Adaptive eigenvalue decomposition algorithm for real time acoustic source localization system

Gary Elko

Adaptive eigenvalue decomposition algorithm for passive acoustic source localization

Jacob Benesty
Bell Laboratories, Lucent Technologies, 700 Mountain Avenue, Murray Hill, New Jersey 07974-0636

(Received 6 May 1998; accepted for publication 29 September 1999)

To find the position of an acoustic source in a room, the relative delay between two (or more) microphone signals for the direct sound must be determined. The generalized cross-correlation method is the most popular technique to do so and is well explained in a landmark paper by Knapp and Carter. In this paper, a new approach is proposed that is based on eigenvalue decomposition. Indeed, the eigenvector corresponding to the minimum eigenvalue of the covariance matrix of the microphone signals contains the impulse responses between the source and the microphone signals (and therefore all the information we need for time delay estimation). In experiments, the proposed algorithm performs well and is very accurate. © *2000 Acoustical Society of America.* [S0001-4966(00)03801-7]

PACS numbers: 43.60.Bf, 43.60.Gk, 43.60.Lq [JCB]

INTRODUCTION

Time delay estimation (TDE) between two (or more) microphone signals can be used as a means for the passive localization of the dominant talker in applications such as camera pointing for teleconferencing and microphone array beam steering for suppressing reverberation in all types of communication and voice processing systems. This problem is difficult because of the nonstationarity of speech and of room acoustic reverberation. Isotropic noise in the room can also be a problem if the signal-to-noise ratio becomes smaller than 20 dB, but that rarely occurs in teleconferencing rooms.

Generalized cross-correlation (GCC)[1] is the most commonly used method for TDE. In this technique, the delay estimate is obtained as the time-lag that maximizes the cross-correlation between filtered versions of the received signals. Several authors[2–4] have proposed techniques to improve the GCC in the presence of noise. However, it is believed that the main problem in practical TDE is reverberation as is clearly shown in Refs. 5 and 6. Accordingly, the same authors proposed an interesting method that is based on cepstral prefiltering. The basic idea is to first estimate the reverberation in the cepstral domain, subtract it from the microphone signals, and then use the GCC to determine the delay.[7,8]

Another interesting idea was proposed in Ref. 9 and was tested in a real time implementation in Ref. 10. The principle of this approach is to identify the onset of pitch periods and to cross-correlate these within a certain acceptance window. The problem of this algorithm is that it might take a very long time before it gives an accurate estimation of the time delay.

Most of the existing methods are based on an ideal signal model (single propagation path from the source to the sensors). In this paper, a new approach is proposed based on a real signal model (with reverberation) using eigenvalue decomposition. Indeed, the eigenvector corresponding to the minimum eigenvalue of the covariance matrix of the microphone signals contains the impulse responses between the source and the microphone signals (and therefore all the information we need for time delay estimation).

All of the above methods can be improved by using several pairs of microphones, e.g., by hyperbolic curve fitting of multiple delay estimates[11] or by linear 3-D intersection of multiple bearing lines.[12–14]

Section I discusses two models used for the TDE problem. In Sec. II, the GCC method is described and it is demonstrated how this method can fail. In Sec. III, the new proposed approach is developed which can also be seen as a generalization of the algorithm proposed in Ref. 15 where a single multipath is assumed. Section IV gives some experimental results and comparison of different algorithms.

I. MODELS FOR THE TDE PROBLEM

In this section, two models that are often used for the TDE problem are discussed. First, the ''ideal model'' is described and then the ''real model'' that more accurately describes a real acoustic environment.

A. Ideal model

A simple and widely used signal model for the classical TDE problem is the following. Let $x_i(n)$, $i=1,2$, denote the i-th microphone signal, then:

$$x_i(n)=\alpha_i s(n-\tau_i)+b_i(n), \tag{1}$$

where α_i is an attenuation factor due to propagation effects, τ_i is the propagation time from the unknown source $s(n)$ to microphone i, and $b_i(n)$ is an additive noise signal at the ith microphone. It is further assumed that $s(n)$, $b_1(n)$, and $b_2(n)$ are zero-mean, uncorrelated, stationary Gaussian random processes. The relative delay between the two microphone signals 1 and 2 is defined as

$$\tau_{12}=\tau_1-\tau_2. \tag{2}$$

This model is ideal in the sense that the solution for determining τ_{12} is clear. Indeed, let's first write Eq. (1) in the frequency domain

$$X_i(f)=\alpha_i S(f)e^{-j2\pi f\tau_i}+B_i(f), \tag{3}$$

and then take the (complex) sign of the cross-spectrum $S_{x_1x_2}(f)$ between $X_1(f)$ and $X_2(f)$,

$$\mathrm{sgn}[S_{x_1x_2}(f)]=\mathrm{sgn}[E\{X_1(f)X_2^\star(f)\}]=e^{-j2\pi f\tau_{12}}, \tag{4}$$

where $\mathrm{sgn}(z)=z/|z|$, $E\{\cdot\}$ denotes mathematical expectation, and * denotes complex conjugate. We can easily see that the inverse Fourier transform of Eq. (4) will result in a sharp peak in the time domain corresponding to the delay τ_{12}.

B. Real model

Unfortunately, in a real acoustic environment we must take into account the reverberation of the room and the ideal model no longer holds. Then, a more complicated but more complete model for the microphone signals $x_i(n), i=1,2$, can be expressed as follows:

$$x_i(n)=g_i * s(n)+b_i(n), \tag{5}$$

where $*$ denotes convolution and g_i is the acoustic impulse response between the source $s(n)$ and the ith microphone. Moreover, $b_1(n)$ and $b_2(n)$ might be correlated which is the case when the noise is directional, e.g., from a ceiling fan or an overhead projector.

In this case, we do not have an ''ideal'' solution to the problem, as is the case for the previous model, unless we can very accurately determine the two impulse responses between the source and the two microphones, which is a very challenging problem.

