Two-Channel-Based Noise Reduction in a Complex Spectrum Plane for Hands-Free Communication System

Toshiya Ohkubo, Tetsuya Takiguchi, and Yasuo Ariki

Department of Computer and Systems Engineering
Kobe University, Kobe, Japan
t-ohkubo@me.cs.scitec.kobe-u.ac.jp, {takigu,ariki}@kobe-u.ac.jp

Abstract. For hands-free communication system, this paper describes a noise reduction method using a 2-channel microphone. Recently, the Complex Spectrum Circle Centroid (CSCC) method has been proposed. This method utilizes geometric information and estimates the spectrum of the target signal. The method is advantageous in that no adjustment of the array-processing parameters to the environment is necessary before its operation and it is effective with non-stationary noise. However, the original CSCC method requires at least three microphones to estimate the spectrum of the target signal (center of circle). In this paper, we propose a method which estimates the spectrum of the target signal using only two microphones. In experimental results, the proposed method outperforms the Delay-and-Sum approach and can restore the target signal almost completely in a simulated noisy environment.

1 Introduction

A speech signal is available for hands-free communication system. However, in a real environment, the quality is degraded by the influence of the noise signals that are added to the target speech signal. Thus, it is necessary to reduce the noise signals and to enhance the target speech signals.

A popular method in noise reduction is Delay-and-Sum (DS) [1]-[3]. The advantage of this method is that DS does not require the training of the filter coefficients, but the DS method needs many microphones to improve its performance. Another method is an adaptive type of array processing [4]-[7], such as those proposed by Griffiths-Jim [8], AMNOR [9], where the training of the filter coefficients is required beforehand. The adaptive type methods can achieve better performance than that of DS, but if the test environment is different from the training, the performance decreases severely.

On the other hand, a technique called the Complex Spectrum Circle Centroid (CSCC) method [10] has been proposed recently. This method utilizes the geometric information of the target signal and the observed signal in a complex spectrum plane. This method can reduce noise without training before its operation and also achieve high performance. Furthermore, to process in each frame,

this method can be effective with non-stationary noise. However, this method needs at least three microphones to estimate the spectrum of the target signal. This means the method requires a special device, such as microphone array.

In this paper, we propose a method which estimates the spectrum of the target signal using only two microphones. The proposed method can reduce noise with high performance and achieve results as good as the CSCC method.

The organization of the paper is as follows. In section 2, we describe the basis of the CSCC method, followed by an explanation of the estimation process using only two microphones. In section 3, the experimental results are discussed. Finally, in section 4, we summarize the conclusions of this work.

2 Noise Reduction Processing in a Complex Spectrum Plane

2.1 The Layout of Observed Signals in a Complex Spectrum Plane

We assume that two acoustic signals, target and noise, propagate toward the microphones. The configuration of the two-microphone case is shown in Figure 1. The observed signal $m_i(t)$ of the i-th microphone is defined as follows:

$$m_i(t) = s(t) + n(t - \tau_i) \quad (i = 1...M) \tag{1}$$

where $s(t)$ is the target signal and $n(t)$ is the noise signal at time t, and τ_i denotes the time delay at the i-th microphone in respect to the noise signal, and M denotes the number of the microphones. The Fourier transform of the observed signal of the i-th microphone is described as follows:

$$M_i(\omega) = S(\omega) + N(\omega)e^{-j\omega\tau_i} \quad (i = 1...M) \tag{2}$$

