# Direction-of-arrival estimation of passive acoustic sources in reverberant environments based on the Householder transformation

Gongping Huang and Jingdong Chen[a)]
*Center of Intelligent Acoustics and Immersive Communications and School of Marine Science and Technology, Northwestern Polytechnical University, 127 Youyi West, Road, Xi'an 710072, China*

Jacob Benesty
*INRS-EMT, University of Quebec, 800 de la Gauchetiere Ouest, Montreal, Quebec H5A 1K6, Canada*

(Received 16 March 2015; revised 11 August 2015; accepted 16 October 2015; published online 16 November 2015)

This paper presents an approach to the direction-of-arrival (DOA) estimation problem in acoustic environments using microphone arrays. It works in the short-time Fourier transform (STFT) domain. It first transforms the noisy speech signals received at the array into the STFT domain. A Householder transformation is then constructed and applied to the multichannel STFT coefficients in each subband. This transformation converts the multichannel STFT coefficients into two components: one is a single coefficient that is dominated by the signal of interest and the other consists of the $M-1$ coefficient that is dominated by noise (or even consists of noise-only if there is no reverberation), where $M$ is the number of sensors. A cost function is then formed from the outputs of the Householder transformation and the DOA information can subsequently be obtained by searching the extremum value of this cost function in the angle range between 0° and 180°. Simulation results are provided to illustrate the performance of this approach.
 [http://dx.doi.org/10.1121/1.4934954]

[KGS] Pages: 3053–3060

## I. INTRODUCTION

The direction-of-arrival (DOA) estimation, which serves as the first stage that feeds into subsequent processing blocks of an array system to detect, identify, and localize radiating sources, has plenty of applications in fields as diverse as radar, sonar, acoustics, and voice communications (Nagata *et al.*, 2007; Chen *et al.*, 2008; Pesavento and Gershman, 2001; Chung and Bohme, 2002; Ward *et al.*, 1988; Hyder and Mahata, 2010; Reddy *et al.*, 2014; Qian *et al.*, 2014; Li and Lu, 2007; Ma *et al.*, 2010; Zhang *et al.*, 2013; Tan and Nehorai, 2014; Nesta and Omologo, 2012; Benesty *et al.*, 2008). It has attracted a considerable amount of research attention ever since sensor arrays were introduced to measure a propagating wavefield. This paper deals with the DOA estimation problem in acoustic environments, which is an essential part of many voice communication systems such as multi-party conferencing (Omologo and Svaizer, 1994; Wang and Chu, 1997; Huang *et al.*, 2011). In such a problem as illustrated in Fig. 1, the acoustic source (e.g., a talker or a loudspeaker) radiates a plane wave that propagates through the air. The normal to the wavefront makes an angle $\theta$ with the line joining the sensors in the linear array and the signal received at each microphone is a time delayed version of the signal at the reference sensor. The objective of DOA estimation is then to estimate the angle $\theta$ based on the signals observed at the microphones.

Instead of estimating the angle $\theta$ directly, one can also first estimate the time-difference-of-arrival (TDOA) among different sensors and then $\theta$ can be obtained by solving a trigonometric equation. This can be seen from Fig. 1. If we denote by $\tau_{21}$ the TDOA between the second and first sensors, we have $\tau_{21} = \delta \cos\theta / c$, where $c$ is the sound velocity in the air and $\delta$ is the spacing between the sensors. Therefore, given $\tau_{21}$, one can easily obtain $\theta$ and vice versa. This indicates that the DOA estimation problem is the same as the TDOA estimation (also called time delay estimation) one, and any algorithm that works for the former should work for the latter.

The generalized cross-correlation (GCC) method, proposed by Knapp and Carter (1976), is the most popular

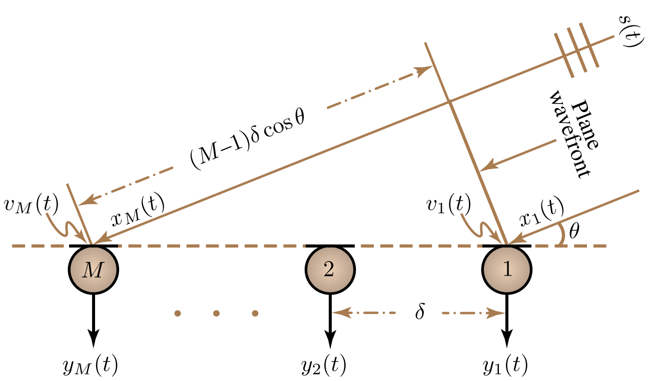


FIG. 1. (Color online) Illustration of a uniform linear array for DOA estimation. $M$ is the number of microphones, $\delta$ is the inter-element spacing, $\theta$ is the source incidence angle, $s(t)$ is the source signal, $x_m(t)$ is the source signal received at the $m$th microphone, $v_m(t)$ is the additive noise at the $m$th microphone, and $y_m(t)$ is the $m$th microphone's output.

[a)]Electronic mail: jingdongchen@ieee.org technique for DOA estimation. With this method, the DOA estimate is obtained as the angle (or the TDOA estimate is obtained as the time-lag) that maximizes the cross-correlation function between filtered versions of the received signals. Since then, many new ideas have been proposed to better deal with noise and reverberation (Benesty, 2000; Moghaddam *et al.*, 2003; Chen *et al.*, 2003; Benesty *et al.*, 2004; Chen *et al.*, 2006; Lombard *et al.*, 2011). However, reverberation remains a challenging problem and in highly reverberant rooms all existing methods do not perform well.