II. THE GCC METHOD

This method is based on the ideal model but is the most commonly used even in very reverberant environments.[14,16–18]

In the GCC technique,[1] the time-delay estimate is obtained as the value of τ that maximizes the generalized cross-correlation function given by

$$\psi_{x_1x_2}(\tau)=\int_{-\infty}^{+\infty}\Phi(f)S_{x_1x_2}(f)e^{j2\pi f\tau}df$$
$$=\int_{-\infty}^{+\infty}\Psi_{x_1x_2}(f)e^{j2\pi f\tau}df, \tag{6}$$

where $\Phi(f)$ is a weighting function and

$$\Psi_{x_1x_2}(f)=\Phi(f)S_{x_1x_2}(f) \tag{7}$$

is the generalized cross-spectrum. Then, the GCC TDE may be expressed as:

$$\hat{\tau}_\phi=\arg\max_\tau \psi_{x_1x_2}(\tau). \tag{8}$$

The choice of $\Phi(f)$ is important in practice. For example, the maximum likelihood (ML) weighting function, which is roughly equivalent to a frequency-dependent SNR, seems to be optimal for a high level of uncorrelated noise. However in practice, this algorithm performs best (within this class of algorithms) only for low SNR ($\leqslant$10 dB), as is shown in Ref. 2. Given that our problem is reverberation, the ML estimator will not help with reasonable SNR ($\geqslant$20 dB) and can even perform poorly.[2]

If we take $\Phi(f)=1$, we obtain the classical cross-correlation (CC) method. In the noiseless case, knowing that $X_i(f)=S(f)G_i(f)$, $i=1,2$, we have

$$\Psi_{x_1x_2}(f)=\Psi_{cc}(f)=G_1(f)E\{|S(f)|^2\}G_2^\star(f). \tag{9}$$

The classical phase transform (PHAT) technique is obtained by taking $\Phi(f)=1/|S_{x_1x_2}(f)|$. In the noiseless case, we show easily that

$$\Psi_{x_1x_2}(f)=\Psi_{pt}(f)=G_1(f)G_2^\star(f)/|G_1(f)G_2^\star(f)| \tag{10}$$

does depend only on the impulse responses and can perform well in a moderately reverberant room.

Another interesting weighting function is $\Phi(f)=1/\sqrt{S_{x_1x_1}(f)S_{x_2x_2}(f)}$. In this case, the corresponding normalized cross-spectrum $\Psi_{cf}(f)$ is simply the complex coherence function. We note that in the noiseless case, $\Psi_{cf}(f)=\Psi_{pt}(f)$, which is not true, in general, in the presence of noise.

The GCC methods can give good results when the reverberation of the room is not very high, but when the reverberation becomes important all of these techniques will fail because they are based on a simple signal model that does not represent reality.

III. THE PROPOSED METHOD

In this section a completely different approach than GCC is proposed. This new method focuses directly on the impulse responses between the source and the microphones in order to estimate the time-delay. First, the principle of this approach is explained and then an algorithm is presented.

A. Principle

We assume that the system (room) is linear and time invariant; therefore, we have the following relation:[19]

$$\mathbf{x}_1^T(n)\mathbf{g}_2=\mathbf{x}_2^T(n)\mathbf{g}_1, \tag{11}$$

where

$$\mathbf{x}_i(n)=[x_i(n) \quad x_i(n-1) \quad \cdots \quad x_i(n-M+1)]^T, \quad i=1,2$$

are vectors of signal samples at the microphone outputs, ''T'' denotes the transpose of a vector or a matrix, and the impulse response vectors of length M are defined as

$$\mathbf{g}_i=[g_{i,0} \quad g_{i,1} \quad \cdots \quad g_{i,M-1}]^T, \quad i=1,2.$$

This linear relation follows from the fact that $x_i=s*g_i$, $i=1,2$, thus $x_1*g_2=s*g_1*g_2=x_2*g_1$.

The covariance matrix of the two microphone signals is:

$$\mathbf{R}=\begin{bmatrix}\mathbf{R}_{x_1x_1} & \mathbf{R}_{x_1x_2}\\ \mathbf{R}_{x_2x_1} & \mathbf{R}_{x_2x_2}\end{bmatrix}, \tag{12}$$

where

$$\mathbf{R}_{x_ix_j}=E\{\mathbf{x}_i(n)\mathbf{x}_j^T(n)\}, \quad i,j=1,2. \tag{13}$$

Consider the $2M\times 1$ vector

$$\mathbf{u}=\begin{bmatrix}\mathbf{g}_2\\ -\mathbf{g}_1\end{bmatrix}.$$

From Eqs. (11) and (12), it can be seen that $\mathbf{Ru}=\mathbf{0}$, which means that the vector $\mathbf{u}$ (containing the two impulse responses) is the eigenvector of the covariance matrix $\mathbf{R}$ corresponding to the eigenvalue 0. Moreover, if the two impulse responses $\mathbf{g}_1$ and $\mathbf{g}_2$ have no common zeros and the autocorrelation matrix of the source signal $s(n)$ is full rank, which is assumed in the rest of this paper, the covariance matrix $\mathbf{R}$ has one and only one eigenvalue equal to 0.[20]

In practice, accurate estimation of the vector $\mathbf{u}$ is not trivial, because of the nature of speech, the length of the impulse responses, the background noise, etc. However, for this application we only need to find an efficient way to detect the direct paths of the two impulse responses. In the following, it is explained how this can be done.

B. Adaptive algorithm

In practice, it is simple to estimate iteratively the eigenvector (here $\mathbf{u}$) corresponding to the minimum (or maximum) eigenvalue of $\mathbf{R}$, by using an algorithm similar to the Frost algorithm which is a simple constrained Least-Mean-Square (LMS),[21] or by using the algorithms proposed in Ref. 22. In the following, we show how to apply these techniques to our problem. Minimizing the quantity $\mathbf{u}^T\mathbf{Ru}$ with respect to $\mathbf{u}$ and subject to $\|\mathbf{u}\|^2=\mathbf{u}^T\mathbf{u}=1$ will give us the optimum filter weights $\mathbf{u}_{\mathrm{opt}}$.

