Signal Processing 90 (2010) 378–384

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Fast communication

An investigation of time delay estimation in room acoustic environment using magnitude ratio

Jwu-Sheng Hu, Chia-Hsin Yang *, Wei-Han Liu

Department of Electrical and Control Engineering, National Chiao-Tung University, Hsinchu, Taiwan

ARTICLE INFO

Article history:
Received 19 January 2009
Received in revised form
20 May 2009
Accepted 8 June 2009
Available online 16 June 2009

Keywords:
Time delay estimation
Magnitude ratio

ABSTRACT

This paper investigates the relation between the nonstationary sound source and the frequency domain magnitude ratio of two microphones based on short-term frequency analysis. The fluctuation level of nonstationary sound sources is modeled by the exponent of polynomials from the concept of moving pole model. According to this model, the sufficient condition for utilizing the fluctuation level and magnitude ratio to estimate the time delay between two microphones is suggested. Simulation results are presented to show the performance of the suggested method.

1. Introduction

Time delay estimation (TDE) between two spatially separated sensors is a useful piece of information for many applications such as source localization, beamforming or sonar systems. This technique has been widely used in room acoustic environments for sound source localization and speech enhancement for the past several years. Natural sound sources are usually nonstationary and the real environment contains reverberations. However, very little research on TDE in the past emphasized on the nonstationary nature of sound sources in a reverberant environment.

Among various TDE techniques, the generalized cross correlation (GCC) proposed by Knapp and Carter is the most popular method [1]. In GCC, the time delay is estimated by finding the time-lag which maximizes the cross correlation between two filtered versions of received signals. The GCC method performs fairly well in non-reverberant environments, however, it has limited performance in reverberant condition [2]. Some algorithms [3,4] have been proposed to improve the GCC method performance in the presence of reverberation or when the interference is directional. Chen et al. proposed the multichannel TDE algorithm based on multichannel cross correlation coefficient (MCCC) [5,6]. This method uses more than two microphones and takes the advantage of redundancy. They also found that the performance in response to noise and reverberation is better as the microphone number increases.

Unlike the cross correlation based methods, Benesty [7] proposed an adaptive eigenvalue decomposition algorithm for TDE. This method focuses directly on identifying the impulse responses between the sound source and the microphones in order to estimate the time delay. In this method, the eigenvector corresponding to the minimum eigenvalue of the correlation matrix of the received signal contains the impulse response information. The time delay is determined by finding the direct paths from the two estimated impulse responses.

For estimating the time delay between two microphones, phase difference between two microphones is an intuitive cue. Magnitude ratio is relatively more unreliable than phase difference for TDE due to its ambiguity problem [8], especially if the source signal is nonstationary signal. Hence, most work on TDE focus on the

* Corresponding author. Lab 905, Engineering Building No. 5, 1001 Ta Hsueh Road, Hsinchu 300, Taiwan. Tel.: +886 3 5712121x54424; fax: +886 3 5715998.

E-mail addresses: jshu@cn.nctu.edu.tw (J.-S. Hu), chyang.ece92g@nctu.edu.tw (C.-H. Yang), lukeliu.ece89g@nctu.edu.tw (W.-H. Liu).

0165-1684/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.sigpro.2009.06.009

phase information process, and very little research estimate time delay using only magnitude ratio. The work in [8] utilizes magnitude ratio information to estimate the sound source location and multiple microphone pairs are employed to solve the magnitude ratio ambiguity problem. Although, it seems unlikely that the time delay can be estimated using only magnitude ratio with two microphones. This paper finds the important relation between magnitude ratio and the nonstationary sound source and presents a preliminary investigation into the possibility of using magnitude ratio for TDE. In this paper, the relation between magnitude ratio and the nonstationary sound source which can be used to estimate time delay is investigated. The idea of moving pole model [9] is employed to model the nonstationary sound source and the acoustic room model is used to simulate the reverberation environment. It is shown that the time delay can be obtained by estimating the slope between magnitude ratio and source fluctuation level parameter using least-square method. The performance of the proposed algorithm is evaluated by simulation and estimation error is also discussed.

2. Nonstationary sound source time delay estimation using magnitude ratio

Before describing the investigation of TDE method using magnitude ratio, the ambiguity problem of TDE using magnitude ratio in a free space environment is presented in Section 2.1.

