Noise reduction using an eyeglass-frame microphone array based on DOA estimation by LASSO

Kenji Ozawa[1], *Shion Yokouchi*[1], *Masanori Moirse*[1]

[1]Faculty of Engineering, University of Yamanashi, Japan
{ozawa, t13cs059, mmorise}@yamanashi.ac.jp

Abstract

In our previous study, we have proposed a practical method for noise reduction using a microphone array. The method initially estimates the direction of arrival (DOA) and waveform of a noise signal, then subtracts the estimated noise from the output of a reference microphone to restore a target signal. However, the method is effective only when there is one noise source. This study expands the method by using the least absolute shrinkage and selection operator (LASSO) algorithm. When there are multiple noise sources, the DOA of the most dominant noise is estimated by the LASSO to reduce the noise effectively. The results from computational simulation experiments show the efficiency of the proposed method when the microphone array is mounted on an eyeglass frame.

Index Terms: microphone array, DOA estimation, LASSO, eyeglass frame

1. Introduction

Microphone arrays are effective for improving speech intelligibility in a noisy environment. As for hearing aids, a simple installation of a microphone array is to mount it on an eyeglass frame. However, classical delay-and-sum (DAS) beamforming provides very small amounts of noise reduction at low frequencies because the array length is comparable to the wavelength of a low frequency component involved in speech. The performance has been progressed by using adaptive array processing [1] and the superdirective array technique [2]. These advanced processing techniques need a high calculation cost, while the cost should be reduced for portable devices such as hearing aids.

We have proposed a practical method to suppress noise using a microphone array [3]. However, the method is effective only when there is one noise source. This study expands the method by using the least absolute shrinkage and selection operator (LASSO) algorithm [4].

2. Proposed system

2.1. System structure

The basic structure of the expanded system is the same as that in [3]. Figure 1 shows the block diagram of the system when there are four microphones. The microphones are located on an eyeglass frame (Fig. 2) where the first one from an edge is defined as the reference microphone ($\mathrm{M_R}$). A sound arriving from perpendicular to the eyeglass frame is regarded as a target signal and sounds arriving from other directions are treated as noise signals. This system first estimates the DOA and waveform of a noise signal observed at the $\mathrm{M_R}$ independently across frequency bins, and then subtracts the estimated waveform of the noise signal from the $\mathrm{M_R}$ output to suppress the noise.

A signal observed at the $\mathrm{M_R}$, $M_\mathrm{R}(\omega)$, is a superimposition

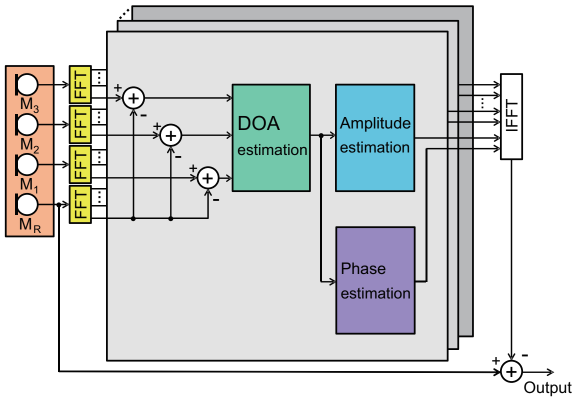


Figure 1: *Block diagram of the proposed system [3].*

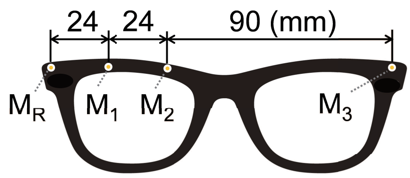


Figure 2: *Microphone array mounted on an eyeglass frame.*

of a target signal, $S(\omega)$, and a noise signal, $N(\omega)$.

