

TIME DELAY ESTIMATION VIA MULTICHANNEL CROSS-CORRELATION

Jingdong Chen, Yiteng (Arden) Huang

Jacob Benesty

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

Université du Québec, INRS-EMT
800 de la Gauchetière Ouest, Suite 6900
Montréal, Québec, H5A 1K6, Canada
benesty@inrs-emt.quebec.ca

ABSTRACT

Time delay estimation (TDE) in a reverberant acoustical environment is a very challenging and difficult problem. This paper tackles the problem by exploiting the redundant information provided by multiple microphone sensors. To do so, the multichannel cross-correlation coefficient (MCCC) is re-derived, in a new way, to connect it to the well-known linear interpolation technique. Some interesting properties and bounds of MCCC are discussed, and a recursive algorithm is then introduced so that MCCC can be estimated and updated efficiently when new data snapshots are available. We then apply the MCCC to the TDE problem, resulting a multichannel cross-correlation algorithm that can be treated as a natural generalization of the generalized cross-correlation (GCC) TDE method to the multichannel case. It is shown that this method can take advantage of the redundancy provided by multiple microphone sensors to improve TDE against both reverberation and noise.

1. INTRODUCTION

Time delay estimation (TDE), which is a fundamental approach for identifying, localizing, and tracking radiating sources, has attracted a considerable amount of research attention for decades. Recently, there has been a growing interest in the use of the TDE technique to locate and track acoustic sources in a conferencing environment, which also serves as the main motivation for this work.

The objective of TDE is to determine the relative time difference of arrival (TDOA) between signals received by different sensors. The generalized cross-correlation (GCC) method is the most popular to do so and is well explained in an informative paper by Knapp and Carter [1]. This technique is quite successful in localizing and tracking a single source in an open-field environment where no multi-path effect is present, but suffers significant performance degradation in the presence of reverberation, which is a common phenomenon in a conferencing room environment [2].

Much attention has been paid to combating reverberation lately. Most of such efforts fall into two categories. The first is to blindly estimate the channel impulse responses from the source to the two microphones [3], [4]. The better this estimate is, the better the relative delay between these two microphone signals can be estimated; but this is a difficult problem and the resulting time delay estimates are sensitive to noise. The second is to use more than two microphones and take advantage of the redundancy [5], [6]. In [7], we developed a spatial correlation based TDE technique, which belongs to the second category. This method can be seen as a natural generalization of the GCC approach to the multichannel case, and is able to improve TDE when more microphones are available. This paper is an extension of the work presented in [7]. The contributions of this paper are threefold. First of all, we re-derive the multichannel cross-correlation coefficient (MCCC), in a new way, to connect it to the well-known linear interpolation technique, and discuss its interesting properties and bounds. Secondly, we introduce a recursive algorithm so that the MCCC can be estimated and updated efficiently when new data snapshots are available. Finally, we apply the MCCC to the TDE problem, resulting in a multichannel correlation algorithm. This algorithm is

able to take advantage of the redundancy provided by multiple microphone sensors to improve TDE against both reverberation and noise.

2. LINEAR INTERPOLATION

We assume that we have L signals $x_0(n), x_1(n), \dots, x_{L-1}(n)$, and we seek to determine how any one of these signals can be interpolated from the others. To interpolate $x_i(n)$ from the rest, we need to minimize the criterion [8]

$$J_i(n) = \sum_{p=0}^n \lambda^{n-p} \left[-\sum_{l=0}^{L-1} c_{il}(n)x_l(p) \right]^2$$

$$= \sum_{p=0}^n \lambda^{n-p} \left[-\mathbf{c}_i^T(n)\mathbf{x}(p) \right]^2 = \mathbf{c}_i^T(n)\mathbf{R}(n)\mathbf{c}_i(n) \quad (1)$$