2.1. Magnitude ratio formulation and ambiguity problem

Consider a sound source and two microphones in a generic free space environment. According to the acoustic inverse-square-law, the i-th microphone can be expressed as

$$y_i(n) = \frac{s(n)}{d_i} \tag{1}$$

where n denotes a discrete time index; $s(n)$ represents the input sound signal and d_i is the distance from the sound source to the i-th microphone. Thus, the energy received by the i-th microphone can be obtained by integrating the square of the discrete time interval $0\sim N_e$:

$$E_i = \sum_{n=0}^{N_e} y_i^2(n) = \frac{1}{d_i^2}\sum_{n=0}^{N_e} s^2(n) \tag{2}$$

Eq. (2) means the received energy decreases as the inverse of the square of the distance to the source. The above equation has the simple relationship between the first and the second microphone:

$$E_1 d_1^2 = E_2 d_2^2 \tag{3}$$

Let (x, y) and (x_i, y_i) be the coordinates of the sound source and the i-th microphone. Then $d_i^2 = (x - x_i)^2 + (y - y_i)^2$. It was shown in [8] that Eq. (3) can be written as a circular equation when $E_1 \neq E_2$:

$$\left(x - \frac{c_x}{c_e}\right)^2 + \left(y - \frac{c_y}{c_e}\right)^2 = \frac{E_1 E_2 d_{12}^2}{c_e^2} \tag{4}$$

where

$$c_e = E_1 - E_2, \quad c_x = E_1 x_1 - E_2 x_2, \quad cy = E_1 y_1 - E_2 y_2$$

and

$$d_{12}^2 = (x_1 - x_2)^2 + (y_1 - y_2)^2$$

According to Eq. (4), the sound source is constrained to lie on a circle centered $(c_x/c_e, c_y/c_e)$ with a radius of $d_{12}\sqrt{E_1 E_2}/c_e$. Besides, when $E_1 = E_2$, Eq. (3) can be written as

$$2c_x x + 2c_y y = E_1(x_1^2 + y_1^2) - E_2(x_2^2 + y_2^2) \tag{5}$$

The equation above represents the sound source is located along the line passing through the mid-point of two

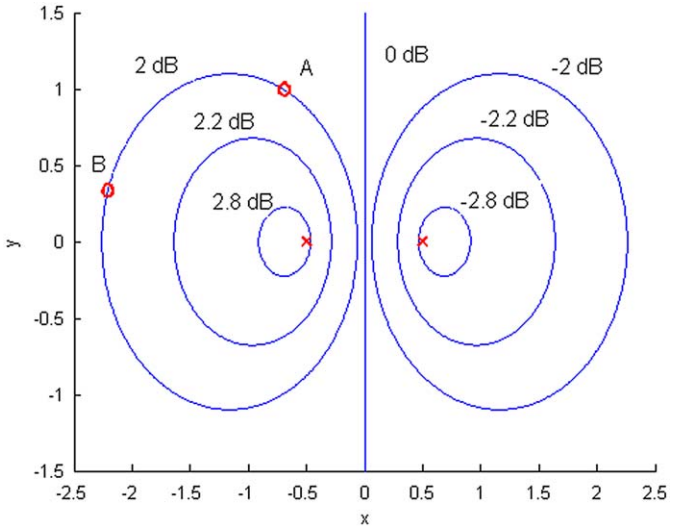

Fig. 1. The relation between $10\log\Delta E$ and sound source location.

microphones and perpendicular to the segment line between two microphones. Let us define the magnitude ratio as $\Delta E = \sqrt{E_1/E_2}$. According to Eqs. (4) and (5), the log value of magnitude ratio ($10 \log \Delta E$) and sound source location relation is shown in Fig. 1. Two microphones are located at (−0.5, 0) and (0.5, 0). As can be seen, the magnitude ratio can be used to judge whether the sound source is from left or right of the microphone pairs but there is an ambiguity problem for TDE using magnitude ratio. For example, points A and B in Fig. 1 have the same magnitude ratio but have different time delay obviously. Hence, with only two microphones magnitude ratio may not be able to estimate the time delay even for the free space environment and this is the reason why rare work estimate time delay between two microphones using only magnitude ratio. With different point of view, this paper investigates the relation between nonstationary sound source and magnitude ratio and some important findings regarding the application of magnitude ratio to calculate the time delay are presented in the next section.

