Distant Sound Suppression Using Spectral Phase Variance for Two Channel Blind Source Separation

Kazuhiro Murakami* , Arata Kawamura† , Youichi Fujisaka‡ , Nobuhiko Hiruma‡ and Youji Iiguni*
* Osaka University, Osaka, Japan
E-mail: murakami@sip.sys.es.osaka-u.ac.jp
† Kyoto Sangyo University, Kyoto, Japan
E-mail: kawamura@cc.kyoto-su.ac.jp
‡ Rion Co., Ltd.

***Abstract*—Various techniques of sound source separation have been proposed. Generally, they may try to separate and extract all the sound sources included in observed signals. On some applications, e.g., speech recognition systems, it requires to extract only desired sound sources, and other sound sources are not important. Hence, it is useful a system that evaluates the importance of each separated sound source and extracts ones with high importance. In this paper, it is assumed that the importance of sound source becomes low according with its distance to the microphone. The proposed method estimates the distance of each sound source by using the spectral phase variance which is calculated with the phase difference between the two microphones. Based on the estimated distance, we suppress distant sound sources and extract ones existing within the desired region.**

I. Introduction

Various techniques for separating observed mixture sound into respective sound sources have been proposed [1]-[6]. They are expected to be used for speech recognition systems, speech communication systems, hearable devices, and so on. The separated sound sources often include desired sound sources and undesired sound sources. Hence, it is useful to establish a system that automatically extracts only desired sounds and removes unnecessary sounds. To achieve such systems, we should evaluate the importance of sound sources. In this paper, it is assume that the desired sound sources exist close to the microphones equipped on a sound separation system, and the importance of sound source becomes low when increasing distance from the sound source position to the microphones. Under this assumption, evaluating the importance of a sound source results in estimating the distance from the sound source to the microphones.

Shimoyama et al. reported about the relation between the distance from a sound source to two microphones and the variance of the spectral phase difference which is calculated from two observed signals captured at respective microphones. Hereafter, we called the distance as SPV (Spectral Phase Variance). It has been reported that SPV differs depending on the distance and the angle from the sound source to the microphones where the zero degrees of the angle denotes the front face of the microphones [7]-[9]. Here, at the same angle, the experimental result of SPV increases as the distance from the sound source to the stereo microphone increases. When the angle is same, SPV increases as the distance increases. When the distance is same, SPV increases as the angle increases. Thus, the distance of the sound source can be obtained from SPV and the angle.

Based on the report [7]-[9], we calculate SPV and the angle of each separated sound source to estimate the distance, i.e., it implies the importance of the separated sound. Unfortunately, to obtain the relation between the distance, SPV and the angle, we have to measure them at various sound source positions in a target room in advance. This is a fatal restriction. To solve this problem, we measure the relation of the distance, SPV, angle for only two sound source positions in advance, under the assumption that the relation can be approximated as a cosine function. Based on the measured values, we obtain a threshold of SPV with the angle that implies the distance threshold. We use the threshold to judge the separated sound source is a desired sound source or not. Here, the separated sound sources is obtained by using a conventional sound source separation method [6]. The proposed system extracts only the desired sound sources whose positions are close to the microphones and removes distant sound sources. To confirm the effectiveness of the proposed method, we carried out sound source separation experiments in an actual room. Experimental results showed that the proposed method can effectively remove distant sound sources.

II. Spectral Phase Variance (SPV)

A. Phase Difference

Phase difference is the difference between two waves having the same frequency and referenced to the same point in time. Let $s(t)$ be a sound source signal at time t. Assuming that $s(t)$ reaches the two microphones, Mic1 and Mic2, as a plane wave as shown in Fig.1. Here, D [m] denotes the length of the space between the two microphones, and θ [rad] denotes the angle of the direction of arrival of $s(t)$. As shown in Fig.1, the sound source $s(t)$ is firstly observed at Mic1 as a digital signal $x_1(n)$, where n denotes discrete time index. After that, it travels $\Delta d = D \sin\theta$ [m], and is observed at Mic2 as $x_2(n)$. We assume that the digital signals consists of K sinusoids given as

