Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics
San Antonio, TX, USA - October 2009

A Modified Cross Power-Spectrum Phase Method Based on Microphone Array for Acoustic Source Localization

Miao Shen, Hong Liu

***Abstract*—The cross power-spectrum phase (CSP) method plays an important role in time delay estimate due to its efficiency in acoustic source localization. Generally, impacted by the spatial noise and reverberation in a room, the positioning accuracy could be greatly enhanced. However, this method could hardly lead to results when the energy of signal is small. In this paper, a new method is proposed: covariance matrix that from the multi-channel signals were applied and the characteristic of the coherent function was employed. Experiments under various conditions were carried out to demonstrate the new method's novelty, including adding background music to simulate the real environments. The 240 groups' data we achieved reached localization accuracy of 97.5% in normal experiments. We also obtained 85.42% and 83.75% correct rate in low and strongly musical environments, respectively.**

I. Instruction

ACOUSTIC source localization based on microphone array [1] has various applications such as video conferencing [2], hands-free mobile telephony [3], vehicle navigation [4], etc. It is also a quite fundamental task in human-robot interaction fields. A microphone array is used in real time voice processing for its good quality in picking up speech and noise, especially in a dynamic working environment. In this environment, robots' applications need more flexibility and adaptability. During human-robot interaction, each order should be located accurately. However, the effect of a single microphone is greatly reduced when environments are complicate with much noise. This problem could be solved effectively using microphones array which is placed on the robot's body to estimate the location of acoustic source.

Time Delay Estimation (TDE) is a widely used method for acoustic source localization (ASL). It is excellent at precision, efficiency, and robustness in estimating delays between the arrival times of the pertinent acoustic waves with two or more sensors [5].

Many methods of TDE are used for ASL, including the general cross-correlation (GCC), the adaptive eigenvalue decomposition (EVD) algorithm [7], the least mean square (LMS) adaptive filter [8] and the cross power-spectrum phase (CSP) [9,10]. GCC is the most straightforward method to estimate the time delay between two signals [6]. The LMS adaptive filter is a Finite Impulse Response (FIR) filter. It could automatically adapt its coefficient to minimize the mean square difference between inputs, even without pre-knowledge of the input spectra. CSP is regarded as the most effective method in environments with weak background noise and room reverberations [11].

The CSP method weights the signal in the domain of the power spectrum and sharpens the peak of the correlation function in the delay. It is a popular time delay estimation method. A more precise TDE can be achieved in an environment with weak noise and medium reverberation. However, the peak of the correlation function will be weakened and the performance will degrade sharply when it's noisy, especially in an environment with low signal-to-noise ratio (SNR) and strong reverberation. At the same time, the function produces a number of peaks. Because the reverberation brought many echoes, the peaks are formed by combination of direct waves and reflected waves. As a result, the desired estimation could not be achieved.

In this paper, a modified CSP method ρ-CSPC is presented: The ρ-Cross Power-Spectrum Phase with Coherence Function. It aims to achieve accurate experimental results in noisy environments. It computes the coherence of two input signals by using the coherence function. Big weight is used when the coherence is high in order to suppress the noise. Also, small weight is used in low coherence. However, when the energy of the signal is small, the error will be enlarged. The methods join multi-channel information acquired by the covariance matrix in the denominator of the weight. Still, a whitening parameter ρ [12] is defined in the cross power-spectrum phase to reduce reverberant effects. Compared with the CSP method, the ρ-CSPC method not only combines many existing modified methods but also has a novel formulation. It produces more precise and robust results.

This paper is organized as follows. In section II, the delay estimation algorithm and the CSP method are formulated, then, the ρ-CSPC method is proposed in section III. Experimental results and discussions are provided in section IV. Finally, the conclusion of this paper is drawn in section Ⅴ.

II. Time Delay Estimation Method

The TDE method usually estimates the time-difference of arrival (TDOA) between each microphone in a pair. It also constrains the acoustic source location to a hyperboloid. In

This work is supported by National Natural Science Foundation of China (NSFC, No. 60675025) and National High Technology Research and Development Program of China (863 Program, No. 2006AA04Z247).

Miao Shen is with the National Laboratory on Machine Perception, Peking University, Beijing, China (phone: 86-10-62759312; fax: 86-10-62755569; e-mail: shenmiao@cis.pku.edu.cn).

Hong Liu is with the National Laboratory on Machine Perception, Peking University, Beijing, China (e-mail: hongliu@pku.edu.cn).