2.2. Proposed time delay estimation method

Now, consider a sound source situated within the reverberant environment. The received signal at the i-th microphone can be expressed as

$$y_i(n) = h_i(n) * s(n) = \sum_{l=0}^{L-1} h_{i,l} s(n-l), \quad h_{i,l} \geq 0 \tag{6}$$

where $h_i(n) = \sum_{l=0}^{L-1} h_{i,l}\delta(n-l)$ is the room impulse response (RIR) with length L between the sound source and the i-th microphone. $h_{i,l}$ are the coefficients of the finite impulse response (FIR) model for RIR. Without loss of generality, the stationary input signal is assumed to be a complex exponential signal with frequency $\hat{\omega}_k$ and constant amplitude A:

$$s(n) = Ae^{j\hat{\omega}_k n} \tag{7}$$

where $\hat{\omega}_k = 2\pi k/N$ represents the sampled frequency of an N-point short-time Fourier transform (STFT) and k is a integer between 0 and $N/2-1$. To analyze the relation between magnitude ratio and nonstationary sound source, a parameterized model for nonstationary sound source is needed. Based on the studies of modeling nonstationary sound source in [9], a nonstationary sound source in an analysis window can be expressed as a sum of moving pole models. In this work, the idea that approximates source signal amplitude as an exponent of polynomial is utilized [9]. Hence, for the nonstationary sound signal, the constant A in Eq. (7) is replaced by time-varying amplitude A_n which can be expressed as

$$A_n = e^{\sum_{t=0}^{N_a} a_t (n/f_s)^t} \tag{8}$$

where N_a is the degree of the polynomial; a_t is the coefficient of the polynomial and f_s denotes the sampling rate. To simplify the analysis, we leave out the terms of $t \geq 2$; therefore, A_n is modeled as

$$A_n = e^{a_0 + (n/f_s)a_1} \tag{9}$$

Therefore, for the defined sound source, the sound received by the i-th microphone can be represented as

$$\begin{aligned} y_i(n) &= \sum_{l=0}^{L-1} h_{i,l} s(n-l) = \sum_{l=0}^{L-1} h_{i,l} A_{n-l} e^{j\hat{\omega}_k (n-l)} \\ &= \sum_{l=0}^{L-1} A_{n-l} h_{i,l} e^{-j\hat{\omega}_k l} e^{j\hat{\omega}_k n} \end{aligned} \tag{10}$$

Take the STFT at frequency $\hat{\omega}_k$:

$$\begin{aligned} Y_i(n, \hat{\omega}_k) &= \sum_{\tau=0}^{N-1} y_i(n+\tau) e^{-j\hat{\omega}_k (n+\tau)} \\ &= \sum_{\tau=0}^{N-1} \sum_{l=0}^{L-1} A_{n+\tau-l} h_{i,l} e^{-j\hat{\omega}_k l} e^{j\hat{\omega}_k (n+\tau)} e^{-j\hat{\omega}_k (n+\tau)} \\ &= \sum_{\tau=0}^{N-1} \sum_{l=0}^{L-1} A_{n+\tau-l} h_{i,l} e^{-j\hat{\omega}_k l} \end{aligned} \tag{11}$$

Substituting $A_n = e^{a_0 + (n/f_s)a_1}$ into Eq. (11), $Y_i(n, \hat{\omega}_k)$ can be rewritten as

$$\begin{aligned} Y_i(n, \hat{\omega}_k) = {} & [e^{a_0 + (n/f_s)a_1} + e^{a_0 + ((n+1)/f_s)a_1} + \cdots \\ & + e^{a_0 + ((n+N-1)/f_s)a_1}] h_{i,0} e^{-j\hat{\omega}_k 0} \\ & + [e^{a_0 + ((n+1)/f_s)a_1} + e^{a_0 + (n/f_s)a_1} + \cdots \\ & + e^{a_0 + ((n+N-2)/f_s)a_1}] h_{i,1} e^{-j\hat{\omega}_k 1} \\ & \vdots \\ & + [e^{a_0 + ((n-(L-1))/f_s)a_1} + e^{a_0 + ((n-(L-2))/f_s)a_1} + \cdots \\ & + e^{a_0 + ((n+(N-L))/f_s)a_1}] h_{i,L-1} e^{-j\hat{\omega}_k (L-1)} \end{aligned} \tag{12}$$

