LOCALIZATION OF MULTIPLE SOUND SOURCES BASED ON A CSP ANALYSIS WITH A MICROPHONE ARRAY

Takanobu Nishiura,Takeshi Yamada**,Satoshi Nakamura*,Kiyohiro Shikano**

* Graduate School of Information Science, Nara Institute of Science and Technology
8916-5 Takayama, Ikoma, Nara, 630-0101 Japan
** Institute of Information Sciences and Electronics, University of Tsukuba
1-1-1 Tennoudai, Tsukuba, Ibaraki, 305-8573 Japan

ABSTRACT

Accurate localization of multiple sound sources is indispensable for the microphone array-based high quality sound capture. For single sound source localization, the CSP (Cross-power Spectrum Phase analysis) method has been proposed. The CSP method localizes a sound source as a crossing point of sound directions estimated using different microphone pairs. However, when localizing multiple sound sources, the CSP method has a problem that the localization accuracy is degraded due to cross-correlation among different sound sources. To solve this problem, this paper proposes a new method which suppresses the undesired cross-correlation by synchronous addition of CSP coefficients derived from multiple microphone pairs. Experiment results in a real room showed that the proposed method improves the localization accuracy when increasing the number of the synchronous addition.

1. INTRODUCTION

High quality sound capture of distant sound sources is very important for teleconference systems and voice control systems. However, background noise and room reverberations seriously degrade the performance of sound capture in real acoustical environments. A microphone array has been applied as one of the promising tools to deal with this problem. A desired signal can be acquired selectively by forming a directive pattern sensitive to the target sound source. However, a reliable sound source localization is necessary to maximize the effect of noise reduction. In particular, accurate sound source localization becomes more important as the directive pattern is sharpened.

For single sound source localization, the CSP (Cross-power Spectrum Phase analysis) method has been proposed [1, 2, 3, 4]. The CSP method localizes a sound source as a crossing point of sound directions estimated using different microphone pairs. However, when localizing multiple sound sources, the CSP method has a problem that the localization accuracy is degraded due to cross-correlation among different sound sources. To solve this problem, this paper proposes a new method which suppresses the undesired cross-correlation by synchronous addition of CSP coefficients derived from multiple microphone pairs.

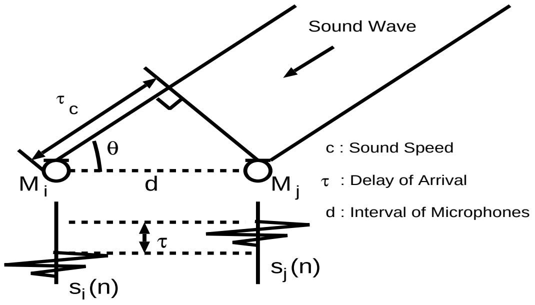


Figure 1: Estimation of DOA with CSP method.

2. CONVENTIONAL LOCALIZATION OF MULTIPLE SOUND SOURCES BY CSP

The procedure for localization of multiple sound sources by the CSP method is as follows:

1. Estimation of delays of arrivals (Estimation of directions of multiple sound sources).
2. Localization of multiple sound sources by clustering delay of arrivals.

2.1. Estimation of Delays of Arrivals

The direction of the sound source can be obtained by estimating the DOA (Delay Of Arrival) between two microphone outputs. The CSP method [1, 2, 3, 4, 5] is widely used for this purpose because of their computational efficiency and stability. Fig.1 illustrates an example of estimation of DOA by the CSP method, where $s_i(n)$ and $s_j(n)$ are the signals received by the microphones i and j. The CSP coefficients are derived from the following equation.

$$csp_{ij}(k) = \mathrm{DFT}^{-1}\left[\frac{\mathrm{DFT}\left[s_i(n)\right]\mathrm{DFT}\left[s_j(n)\right]^*}{\left|\mathrm{DFT}\left[s_i(n)\right]\right|\left|\mathrm{DFT}\left[s_j(n)\right]\right|}\right], \tag{1}$$

where n and k is the time index, DFT $[\cdot]$ (or $\mathrm{DFT}^{-1}[\cdot]$) is the discrete Fourier transform (or the inverse discrete Fourier transform) and * is the complex conjugate. When there is a single sound source, the DOA can be estimated by finding the maximum value of the CSP coefficients.

