INTERSPEECH 2009 BRIGHTON

A Study on Multiple Sound Source Localization with a Distributed Microphone System

Kook Cho, Takanobu Nishiura, and Yoichi Yamashita

College of Information Science and Engineering, Ritsumeikan University, Kusatsu-shi, Japan

{cho@slp.is, nishiura@is, yama@media}.ritsumei.ac.jp

10.21437/Interspeech.2009-421

Abstract

This paper describes a novel method for multiple sound source localization and its performance evaluation in actual room environments. The proposed method localizes a sound source by finding the position that maximizes the accumulated correlation coefficient between multiple channel pairs. After the estimation of the first sound source, a typical pattern of the accumulated correlation for a single sound source is subtracted from the observed distribution of the accumulated correlation. Subsequently, the second sound source is searched again. To evaluate the effectiveness of the proposed method, experiments of multiple sound source localization were carried out in an actual office room. The result shows that multiple sound source localization accuracy is about 99.7%. The proposed method could realize the multiple sound source localization robustly and stably.

Index Terms: real environment, distributed microphone system, multiple sound sources, sound source localization, TDOA

1. Introduction

In recent years, significant research efforts have been devoted to the development of sound source localization in various environments and situations. Precise sound source localization is useful in a variety of domains, including hands-free speech recognition systems, online conferencing systems, meeting analysis, and camera steering.

A microphone array system enables estimation of the direction of arrival (DOA) of the observed speech signal based on the time delay of arrival (TDOA) between multiple captured signals. The minimum variance (MV) method [1], the multiple signal classification (MUSIC) method [2], the cross-correlation (CC) method [3] and the cross-power spectrum phase (CSP) analysis [3, 4] are popular DOA estimation methods.

A sound source can be theoretically localized using two sets of microphone array by combining two independent directions. However, this approach degrades the performance of sound source localization in noisy and reverberant environments, because small errors of direction estimation may result in a large error of position estimation. In addition, the MV and the MUSIC method are very difficult to process in real time because of their heavy computational load and complexity. Although beamforming such as the MV method can give good results, the computation is generally too expensive to allow the likelihoods to be computed at all possible locations. Although time-delay estimation methods such as the CC method and the CSP method are fast, they generally perform poorly in highly reverberant environments. Recently an accumulated correlation algorithm [5, 6, 7, 8] was proposed to combine the advantages of these two approaches. Instead of taking the peak of each correlation vector, all the correlation values from all the vectors are accumulated in a common coordinate system.

More than one sound source may exist in real environments. For example, several persons sometimes talk simultaneously in a meeting or discussion. A technique of the sound source localization is required to work in multiple sound source situations. To solve this problem, this paper proposes a new method of multiple sound source localization using a distributed microphone system that is widely distributed and placed under the ceiling of a room. The distributed microphone system can localize sound sources with optimization for a wide area of the space. A number of microphone pairs give a series of correlation coefficients as a function of the time delay. The accumulated correlation algorithm localizes the sound source based on correlation of two channel signals that are delayed with the time delay of arrival for a hypothetical sound source. The correlation coefficients for hypothetical sound sources are accumulated over many microphone pairs. The proposed method localizes a sound source by finding the position that maximizes the accumulated correlation coefficient between multiple channel pairs. After the estimation of the first sound source, a typical pattern of the accumulated correlation for a single sound source is subtracted from the observed distribution of the accumulated correlation, and the second sound source is localized by finding the maximum correlation. The position of multiple sound sources can be estimated by adapting the proposed method repeatedly.

2. Sound source localization method

2.1. Estimation of TDOA with the CSP method

The direction of the sound source can be obtained by estimating a TDOA between two microphone outputs. The CSP coefficients are calculated by the following equation.

$$CSP_{ij}(k) = IDFT\left[\frac{DFT[s_i(n)]DFT[s_j(n)]^*}{|DFT[s_i(n)]||DFT[s_j(n)]|}\right] \quad (1)$$

$$\tau = \underset{k}{\operatorname{argmax}}(CSP_{ij}(k)) \quad (2)$$

where $s_i(n)$ and $s_j(n)$ are the signals acquired through the i-th and j-th microphones, n and k are the time index, DFT[·] is the discrete Fourier transform, IDFT[·] is the inverse discrete Fourier transform, the symbol * is the complex conjugate, $CSP_{ij}(k)$ is the CSP coefficients, and τ is an estimated TDOA. The TDOA can be estimated by finding the maximum value of the CSP coefficients.