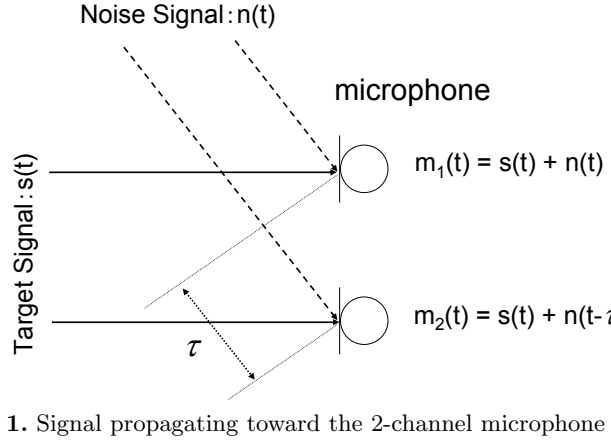


Fig. 1. Signal propagating toward the 2-channel microphone

where ω denotes angular frequency, and $M_i(\omega)$, $S(\omega)$ and $N(\omega)$ indicate Fourier transforms of $m_i(t)$, $s(t)$ and $n(t)$, respectively. Figure 2 gives a graphic representation of Equation (2). Each $M_i(\omega)$ lies on a circle with radius $\parallel N(\omega) \parallel$ and at a center $S(\omega)$. The value $\omega\tau_i$ denotes a deflection angle.

In the Complex Spectrum Circle Centroid (CSCC) method [10], the circle location is estimated by using only $M_i(\omega)$, and the center of the circle is the spectrum of the target signal.

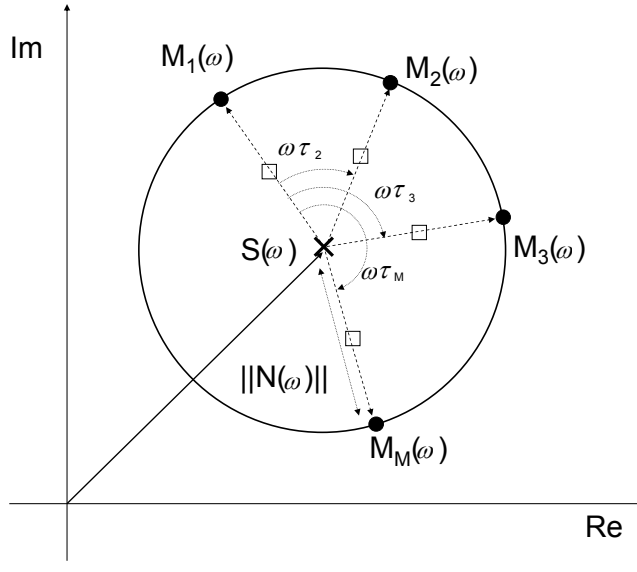


Fig. 2. Distribution of the observed signal by multiple microphones in a complex spectrum plane

2.2 Estimation of the Target Signal Spectrum Using a 2-channel Microphone

The original CSCC method requires at least three microphones because we need to estimate the location of the circle. This means that the method requires a special device, such as a microphone array.

Here, we propose a method to estimate the spectrum of the target signal using only two microphones. As shown in Figure 1, if the signals propagate as a plane wave, the spectrums of the signal observed using a 2-channel microphone are given as follows:

$$M_1(\omega) = S(\omega) + N(\omega) \tag{3}$$

$$M_2(\omega) = S(\omega) + N(\omega)e^{-j\omega\tau} \tag{4}$$

where $M_1(\omega)$ and $M_2(\omega)$ are the spectrums of the observed signal, $S(\omega)$ and $N(\omega)$ denote the spectrums of the target signal and the noise signal, respectively. The value τ denotes the time delay between the two microphones in respect to the noise signal.

As discussed in Section 2.1, $S(\omega)$ is located at an equal distance from $M_1(\omega)$ and $M_2(\omega)$, and the distance is $\parallel N(\omega) \parallel$. Subtracting Equation (4) from Equation (3) gives the value of $N(\omega)$ as

$$\parallel N(\omega) \parallel = \frac{\parallel M_1(\omega) - M_2(\omega) \parallel}{\parallel 1 - e^{-j\omega\tau} \parallel}. \tag{5}$$

Figure 3 shows the process used to estimate $S(\omega)$ using two microphones. First, we draw a perpendicular bisector toward a straight line connecting $M_1(\omega)$ and $M_2(\omega)$ in a complex spectrum plane. Next, we draw a circle with the radius $\parallel N(\omega) \parallel$ shown in Equation (5) and its center at $M_1(\omega)$. The coordinates of each spectrum in Figure 3 are defined as follows:

– The spectrum of the observed signal:

$$\begin{cases} M_1(\omega) = (M_{1x}, M_{1y}) \\ M_2(\omega) = (M_{2x}, M_{2y}) \end{cases}$$

– The candidate for the target signal spectrum:

$$\tilde{S}(\omega) = \{S_1(\omega), S_2(\omega)\} \ , \ \begin{cases} S_1(\omega) = (S_{1x}, S_{1y}) \\ S_2(\omega) = (S_{2x}, S_{2y}) \end{cases}$$

– The midpoint:

$$C(\omega) = (C_x, C_y) = (\frac{M_{1x} + M_{2x}}{2}, \frac{M_{1y} + M_{2y}}{2})$$

where subscript x and y denote the coordinates of the real part and the imaginary part, respectively.

