MULTICHANNEL TIME DELAY ESTIMATION FOR ACOUSTIC SOURCE LOCALIZATION VIA ROBUST ADAPTIVE BLIND SYSTEM IDENTIFICATION

Hongsen He[1]*, Jingdong Chen*[2]*, Jacob Benesty*[3]*, and Tao Yang*[1]

[1]School of Information Engineering,
Robot Technology Used for Special Environment
Key Laboratory of Sichuan Province
Southwest University of Science and Technology
Mianyang 621010, China
{hongsenhe@gmail.com, yangtao98@tsinghua.org.cn}

[2]Center of Immersive and
Intelligent Acoustics
Northwestern Polytechnical University
127 Youyi West Road
Xi'an 710072, China
{jingdongchen@ieee.org}

[3]INRS-EMT
University of Quebec
800 de la Gauchetiere Ouest
Suite 6900, Montreal
QC H5A 1K6, Canada
{benesty@emt.inrs.ca}

ABSTRACT

In the problem of acoustic source localization, time difference of arrival (TDOA) among multiple sensors is needed, which is often obtained through time delay estimation (TDE) techniques. Among the multiple TDE methods developed in the literature, the normalized multichannel frequency-domain least-mean-square (NMCFLMS) algorithm is shown robust to reverberation. The performance of this algorithm, however, deteriorates in non-Gaussian and low signal-to-noise ratio (SNR) Gaussian noise environments. In this paper, we re-derive a robust normalized multichannel frequency-domain least-mean-M-estimate (RNMCFLMM) algorithm to estimate TDOAs for acoustic source localization. The proposed algorithm exploits the non-sensitivity of an M-estimator to non-Gaussian noise and makes a tradeoff between the least-squares and least-absolute criteria to improve the robustness of TDE with respect to non-Gaussian and Gaussian noises. The effectiveness of the proposed algorithm is demonstrated in real acoustic environments.

***Index Terms*—** Acoustic source localization, microphone arrays, robust normalized multichannel frequency-domain least-mean-M-estimate (RNMCFLMM), multichannel time delay estimation.

1. INTRODUCTION

Acoustic source localization, which aims at estimating the position of radiating sound sources, plays an important role in many applications such as hands-free voice communication. Typically, source localization is achieved in two steps [1]: 1) TDOAs among different sensors are estimated and 2) triangulation principle is then used to determine the source position. In this two-step approach, the most critical part is the TDOA estimation, which is also called time delay estimation (TDE).

Commonly used TDE algorithms include the cross-correlation (CC) method, the generalized cross-correlation (GCC) algorithm [1], [2], the multichannel cross-correlation coefficient (MCCC) algorithm [3], the multichannel spatio-temporal prediction (MCSTP) technique [4], the multichannel sparse linear prediction method [5], the information theory based algorithms [6], [7], the methods exploiting some characteristics of speech signals [8], [9], etc. These algorithms, however, are found to be easily affected by background noise and reverberation in room acoustic environments. An effective way to improve the robustness of TDE against reverberation is through channel identification as shown in the the adaptive eigenvalue decomposition (AED) method [10], [11] derived from [12], [13]. This algorithm first blindly identifies the impulse responses of a two-channel system, and then determines TDOA using the relative position between the two direct-path components. This algorithm was also generalized to the multichannel case, resulting in an NMCFLMS algorithm that can produce better TDE performance than the AED method [14], [15]. Both the AED and NMCFLMS algorithms are found robust to reverberation; however, their performance deteriorates in non-Gaussian and low SNR Gaussian noises.

The M estimator is widely used in robust signal processing [16], [17], e.g., it has been applied to adaptive filtering for echo cancelation [18]. In our recent work, we exploited the Huber estimator to construct a robust frequency-domain adaptive filter to blindly identify the impulse responses of an acoustic single-input multiple-output (SIMO) system in room acoustic environments [19], which greatly improves the robustness of blind multichannel identification to non-Gaussian and Gaussian noises. The major differences between the algorithm in [18] and that in [19] are: 1) the former is a non-blind method dealing with single channel identification while the latter is a blind method that copes with blind multichannel identification, and 2) the former considers only the case of impulsive noise while the latter uses an adaptive Huber estimator and introduces a spectral constraint so that the filter is robust in non-Gaussian and Gaussian noises.

In this paper, we re-derive the RNMCFLMM algorithm but from a different perspective and apply it to the problem of TDOA estimation for acoustic source localization. This algorithm employs the non-sensitivity of an M-estimator to non-Gaussian noise and makes a tradeoff between the least-squares and least absolute criteria to improve the robustness of TDE with respect to non-Gaussian and Gaussian noises. In comparison with the algorithm in [19], the major contributions of this papers are: 1) we construct a new recursive cost function to re-derive the RNMCFLMM algorithm and the new derivation process is more rigorous than that in [19], and 2) this RNMCFLMM algorithm is applied to the problem of multichannel TDE, which is robust in reverberant environments with non-Gaussian and Gaussian noises.

