ENHANCEMENT OF SOUND SOURCE SURROUNDED BY AMBIENT NOISE USING PAIR OF MICROPHONE ARRAYS

[1]*Yusuke Hioka, Kazunori Kobayashi, Ken'ichi Furuya, and Akitoshi Kataoka*

[1]NTT Cyber Space Laboratories, NTT Corporation
3-9-11 Midori-cho, Musashino, Tokyo 180-8585, Japan

ABSTRACT

We propose a method for ambient noise reduction using a pair of small microphone arrays separated from each other to avoid using a large aperture array. The proposed method includes the idea of applying a nonlinear filter in the frequency domain, which is designed from several beamforming outputs in order to use the spatial information about sound sources, to suppress noise components. This idea enables the suppression of ambient noise that comes from behind the target using *small* microphone arrays, which has been a challenging issue. Thus, noise reduction performance is expected to be improved.

We performed experiments in a room with slight reverberation and confirmed that the method succeeds in suppressing the ambient noise to about 13dB, which is a much better effect than that of the conventional beamforming method using a large-aperture microphone array.

1. INTRODUCTION

With the popularization of videophone and videoconference systems, hands-free microphone systems have become indispensable devices due to their convenience. Compared to conventional handset systems, the hands-free system often suffers from SNR deterioration caused by various types of ambient noises. Using a microphone array [1] is a well-known method of recovering degraded speech signals that exploits the spatial information of each sound source for separating signal sources. Various ambient noise sources are normally scattered behind the target speaker. Therefore, ambient noise sources and the target source are often located in the same direction but at different distances from the array.

Although many previous studies on noise reduction using a microphone array have been performed, most of them are only designed to separate sources located in different directions [2]. Nevertheless, some studies have worked on the reduction of noise sources located in the same direction. The fixed beamforming has an advantage in its stability, that is, robustness against changes in the location of noise sources. Optimal weights in filtering equations for fixed beamforming have been proposed [3] [4]. However, the reduction in noise level is limited to a few dBs. Meanwhile, adaptive beamforming achieves better noise reduction performance, but it is sometimes severely degraded by noise source movement due to the convergence speed falls behind the noise position changes. There is also a method based on ICA (Independent Component Analysis) [5]. This method assumes the "blind" situation, so source and microphone positions *a priori* need not be known. However, their paper does not refer to performance in an actual situation. Furthermore, due to a theoretical reason, the array aperture should be sufficiently large to discriminate the distances from sound sources. This inherent limitation causes difficulties in carrying and setting up the array for use.

Motivated by this situation, the purpose of this work is to achieve (A) sufficient reduction of ambient noise located behind the target using small microphone arrays, and (B) less sensitivity to noise position changes.

In this work, we use two small microphone arrays separated from each other to avoid using a large aperture array. The proposed method includes the idea of applying a nonlinear filter in the frequency domain, which is designed from several beamforming outputs in order to use the spatial information about sound sources, to suppress noise components. This idea enables the suppression of ambient noise that comes from behind the target, and thus, we expect the noise reduction performance to be improved. Furthermore, stability is assured because no adaptive processing is applied.

This report is organized as follows. In Sec. 2, we define the problem of positioning microphone arrays as well as principles of the proposed method. Then, we explain detailed procedures of the proposed method in Sec. 3, and the experimental results are shown in Sec. 4. Finally, in Sec. 5, we conclude this report with some comments.

2. SPATIAL NONLINEAR FILTERING

2.1. Problem Definition

The problem discussed in this study is to record a target source signal while suppressing various ambient noise sources and types of noise, as shown in Fig. 1. To facilitate carrying and setting up the array, we use a pair of microphone arrays, array-L and array-R, whose aperture sizes are sufficiently small, e.g., 15cm, and they are roughly placed apart each other. The desirable inter-array distance is approximately same as the distance between the arrays and the target source, but anyway it is set to be unknown. The target signal and each noise are assumed to be uncorrelated. We also set some restrictions such that each microphone array knows the target source direction *a priori*, and the distance between the target and each array is nearly equal. The restrictions did not cause a loss in generality in determining the locations and types of noise sources.