with the constraint

$$\mathbf{c}_i^T(n)\mathbf{u}_i = c_{ii} = -1, \quad (2)$$

where λ ($0 < \lambda \leq 1$) is a forgetting factor,

$$\mathbf{c}_i(n) = [c_{i0}(n) \ c_{i1}(n) \ \dots \ c_{i(L-1)}(n)]^T$$

is a vector used to compute the interpolation error (this vector without the component c_{ii} is the i th ($0 \leq i \leq L-1$) interpolator of the vector signal),

$$\mathbf{x}(n) = [x_0(n) \ x_1(n) \ \dots \ x_{L-1}(n)]^T,$$

$$\mathbf{u}_i = [0 \ \dots \ 0 \ 1 \ 0 \ \dots \ 0]^T$$

is a vector of length L where its i th component is equal to one and all others are zero, and

$$\mathbf{R}(n) = \sum_{p=0}^n \lambda^{n-p} \mathbf{x}(p)\mathbf{x}^T(p) \quad (3)$$

is an estimate of the signal covariance matrix. Matrix $\mathbf{R}(n)$ is positive semi-definite; but in the rest, we suppose that it is positive definite so it is invertible.

By using a Lagrange multiplier, it is easy to see that the solution to this optimization problem is:

$$\mathbf{R}(n)\mathbf{c}_i(n) = -E_i(n)\mathbf{u}_i, \quad (4)$$

where

$$E_i(n) = \mathbf{c}_i^T(n)\mathbf{R}(n)\mathbf{c}_i(n) = \frac{1}{\mathbf{u}_i^T \mathbf{R}^{-1}(n)\mathbf{u}_i} \quad (5)$$

is the interpolation error energy. Since

$$-\frac{\mathbf{c}_i(n)}{E_i(n)} = \mathbf{R}^{-1}(n)\mathbf{u}_i, \quad (6)$$

then the i th column of $\mathbf{R}^{-1}(n)$ is $-\mathbf{c}_i(n)/E_i(n)$. We deduce that $\mathbf{R}^{-1}(n)$ can be factorized as follows:

$$\mathbf{R}^{-1}(n) = \begin{bmatrix} 1 & -c_{10}(n) & \dots & -c_{(L-1)0}(n) \\ -c_{01}(n) & 1 & \dots & -c_{(L-1)1}(n) \\ \vdots & \vdots & \ddots & \vdots \\ -c_{0(L-1)}(n) & -c_{1(L-1)}(n) & \dots & 1 \end{bmatrix}$$

$$\begin{bmatrix} 1/E_0(n) & 0 & \cdots & 0 \\ 0 & 1/E_1(n) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1/E_{L-1}(n) \end{bmatrix} \triangleq \mathbf{C}^T(n)\mathbf{D}_E^{-1}(n). \quad (7)$$

Since $\mathbf{R}^{-1}(n)$ is a symmetric matrix, (7) can also be written as:

$$\mathbf{R}^{-1}(n) = \mathbf{D}_E^{-1}(n)\mathbf{C}(n). \quad (8)$$

The first and last columns of $\mathbf{R}^{-1}(n)$ contain respectively the normalized forward and backward predictors and all the columns between contain the normalized interpolators. $\mathbf{C}(n)$ is simply the matrix of the interpolators and $\mathbf{D}_E(n)$ is a diagonal matrix containing all the respective interpolation error energies.

3. MULTICHANNEL CROSS-CORRELATION COEFFICIENT

The definition of multiple coherence function, derived from the concepts of the ordinary coherence function between two signals and the partial (conditioned) coherence function, was presented in [9] to measure the correlation between the output of a MISO (multiple-input/single-output) system and its inputs. In this section, we re-derive the multichannel cross-correlation coefficient (MCCC) in a new way such that it is related to the multichannel correlation matrix. With this new definition, the MCCC can be treated as a generalization of the classical cross-correlation coefficient to the case where we have more than two processes. The covariance matrix can be factorized as follows:

$$\mathbf{R}(n) = \mathbf{D}_x^{1/2}(n)\tilde{\mathbf{R}}(n)\mathbf{D}_x^{1/2}(n), \quad (9)$$

where

$$\mathbf{D}_x^{1/2}(n) = \begin{bmatrix} \sqrt{r_{00}(n)} & 0 & \cdots & 0 \\ 0 & \sqrt{r_{11}(n)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \sqrt{r_{(L-1)(L-1)}(n)} \end{bmatrix}, \quad (10)$$

$$\tilde{\mathbf{R}}(n) = \begin{bmatrix} 1 & \rho_{10}(n) & \cdots & \rho_{(L-1)0}(n) \\ \rho_{10}(n) & 1 & \cdots & \rho_{(L-1)1}(n) \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{(L-1)0}(n) & \rho_{(L-1)1}(n) & \cdots & 1 \end{bmatrix}, \quad (11)$$

$$r_{ij}(n) = \sum_{p=0}^n \lambda^{n-p} x_i(p)x_j(p), \quad i, j = 0, 1, \dots, L-1, \quad (12)$$

and

$$\rho_{ij}(n) = \frac{r_{ij}(n)}{\sqrt{r_{ii}(n)r_{jj}(n)}}, \quad i, j = 0, 1, \dots, L-1. \quad (13)$$

$\rho_{ij}(n)$ is the cross-correlation coefficient between $x_i(n)$ and $x_j(n)$.

Since matrix $\tilde{\mathbf{R}}(n)$ is symmetric, positive definite, and its diagonal elements are all equal to one, it can be shown that:

$$0 < \det[\tilde{\mathbf{R}}(n)] \leq 1, \quad (14)$$

where “det” stands for determinant.

We can now define the squared MCCC among the L signals $x_0(n), x_1(n), \dots, x_{L-1}(n)$, as:

$$\rho_L^2(n) \triangleq 1 - \det[\tilde{\mathbf{R}}(n)] = 1 - \frac{\det[\mathbf{R}(n)]}{\prod_{l=0}^{L-1} r_{ll}(n)}. \quad (15)$$

This definition is identical to the one given in [10] using the Gram determinant. For two ($L = 2$) processes $x_0(n)$ and $x_1(n)$, we have:

$$\rho_2^2(n) = \frac{r_{01}^2(n)}{r_{00}(n)r_{11}(n)}, \quad (16)$$

which is the classical definition of the squared cross-correlation coefficient. It can be shown that $\rho_L^2(n)$ has the following properties. (Proofs are omitted here due to the limited space available.)

- $0 \leq 1 - \frac{E_i(n)}{r_{ii}(n)} \leq \rho_L^2(n) \leq 1 - \prod_{l=0}^{L-1} \frac{E_l(n)}{r_{ll}(n)} \leq 1, \forall i \in [0, L-1]$.
- If two or more signals are perfectly correlated, then $\rho_L^2(n) = 1$.
- If all the processes are completely uncorrelated with each other, then $\rho_L^2(n) = 0$.
- If one of the signals is completely uncorrelated with the $L-1$ other signals, then MCCC will measure the correlation among those $L-1$ remaining signals.

Now if we define $\rho_{i,j}^2$ as

$$\rho_{i,j}^2 = 1 - \det[\tilde{\mathbf{R}}_{i,j}(n)], \quad (17)$$

where $j \geq i$, and $\tilde{\mathbf{R}}_{i,j}(n)$ is a $(j-i) \times (j-i)$ matrix whose elements are taken as a block from the (i,i) -position to the (j,j) -position from the matrix $\tilde{\mathbf{R}}(n)$. We then can derive that:

$$\rho_{0:L-1}^2(n) = \rho_L^2(n) = 1 - \prod_{l=0}^{L-1} \frac{E_{l,l:L-1}(n)}{r_{ll}(n)}, \quad (18)$$