Eq. (12) can be rearranged as

$$\begin{aligned} Y_i(n, \hat{\omega}_k) = {} & e^{a_0 + (n/f_s)a_1} [1 + e^{(1/f_s)a_1} + \cdots \\ & + e^{((N-1)/f_s)a_1}] h_{i,0} e^{-j\hat{\omega}_k 0} \\ & + e^{(-a_1/f_s)} e^{a_0 + (n/f_s)a_1} [1 + e^{(1/f_s)a_1} + \cdots \\ & + e^{((N-1)/f_s)a_1}] h_{i,1} e^{-j\hat{\omega}_k 1} \\ & \vdots \\ & + e^{-(L-1)(a_1/f_s)} e^{a_0 + (n/f_s)a_1} [1 + e^{(1/f_s)a_1} + \cdots \\ & + e^{((N-1)/f_s)a_1}] h_{i,L-1} e^{-j\hat{\omega}_k (L-1)} \\ = {} & e^{a_0 + (n/f_s)a_1} \left(1 - e^{(N(a_1/f_s))}\right) \Big/ \left(1 - e^{(a_1/f_s)}\right) \\ & \times \left(\sum_{l=0}^{L-1} e^{(-a_1/f_s)l} h_{i,l} e^{-j\hat{\omega}_k l} \right) \end{aligned} \tag{13}$$

Therefore, the natural logarithm of magnitude ratio between two microphones is

$$M(n, \hat{\omega}_k) = \ln \left| \frac{Y_1(n, \hat{\omega}_k)}{Y_2(n, \hat{\omega}_k)} \right| = \ln \left| \frac{\sum_{l=0}^{L-1} e^{(-a_1/f_s)l} h_{1,l} e^{-j\hat{\omega}_k l}}{\sum_{l=0}^{L-1} e^{(-a_1/f_s)l} h_{2,l} e^{-j\hat{\omega}_k l}} \right| \tag{14}$$

By observing Eq. (14), we can find that the values of the magnitude ratio depend on the coefficient of the room impulse response models $h_{i,l}$ and the value of a_1, which is the slope of the natural logarithm of A_n. This result concludes that the magnitude ratio between two microphones is still influenced by the reverberations in the room. However, the term $e^{(-a_1/f_s)l}$ in Eq. (14) decreases with the increase of l when a_1 is positive. This means the reflection part in the channel model is less weighted and the influence of direct path is becoming significant. Notice that the numerator or denominator of Eq. (14) is the linear

combination of L vectors. The vector direction is decided by the frequency $\hat{\omega}_k$ and l and the magnitude is controlled by $e^{(-a_1/f_s)l}h_{i,l}$. Since the values $e^{(-a_1/f_s)l}$ and $h_{i,l}$ decrease with the increase of l, the direct path vector, $e^{(-a_1/f_s)l_{i,D_1}}h_{i,l_{i,D_1}}e^{-j\hat{\omega}_k l_{i,D_1}}$, is less influenced by the reflection vector $(e^{(-a_1/f_s)l_{i,Dm}}h_{i,l_{i,D_m}}e^{-j\hat{\omega}_k l_{i,D_m}}, m \geq 2)$. Hence, when a_1 is positive, $M(n,\hat{\omega})$ can be approximated by

$$\begin{aligned} M(n,\hat{\omega}_k) &\approx \ln\left|\frac{e^{(-a_1/f_s)l_{1,D_1}}h_{1,l_{1,D_1}}e^{-j\hat{\omega}_k l_{1,D_1}}}{e^{(-a_1/f_s)l_{2,D_1}}h_{2,l_{2,D_1}}e^{-j\hat{\omega}_k l_{2,D_1}}}\right| \\ &= \ln\left|\frac{e^{(-a_1/f_s)l_{1,D_1}}h_{1,l_{1,D_1}}}{e^{(-a_1/f_s)l_{2,D_1}}h_{2,l_{2,D_1}}}\right| \\ &= \frac{-(l_{1,D_1}-l_{2,D_1})}{f_s}a_1 + \ln\frac{h_{1,l_{1,D_1}}}{h_{2,l_{2,D_1}}} \end{aligned} \tag{15}$$