$$\tau = \underset{k}{\mathrm{argmax}}(\mathrm{CSP}_{ij}(k)). \tag{2}$$

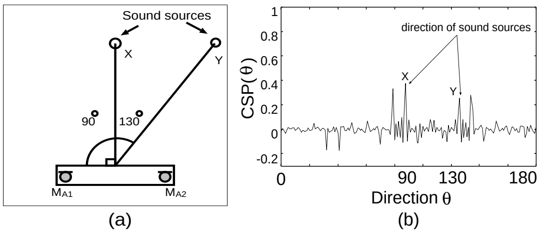


Figure 2: CSP coefficients derived from microphone pairs M_{A1} and M_{A2} when two signals are highly correlated.

Then the sound source direction is derived from the following equation.

$$\theta = \cos^{-1}\left(\frac{c \cdot \tau / F_s}{d}\right), \qquad (3)$$

where d is the distance between two adjacent microphones, c is the sound propagation speed and F_s is the sampling frequency. However, when there are multiple sound sources, the estimation of the DOAs is very difficult due to cross-correlation among different sound sources. For example, let $x(t)$ and $y(t)$ be the signals come from two sound sources. Then $s_i(n)$ and $s_j(n)$ at the microphone i and j are written as follows:

$$s_i(n) = a_i x(m + \varphi_i) + b_i y(m + \xi_i), \qquad (4)$$

$$s_j(n) = a_j x(m + \varphi_j) + b_j y(m + \xi_j), \qquad (5)$$

where m is the time index, φ and ξ are the time delays of arrivals, a and b are the distance attenuation coefficients. The numerator of Eq. 1 can be rewritten as follows:

$$\begin{aligned}&\mathrm{DFT}(s_i(n))\mathrm{DFT}(s_j(n))^* = \\ &\quad a_i a_j X(w)^2 e^{-jw(\varphi_i - \varphi_j)} + b_i b_j Y(w)^2 e^{-jw(\xi_i - \xi_j)} + \\ &\quad X(w)Y(w)(a_i b_j e^{-jw(\varphi_i - \xi_j)} + a_j b_i e^{-jw(\xi_i - \varphi_j)}). \qquad (6)\end{aligned}$$

The CSP method can accurately estimate the DOAs when two signals are uncorrelated. However, when two signals are correlated as in the real environments, the CSP method fails to estimate the correct DOAs. Fig.2(b) illustrates an example of the CSP coefficients when two signals $X(w)$ and $Y(w)$, which are highly correlated, come from an angle of 90 degrees and an angle of 130 degrees. It can be seen that the CSP function has many peaks not only in correct directions but also in incorrect directions because of the cross-correlation. Fig.3(b) illustrates another example of the CSP coefficients when three signals $X(w)$, $Y(w)$ and $Z(w)$, which are highly correlated, come from an angle of 90 degrees, an angle of 130 degrees and an angle of 70 degree, respectively. It can been seen that the incorrect DOA $Y^{`}$ is selected instead of the correct DOA Y. Thus, it is necessary to suppress undesired cross-correlation to estimate multiple sound source directions accurately.

2.2. Localization of multiple sound sources by clustering delay of arrivals

Multiple sound sources can be localized by finding a cross point of DOAs from microphone pairs.

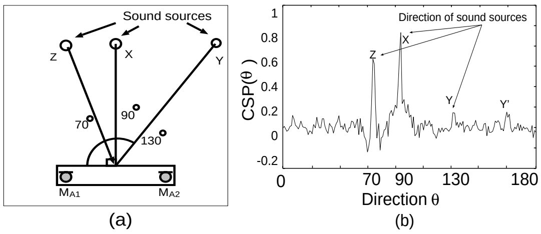


Figure 3: CSP coefficients of microphone pairs M_{A1} and M_{A2} when the magnitude of the sound source $Y(w)$ is lower than that of the other sound sources.