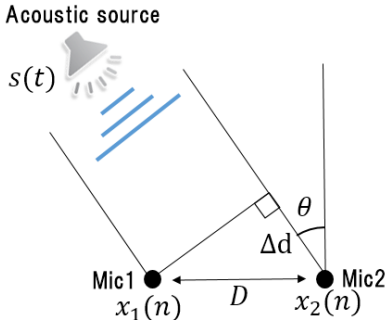

Fig. 1. Plane wave radiated from acoustic source reaches the stereo microphone at the angle θ.

$$x_1(n) = \sum_{k=1}^{K} A_{1k} \sin(2\pi f_k(n-\varphi_k)), \quad (1)$$

$$x_2(n) = \sum_{k=1}^{K} A_{2k} \sin(2\pi f_k(n-\tau-\varphi_k)). \quad (2)$$

Where f_k and φ_k denote the frequency and initial phase of the kth sinusoid, A_{ik} denotes the amplitude of the kth sinusoid at Mic i, and τ denotes the time required for traveling ΔD. Then, phase difference is defined as

$$\phi(f) = 2\pi\frac{\Delta d}{v}f + 2n\pi \quad (\exists n \in \mathbb{R}, |\phi(f)| \leq \pi/2). \quad (3)$$

The phase difference is proportional to f. Then, $x_2(n)$ is rewritten as

$$x_2(n) = \sum_{k=1}^{K} A_{ik} \sin(2\pi f_k n - \varphi_k - \phi(f_k)). \quad (4)$$

B. Spectral Phase Variance

In the previous section, we showed a linear relationship between ϕ and f when only direct sound exists. On the other hand, actual rooms usually generate a reverberation. Due to the effect of the reverberation, the linear relationship between ϕ and f is disturbed and variance occurs. To confirm this phenomena, we measured the phase difference of signals captured at two microphones in an actual room. Experimental conditions are shown in TABLE I, where the shift size of STFT analysis frame was put to a half of the STFT frame size. The calculated $\phi(f)$ from the observed signals are plotted in Fig. 2, where the horizontal axis indicates f [Hz] and the vertical axis indicates ϕ [rad]. Fig. 2 shows that a linear relationship between ϕ and f exists although a variance is included. Three solid lines in Fig. 2 are obtained from a regression line of this data, where these lines satisfy (3). The slope a of the solid lines is given as

$$a = \arg\min_k \mathrm{J}(k), \quad (5)$$

$$\mathrm{J}(k) = \sum_{i=1}^{M} \{\arg\min_{n\in\mathbb{Z}} (\phi_i - kf_i + 2n\pi)\}^2, \quad (6)$$

TABLE I
CONDITIONS OF EXPERIMENT

sampling frequency	16kHz
frame size of STFT	4096
frame sift size of STFT	2048

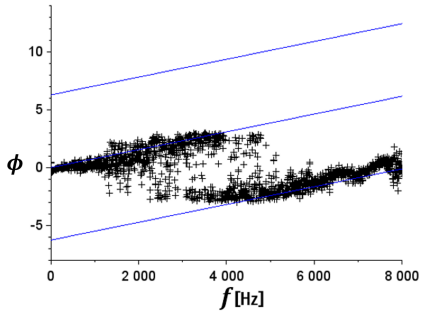

Fig. 2. Phase difference at each frequency and regression line (sound reaching from π/6 [rad] respect to stereo microphone).

where M is the number of frequencies. Also, SPV is defined as

$$\mathrm{SPV} = \frac{\mathrm{J(a)}}{M}. \quad (7)$$

C. SPV Varies Depending on Sound Source Position

It has been reported that the SPV, shown in the equation (7), varies depending on the positional relationship between the stereo microphone and sound source under reverberant field [7][8]. In order to confirm this, we measured SPV in a lecture room. The outline of experiment is shown in Fig. 3. The distance to microphone is 1m, 2m, 3m, the angle θ changes every distance to $\pi/18$ [rad] in the range of $-\pi/2 \sim \pi/2$. The result is shown in Fig. 4, where the horizontal axis indicates θ [rad] and the vertical axis indicates SPV. As distance and angle increase, SPV increases. For θ, SPV is smallest at $\theta = 0$ for all distance. The amount of change of SPV is large around $\pm\pi/4$ radian and small at around 0 or $\pm\pi/2$ rad. Also, SPV is approximately line symmetric with $\theta = 0$. In Section 3, we construct distant sound sources suppression method based on these considerations.