where $E_{l,l:L-1}(n)$ is the forward prediction error energy (of order $L-1-l$) using the signals $x_l(n), x_{l+1}(n), \dots, x_{L-1}(n)$ and $E_{0,0:L-1}(n) = E_0(n)$. This shows how the MCCC is related to the different orders of the forward linear prediction energies. It can also be shown with the different orders of the backward linear prediction energies:

$$\rho_{L-1:0}^2(n) = \rho_L^2(n) = 1 - \prod_{l=0}^{L-1} \frac{E_{L-1-l,L-1-l:0}(n)}{r_{ll}(n)}, \quad (19)$$

where $E_{L-1-l,L-1-l:0}(n)$ is the backward prediction error energy (of order $L-1-l$) and $E_{L-1,L-1:0}(n) = E_{L-1}(n)$. Obviously, we can generalize this approach to any linear interpolator. For two ($L = 2$) processes $x_0(n)$ and $x_1(n)$, we have:

$$\rho_2^2(n) = \frac{r_{01}^2(n)}{r_{00}(n)r_{11}(n)} = 1 - \frac{E_0(n)}{r_{00}(n)} = 1 - \frac{E_1(n)}{r_{11}(n)}. \quad (20)$$

4. APPLICATION TO TIME DELAY ESTIMATION

4.1. Signal Model

Suppose that we have a linear array, which consists of L microphones whose outputs are denoted as $x_l(n), l = 0, 1, \dots, L-1$. Without loss of generality, we select Microphone 0 as the reference point and consider the following propagation model:

$$x_l(n) = \alpha_l s[n - t - f_l(\tau)] + w_l(n), \quad (21)$$

where α_l , $l = 0, 1, 2, \dots, L-1$, are the attenuation factors due to propagation effects, t is the propagation time from the unknown source $s(n)$ to Microphone 0, $w_l(n)$ is an additive noise signal at the l th microphone, τ is the relative delay between Microphones 0 and 1, and $f_l(\tau)$ is the relative delay between microphones 0 and l , with $f_0(\tau) = 0$ and $f_1(\tau) = \tau$. The function f_l (for $l > 1$) depends on τ but also on the microphone array geometry. It can be specified for arbitrary arrays in one, two, or three dimensions. In this paper we are considering only linear arrays. In the far-field case (i.e., plane wave propagation), if the array is equispaced, we have $f_l(\tau) = l\tau$, and if it is not equispaced, we have $f_l(\tau) = [\sum_{i=0}^{l-1} d_i/d_0]\tau$, where d_i is the distance between Microphones i and $i+1$, $i = 0, 1, 2, \dots, L-2$. In the near-field case, f_l depends also on the position of the source. In this paper, we focus only on the far-field case. In such a situation, 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. It is further assumed that $w_l(n)$ is a zero-mean Gaussian random process that is uncorrelated with $s(n)$ and the noise signals at other microphones. It is also assumed that $s(n)$ is reasonably broadband.

4.2. Time Delay Estimation Based on the Squared MCCC

We are interested in estimating only one time delay (τ) from multiple sensors. Consider the following vector:

$$\mathbf{x}(n, m) = [x_0(n) \ x_1[n + f_1(m)] \ \dots \ x_{L-1}[n + f_{L-1}(m)]]^T. \quad (22)$$

We can check that for $m = \tau$, all the signals $x_l[n + f_l(\tau)]$, $l = 0, 1, \dots, L-1$, are aligned. This observation is essential because it already gives an idea on how to estimate τ . An estimate of the covariance matrix corresponding to the signal $\mathbf{x}(n, m)$ is:

$$\begin{aligned} \mathbf{R}(n, m) &= \sum_{p=0}^n \lambda^{n-p} \mathbf{x}(p, m) \mathbf{x}^T(p, m) \\ &= \lambda \mathbf{R}(n-1, m) + \mathbf{x}(n, m) \mathbf{x}^T(n, m). \end{aligned} \quad (23)$$