The perpendicular bisector and the circle are given as follows:

$$\tilde{S}_y(\omega) - C_y(\omega) = \frac{M_{1x}(\omega) - M_{2x}(\omega)}{M_{2y}(\omega) - M_{1y}(\omega)}(\tilde{S}_x(\omega) - C_x(\omega)) \tag{6}$$

$$(\tilde{S}_x(\omega) - M_{1x}(\omega))^2 + (\tilde{S}_y(\omega) - M_{1y}(\omega))^2 = \parallel N(\omega) \parallel^2 . \tag{7}$$

The spectrum of the target signal, $S(\omega)$, is located at the intersecting points between the perpendicular bisector and the circle. Hence, $S_1(\omega)$ and $S_2(\omega)$ are obtained by solving the simultaneous equations between Equation (6) and Equation (7). We replace the gradient in Equation (6) with d, which is shown as follows:

$$d = \frac{M_{1x}(\omega) - M_{2x}(\omega)}{M_{2y}(\omega) - M_{1y}(\omega)}. \tag{8}$$

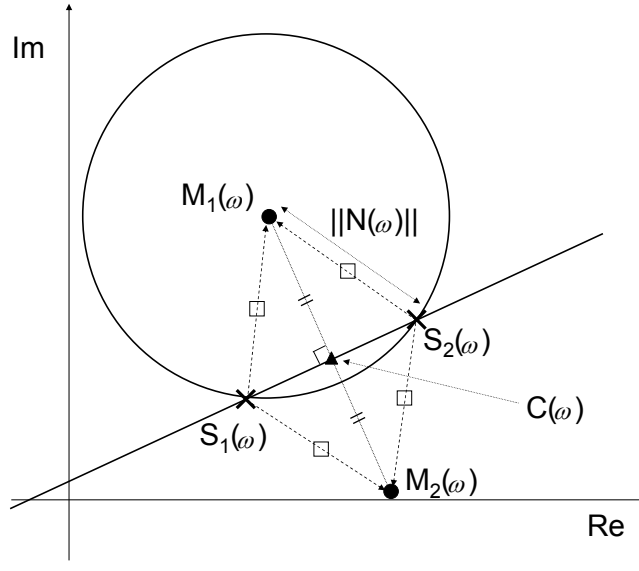


Fig. 3. The estimation process of the target signal spectrum using the 2-channel microphone in a complex spectrum plane

Using this equation, we define the constants as follows :

$$\begin{cases} a = 1 + d^2 \\ b = -2(1 + d^2)C_x(\omega) \\ c = (dC_x(\omega) - C_y(\omega) + M_{1y}(\omega))^2 - \parallel N(\omega) \parallel^2, \end{cases} \tag{9}$$

and calculate $\tilde{S}_x(\omega)$ as follows:

$$\tilde{S}_x(\omega) = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}. \tag{10}$$

Substituting the obtained $\tilde{S}_x(\omega)$ into Equation (6), we are able to obtain $\tilde{S}_y(\omega)$.

Finally, we must choose the proper spectrum from two among the candidates for the target signal. In this paper, we chose the candidate whose spectrum power is smaller, since we considered that the power of the estimated clean signal will be smaller than that of the observed noisy signal. In the case shown in Figure 3, $S_1(\omega)$ is chosen as the target signal spectrum.