One effective way to improve the robustness of DOA estimation with respect to reverberation is by taking advantage of the redundancy from multiple microphones (Chen *et al.*, 2003; Benesty *et al.*, 2004). Following this principle, we present in this paper a new multichannel DOA estimation approach based on the Householder transformation (HT) (de Campos *et al.*, 1999, 2002; Golub and Loan, 1996). This approach first transforms the multichannel signals into the short-time Fourier transform (STFT) domain. A Householder transformation is then constructed and applied to the multichannel STFT coefficients in each subband. This transformation converts the multichannel STFT coefficients into two components: one is a single coefficient that is dominated by the signal of interest and the other consists of $M-1$ coefficient that is dominated by noise (or even consists of noise-only if there is no reverberation), where $M$ is the number of sensors. A cost function is then formed from the outputs of the Householder transformation and the DOA information is subsequently obtained by searching the extremum value of this cost function in the angle range between $0^\circ$ and $180^\circ$.

The major contributions of this paper are twofold. First, it introduces the Householder transformation to the problem of DOA estimation. Second, based on this transformation, a DOA estimator is developed for either narrowband or broadband cases. It can achieve DOA estimation using either two or multiple microphones. In the multiple-microphone case, the DOA estimation performance in noise and reverberation increases with the number of sensors.

The rest of this paper is organized as follows. In Sec. II, we present the signal model and the problem formulation. In Sec. III, we briefly introduce the Householder transformation in the context of microphone arrays. Then, we discuss how to utilize it in the DOA estimation problem in Sec. IV. In Sec. V, simulations are provided to illustrate the performance of the developed method in different environments with noise and reverberation. Finally, some conclusions are drawn in Sec. VI.

## II. SIGNAL MODEL AND PROBLEM FORMULATION

We consider a uniform linear array (ULA) consisting of $M$ omnidirectional microphones, where the spacing between two successive sensors is equal to $\delta$ as illustrated in Fig. 1. Let us first assume that the environment is free of reverberation and the speech source is in the far field. If we choose the first microphone as the reference, the received signals, at the discrete-time index $t$, can be expressed as (Benesty *et al.*, 2008; Chen *et al.*, 2006)

$$\begin{aligned} y_m(t) &= x_1(t-\tau_{m1}) + v_m(t) \\ &= x_m(t) + v_m(t), \quad m = 1,2,\ldots,M, \end{aligned} \tag{1}$$

where $x_1(t)$ is the speech source signal at the first microphone, $x_m(t) = x_1(t-\tau_{m1})$ is the signal received at the $m$th microphone, $\tau_{m1}$ is the relative time delay between microphones $m$ and 1, and $v_m(t)$ is the additive noise at the $m$th microphone. All signals are considered to be real and zero-mean random processes, and $x_m(t)$ and $v_m(t)$ are assumed to be independent. Denoting by $\theta$ the source incidence angle, we have

$$\tau_{m1} = (m-1)\tau_0 \cos\theta, \quad m = 1,2,\ldots,M,$$

where $\tau_0 = \delta/c$ is the delay between two successive microphones at the angle 0.

In the frequency domain, at the frequency index $f$, Eq. (1) can be written as (Benesty *et al.*, 2008)

$$\begin{aligned} Y_m(f) &= X_m(f) + V_m(f) \\ &= e^{-j2(m-1)\pi f \tau_0 \cos\theta} X_1(f) + V_m(f), \quad m = 1,2,\ldots,M, \end{aligned} \tag{2}$$

where $Y_m(f)$, $X_m(f)$, and $V_m(f)$ are the frequency-domain representations of $y_m(t)$, $x_m(t)$, and $v_m(t)$, respectively, and $j$ is the imaginary unit with $j^2 = -1$.

The objective of DOA estimation is to estimate the incidence angle $\theta$ from the $M$ observations $Y_m(f)$, $m = 1, 2, \ldots, M$. To achieve this goal, it is more convenient to write the $M$ frequency-domain microphone signals in a vector notation

$$\begin{aligned} \mathbf{y}(f) &= \mathbf{x}(f) + \mathbf{v}(f) \\ &= \mathbf{d}(f,\theta) X_1(f) + \mathbf{v}(f), \end{aligned} \tag{3}$$

where

$$\begin{aligned} \mathbf{y}(f) &\triangleq [\, Y_1(f) \quad Y_2(f) \quad \cdots \quad Y_M(f) \,]^T, \\ \mathbf{x}(f) &\triangleq [\, X_1(f) \quad X_2(f) \quad \cdots \quad X_M(f) \,]^T, \\ \mathbf{v}(f) &\triangleq [\, V_1(f) \quad V_2(f) \quad \cdots \quad V_M(f) \,]^T, \end{aligned}$$

where the superscript $^T$ is the transpose operator, and

$$\mathbf{d}(f,\theta) \triangleq [\, 1 \quad e^{-j2\pi f \tau_0 \cos\theta} \quad \cdots \quad e^{-j2(M-1)\pi f \tau_0 \cos\theta} \,]^T, \tag{4}$$

is a phase-delay vector of length $M$ (its form is the same as the steering vector used in traditional beamforming).

Now, if there is reverberation, the received signals can be written in the time domain as

$$\begin{aligned} y_m(t) &= g_m(t) * s(t) + v_m(t) \\ &= x_m(t) + v_m(t), \quad m = 1,2,\ldots,M, \end{aligned} \tag{5}$$

where $*$ stands for linear convolution, $g_m(t)$ is the aoustic impulse response from the position of $s(t)$ to the $m$th microphone, $s(t)$ is the unknown speech source, and $x_m(t) = g_m(t) * s(t)$ is the convolved speech signal at the $m$th sensor. With this model, the DOA information is related to the time difference between the direct paths of the channel impulse responses.