Let us define the error signal:

$$e(n)=\frac{\mathbf{u}^T(n)\mathbf{x}(n)}{\|\mathbf{u}(n)\|}, \tag{14}$$

where $\mathbf{x}(n)=[\mathbf{x}_1^T(n) \quad \mathbf{x}_2^T(n)]^T$. Note that minimizing the mean square value of $e(n)$ is equivalent to solving the above eigenvalue problem. Taking the gradient of $e(n)$ with respect to $\mathbf{u}(n)$ gives

$$\nabla e(n)=\frac{1}{\|\mathbf{u}(n)\|}\left[\mathbf{x}(n)-e(n)\frac{\mathbf{u}(n)}{\|\mathbf{u}(n)\|}\right], \tag{15}$$

and we obtain the gradient-descent constrained LMS algorithm:

$$\mathbf{u}(n+1)=\mathbf{u}(n)-\mu e(n)\nabla e(n), \tag{16}$$

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

Substituting Eqs. (14) and (15) into Eq. (16) gives

$$\mathbf{u}(n+1)=\mathbf{u}(n) - \frac{\mu}{\|\mathbf{u}(n)\|}\left[\mathbf{x}(n)\mathbf{x}^T(n)\frac{\mathbf{u}(n)}{\|\mathbf{u}(n)\|}-e^2(n)\frac{\mathbf{u}(n)}{\|\mathbf{u}(n)\|}\right] \tag{17}$$

and taking mathematical expectation after convergence, we get

$$\mathbf{R}\frac{\mathbf{u}(\infty)}{\|\mathbf{u}(\infty)\|}=E\{e^2(n)\}\frac{\mathbf{u}(\infty)}{\|\mathbf{u}(\infty)\|}, \tag{18}$$

which is what is desired: the eigenvector $\mathbf{u}(\infty)$ corresponding to the smallest eigenvalue $E\{e^2(n)\}$ of the covariance matrix $\mathbf{R}$.

In practice, it is advantageous to use the following adaptation scheme to avoid roundoff error propagation:[23]

$$\mathbf{u}(n+1)=\frac{\mathbf{u}(n)-\mu e(n)\nabla e(n)}{\|\mathbf{u}(n)-\mu e(n)\nabla e(n)\|}. \tag{19}$$

Note that if this trick is used, then $\|\mathbf{u}(n)\|$ [which appears in $e(n)$ and $\nabla e(n)$] can be removed, since we will always have $\|\mathbf{u}(n)\|=1$.

C. A simplified algorithm

The algorithm Eq. (19) presented above is a little bit complicated and is very general to find the eigenvector corresponding to the smallest eigenvalue of any matrix $\mathbf{R}$. If the smallest eigenvalue is equal to zero, which is the case here, the algorithm can be simplified as follows:

$$e(n)=\mathbf{u}^T(n)\mathbf{x}(n) \tag{20}$$

and

$$\mathbf{u}(n+1)=\frac{\mathbf{u}(n)-\mu e(n)\mathbf{x}(n)}{\|\mathbf{u}(n)-\mu e(n)\mathbf{x}(n)\|}. \tag{21}$$

Note that this algorithm can be seen as an approximation of the previous one by neglecting the terms in $e^2(n)$ [in Eq. (19)], which is reasonable (since the smallest eigenvalue is equal to zero). In this application, the two algorithms Eqs. (19) and (21) should have the same performance after convergence even with low SNRs. Moreover, in all experiments the unconstrained frequency-domain adaptive filter (UFLMS)[24] is used to implement the impulse response estimation algorithm. Note that this algorithm is still efficient from a complexity point of view but it requires seven FFT operations per block (because we need to go back to the time-domain to apply the norm constraint), while the PHAT requires only three FFT operations per block.

D. Discussion

The goal here is not to accurately estimate the two impulse responses g_1 and g_2 but rather the time-delay. For that, we need to detect the two direct paths. Thus initialization of the proposed algorithm is a key issue. During adaptation, the first half of vector $\mathbf{u}$ will contain an estimate of the impulse response g_2; initializing by 1 (at time $n=0$) a tap of $\mathbf{u}$ somewhere in the middle of the first half part (in order to take into account negative and positive relative delays), say $u_{M/2}(0)=1$, and adjusting the parameters of the adaptive algorithm (which is easy to do) in such a way that this (positive) peak will always be dominant (during adaptation) in comparison with the $M-1$ other taps of the first half of $\mathbf{u}$, $u_{M/2}(n)$ will be considered an estimate of the direct path of g_2. A ''mirror'' effect will appear in the second half of $\mathbf{u}$ (containing an estimation of the impulse response $-g_1$): a negative peak will dominate and that will be an estimation of the direct path of $-g_1$. Thus the relative delay will simply be the difference between the indices corresponding to these two

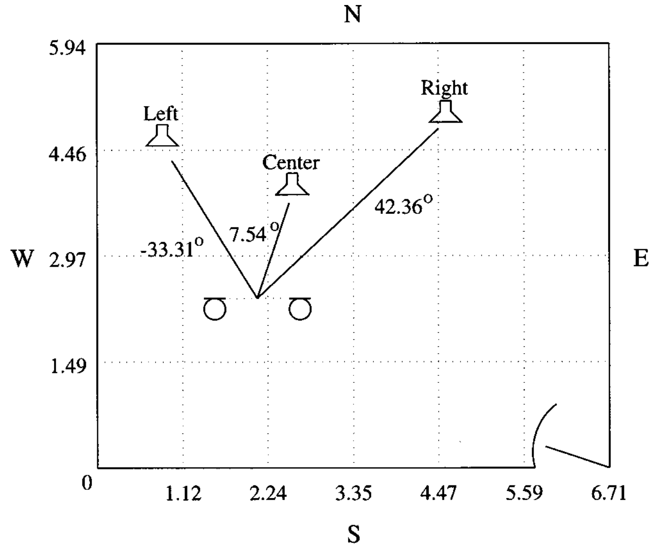

FIG. 1. Varechoic chamber floor plan (coordinate values measured in meters) with the position of the two microphones and the sources.

peaks. If reality corresponds to the ideal model, it is clear that the vector **u** will converge to the two peaks and the other taps will be close to zero.