978-1-4244-2794-9/09/$25.00 ©2009 IEEE

space, the hyperboloid is made up by a set of points. These points maintain the same distance with each other. Two microphones are located on two respective fixed points in multi-dimensions [13]. Actually, if the hyperboloids of multiple microphone pairs are acquired, the range of the acoustic source location could be easily pinpointed from the intersection of the hyperboloids. Using microphone array, the hyperboloids of several microphone pairs can be calculated. Further, the difference of distance can also be obtained from the time delay. Consequently, the purpose of the time delay estimation is to determine the location of the acoustic source based on a selected set of TDOA's [14].

A. Signal Model

The acoustic source signal s can be acquired by a microphone m_i and the output from the microphone can be expressed as the signal $x_i(t)$.

$$\begin{aligned} x_i(t) &= h_i(t) * s(t) + \omega_i(t) \\ &= \alpha_i s(t-\tau_i) + \omega_i(t) \end{aligned} \tag{1}$$

Where $h_i(t)$ is the acoustic impulse response between S and $x_i(t)$, α is the attenuation factor due to propagation effects, τ_i is the time it takes sound to propagate from the source to microphone m_i, and $\omega_i(t)$ is the additive noise signal which can also be correlated with s in a reverberant environment.

B. Cross Correlation Function (CCF)

Considering two microphones m_i and m_j, the time delay τ_{ij} is defined to express expediently as follows:

$$\tau_{ij} = \tau_i - \tau_j \tag{2}$$

Correlation function is selected here for its location of peak can reflect the time delay and the relevance of two signals. The CCF of signal x_i and x_j can be expressed as:

$$\begin{aligned} R_{ij}(\tau) &= E\left[x_i(t) x_j(t-\tau)\right] \\ &= \alpha_i \alpha_j E\left[s(t-\tau_i) s(t-\tau-\tau_j)\right] \\ &= \alpha_i \alpha_j R_{ss}(\tau-\tau_{ij}) \end{aligned} \tag{3}$$

Where R_{ss} is the autocorrelation of signal s.

According to the characteristic of the autocorrelation function, the R_{ss} can get the maximization when:

$$\tau = \tau_{ij} \tag{4}$$

Therefore, the time delay is corresponded by the peak of the CCF which has the mapping with the cross power spectrum in frequency domain. Then the function can be formulated as:

$$R_{ij}(\tau, n) = \int_{-\infty}^{+\infty} \Phi_{x_i x_j}(\omega, n) e^{j2\pi\omega\tau} d\omega \tag{5}$$

Where $\Phi_{xixj}(\omega,n)$ is the cross power spectrum of signal x_i and x_j.

C. Cross Power-Spectrum Phase (CSP)

In this method, the weighting is used to sharpen the cross-correlation peak. The modified CCF can be defined as:

$$R_{ij}^{g}(\tau, n) = \int_{-\infty}^{+\infty} \psi_{x_i x_j}(\omega, n) \Phi_{x_i x_j}(\omega, n) e^{j2\pi\omega\tau} d\omega \tag{6}$$

Where $\psi_{xixj}(\omega,n)$ is a frequency weighting. Some spectral information is emphasized on. The information depends on characteristics of signal and noise. The CSP, "whitening" the input signals, is a typical weighting to sharpen the cross-correlation peak. Here the weighting is chosen as:

$$\psi_{x_i x_j}(\omega, n) = \frac{1}{|\Phi_{x_i x_j}(\omega, n)|} \tag{7}$$

Thereby, the modified cross-power spectrum is:

$$\begin{aligned} &\psi_{x_1 x_2}(\omega,n)\Phi_{x_1 x_2}(\omega,n) \\ &= \frac{\Phi_{x_1 x_2}(\omega,n)}{|\Phi_{x_1 x_2}(\omega,n)|} \\ &= \frac{|\Phi_{x_1 x_2}(\omega,n)| \exp(-j\varphi(\omega,n))}{|\Phi_{x_1 x_2}(\omega,n)|} \\ &= \exp(-j\varphi(\omega,n)) \end{aligned} \tag{8}$$

In the ideal situation of uncorrelated noise and signals, the phase of $\varphi(\omega,n)$ is a linear J even if ω and n are different. Then the cross-correlation function can be expressed as:

$$R_{x_1 x_2}(\tau, n) = \delta(\tau - J) \tag{9}$$

Eq.(9) represents the CSP method has an advantage. It doesn't require the modeling of the statistical characteristics of the acoustic source signal and noise. However, in a real situation, the small signal energy and high noise slow down the detecting accuracy of the time delay. As a result, an improved method ρ-CSPC is put forward in next subsection.