Again, in the frequency domain, Eq. (5) can be expressed as

$$\begin{aligned} Y_m(f) &= G_m(f)S(f) + V_m(f) \\ &= X_m(f) + V_m(f), \quad m = 1, 2, \ldots, M, \end{aligned} \tag{6}$$

where $Y_m(f)$, $G_m(f)$, $S(f)$, $X_m(f)$, and $V_m(f)$ are the frequency-domain representations of $y_m(t)$, $g_m(t)$, $s(t)$, $x_m(t)$, and $v_m(t)$, respectively. In a vector form, Eq. (6) is written as

$$\begin{aligned} \mathbf{y}(f) &= \mathbf{g}(f)S(f) + \mathbf{v}(f) \\ &= \mathbf{x}(f) + \mathbf{v}(f) \\ &= \mathbf{d}(f)X_1(f) + \mathbf{v}(f), \end{aligned} \tag{7}$$

where

$$\mathbf{g}(f) \triangleq [\, G_1(f) \quad G_2(f) \quad \cdots \quad G_M(f)\,]^T \tag{8}$$

and

$$\mathbf{d}(f) \triangleq \left[ 1 \quad \frac{G_2(f)}{G_1(f)} \quad \cdots \quad \frac{G_M(f)}{G_1(f)} \right]^T = \frac{\mathbf{g}(f)}{G_1(f)}. \tag{9}$$

## III. HOUSEHOLDER TRANSFORMATION

We first define the Householder transformation (Golub and Loan, 1996) associated with $\mathbf{d}(f, \theta)$ as

$$\mathbf{T}(f, \theta) \triangleq \mathbf{I}_M - \frac{2}{\mathbf{b}^H(f, \theta)\mathbf{b}(f, \theta)} \mathbf{b}(f, \theta)\mathbf{b}^H(f, \theta), \tag{10}$$

where $\mathbf{I}_M$ is the $M \times M$ identity matrix, the superscript $^H$ denotes the conjugate-transpose operator, and

$$\mathbf{b}(f, \theta) \triangleq \mathbf{d}(f, \theta) + \sqrt{M}\, \mathbf{i}_1, \tag{11}$$

with $\mathbf{i}_1$ being the first column of the identity matrix $\mathbf{I}_M$, i.e.,

$$\mathbf{i}_1 = [\,1 \quad 0 \quad \cdots \quad 0\,]^T. \tag{12}$$

It can be checked that $\mathbf{T}(f, \theta)$ is Hermitian and unitary, i.e., $\mathbf{T}^H(f, \theta) = \mathbf{T}(f, \theta)$ and $\mathbf{T}(f, \theta)\mathbf{T}^H(f, \theta) = \mathbf{I}_M$.

It is easy to verify that

$$\mathbf{T}(f, \theta)\mathbf{x}(f) = -\sqrt{M} X_1(f)\, \mathbf{i}_1. \tag{13}$$

So, the Householder transformation projects the vector $\mathbf{x}(f)$ into another vector that has zeros in all positions but one, and the only nonzero entry is the first element of $\mathbf{T}(f, \theta)\mathbf{x}(f)$, which is equal to $-\sqrt{M} X_1(f)$.

Left-multiplying both sides of Eq. (3) by $-\mathbf{T}(f, \theta)/\sqrt{M}$, we get

$$\begin{aligned} \mathbf{y}'(f) &= -\frac{1}{\sqrt{M}} \mathbf{T}(f, \theta)\mathbf{y}(f) \\ &= \mathbf{i}_1 X_1(f) - \frac{1}{\sqrt{M}} \mathbf{T}(f, \theta)\mathbf{v}(f) \\ &= \mathbf{i}_1 X_1(f) + \mathbf{v}'(f), \end{aligned} \tag{14}$$

or, alternatively,

$$\begin{bmatrix} Y_1'(f) \\ \mathbf{y}_2'(f) \end{bmatrix} = \begin{bmatrix} X_1(f) \\ 0_{(M-1)\times 1} \end{bmatrix} + \begin{bmatrix} V_1'(f) \\ \mathbf{v}_2'(f) \end{bmatrix}. \tag{15}$$

We see how the Householder transformation gives a clear noise reference signal. Indeed, $Y_1'(f) = X_1(f) + V_1'(f)$ is the sum of the desired signal and noise, while the $(M-1)$-dimensional vector $\mathbf{y}_2'(f) = \mathbf{v}_2'(f)$ contains noise only. Note that in the presence of reverberation, $\mathbf{y}_2'(f)$ will contain some speech, but noise will still be dominant.

## IV. DOA ESTIMATION BASED ON THE HOUSEHOLDER TRANSFORMATION

In practice, the DOA $\theta$ is not known and needs to be estimated. Now, let us consider any angle $\theta_1$ $(0° \leq \theta_1 \leq 180°)$. The Householder transformation associated with the steering vector, $\mathbf{d}(f, \theta_1)$, is

$$\mathbf{T}(f, \theta_1) \triangleq \mathbf{I}_M - \frac{2}{\mathbf{b}^H(f, \theta_1)\mathbf{b}(f, \theta_1)} \mathbf{b}(f, \theta_1)\mathbf{b}^H(f, \theta_1), \tag{16}$$

where $\mathbf{b}(f, \theta_1)$ is defined in the same manner as $\mathbf{b}(f, \theta)$ in Eq. (11). Left-multiplying both sides of Eq. (3) by $-\mathbf{T}(f, \theta_1)/\sqrt{M}$, we get

$$\begin{bmatrix} Y_1'(f, \theta_1) \\ \mathbf{y}_2'(f, \theta_1) \end{bmatrix} = \begin{bmatrix} X_1'(f, \theta_1) \\ \mathbf{x}_2'(f, \theta_1) \end{bmatrix} + \begin{bmatrix} V_1'(f, \theta_1) \\ \mathbf{v}_2'(f, \theta_1) \end{bmatrix}. \tag{17}$$