As we noted in the introduction, a small aperture array cannot separate a signal arriving from the same direction as that of sound sources. The proposed method designs a nonlinear filter in the frequency domain by the use of various fixed beamforming outputs. The remainder of this section describes the principle

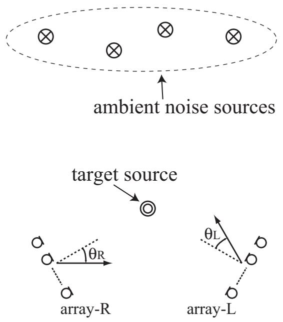

Figure 1: Problem definition

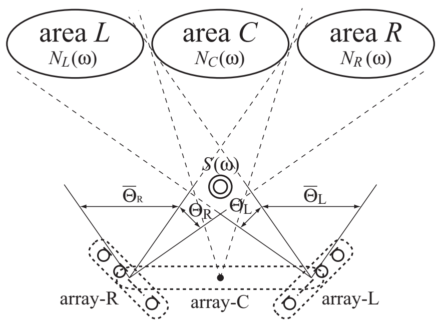

Figure 2: Area division using beamforming

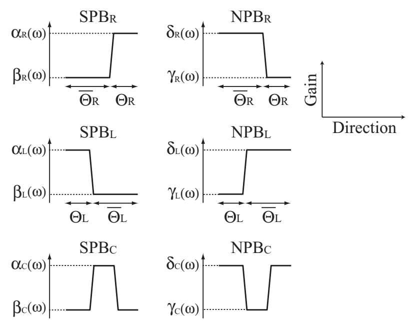

Figure 3: Spatial responses of beamformers

of noise suppression using the nonlinear filter calculation.

2.2. Signal Enhancement Filter

As shown in Fig. 2, we assume that every noise source is located in one of three areas, R, C, or L that are located sufficiently farther from the array than the target source. From the perspective of array-L or array-R, areas L or R, respectively, are located in the same directions as that of the target, and area C is located between areas L and R. For signals received by an array, we assume that a signal from each noise source reaches all microphones at the same power level. Although this is a rough assumption, generality is not lost because noise sources are located sufficiently far away to enable us to ignore the difference in propagation path length. Thus, if the input signal of arbitrary microphone m is represented as

$$Z_m(\omega) = S(\omega) + N_L(\omega) + N_C(\omega) + N_R(\omega), \quad (1)$$

where $N_{\mathtt{a}}(\omega)$ ($\mathtt{a} \in \{R, L, C\}$) denotes the summation of received signals from area $\mathtt{a}$, we can approximate the power of array input signal $|Z(\omega)|^2$ using the terms in the right side of Eq.(1) due to the uncorrelation of target and noise signals, given by

$$\begin{aligned} |Z(\omega)|^2 &\simeq |Z_m(\omega)|^2 \\ &\simeq |S(\omega)|^2 + |N_L(\omega)|^2 + |N_C(\omega)|^2 + |N_R(\omega)|^2. \quad (2) \end{aligned}$$

Based on these models, our goal is estimating the optimal nonlinear filter, $G(\omega)$, defined by

$$\begin{aligned} G(\omega) &= \sqrt{\frac{|S(\omega)|^2}{|Z(\omega)|^2}} \\ &\simeq \sqrt{\frac{|S(\omega)|^2}{|S(\omega)|^2 + |N_L(\omega)|^2 + |N_C(\omega)|^2 + |N_R(\omega)|^2}}. \quad (3) \end{aligned}$$

Multiplying $G(\omega)$ by the magnitude of the array input signal $|Z(\omega)|$, we obtain the power spectrum of the target signal without noise.

