TIME DELAY ESTIMATION USING SPATIAL CORRELATION TECHNIQUES

Jingdong Chen, Jacob Benesty†, and Yiteng (Arden) Huang**

*: Bell Laboratories, Lucent Technologies
600 Mountain Avenue,
Murray Hill, NJ 07974, USA
{jingdong,arden}@research.bell-labs.com

†: Universite du Quebec, INRS-EMT
800 de la Gauchetiere Ouest, Suite 6900,
Montreal, Quebec, H5A 1K6, Canada
benesty@inrs-emt.uquebec.ca

ABSTRACT

Recently there has been an increasing interest in the use of the TDE technique to locate and track acoustic sources in a conferencing environment. Typically, the delay estimate is obtained as the time-lag that maximizes the cross-correlation function between the filtered versions of two received signals. This so-called generalized cross-correlation technique, based on the measurements provided by two sensors, however, suffers significant performance degradation in the presence of reverberation. In this paper, the idea of cross-correlation coefficient between two signals is generalized to the multichannel case by using the notion of spatial prediction. The multichannel spatial correlation matrix is then deduced and applied for the purpose of TDE. It is shown that this new method can take advantage of the redundant information provided by multiple microphone sensors to help the estimator to better cope with reverberation and noise.

1. INTRODUCTION

Time delay estimation (TDE) has been an area of great research interest since the invention of the radar in the 1930s. It has plenty of applications in fields as diverse as radar, sonar, seismology, geophysics, and ultrasonics for identifying and localizing radiating sources [1][2]. Recently there has been an increasing interest in use of the TDE technique to locate and track acoustics sources in a conferencing environment [3]–[5], which also serves as the main motivation for this work.

The objective of the TDE problem is to determine the relative time difference of arrival (TDOA) between signals received by different sensors. By assuming that the signals picked up by two sensors are delayed and attenuated versions of the original source signal, the classical TDE algorithms estimate the relative delay as the time-lag that maximizes the cross-correlation between filtered versions of the received signals. This so-called generalized cross-correlation (GCC) method is well explained in an informative landmark paper by Knapp and Carter [1]. Since then, many efforts have been devoted to evaluate and improve this technique in various application scenarios, and it is shown that GCC is quite successful in localizing and tracking a single source in an open-field environment where no multi-path effect is present. This technique, however, suffers significant performance degradation in the presence of reverberation, which is a common phenomenon in a room environment [6]. The deterioration of performance derives from the inadequacy of the ideal propagation model in describing the TDE problem in the presence of reverberation.

Recently, a more realistic convolutive model has been proposed to describe the TDE problem in a reverberant environment, and the blind channel identification technique has been then applied to solve the TDE problem [7]. This new technique, also using signals received by two sensors, has shown better performance than GCC in the presence of reverberation.

Instead of using only two sensors (microphones), this paper generalizes the idea of cross-correlation to multichannel case, and proposes an approach to estimate one time delay using multiple sensors based on the notion of spatial correlation. This new approach can take advantage of the redundancy among multiple microphones to help the estimator to better deal with both noise and reverberation. Experimental results show that the TDE performance of the new method increases with the number of microphones in the presence of noise and reverberation.

2. SIGNAL MODEL

In an ideal propagation environment where no multipath effect is present, the signals acquired by a microphone array can be expressed as:

$$x_l[n] = \alpha_l s[n - t - f_l(\tau)] + w_l[n], \quad (1)$$

where $\alpha_l, l = 0, 1, 2, ..., L$, are the attenuation factors due to propagation effects, $L+1$ is the total number of sensors, t is the propagation time from the unknown source $s[n]$ to microphone 0, $w_l[n]$ is an additive noise signal at the lth microphone, which is assumed to be a (real) zero-mean stationary Gaussian random process and to be uncorrelated with both the source signal and the noise signals from other microphones, τ is the relative delay between microphones 0 and 1, and $f_l(\tau)$ is the relative delay between microphones 0 and l. The function f_l depends of τ but also of the microphone array geometry. For example, in the far-field case (plane wave propagation), for a linear equispaced array, we have:

$$f_l(\tau) \quad = \quad l\tau, \quad (2)$$

and for a linear non-equispaced array, we have:

$$f_l(\tau) \quad = \quad \frac{\sum_{i=0}^{l-1} d_i}{d_0}\tau, \quad (3)$$

where d_i is the distance between microphones i and $i+1$, $i = 0, 1, 2, ..., L-1$. In the near-field case, f_l depends also on the position of the source. In general τ is not known, but the geometry of the antenna is known such that the exact mathematical relation of the relative delay between microphones 0 and l is well defined and given.