Now, we show how the Householder transformation can be used to estimate the DOA by scanning the space from 0° to 180°. As a matter of fact, it is easy to check that $\mathbf{y}_2'(f, \theta_1) = 0_{(M-1)\times 1}$ when $\theta_1 = \theta$ and with no reverberation. It follows naturally that the DOA estimation can be obtained from $Y_m'(f, \theta_1)$, $m = 2, 3, \ldots, M$, the elements of $\mathbf{y}_2'(f, \theta_1)$. In practice, the DOA estimation can be achieved by finding the minimum of the statistics of $Y_m'(f, \theta_1)$, e.g., $\phi_{Y_m', \beta}(f, \theta_1) = E[|Y_m'(f, \theta_1)|^\beta]$, with $E[\cdot]$ denoting mathematical expectation and $\beta$ is a positive real constant. Indeed, it is obvious that for $\theta_1 = \theta$, we have

$$\begin{aligned} \phi_{Y_m', \beta}(f, \theta_1) &= E[|Y_m'(f, \theta_1)|^\beta] \\ &= \phi_{V_m', \beta}(f, \theta_1), \quad m = 2, 3, \ldots, M, \end{aligned} \tag{18}$$

where $\phi_{V_m', \beta}(f, \theta_1)$ is defined in a similar way to $\phi_{Y_m', \beta}(f, \theta_1)$. As a result, the DOA $\theta$ can be determined as

$$\hat{\theta} = \arg\min_{\theta_1} \phi_{Y_m', \beta}(f, \theta_1), \quad m = 2, 3, \ldots, M, \tag{19}$$

or, more effectively, as

$$\hat{\theta} = \arg\min_{\theta_1} \bar{\phi}_{Y', \beta}(f, \theta_1), \tag{20}$$

where

$$\bar{\phi}_{Y', \beta}(f, \theta_1) \triangleq \frac{1}{M-1} \sum_{m=2}^{M} \phi_{Y_m', \beta}(f, \theta_1). \tag{21}$$

The previous two estimators are defined on a narrowband basis. But they can be extended to process broadband signals such as speech. For example, one can define a broadband counterpart of $\bar{\phi}_{Y',\beta}(f,\theta_1)$ as

$$\tilde{\phi}_{Y',\beta}(\theta_1) = \int_{f_1}^{f_2} \frac{\bar{\phi}_{Y',\beta}(f,\theta_1)}{\int_{\theta} \bar{\phi}_{Y',\beta}(f,\theta)d\theta} df, \tag{22}$$

where $f_1$ and $f_2$ are the lower and upper cutoff frequencies of the band of interest. Note that we normalize $\bar{\phi}_{Y',\beta}(f,\theta_1)$ by $\int_{\theta}\bar{\phi}_{Y',\beta}(f,\theta)d\theta$ in defining the broadband cost function $\tilde{\phi}_{Y',\beta}(\theta_1)$ to make sure that all the subbands make equal contributions to the DOA estimation. Then, the DOA $\theta$ is determined as

$$\hat{\theta} = \arg\min_{\theta_1} \tilde{\phi}_{Y',\beta}(\theta_1). \tag{23}$$

It can be checked that the DOA can also be determined by maximizing $\phi_{Y'_1,\beta}(f,\theta_1)$ or extremizing the combination between $\phi_{Y'_1,\beta}(f,\theta_1)$ and $\phi_{Y'_m,\beta}(f,\theta_1)$, which will not be discussed in detail in this paper.

The previous estimators in Eqs. (19), (20), and (23) are derived based on microphone 1 as the reference sensor. In practice, due to different reasons such as sensor mismatch, noise, and reverberation, using microphone 1 as the reference may not always produce the best performance. One way to improve this is to compute the cost functions for all the $M$ cases, where in each case we take microphone $m$, $m = 1, 2, \ldots, M$, as the reference and then combine those cost functions together to form the overall cost function for DOA estimation. This would increase the complexity by a factor of $M$, but can help improve the robustness of DOA estimation with respect to noise, reverberation, and the array imperfection.

## V. SIMULATIONS

The image model method is used to simulate reverberant acoustic environments (Allen and Berkley, 1979; Huang *et al.*, 2006). The room size is $3\,\text{m} \times 3\,\text{m} \times 3\,\text{m}$. A linear array is used, which consists of 8 omnidirectional microphones placed at $(x, 2.0, 1.6)$, where $x = 1.72 : 0.04 : 2.0$. To simulate the source, a loudspeaker is placed at the position $(2.5, 2.86, 2.0)$, playing back a pre-recorded speech signal. The sampling rate is 16 kHz. In the simulations, the acoustic channel impulse responses from the source to the microphones are generated with the image model method