2.3. Power Spectra Estimation Using Beamforming Outputs

The powers of $S(\omega)$ and $N_{\mathtt{a}}(\omega)$ are required for estimating the $G(\omega)$. Now, we introduce two fixed beamformers called SPB (signal pass beamformer) and NPB (noise pass beamformer), whose spatial responses are designed to be complementary to each other. They are applied to array-L and array-R respectively, and furthermore, another set of them is applied to array-C, which consists of the centre microphones of array-L and array-R. If beamformers are designed to have ideally flat spatial responses, as shown in Fig. 3, the powers of the beamformer outputs in the vector form are given by Eq.(4).

$$\underbrace{\begin{bmatrix} |Y_{\mathrm{SL}}(\omega)|^2 \\ |Y_{\mathrm{NL}}(\omega)|^2 \\ |Y_{\mathrm{SR}}(\omega)|^2 \\ |Y_{\mathrm{NR}}(\omega)|^2 \\ |Y_{\mathrm{SC}}(\omega)|^2 \\ |Y_{\mathrm{NC}}(\omega)|^2 \end{bmatrix}}_{\mathbf{Y}(\omega)} \simeq \underbrace{\begin{bmatrix} \alpha_L^2 & \alpha_L^2 & \beta_L^2 & \beta_L^2 \\ \gamma_L^2 & \gamma_L^2 & \delta_L^2 & \delta_L^2 \\ \alpha_R^2 & \beta_R^2 & \beta_R^2 & \alpha_R^2 \\ \gamma_R^2 & \delta_R^2 & \delta_R^2 & \gamma_R^2 \\ \alpha_C^2 & \beta_C^2 & \alpha_C^2 & \beta_C^2 \\ \gamma_C^2 & \delta_C^2 & \gamma_C^2 & \delta_C^2 \end{bmatrix}}_{\mathbf{T}(\omega)} \underbrace{\begin{bmatrix} |S(\omega)|^2 \\ |N_L(\omega)|^2 \\ |N_C(\omega)|^2 \\ |N_R(\omega)|^2 \end{bmatrix}}_{\mathbf{Z}(\omega)} \quad (4)$$

For simplicity, we omitted the frequency variable (ω) from every component of $\mathbf{T}(\omega)$. In practice, such flat spatial responses cannot be achieved, so we defined $\mathbf{Y}(\omega)$ and $\mathbf{T}(\omega)\mathbf{Z}(\omega)$ as being $\simeq$, nearly equal. The components of $\mathbf{Y}(\omega)$ are derived from beamformer outputs, and each component of $\mathbf{T}(\omega)$ is calculated from the beampattern *a priori*. Therefore, we can calculate the estimated value of $\mathbf{Z}(\omega)$ by using Eq.(5).

$$\begin{aligned} \overline{\mathbf{Z}(\omega)} &= \left[\, |\overline{S(\omega)}|^2 \ |\overline{N_L(\omega)}|^2 \ |\overline{N_C(\omega)}|^2 \ |\overline{N_R(\omega)}|^2 \,\right]^T \\ &= \mathbf{T}^+(\omega)\mathbf{Y}(\omega), \quad (5) \end{aligned}$$

Here, the superscript $^+$ denotes the pseudo inverse and $\overline{\cdot}$ means the estimated value.

3. PROCEDURES OF THE PROPOSED METHOD

The signal flow of the proposed method is shown in Fig. 4. The method consists of two parts, “beamforming” and “nonlinear filtering”. For simplicity, we express signals in the frequency domain.

3.1. Input Signal Modeling

The input signal vectors of array-L and array-R are defined by

$$\mathbf{X}_L(\omega) = \left[X_{L1}(\omega) \cdots X_{LM_L}(\omega)\right]^T \quad (6)$$

$$\text{and} \quad \mathbf{X}_R(\omega) = \left[X_{R1}(\omega) \cdots X_{RM_R}(\omega)\right]^T, \quad (7)$$

where M_L and M_R are the numbers of microphones of array-L and array-R respectively.