2. MULTICHANNEL TIME DELAY ESTIMATION VIA ROBUST BLIND SYSTEM IDENTIFICATION

2.1. Robust Adaptive Blind Multichannel Identification

The input-output relationship of a SIMO system is given by

$$x_k(n) = s(n) * h_k(n) + v_k(n), \quad k = 1, 2, \dots, M, \tag{1}$$

This work was supported in part by the National Science Foundation of China (NSFC) (Grant No. 61571376), the NSFC "Distinguished Young Scientists Fund" (Grant No. 61425005), and the Incubation Program for the Distinguished Youth Foundation of Sichuan Province of China (Grant No. 2014JQ0042).

978-1-5090-2007-2/16/$31.00 ©2016 IEEE

where $s(n)$ is the source signal, $*$ denotes linear convolution, $h_k(n)$ is a finite-impulse-response (FIR) filter, which denotes the channel impulse response between the sound source and the kth microphone, $v_k(n)$ is the additive noise at the kth microphone, and M is the number of microphones. If we neglect the noise term in (1), the following relation can be obtained

$$\begin{aligned} x_i(n) * h_j(n) &= s(n) * h_i(n) * h_j(n) \\ &= x_j(n) * h_i(n), \;\; i, j = 1, 2, \ldots, M, \;\; i \neq j. \end{aligned} \tag{2}$$

This relation can be written in a matrix-vector form as

$$\mathbf{x}_i^T(n)\mathbf{h}_j - \mathbf{x}_j^T(n)\mathbf{h}_i = 0, \tag{3}$$

where

$$\mathbf{x}_i(n) = [x_i(n) \;\; x_i(n-1) \;\; \cdots \;\; x_i(n-L+1)]^T \tag{4}$$

is the observation signal vector of channel i with length of L,

$$\mathbf{h}_i = [h_{i,0} \;\; h_{i,1} \;\; \cdots \;\; h_{i,L-1}]^T \tag{5}$$

is the impulse response vector of channel i with length of L, and $(\cdot)^T$ stands for the transpose of a vector or a matrix.

When noise is present and/or the estimate of the impulse responses deviates from their true values, the right-hand side of (3) is no longer zero and an *a priori* error signal between the ith and jth channels is produced as

$$e_{ij}(n) = \mathbf{x}_i^T(n)\widehat{\mathbf{h}}_j(n) - \mathbf{x}_j^T(n)\widehat{\mathbf{h}}_i(n), \tag{6}$$

where $\widehat{\mathbf{h}}_i(n)$ is an estimate of $\mathbf{h}_i$ at time n. This error signal can then be used to define a cost function that can be minimized to find an optimal estimate of the impulse responses. In an earlier work, we proposed an RNMCFLMM algorithm to blindly estimate the multichannel impulse responses. Some derivations in RNMCFLMM, however, are empirical, which were not rigorously deduced.

This paper uses an M-estimator [16], which is a generalization of maximum likelihood-type estimators, to define a recursive cost function, based on which we re-derive the RNMCFLMM algorithm to blindly identify the SIMO system so that the update equations are directly deduced. The cost function is then defined as

$$\mathcal{J}(m) = (1-\lambda)\sum_{r=0}^{m} \lambda^{m-r}\mathcal{J}_\rho(r), \tag{7}$$

where m is the block-time index, $0 < \lambda < 1$ is a forgetting factor,

$$\mathcal{J}_\rho(r) = \sum_{i=1}^{M-1}\sum_{j=i+1}^{M}\sum_{n=rL}^{rL+L-1} \rho\left[e_{ij}(n)\right], \tag{8}$$

$\rho(\cdot)$ is an M-estimator. In this paper, we use the Huber estimator [16], [19] as the M-estimator, and then the corresponding estimate function between the ith and jth channels is written as

$$\rho\left[e_{ij}(n)\right] = \begin{cases} e_{ij}^2(n)/2, & |e_{ij}(n)| < \xi_{ij} \\ \xi_{ij}\left[|e_{ij}(n)| - \xi_{ij}/2\right], & |e_{ij}(n)| \geq \xi_{ij} \end{cases}, \tag{9}$$

where ξ_{ij} is an adaptive threshold parameter between channels i and j (see e.g., [19]). If $|e_{ij}(n)| \leq \xi_{ij}$, the cost function employs the least-squared error (LSE) criterion. However, if $|e_{ij}(n)| \geq \xi_{ij}$, the cost function employs the least-absolute error (LAE) criterion to deemphasize the effect of outliers, thereby ensuring that the corresponding adaptive filter is robust not only to Gaussian noise but to non-Gaussian noise due to its compromise between the LSE and LAE criteria.