2.2. Sound sources localization based on accumulated the inter-channel correlation

In real environments, the presence of ambient noises and room reverberations seriously degrades the accuracy in sound source

localization. This paper proposes a new localization algorithm for multiple sound sources based on the inter-channel correlation calculated by the CSP method.

2.2.1. Single sound source localization

The procedure for single sound source localization by the inter-channel correlation method is as follows:

1. Make a set of hypothetical sound sources.
2. Calculate correlation coefficients for various TDOA from received signals in each microphone pair.
3. Calculate a TDOA using a transmission path between a hypothetical sound source and two microphones, and the correlation coefficients between two channel signals delayed with the TDOA are accumulated over all the microphone pairs.
4. The sound source is localized as the hypothetic position that maximizes the accumulated correlation coefficients.

The correlation coefficients for many microphone pairs are calculated by the CSP method. A TDOA, k_{ijp}, between the i-th and j-th microphones for the p-th hypothetical sound source is derived from Eq. (3).

$$k_{ijp} = \frac{|\boldsymbol{m_i} - \boldsymbol{s_p}| - |\boldsymbol{m_j} - \boldsymbol{s_p}|}{c} \tag{3}$$

where $\boldsymbol{m_i}$ is the position coordinate of the i-th microphone, $\boldsymbol{s_p}$(p=$1, 2, \ldots, P$) is the p-th hypothetical sound source position coordinate, c is the sound propagation speed. Then the accumulated CSP coefficient in the p-th hypothetical sound source, $CSP_{acc}(p)$, is derived from Eq. (4) and Eqs. (5), (6), (7).

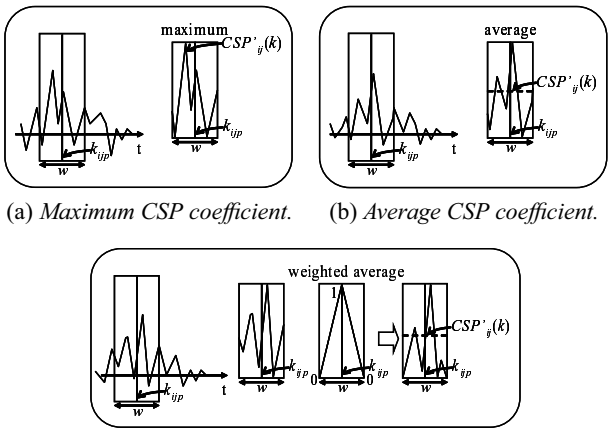
$$CSP_{acc}(p) = \sum_{(i,j)\in S} CSP'_{ij}(k_{ijp}) \tag{4}$$

$$CSP'_{ij}(k) = \max[CSP_{ij}(k-w), \ldots, CSP_{ij}(k+(w-1)), CSP_{ij}(k+w)] \tag{5}$$

$$CSP'_{ij}(k) = \frac{\sum_{t=-w}^{w} CSP_{ij}(k+t)}{2w+1} \tag{6}$$

$$CSP'_{ij}(k) = \frac{\sum_{t=-w}^{w} \frac{w-|t|}{w} CSP_{ij}(k+t)}{2w+1} \tag{7}$$

where $CSP_{ij}(k)$ is the CSP coefficient of the i-th and j-th microphone pair for TDOA, k, as shown in Eq. (1). S is a set of microphone pairs. The delay k_{ijp} is a theoretical value of the time delay between the i-th and j-th microphone pair for the p-th hypothetical sound source, and it is calculated based on the microphone positions, shown as Eq. (3). $CSP'_{ij}(k_{ijp})$ is an adjusted CSP correlation and it is accumulated in stead of a raw CSP correlation, $CSP_{ij}(k_{ijp})$, to consider measurement errors of the microphone positions. We investigate three types of the adjusted CSP correlations which are defined by Eqs. (5), (6), and (7). Fig. 1 shows these adjusted CSP correlations schematically. Eqs. (5), (6), and (7) are called the maxCSP, avgCSP, and WavgCSP, shown in Fig. 1 (a), (b), and (c), respectively. The parameter, w, is a CSP window width that controls a search range in time domain. The CSP window is defined as $|k_{ijp} - t| \leq w$. The sound source positions can be estimated by finding the maximum values of the accumulated CSP coefficients by Eq. (8).