In a reverberant environment, each microphone receives delayed and attenuated replicas of source signal due to reflections of the source wave from boundaries in addition to the direct path signal. In order to model the TDE problem in such a condition, (1) is generalized to the following convolutive model:

$$x_l[n] = h_l * s[n] + w_l[n], \quad (4)$$

where $*$ denotes convolution, and h_l is the channel response between the source and the lth microphone. Note that in the ideal propagation model $h_l = \alpha_l \delta[n - t - f_l(\tau)]$ and (4) reduces to (1).

3. THE GCC METHOD

The GCC method is based on the ideal propagation model with two microphones, *i.e.*, (1) with $l = 0, 1$. The delay estimate is obtained as

$$\begin{aligned} \hat{\Psi}_{\text{GCC}}[n] \quad &= \quad \sum_{k=0}^{N-1} \Phi[k] X_{x_0 x_1}[k] e^{\frac{j2\pi nk}{N}}, \\ \hat{\tau}_{\text{GCC}} \quad &= \quad \arg\max_n \hat{\Psi}_{\text{GCC}}[n], \end{aligned} \quad (5)$$

where $\hat{\Psi}_{\text{GCC}}[k]$ is the generalized cross-correlation function (GCCF), $X_{x_0 x_1}[k] = E\{X_0[k] X_1^*[k]\}$ is the cross-spectrum, $E\{\cdot\}$ and *

stand respectively for the expectation and complex conjugate operator, $X_l[k]$ is the discrete Fourier transform (DFT) of the signal $x_l[n]$, $\Phi[k]$ is a weighting function, and N denotes the number of observation samples during the observation interval. The choice of $\Phi[k]$ is of great importance in practice. The classical cross-correlation method is obtained by taking $\Phi[k] = 1$. In the noise-free condition, knowing that $X_l[k] = S_l[k]H_l[k]$, where $S_l[k]$ is the DFT of the source signal and $H_l[k]$ represents the channel transfer function, we immediately have

$$X_{x_0x_1}[k] = H_0[k]E\{|S[k]|^2\}H_1^*[k]. \tag{6}$$

One can see that the cross-spectrum depends not only on the channel response but on the source signal as well. The presence of the source signal in the cross-spectrum could be problematic for TDE.

In the so-called phase transform (PHAT) algorithm [1], the weighting function is chosen as $\Phi[k] = 1/|X_{x_0x_1}[k]|$. A simple calculation will show that this selection of $\Phi[k]$ makes the weighted cross-spectrum depend only on the channel response. Many reported results show that the PHAT approach is superior to the classical cross-correlation method.

4. ADAPTIVE EIGENVALUE DECOMPOSITION ALGORITHM

The adaptive eigenvalue decomposition algorithm (AEDA) is also a two-sensor technique, which is based on the convolutive model, *i.e.*, (4) with $l = 0, 1$. By following (4) and the fact that

$$x_0[n] * h_l = x_1[n] * h_0, \tag{7}$$

in the noise-free case, the following relation holds at time n [7]:

$$\mathbf{x}^T[n]\mathbf{u} = \mathbf{x}_0^T[n]\mathbf{h}_1 - \mathbf{x}_1^T[n]\mathbf{h}_0 = 0, \tag{8}$$

where ^{T} indicates transpose, and

$$\begin{aligned}
\mathbf{x}_l[n] &= [x_l[n], x_l[n-1], \ldots, x_l[n-M+1]]^T, \\
\mathbf{h}_l &= [h_{l,0}, h_{l,1}, \ldots, h_{l,M-1}]^T, \\
\mathbf{x}[n] &= \left[\mathbf{x}_0^T[n], \mathbf{x}_1^T[n]\right]^T, \\
\mathbf{u} &= \left[\mathbf{h}_1^T, -\mathbf{h}_0^T\right]^T,
\end{aligned}$$