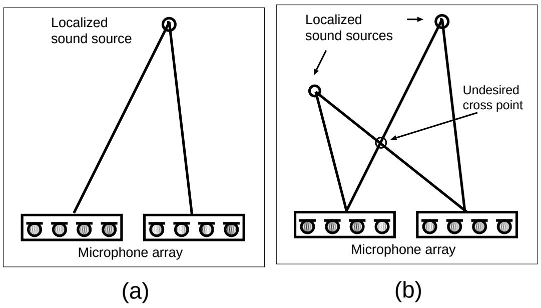


Figure 4: Source Localization of single sound source and multiple sound sources by finding cross points of the DOAs.

- Single sound source : The sound source can be localized by finding a cross point of the DOAs from microphone pairs as shown in Fig.4(a).
- Multiple sound sources : It can be seen that not only desired cross points but also undesired cross points appear as shown in Fig.4(b). Therefore, it is necessary to remove the undesired cross points for localization of multiple sound sources.

3. THE PROPOSED METHOD

Although the conventional method is effective to localize a single sound source, it is very difficult to localize multiple sound sources because of cross-correlation among sound sources. This paper proposes a new localization algorithm for multiple sound sources based on the CSP method which suppresses cross-correlation effects. Fig.5 shows an overview of the proposed algorithm.

3.1. Synchronous Addition of CSP Coefficients

If microphones position is changed, DOAs both for correct directions and directions for cross-correlation will be changed. However, if microphone are located so that the center positions of the microphone pairs are the same, DOAs for correct directions must be the same while DOAs for directions of cross-correlation might be different. Fig.6(a) and Fig.6(b) illustrate microphone positions used in the conventional method and the proposed method. Microphone pairs M_B and M_C inside of the microphone pair M_A are arranged so that center positions of those pairs are the same as the

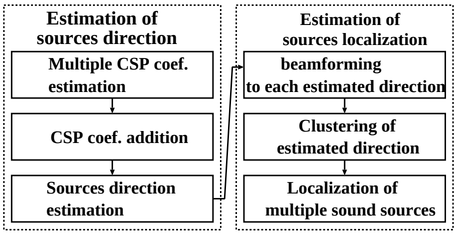

Estimation of sources direction
Multiple CSP coef. estimation
CSP coef. addition
Sources direction estimation
Estimation of sources localization
beamforming to each estimated direction
Clustering of estimated direction
Localization of multiple sound sources

Figure 5: An overview of the proposed algorithm.

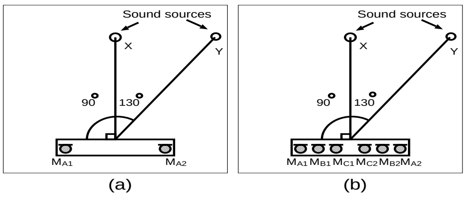


Figure 6: Positions of microphones. (a) the conventional method. (b) the proposed method).

center position of M_{A_1} and M_{A_2}. Fig.7(a)(b)(c) illustrate the CSP coefficients derived from microphone pairs M_A, M_B and M_C, respectively. Fig.7(d) illustrates a result of synchronous addition of the CSP coefficients derived from the microphone pairs M_A, M_B and M_C. It can be seen that the CSP coefficients for correct sound source directions (90° and 130°) are emphasized. In this way, only the CSP coefficients derived from the correct sound source direction can be emphasized by adding CSP coefficients from M_A, M_B and M_C. Therefore, the proposed method can be useful for accurate estimation of multiple sound source directions.

3.2. DOA Clustering with Beam-forming

If sound source directions are obtained, the sound source position is estimated as a crossing point of the directions as shown in Fig.4(a). However, in case that there are multiple sound sources, not only correct sound source positions but also undesired positions are detected as shown in Fig.4(b).