(a) *Maximum CSP coefficient.* (b) *Average CSP coefficient.*

(c) *Weighted average CSP coefficient.*

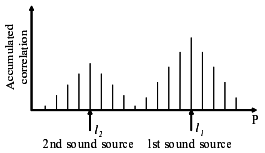
Figure 1: *Adjusted CSP correlations.*

$$\hat{l} = \underset{p}{\operatorname{argmax}}(CSP_{acc}(p)) \tag{8}$$

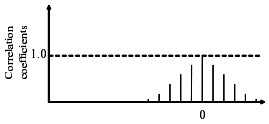
where $\hat{l}$ is the estimated position of the sound source.

2.2.2. Multiple sound source localization

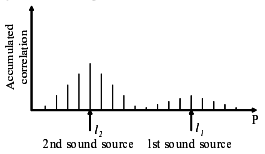
The multiple sound source positions might be repeatedly estimated by finding the maximum values of the accumulated correlation coefficients in descending order. Fig. 2 illustrates examples of accumulated correlation distribution in a one-dimensional space. The accumulated correlation peak for the second sound source is not necessarily the second largest peak in observed accumulation correlation distribution, $CSP_{acc}(p)$, shown in Fig. 2 (a). The second largest peak of the accumulated correlation does not locate in the second sound source, $\hat{l}_2$, but also in the neighbor of the first sound source, $\hat{l}_1$. The proposed method introduces the subtraction of the accumulated correlation in order to avoid such a localization error of the second sound source. The average distribution of the accumulated correlation is obtained with accumulated correlation distribution for a single sound source, and it is called the Single Source model (SS-model), shown in Fig. 2 (b). Fig. 3 illustrates examples of the SS-model of the accumulated correlation that is obtained with training data. The SS-model is obtained with the various accumulated correlation distributions for a single sound source shown in Fig. 3 (a). Specifically, the SS-models for three CSP window types and various window widths are obtained with 128 training data of a single sound source. The SS-model is calculated by averaging the various accumulated correlation distributions after they are normalized so that the peak of correlation is 1, as shown in Fig. 3 (b). The accumulated correlation distribution for multiple sound sources is modified by the



(a) *An example of observed distribution of the accumulated correlation for two sound sources.*



(b) *The average distribution of the accumulated correlation that was obtained with accumulated correlation distribution for a single sound source.*



(c) *An example of reformed distribution of the accumulated correlation for two sound sources.*

Figure 2: *Examples of accumulated correlation distribution.*

subtraction of the SS-model shown in Fig. 3 (c). The modified distribution, $CSP'_{acc}(p)$, is calculated by

$$CSP'_{acc}(p) = CSP_{acc}(p) - CSP_{acc}(\hat{l_1})Peak(p - \hat{l_1}) \quad (9)$$

using the estimated position of the first sound source, $\hat{l_1}$, and the SS-model. $Peak(p)$ is the correlation distribution of the SS-model. Fig. 2 (c) shows an example of the modified distribution of the accumulated correlation. The second source can be successfully identified by finding a correlation peak in the modified distribution since the peak of the first sound source was removed by the subtraction of the SS-model. The estimated position of the second sound source, $\hat{l_2}$, is obtained by

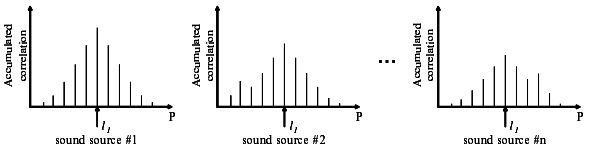
$$\hat{l_2} = \underset{p}{\operatorname{argmax}}(CSP'_{acc}(p)) \quad (10)$$

In the case of more than two sound sources, sound source positions can be repeatedly estimated by modifying the accumulated correlation distribution based on the earlier estimated sound position and the SS-model subtraction.

