and

$$g(j) = c(j)\left[u\left(n, j+\frac{M}{2}\right) + u\left(n, \frac{M}{2}-j\right)\right], \quad \text{for } 0 \le j \le \frac{M}{2}, \tag{18}$$

$$g(j) = c(j)\left[u\left(n, j+\frac{M}{2}\right) - u\left(n, \frac{5M}{2}-j\right)\right], \quad \text{for } \left(\frac{M}{2}+1\right) \le j \le (M-1). \tag{19}$$

Since (16) is an $M$-point DCT-III [5], any of the famous fast DCT algorithms [5]–[7] can be used to reduce the computational complexity to $O(\frac{M}{2}\log_2 M)$. Here, we can use the mixed DIF and DIT fast DCT method to reduce further the number of multiplications.

After performing the first DIT procedure [7] to (16), we obtain

$$G(i) = \sum_{j=0}^{M/2-1} c(2j)g(2j)\cos\left(\frac{(2i+1)j\pi}{M}\right) + Z(i) \tag{20}$$

where

$$Z(i) = \sum_{j=0}^{M/2-1} c(2j+1)g(2j+1) \times \cos\left(\frac{(2i+1)(2j+1)\pi}{2M}\right). \tag{21}$$

After the above decomposition, the even-index summations stated in (20) can be repeatedly decomposed by the same DIT procedure. The odd-index summation in a recursive formula can be expressed by [6]

$$Z(i) = \sum_{j=0}^{M/2-1} 2c(2j+1)g(2j+1)\cos\left(\frac{(2j+1)\pi}{2M}\right) \times \cos\left(\frac{(2j+1)i\pi}{M}\right) - Z(i-1). \tag{22}$$

If we treat

$$z(j) = 2c(2j+1)g(2j+1)\cos\left(\frac{(2j+1)\pi}{2M}\right) \tag{23}$$

as a new input data, the summation in (22) can be achieved by using the DIF fast DCT algorithm [6]. It is noted that the multiplication of $2\cos(\frac{(2j+1)\pi}{2M})$ to $g(2j+1)$ in (23) depends on the indexes of polyphase filters. Hence, according to (18) and (19), we can merge the factors $2\cos(\frac{(2j+1)\pi}{2M})$ into the corresponding coefficients of polyphase filters, as Fig. 2 shows. Then, we can save those $\frac{M}{2}$ multiplications in the odd-index recursive formula (22) of the first DIT process. Therefore, by merging all possible constant factors with the polyphase filter coefficients, we can eliminate

$$\frac{M}{2} + \frac{M}{4} + \cdots + 4 + 2 = M - 2 \tag{24}$$

multiplications from the fast $M$-point DCT algorithms. Applying the polyphase recursive algorithm and the mixed DCT procedures, we obtain the structure shown in Fig. 2 for cosine-modulated subband filtering in the encoder. To obtain $M$ decimated-outputs from $N$ input-samples, we require only $(\frac{N}{2} + \frac{M}{2}\log_2 M + 2)$ multiplications. Table I shows the computations required for filtering 512 inputs to obtain 32 desired subband outputs. We found that the proposed fast algorithms require much less computation than the ISO suggestion [3].

## V. Conclusion

In this paper, we propose two fast algorithms for the implementation of multirate subband filtering. Using the recursive formula and the mixed fast DCT algorithm, we can reduce the computation of the cosine-modulated subband filtering in the MPEG audio encoder to about 13.2% of multiplications and 41.6% of additions. Since the proposed algorithms possess highly regular configurations, we believe that the recursive formula combined with the mixed DCT algorithm should be a better candidate for VLSI implementation of the MPEG audio coder.

TABLE I
Computations Required for Filtering 512 Inputs to Obtain 32 Desired Outputs

| Algorithms | Multiplications | Additions |
|---|---|---|
| MPEG Proposed Method | 2560 | 2464 |
| Recursive Algorithm + Mixed DCT Method | 338 | 1042 |

## References


[1] H. J. Nussbaumer and M. Vetterli, "Computationally efficient QMF filter banks," in *Proc. ICASSP*, 1984, pp. 11.3.1–11.3.4.

[2] P. P. Vaidyanathan, *Multirate Systems and Filter Banks*. Englewood Cliffs, NJ: Prentice-Hall, 1993, pp. 373–391.

[3] ISO. MPEG Std. ISO-IEC JTCI/SC2/WG11, MPEG 91.

[4] G. Theile, G. Stoll, and M. Link, "Low bit-rate coding of high-quality audio signals—An introduction to the MUSICAM system," *EBU Rev.-Tech.*, no. 230, pp. 158–181, Aug. 1988.

[5] K. R. Rao and P. Yip, *Discrete Cosine Transform: Algorithms, Advantages, Applications*. New York: Academic, 1990, pp. 15–16.

[6] S. C. Chan and K. L. Ho, "A new two-dimensional fast cosine transform algorithm," *IEEE Trans. Signal Processing*, vol. 39, no. 2, pp. 481–485, Feb. 1991.