III. Suppress Distant Sound Sources

Shown in Fig. 4, SPV increases as the distance and angle increases. In this paper, we assume the importance of sound source becomes low according with its distance to the microphone. Here, we consider that all desired sound sources exist within L [m] from the microphone. For the property that SPV increases according to the distance, sound sources are distinguished whether a distant one or not. We can distinguish when "SPV and θ of each sound source" and "SPV for all θ at L[m]" are acquired. In order that acquiring "SPV for all θ at L[m]" is extremely difficult, we consider a method of curve fitting from a finite number of measured data. Furthermore, we also consider the method of estimate the existence angle θ of each separated sound source.

Fig. 3. Positional relationship between acoustic source and stereo microphone.

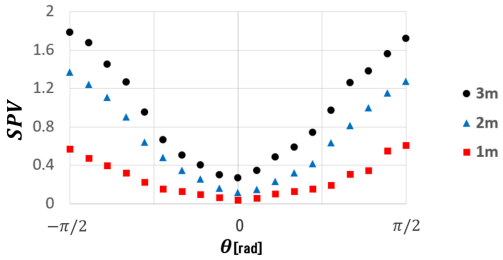

Fig. 4. SPV mesured in various positional relationship shown in Fig. 3. Horizontal axis indicates θ and same distance is indicated in the same symbol.

A. Curve Fitting for SPV

In this section, we define a model of a function representing the relationship between SPV and θ. From the result shown in Fig. 4, the following considerations are obtained

1) SPV is smallest at $\theta = 0$.
2) The amount of change of SPV is large around $\pm\pi/4$ radian and small at around 0 or $\pm\pi/2$ rad.
3) SPV is approximately line symmetric with $\theta = 0$.

Also, since the stereo microphone and the sound source have the positional relationship shown in Fig. 1, we predicted that there is periodicity with one period as $-\pi/2 \sim \pi/2$ for theta. From these considerations, we consider approximating SPV for θ with following a cosine function

$$V = -p\cos(2\theta) + q, \tag{8}$$

where the unit of θ is radian and p and q is unknown parameter.

Here, we consider how to calculate unknown parameters p and q. When the data sets of angle and SPV $([\theta_1, V_1] \sim [\theta_N, V_N])$ are obtained, the sum of squares of the observed value and the residual of the model is given as

$$S = \sum_{i=1}^{N}(V_i + p\cos(2\theta_i) - q)^2. \tag{9}$$

The partially differentiate equation (9) with p, q, set the result to zero, is given as

$$\frac{\partial S}{\partial p} = 2\sum_{i=1}^{N}(V_i + p\cos(2\theta_i) - q)\cos(2\theta_i) = 0, \tag{10}$$

$$\frac{\partial S}{\partial q} = -2\sum_{i=1}^{N}(V_i + p\cos(2\theta_i) - q) = 0. \tag{11}$$

The p and q are obtained by simultaneously solving equation (10) and (11). The result is given as

$$p = \frac{N\Sigma V_i\cos(2\theta_i) + \Sigma\cos(2\theta_i)\Sigma V_i}{N\Sigma(\cos(2\theta_i))^2 - (\Sigma\cos(2\theta_i))^2}, \tag{12}$$

$$q = \frac{\Sigma V_i + p\Sigma(\cos(2\theta_i))^2}{N}. \tag{13}$$

In order to verify the approximate accuracy of the equation (8), we conducted experiment in actual environments. This experiment was done in three rooms shown in Fig. 5∼7. Distance of sound source and microphone was 0.5m, 1m, 1.5m, 2m, and θ was every $\pi/18$ radian in the range of $-\pi/2 \sim \pi/2$ radian to get the curve shape. Fig. 8∼10 show the measurement result and approximate curves, here, the horizontal axis is θ[rad] and the vertical axis is SPV. Also, measured discrete points are plotted and the approximate curves are represented by a solid line of the same color as the plot. These figures shows that the relation between SPV and θ can be approximated by the equation (8) in all environments. Furthermore, it can be seen that the difference from the approximate curve increases as the distance from the microphone increases. Here, it was found that the approximation accuracy is relatively high at the short distance of 1.5m or less. As a whole, it is expected that an approximate curve can be used effectively if the threshold distance is within about 1.5m.