Therefore, the squared MCCC is:

$$\rho_L^2(n, m) = 1 - \frac{\det[\mathbf{R}(n, m)]}{\prod_{l=0}^{L-1} r_{ll}(n, m)}, \quad (24)$$

where

$$r_{ll}(n, m) = \sum_{p=0}^n \lambda^{n-p} x_l^2[p + f_l(m)], \quad l = 0, 1, \dots, L-1. \quad (25)$$

The value of m that gives the maximum of $\rho_L^2(n, m)$, for different m , corresponds to the time delay between microphones 0 and 1. Hence, the solution to our problem is:

$$\hat{\tau} = \arg \max_m \rho_L^2(n, m), \quad (26)$$

where $m \in [-\tau_{\max}, \tau_{\max}]$, and τ_{\max} is the maximum possible delay. When there are only two microphones available, (26) becomes

$$\hat{\tau} = \arg \max_m \rho_2^2(n, m) = \arg \max_m \frac{r_{01}^2(n, m)}{r_{00}(n, m) r_{11}(n, m)}. \quad (27)$$

In this case, the approach is similar to the generalized cross-correlation method proposed by Knapp and Carter [1]. When there are more than two microphones available, the approach can be seen as a multichannel cross-correlation method, which can 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-squares sense.

4.3. Recursive Estimation of the Squared MCCC

There are many different ways to estimate the squared MCCC. Here, we propose to estimate the elements of $\rho_L^2(n, m)$ recursively. The recursive estimation of $r_{ll}(n, m)$ is straightforward. Indeed, we have:

$$\begin{aligned} r_{ll}(n, m) &= \lambda r_{ll}(n-1, m) + x_l^2[n + f_l(m)], \\ & \quad l = 0, 1, \dots, L-1, \end{aligned} \quad (28)$$

it is then easy to compute $\prod_{l=0}^{L-1} r_{ll}(n, m)$. From (23), we have:

$$\begin{aligned} \frac{1}{\lambda} \mathbf{R}(n, m) \mathbf{R}^{-1}(n-1, m) \\ = \mathbf{I} + \frac{1}{\lambda} \mathbf{x}(n, m) \mathbf{x}^T(n, m) \mathbf{R}^{-1}(n-1, m). \end{aligned} \quad (29)$$

One can notice that the right-hand side of (29) is of the form $\mathbf{I} + \mathbf{y}\mathbf{z}^T$. So it has one eigenvalue equal to $1 + \mathbf{y}^T\mathbf{z}$, and the rest all equal to unity. The determinant, which is the product of all the eigenvalues is therefore equal to $1 + \mathbf{y}^T\mathbf{z}$. We then have:

$$\begin{aligned} \frac{1}{\lambda^L} \det[\mathbf{R}(n, m) \mathbf{R}^{-1}(n-1, m)] \\ = 1 + \frac{1}{\lambda} \mathbf{x}^T(n, m) \mathbf{R}^{-1}(n-1, m) \mathbf{x}(n, m) \triangleq \frac{1}{\varphi(n, m)}, \end{aligned} \quad (30)$$

and finally we get:

$$\det[\mathbf{R}(n, m)] = \frac{\lambda^L}{\varphi(n, m)} \det[\mathbf{R}(n-1, m)]. \quad (31)$$

The inverse of matrix $\mathbf{R}(n, m)$ that appears in $\varphi(n, m)$ can also be calculated recursively:

$$\begin{aligned} \mathbf{R}^{-1}(n, m) &= \lambda^{-1} \mathbf{R}^{-1}(n-1, m) \\ & \quad - \lambda^{-2} \varphi(n, m) \mathbf{k}'(n, m) \mathbf{k}'^T(n, m), \end{aligned} \quad (32)$$