The proposed algorithm can be seen as a generalization of the LMS TDE proposed in Refs. 25 and 26, where the minimization criterion of the direct paths as well as echoes are accounted for, whereas the LMS TDE considers only the two direct paths. Mathematically speaking, the LMS TDE is based on a criterion using the following error signal:

$$e(n)=x_1(n-D)-\mathbf{w}_2^T(n)\mathbf{x}_2(n), \tag{22}$$

whereas the proposed method is based on a criterion using the error signal

$$e(n)=\mathbf{w}_1^T(n)\mathbf{x}_1(n)-\mathbf{w}_2^T(n)\mathbf{x}_2(n). \tag{23}$$

The two algorithms will have exactly the same performance in an environment that corresponds to the ideal model. In practice, the LMS TDE performs like the CC method.

IV. EXPERIMENTAL RESULTS

Now, the performance of the proposed method is compared to PHAT, CC, and the Fischell–Coker (FC) algorithm. All the measurements were made in the Varechoic chamber at Bell Labs, which is a room with computer-controlled absorbing panels which vary the acoustic absorption of the walls, floor, and ceiling.[27] Each panel consists of two perforated sheets whose holes, if aligned, expose sound absorbing material behind, but if shifted to misalign, form a highly reflective surface. The panels are individually controlled so that the holes are either fully open (absorbing state) or fully closed (reflective state). Three different panel configurations were selected: panels all open, half of the panels open (ran-

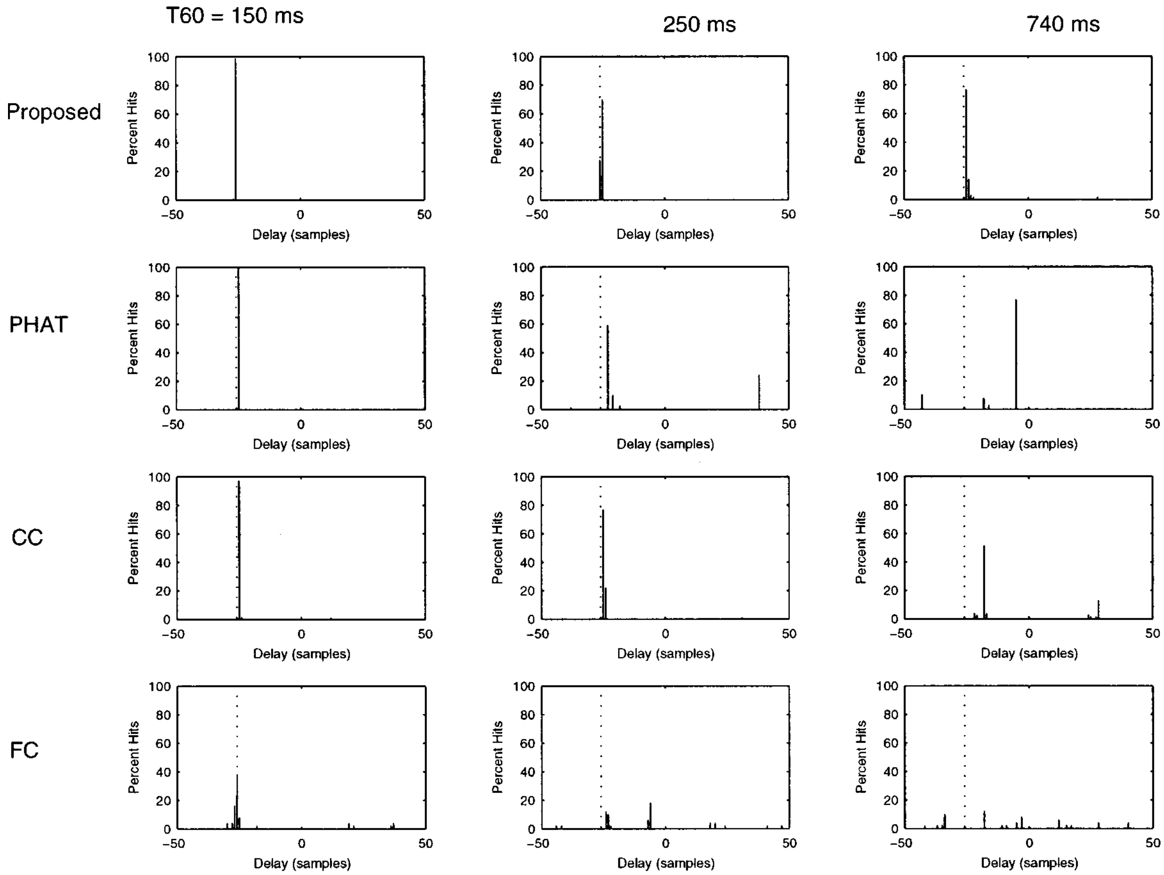

FIG. 2. Histograms of TDE with a pair of cardioid microphones. The source is a speech signal and its position is on the right in Fig. 1. The first, second, and third columns correspond, respectively, to a reverberation time of 150 ms, 250 ms, and 740 ms. The first, second, third, and fourth lines correspond, respectively, to the TDE by the proposed algorithm, PHAT, CC, and FC. The true delay is plotted with a dotted line.