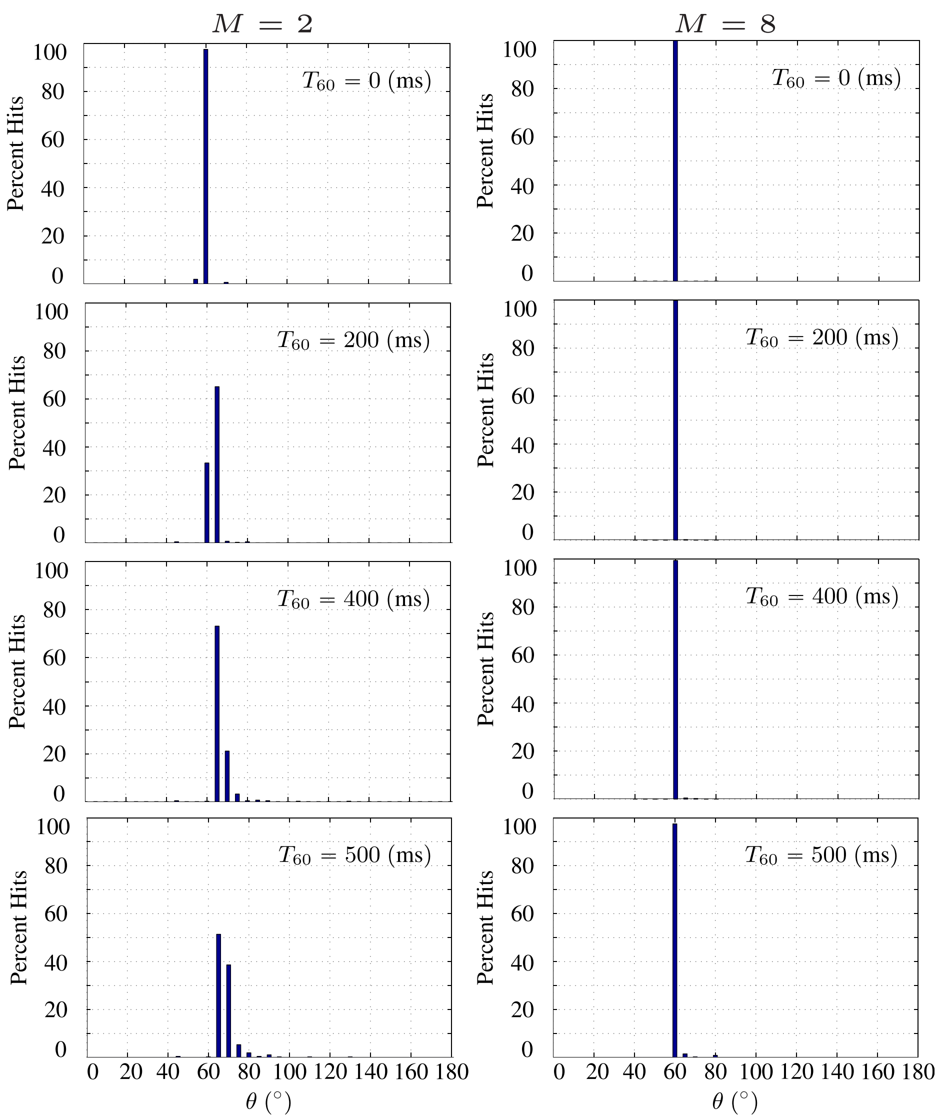


FIG. 2. (Color online) Histograms of the DOA estimates with the HT algorithm in both reverberant and noisy environments: SNR = 10 dB, $\beta = 2$, and the true DOA is at 60°.

(Allen and Berkley, 1979; Huang *et al.*, 2006). Then, the microphone signals are generated by convolving the source signal with the corresponding impulse responses and white Gaussian noise is subsequently added to control the signal-to-noise ratio (SNR).

DOA estimation is carried out in the STFT domain. The array signals are partitioned into non-overlapping time frames of size 64 ms and each frame is then transformed into the STFT domain using a 1024-point fast Fourier transform (FFT). DOA estimates are obtained on a frame-by-frame basis. The statistics $\phi_{Y'_m,\beta}(f,\theta_1)$, $m=1,2,\ldots,M$, at the $k$th frame are computed from $Y'_m(f,\theta_1,k)$ [where $Y'_m(f,\theta_1,k)$ denotes $Y'_m(f,\theta_1)$ computed at the $k$th frame] with a recursive method as

$$\hat{\phi}_{Y'_m,\beta}(f,\theta_1,k) = \lambda\hat{\phi}_{Y'_m,\beta}(f,\theta_1,k-1) + (1-\lambda)|Y'_m(f,\theta_1,k)|^{\beta}, \quad (24)$$

where $\lambda$ is a forgetting factor that controls the influence of the previous data samples on the current estimate. In our simulations, we set $\lambda = 0.90$. Note that the optimal value of this parameter can be found through experiments, which is beyond the scope of this paper. Now, substituting the estimated statistics $\hat{\phi}_{Y'_m,\beta}(f,\theta_1,k)$, $m=2,\ldots,M$, into Eq. (23), we can perform DOA estimation for broadband speech signals.

The performance of the developed algorithm is evaluated in both noisy and reverberant environments where the noise is spatially white with a 10-dB SNR. Regarding reverberation, we use the image model to simulate four conditions: anechoic condition with no reflections, a light reverberation condition with all the six-wall reflection coefficients being 0.8 and the reverberation time $T_{60}$ is approximately 200 ms, and two very reverberant environments with the reflections coefficients being 0.9 ($T_{60} \approx 400$ ms) and 0.95 ($T_{60} \approx 500$ ms), respectively.

The histograms of the DOA estimates with $\beta = 2$ are plotted in Fig. 2. As seen, reverberation can significantly affect the DOA estimation and the performance of the HT based DOA estimation algorithm with a given number of microphones decreases as the reverberation time increases. It is also seen that in the same environment, the DOA estimation performance with eight microphones is much better than that with two microphones. This shows that the HT algorithm can take advantage of the redundant information provided by multiple sensors to improve the DOA performance.