where l_{1,D_1} and l_{2,D_1} denote the propagation delay sample of the direct path from the sound source to the microphones. Consequently, the relation between the natural logarithm of magnitude ratio between microphones and a_1 is approximately linear with a slope of $-(l_{1,D_1}-l_{2,D_1})/f_s$ and $l_{1,D_1}-l_{2,D_1}$ is the time delay sample between microphones. To estimate the time delay between microphones is identical to estimate the slope of the linear relation between $M(n,\hat{\omega}_k)$ and a_1.

In summary, to estimate the TDE between two microphones, a set of sound sources with T values of a_1 is emitted. Because the T values of a_1 are decided by us. Therefore, we choose positive a_1 to suppress the reflection part influence in Eq. (14)

$$\begin{bmatrix} \hat{M}(n_{a_1(1)},\hat{\omega}_k) \\ \vdots \\ \hat{M}(n_{a_1(T)},\hat{\omega}_k) \end{bmatrix} = \begin{bmatrix} a_1(1) & 1 \\ \vdots & \vdots \\ a_1(T) & 1 \end{bmatrix} \begin{bmatrix} \dfrac{-(l_{1,D_1}-l_{2,D_1})}{f_s} \\ \ln\dfrac{h_{1,l_{1,D_1}}}{h_{2,l_{2,D_1}}} \end{bmatrix} \tag{16}$$

where $a_1(t)$, $t=1,\ldots,T$, denotes a set of a_1, $M(n_{a_1(t)},\hat{\omega})$, $t=1,\ldots,T$, denotes the magnitude ratio obtained with $a_1(t)$. To simplify the expression, we let

$$l_{1,D_1}-l_{2,D_1}=D,\quad \begin{bmatrix} \hat{M}(n_{a_1(1)},\hat{\omega}_k) \\ \vdots \\ \hat{M}(n_{a_1(T)},\hat{\omega}_k) \end{bmatrix} = \hat{\boldsymbol{Y}} \quad \text{and} \quad \begin{bmatrix} a_1(1) & 1 \\ \vdots & \vdots \\ a_1(T) & 1 \end{bmatrix} = \boldsymbol{X}$$

Finally, the time delay sample D can be estimated by the least-square method:

$$\hat{D} = [-f_s\ 0]\times(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\hat{\boldsymbol{Y}} \tag{17}$$

2.3. Estimation error analysis

Eq. (14) can be approximated by Eq. (15) due to the fact that $e^{(-a_1/f_s)l}$ and $h_{i,l}$ are decreasing when l is increasing. However, the delay estimation error occurs when the reflection is strong. The delay estimation error can be defined as

$$\begin{aligned} D-\hat{D} &= [-f_s\ 0]\times(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T(\boldsymbol{Y}-\hat{\boldsymbol{Y}}) \\ &= [-f_s\ 0]\times(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T \begin{bmatrix} C(a_1(1)) - \ln\left|\dfrac{\sum_{l=0}^{L-1}e^{(-a_1(1)/f_s)l}h_{1,l}e^{-j\hat{\omega}_k l}}{\sum_{l=0}^{L-1}e^{(-a_1(1)/f_s)l}h_{2,l}e^{-j\hat{\omega}_k l}}\right| \\ \vdots \\ C(a_1(T)) - \ln\left|\dfrac{\sum_{l=0}^{L-1}e^{(-a_1(T)/f_s)l}h_{1,l}e^{-j\hat{\omega}_k l}}{\sum_{l=0}^{L-1}e^{(-a_1(T)/f_s)l}h_{2,l}e^{-j\hat{\omega}_k l}}\right| \end{bmatrix} \\ &= \frac{1}{\Delta}\left[-Tf_s a_1(1)+f_s\sum_{i=1}^{T}a_1(i),\quad \ldots\quad ,-Tf_s a_1(T)+f_s\sum_{i=1}^{T}a_1(i)\right] \\ &\quad \times \begin{bmatrix} C(a_1(1))-P(a_1(1))+Q(a_1(1)) \\ \vdots \\ C(a_1(T))-P(a_1(T))+Q(a_1(T)) \end{bmatrix} \end{aligned} \tag{18}$$