where

$$\mathbf{k}'(n, m) = \mathbf{R}^{-1}(n-1, m) \mathbf{x}(n, m). \quad (33)$$

5. EXPERIMENTS

5.1. Experimental 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.

A linear microphone array which consists of six omnidirectional microphones was employed in the measurement. The six microphone positions are M1 (2.437, 5.600, 1.400), M2 (2.537, 5.600, 1.400), M3 (2.637, 5.600, 1.400), M4 (2.737, 5.600, 1.400), M5 (2.837, 5.600, 1.400), and M6 (2.937, 5.600, 1.400), respectively (coordinates with reference to the lower southwest corner and corresponding to meters along the South wall, West wall, floor). The source was simulated by placing a loudspeaker at (1.337, 4.162, 1.600). The transfer functions of the acoustic channels between the loudspeaker and six microphones were measured at a 48 kHz sampling rate. Then the obtained channel impulse responses were downsampled to a 16 kHz sampling rate and truncated to 4096 samples. These measured impulse responses will be treated as the actual impulse responses in the TDE experiments.

5.2. Experimental Results

The source signal is a speech (from a female speaker) sampled at 16 kHz and of duration 4 minutes. The six-channel observation signals are obtained by convolving the speech source with the corresponding measured channel impulse responses and adding a zero-mean, white, Gaussian noise to each one of these outputs for a given signal-to-noise ratio (SNR).

Two experimental conditions are considered. One pertains to a light reverberant environment where reverberation time, T_{60} , is approximately 240 ms. The other relates to a heavily reverberant environment where $T_{60} = 580$ ms. In both cases, $\text{SNR} = -5$ dB. The multichannel signals are partitioned into non-overlapping frames with a frame size of 128 milliseconds. For each frame, a delay estimate is measured according to the estimator given by (26) with a forgetting factor $\lambda = 0.95$. Therefore, with a 4-minute speech sequence, a total of 1875 time delay estimates are yielded. To evaluate the performance, we classify an estimate into two comprehensive categories: the class of success and the class of failure [2]. An estimate $\hat{\tau}_i$ for which the absolute error $|\hat{\tau}_i - \tau_i|$ exceeds $T_c/2$, where T_c is the signal correlation time, and τ_i the true delay, is identified as a failure or an anomaly which follows the terminology used in [2]. Otherwise, an estimate would be deemed as a success or a nonanomalous one. In this paper, T_c is defined as the width of the main lobe of the source signal autocorrelation function (taken between the -3 -dB points). For the particular speech signal used here, which is sampled at 16 kHz, T_c is equal to 4.0 samples (0.25 ms). After time delay estimates are classified into the two classes, the TDE performance is evaluated in terms of the percentage of anomalies over the total estimates, and the mean square error (MSE) of the nonanomalous estimates.

The experimental results are graphically portrayed in Fig. 1. We found that the percentages of anomalies in both conditions are rather small, therefore not plotted. As seen from Fig. 1 (a), the estimator yields reasonably good performance in the light reverberation condition. The MSE is approximately -12 dB when only two sensors are used (in this case, the estimator is equivalent to the classical cross-correlation method, one member of the GCC family). It is reduced to -25 dB when one more microphone is added, and diminishes when more than four sensors are available. This demonstrates the effectiveness of the algorithm in taking advantage of the redundant information provided by multiple microphones to mitigate the effect of noise and reverberation.

Comparing Fig. 1 (a) with Fig. 1 (b), one can see that the MSE of the estimator deteriorates significantly when reverberation time becomes longer. This is understandable. As reverberation becomes stronger, more reflections (some have a stronger energy level, and some have a longer delay) will reach the microphone sensors. As a result, the peak of the cost function shifts away from the true delay, which will eventually lead to performance degradation. It is remarkable that, even in the heavily reverberant environment, the TDE accuracy increases with the number of microphones, corroborating the powerfulness of the multichannel TDE approach in exploiting redundancy to combat distortion.