In this work, we use the iterative Newton method [18] to derive the adaptive algorithm that minimizes the cost function $\mathcal{J}(m)$. Then, we have the following recursive solution to the optimization problem:

$$\begin{aligned} \underline{\widehat{\boldsymbol{h}}}_k(m+1) = \underline{\widehat{\boldsymbol{h}}}_k(m) - 2\mu(1-\lambda)\boldsymbol{\mathcal{S}}_k^{-1}(m)\nabla\mathcal{J}_\rho(m), \\ k = 1, 2, \ldots, M, \end{aligned} \tag{10}$$

$$\boldsymbol{\mathcal{S}}_k(m) = \lambda\boldsymbol{\mathcal{S}}_k(m-1) + (1-\lambda)\nabla\nabla^H\mathcal{J}_\rho(m), \tag{11}$$

where $\underline{\widehat{\boldsymbol{h}}}_k(m) = \mathbf{F}_{L\times L}\widehat{\mathbf{h}}_k(m)$ with the Fourier matrix $\mathbf{F}_{L\times L}$ of size $L \times L$, μ is a step size, $\nabla\mathcal{J}_\rho(m)$ is the gradient of $\mathcal{J}_\rho(m)$ with respect to $\underline{\widehat{\boldsymbol{h}}}_k^*(m)$, the superscript $(\cdot)^*$ denotes the conjugate operator, $\nabla\nabla^H\mathcal{J}_\rho(m)$ is the corresponding Hessian matrix, and $(\cdot)^H$ denotes the conjugate transpose of a vector or matrix. Note that the goal of the calculation in the Fourier domain is to reduce the computational complexity by using the fast Fourier transform (FFT).

To calculate $\nabla\mathcal{J}_\rho(m)$ and $\nabla\nabla^H\mathcal{J}_\rho(m)$, let us take a block (of L samples) of the signal $e_{ij}(n)$ given in (6) as follows:

$$\begin{aligned} \mathbf{e}_{ij}(m) &= \mathbf{W}_{L\times 2L}^{01}\mathbf{F}_{2L\times 2L}^{-1} \\ &\times \left[\boldsymbol{\mathcal{D}}_{x_i}(m)\boldsymbol{\mathcal{G}}_{2L\times L}^{10}\underline{\widehat{\boldsymbol{h}}}_j(m) - \boldsymbol{\mathcal{D}}_{x_j}(m)\boldsymbol{\mathcal{G}}_{2L\times L}^{10}\underline{\widehat{\boldsymbol{h}}}_i(m)\right], \end{aligned} \tag{12}$$

where

$$\mathbf{e}_{ij}(m) = [e_{ij}(mL) \;\; e_{ij}(mL+1) \;\; \cdots \;\; e_{ij}(mL+L-1)]^T, \tag{13}$$

$$\mathbf{W}_{L\times 2L}^{01} = \left[\begin{array}{cc} \mathbf{0}_{L\times L} & \mathbf{I}_{L\times L} \end{array}\right], \tag{14}$$

$$\boldsymbol{\mathcal{D}}_{x_i}(m) = \mathrm{diag}\{\mathbf{F}_{2L\times 2L}\mathbf{x}_{i,2L}(m)\}, \tag{15}$$

$$\begin{aligned} \mathbf{x}_{i,2L}(m) = [&x_i(mL-L) \;\; x_i(mL-L+1) \\ &\cdots \;\; x_i(mL+L-1)]^T, \end{aligned} \tag{16}$$

$$\boldsymbol{\mathcal{G}}_{2L\times L}^{10} = \mathbf{F}_{2L\times 2L}\mathbf{W}_{2L\times L}^{10}\mathbf{F}_{L\times L}^{-1}, \tag{17}$$

$$\mathbf{W}_{2L\times L}^{10} = \left[\begin{array}{cc} \mathbf{I}_{L\times L} & \mathbf{0}_{L\times L} \end{array}\right]^T, \tag{18}$$

$\mathbf{0}_{L\times L}$ is the null matrix of size $L \times L$, $\mathbf{I}_{L\times L}$ is the identity matrix of size $L \times L$, $\mathrm{diag}(\cdot)$ denotes a diagonal matrix with indicated vector along the diagonal.

Having carried out some simple derivations, we can achieve the gradient of $\mathcal{J}_\rho(m)$ with respect to $\underline{\widehat{\boldsymbol{h}}}_k^*(m)$ as follows:

$$\begin{aligned} \nabla\mathcal{J}_\rho(m) &= 2\frac{\partial\mathcal{J}_\rho(m)}{\partial\underline{\widehat{\boldsymbol{h}}}_k^*(m)} \\ &= 2\sum_{i=1}^{M-1}\sum_{j=i+1}^{M}\sum_{n=mL}^{mL+L-1}\frac{\partial\rho\left[e_{ij}(n)\right]}{\partial\underline{\widehat{\boldsymbol{h}}}_k^*(m)} \\ &= \frac{2}{L}\sum_{i=1}^{M}\boldsymbol{\mathcal{G}}_{L\times 2L}^{10}\boldsymbol{\mathcal{D}}_{x_i}^*(m)\mathbf{F}_{2L\times 2L}\mathbf{W}_{2L\times L}^{01}\boldsymbol{\varphi}\left[\mathbf{e}_{ik}(m)\right], \end{aligned} \tag{19}$$