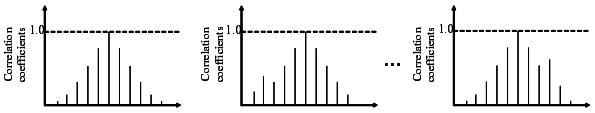
3. Experimental evaluation

3.1. Experimental conditions

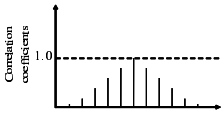
We recorded data in an actual room and evaluated the effectiveness of the proposed method. Fig. 4 shows the layout of sound sources and microphones in an experimental environment. The number of the microphones is 16 in our distributed microphone system which is installed in a 4×4 lattice condition under the ceiling. The distance between the microphones was 135[cm], and the height of the microphones was 233[cm]. As shown in Fig. 4, several noise sources such as a server and workstations existed in the experimental environment. Room reverberation ($T_{[60]}$) was 0.4[sec] and ambient noise level was 48.2∼56.0[dBA]. Thus, this room is a highly noisy environment.



(a) *A various accumulated correlation distributions for a single sound source.*



(b) *The normalization of the various accumulated correlation distributions.*



(c) *The SS-model is obtained with averaging the various accumulated correlation distributions.*

Figure 3: *Single Source model (SS-model).*

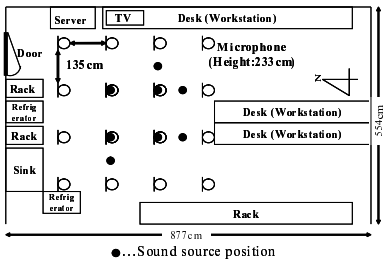


Figure 4: *An experimental environment.*

The number of the sound sources is two in this experiment. Speech materials consist of four Japanese sentences spoken by a male speaker and a female speaker, and they are played through two loudspeakers and recorded by the distributed microphone system. Direction of the loudspeakers was set to one of four directions; north, east, south, and west. Angle of the loudspeakers was set to the horizontal direction. Eight positions indicated by small black circles in Fig. 4 are evaluated as sound source positions. Two loudspeakers are put in two positions which are selected among the eight sound source positions. Thus, changing the setting of the loudspeaker, we recorded the data of 96 sentences in total.

The sampling frequency was 16[kHz], and quantization

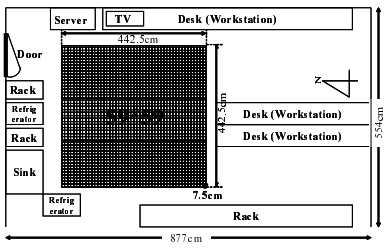


Figure 5: *Hypothetical sound sources.*

was 16[bits]. We tried to evaluate the proposed method with 1,024[msec] frame length. Sound source localization estimation is conducted for speech periods with 100[msec] shift interval. The height of the loudspeaker was 108[cm]. A sound source is localized under a condition that the height of the source is given. We investigated the accuracy of sound source localization for 2-dimensional lattices of hypothetical sound sources; 7.5[cm] (3481 point), as shown in Fig. 5.

3.2. Experimental results

The performance of the proposed method is controlled by the CSP window, w. In order to set up the suitable CSP window, the accuracy of multiple sound source localization was investigated changing the CSP window by three methods; maxCSP, avgCSP, and WavgCSP which are defined by Eqs. (5), (6), and (7). Fig. 6 shows the correct estimation rate of a multiple sound sources. The correct estimation is defined such that the distance between the estimated and the correct positions is less than 20[cm]. Fig. 7 shows the average error distances between the correct and the estimated sound sources. In Fig. 6 (b) and Fig. 7 (b), the method-1 indicates results of the proposed method. In Fig. 6 (a) and Fig. 7 (a), the method-0 indicates results of the baseline method which identifies the second largest peak in the original accumulated correlation distribution as the second sound source.

