.

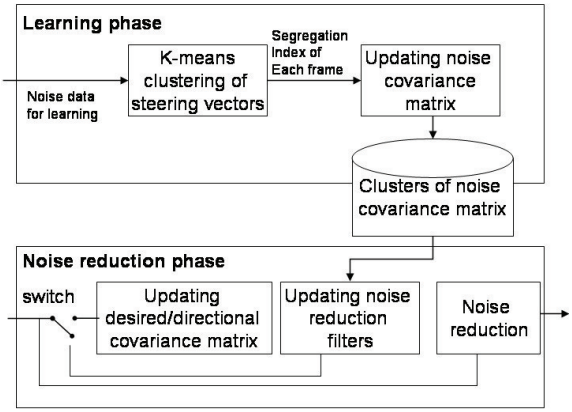

Fig. 1. Block diagram of proposed method

The proposed method is composed of two phases. The first phase is the learning phase, in this phase, the microphone input signal is only the noise signal. Discretized noise covariance matrices are learned by using k-means clustering of the normalized steering vectors. The second phase is the noise reduction phase. A microphone input signal is assumed to be mixed with the desired source and the noise sources. The multichannel desired source covariance matrix and the multichannel directional noise source covariance matrix are obtained by the sparseness based segregation of the microphone input signal. The noise signal in input signal is reduced by selecting the noise reduction filter which minimizes the output power after filtering.

4. EXPERIMENT

The proposed method was evaluated by two types of real mechanical noises. The first mechanical noise is mechanical noise reduction

on a digital camera when optical zoom is active. The second one is mechanical noise reduction on a communication robot when it moves its arm. A communication robot which was used in this experiment was EMIEW2 [15], which has been developed in Hitachi, Ltd. The proposed method is compared with the conventional noise reduction method with single noise reduction filter. The evaluation measures are Source-to-Interferences Ratio (SIR), Source-to-Distortion Ratio (SDR). These measures are defined in BSS_EVAL [16]. Noise reduction performance of the proposed method depends on the number of the noise reduction filters. Therefore, the proposed method was evaluated with various number of the noise reduction filters. Furthermore, the proposed method also depends on signal-to-noise ratio of the input signal. Therefore, the proposed method was evaluated under various SNR conditions. The desired source signal was set to be a male speech. At first, the mechanical noise reduction of a digital camera when optical zoom is active is shown. The reverberation time of the experimental room was about 100 ms. The number of the microphones used in this experiment was 2, and the sampling rate was 48 kHz. These are common settings for recoding on a digital camera. In Fig. 2, a spectrogram of the noise signal is shown. At first, the experimental result when there are no

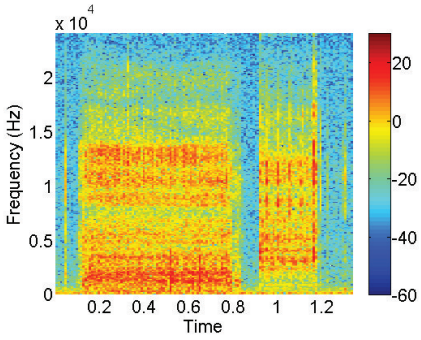

Fig. 2. A spectrogram of noise sources when optical zoom is active

directional noise is shown. The evaluation results of SDR is shown in Fig. 3, SIR in Fig. 4.

It is shown that SIR is increasing with the number of the noise reduction filters. When SNR of the input signal is low, SDR results are also increasing. On the other hand, the improvement of SDR with respect to the number of the noise reduction filters decreases at higher-SNR results. That is because there is less noise signal at the high SNR, so the noise signal can be reduced with the small number of the noise reduction filters and using the excess number of the noise reduction filters leads to degradation of SDR. However, the optical zoom noise happens typically near the microphones, so SNR tends to be less than 0 dB.

The proposed method was also evaluated under the condition that there is a directional noise source on a communication robot,