where

$$\boldsymbol{\mathcal{G}}_{L\times 2L}^{10} = \frac{1}{2}(\boldsymbol{\mathcal{G}}_{2L\times L}^{10})^H, \tag{20}$$

$$\mathbf{W}_{2L\times L}^{01} = (\mathbf{W}_{L\times 2L}^{01})^T, \tag{21}$$

$$\boldsymbol{\varphi}\left[\mathbf{e}_{ik}(m)\right] = \left[\begin{array}{c} \rho'[e_{ik}(mL)] \\ \rho'[e_{ik}(mL+1)] \\ \vdots \\ \rho'[e_{ik}(mL+L-1)] \end{array}\right], \tag{22}$$

and $\rho'(\cdot)$ is the first-order derivative of $\rho(\cdot)$. The Hessian matrix is then

$$\begin{aligned}\nabla\nabla^H \mathcal{J}_\rho(m) &= 2\frac{\partial}{\partial \underline{\widehat{\boldsymbol{h}}}_k^*(m)}\left[\nabla \mathcal{J}_\rho(m)\right]^H \\ &= \frac{2}{L}\frac{\partial}{\partial \underline{\widehat{\boldsymbol{h}}}_k^*(m)}\sum_{i=1}^{M}\boldsymbol{\varphi}^H\left[\mathbf{e}_{ik}(m)\right]\mathbf{W}_{L\times 2L}^{01}\mathbf{F}_{2L\times 2L}^H\boldsymbol{\mathcal{D}}_{x_i}(m)\boldsymbol{\mathcal{G}}_{2L\times L}^{10} \\ &= \frac{2}{L}\boldsymbol{\mathcal{G}}_{L\times 2L}^{10}\boldsymbol{\mathcal{P}}_k(m)\boldsymbol{\mathcal{G}}_{2L\times L}^{10}, \qquad (23)\end{aligned}$$

where

$$\begin{aligned}\boldsymbol{\mathcal{P}}_k(m) &= 2\sum_{i=1,i\neq k}^{M}\boldsymbol{\mathcal{D}}_{x_i}^*(m)\mathbf{F}_{2L\times 2L}^{-H}\mathbf{W}_{2L\times L}^{01}\mathbf{T}_{ik}(m) \\ &\quad\times \mathbf{W}_{L\times 2L}^{01}\mathbf{F}_{2L\times 2L}^H\boldsymbol{\mathcal{D}}_{x_i}(m), \qquad (24)\end{aligned}$$

$$\mathbf{T}_{ik}(m) = \mathrm{diag}\left\{\rho''\left[e_{ik}(mL)\right],\cdots,\rho''\left[e_{ik}(mL+L-1)\right]\right\}, \qquad (25)$$

and $\rho''(\cdot)$ is the second-order derivative of $\rho(\cdot)$.

From (11), we can obtain

$$\begin{aligned}\boldsymbol{\mathcal{S}}_k(m) &= \lambda\boldsymbol{\mathcal{S}}_k(m-1) + (1-\lambda)\nabla\nabla^H\mathcal{J}_\rho(m) \\ &= \lambda(1-\lambda)\sum_{i=0}^{m-1}\lambda^{m-1-i}\nabla\nabla^H\mathcal{J}_\rho(i) \\ &\quad + (1-\lambda)\nabla\nabla^H\mathcal{J}_\rho(m) \\ &= (1-\lambda)\sum_{i=0}^{m}\lambda^{m-i}\nabla\nabla^H\mathcal{J}_\rho(i). \qquad (26)\end{aligned}$$

Substituting (23) into (26) produces

$$\begin{aligned}\boldsymbol{\mathcal{S}}_k(m) &= (1-\lambda)\sum_{i=0}^{m}\lambda^{m-i}\frac{2}{L}\boldsymbol{\mathcal{G}}_{L\times 2L}^{10}\boldsymbol{\mathcal{P}}_k(i)\boldsymbol{\mathcal{G}}_{2L\times L}^{10} \\ &= \frac{2}{L}\boldsymbol{\mathcal{G}}_{L\times 2L}^{10}\boldsymbol{\mathcal{Q}}_k(m)\boldsymbol{\mathcal{G}}_{2L\times L}^{10}, \qquad (27)\end{aligned}$$

where

$$\begin{aligned}\boldsymbol{\mathcal{Q}}_k(m) &= (1-\lambda)\sum_{i=0}^{m}\lambda^{m-i}\boldsymbol{\mathcal{P}}_k(i) \\ &= \lambda\boldsymbol{\mathcal{Q}}_k(m-1) + (1-\lambda)\boldsymbol{\mathcal{P}}_k(m). \qquad (28)\end{aligned}$$