III. ρ-CSPC Method

By utilizing the characteristics of coherence function and the covariance matrix, the ρ-CSPC method is proposed here to improve the weighting. It also combines with existing modified methods and real experimental environment

A. The Covariance Matrix

Cross power spectral density, the most important factor in CSP method, is the description of random signals in the frequency domain. The satisfactory time difference will be obtained when it is estimated correctly. Actually, the signal-to-noise ratio (SNR) is always computed as a principle speech detecting accuracy [15]. We assume the noise is independent of the acoustic source and the impulse response of the reverberant channel is time invariant and linear. A good method is to estimate the variable SNR by classifying the covariance matrix. The covariance matrix can characterize the variance of various vector elements into speech and non-speech parts. The former one can be used to compute the cross power spectral density $\Phi_{xixj}(\omega,n)$.

The output vector of array in the frequency domain can be described as:

$$X(\omega) = H(\omega) S(\omega) + N(\omega) \tag{10}$$

where $H(\omega)$ represents the spectrum of the impulse response in the reverberant channel. The covariance matrix which serves also as the cross correlation function can be modeled in the frequency domain [15].

$$\begin{aligned} r(\omega) &= E\left[X(\omega)X^H(\omega)\right] \\ &= \left|S(\omega)\right|^2 H(\omega)H^H(\omega)+\sigma^2(\omega)I \end{aligned} \tag{11}$$

Where $|S(\omega)|^2$ represents the power spectral density of acoustic source signal and $\sigma^2(\omega)$ represents environmental noise. $H^H(\omega)$ is the conjugate matrix of $H(\omega)$ and I is the unit matrix.

As deducted above, the noise in different frequency is also assumed to be uncorrelated, so the non-diagonal elements of the covariance matrix are used to modify $\Phi_{xixj}(\omega,n)$. The cross power spectral density of the speech signals can be obtains from the multi-channel signals as:

$$\begin{aligned} \left|S(\omega)\right|^2 &= \Phi'_{x_ix_j}(\omega,n) \\ &= \frac{2}{M^2-M}\sum_{i=2}^{M}\sum_{j}^{i-1}|r_{ij}(\omega)| \end{aligned} \tag{12}$$

Where M is the number of the microphones, $r(\omega)$ is the covariance matrix of the microphone array output vector.

This method doesn't need a prior knowledge of background noise and it can be performed easily. In eq. (10), if M=2, the cross power spectral density just corresponds to the single channel model.

B. The Effect of Noise and Reverberations

From the algorithm described above, we find SNR has different values in different segments. Here we assume that the statistical behavior of both signal and noise is uniform through the entire spectrum [12]. On account of large amounts of non-speech is below 200 Hz, a whitening parameter ρ is set to perform only a partial whitening of the spectrum. It could discard the portion below 200 Hz in the CSP method [12]. And the weighting is modified as follows:

$$\psi_{x_ix_j}(\omega,n)=\frac{1}{|\Phi'_{x_ix_j}(\omega,n)|^{\rho}},\quad 0\le\rho\le1 \tag{13}$$

The value of ρ can be determined by characteristics of noise and reverberations in a room. And always, a good value of ρ can be estimated experimentally for several different enclosures in normal rooms. If ρ is set to 0, the algorithm becomes un-normalized cross correlation, while, if it equals to 1, the cross power spectral density is just like the eq. (12).

C. Coherence Function

If the $|\Phi_{xixj}(\omega,n)|$ is used as a weight in eq.(7), the denominator of the weight will become 0 when the signal energy is relatively small. The method could be improved to reduce the increased errors. A constant can be added in the denominator in the weight. The coherence function [16] is proposed here and the novel point in this paper is our proposed method of choosing the constant.