$$M_\mathrm{R}(\omega) = S(\omega) + N(\omega) \tag{1}$$

If we assume that one noise sources is located far away from the array, the signal observed at the i-th microphone, $M_i(\omega)$, can be described as follows:

$$M_i(\omega) = S(\omega) + N(\omega)e^{-j\omega\tau_i} \tag{2}$$

where τ_i is the time delay from the $\mathrm{M_R}$. By subtracting $M_\mathrm{R}(\omega)$ from $M_i(\omega)$, the residual signal from the i-th microphone, $R_i(\omega)$, is given by

$$\begin{aligned} R_i(\omega) &= M_i(\omega) - M_\mathrm{R}(\omega) \\ &= N(\omega)\left(e^{-j\omega\tau_i} - 1\right) \\ &= 2\left|N(\omega)\right| \sin\left(-\frac{\omega\tau_i}{2}\right) e^{j\left(\mathrm{Arg}[N(\omega)] - \frac{\omega\tau_i}{2} + \frac{\pi}{2}\right)} \end{aligned} \tag{3}$$

where $|N(\omega)|$ and $\mathrm{Arg}[N(\omega)]$ describe the amplitude and phase of the noise, respectively. Thus, the amplitude and phase of $R_i(\omega)$ are given as follows:

$$\begin{aligned} \left|R_i(\omega)\right| &= 2\left|N(\omega)\right| \sin\left(-\frac{\omega\tau_i}{2}\right), \\ \mathrm{Arg}[R_i(\omega)] &= \mathrm{Arg}[N(\omega)] - \frac{\omega\tau_i}{2} + \frac{\pi}{2}. \end{aligned} \tag{4}$$

First, we estimate the DOA of the noise signal, θ_N, by

$$\theta_N = \sin^{-1}\left(\frac{-2c(\mathrm{Arg}[R_{i+1}(\omega)] - \mathrm{Arg}[R_i(\omega)])}{\omega(d_{i+1} - d_i)}\right) \quad (5)$$

where d_i denotes the distance between $\mathrm{M_R}$ and M_i [3].

Next, we estimate $N(\omega)$ based on Eq. (4) using $\tau_i = d_i \sin(\theta_N)/c$ where c is the velocity of sound.

$$\begin{aligned} |N(\omega)| &= \frac{|R_i(\omega)|}{2\sin\left(-\frac{\omega\tau_i}{2}\right)}, \\ \mathrm{Arg}[N(\omega)] &= \mathrm{Arg}[R_i(\omega)] + \frac{\omega\tau_i}{2} - \frac{\pi}{2} \end{aligned} \quad (6)$$

The above derivation indicates that the noise signal at the $\mathrm{M_R}$ is easily estimated. Thus, the calculation cost is low.

Finally, the waveform of the noise is synthesized by superimposing an inverse fast Fourier transform, IFFT[$N(\omega)$], of all frequency bins onto it, which is then subtracted from the output of the $\mathrm{M_R}$ to restore the target signal.

2.2. Expansion of the system

As described above, the DOA of noise, θ_N, is given by Eq. (5). When there are two noise sources at the same time, however, this equation estimates an intermediate direction of the two sources. As a result, the amplitude and phase of the superimposed noise cannot be estimated correctly by Eq. (6) and the performance of noise reduction is degraded.

To overcome this problem, this study expands the method to focus the most dominant noise in a temporal frame in signal processing. Because speech signals are sparse in the time-frequency domain, the dominant signal can be specified for every frequency bin in a temporal frame. Thus, the DOA of a dominant signal can be regarded as a function of frequency, $\theta_N(\omega)$. We decided to adopt the LASSO algorithm to specify the DOA of a dominant noise source. If the experimental conditions such as the number of microphones are enough, this algorithm estimates not only the DOA but also the amplitude and phase of the noise [5]. However, we estimate only the DOA of a dominant source because the present condition of four microphones with a short array seems not enough. Based on the estimated DOA, the amplitude and phase of the dominant noise are calculated by Eq. (6) for every frequency bin independently.