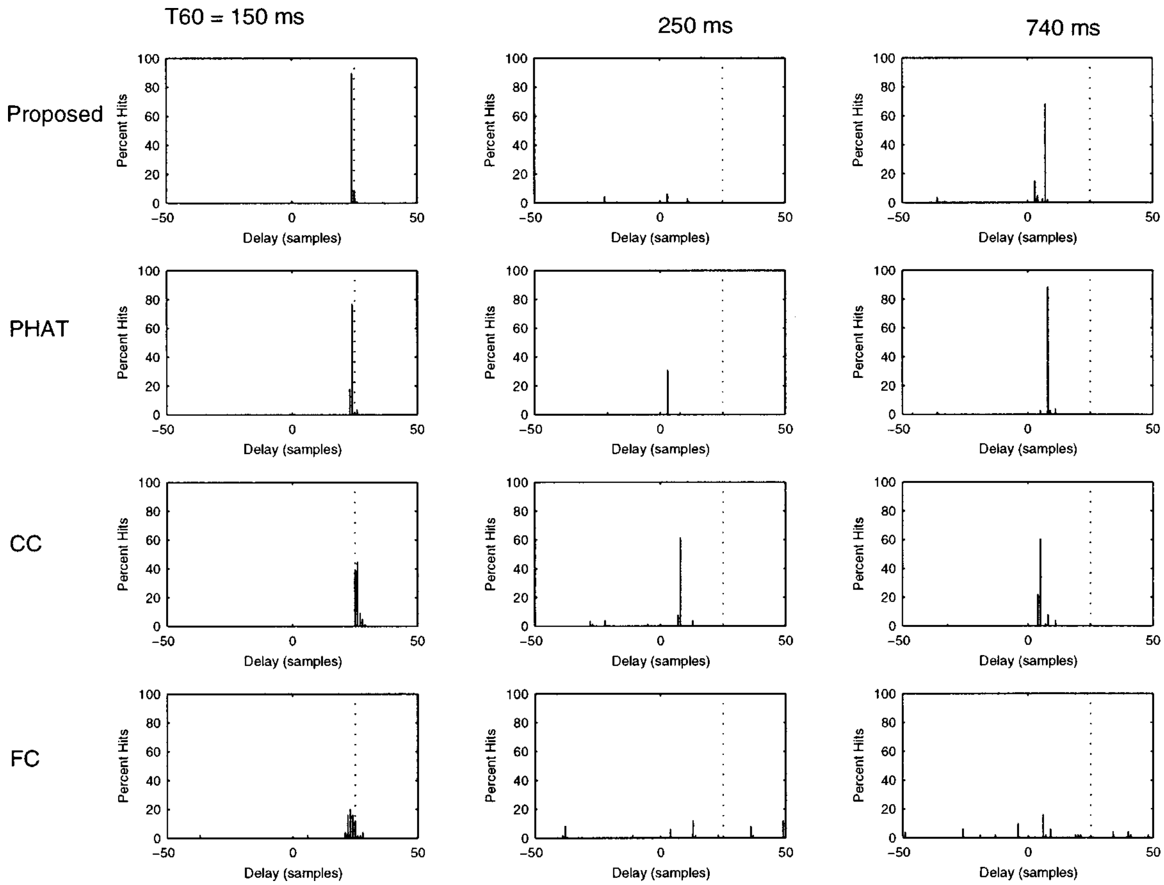

FIG. 3. Histograms of TDE with a pair of omni microphones. The source is a speech signal and its position is on the left. Presentation the same as in Fig. 2.

dom selection), and panels all closed; and the corresponding 60 dB reverberation time in the 400–1600 Hz band is, respectively, 150 ms, 250 ms, and 740 ms. Two different pairs of microphones were used: cardioid and omni. The distance between the two microphones was about 95 cm (37.25 in.). The source was simulated by placing a loudspeaker in four different positions (left, center, right, and far-right). Figure 1 shows a diagram of the floor plan layout with the position of the sources and the two microphones. A total of 14 different pairs of microphone signals were recorded, 12 with a speech source signal, and 2 with a white noise source signal. The original sampling rate was 48 kHz, but all the files were subsequently converted to a 16 kHz sampling rate. Each recording was about 5 s long.

For PHAT and CC, a 64 ms Kaiser window was used for the analysis frame. These two algorithms were well optimized in order to get the best results in terms of accuracy of TDE. For FC, a 100 ms window was used.[10] For the new method, we have used the UFLMS[24] with $\mu = 0.003$. The length of the adaptive vector u is taken as $L = 2M = 512$. The power of the two microphone signals $X_i(f,n), i = 1,2$, was estimated in the frequency-domain as follows:

$$P_i(f,n) = \gamma P_i(f,n-1) + (1-\gamma)|X_i(f,n)|^2, \tag{24}$$

with $\gamma = 0.999$. We initialized the algorithm with $P_i(f,0) = 2L\sigma_{x_i}^2$ ($\sigma_{x_i}^2$ is the average power of x_i).

It is not always easy to compare fairly different algorithms and the proposed choice of parameters of the previous algorithms may be argued, since in practice another set may be used for a better tracking. But this choice was done deliberately. We tried to find these parameters in order to have the best accuracy possible for all the algorithms. Note that this choice does not favor our algorithm. All of these simulations were made in a ''blind'' way, without using any information about the true delay. Moreover, we have used the same parameters for all the examples.

Figure 2 shows histograms of TDE with a pair of cardioid microphones. The source is a speech signal and its position is on the right. The first, second, and third columns correspond, respectively, to a reverberation time of 150 ms, 250 ms, and 740 ms. The first, second, third, and fourth lines correspond, respectively, to the TDE by the proposed algorithm, PHAT, CC, and FC. The true delay is plotted with a dotted line. It can be seen from this example that the new method performs better and is the most accurate. Notably with a 740 ms reverberation time, all methods fail except for the proposed algorithm.

Figures 3, 4, and 5 show histograms of TDE with a pair of omni directional microphones, which presents a more difficult problem. The source is again a speech signal and the presentation is the same as in Fig. 2. For Fig. 3, the source is on the left. All the methods fail for reverberation time of 250 ms and 740 ms. In Fig. 4, the source is in the center and here again the proposed algorithm seems to give better results. For Fig. 5, the source is on the right. It can be noticed that the new method fails completely only for a 250 ms reverberation time, whereas the other methods fail for 250 ms and 740 ms reverberation time.