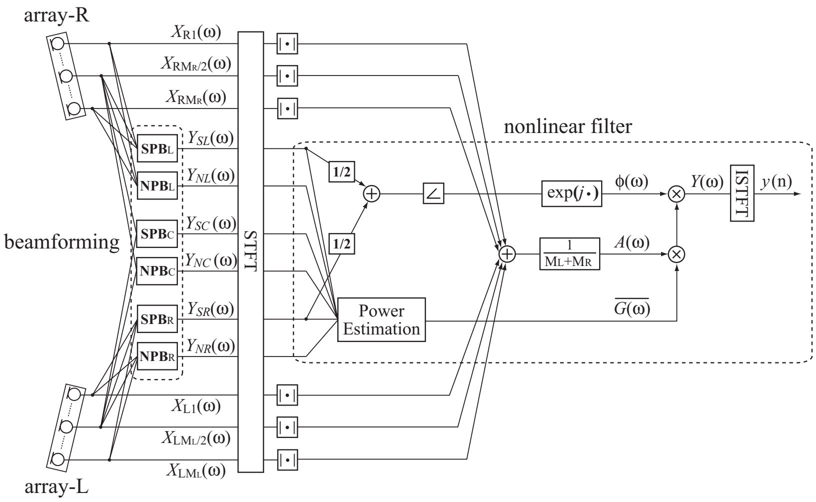

Figure 4: Signal flow of proposed method

3.2. Beamforming

In the beamforming stage, we applied three sets of complementary beamformers, SPB and NPB, designed to imitate the spatial responses for array-L, array-R, and array-C, as shown in Fig. 3.

The output signals of the beamformers for the array-L and array-R are given by

$$\begin{aligned} Y_{SL}(\omega) &= \mathbf{W}^H_{\mathrm{SPB_L}}(\omega)\mathbf{X}_L(\omega), && (8)\\ Y_{NL}(\omega) &= \mathbf{W}^H_{\mathrm{NPB_L}}(\omega)\mathbf{X}_L(\omega), && (9)\\ Y_{SR}(\omega) &= \mathbf{W}^H_{\mathrm{SPB_R}}(\omega)\mathbf{X}_R(\omega), && (10)\\ Y_{NR}(\omega) &= \mathbf{W}^H_{\mathrm{NPB_R}}(\omega)\mathbf{X}_R(\omega). && (11)\end{aligned}$$

$\mathbf{W}_{\mathrm{SPB}_\mathtt{b}}(\omega)$ and $\mathbf{W}_{\mathrm{NPB}_\mathtt{b}}(\omega)$ denote the weight vectors of respective $\mathrm{SPB}_\mathtt{b}$ and $\mathrm{NPB}_\mathtt{b}$ ($\mathtt{b} \in \{\mathrm{L}, \mathrm{R}\}$), which is designed so that its spatial response,

$$\begin{aligned} H_{\mathrm{SPB}_\mathtt{b}}(\omega,\theta_\mathtt{b}) &= \mathbf{W}^H_{\mathrm{SPB}_\mathtt{b}}(\omega)\cdot \mathbf{e}_\mathtt{b}(\omega,\theta_\mathtt{b}), && (12)\\ H_{\mathrm{NPB}_\mathtt{b}}(\omega,\theta_\mathtt{b}) &= \mathbf{W}^H_{\mathrm{NPB}_\mathtt{b}}(\omega)\cdot \mathbf{e}_\mathtt{b}(\omega,\theta_\mathtt{b}), && (13)\end{aligned}$$

follows the constraints given by

$$\begin{aligned} &H_{\mathrm{SPB}_\mathtt{b}}(\omega,\theta_\mathtt{b})|_{\theta_\mathtt{b}\in\Theta_\mathtt{b}} = 1 \quad \text{and} \quad H_{\mathrm{SPB}_\mathtt{b}}(\omega,\theta_\mathtt{b})|_{\theta_\mathtt{b}\in\overline{\Theta_\mathtt{b}}} = 0\\ &H_{\mathrm{NPB}_\mathtt{b}}(\omega,\theta_\mathtt{b})|_{\theta_\mathtt{b}\in\Theta_\mathtt{b}} = 0 \quad \text{and} \quad H_{\mathrm{NPB}_\mathtt{b}}(\omega,\theta_\mathtt{b})|_{\theta_\mathtt{b}\in\overline{\Theta_\mathtt{b}}} = 1,\end{aligned}$$

where $\mathbf{e}_\mathtt{b}(\omega,\theta_\mathtt{b})$ denotes the steering vector of array-$\mathtt{b}$ and the superscript ^{H} is the Hermitian conjugate.