When L is large, substituting (19) and (27) into (10), pre-multiplying both sides by $\boldsymbol{\mathcal{G}}_{2L\times L}^{10}$, and with some simple manipulations, we can deduce the RNMCFLMM algorithm as follows:

$$\begin{aligned}\underline{\widehat{\boldsymbol{h}}}_k^{10}(m+1) &= \underline{\widehat{\boldsymbol{h}}}_k^{10}(m) - \mu_{\mathrm{f}}\boldsymbol{\mathcal{Q}}_k^{-1}(m)\sum_{i=1}^{M}\boldsymbol{\mathcal{D}}_{x_i}^*(m)\underline{\boldsymbol{\varphi}}^{01}\left[\mathbf{e}_{ik}(m)\right], \\ &\quad k = 1,2,\ldots,M, \qquad (29)\end{aligned}$$

where μ_{f} is a new step size, and

$$\underline{\widehat{\boldsymbol{h}}}_k^{10}(m) = \boldsymbol{\mathcal{G}}_{2L\times L}^{10}\underline{\widehat{\boldsymbol{h}}}_k(m), \qquad (30)$$

$$\underline{\boldsymbol{\varphi}}^{01}\left[\mathbf{e}_{ik}(m)\right] = \mathbf{F}_{2L\times 2L}\mathbf{W}_{2L\times L}^{01}\boldsymbol{\varphi}\left[\mathbf{e}_{ik}(m)\right]. \qquad (31)$$

It can be seen from (29) that the update equations have high computational complexity due to the need of computing the inverse of matrix $\boldsymbol{\mathcal{Q}}_k(m)$ in (28). When L is large, we can formulate a computationally efficient RNMCFLMM algorithm. To this end, let us approximate $\mathbf{T}_{ik}(m)$ in (25) using

$$\mathbf{T}_{ik}(m) = \rho''\left[e_{ik}(m)\right]_{\max}\mathbf{I}_{L\times L}, \qquad (32)$$

where

$$\rho''\left[e_{ik}(m)\right]_{\max} = \max_{0\leq l\leq L-1}\left\{\rho''\left[e_{ik}(mL+l)\right]\right\}. \qquad (33)$$

So, (24) can be simplified as

$$\boldsymbol{\mathcal{P}}_k(m) = \sum_{i=1,i\neq k}^{M}\rho''\left[e_{ik}(m)\right]_{\max}\boldsymbol{\mathcal{D}}_{x_i}^*(m)\boldsymbol{\mathcal{D}}_{x_i}(m). \qquad (34)$$

It can be seen from (34) that $\boldsymbol{\mathcal{P}}_k(m)$ is simplified to a diagonal matrix, which largely reduces the computational load. It can also be seen from (28) and (34) that the recursive power spectrum matrix is directly deduced from the cost function rather than an empirical formulation [14], [19].

When there are large bursts present in the microphone signals, the normalization by $\boldsymbol{\mathcal{P}}_k(m)$ in (34) greatly diminishes the gradient by exploiting the maximum element in $\{\rho''\left[e_{ik}(mL+l)\right]\}$, which smoothes out the fluctuation due to large bursts of the microphone signals. Furthermore, unlike the NMCFLMS-type algorithms that are based on the use of $\mathbf{e}_{ik}(m)$, the RNMCFLMM algorithm uses $\boldsymbol{\varphi}\left[\mathbf{e}_{ik}(m)\right]$, which can limit the adverse effect of large bursts on the update equations when the error signal becomes very large.

To make the proposed algorithm immune to Gaussian noise, we also introduce a similar spectral constraint as in [20] into the proposed algorithm. So, the update equations of the final RNMCFLMM algorithm are as follows:

$$\begin{aligned}\underline{\widehat{\boldsymbol{h}}}_k^{10}(m+1) &= \underline{\widehat{\boldsymbol{h}}}_k^{10}(m) - \mu_{\mathrm{f}}\nabla\mathcal{J}_{\mathrm{NFM},k}^{01}(m) \\ &\quad + \mu_{\mathrm{f}}\beta(m)\nabla\mathcal{J}_{\mathrm{SC},k}^{10}(m),\ \ k=1,2,\ldots,M, \qquad (35)\end{aligned}$$

where

$$\nabla\mathcal{J}_{\mathrm{NFM},k}^{01}(m) = \boldsymbol{\mathcal{Q}}_k^{-1}(m)\sum_{i=1}^{M}\boldsymbol{\mathcal{D}}_{x_i}^*(m)\underline{\boldsymbol{\varphi}}^{01}\left[\mathbf{e}_{ik}(m)\right], \qquad (36)$$

$$\nabla\mathcal{J}_{\mathrm{SC},k}^{10}(m) = 2\underline{\widehat{\boldsymbol{h}}}_k^{10}(m)\oslash\left(\mathbf{1}_{2L\times 1} + \left|\underline{\widehat{\boldsymbol{h}}}_k^{10}(m)\right|^2\right), \qquad (37)$$