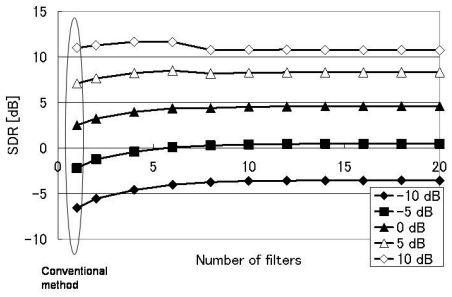

Fig. 3. SDR results of noise reduction for digital camera

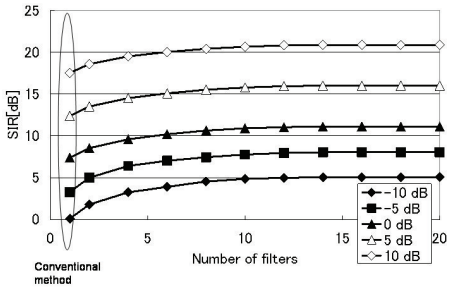

Fig. 4. SIR results of noise reduction for digital camera

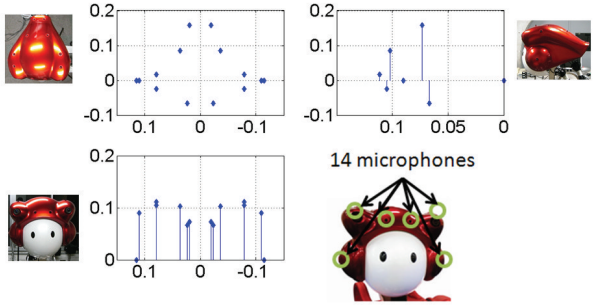

Fig. 5. Microphone alignment of EMIEW2

EMIEW2. EMIEW2 has 14 microphones. The microphone alignment of EMIEW2 is shown in Fig. 5. The sampling rate is 8 kHz. The reverberation time of the experimental room was about 300 ms. The desired speech source was located in front of EMIEW2. The distance between the desired source and EMIEW2 was 1 m. The mechanical noise that is used in this evaluation is the mechanical noise which occurs when EMIEW2 moves its arm. The directional noise source was located just beside EMIEW2. The distance between the directional noise source and EMIEW2 was 1 m. Averaged power of the directional noise is set to be equivalent to that of the desired speech source. The evaluation result under the condition that the number of the noise reduction filters is 2 is shown in Table. 1. γ is defined in Eq. 17. When the γ is a big value, noise reduction performance for the directional noise source degrades. "dir" is the multichannel noise reduction result when the mechanical noise is regarded as one of the directional noise sources and reduced by a conventional maximum SNR beamformer which maximizes the ratio between the desired source and the directional noise sources in the output signal. The noise reduction performance of the proposed method is shown to be higher than this a conventional maximum SNR beamformer. By comparison $\gamma = 1.0$ with $\gamma = 0.1$ or $\gamma = 0.5$, SNR is shown to be improved by using a combined multichannel noise covariance matrix defined in Eq. 17.

Table 1. Evaluation result for both a directional noise source and mechanical noise reduction: evaluation measure is SIR [dB].

	dir	$\gamma = 0.1$	$\gamma = 0.5$	$\gamma = 1.0$
SNR=-10 dB	-4.2	12.0	13.7	11.4
0 dB	5.8	15.0	14.3	11.4
10 dB	13.3	15.6	14.4	11.4

A sample of the output signal is shown in Fig. 6. In "Section A", the mechanical noise is mixed into the input signal. By using the proposed method, the mechanical noise is shown to be reduced. In

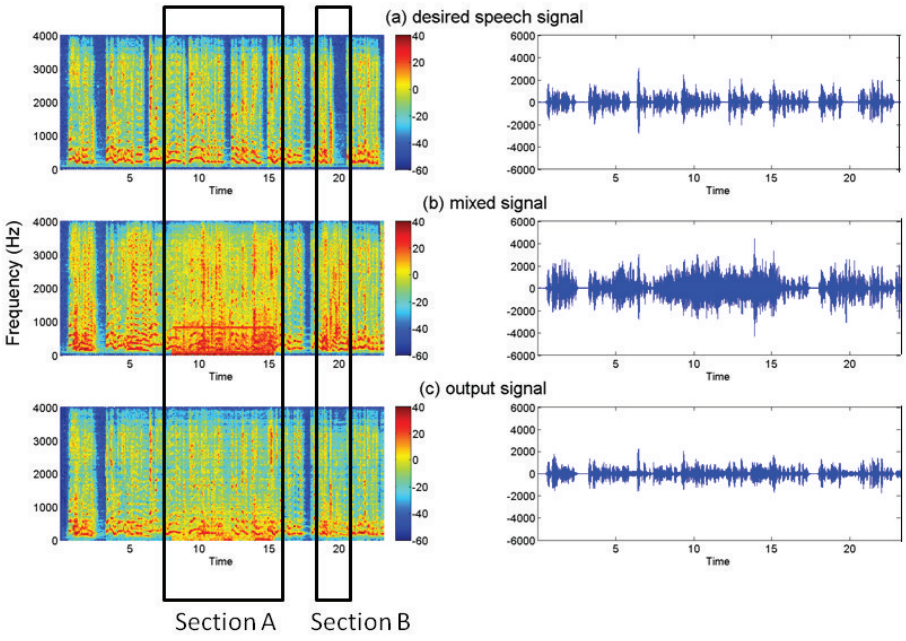

Fig. 6. A sample of output signal by proposed method

"Section B", the directional noise source is dominant, but the directional noise source is reduced in the output signal of the proposed method.