On the other hand, simple summation and subtraction beamforming is applied to array-C.

$$\begin{aligned} Y_{SC}(\omega) &= \tfrac{1}{2}(X_{LM_L/2}(\omega) + X_{RM_R/2}(\omega)) && (14)\\ Y_{NC}(\omega) &= X_{LM_L/2}(\omega) - X_{RM_R/2}(\omega) && (15)\end{aligned}$$

From the weight vectors of the beamformers given above, we derive each component of $\mathbf{T}(\omega)$ by calculating the average of the magnitude of spatial response given by

$$\begin{aligned} \alpha_\mathtt{b}(\omega) &= E[H_{\mathrm{SPB}_\mathtt{b}}(\omega,\theta_\mathtt{b})]|_{\theta_\mathtt{b}\in\Theta_\mathtt{b}} && (16)\\ \beta_\mathtt{b}(\omega) &= E[H_{\mathrm{SPB}_\mathtt{b}}(\omega,\theta_\mathtt{b})]|_{\theta_\mathtt{b}\in\overline{\Theta_\mathtt{b}}} && (17)\\ \gamma_\mathtt{b}(\omega) &= E[H_{\mathrm{NPB}_\mathtt{b}}(\omega,\theta_\mathtt{b})]|_{\theta_\mathtt{b}\in\Theta_\mathtt{b}} && (18)\\ \delta_\mathtt{b}(\omega) &= E[H_{\mathrm{NPB}_\mathtt{b}}(\omega,\theta_\mathtt{b})]|_{\theta_\mathtt{b}\in\overline{\Theta_\mathtt{b}}}. && (19)\end{aligned}$$

3.3. Nonlinear Filter

Nonlinear filter is estimated by substituting components in the $\overline{\mathbf{Z}(\omega)}$, which is estimated from beamforming outputs, into Eq.(3), given by

$$\overline{G(\omega)} = \sqrt{\frac{|\overline{S(\omega)}|^2}{|\overline{S(\omega)}|^2 + |\overline{N_L(\omega)}|^2 + |\overline{N_C(\omega)}|^2 + |\overline{N_R(\omega)}|^2}}. \quad (20)$$

The magnitude of the final output signal is derived by multiplying $\overline{G(\omega)}$ and the average magnitude of the array input signal $A(\omega)$ while using $\mathrm{SPB}_\mathtt{b}$ outputs for the phase components of $Y(\omega)$:

$$Y(\omega) = \overline{G(\omega)}\cdot A(\omega)\cdot \exp(j\phi(\omega)), \quad (21)$$

where

$$\begin{aligned} A(\omega) &= \frac{\sum_{m_L}|X_{m_L}(\omega)| + \sum_{m_R}|X_{m_R}(\omega)|}{M_R+M_L}\\ \text{and}\quad \phi(\omega) &= \angle(Y_{SR}(\omega) + Y_{SL}(\omega)).\end{aligned}$$

4. EXPERIMENTAL RESULTS

To verify the effectiveness of the proposed method, some experiments were performed in a room with reverberation (T_{60} = 150[ms]). The arrangement for sound sources and microphone arrays is shown in Fig. 5 and Table 1 gives parameter settings. In the conventional method for comparison, we use the delay-and-sum beamformer [3] giving the distance between array-L and array-R, i.e., large aperture microphone array. For the quantitative evaluation, we adopt the NRR (Noise Reduction Rate) [6], which is adapted to nonlinear filtering defined in [7].

As seen in Figs. 6 and 7, the proposed method suppresses noise signals that are not sufficiently removed by conventional methods. A small difference can be found in comparing the results of Cases I and II obtained by using the proposed method. This finding indicates that the method is robust against noise source movement.

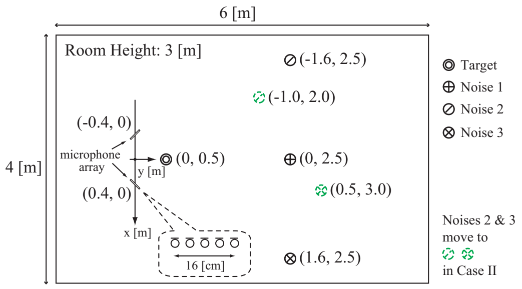

Figure 5: Sound source and microphone array arrangement

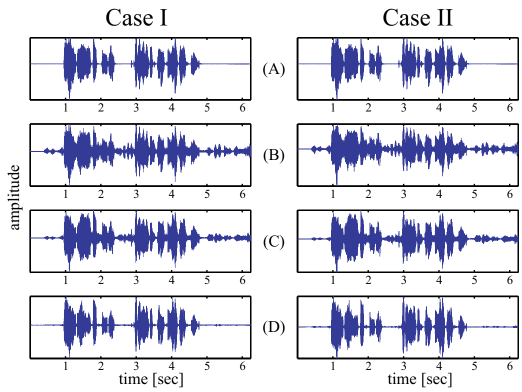

Figure 6: Sound waveforms (A) original target signal (B) measured signal (C) delay-sum beamforming (D) proposed method