3 Experiments

3.1 Experimental Conditions

To evaluate the proposed method, we used two evaluation measures. One is a cross-correlation value ($CC\text{-}Value$) between the target signal and the noise

removed signal. It is defined as follows:

$$CC\text{-}Value = \frac{\sum_t^T s(t)\tilde{s}(t)}{\sqrt{\sum_t^T (s(t))^2}\sqrt{\sum_t^T (\tilde{s}(t))^2}} \tag{11}$$

where T denotes the length of the signal, and t is the variable of time. $\tilde{s}(t)$ is the noise removed signal.

The other evaluation measure for performance in speech recognition is the cepstrum distance ($CepDist$), which is defined as follows:

$$CepDist = \frac{1}{N}\frac{1}{P}\sum_n^N\sum_p^P(\parallel S(n,p) - \tilde{S}(n,p) \parallel) \tag{12}$$

where N and P denote the total number of analysis frames and the dimension of the cepstrum, respectively, and n and p are the variables of the frames and the cepstrum dimension, respectively.

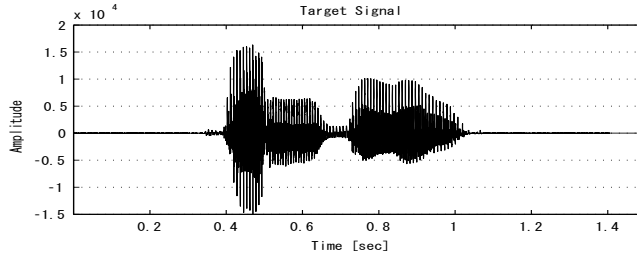
In the experiments, the target source and the noise source were located at 90 degrees and 30 degrees from the line connecting the microphones, respectively. The microphones were uniformly spaced at 2.83-cm intervals. We used 10 Japanese utterances in the "ATR SPEECH DATABASE" as the target signals and three other utterances in the same database as the noise signals, and mixed them to produce the observed signals with a signal-to-noise ratio (SNR) of 0 dB.

The observed signal was sampled at 16 kHz and windowed with a 20-ms Hamming window every 10-ms. Then a 320-point DFT was used to compute 16-order MFCCs (mel-frequency cepstral coefficients). We compared the performance of the conventional method of Delay-and-Sum (using 2 or 14 microphones).

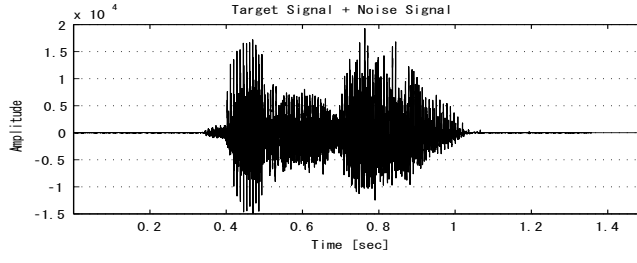
The proposed method requires the noise source direction to calculate $\parallel N(\omega) \parallel$ in Equation (5). The direction can be estimated by the inference method of the sound source, such as Cross-power Spectrum Phase analysis (CSP). However, in this experiments, we gave the noise source direction to evaluate the noise reduction performance of the proposed method.

3.2 Simulated Environment

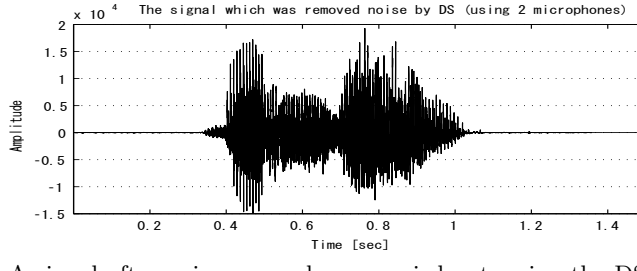
In this experiment, microphone-array data was generated by simulation considering only the time delay. Therefore, the target signals and the noise signals are propagated toward the microphones without degradation. Table 1 and Table 2 show the results for $CC\text{-}Value$ and $CepDist$. From these results, we see that the proposed method outperforms the Delay-and-Sum method. An example of the signal waveforms is shown in Figure 4. This example uses an utterance as the noise signal. And another example of the signal waveforms which uses a music as the noise is shown in Figure 5. These result shows that the proposed method is effective for the noise both in speech regions and non-speech regions. In addition, these results lead us to the conclusion that the proposed method can restore the target signal almost completely.