and M is the length of the impulse responses. Left multiplying (8) by $\mathbf{x}[n]$ and taking expectation yields

$$\mathbf{R}[n]\mathbf{u} = 0, \tag{9}$$

where $\mathbf{R}[n] = E\{\mathbf{x}[n]\mathbf{x}^T[n]\}$ is the covariance matrix of the microphone signals. This means that the vector $\mathbf{u}$ (consisting of two impulse responses) is the eigenvector of $\mathbf{R}$ corresponding to the eigenvalue 0. [7] presented several adaptive algorithms to search for the eigenvector $\mathbf{u}$ corresponding to the minimum eigenvalue of $\mathbf{R}$ providing that the acoustic channels are identifiable. A simple one is to iteratively update $\mathbf{u}$ through:

$$e[n] = \mathbf{u}^T[n]\mathbf{x}[n], \tag{10}$$

and

$$\mathbf{u}[n+1] = \frac{\mathbf{u}[n] - \mu e[n]\mathbf{x}[n]}{||\mathbf{u}[n] - \mu e[n]\mathbf{x}[n]||}, \tag{11}$$

where μ, the adaptation step, is a positive constant.

In practice, (11) may not produce accurate estimation of the vector $\mathbf{u}$ because of the nonstationarity of speech, the background noise, and the unknown length of the impulse responses. However, it yields a solution accurate enough for the purpose of TDE since such an application only needs to detect the direct paths of the two impulse responses.

5. THE PROPOSED ALGORITHM

As opposed to GCC and AEDA, the proposed algorithm is a multi-sensor technique which aims to estimate one delay using an array of microphones. Without loss of generality, we assume to estimate τ (relative delay between microphone 0 and 1) in the ideal propagation model given in (1) with $L + 1$ microphones.

5.1. Spatial Prediction Technique

The notion of spatial prediction was presented in [8] but in the simple case that makes the spatial prediction equivalent to the classical linear prediction. In this section, we generalize this idea in a way that the geometry of the array is taken into account as well as the relative delay among the elements of this array.

Spatial prediction can be formulated as spatial forward prediction which is to align successive observation samples of microphone 0 with spatial samples from the L other microphone signals, spatial backward prediction which tries to align successive observation samples of the Lth microphone with spatial samples from the L other microphone signals, and more generally, spatial interpolation which aligns successive observation samples of the lth microphone with spatial samples from the L other microphone signals. For the limited space, here we only consider the spatial forward prediction. It is trivial to generalize the idea of spatial forward prediction to spatial backward prediction and spatial interpolation.

It is clear that $x_0[n - f_L(\tau)]$ is in-phase with the signals $x_l[n - f_L(\tau) + f_l(\tau)]$, $l = 1, 2, ..., L$. From these observations, we define the following forward spatial prediction error signal:

$$e_0[n - f_L(m)] = x_0[n - f_L(m)] - \mathbf{x}_{1:L}^T[n - f_L(m)]\mathbf{a}_m, \tag{12}$$

where m is any guessed relative delay,

$$\mathbf{x}_{1:L}[n - f_L(m)] = [x_1[n - f_L(m) + f_1(m)] \ \cdots \ x_L[n]]^T,$$

and

$$\mathbf{a}_m = [\ a_{m,1} \quad a_{m,2} \quad \cdots \quad a_{m,L}\ ]^T$$

is the linear forward spatial predictor. The minimization of the criterion:

$$J_{m,0} = E\{e_0^2[n - f_L(m)]\}, \tag{13}$$

gives

$$\mathbf{R}_{m,1:L}\mathbf{a}_m = \mathbf{r}_{m,1:L}, \tag{14}$$

where

$$\mathbf{R}_{m,1:L} = E\{\mathbf{x}_{1:L}[n - f_L(m)]\mathbf{x}_{1:L}^T[n - f_L(m)]\}, \tag{15}$$

is the spatial correlation (SC) matrix, and

$$\mathbf{r}_{m,1:L} = E\{\mathbf{x}_{1:L}[n - f_L(m)]x_0[n - f_L(m)]\}, \tag{16}$$

is the spatial correlation vector.