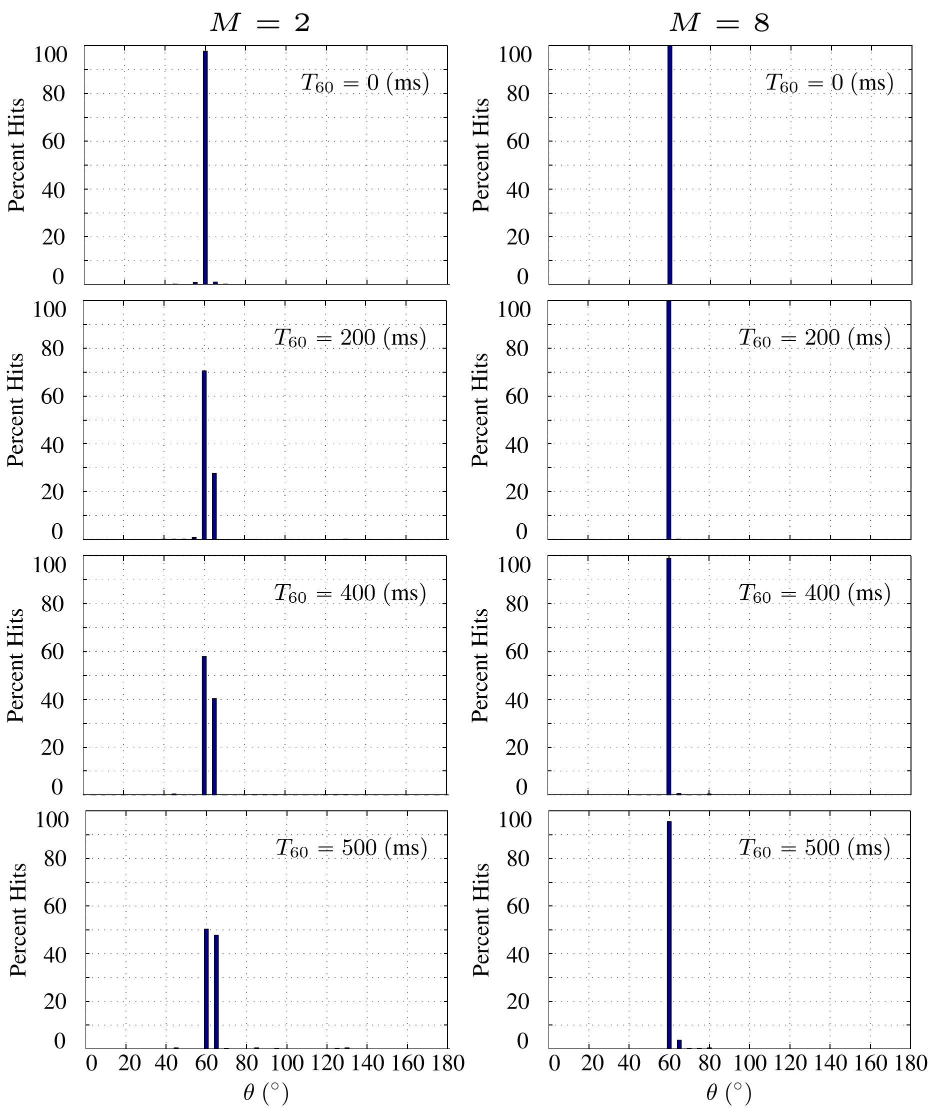


FIG. 3. (Color online) Histograms of the DOA estimates with the HT algorithm in both reverberant and noisy environments: SNR = 10 dB, $\beta = 4$, and the true DOA is at 60°.

The parameter $\beta$ may also play an important role on the DOA estimation performance. To illustrate this, we repeat the previous simulation but with $\beta = 4$. Comparing Figs. 2 and 3, one can see that the DOA estimation performance with $\beta = 4$ is consistently better than that with $\beta = 2$. The optimal value of $\beta$ can be found through simulations in practice, which will not be discussed in detail due to the limit in space.

We also compare the developed algorithm with the popularly used PHAT algorithm (Knapp and Carter, 1976; Chen *et al.*, 2006). Since PHAT is typically used for TDOA estimation (the DOA estimates can be obtained according to the relationship $\tau_{21} = \delta \cos\theta / c$), we compare the results of the TDOA estimation. The results in two reverberant ($T_{60} = 200$ ms and 500 ms) and noisy (spatially white noise with SNR = 10 dB) environments are plotted in Fig. 4. It is seen that the TDOA performance with both the HT and PHAT methods decreases as reverberation becomes stronger. When two microphones are used, the HT algorithm achieves a performance comparable to (if $\beta = 2$) or slightly better than (if $\beta = 4$) that of the PHAT method in all the studied conditions. When eight microphones are used, the HT algorithm yields a significantly better performance than the PHAT method.

In the last simulation, we compare the developed algorithm with the well-known steered response power (SRP) algorithm (Do *et al.*, 2007), in which the SRP cost function is computed by summing the PHAT based GCC functions