[7] B. G. Lee, "A new algorithm to compute the discrete cosine transform," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-32, no. 6, pp. 1243–1245, Dec. 1984.

[8] R. E. Blahut, *Fast Algorithm for Digital Signal Processing*. Reading, MA: Addison-Wesley, 1985.

[9] S.-G. Chen and R.-J. Tsai, "A novel fast FIR filtering algorithm and its implementation," in *Proc. EUSIPCO*, 1992, pp. 945–948.


# Performance of Time-Delay Estimation in the Presence of Room Reverberation

Benoît Champagne, Stéphane Bédard, and Alex Stéphenne

***Abstract*— In this correspondence, synthetic microphone signals generated with the image model technique are used to study the effects of room reverberation on the performance of the maximum likelihood (ML) estimator of the time delay, in which the estimate is obtained by maximizing the cross correlation between filtered versions of the microphone signals. The results underscore the adverse effects of reverberation on the bias, variance and probability of anomaly of the ML estimator. Explanations of these effects are provided.**

Manuscript received October 14, 1994; revised October 2, 1995. This work was supported by NSERC under Grant OGP0105533 and by FCAR under Grant 93-ER-1577. Part of this work has been presented at the *IEEE 1994 Int. Conf. on Acoust., Speech, Signal Processing*, Adelaide, Australia, April 1994. The associate editor coordinating the review of this paper and approving it for publication was Dr. James H. Snyder.

B. Champagne and A. Stéphenneare with INRS-Télécommunications, Université du Québec, Verdun, Québec, Canada H3E 1H6

S. Bédard is with Bell SYGMA Telecom Solutions, Montréal, Québec, Canada H3B 4L1.

Publisher Item Identifier S 1063-6676(96)02442-X.

## I. Introduction

Microphone arrays, which have the ability to respond to a signal originating from a desired look-direction while discriminating against noises from other directions, are being used increasingly for speech signal transduction under reverberant and noisy conditions [1]. In many applications, a fundamental difficulty remains the localization of the dominant talker in real-time so as to continuously steer the array in his/her direction. To solve this problem, several passive localization techniques based on time-delay estimation (TDE) between the direct path signals received by pairs of microphones have been proposed recently [2]-[3]. These techniques typically rely on the use of generalized cross correlation (GCC) methods, in which the delay estimate is obtained as the time lag which maximizes the cross correlation between filtered versions of the received signals [4].

GCC methods are very popular in signal processing applications because of their accuracy and moderate computational requirements. Their statistical performance has been extensively studied under the assumption of single-path propagation (i.e., no reverberation) and is generally well understood [4]-[6]. In this respect, one member of the GCC family known as the maximum likelihood (ML) estimator is of particular interest in both practical applications and theoretical studies for it achieves the fundamental performance limits predicted in these works. Considerably less is known about the use of GCC methods in reverberant environments, such as a small office or a teleconference room. In the past, some studies have investigated the TDE problem in the presence of a few correlated additive echoes (e.g., [7]). However, the results obtained cannot be used to predict the effects of reverberation on TDE performance since reverberation consists in the superposition of a very large number of closely spaced echoes, a phenomenon that is more adequately modeled as multiplicative noise in the frequency domain (i.e., convolutional smearing).

In this correspondence, the effects of room reverberation on the performance of GCC methods for TDE are investigated under controlled conditions via Monte Carlo simulations. Emphasis is given to the maximum likelihood (ML) estimator because of its optimality within the GCC family. In the study, synthetic microphone signals corresponding to different levels of reverberation in a rectangular room are generated with the image model technique [8]. These signals are used to study the bias, variance and probability of anomaly of the ML estimator as a function of the reverberation time of the room and other physical parameters of interest. The results clearly demonstrate the adverse effects of reverberation on MLTDE performance and also reveal the existence of a threshold phenomenon similar to that identified in [9] for certain array-based direction finding methods. Qualitative and quantitative explanations of these effects are provided.

## II. Review of MLTDE for Single-Path Propagation

A widely used signal model for the TDE problem is given by

$$\begin{aligned} x_1(t) &= s(t) + n_1(t) \qquad 0 \le t \le T \\ x_2(t) &= s(t+\tau) + n_2(t) \end{aligned} \tag{1}$$

where $x_i(t)$, $i \in \{1,2\}$, denotes the output signal of the $i$th receiver (e.g., microphone), $s(t)$ is the desired source signal, $\tau$ is a free parameter representing the unknown delay, $n_i(t)$ is the additive noise at the $i$th receiver and $T$ denotes the duration of the observation interval. Furthermore, it is assumed that $s(t)$, $n_1(t)$, and $n_2(t)$ are (real) zero-mean, uncorrelated, stationary Gaussian random processes. This model corresponds to an ideal situation in which the signal propagation from the source to each receiver occurs along a single direct path, without attenuation (e.g., plane waves), in a nondispersive medium. In a typical application, the unknown delay $\tau$ is equal to the difference in travel time of a wavefront propagating from the source to the two receivers. Hence, multiple TDE's from distinct microphone pairs can be used to locate the source via triangulation techniques.