Note that the spatial correlation matrix given in (15) is not Toeplitz in general, except for some particular cases.

For $m = \tau$ and for the noise-free case where $w_l[n] = 0$, $l = 1, 2, ..., L$, it can easily be checked that with the ideal propagation model given in (1), the rank of matrix $\mathbf{R}_{\tau,1:L}$ is equal to 1. This means that the sample $x_0[n - \tau]$ can be perfectly predicted from any one of the other microphone samples. However, the noise is never zero in practice and is in general isotropic. The energy of the different noise components at the microphones will be added at the main diagonal of the correlation matrix $\mathbf{R}_{\tau,1:L}$. This regularized matrix will then become positive definite (which we suppose in the rest of this paper). A unique solution to (14) is then guaranteed whatever the number of microphones would be. This solution is optimal from a Wiener theory point of view.

5.2. Application to Time Delay Estimation

Let $J_{m,0;\min}$ denote the minimum mean-squared error, for the value m, defined by

$$J_{m,0;\min} = E\{e_{0;\min}^2[n - f_L(m)]\}. \tag{17}$$

If we replace $\mathbf{a}_m$ by $\mathbf{R}_{m,1:L}^{-1}\mathbf{r}_{m,1:L}$ in (12), we get:

$$e_{0;\min}[n - f_L(m)] = x_0[n - f_L(m)] - \mathbf{x}_{1:L}^T[n - f_L(m)]\mathbf{R}_{m,1:L}^{-1}\mathbf{r}_{m,1:L}. \tag{18}$$

We deduce that:

$$J_{m,0;\min} = E\{x_0^2[n - f_L(m)]\} - \mathbf{r}_{m,1:L}^T \mathbf{R}_{m,1:L}^{-1} \mathbf{r}_{m,1:L}. \quad (19)$$

The value of m that gives the minimum $J_{m,0;\min}$, for different m, corresponds to the time delay between microphone 0 and 1. Mathematically, the solution to our problem is then given by

$$\hat{\tau} = \arg\min_m J_{m,0;\min}, \quad (20)$$

where $\hat{\tau}$ is an estimate of τ.

Particular case: Two microphones ($L = 1$). In this case, we have

$$\begin{aligned} \hat{\tau} &= \arg\min_m \{1 - \rho_{m,01}^2\} \\ &= \arg\max_m (\rho_{m,01}^2), \end{aligned} \quad (21)$$

where $\rho_{m,01}$ ($\rho_{m,01}^2 \leq 1$) is the cross-correlation coefficient between $x_0[n-m]$ and $x_1[m]$. When the cross-correlation coefficient is close to 1, this means that the two signals that we compare are highly correlated which happens when the signals are in-phase, i.e. $m \approx \tau$ and this implies that $J_{\tau,0;\min} \approx 0$. This approach is similar to the GCC method [1]. Note that in the general case with any number of microphones, the proposed approach can be seen as a cross-correlation method, but we take advantage of the knowledge of the microphone array to estimate only one time delay (instead of estimating multiple time delays independently) in an optimal way in a least mean square sense.

5.3. Other Information from the Spatial Correlation Matrix

Consider the $L+1$ microphone signals $x_l,\ l = 0, 1, ..., L$, the corresponding spatial correlation matrix is:

$$\begin{aligned} \mathbf{R}_{m,0:L} &= \mathbf{R}_m \\ &= E\{\mathbf{x}_{0:L}[n - f_L(m)]\mathbf{x}_{0:L}^T[n - f_L(m)]\}. \end{aligned} \quad (22)$$