The equation (8) has two unknown parameters, so it is possible to calculate unknown parameters from at least two measured data. We verified which two angle of data minimize misrecognitions. For data shown in Fig. 8∼10, we researched the number of false recognition. In this case, the number of false recognition is the sum of the number of sound sources which is not judged to be far despite it is 2m and the number of it which in judged far despite it is 0.5m or 1m. The sum in the three environments is shown in Fig. 11. The first row and the first column are angles[°], and others are the number of false recognitions. Since it is redundant, the upper right is blank. The hatched cell has no value due to the nature of the equation (12). Unfortunately, there is no zero. This indicates that proposed method is not always valid. However, it can be said that proposed method is sufficiently effective when using a pair of angles indicated by a cell painted in red in Fig. 11.

B. Angle Estimation

According to Fig. 8∼10, if SPV and θ are obtained, we can estimate the sound source distance L. Here, we consider the estimation method of the angle θ of the sound source.

In Fig. 1, when the arrival angle of the sound wave is θ, Δd is expressed as in

$$\Delta d = av/2\pi, \tag{14}$$

where a is the slope calculated in (5) and v is the speed of sound. Here, θ is given as $\theta = \sin^{-1}(\Delta d/D)$. From this and (14), estimated angle $\hat{\theta}$ is given as

$$\hat{\theta} = \sin^{-1}(av/2\pi D). \tag{15}$$

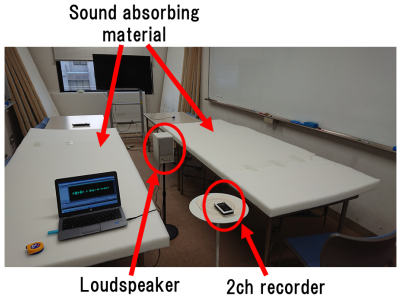

Fig. 5. Environment 1, small conference room. Sound absorbing materials reduce reflected sound from table.

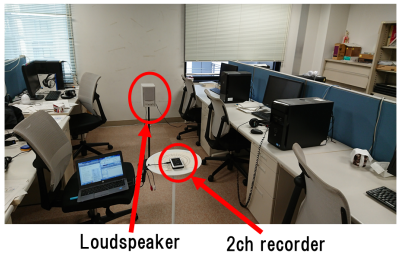

Fig. 6. Environment 2, laboratory. There are several tables and chairs.

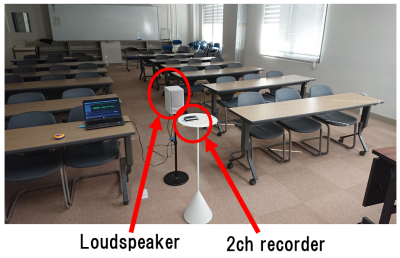

Fig. 7. Environment 3, lecture room. There are many tables and chairs.

The result of angle estimation in environment 1 is shown in Fig. 12. The $\hat{\theta}$ is calculated by equation (15) with $v = 340$m/s and $D = 0.05$m. The horizontal axis indicates $\theta[^\circ]$ and the vertical axis indicates $\hat{\theta}[^\circ]$. The black line indicates true value.

As the angle became larger, the error from the true value increased. In particular, the estimated angle for $\pm\pi/2$ rad is more than $\pm\pi/6$ rad apart. Probably, v is not accurate and the microphone itself has a width so D is not accurate. It is difficult to measure these value precisely every time, so $\hat{\theta}$ has to be corrected.