5. CONCLUSION

In this paper, we proposed a noise reduction method for a mechanical noise whose impulse response is time-variant. The multichannel noise covariance matrix of a mechanical noise is also regarded as a time-variant matrix. In the learning phase, the proposed method discritizes the multichannel noise covariance matrix by k-means clustering and obtained multiple noise covariance matrices. Unlike conventional multichannel noise reduction methods with single noise reduction filter, the proposed method use multiple noise reduction filter. Each noise reduction filter is made from the corresponding noise covariance matrix in the learning phase. In the noise reduction phase, the proposed method selects a filter which minimizes output power after noise reduction, and the output signal of this filter is regarded as output signal of the proposed method. We applied the proposed method for noise reduction of a digital camera and a communication robot. Experimental results show that the proposed method is superior to conventional maximum SNR beamformer with single noise reduction filter especially at lower SNR.

REFERENCES

[1] I. Cohen, "Optimal speech enhancement under signal presence uncertainty using log-spectral amplitude estimator," IEEE Signal Processing Lett., vol. 9, no. 4, pp. 113-116, Apr. 2002.

[2] A. Abramson and I. Cohen, "Enhancement of speech signals under multiple hypotheses using an indicator for transient noise presence," *Proc. ICASSP2007*, vol. IV, pp. 553-556, 2007.

[3] O. L. Frost, III, "An algorithm for linearly constrained adaptive array processing," In *Proc. IEEE*, vol. 60, no. 8, pp. 926-935, 1972.

[4] L. J. Griffith and C. W. Jim, "An alternative approach to linearly constrained adaptive beamforming," *IEEE Trans. AP*, vol.30, i.1, pp.27-34, 1982.

[5] O. Hoshuyama, A. Sugiyama, and A. Hirano, "A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters," *IEEE Trans. SP,* vol. 47, no. 10, pp. 2677-2684, Oct. 1999

[6] S. Gannot, D. Burshtein, and E. Weinstein, "Signal enhancement using beamforming and non-stationarity with applications to speech," *IEEE Trans. SP*, vol. 49, no. 8, pp. 1614–1626, 2001.

[7] A. Spriet, M. Moonen, and J. Wouters, "Spatially pre-processed speech distortion weighted multi-channel Wiener filtering for noise reduction," *Signal Processing*, vol. 84, no. 12, pp. 2367–2387, 2004.

[8] A. Hyvärinen, J. Karhunen, and E. Oja, "Independent component analysis," John Wiley & Sons, 2001.

[9] J. Even, H. Saruwatari, K. Shikano, "Frequency domain semi-blind signal separation: applications to the rejection of internal noises," *IEEE ICASSP 2008*, pp. 157-160, 2008.

[10] M. Matsumoto, T Abe, and S. Hashimoto, "Internal noise reduction combining microphones and a pezoelectric device under blind condition," *IEEE MFIIS 2008*, pp. 498-502, 2008.

[11] E. Warsitz and R. Haeb-Unbach, "Controlling speech distortion in adaptive frequency-domain principal eigenvector beamforming," in *Proc. IWAENC 2006*, 2006.

[12] S. Araki, H. Sawada, and S. Makino, "Blind Speech Separation in a Meeting Situation," *Proc. ICASSP2009*, vol. I, pp. 41–45, 2007.

[13] M. Togami, Y. Obuchi, and A. Amano, "Automatic Speech Recognition of Human-Symbiotic Robot EMIEW," in "Human-Robot Interaction", pp. 395-404, I-tech Education and Publishing, 2007.

[14] Ö. Yılmaz and S. Rickard, "Blind separation of speech mixtures via time frequency masking," *IEEE* Trans. SP, vol.52, no.7, pp. 1830-1847, Jul. 2004.

[15] M. Togami, A. Amano, T. Sumiyoshi, and Y. Obuchi, "DOA estimation method based on sparseness of speech sources for human symbiotic robots," *Proc. ICASSP2009*, pp. 3693–3696, 2009.

[16] C. Fevotte, R. Gribonval, and E. Vincent, "BSS_EVAL toolbox user guide -Revision 2.0," Tech. Rep. 1706, IRISA, 2005.