It can be shown that $\mathbf{R}_m$ can be factored as:

$$\mathbf{R}_m = \mathbf{D}\widetilde{\mathbf{R}}_m\mathbf{D}, \quad (23)$$

where

$$\mathbf{D} = \begin{bmatrix} \sqrt{E\{x_0^2[n]\}} & 0 & \cdots & 0 \\ 0 & \sqrt{E\{x_1^2[n]\}} & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \sqrt{E\{x_L^2[n]\}} \end{bmatrix} \quad (24)$$

is a diagonal matrix,

$$\widetilde{\mathbf{R}}_m = \begin{bmatrix} 1 & \rho_{m,01} & \cdots & \rho_{m,0L} \\ \rho_{m,01} & 1 & \cdots & \rho_{m,1L} \\ \vdots & \ddots & \ddots & \vdots \\ \rho_{m,0L} & \cdots & \rho_{m,L-1L} & 1 \end{bmatrix} \quad (25)$$

is a symmetric matrix, and

$$\rho_{m,kl} = \frac{E\{x_k[n - f_l(m)]x_l[n - f_k(m)]\}}{\sqrt{E\{x_k^2[n]\}E\{x_l^2[n]\}}}, \quad k, l = 0, 1, ..., L, \quad (26)$$

is the cross-correlation coefficient between $x_k[n - f_l(m)]$ and $x_l[n - f_k(m)]$.

It can be shown that the determinant of the matrix $\widetilde{\mathbf{R}}_m$ satisfies:

$$0 < \det\left(\widetilde{\mathbf{R}}_m\right) \leq 1, \quad (27)$$

where "det" stands for *determinant*, and

$$\det\left(\widetilde{\mathbf{R}}_m\right) \leq \frac{J_{m,0;\min}}{E\{x_0^2[n]\}} \leq 1. \quad (28)$$

These two inequalities gives the boundaries of $\det\left(\widetilde{\mathbf{R}}_m\right)$, which is useful for TDE.

As we can see, the determinant of the spatial correlation matrix is related to the minimum mean-squared error and to the correlation of the signals. Let's take the two-channel case. It is obvious that the cross-correlation coefficient between the two signals x_0 and x_1 is linked to the determinant of the corresponding spatial correlation matrix:

$$\rho_{m,01}^2 = 1 - \det\left(\widetilde{\mathbf{R}}_{m,0:1}\right). \quad (29)$$

By analogy to the cross-correlation coefficient definition between two signals, we define the multichannel correlation coefficient among the signals $x_l,\ l = 0, 1, ..., L$, as:

$$\rho_{m,0:L}^2 = 1 - \det\left(\widetilde{\mathbf{R}}_{m,0:L}\right). \quad (30)$$

From (28), we give a new bound for $\rho_{m,0:L}^2$:

$$1 - \frac{J_{m,0;\min}}{E\{x_0^2[n]\}} \leq \rho_{m,0:L}^2 \leq 1. \quad (31)$$

Basically, the coefficient $\rho_{m,0:L}$ will measure the amount of correlation among all the channels. This coefficient has some interesting properties. For example, if one of the signals, say x_0, is completely decorrelated from the others because the microphone is defective, or it picks up only noise, or the signal is saturated, this signal will not affect $\rho_{m,0:L}$ since $\rho_{m,0l} = 0,\ \forall l$. In this case:

$$\rho_{m,0:L}^2 = \rho_{m,1:L}^2. \quad (32)$$

In other words, the measure "drops" the signals who have no correlation with the others. This makes sense from a correlation point of view, since we want to measure the degree of correlation only from the channels who have something in common. In the extreme cases where all the signals are uncorrelated, we have $\rho_{m,0:L}^2 = 0$, and where any two signals (or more) are perfectly correlated, we have $\rho_{m,0:L}^2 = 1$.

Obviously, the multichannel coefficient $\rho_{m,0:L}^2$ can be used for time delay estimation in the following way:

$$\begin{aligned} \hat{\tau} &= \arg\max_m \left(\rho_{m,0:L}^2\right) \\ &= \arg\min_m \left[\det\left(\widetilde{\mathbf{R}}_{m,0:L}\right)\right]. \end{aligned} \quad (33)$$

This method can be seen as a multichannel correlation approach for the estimation of time delay and it is clear that (33) is equivalent to (20).