Fig. 13 shows the outline of method to correct angles. Now, the relationship seems to be $\hat{\theta} = k\theta$, so we have to estimate slope $\hat{k}$. The $\hat{k}$ is estimated from the set of $\hat{\theta}$ and θ at two points $\left([\hat{\theta}_1, \theta_1], [\hat{\theta}_2, \theta_2]\right)$ as in

$$\hat{k} = \frac{\hat{\theta}_2 - \hat{\theta}_1}{\theta_2 - \theta_1}. \tag{16}$$

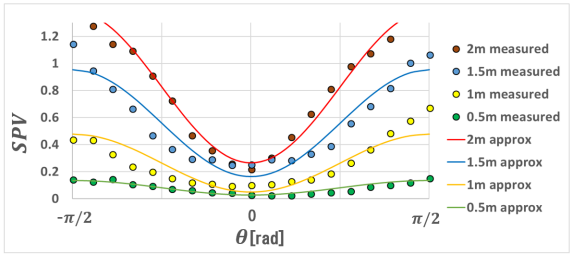

Fig. 8. SPV for each positions and approximate functions (environment 1). The same color corresponds to the same distance.

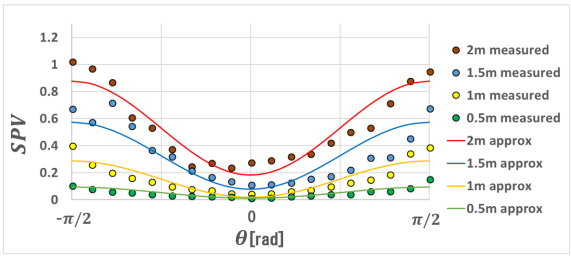

Fig. 9. SPV for each positions and approximate functions (environment 2). The same color corresponds to the same distance.

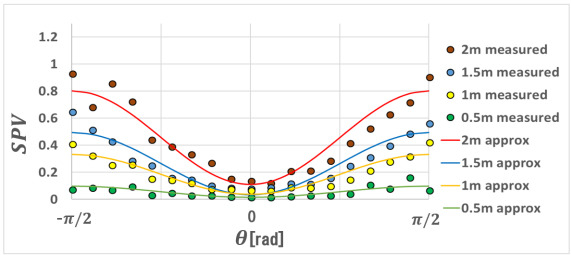

Fig. 10. SPV for each positions and approximate functions (environment 3). The same color corresponds to the same distance.

	90	80	70	60	50	40	30	20	10	0	-10	-20	-30	-40	-50	-60	-70	-80	-90
90																			
80	76																		
70	72	64																	
60	55	47	32																
50	43	37	19	20															
40	33	28	21	21	23														
30	18	9	11	9	11	17													
20	2	1	3	4	5	13	19												
10	3	2	5	6	7	14	18	16											
0	4	2	5	6	7	14	17	10	18										
-10	3	1	4	5	6	13	16	19		23									
-20	3	2	5	6	8	16	28		19	10	8								
-30	9	7	7	6	6	26		13	11	9	7	13							
-40	22	22	29	30	42		30	13	12	11	10	11	11						
-50	39	29	35	34		28	15	4	5	5	4	6	2	13					
-60	39	48	48		26	29	17	3	5	6	5	4	3	19	30				
-70	54	51		53	35	33	19	4	8	8	7	8	10	28	40	48			
-80	76		65	50	42	32	13	1	2	3	1	3	7	20	28	45	51		
-90		86	74	57	45	36	19	3	6	7	5	3	12	24	41	41	60	81	

Fig. 11. Sum of the number of false recognition. Small numbers are better. Red cells indicate the best results.

When $\hat{\theta}$ obtained by (15), the corrected angle $\hat{\theta}'$ is given as

$$\hat{\theta}' = \hat{\theta}/\hat{k}. \tag{17}$$

It was confirmed that the accuracy of angle that is corrected by two point of 1.5m, $\pi/3$ and $-\pi/6$. Fig. 14 shows the

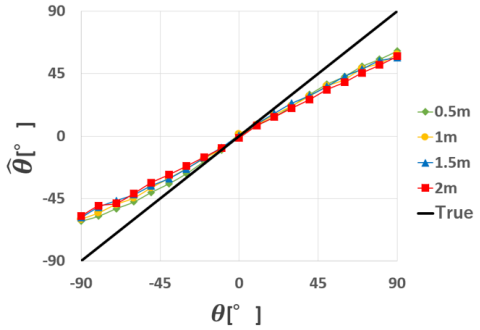

Fig. 12. Result of angle estimation (environment 1). θ indicates true angle and $\hat{\theta}$ indicates estimated angle.