3. Evaluation of the proposed system

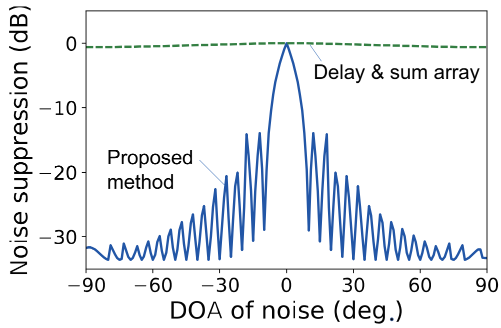
We conducted computer simulation experiments to evaluate the performance of the expanded system. The target, noise, and jammer signals were "Thank you very much. (female)", "Hello, hello. (male)", and "Welcome to Japan. (female)", respectively. Off-axis noises were made with phase-shifting digital filters. The sampling frequency was 16 kHz, and the FFTs were conducted for 512-point temporal frames with the hanning window (frame shift: 256 points). The noise DOAs were estimated independently across frequency bins for higher than 1500 Hz, and their median was set for lower frequency bins. The LASSO estimation was implemented using the function of MultiTaskLasso (alpha = 0.5) in a Python machine learning library, scikit-learn. The bases of LASSO consist of the following vectors:

$$\left[0,\ e^{-j\omega\tau_1} - 1,\ e^{-j\omega\tau_2} - 1,\ e^{-j\omega\tau_3} - 1\right]^T$$

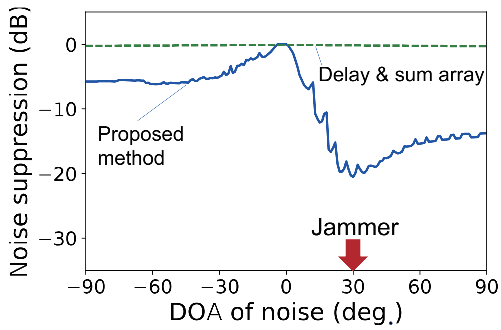
where these vectors were prepared for every 5°.

Figure 3 shows the obtained amount of noise suppression that is defined as the decrease in noise power in the $\mathrm{M_R}$ output. When there is one noise source, the performance is much better than the DAS beamforming (Fig. 3(a)). Because the LASSO bases were prepared in 5° steps, the amount of noise reduction is larger if the DOA of noise is near one of the bases. When another noise source (jammer) is fixed to 30°, the performance becomes poorer for the contralateral bearings (Fig. 3(b)). However, it is better than our previous system [3] in which little suppression is observed for the contralateral bearings. Thus, we can conclude that the expansion was succeeded.

Because the system was implemented in Python interpreter language to use a smart library of machine learning, the processing is not realtime operation. The next step of the project will be implementation of the system using a compiler language.



(a) There was one noise source.



(b) There were two noise sources where one was fixed at 30°.

Figure 3: *Performance of noise suppression as a function of the DOA of a noise signal.*

4. Acknowledgements

This work was supported by JSPS KAKENHI Grant (JP16K06384) and the Cooperative Res. Project (H28/A10) of the Res. Inst. of Electr. Commun., Tohoku University.

5. References

[1] J. E. Greenberg and P. M. Zurec, "Evaluation of an adaptive beamforming method for hearing aids," *J. Acoust. Soc. America*, vol. 91, no. 3, pp. 1663–1676, 1992.

[2] J. M. Kates, "Superdirective arrays for hearing aids," *J. Acoust. Soc. America*, vol. 94, no. 4, pp. 1930–1933, 1993.

[3] K. Ozawa, K. Amano, M. Morise, G. Shimizu, S. Oode and K. Ono, "Superdirective microphone array based on DOA and waveform estimations of noise," *Proc. of the 5th IEEE Global Conference on Consumer Electronics (GCCE 2016)*, pp. 119–120, 2016.

[4] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Royal Statistical Soc. Series B*, vol. 58, no. 1, pp. 267–288, 1996.

[5] A. Xenaki and P. Gerstoft, "Compressive beamforming," *J. Acoust. Soc. America*, vol. 136, no. 1, pp. 260–271, 2014.