where

$$\Delta = T\cdot\sum_{i=1}^{T}a_1(i)^2 - \left(\sum_{i=1}^{T}a_1(i)\right)^2$$

$$C(a_1(i)) = \frac{-(l_{1,D_1}-l_{2,D_1})}{f_s}a_1(i) + \ln\frac{h_{1,l_{1,D_1}}}{h_{2,l_{2,D_1}}}$$

$$\begin{aligned} P(a_1(i)) &= \frac{1}{2}\ln\left[\left(\sum_{l=0}^{L-1}e^{(-a_1(i)/f_s)l}h_{1,l}\cos(\hat{\omega}_k l)\right)^2 \right. \\ &\quad \left. + \left(\sum_{l=0}^{L-1}e^{(-a_1(i)/f_s)l}h_{1,l}\sin(\hat{\omega}_k l)\right)^2\right] \\ Q(a_1(i)) &= \frac{1}{2}\ln\left[\left(\sum_{l=0}^{L-1}e^{(-a_1(i)/f_s)l}h_{2,l}\cos(\hat{\omega}_k l)\right)^2 \right. \\ &\quad \left. + \left(\sum_{l=0}^{L-1}e^{(-a_1(i)/f_s)l}h_{2,l}\sin(\hat{\omega}_k l)\right)^2\right] \end{aligned}$$

Eq. (18) can be considered as a function of $\hat{\omega}_k$ and only the term $\hat{\boldsymbol{Y}}$ is $\hat{\omega}_k$ dependent. Hence, for different frequency, Eq. (18) is the combination of constant values, cosine signals and sine signals under the fixed room impulse response. It means that the delay estimation error is varying with different $\hat{\omega}_k$. Different frequency would cause the different estimation error when the impulse response is unchanged. Moreover, it is easy to see that the estimation error should oscillate with frequency. Strong reflected environment would cause the larger oscillation amplitude.

3. Simulation results

This section provides the simulation results to access the capability of the time delay estimation using magnitude ratio proposed in this paper. In these simulations, the image method [10] is adopted to model the room impulse response and the reflection coefficient is varying between 0 and 1. The sampling rate is 16 kHz. To test the proposed approach carefully, the source signal is the synthetic signal with the known parameters (a_1 and $\hat{\omega}_k$). The values of a_1 are selected to be ten values ($a_1 = 26, 27, \ldots, 35$, $T = 10$). The room impulse response is convolved with

source signals to generate microphone signals and the STFT size is 1024. The enclosure room size is 10 m × 6 m × 3.6 m with different reflection coefficients and two microphones with 10 cm spacing are located at (5 2 1.2) and (5.1 2 1.2). Three experiments are carried out in this section and one performance index, root mean square error (RMSE), is defined below to evaluate the performance of the suggested method:

$$RMSE = \sqrt{\frac{1}{N_T}\sum_{i=1}^{N_T}(\hat{D}_i - D_i)^2} \tag{19}$$

where N_T is the total number of estimation; $\hat{D}_i$ is the i-th time delay estimation and D_i is the i-th correct delay sample with a integer. The smaller the RMSE is, the better the estimator is.

3.1. Reverberant environment

The first experiment is performed in a reverberant environment and the sound source is placed at a distance of 60 cm from the mid-point of two microphones for several directions. For each testing, the source frequency is chosen from the range 100–1000 Hz. The noise is absence in this experiment and the room reverberation time T_{60} is computed by Sabine's formulas. Fig. 2 illustrates the RMSE as the function of the reverberation time. The total estimation number N_T is 300. As can be seen from Fig. 2, in the non-reverberant environment ($T_{60} = 0$ s), the proposed method can accurately identify the time delay. This is because when the environment is non-reverberant, the impulse response coefficients only contain one value $h_{i,i_{D_1}}$. Hence, Eq. (14) can be equal to Eq. (15) exactly. The estimation error occurs as T_{60} increases. This can be explained by the fact that the strong reflection vector would influence the magnitude of the direct path vector and cause the approximation error of Eq. (15). Fig. 2 also shows that the proposed method has a small RMSE for slight reverberant environment.