The coherence function is defined in terms of power spectral densities and the cross-spectral density. It relates to cross-correlation.

$$\gamma^2_{x_ix_j}(\omega)=\frac{|\Phi_{x_ix_j}(\omega)|^2}{\Phi_{x_i}(\omega)\Phi_{x_j}(\omega)},\quad 0<\gamma^2_{x_ix_j}(\omega)<1$$

(14)

It is a real function between 0 and 1 which gives a measure of correlation between x_i and x_j at each frequency ω. When x_i and x_j are uncorrelated, the sample coherence converges to 0 at all frequencies. Then, when the value of the coherence function is 1, the input signal and output signals are fully coherent. Furthermore, if the coherence function is between 0 and 1, there are three possibilities: the outside noise interference is existed, the output is integrated by many inputs or it is a non-linear system. A common application for the coherence function is the validation of input or output data in an acoustics experiment with purpose of system identification.

The minimum value of coherence function of two signals is the best constant in the denominator in the weight because of its characteristics and many experimental factors.

To sum up, the modified weight can be expressed as follows:

$$\psi^M_{x_ix_j}(\omega,n)=\frac{1}{|\Phi'_{x_ix_j}(\omega,n)|^{\rho}+\min\left(\gamma^2_{x_ix_j}(\omega)\right)}\qquad 0\le\rho\le1 \tag{15}$$

Once two time differences from two microphone pairs are acquired by using the modified weight function in the ρ-CSPC method, the reliable localization angle of the acoustic source can be computed by the geometrical model.

IV. Experiments and Discussions

A. The Experimental Environment

A real-time localization system is established for the application in robot auditory system. In this system, an array of four microphones classified into two pairs is arranged in linear; the distance of two adjacent microphones is 10 cm; they are placed in an empty room of 8m × 3m × 3.5m without obstacles. Ten students are arranged to walk around the microphone array on the robot. Eight positions are designated for detecting the location of the acoustic source conveniently. When the person gets into the appointed position with constant velocity, he must say a word or sentence. Three laps are requested for each person. Consequently, the system gets 240 groups of experimental data. Here, the data is divided into three parts equally according to different background noise. The first part with 80 groups' data is in normal environment. The second part is in low volume musical background environment. And the last 80 groups' data are created in big musical environment. Fig.1 shows the experimental environment. And it is considered to be correct when the error in direction is 10 degrees or less in humans.

Fig.1. the experimental environment consists of four microphones and the shelf which can be applied in robot.

In Fig.1, four microphones are arranged on the shelf. The height is 80 cm and the interval is 10 cm. The model of the microphones is MPA416, A German made Marian Trace 8 sound card is used as the all-analog multi-channel in this localization system

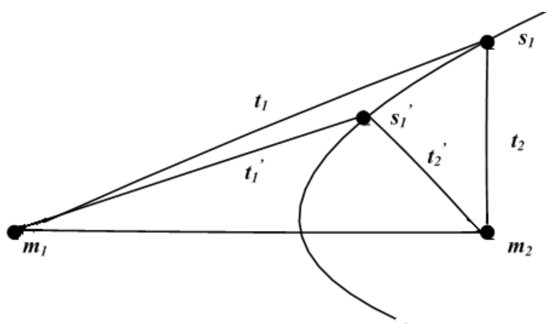

Fig.2. the sketch map of sound propagation from acoustic source to microphones.

As shown in Fig.2, s_1 and s_1' are acoustic sources, m_1, m_2 are two microphones which make up a pair and t_1, t_2, t_1', t_2' are the time interval when acoustic sources transform to microphone m_1, m_2. Then the microphone pair m_1, m_2 with a distance of 10 cm are considered. The transforming time from acoustic source s_1 to microphones m_1, m_2 is t_1, t_2, respectively. While the time difference defined as D_{12} is computed by TDOA estimate is t_1-t_2.

Accordingly, a hyperboloid H_1 consists of all points with the same D_{12} in a three dimensional space. D_{12} is defined as the eq. (16). The acoustic source S_1 can be estimated on this hyperboloid:

$$\frac{D(s\text{-}m_1)\text{-}D(s\text{-}m_2)}{V}=D_{12} \tag{16}$$

Where $D(s\text{-}m_k)$ refers the distance and V refers the speed of sound. V is always a constant as 340 meters per second at environments with normal temperature and pressure. Here S means a random point as an acoustic source lying on H_1. Different microphone pairs can estimate different hyperboloids according to different time difference. The same acoustic source S must satisfy the following formula:

$$\frac{D(s\text{-}m_i)\text{-}D(s\text{-}m_j)}{V}=D_{ij} \tag{17}$$

The location of the acoustic source is the intersection of all hyperboloids involved in the above formulas in theory. The reference [12] provides a more particular expatiation by inducting the error.