6. CONCLUSIONS

Time delay estimation in reverberant environments remains a difficult challenge and further research efforts are indispensable. This paper has dealt with TDE, with emphasis on combating reverberation. Starting with the theory of linear interpolation, it has introduced the concepts of multichannel correlation matrix and multichannel cross-correlation coefficient. Some interesting properties and bounds of the MCCC were discussed. An efficient recursive algorithm was proposed to estimate and update the MCCC when new data snapshots are available. This new definition of the MCCC was then applied to the problem of time delay estimation, resulting in a multichannel TDE algorithm. It was shown that this new approach is equivalent to the classical cross-correlation method, one member of the GCC family, in the two-sensor case. It

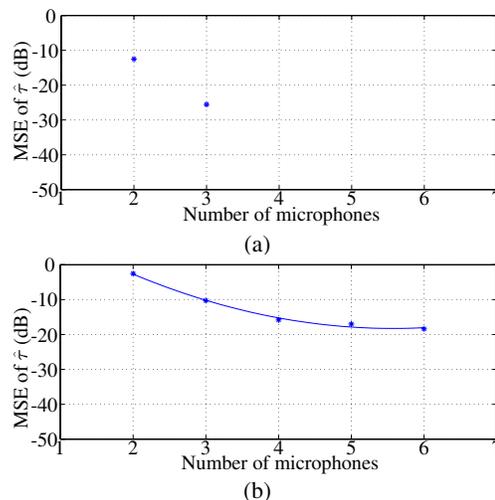


Fig. 1. MSE of the nonanomalous delay estimates in reverberant and noisy environments: (a) $T_{60} = 240$ ms, and $\text{SNR} = -5$ dB; (b) $T_{60} = 580$ ms, and $\text{SNR} = -5$ dB. (Note: in (a), when four or more microphones are used, all the time delays are correctly identified, and MSE of the nonanomalous estimates becomes zero. Thus $10\log_{10}(\text{MSE})$ becomes minus infinity, which is not displayed in the figure.)

can be treated as a natural generalization of the cross-correlation method to the multichannel case when more than two sensors are available. An appealing property of this new algorithm is that it can fully utilize the redundant information provided by multiple sensors to enhance the TDE performance against distortion. Experiments confirmed that the delay estimation accuracy increases with the number of sensors.

7. 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] B. Champagne, S. Bédard, and A. Stéphenne, "Performance of time-delay estimation in presence of room reverberation," *IEEE Trans. Speech Audio Processing*, vol. 4, pp. 148–152, Mar. 1996.
- [3] J. Benesty, "Adaptive eigenvalue decomposition algorithm for passive acoustic source localization," *J. Acoust. Soc. Am.*, vol. 107, pp. 384–391, Jan. 2000.
- [4] Y. Huang and J. Benesty, "A class of frequency-domain adaptive approaches to blind multichannel identification," *IEEE Trans. Signal Processing*, vol. 51, pp. 11–24, Jan. 2003.
- [5] S. M. Griebel and M. S. Brandstein, "Microphone array source localization using realizable delay vectors," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2001.
- [6] B. G. Ferguson, "Improved time-delay estimates of underwater acoustic signals using beamforming and prefiltering techniques," *IEEE J. Oceanic Eng.*, vol. OE-14, pp. 238–244, July 1989.
- [7] J. Chen, J. Benesty, and Y. Huang, "Robust time delay estimation exploiting redundancy among multiple microphones," *IEEE Trans. Speech Audio Processing*, vol. 11, pp. 549–557, Nov. 2003.
- [8] S. Kay, "Some results in linear interpolation theory," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 31, pp. 746–749, June 1983.
- [9] J. S. Bendat and A. G. Piersol, *Random Data Analysis and Measurement Procedures*, John Wiley & Sons, New York, NY, 1986.
- [10] D. Cochran, H. Gish, and D. Sinno, "A geometric approach to multichannel signal detection," *IEEE Trans. Signal Processing*, vol. 43, pp. 2049–2057, Sept. 1995.