$\oslash$ denotes element-by-element division of two vectors, $\beta(m)$ is the Lagrange multiplier similar to that in the robust NMCFLMS (RNMCFLMS) algorithm [20], $\mathbf{1}_{2L\times 1}$, a vector of length $2L$ with all the elements being 1,

2.2. Multichannel Time Delay Estimation for Acoustic Source Localization

As mentioned previously, we need to estimate the TDOAs between the multiple microphones to locate an acoustic source. Since we obtain the impulse responses via blind system identification, we can determine the time delays by comparing the time differences of the direct-path components between different channels. The TDOA estimate between any two different channels can then be obtained as

$$\widehat{\tau}_{ij} = \arg\max_{l}\ |\widehat{h}_{j,l}| - \arg\max_{l}\ |\widehat{h}_{i,l}|. \qquad (38)$$

3. EXPERIMENTS

3.1. Experimental Setup

The measurements used in this paper were made in the Varechoic chamber at Bell Labs [21]. The dimension of the chamber is 6.7 m×6.1 m×2.9 m. For convenience, positions in the room are designated by (x, y, z) coordinates with reference to the northwest corner of the chamber floor. An equispaced linear array which consists of three omnidirectional microphones is employed in the measurement. The three microphones of the array are situated at (2.437, 0.500, 1.400), (3.137, 0.500, 1.400), and (3.837, 0.500, 1.400), respectively. A sound source is placed at (0.337, 3.938, 1.600). The transfer functions of the acoustic channels between the source and microphones were measured at a 48 kHz sampling rate when 89% panels on the walls were open (the corresponding reverberation time is 280 ms). Then the obtained channel impulse responses are down-sampled to a 16 kHz sampling rate and truncated to 1024 samples. The measured impulse responses will be treated as the true impulse responses in the blind multichannel identification experiments.

The sound source signal is sampled at 16 kHz, the former half of which is recorded from a male English speaker, while the latter half is recorded from a female English speaker. The multichannel system outputs are computed by convolving the source signal with the corresponding measured channel impulse responses and adding noise to the results at a given SNR. The noise used in this paper is modeled by the symmetric α-stable (SαS) distribution [22], [23], where $0 < \alpha < 2.0$ corresponds to non-Gaussian noise and $\alpha = 2.0$ corresponds to Gaussian noise. In the experiments, all the parameters setting is the same as those in [19].

In the experiments, an estimate is yielded every frame with a frame size of 64 ms. Two performance metrics, namely the probability of anomalous estimates and the root mean-squared error (RMSE) of nonanomalous estimates [3], [4], are used to evaluate the performance of the proposed algorithm. The total number of frames is 768. The true time delays from the sound source to the three microphone pairs are respectively $\tau_{12} = 19$ samples, $\tau_{13} = 42$ samples, and $\tau_{23} = 23$ samples.

3.2. Results

The TDE results of the proposed RNMCFLMM algorithm are presented in Table 1. For comparison, the results of the phase transform (PHAT) [1], NMCFLMS [15], and RNMCFLMS [20] algorithms are also presented. As can be seen, the NMCFLMS algorithm outperforms the PHAT algorithm in non-Gaussian and Gaussian noises due to the ability of the former to identify the direct path in reverberant and noisy environments. The RNMCFLMS algorithm is more robust to Gaussian noise than the PHAT and NMCFLMS algorithms since a spectral energy constraint is imposed. Among the four studied algorithms, the proposed RNMCFLMM algorithm obtains the best performance in non-Gaussian and Gaussian noise environments. This is mainly due to the use of an M-estimator and the alternate employment of the LAE and LSE criteria in the M-estimator [19].

Table 1. The probability of anomalous time delay estimates and RMSE of nonanomalous time delay estimates of the four investigated TDE algorithms under different levels of SNR and noises.