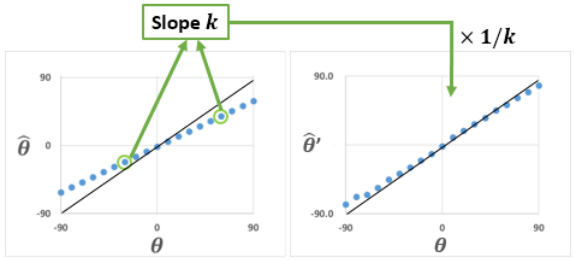

Fig. 13. Outline of angle correction. The relationship seems to be $\hat{\theta} = k\theta$. Estimate slope k from two points and correct the relationship to $\hat{\theta}' = \hat{\theta}/k$.

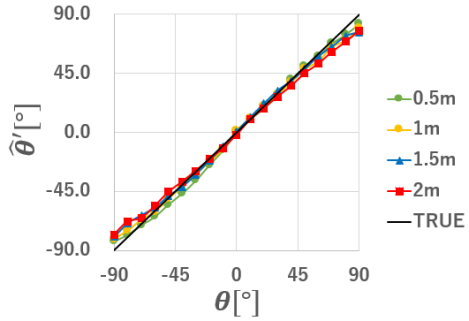

Fig. 14. Result of angle correction (environment 1). θ indicates true angle and $\hat{\theta}'$ indicates corrected angle.

result in environment 1. It shows that the corrected angle is close to the true value. Also, the result of environment 2 and environment 3 are shown in Fig. 15 and 16. From these results, it is considered that angle correction is effective. In this paper, we estimate angles using this.

C. Construction of Distant Sound Source Removal System

From the verification result so far, we construct a distant sound sources removal system. First, set the threshold distance L [m] and measure SPV at 2 angles. An approximate curve is created from the value by a least squares method and is used as a threshold value. By comparing SPV at estimated angle θ of each separated sound source with the threshold value, the desired sound sources are acquired, and distant sound sources are removed.

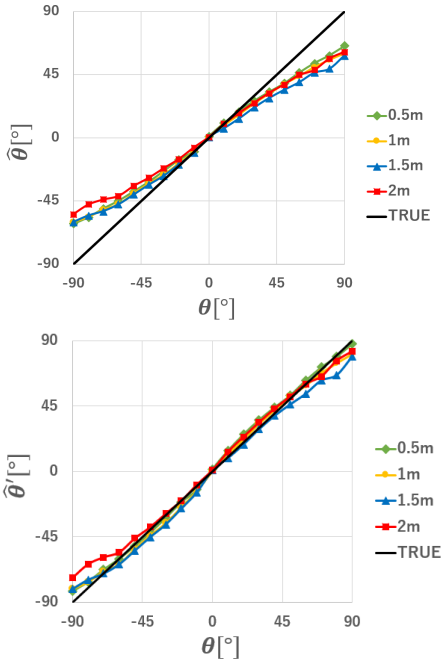

Fig. 15. Estimated angle (above) and corrected angle (below) (environment 2). θ indicates true angle, $\hat{\theta}$ indicates estimated angle, and $\hat{\theta}'$ indicates corrected angle.

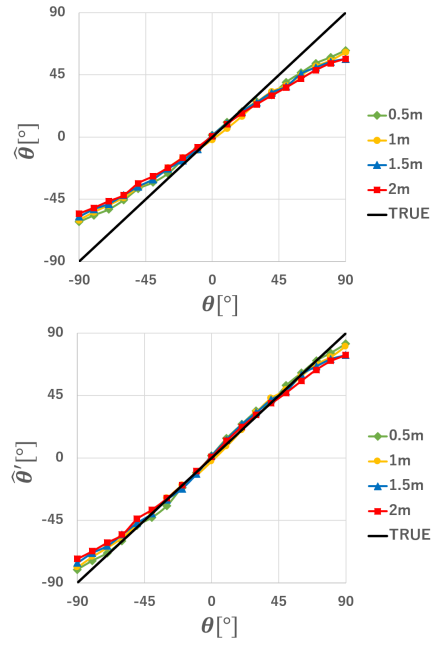

Fig. 16. Estimated angle (above) and corrected angle (below) (environment 3). θ indicates true angle, $\hat{\theta}$ indicates estimated angle, and $\hat{\theta}'$ indicates corrected angle.