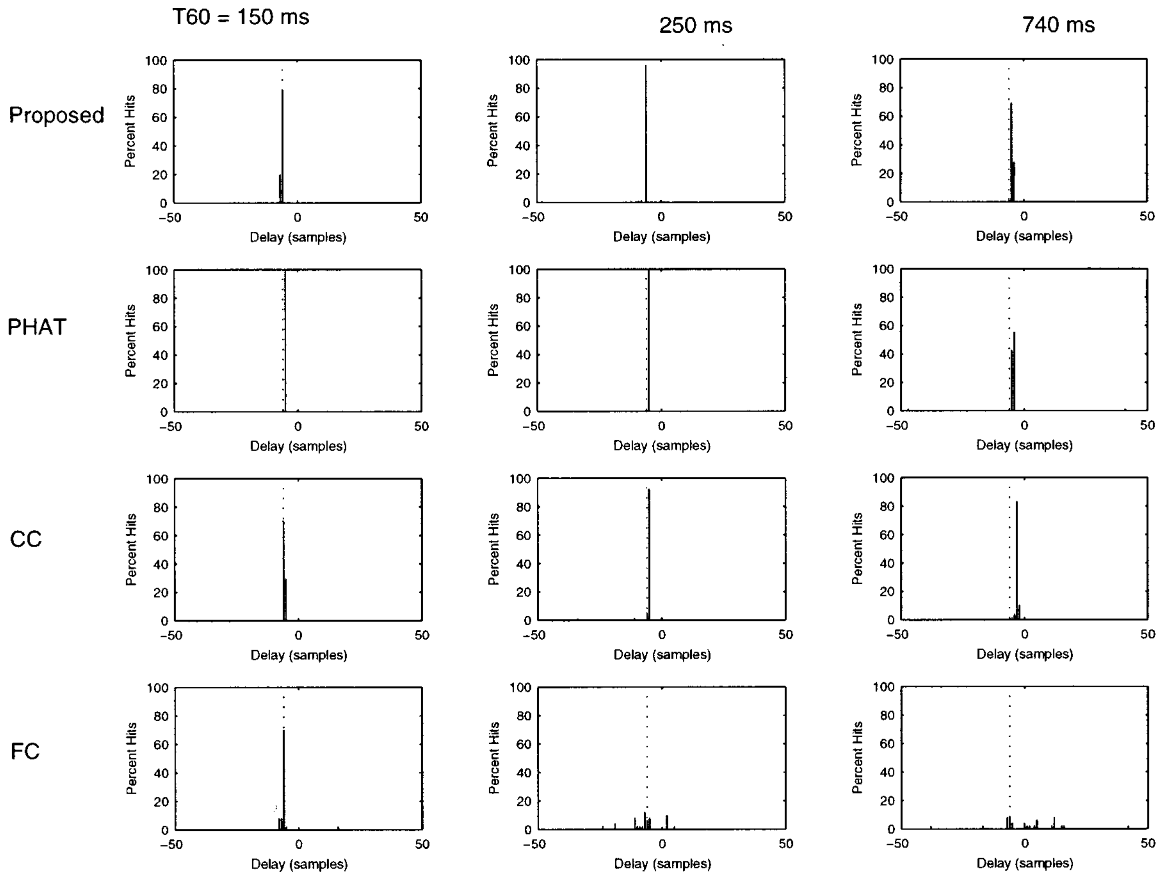

FIG. 4. Histograms of TDE with a pair of omni microphones. The source is a speech signal and its position is on the center. Presentation the same as in Fig. 2.

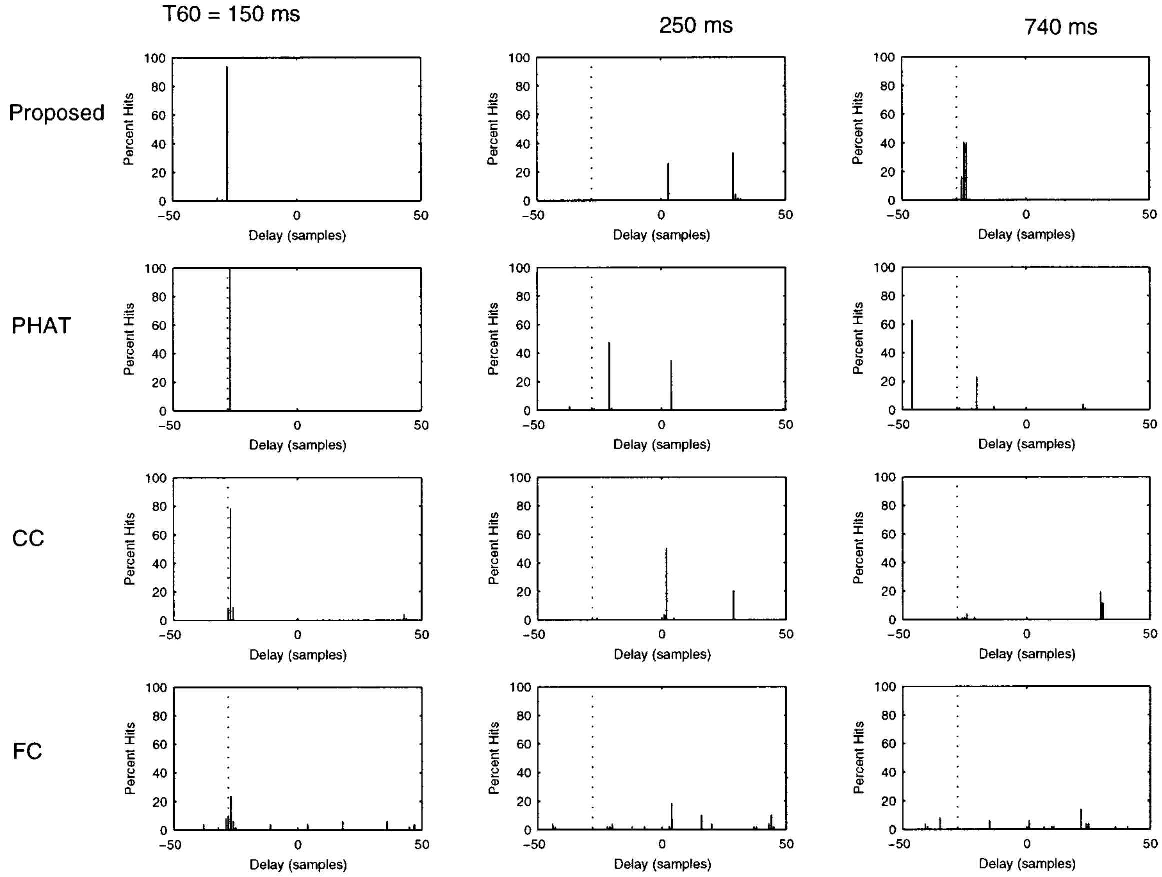

FIG. 5. Histograms of TDE with a pair of omni microphones. The source is a speech signal and its position is on the right. Presentation the same as in Fig. 2.