3.2. Noisy and non-reverberant environment

In this section, we will evaluate the performance of the proposed algorithm in the non-reverberant but noisy environment. The white noise is properly scaled and added to each microphone signal to control the signal-to-noise ratio (SNR). The total estimation number N_T is 300 and the source frequency is 100–1000 Hz. Fig. 3 presents the RMSE with respect to varying SNR. The result states that the RMSE decreases when SNR is increased. The RMSE is <1 even at the lower SNR. It can further be noticed that by comparing Fig. 2 with Fig. 3, the proposed method is significantly affected by the reverberation time and is relatively insensitive to the noise. In addition, the noise is also created by the speech source. As can be seen from Fig. 3, the nonstationary noise would affect the performance more serious than the stationary noise.

3.3. Estimation error versus frequency analysis

The RMSE results of Sections 3.1 and 3.2 are the statistical results for different source frequencies. In fact, different source frequency will lead to the different estimation error under the fixed impulse response condition. This section will analyze the relation between estimation error and source frequency. The estimation error is defined in Eq. (18) and the simulation result is depicted in Fig. 4. The source is located at (4.7, 2.52, 1.2). As can be seen, the estimation error remains at zero for different frequencies when $T_{60} = 0$ s. This is expected since Eq. (18) becomes frequency independent when the environment has no reverberation. However, the estimation error starts to oscillate with frequency when $T_{60} > 0$ s and this is because the magnitude ratio components are the combination of some exponential signals. The oscillation amplitude becomes large as the reverberation time is increased. Fig. 4 also demonstrates that if the impulse response is fixed, there exist some frequencies which can make no estimation error.

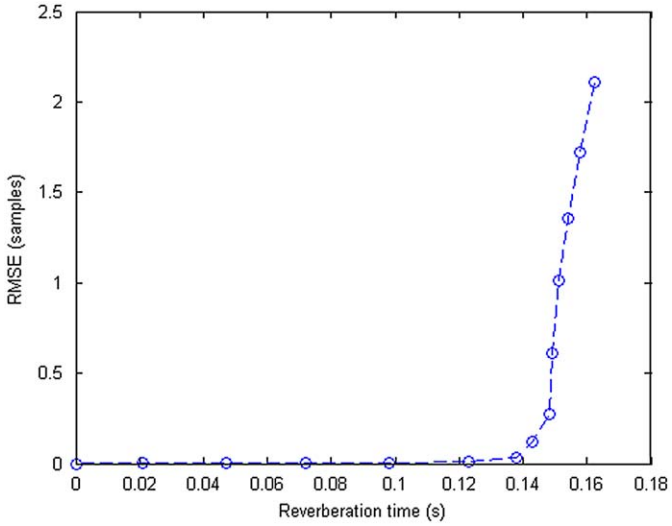

Fig. 2. RMSE versus reverberation time.

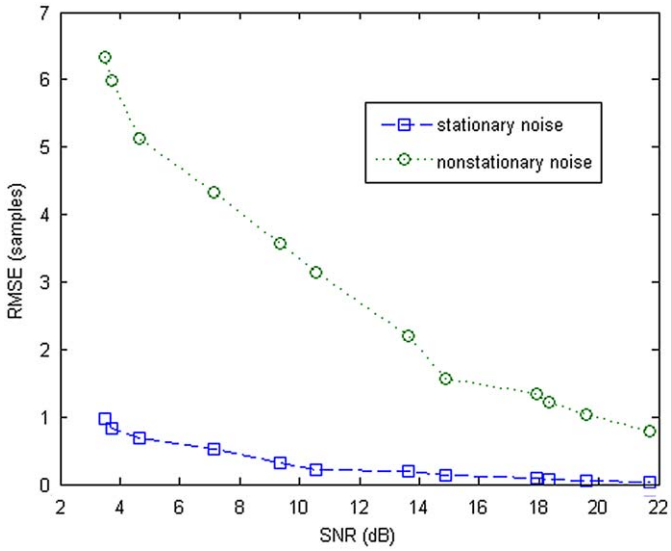

Fig. 3. RMSE versus SNR.