IV. Experiment

A. Conditions of Experiment

In an actual environment, we recorded a sound using a stereo microphone. Here, three sound sources were set at predetermined positions and sounded each sound. The arrangement of the microphone and speaker at the time of recording is shown in Fig. 17. Three sound sources are referred to as S1, S2, and S3. Experimental condition is as shown in TABLE II. The observation signal is shown in Fig. 18. The horizontal axis indicates time[s], and the vertical axis indicates amplitude. The upper waveform is the sound observed with Mic1 and the bottom waveform is the sound observed with Mic2. The arrows on the waveform indicate the range of each sound sources.

B. Results and Discussion

The SPV and angle estimation of each separated sound source are shown in TABLE III. The error between the estimated angle and the actual angle was $-12.2°$ in S1, $1.2°$ in S2, and $1.3°$ in S3. Estimated angle of S1 has a large error. This error will be a future research. In Fig. 19, the result of TABLE III is plotted and the previously acquired threshold value is drawn with a black curve. The horizontal axis is $\theta[°]$, and the vertical axis is SPV. From the same figure, the S2 exists above the curve, so regarded as a distant sound source and removed. As a result, separated sound sources of S1 and S3 shown in Fig. 20 were extracted.

TABLE II
CONDITIONS OF EXPERIMENT

number of sound sources	3
number of microphones	2
sampling frequency	16kHz
frame size of STFT	4096
frame sift size of STFT	2048
threshold Distance	1.5m
threshold creation angle	$80°$ and $-10°$

TABLE III
θ, $\hat{\theta}'$, AND SPV OF EACH SEPARATED SOUNDS

sound source	$\theta[°]$	$\hat{\theta}'[°]$	SPV
S1	-30	-42.2	0.20
S2	0	1.2	0.42
S3	60	61.3	0.43

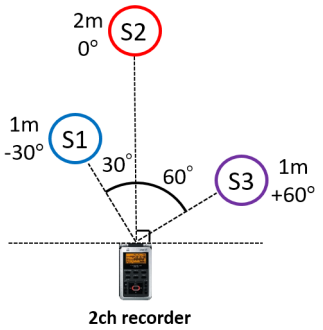

Fig. 17. Speaker arrangement respect to 2ch recorder.

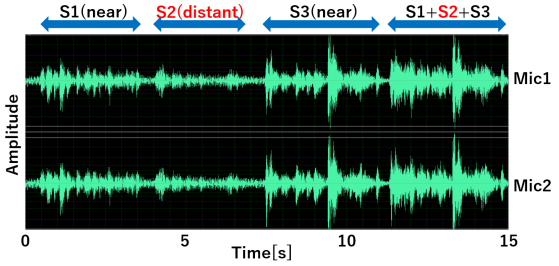

Fig. 18. Observation signal. The upper waveform is sound observed with Mic1 and the bottom waveform is sound observed with Mic2. S1, S2, S3 denote sound sources, respectively. Here, S1 and S3 exist near and S2 exists distant from the stereo microphone.

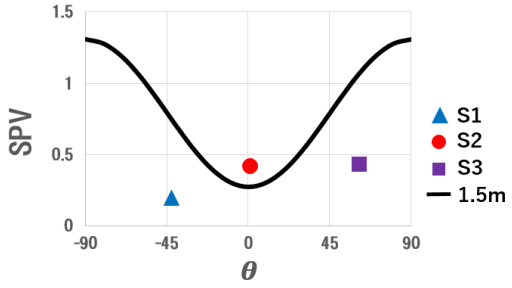

Fig. 19. Estimation results of sound source distance. Solid line indicates approximate curve at threshold distance. SPV and estimated angle of each sound sources are plotted. Triangle indicates S1, circle indicates S2, and square indicates S3.