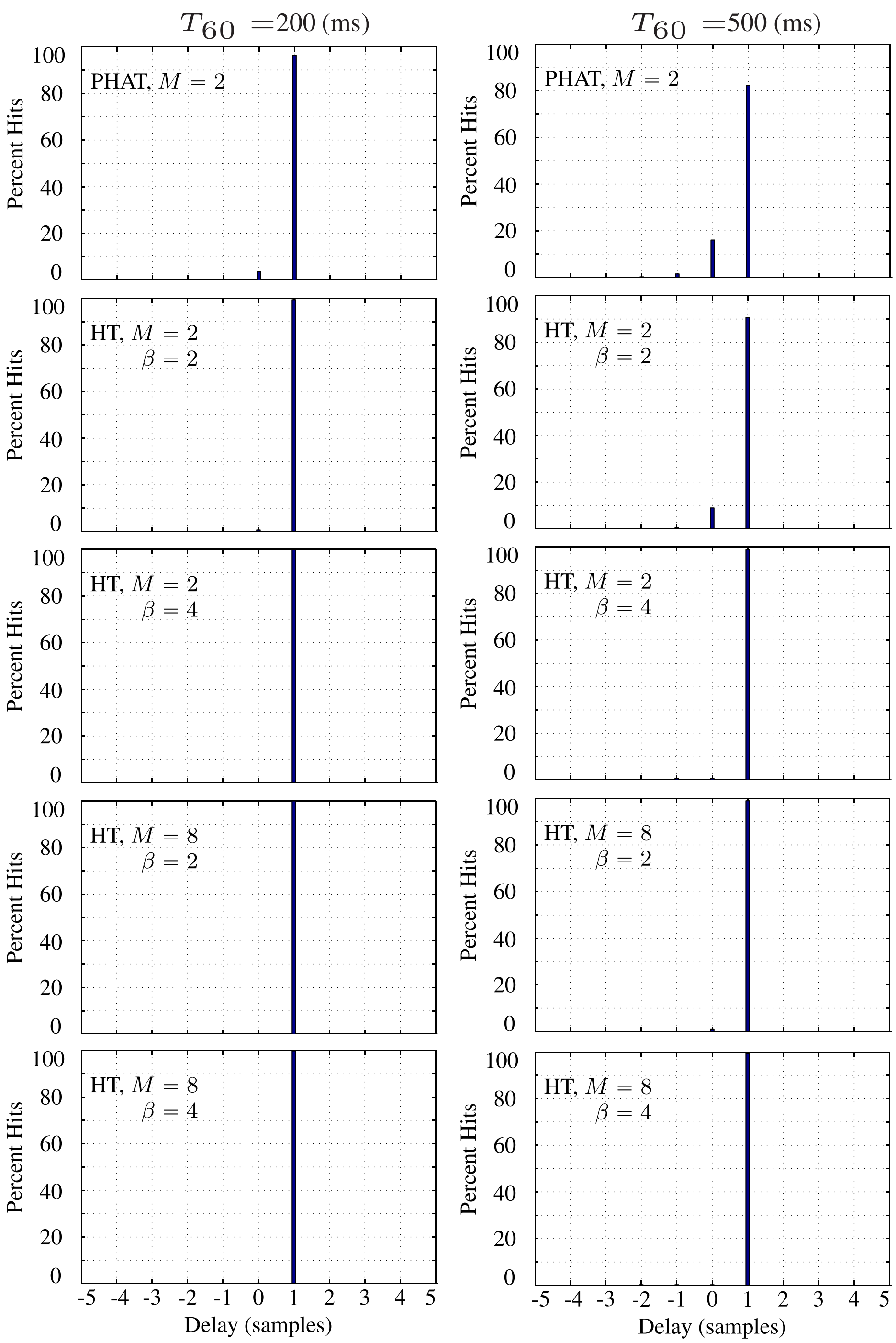


FIG. 4. (Color online) Histograms of the TDOA estimates with the PHAT and HT methods in both reverberant and noisy environments: SNR = 10 dB and the true TDOA is at 60°.

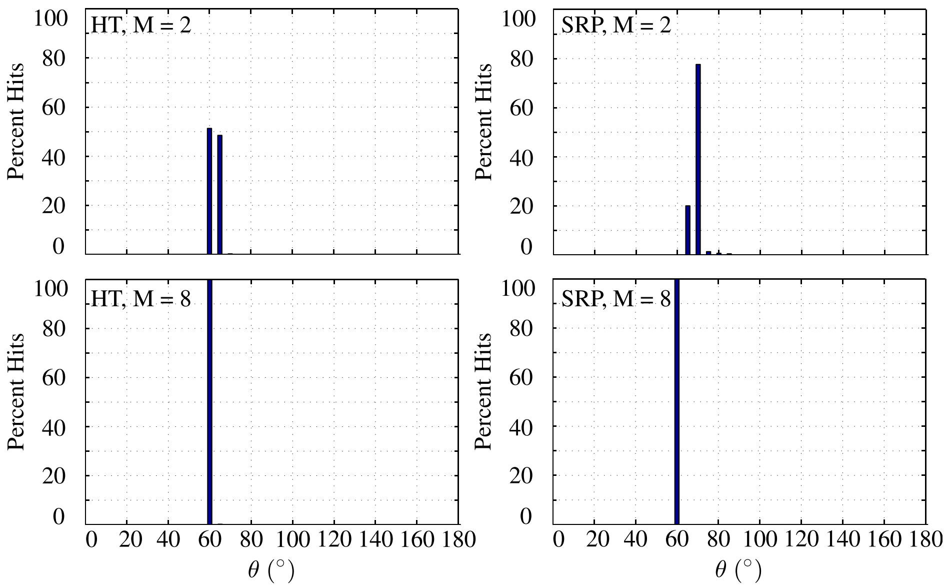


FIG. 5. (Color online) Histograms of the DOA estimates with the HT and SRP methods in both reverberant and noisy environments: SNR = 10 dB, $T_{60} \approx 500$ ms, and the true TDOA is at 60°.

from all the pairs of microphones. The input SNR is 10 dB and the reverberation time is approximately 500 ms. We set $\beta = 4$ and keep all the other parameters the same as in the previous experiments. The results are plotted in Fig. 5. It is seen that when two microphones are used, the developed algorithm outperforms the SRP method in the studied environment. When $M = 8$, both algorithms yield an accurate DOA estimation.

## VI. CONCLUSIONS

This paper dealt with the DOA estimation problem in acoustic environments with multiple microphones. We presented an approach based on the Householder transformation. This approach first transforms the noisy speech signals received at the array into the STFT domain. A Householder transformation is then constructed and applied to the multichannel STFT coefficients in each subband. This transformation, when aligned with the source incidence angle, converts the multichannel STFT coefficients into two components: one is a single coefficient that is dominated by the signal of interest and the other consists of the $M - 1$ coefficient that is are dominated by noise (or even consist of noise-only if there is no reverberation), where $M$ is the number of sensors. A DOA estimator was subsequently constructed to estimate the DOA information by searching the extremum value of this cost function in the angle range between 0° and 180°. A number of simulations were conducted to validate the performance of the developed algorithm in both noisy and reverberant environments. Results showed that the developed approach yielded a reasonably good DOA estimation and its performance as well as its robustness with respect to noise and reverberation increase with the number of microphones.

## ACKNOWLEDGMENT


This work was supported in part by the NSFC "Distinguished Young Scientists Fund" under Grant No. 61425005. The authors would like to thank the two anonymous reviewers for carefully reading the manuscript and providing many constructive comments and suggestions that have improved the clarity and quality of this paper.