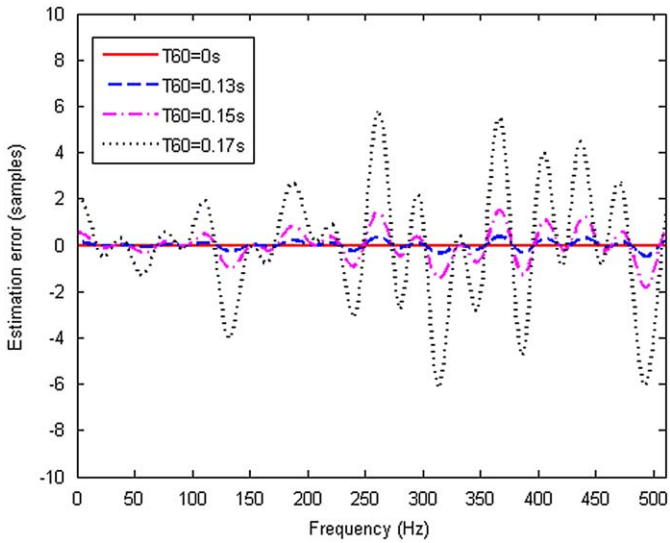

Fig. 4. Estimation error versus frequency.

In summary, by observing the simulation results, the proposed method can estimate the time delay exactly using only two microphones and magnitude ratio information in the non-reverberant environment but the performance degrades as the reverberation is present. In this paper, we present a preliminary investigation into the possibility of using magnitude ratio for TDE and the moving pole model with the known parameters (a_1 and $\hat{\omega}_k$) is needed to be the sound source. In order to apply the proposed method to handle the real nonstationary sound source (such as speech) or to be more robust to the reverberant environment, the more complex models may be incorporated. This is left as a further research topic.

4. Conclusion

This paper investigates the relation between nonstationary sound source and magnitude ratio when STFT is utilized. From the investigation, a method which can be used to estimate the time delay is suggested. In this method, the time delay can be obtained by estimating the

slope between magnitude ratio and a parameter of the moving pole model of the nonstationary sound source. The performance of the proposed method in different reverberation environments and SNR is presented with simulation and the relation between the performance and source signal frequency is also discussed.

References

[1] C.H. Knapp, G.C. Carter, The generalized correlation method for estimation of time delay, IEEE Trans. Acoust. Speech Signal Process. 24 (1976) 320–327.
[2] B. Champagne, S. Bedard, A. Stephenne, Performance of time-delay estimation in the presence of room reverberation, IEEE Trans. Speech Audio Process. 4 (2) (1996) 148–152.
[3] M. Brandstein, H. Silverman, A robust method for speech signal time-delay estimation in reverberant rooms, in: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, 1997, pp. 375–378.
[4] C. Nikias, R. Pan, Time delay estimation in unknown Gaussian spatially correlated noise, IEEE Trans. Acoust. Speech Signal Process. 36 (1988) 1706–1714.
[5] J. Chen, J. Benesty, Y. Huang, Robust time delay estimation exploiting redundancy among multiple microphones, IEEE Trans. Speech Audio Process. 11 (2003) 549–557.
[6] J. Benesty, Y. Huang, J. Chen, Time delay estimation via minimum entropy, IEEE Signal Process. Lett. 14 (2007) 157–160.
[7] J. Benesty, Adaptive eigenvalue decomposition algorithm for passive acoustic source localization, J. Acoust. Soc. Am. 107 (1) (2000) 384–391.
[8] S.T. Birchfield, R. Gangishetty, Acoustic localization by interaural level difference, IEEE Int. Conf. Acoust. Speech Signal Process. 4 (2005) 1109–1112.
[9] F. Casacuberta, E. Vidal, A nonstationary model for the analysis of transient speech signals, IEEE Trans. Acoust. Speech Signal Process. 35 (2) (1987) 226–228.
[10] J.B. Allen, D.A. Berkley, Image method for efficiently simulating small-room acoustics, J. Acoust. Soc. Am. 65 (1978) 943–950.