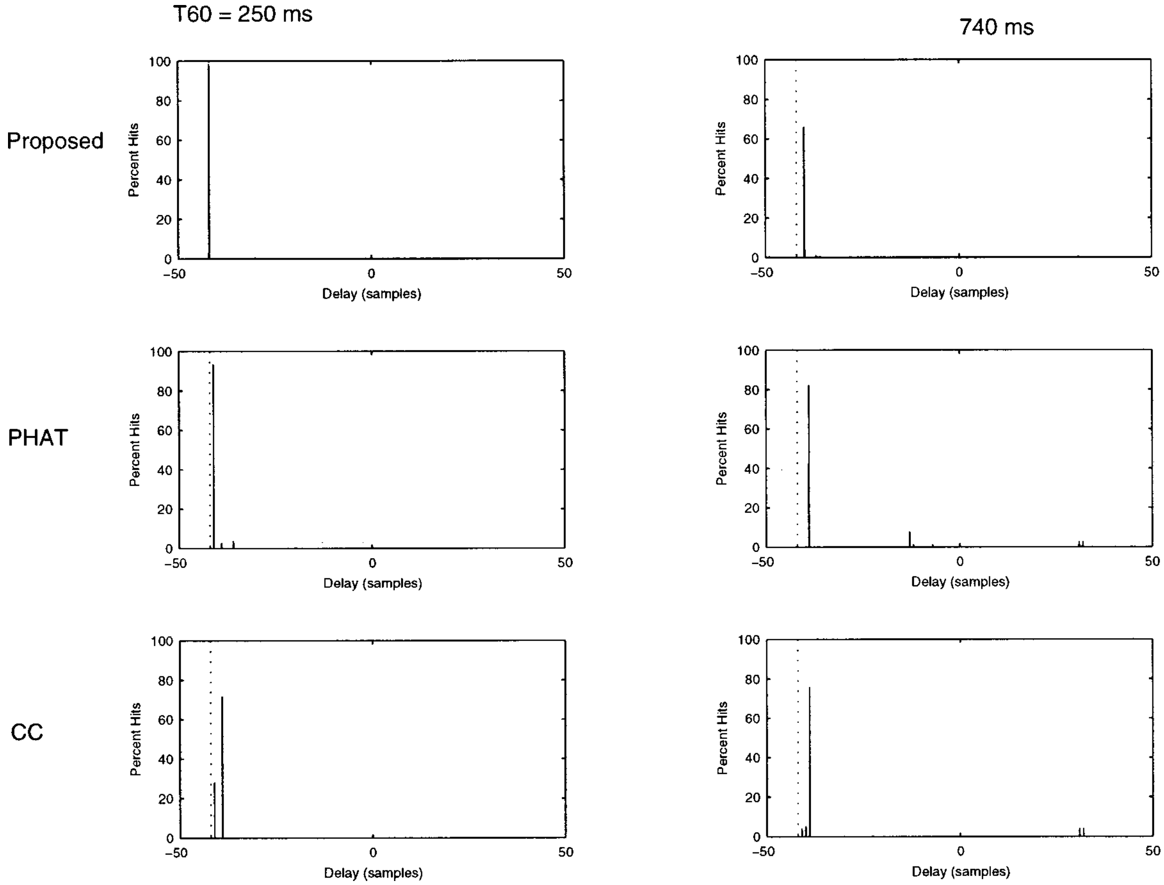

FIG. 6. Histograms of TDE with a pair of omni microphones. The source is a white noise signal and its position is on the far-right. The first and second columns correspond, respectively, to a reverberation time of 250 ms and 740 ms. The first, second, and third lines correspond, respectively, to the TDE by the proposed algorithm, PHAT, and CC. The true delay is plotted with a dotted line.

Figure 6 shows histograms of TDE with a pair of omni microphones, and a white noise source positioned on the far-right (not shown in Fig. 1). The first and second columns correspond, respectively, to a reverberation time of 250 ms and 740 ms. The first, second, and third lines correspond, respectively, to the TDE by the proposed algorithm, PHAT, and CC. Results for FC are not shown, since it is a pitch-based algorithm. Again, the true delay is plotted with a dotted line. All the algorithms are close to the solution but the new one is by far the most accurate.

The proposed algorithm converges very fast to a good time-delay estimate, it converges in less than 250 ms. Moreover, it is very robust to noise even with an SNR as low as 10 dB. Figure 7 compares the proposed algorithm to the PHAT, with a speech source on the right, a pair of cardiod microphones, a 250 ms reverberation time, and a white noise source in the center with an SNR equal to 10 dB.

V. CONCLUSION

In this paper, a new and simple approach to time-delay estimation has been proposed. The method consists of detecting the direct paths of the two impulse responses between the source signal and the microphones that are estimated in the eigenvector corresponding to the smallest eigenvalue of the covariance matrix of the microphone signals. In comparison with other methods, the proposed one seems to be more efficient in a reverberant environment and much more accurate.

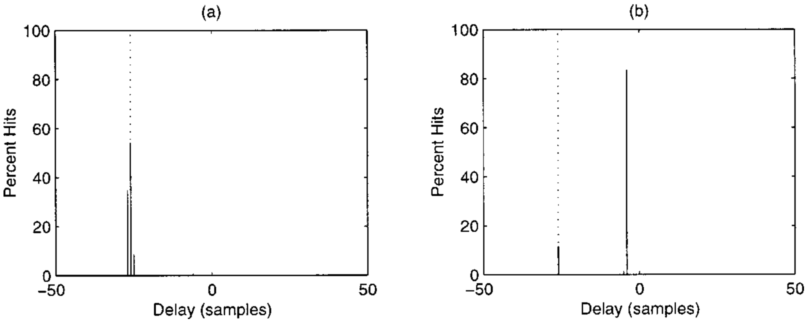

FIG. 7. Comparison of (a) the proposed algorithm to (b) the PHAT, with a speech source on the right, a pair of cardiod microphones, a 250 ms reverberation time, and a white noise source in the center with an SNR equal to 10 dB.