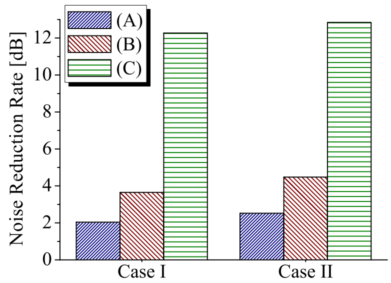

Figure 7: Noise reduction rate (A)delay-sum (B)SPB output (C)proposed method

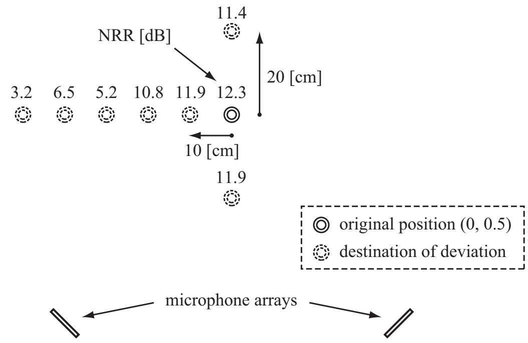

Figure 8: NRRs for the target source deviated from the original position

In addition, the effect of target source movement in Case I is investigated. The result shown in Fig. 8 indicates that the method allows slight target deviation, but once the target moves outside the area, the performance deteriorates rapidly.

5. CONCLUSION

A method for ambient noise reduction using a pair of microphone arrays was proposed. The method estimates the nonlinear filter for reducing noise using several beamformer outputs. According to experimental results, we confirmed that the proposed method achieves better noise reduction performance than that of conventional beamforming. Future subjects of this research are to improve the noise reduction rate and to enlarge the noise area that can be effectively suppressed. It is also required to evaluate the signal distortion caused by the proposed method quantitatively, and furthermore, some subjective evaluations need to be performed.

Table 1: Experimental parameters

Sampling Rate	16 kHz	Frame Length	256 sample
M_R	5	Frame Shift	128 sample
M_L	5	FFT Points	256 sample

6. REFERENCES

[1] M. Brandstein and D. Ward, “Microphone Arrays,” Springer-Verlag, Berlin, 2001.

[2] O. Hoshuyama, B. Begasse, and A. Sugiyama, “A New Adaptation-Mode Control Based on Cross Correlation for a Robust Adaptive Microphone Array,” IEICE Trans. Fundamentals, Vol.E84-A, No.2, pp.406–413, 2001.

[3] H. Nomura, Y. Kaneda, and J. Kojima, “Microphone array for near sound field,” Journal of the Acoustical Society of Japan, Vol. 53, No. 2, pp. 110–116, 1997. (in Japanese)

[4] J.G. Ryan and R.A. Goubran, “Near-Field Beamforming for Microphone Arrays,” Proc. of ICASSP1997, pp. 363–366, 1997.

[5] A. Ando, M. Iwaki, K. Ono, and K. Kurozumi, “Separation of Sound Sources Propagated in the Same Direction,” IEICE Trans. Fundamentals, Vol. E88-A, No. 7, pp. 1665–1672, 2005.

[6] S. Araki, S. Makino, T. Nishikawa, and H. Saruwatari, “Fundamental limitation of frequency domain blind source separation for convolved mixture of speech,” Proc. ICASSP2001, Vol. V, pp. 2737–2740, 2001.

[7] M. Aoki, K. Furuya, and A. Kataoka, “Improvement of “SAFIA”Source Separation Method under Reverberant Conditions,” Electronics & Communications in Japan, Part 3, Vol. 89, No. 3, pp. 22–37, 2006.