Thus, this paper proposes a new clustering method to find correct sound positions. Fig.8 illustrates a new clustering method with beam-forming. The procedure is as follows:

1. Steer microphone array directivities by the delay-and-sum beamformer[6] to each estimated direction. In Fig.8, beam1 and beam2 focus on the same sound source.
2. Calculate correlation coefficients between the beamformed signals for the same sound source (ex. beam1 and beam2).
3. Cluster the DOAs (or sound source directions) based on correlation coefficients to find correct sound source positions.

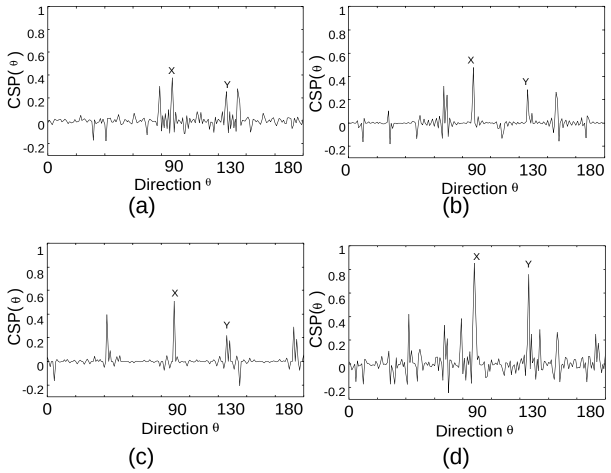


Figure 7: CSP coefficients derived from microphone pairs M_A, M_B and M_C, respectively.

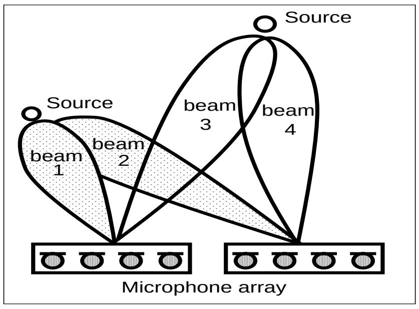


Figure 8: DOA clustering with beam-forming.

For example, beam1 and beam2, and beam3 and beam4 can be clustered in the same clusters as shown in Fig.8. In this paper, sound sources are localized by finding cross points of beam pairs with the maximum cross-correlation in the same cluster.

4. EVALUATION EXPERIMENTS

4.1. Experiment Conditions

The proposed method is evaluated by data recorded in a real room as shown in Fig. 9. Reverberation time of this room is 0.18sec. Fifteen positions indicated by A~O in Fig.9 are evaluated as the source positions. The omni-directional microphones are used. Acoustic source signals are Japanese words uttered by three talkers. The sampling frequency is 48kHz.

More than two microphone arrays are required for the proposed method. We used two sets of microphone elements such as array1 and array2 in Table.1. Each microphone array has 14 microphone elements. Table.1 also shows a configuration of microphone pairs. In case of "synchronous addition" $pair1 \cdots pair5$ are used, while only $pair1$ is used in "no addition", which is equivalent to the conventional method. The number of synchronous addition of the CSP coefficients is five times at most. The method is evaluated in the three cases that the number of the sound sources is 1, 2 and 3.

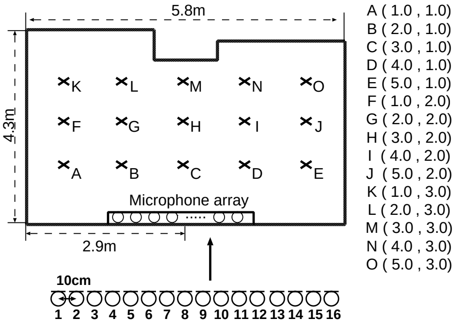


Figure 9: An arrangement of sound sources and a microphone array.

Table 1: A configuration of microphone arrays.

	array1($1 \sim 14$)	array2($3 \sim 16$)
*pair*1	1-14	3-16
*pair*2	2-13	4-15
*pair*3	3-12	5-14
*pair*4	4-11	6-13
*pair*5	5-10	7-12

4.2. Results of Experiments

Fig.10 illustrates the accuracy of sound source localization. The accuracy is evaluated by the localization rate (P_{cor}), which is based on the distance between the estimated sound position (Q_{est}) and the correct sound position (Q_{tru}) by the following equations. E_r is the allowance range of the error. N is the number of samples.