In Fig. 6 (b), when the CSP window, w, is increased, localization accuracy is improved. However, too large w decreases the accuracy. The accuracy in the multiple sound source localization was maximized by 250[micro sec] of the avgCSP. The accuracy of the multiple sound source localization accuracy is 99.7% for optimized parameters, as shown in Fig. 6 (b). In Fig. 7 (b), average error distances of the multiple sound sources are less than 13.7cm. Even if a sound source was incorrectly localized, estimated positions were in the vicinity of the correct position. The method-1 clearly has better results than the method-0. The method-1 subtracts the accumulated correlation distribution around the first sound source from the observed distribution of the accumulated correlation and the subtraction of the accumulated correlation is effective to find the second sound source. These results show that the proposed method accurately estimates the multiple sound sources.

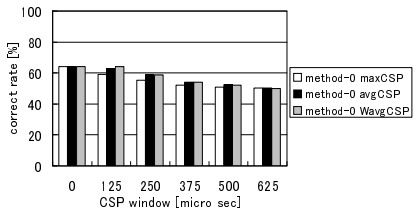
4. Conclusions

This paper proposes a new multiple sound source localization method based on the accumulated inter-channel correlation using a distributed microphone system. The experiments were carried out to evaluate the proposed method in a real environment. As a result of evaluation experiments, we confirmed that the multiple sound source localization estimation performance of the proposed method is superior. In addition, the performance of the proposed method is improved by the CSP window.

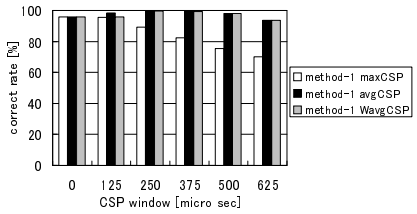
In the future, we will investigate the subtraction of the accumulated correlation in order to improve the localization error by the reflection sound. The accumulated correlation is improved by the subtraction of the accumulated correlation coefficient model corresponding to the reflection sound from the observed distribution. Our final goal is to acquire source separation by using results of sound source localization.

5. References

[1] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," Proc. IEEE, Vol. 57, No. 8, pp.1408–1418, Aug. 1969.

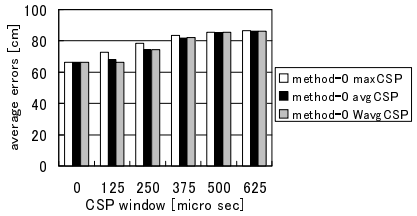


(a) *Correct estimation rate of a method-0.*

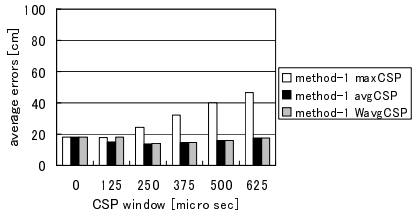


(b) *Correct estimation rate of a method-1.*

Figure 6: *Correct estimation rate of a multiple sound sources.*



(a) *Average errors of the estimated method-0.*



(b) *Average errors of the estimated method-1.*

Figure 7: *Average errors of the estimated multiple sound sources.*

[2] R.O. Schmidt, "Multiple emitter location and signal parameter estimation," IEEE Trans. on Antennas and Propag., Vol. AP-34, No. 3, pp.276–280, Mar. 1986.

[3] C.H. Knapp, et al., "The generalized correlation method for estimation of time delay," IEEE Trans, Acoust. Speech Signal Process., vol. ASSP-24, no. 4, pp.320–327, Aug. 1976.

[4] M. Omologo, et al., "Acoustic source location in noisy and reverberant environment using CSP analysis," Proc. ICASSP96, Vol. 2, pp.921–924, Atlanta, GA, USA, May 1996.

[5] K. Cho, et al., "Sound source localization using a distributed microphone system in real environments," 4th Joint Meeting of the ASA and the ASJ, Vol. 120, 3pSP42, Honolulu, USA, Nov. 2006.

[6] K. Cho, et al., "3-Dimensional sound source localization using a distributed microphones system," Proc. ICA2007, CAS-04-007, Madrid, Spain, Sep. 2007.

[7] S. T. Birchfield, et al., "Acoustic source direction by hemisphere sampling," Proc. ICASSP'01, Vol. 5, pp.3053–3056, Salt Lake City, UT, USA, May 2001.

[8] S. T. Birchfield, "A unifying framework for acoustic localization," Proc. EUSIPCO'04, Vienna, Austria, Sep. 2004.