It is obvious that time difference is the most important factor in acoustic source localization when the distance between a microphone pair is mixed. As a result, the experimental results in this paper are characterized by using time delay.

B. *Experimental Performance in Real Normal Environment*

First, considering the two signals from one group of random experimental data in the first part, the cross correlation functions of the CSP and ρ-CSPC are shown as follows to check the peak value and then to compute the time delay. In this situation, with little noise, the experiment can be dealt with noise-free environment by filtering through the filter. The value of whitening parameter ρ is 0.9 experientially.

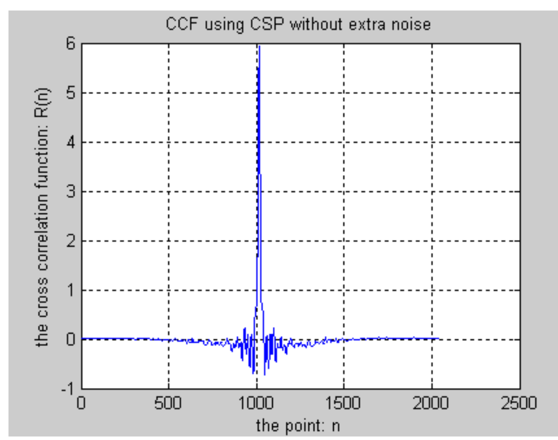

Fig.3. cross correlation function using CSP without extra noise

What shows in Fig.3 is the cross correlation function acquired by the CSP method based on two signals from one group of random experimental data, when ρ is 0.9. The point n corresponds to the maximum value of the cross correlation function is 1017. The time delay is 0.16 ms. In the same condition, Fig.4 shows the result with ρ-CSPC method.

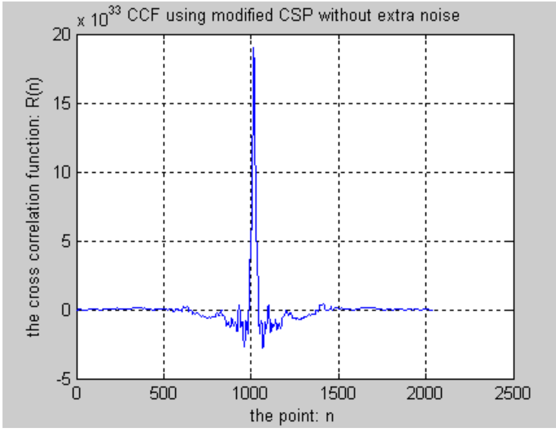

Fig.4. cross correlation function using modified CSP without extra noise

Fig.3 also refers to the values of cross correlation function in different points using CSP method while the whitening parameter is 0.9 in a state of quiet. In Fig.4, with the same state, the cross correlation function is just computed making use of the ρ-CSPC method. Both the two methods can calculate the exact result in this situation. The modified CSP even can detect the time delay with two points accurately. Combining with sampling rates, if the time delay of arrival is one point, it means that the time of the estimation is approximate 0.23 ms, which, calculated as follows:

$$T=\frac{|n-1024|}{N} \tag{18}$$

Where n is the number of the points and N is the number of the sampling points which is also the sampling rates.

C. Experimental Performance in Musical Environment

In order to test the localization performance, some songs have been added to the experimental environment as the background noise while the definition of signal-noise rate (SNR) is advanced here to estimate the noise accurately. In this experiment, two songs with different volumes are applied in background environments.