α	SNR (dB)	TDE algorithms	Anomalies(%)			RMSE (samples)		
			$\widehat{\tau}_{12}$	$\widehat{\tau}_{13}$	$\widehat{\tau}_{23}$	$\widehat{\tau}_{12}$	$\widehat{\tau}_{13}$	$\widehat{\tau}_{23}$
0.8	−10	PHAT	99.5	99.1	99.3	1.4	1.3	1.4
		NMCFLMS	98.3	97.7	99.5	0.4	0.6	1.6
		RNMCFLMS	99.1	99.3	97.0	0.5	0.7	0.7
		RNMCFLMM	0.4	1.0	0.0	0.2	0.6	0.5
	0	PHAT	65.6	99.8	77.8	0.9	1.4	1.2
		NMCFLMS	7.7	14.5	14.7	0.2	0.7	0.1
		RNMCFLMS	2.9	99.6	99.2	0.1	0.3	0.4
		RNMCFLMM	0.4	0.5	0.5	0.1	0.1	0.3
	10	PHAT	36.1	99.2	63.3	0.5	0.9	0.9
		NMCFLMS	0.5	0.6	0.6	0.2	0.1	0.2
		RNMCFLMS	0.4	0.4	0.1	0.1	0.1	0.1
		RNMCFLMM	0.4	0.1	0.5	0.1	0.1	0.1
	20	PHAT	20.3	66.0	59.6	0.3	0.6	0.7
		NMCFLMS	0.7	0.5	0.9	0.1	0.4	0.1
		RNMCFLMS	0.4	0.1	0.4	0.1	0.1	0.1
		RNMCFLMM	0.3	0.2	0.4	0.0	0.1	0.1
1.2	−10	PHAT	99.5	98.8	99.2	1.4	1.3	1.4
		NMCFLMS	96.6	96.9	99.7	1.0	1.1	1.7
		RNMCFLMS	99.3	99.2	98.2	0.9	0.5	0.9
		RNMCFLMM	0.4	0.9	0.9	0.1	0.3	0.3
	0	PHAT	66.3	99.6	79.8	0.9	1.5	1.2
		NMCFLMS	7.0	8.9	11.6	0.1	0.6	0.6
		RNMCFLMS	7.2	3.0	7.3	0.0	0.6	0.5
		RNMCFLMM	0.4	0.7	0.7	0.1	0.1	0.2
	10	PHAT	37.9	87.2	64.7	0.5	0.7	0.9
		NMCFLMS	0.6	0.8	0.9	0.2	0.7	0.2
		RNMCFLMS	0.3	0.1	0.3	0.1	0.1	0.1
		RNMCFLMM	0.4	0.4	0.5	0.1	0.1	0.1
	20	PHAT	20.3	55.5	60.9	0.3	0.7	0.7
		NMCFLMS	0.6	0.7	0.9	0.1	0.1	0.1
		RNMCFLMS	0.4	0.1	0.4	0.1	0.1	0.1
		RNMCFLMM	0.4	0.5	0.5	0.1	0.1	0.0
1.6	−10	PHAT	89.6	98.4	99.5	1.4	1.6	1.4
		NMCFLMS	80.5	61.1	94.5	1.6	1.1	0.8
		RNMCFLMS	99.8	98.1	99.0	0.2	0.4	0.3
		RNMCFLMM	0.7	5.5	5.5	0.1	0.1	0.4
	0	PHAT	66.2	95.8	78.1	0.9	0.9	1.3
		NMCFLMS	2.6	1.4	31.9	0.9	0.5	0.6
		RNMCFLMS	23.2	0.8	22.1	0.3	0.9	0.9
		RNMCFLMM	0.4	0.7	0.8	0.2	0.0	0.2
	10	PHAT	33.2	69.7	61.2	0.5	0.8	0.9
		NMCFLMS	0.7	1.0	0.9	0.1	0.1	0.1
		RNMCFLMS	0.4	0.3	0.3	0.1	0.6	0.6
		RNMCFLMM	0.3	0.5	0.5	0.1	0.1	0.0
	20	PHAT	19.1	37.5	58.9	0.3	0.6	0.7
		NMCFLMS	0.7	0.5	0.9	0.1	0.1	0.1
		RNMCFLMS	0.4	0.2	0.4	0.1	0.3	0.1
		RNMCFLMM	0.4	0.3	0.4	0.1	0.1	0.0
2.0	−10	PHAT	87.8	95.8	88.9	1.3	1.4	1.4
		NMCFLMS	45.1	23.2	45.3	0.8	0.8	1.2
		RNMCFLMS	33.4	11.5	41.8	0.6	0.2	0.6
		RNMCFLMM	9.2	2.7	9.2	0.0	0.1	0.1
	0	PHAT	58.6	77.2	74.1	1.1	0.9	1.3
		NMCFLMS	1.2	1.2	1.7	0.3	0.5	0.6
		RNMCFLMS	0.7	0.3	0.1	0.1	0.9	0.9
		RNMCFLMM	0.7	0.4	0.1	0.1	0.6	0.6
	10	PHAT	30.1	37.8	58.5	0.6	0.8	0.9
		NMCFLMS	0.5	0.8	0.6	0.2	0.1	0.2
		RNMCFLMS	0.3	0.2	0.1	0.1	0.1	0.1
		RNMCFLMM	0.4	0.3	0.2	0.1	0.1	0.1
	20	PHAT	17.8	22.4	57.1	0.3	0.7	0.7
		NMCFLMS	0.6	0.7	1.0	0.1	0.1	0.1
		RNMCFLMS	0.4	0.1	0.4	0.1	0.1	0.1
		RNMCFLMM	0.2	0.2	0.4	0.1	0.1	0.0

4. CONCLUSIONS

In this paper, we developed a robust multichannel TDE algorithm. This algorithm uses an M-estimator to construct a recursive cost function of the multichannel frequency-domain blind identification algorithm and exploits the property of the M-estimator to deal with outliers in the microphones' signals, thereby ensuring the robustness of TDE with respect to non-Gaussian noises. Moreover, this TDE algorithm is also robust to Gaussian noise thanks to the alternate use of the LSE and LAE criteria. The experimental results confirmed that the developed algorithm is robust to noise in different SNR conditions regardless of whether the noise is Gaussian and non-Gaussian.