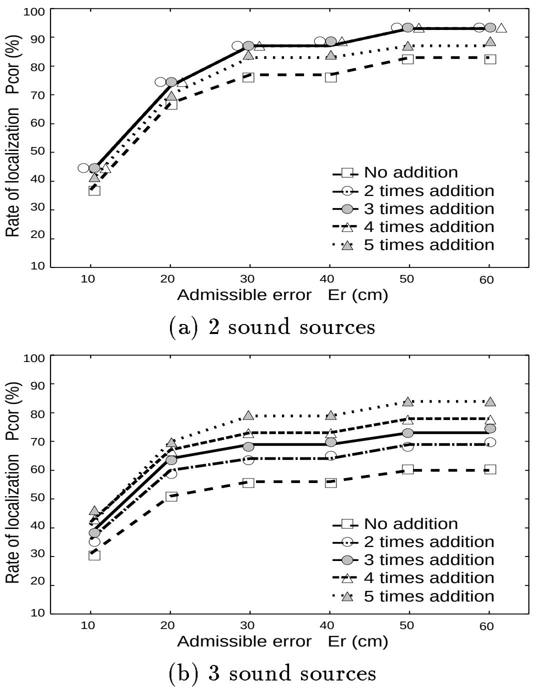
$$P_{cor} = \frac{\sum_{n=0}^{N} I_{cor}[n]}{N}. \tag{7}$$

$$I_{cor}[n] = \begin{cases} 1 & \|Q_{est} - Q_{tru}\| \leq Er. \\ 0 & \|Q_{est} - Q_{tru}\| > Er. \end{cases} \tag{8}$$

These results show that the proposed method precisely estimates the sound sources compared with the conventional method. The more number of addition increases, the more localization becomes accurate. Because the CSP coefficients for the correct directions are emphasized by increasing number of synchronous addition. Thus it is confirmed the proposed method provides the effective localization of multiple sound sources.

5. CONCLUSION

It is necessary to localize multiple sound sources accurately to capture distant talking speech using a microphone array. However, the conventional method has a problem that accuracy of localization seriously degrades because of the cross-correlation of difference sound sources. Thus this paper proposes a new method which suppresses the undesired cross-correlation by synchronous addition of CSP coefficients extracted from different microphone pairs. The experiments were carried out to evaluate the proposed method in real acoustical environments. The results showed the fact that the proposed method accurately estimates positions of the multiple sound sources. Also the performance improvement depends on the number of synchronous addition of the CSP coefficients. In future works, speech recognition evaluation of the multiple distant-talking speech will be investigated.



(a) 2 sound sources

(b) 3 sound sources

Figure 10: Accuracy of sound sources localization.

6. REFERENCES

[1] C. H. Knapp, G. C. Carter, "The generalized correlation method for estimation of time delay", IEEE Trans, ASSP, vol.24, no.4, pp. 320–327, 1976.

[2] M. Omologo, P. Svaizer, "Acoustic Event Localization using a Crosspower-Spectrum Phase based Technique",Proc. ICASSP94, pp. 273–276, 1994.

[3] M. Omologo, P. Svaizer, "Acoustic source location in noisy and reverberant environment using CSP analysis", Proc. ICASSP96, pp. 921–924, 1996.

[4] P. Svaizer, M. Matassoni, M. Omologo, "Acoustic Source Location Three-dimensional Space Using Crosspower Spectrum Phase, " Proc. ICASSP97, pp. 231–234, 1997.

[5] M. Brandstein, J.Adcock, H.Silverman, "A closed-form method for finding source locations from microphone-array time-delay estimates", Proc. ICASSP95, pp. 3019–3022, 1995.

[6] S. U. Pillai, "Array Signal Processing",Springer-Verlag, New York, 1989.