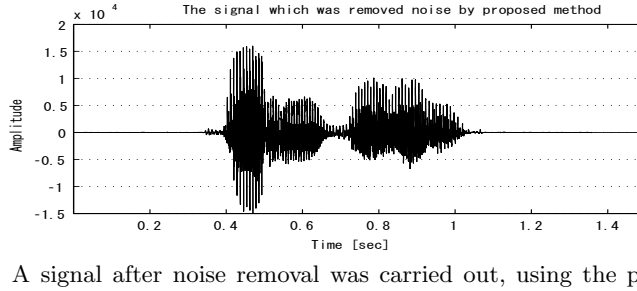
(a) A target signal (the utterance is /danbou/)



(b) An observed signal (the utterance of noise signal is /shisyuu/) (SNR:0dB)

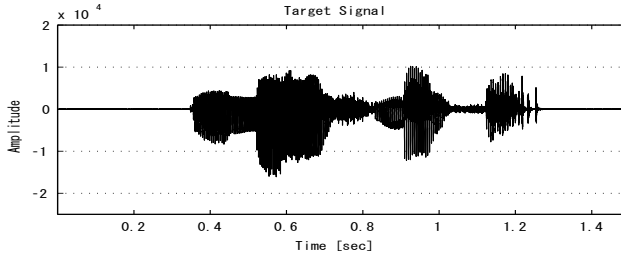


(c) A signal after noise removal was carried out, using the DS(mic:2) method

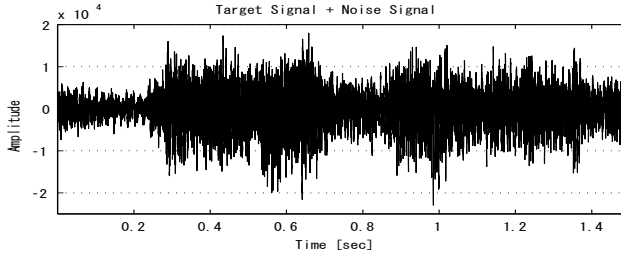


(d) A signal after noise removal was carried out, using the proposed method

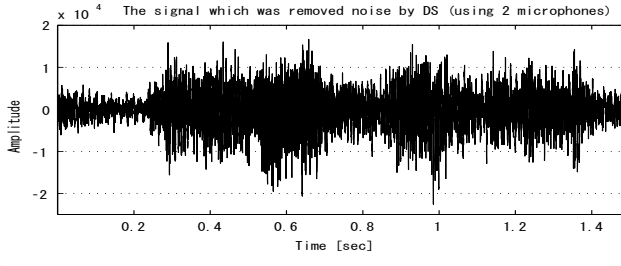
Fig. 4. The waveforms of a target signal, an observed signal and a noise removed signal. (The noise is an utterance.)



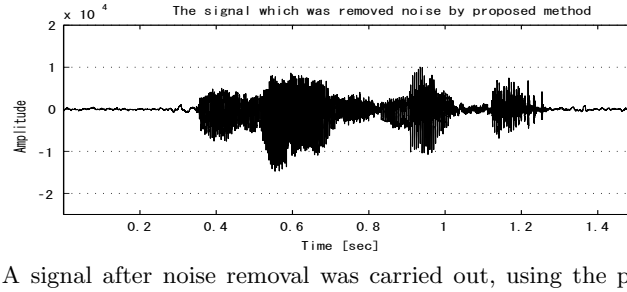
(a) A target signal (the utterance is /mimiwosumase/)



(b) An observed signal (the noise signal is a music) (SNR:0dB)



(c) A signal after noise removal was carried out, using the DS(mic:2) method



(d) A signal after noise removal was carried out, using the proposed method

Fig. 5. The waveforms of a target signal, an observed signal and a noise removed signal. (The noise is a music.)

Table 1. Comparison of cross-correlation value when using simulation data.