6. PERFORMANCE EVALUATION

6.1. Experiment Setup

Experiments were carried out in the Varechoic Chamber which is a unique facility at Bell Laboratories. The chamber is a $6.7 \times 6.1 \times 2.9$ m room whose surfaces are covered by a total of 369 active panels which can be controlled digitally. Each panel consists of two perforated sheets. When the holes in the sheets are aligned, absorbing material behind the sheets will be exposed to the sound field, whereas a highly reflective surface can be formed if the holes are shifted to misalignment. Combination of open and closed panels can produce 2^{369} different acoustic environments where the 60-dB reverberation time T_{60} can change from 0.2 to almost 1 second. See [9] for more details.

A linear microphone array which consists of 22 omnidirectional Panasonic WM-61A microphones was mounted at the distance of 0.5 m from the north wall of the chamber and approximately at the center of the wall. The 22 microphones are uniformly distributed along an aluminum rod whose diameter is 1 cm. The spacing between adjacent microphones is 10 cm. The source signal is played by a Cabasse Baltic Murale loudspeaker in 46 different positions. An illustration of this setup is shown in Fig. 1.

For the purpose of data reusability, the impulse response from each source location to each microphone was measured [10]. The observed signal is then obtained by convolution of a recorded speech signal with the measured impulse responses. The measurement of the impulse responses were performed using the built-in measurement tool of the Huron Lake system. A 65536-point long logarithmic sweep signal digitized at a sampling rate of 48 kHz was used as the excitation signal. From each source location to each microphone, the excitation is played and recorded. An estimate of the transfer function is obtained by spectral division between the original source excitation and the recorded microphone signal.

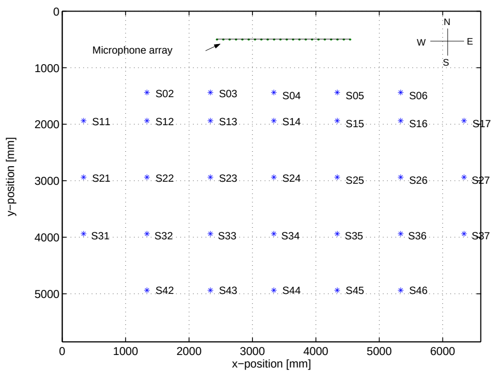

Fig. 1. Layout of the microphone array and source positions in the Varechoic Chamber.

6.2. TDE performance

A series of experiments were conducted to compare the proposed, the CC, the PHAT, and the AEDA algorithms, as well as how the performance of the proposed method is affected by the number of microphones. For brevity, we report two sets of experimental results: one involves a set in a light reverberation condition where the reverberation time, T_{60}, which is defined as the time for the sound to die away to a level 60 decibels below its original level and measured by the Schroeder's method [11], is approximately 240 ms, and the other pertains to a highly reverberant environment where $T_{60} = 580$ ms. The observed signal is obtained by convolution of 2-minute speech (sampled at 16 kHz) from a female speaker with the down-sampled impulse responses. Computer-generated white Gaussian noise is then added to the signal to control the signal-to-noise ratio (SNR) to be 0 dB. A delay estimate is obtained every 128 milliseconds. Figure 2 plots the mean square errors (MSE) for the delay estimates obtained by the CC, PHAT, AEDA, and the proposed SC algorithms respectively.

The CC and PHAT algorithms are based on the ideal propagation model. They work fairly well in lightly reverberant environments. The PHAT method generally outperforms the CC approach in weak noisy conditions. In strong noisy, or highly reverberant environment, however, these two methods deliver almost similar performance, as can be seen in Fig. 2.

The AEDA algorithm is derived from the convolutive model, which accurately describes the signal propagation in reverberation conditions. It works well from nonreverberant to highly reverberant environments. The AEDA algorithm is superior to both CC and PHAT in most situations.

As opposed to CC, PHAT, and AEDA, the SC algorithm is a multi-sensor technique, which employs multiple microphones to estimate one delay. Although spawned from the ideal propagation model, it can be seen from Fig. 2 that the SC algorithm can take advantage of the redundancy among multiple microphones to produce better performance in noise and reverberation conditions. Although in two-microphone case (equivalent to CC), its performance is worse than that of AEDA, it is remarkable that the SC approach outperforms AEDA when one more microphone is available in the studied conditions.