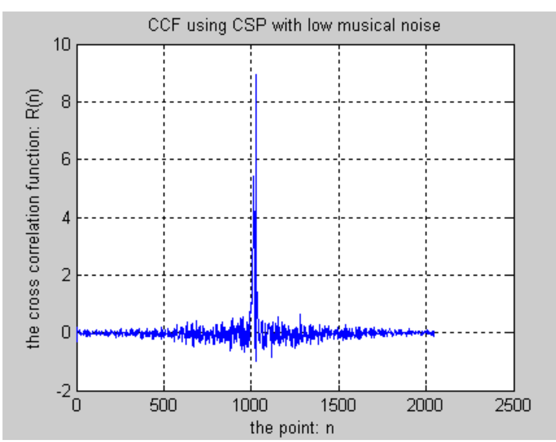

Fig.5. cross correlation function using CSP in musical environments

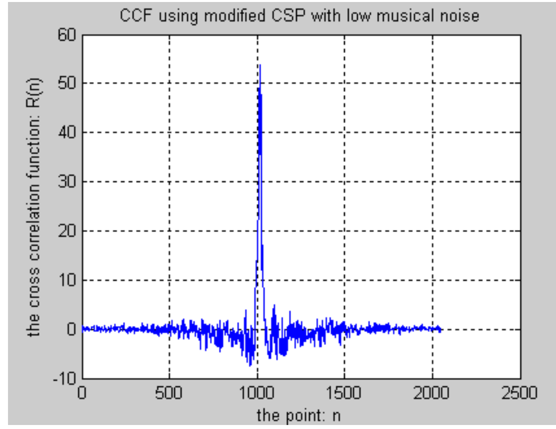

Fig.6. cross correlation function using modified CSP in musical environments while ρ is 0.9

Fig.5, Fig.6 and Fig.7 are used to check the peak value to get the exact time delay. Fig.5 presents the cross correlation function with CSP of two signals from one group of random experimental data. The point n that corresponds to the max value of the cross correlation function is 1024, while as ρ is 0.9 in Fig.6, the cross correlation function with ρ-CSPC of two signals from one group of random experimental data. The point n that corresponds to the max value of the cross correlation function is 1020.The time delay is 0.09 ms. Then as ρ is 0.78 in Fig.7, the cross correlation function with ρ-CSPC of two signals from one group of random experimental data. The point n that corresponds to the max value of the cross correlation function is 1015. The time delay is 0.20 ms. Combining with the actual location of the acoustic source, it is found that the best value of ρ is 0.78 in the experimental environment in this filtering system.

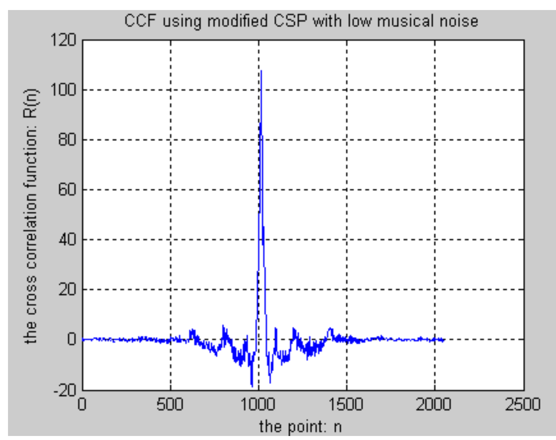

Fig.7. cross correlation function using modified CSP in musical environments while ρ is 0.78

It refers to the CSP method cannot get accurate result in this situation because of impact of noise. At 10dB, for all subsequent experiments, CSP method can only check 11 points at most while ρ-CSPC can get 5 points.

Three tables below list show the localization accuracy of CSP and modified CSP method in different musical intensities. They also show the number of experimental data which can keep the error below 5 degrees and 10 degrees respectively. They show that the modified CSP method ρ-CSPC gets better localization results than the normal CSP method with songs of various volumes. Still, ρ-CSPC method obtains the number of data with accuracy below 5 degrees, which is more precise than CSP method.

TABLE I
THE COMPARISON WITH CSP AND P-CSPC OF THE CORRECT RATE IN NORMAL ENVIRONMENT (%)

METHOD		30dB	<5°	<10°
CSP		92.08	31	43
ρ_CSPC	ρ=0.9	*97.50*	*64*	*14*
	ρ=0.78	*95.42*	*60*	*16*

TABLE II
THE COMPARISON WITH CSP AND P-CSPC OF THE CORRECT RATE IN LOW MUSICAL ENVIRONMENT (%)

METHOD		10dB	<5°	<10°
CSP		72.08	27	31
ρ_CSPC	ρ=0.9	*82.92*	*44*	*22*
	ρ=0.78	*85.42*	*51*	*17*

TABLE III
THE COMPARISON WITH CSP AND P-CSPC OF THE CORRECT RATE IN HIGH MUSICAL ENVIRONMENT (%)

METHOD		0dB	<5°	<10°
CSP		44.17	11	24
ρ_CSPC	ρ=0.9	*69.12*	*32*	*23*
	ρ=0.78	*83.75*	*51*	*16*

The accuracy of localization is influenced by SNR. Low SNR means much noise exists, the correct rate is lower than the situation with high SNR and it can be improved when the value of whitening parameter is 0.78. When the signal energy is more powerful than the noise, 0.9 is a better value for the whitening parameter. It also shows the correct rate of time delay estimate in two methods with different values in both whitening parameter and the circumstance noise. Besides that the value of ρ is impacted by the noise. When the noise increases, experimental results will be better with smaller ρ.