ACKNOWLEDGMENTS

I would like to thank Gary W. Elko and Jens Meyer for recording the data in the Varechoic chamber, and Dennis R. Morgan and again G. W. Elko for a careful reading of a preliminary draft and providing many useful comments for improvement. Also thanked are the anonymous reviewers for helpful suggestions.

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

[2] M. S. Brandstein, ''A pitch-based approach to time-delay estimation of reverberant speech,'' in Proceedings of the IEEE ASSP Workshop Applications on Signal Processing Audio Acoustics, New Paltz, NY, 1997.

[3] P. G. Georgiou, C. Kyriakakis, and P. Tsakalides, ''Robust time delay estimation for sound source localization in noisy environments,'' in Proceedings of the IEEE ASSP Workshop Applications on Signal Processing Audio Acoustics, New Paltz, NY, 1997.

[4] H. Wang and P. Chu, ''Voice source localization for automatic camera pointing system in video-conferencing,'' in Proceedings of the IEEE ASSP Workshop Applications on Signal Processing Audio Acoustics, New Paltz, NY, 1997.

[5] S. Bédard, B. Champagne, and A. Stéphenne, ''Effects of room reverberation on time-delay estimation performance,'' in Proceedings of the IEEE ICASSP, Adelaide, Australia, 1994, pp. II-261–II-264.

[6] B. Champagne, S. Bédard, and A. Stéphenne, ''Performance of time-delay estimation in the presence of room reverberation,'' IEEE Trans. Speech Audio Process. **4**, 148–152 (1996).

[7] A. Stéphenne and B. Champagne, ''Cepstral prefiltering for time delay estimation in reverberant environments,'' in Proceedings of IEEE ICASSP, Detroit, MI, 1995, pp. 3055–3058.

[8] A. Stéphenne and B. Champagne, ''A new cepstral prefiltering technique for time delay estimation under reverberant conditions,'' Signal Process. **59**, 253–266 (1997).

[9] D. R. Fischell and C. H. Coker, ''A speech direction finder,'' in Proceedings of IEEE ICASSP, 1984, pp. 19.8.1–19.8.4.

[10] D. R. Morgan, V. N. Parikh, and C. H. Coker, ''Automated evaluation of acoustic talker direction finder algorithms in the varechoic chamber,'' J. Acoust. Soc. Am. **102**, 2786–2792 (1997).

[11] H. F. Silverman and S. E. Kirtman, ''A two-stage algorithm for determining talker location from linear microphone array data,'' Comput. Speech Lang. **6**, 129–152 (1992).

[12] M. S. Brandstein, J. E. Adcock, and H. F. Silverman, ''A closed-form method for finding source locations from microphone-array time-delay estimates,'' in Proceedings of IEEE ICASSP, Detroit, MI, 1995, pp. 3019–3022.

[13] M. S. Brandstein, J. E. Adcock, and H. F. Silverman, ''A closed-form location estimator for use with room environment microphone arrays,'' IEEE Trans. Speech Audio Process. **5**, 45–50 (1997).

[14] D. V. Rabinkin, R. J. Renomeron, A. Dahl, J. C. French, and J. L. Flanagan, ''A DSP implementation of source location using microphone arrays,'' Proc. SPIE **2846**, 88–99 (1996).

[15] P. C. Ching, Y. T. Chan, and K. C. Ho, ''Constrained adaptation for time delay estimation with multipath propagation,'' IEE Proc. F, Radar Signal Process. **138**, 453–458 (1991).

[16] M. Omologo and P. Svaizer, ''Acoustic event localization using a crosspower-spectrum phase based technique,'' in Proceedings of IEEE ICASSP, Adelaide, Australia, 1994, pp. II-273–II-276.

[17] M. Omologo and P. Svaizer, ''Acoustic source location in noisy and reverberant environment using CSP analysis,'' in Proceedings of IEEE ICASSP, Atlanta, GA, 1996, pp. 921–924.

[18] D. V. Rabinkin, R. J. Renomeron, J. C. French, and J. L. Flanagan, ''Estimation of wavefront arrival delay using the cross-power spectrum phase technique,'' J. Acoust. Soc. Am. **104**, 2697(A).

[19] J. Benesty, F. Amand, A. Gilloire, and Y. Grenier, ''Adaptive filtering algorithms for stereophonic acoustic echo cancellation,'' in Proceedings of IEEE ICASSP, Detroit, MI, 1995, pp. 3099–3102.

[20] L. Tong, G. Xu, and T. Kailath, ''Fast blind equalization via antenna arrays,'' in Proceedings of IEEE ICASSP, Minneapolis, MN, 1993, Vol. IV, pp. 272–275.

[21] O. L. Frost III, ''An algorithm for linearly constrained adaptive array processing,'' Proc. IEEE **60**, 926–935 (1972).

[22] N. L. Owsley, ''Adaptive data orthogonalization,'' in Proceedings of IEEE ICASSP, 1978, pp. 109–112.

[23] M. Bellanger, *Analyse des signaux et filtrage numérique adaptatif* (Masson et CNET-ENST, Paris, 1989).

[24] D. Mansour and A. H. Gray, Jr., ''Unconstrained frequency-domain adaptive filter,'' IEEE Trans. Acoust., Speech, Signal Process. **ASSP-30**, 726–734 (1982).

[25] F. A. Reed, P. L. Feintuch, and N. J. Bershad, ''Time delay estimation using the LMS adaptive filter-static behavior,'' IEEE Trans. Acoust., Speech, Signal Process. **ASSP-29**, 561–571 (1981).

[26] D. H. Youn, N. Ahmed, and G. C. Carter, ''On using the LMS algorithm for time delay estimation,'' IEEE Trans. Acoust., Speech, Signal Process. **ASSP-30**, 798–801 (1982).

[27] W. C. Ward, G. W. Elko, R. A. Kubli, and W. C. McDougald, ''The new Varechoic chamber at AT&T Bell Labs,'' in Proceedings of the Wallace Clement Sabine Centennial Symposium, Acoustical Society of America, Woodbury, NY, 1994, pp. 343–346.