5. REFERENCES

[1] Y. Huang, J. Benesty, and J. Chen, *Acoustic MIMO Signal Processing*. Boston, MA: Springer, 2006.

[2] C. H. Knapp and G. C. Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-24, no. 8, pp. 320–327, Aug. 1976.

[3] J. Chen, J. Benesty, and Y. Huang, "Robust time delay estimation exploiting redundancy among multiple microphones," *IEEE Trans. Speech Audio Process.*, vol. 11, no. 6, pp. 549–557, Nov. 2003.

[4] H. He, L. Wu, J. Lu, X. Qiu, and J. Chen, "Time difference of arrival estimation exploiting multichannel spatio-temporal prediction," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 21, no. 3, pp. 463–475, Mar. 2013.

[5] H. He, J. Chen, J. Benesty, and T. Yang, "On time delay estimation based on multichannel spatiotemporal sparse linear prediction," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2016, pp. 390–394.

[6] F. Talantzis, A. G. Constantinides, and L. C. Polymenakos, "Estimation of direction of arrival using information theory," *IEEE Signal Process. Lett.*, vol. 12, pp. 561–564, Aug. 2005.

[7] J. Benesty, Y. Huang, and J. Chen, "Time delay estimation via minimum entropy," *IEEE Signal Process. Lett.*, vol. 14, pp. 157–160, Mar. 2007.

[8] M. S. Brandstein, "A pitch-based approach to time-delay estimation of reverberant speech," in *Proc. IEEE Workshop Applicat. Signal Process. Audio Acoust. (WASPAA)*, 1997.

[9] T. G. Dvorkind and S. Gannot, "Time difference of arrival estimation of speech source in a noisy and reverberant environment," *Elsevier Signal Process.*, vol. 85, pp. 177–204, Jan. 2005.

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

[11] Y. Huang, J. Benesty, and G. W. Elko, "Adaptive eigenvalue decomposition algorithm for realtime acoustic source localization system," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 1998, vol. 2, pp. 937–940.

[12] L. Tong, G. Xu, and T. Kailath, "Blind identification and equalization based on second-order statistics: A time domain approach," *IEEE Trans. Inf. Theory*, vol. 40, no. 2, pp. 340–349, Mar. 1994.

[13] G. Xu, H. Liu, L. Tong, and T. Kailath, "A least-squares approach to blind channel identification," *IEEE Trans. Signal Process.*, vol. 43, no.12, pp. 2982–2993, Dec. 1995.

[14] Y. Huang and J. Benesty, "A class of frequency-domain adaptive approaches to blind multichannel identification," *IEEE Trans. Signal Process.*, vol. 51, no. 1, pp. 11–24, Jan. 2003.

[15] J. Chen, Y. Huang, and J. Benesty, "An adaptive blind SIMO identification approach to joint multichannel time delay estimation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2004, pp. IV-53–IV-56.

[16] P. J. Huber, *Robust Statistics*. New York, NY: Wiley, 1981.

[17] S. C. Chan and Y. X. Zou, "A recursive least M-estimate algorithm for robust adaptive filtering in impulse noise: Fast algorithm and convergence performance analysis," *IEEE Trans. Signal Process.*, vol. 52, no. 4, pp. 975–991, Apr. 2004.

[18] H. Buchner, J. Benesty, T. Gänsler, and W. Kellermann, "Robust extended multidelay filter and double-talk detector for acoustic echo cancellation," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, no. 5, pp. 1633–1644, Sep. 2006.

[19] H. He, J. Lu, J. Chen, X. Qiu, and J. Benesty, "Robust blind identification of room acoustic channels in symmetric alpha-stable distributed noise environments," *J. Acoust. Soc. Amer.*, vol. 136, no. 8, pp. 693–704, Aug. 2014.

[20] M. A. Haque and M. K. Hasan, "Noise robust multichannel frequency-domain LMS algorithms for blind channel identification," *IEEE Signal Process. Lett.*, vol. 15, pp. 305–308, 2008.

[21] A. Härmä, "Acoustic measurement data from the varechoic chamber," Tech. Memorandum 110101, Agere Systems, Allentown, Pa, USA, 2001.

[22] C. L. Nikias and M. Shao, *Signal Processing With Alpha-Stable Distributions and Applications*. New York, NY: Wiley, 1995.

[23] J. M. Chambers, C. L. Mallows, and B. W. Stuck, "A method for simulating stable random variables," *J. Amer. Statist. Assoc.*, vol. 71, pp. 340–344, Jun. 1976.