Allen, J. B., and Berkley, D. A. (**1979**). "Image method for efficiently simulating small-room acoustics," J. Acoust. Soc. Am. **65**, 943–950.

Benesty, J. (**2000**). "Adaptive eigenvalue decomposition algorithm for passive acoustic source localization," J. Acoust. Soc. Am. **107**, 384–391.

Benesty, J., Chen, J., and Huang, Y. (**2004**). "Time-delay estimation via linear interpolation and cross correlation," IEEE Trans. Audio, Speech, Lang. Process. **12**, 509–519.

Benesty, J., Chen, J., and Huang, Y. (**2008**). *Microphone Array Signal Processing* (Springer-Verlag, Berlin, Germany), pp. 1–240.

Chen, J., Benesty, J., and Huang, Y. (**2003**). "Robust time delay estimation exploiting redundancy among multiple microphones," IEEE Trans. Speech, Audio Process. **11**, 549–557.

Chen, J., Benesty, J., and Huang, Y. (**2006**). "Time delay estimation in room acoustic environments: An overview," EURASIP J. Appl. Signal Process. **2006**, 1–19.

Chen, C., Lorenzelli, F., Hudson, R., and Yao, K. (**2008**). "Stochastic maximum-likelihood DOA estimation in the presence of unknown nonuniform noise," IEEE Trans. Signal Process. **56**, 3038–3044.

Chung, P. J., and Bohme, J. F. (**2002**). "DOA estimation using fast EM and SAGE algorithms," Signal Process. **82**, 1753–1762.

de Campos, M. L. R., Werner, S., and Apolinário, J. A., Jr. (**1999**). "Householder-transform constrained LMS algorithms with reduced-rank updating," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 1857–1860.

de Campos, M. L. R., Werner, S., and Apolinário, J. A., Jr. (**2002**). "Constrained adaptation algorithms employing Householder transformation," IEEE Trans. Signal Process. **50**, 2187–2195.

Do, H., Silverman, H. F., and Yu, Y. (**2007**). "A real-time SRP-PHAT source location implementation using stochastic region contraction (SRC) on a large-aperture microphone array," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 121–124.

Golub, G. H., and Loan, C. F. V. (**1996**). *Matrix Computations*, 3rd ed. (The Johns Hopkins University Press, Baltimore, MD), pp. 206–274.

Huang, Y., Benesty, J., and Chen, J. (**2006**). *Acoustic MIMO Signal Processing* (Springer, Berlin, Germany), pp. 215–259.

Huang, Y., Chen, J., and Benesty, J. (**2011**). "Immersive audio schemes: The evolution of multiparty teleconferencing," IEEE Signal Process. Mag. **28**, 28–32.

Hyder, M. M., and Mahata, K. (**2010**). "Direction-of-arrival estimation using a mixed norm approximation," IEEE Trans. Signal Process. **58**, 4646–4655.

Knapp, C. H., and Carter, G. C. (**1976**). "The generalized correlation method for estimation of time delay," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-24**, 320–327.

Li, M. H., and Lu, Y. L. (**2007**). "A refined genetic algorithm for accurate and reliable DOA estimation with a sensor array," Wireless Personal Commun. **43**, 533–547.

Lombard, A., Zheng, Y., Buchner, H., and Kellermann, W. (**2011**). "TDOA estimation for multiple sound sources in noisy and reverberant environments using broadband independent component analysis," IEEE Trans. Audio, Speech, Lang. Process. **19**, 1490–1503.

Ma, W. K., Hsieh, T. H., and Chi, C. Y. (**2010**). "Direction-of-arrival estimation of quasi-stationary signals with less sensors than sources and unknown spatial noise covariance: A Khatri-Rao subspace approach," IEEE Trans. Signal Process. **58**, 2168–2180.

Moghaddam, P. P., Amindavar, H., and Kirlin, R. L. (**2003**). "A new time-delay estimation in multipath," IEEE Trans. Signal Process. **51**, 1129–1142.

Nagata, Y., Fujioka, T., and Abe, M. (**2007**). "Two-Dimensional DOA estimation of sound sources based on weighted Wiener gain exploiting two-directional microphones," IEEE Trans. Speech, Audio Process. **15**, 416–429.

Nesta, F., and Omologo, M. (**2012**). "Generalized state coherence transform for multidimensional TDOA estimation of multiple sources," IEEE Trans. Audio, Speech, Lang. Process. **20**, 246–260.

Omologo, M., and Svaizer, P. (**1994**). "Acoustic event localization using a crosspower-spectrum phase based technique," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. II/273–II/276.

Pesavento, M., and Gershman, A. M. (**2001**). "Maximum-likelihood direction-of-arrival estimation in the presence of unknown nonuniform noise," IEEE Trans. Signal Process. **49**, 1310–1324.

Qian, C., Huang, L., and So, H. C. (**2014**). "Improved unitary root-MUSIC for DOA estimation based on pseudo-noise resampling," IEEE Signal Process. Lett. **21**, 140–145.

Reddy, V., Khong, A., and Poh Ng, B. (**2014**). "Unambiguous speech DOA estimation under spatial aliasing conditions," IEEE/ACM Trans. Audio, Speech, Lang. Process. **22**, 2133–2145.

Tan, Z., and Nehorai, A. (**2014**). "Sparse direction of arrival estimation using co-prime arrays with off-grid targets," IEEE Signal Process. Lett. **21**, 26–29.

Wang, H., and Chu, P. (**1997**). "Voice source localization for automatic camera pointing system in videoconferencing," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 187–190.

Ward, D. B., Ding, Z., and Kennedy, R. A. (**1988**). "Broadband DOA estimation using frequency invariant beamforming," IEEE Trans. Signal Process. **46**, 1463–1469.

Zhang, Y. D., Amin, M. G., and Himed, B. (**2013**). "Sparsity-based DOA estimation using co-prime arrays," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 3967–3971.