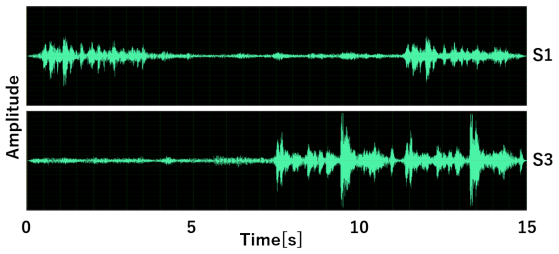

Fig. 20. Waveforms of separated sound sources judged as near position.

V. Summary

In this paper, we proposed the method that suppresses distant sound sources. In the proposed method, distant sound sources are suppressed as below. First, acquire SPV at two angles with different threshold distances in advance. Second, the observed mixture sound is separated into respective sound sources and also acquire SPV of each sound sources. Finally, based on each SPV and threshold SPV, distant sound sources are removed and only sound sources that exist in desired region are output. Experiment in actual room clarified the effectiveness of the proposed method.

References

[1] J. L. Flanagan, J. D. Johnston, R. Zahn, and G. W. Eiko, "Computer-streered microphone arrays for sound transduction in large morns," J. Acoust. Soc. Amer. 78, pp.1508-1518, July, 1985.

[2] Te-Won Lee, Michael S. Lewicki, Mark Girolami, and Terrence J. Sejnowski, “Blind Source Separation of More Sources Than Mixtures Using Overcomplete Representations,” IEEE SIGNAL PROCESSING LETTERS, VOL. 6, NO. 4, APRIL 1999.
[3] Ö.Yilmaz, and S. Rickard, “Blind separation of speech mixtures via time-frequency masking,” IEEE Trans. on SP, vol. 52, no. 7, pp. 1830-1847, July 2004.
[4] D. Kitamura, H. Sumino, N. Takamune, S. Takamichi, H. Saruwatari, and N. Ono, “Experimental Evaluation of Multichannel Audio Source Separation Based on IDLMA,” IEICE Technical Report, pp.13-20, May, 2008.
[5] K. Murase, J. Ono, S. Miyabe, T. Yamada, and S. Makino, “Far-noise suppression by transfer-function-gain non-negative matrix factorization in ad hoc microphone array,” Acoustical Society of Japan, volume 73, issue 9, pp.563-570, Mar. 2017.
[6] A. Matsuda, A. Kawamura, Y. Iiguni, “Low computational two-channel blind source separation using single voice activity segment for unknown number of sources,” Acoustical Society of Japan, volume 72, issue 3, pp.115-122, Mar. 2016.
[7] R. Shimoyama, “Distant Estimation to a Sound Source Off-Centered Using Phase Difference Spectrum Images,” FIT2009, pp237-238(issue 3), Sept, 2009.
[8] R. Shimoyama, “Distant estimation to a sound source using phase difference spectrum images,” College of Indastrial Technology. Nihon University 43rd Academic Lecture, Dec. 2010.
[9] R. Shimoyama, K. Yamazaki, “The Phase Difference Spectrum Images for Front-Back Detection on the Source Localization,” Institute of Electronics, Information and Communication Engineers (Japan), pp238, D-12-107, 2008.
[10] H. Sawada, S. Araki and S. Makino, “Underdetermined convolutive blind source separation via frequency bin-wise clustering and permutation alignment,” IEEE Trans. Audio Speech Lang. Process, 19, 516-527(2010).
[11] T. Miyauchi, D. Kitamura, H. Saruwatari, “Depth estimation of sound images using direction of arrival distribution and activation-shared non-negative matrix factorization,” Acoustical Society of Japan, pp.947-950, Mar. 2014.
[12] K. Suzuki, T. Koga, J. Hirokawa, H. Ogawa, N. Matsuhira, “Clustering of sound-source signals using Hough transformation, and application to omni-directional acoustic sense for robots,” JSAI Technical Report, pp53-54, Oct, 2005.