Data No.	Observed Signal	DS(microphone:2)	DS(microphone:14)	Proposed Method
1	0.69	0.70	0.79	**0.96**
2	0.72	0.73	0.80	**0.94**
3	0.71	0.72	0.84	**0.99**
4	0.66	0.67	0.81	**0.96**
5	0.68	0.69	0.80	**0.97**
6	0.65	0.66	0.81	**0.99**
7	0.71	0.75	0.82	**0.98**
8	0.70	0.74	0.81	**0.99**
9	0.70	0.74	0.81	**0.91**
10	0.69	0.70	0.79	**0.97**
Average	0.69	0.71	0.81	**0.97**

*The figure in () means the number of microphones.

Table 2. Comparison of cepstrum distance when using simulation data.

Data No.	Observed Signal	DS(microphone:2)	DS(microphone:14)	Proposed Method
1	9.18	8.88	6.38	**4.44**
2	8.04	7.65	5.61	**4.06**
3	12.46	12.24	11.34	**6.97**
4	8.60	8.62	7.51	**5.68**
5	8.16	8.35	6.73	**5.13**
6	9.28	9.31	8.11	**5.93**
7	9.66	9.26	5.58	**3.40**
8	9.91	9.63	6.69	**4.98**
9	9.38	9.02	5.46	**3.60**
10	8.54	8.12	5.76	**4.01**
Average	9.32	9.11	6.92	**4.82**

*The figure in () means the number of microphones.

Also an example of the spectrum estimation of the target signal in a complex spectrum plane is shown in Figure 6. In Figure 6, the points of the observed signal spectrum by a 2-channel micorphone are shown as circle symbols, and the candidates of the target signal spectrum which estimated by the observed signal spectrums are shown as square symbols. Here, we chose the bottom-left spectrum as the target signal from the candidates because its power is smaller than the other, as shown in Figure 6. It should be noticed that the estimated target spectrum located in the almost same location as the spectrum of the real target signal which are shown as the cross symbol.

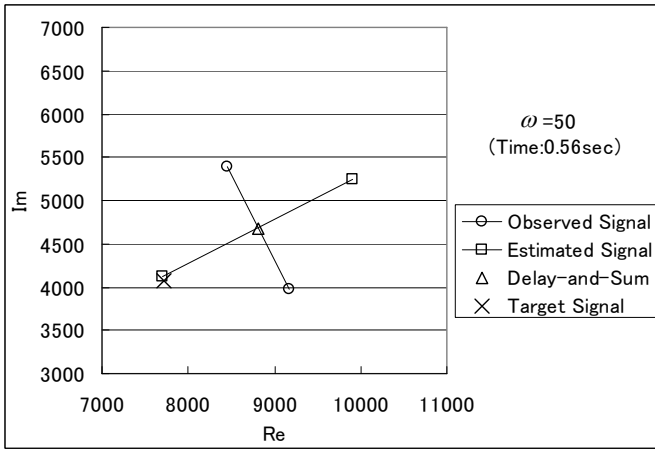


Fig. 6. An example of the target signal spectrum in a complex spectrum plane estimated from the simulated noisy data.

3.3 Reverberant Environment

In an experiment in a reverberant environment, we produced the observed signals with reverberant acoustic characteristic using the following steps. First, we convoluted the impulse responses with the target signals and the noise signals. Next, we added the noise signals to the target signals at each microphone. We used impulse responses from "RWCP Sound Scene Database in Real Acoustical Environments" [12], where the reverberation time was 300-ms and the distance between the sound source and the microphone was 2 meters.

Table 3. Comparison of cross-correlation value when using reverberant data.

Data No.	Observed Signal	DS(microphone:2)	DS(microphone:14)	Proposed Method
1	0.57	0.57	0.59	0.48
2	0.69	0.69	0.74	0.48
3	0.71	0.72	0.76	0.51
4	0.65	0.66	0.72	0.44
5	0.66	0.67	0.77	0.46
6	0.63	0.64	0.73	0.51
7	0.66	0.68	0.74	0.50
8	0.73	0.75	0.81	0.53
9	0.66	0.68	0.74	0.53
10	0.65	0.65	0.69	0.45
Average	0.66	0.67	0.73	0.49

*The figure in () means the number of microphones.

Table 4. Comparison of cross-correlation value when using reverberant data.