P-CSPC method always brings with a sharper peak and a higher accuracy rate even with great noise, it still able to achieve desired effect in real-time. Therefore, it is applied in robots although the calculation of covariance of multi-channel signal is increased.

V. CONCLUSIONS

In this paper, we introduced a real-time acoustic localization algorithm which makes use of coherence function and covariance matrix based on CSP method. Whitening parameter utilized to improve the accuracy of localization and reduce the impact of noise. Covariance matrix computed through the multi-channel signals is also applied to weaken the noise and reverberation. Still, the coherence function with a constant added in the denominator in the weight, in order to reduce the increased errors. Experiments under various conditions show that this algorithm can gain better positioning accuracy compared to the general CSP method. All the 240 groups of data come up with a best localization correct rate of 97.50% in normal environments, 85.42% in low musical condition and 83.75% in a strong musical environment. However, there are still areas that need be improved. It is possible to get better accuracy by applying the algorithm twice in a very short time interval. This indicates that when the person goes quickly and passes the specified position before he (she) even finishes a word. Furthermore, the localization system may be reinforced by adjusting the value of the sampling rate.

REFERENCES

[1] J.L. Flanagan, H.F. Silverman, "Workshop on Microphone Arrays: Theory, Design & Application", CAIP Center-Rutgers University, October 1994.

[2] H. Wang, P. Chu, "Voice source localization for automatic camera pointing system in videoconferencing", Proceedings of ICASSP, Germany, 1997, pp. 187-189.

[3] J.G. Ryan, R.A. Goubran, "Application of near-field optimum microphone arrays to hands-free mobile telephony", IEEE Transactions on Vehicular Technology, Canada, 2003, pp. 390-400.

[4] D. Pomerleau, T. Jochem, "Rapidly adapting machine vision for automated vehicle steering", IEEE Expert: Special Issue on Intelligent System and their Applications, 1996, pp. 19-27.

[5] G.C. Carter, "Special Issue on Time Delay Estimation", IEEE Transactions on Acoustic, Speech, and Signal Processing Newsletter, London, 2003, pp. 12.

[6] C.H. Knapp, G.C. Carter, "The generalized correlation method for estimation of time delay", IEEE Transactions on Acoustics, Speech and Signal Processing, 1976, pp. 320-327.

[7] J. Benesty, "Adaptive eigenvalue decomposition algorithm for passive acoustic source localization", Acoustic Society of America, 2000, pp. 384-391.

[8] S.N. Lin, S.J. Chern, "A new adaptive constrained LMS time delay estimation algorithm", Signal Processing, Volume 71, Issue 1, November 1998, pp. 29-44

[9] M. Omologo, P. Svaizer, "Acoustic event localization using a crosspower-spectrum phase based technique", Proceedings of ICASSP, Australia, 1994, pp. 273-276.

[10] M. Omologo, P.Svaizer, "Use of the crosspower-spectrum phase in acoustic event location", IEEE Transactions on Speech Audio Process, 1997, pp. 288-292.

[11] M. Onologo, P. Svaizer, "Acoustic source location in noisy and reverberant environment using CSP analysis", Proceedings of ICASSP, Atlanta America, 1996, pp. 921-924.

[12] D.V. Rabinkin, R.J. Renomeron, A. Dahl, J.C. French, J.L. Flanagan and M.H. Bianchi. "A DSP Implementation of Source Location Using Microphone Arrays", The Journal of the Acoustical Society of America, Volume 99, Issue 4, April 1996, pp. 2510-2527.

[13] D. Nguyen, P. Aarabi and A. Sheikholeslami, "Real-time localization using field-programmable gate arrays", Proceedings of ICASSP, 2003.

[14] A.G. Piersol "Time delay estimation using phase [J]", IEEE TransferAcoustic, Speech, Singnal, Processing, 1981, ASS-19(3), pp. 471-477.

[15] J.F. Chen, W. Ser, "Speech detection using microphone array", Electronic, Letters, 2000, pp: 181-182.

[16] H.F. Silverman, S.E. Kirtman, "A Two-stage Algorithm for Determining Talker Location from Linear Microphone Array Data", Computer Speech and Language (1992) 6, pp. 128-152.