7. CONCLUSIONS

Estimating TDOA in reverberant environments remains to be a challenging problem. This paper presented a multi-sensor algorithm based on the spatial correlation technique for the purpose of TDE. This new algorithm can be treated as a natural generalization of the classical cross-correlation method to multichannel case. Experimental results demonstrated that this new approach can take advantage of the redundant information provided by multiple microphones to make the delay estimator more robust with respect to both noise and reverberation. Comparison with CC, PHAT, and AEDA also justified the effectiveness of the new algorithm.

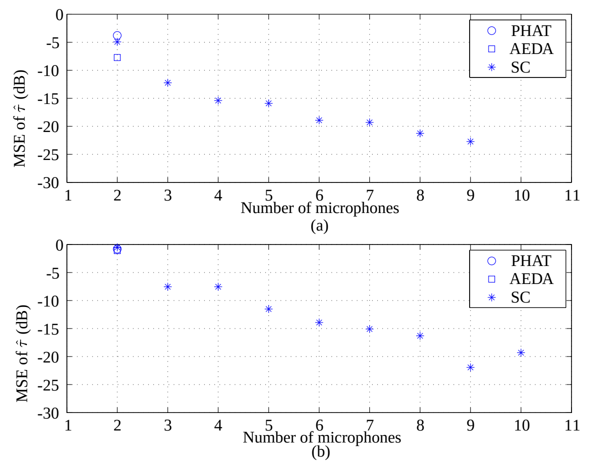

Fig. 2. The TDE performances of the CC, PHAT, AEDA, and SC methods in noisy and reverberant environments: (a) SNR = 0 dB, $T_{60} = 240$ ms, sound source is in S31; (b) SNR = 0 dB, $T_{60} = 580$ ms, sound source is in S32.

8. REFERENCES

[1] C. H. Knapp and G. C. Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 24, pp. 320–327, Aug. 1976.

[2] R. J. Vaccaro, "The past, present, and future of underwater acoustic signal processing," *IEEE signal Processing Magazine*, vol. 15, pp. 21–51, July 1998.

[3] D. R. Fischell and C. H. Coker, "A speech direction finder," in *Proc. IEEE ICASSP*, 1984, pp. 19.8.1–19.8.4.

[4] M. Omologo and P. Svaizer, "Acoustic event localization using a crosspower-spectrum phase based technique," in *Proc. IEEE ICASSP*, 1994, pp. II-273–II-276.

[5] Y. Huang, J. Benesty, and G. W. Elko, "Microphone arrays for video camera steering," in *Acoustic Signal Processing for Telecommunication*, S. L. Gay and J. Benesty, eds., Kluwer Academic Publishers, chap. 11, pp. 239–259, 2000.

[6] B. Champagne, S. Bédard, and A. Stéphenne, "Performance of time-delay estimation in presence of room reverberation," *IEEE Trans. Speech and Audio Processing*, vol. 4, pp. 148-152, Mar. 1996.

[7] J. Benesty, "Adaptive eigenvalue decomposition algorithm for passive acoustic source localization," *J. Acoust. Soc. Am.*, vol. 107, pp. 384–391, Jan. 2000.

[8] S. Haykin, "Radar array processing for angle of arrival estimation," in *Array Signal Processing*, S. Haykin, ed., Prentice-Hall, chap. 4, pp. 194–292, 1985.

[9] W. C. Ward, G. W. Elko, R. A. Kubli, and W. C. McDougald, "The new varechoic chamber at AT&T Bell Labs," in *Proc. Wallance Clement Sabine Centennial Symposium*, 1994, pp. 343-346.

[10] A. Harma, "Acoustic measurement data from the varechoic chamber," *Technical Memorandum*, No. 110101, Agere Systems, 2001.

[11] M. R. Schroeder, "New method for measuring reverberation time," *J. Acoust. Soc. Am.*, Vol. 37, pp. 409–412, 1965.