Data No.	Observed Signal	DS(microphone:2)	DS(microphone:14)	Proposed Method
1	10.52	10.43	9.13	9.90
2	8.30	7.95	6.47	8.29
3	12.22	11.97	11.60	11.64
4	9.68	9.67	8.74	9.82
5	10.28	10.24	9.20	10.06
6	11.33	11.34	10.25	11.08
7	9.48	9.51	7.69	9.47
8	10.07	9.96	7.64	9.81
9	9.29	9.31	7.73	9.56
10	9.37	9.26	7.57	9.13
Average	10.05	9.96	8.60	9.98

*The figure in () means the number of microphones.

Table 3 and Table 4 show the results of $CC\text{-}Value$ and $CepDist$. From these results, the performance of the proposed method degrades below DS. We consider that the failed estimation is attributable to the signal degradation caused by reverberation.

4 Conclusions

In this work, we have presented a noise reduction method in a complex spectrum plane using only two microphones. The method utilizes the geometric information and restores the target signal to estimate its spectrum. The method is advantageous in that no training time is necessary before its operation and it is effective with non-stationary noise.

The experiment results showed that the proposed method outperformed the Delay-and-Sum method and can restore the target signal almost completely in the simulated noisy environment. On the other hand, in the reverberant environment, the performance degraded. We consider this failed estimation of the spectrum of the target signal was caused by the reverberation and the propagation degradation.

In future work, we will investigate the affects in a reverberant environment, and will try to improve the performance of the proposed method in a real environment. Furthermore, we will investigate a way to estimate the noise propagation direction, as the proposed method requires that we do so.

References

1. J. L. Flanagan, J. D. Johnston, R. Zahn, an G. W. Elko, "Computer-steered microphone arrays for sound transduction in large rooms," J. Acoust. Soc. Am., vol. 78, no. 5, pp. 1508-1518, 1985.

2. K. Kiyohara, Y. Kaneda, S. Takahashi, H. Nomura, and J. Kojima, "A microphone array system for speech recognition," Proc. ICASSP97, vol. 1, pp. 215-218, 1997.
3. M. Omologo, M. Matassoni, P. Svaizer, and D. Giuliani, "Microphone array based speech recognition with different talker-array positions," Proc. ICASSP97, vol. 1, pp. 227-230, 1997.
4. Michael L. Seltzer and Richard M. Stern, "Subband parameter optimization of microphone arrays for speech recognition in reverberant environment," Proc. ICASSP2003, vol. I, pp.408-411, 2003.
5. Alan Davis, Siow Yong Low, Sven Nordholm, "A subband space constrained beamformer incorporating voice activity detection," Proc. ICASSP2005, vol. III, pp.65-68, 2005.
6. W. Herbordt, S. Nakamura and W. Kellermann, "Joint optimization of LCMV beamforming and acoustic echo cancellation for automatic speech recognition," Proc. ICASSP2005, vol. III, pp.77-80, 2005.
7. Xianxian Zhang and John H. L. Hansen, "CFA-BF : A Novel Combined Fixed / Adaptive Beamforming for Robust Speech Recognition In Real Car Environments," Proc. EUROSPEECH2003, pp.1289-1292, 2003.
8. L. J. Griffiths and C. W. Jim, "An alternative approach to linearly constrained adaptive beam forming," IEEE Trans. Antennas Propag., vol. AP-30, no. 1, pp. 27-34, 1982.
9. Y. Kaneda and J. Ohga, "Adaptive microphone-array system for noise reduction," IEEE Trans. Antennas Propag., vol. ASSP-34, no. 6, pp. 1391-1400, 1986.
10. S. Sagayama, T. Okajima, Y. Kamamoto, T. Nishimoto, "Complex Spectrum Circle Centroid for Microphone-Array-Based Noisy Speech Recognition," Proc. ICSLP2004, WeA1705o.5, 2004.
11. M. Omologo and P. Svaizer, "Acoustic Event Localization in Noisy and Reverberant Environment Using CSP Analysis," Proc. ICASSP96, pp. 921-924, 1996.
12. S. Nakamura, "Acoustic sound database collected for hands-free speech recognition and sound scene understanding," International Workshop on Hands-Free Speech Communication, pp. 43-46, 2001.