One of the most popular TDE methods for single-path propagation models is that of maximum likelihood. By definition, the ML estimator of the time delay, denoted $\hat{\tau}_{ML}$, is the value of $\tau$ in (1) which maximizes the likelihood function of the observed data, consisting here of the signals $x_i(t)$ for $i = 1,2$, and $0 \le t \le T$. Using (1) together with the statistical assumptions made on the signal and noises, the following expressions can be derived for $\hat{\tau}_{ML}$ [4]:

$$\hat{\tau}_{ML} = \arg\max_{\tau \in \mathcal{D}} R_{ML}(\tau) \tag{2}$$

$$R_{ML}(\tau) = \int_{-\infty}^{\infty} \psi_{ML}(f) X_1(f) X_2^*(f) e^{j2\pi f\tau} df \tag{3}$$

$$\psi_{ML}(f) = \frac{S(f)}{N_1(f)N_2(f)} \{1 + \frac{S(f)}{N_1(f)} + \frac{S(f)}{N_2(f)}\}^{-1} \tag{4}$$

where $\mathcal{D}$ is the set of *a priori* delay values, $X_i(f)$ is the Fourier transform of $x_i(t)$ over the interval $0 \le t \le T$, the superscript $^*$ denotes the complex conjugate and $S(f)$, $N_1(f)$, and $N_2(f)$ are the power spectral densities of $s(t)$, $n_1(t)$, and $n_2(t)$, respectively. The choice of the set $\mathcal{D}$ used for the search in (2) is generally based on physical considerations.

The function $R_{ML}(\tau)$ (3) is a particular form of the GCC, for it can be expressed as the cross correlation between filtered versions of the observed signals $x_1(t)$ and $x_2(t)$, with filter transfer functions given by $H_1(f) = H_2(f) = \sqrt{\psi_{ML}(f)}$, respectively. The effect of the frequency weighting $\psi_{ML}(f)$ in (4), or equivalently the filters $H_i(f)$ above, is to improve the accuracy of the delay estimate by attenuating signal components fed to the cross correlator in spectral regions where the signal-to-noise ratio is the lowest. In practice, the power spectral densities $S(f)$ and $N_i(f)$ in (4) are unknown and must be estimated from the data, a procedure sometimes referred to as approximate MLTDE [4].

For the single-path model (1) with uncorrelated, stationary Gaussian signal and noises, MLTDE is asymptotically unbiased and efficient in the limit of long observation intervals. That is, in the limit $T \to \infty$, the expected value of the estimation error converges to zero while the ratio of the estimator variance to the Cramer–Rao lower bound (CRLB) converges to one. Here, the CRLB is given by [4]

$$\sigma_{cr}^2 = \{8\pi^2 T \int_0^{\infty} \psi_{ML}(f) S(f) f^2 df\}^{-1} \tag{5}$$

and is a function of the observation interval, the processing bandwidth of the system (subsequently denoted by $W$) and the in-band signal-to-noise ratio (SNR).

Although a mathematical analysis of MLTDE performance is quite involved for finite $T$, the overall behavior can be characterized in terms of a few fundamental parameters [5]-[6]. For instance, in the case of baseband signals with large time-bandwidth product (i.e., $WT \gg 1$), which is of particular interest here, the SNR domain can be partitioned into three disjoint regions, with distinct behavior of the ML estimator in each region. At high SNR, $\hat{\tau}_{ML}$ is unbiased and efficient. Below a so-called threshold SNR, denoted $\mathrm{SNR}_{th}$, the variance of the estimator suddenly departs from the CRLB. This behavior is due to a rapid increase in the probability of making a large error, or an anomalous estimate, which results from selecting the wrong peak of $R_{ML}(\tau)$ (4) owing to the presence of noise. At low SNR, the estimation process is entirely dominated by the noise and $\hat{\tau}_{ML}$ is uniformly distributed in the *a priori* interval $\mathcal{D}$.

## III. Simulation Experiments

The transmission of an acoustic signal between a source and two microphones in a room surrounded by reflective boundaries (walls, ceiling and floor) is not accurately modeled by (1). In this case, due to multiple reflections of the sound waves on the boundaries, several delayed and attenuated replicas of the source signal (i.e., echos) are received by each microphone in addition to the direct path signal, a perceivable phenomenon known as reverberation. While the received echos can formally be included in the noise terms $n_i(t)$ in (1), this would violate the assumption of uncorrelated signal and noises which is made in the derivation and performance analysis of the classical MLTDE method. Hence, without further knowledge, the results reported at the end of Section II cannot be used to predict the performance of MLTDE between microphone signals contaminated by room reverberation. In this section, we describe the methodology of a Monte-Carlo computer simulation experiment designed to study the effects of room reverberation on MLTDE performance under controlled conditions.

For the simulations, we consider a rectangular room with plane reflective boundaries. Each boundary is characterized by a uniform reflection coefficient, say $\beta_j (j \in \{1, ..., 6\})$ with $0 \leq \beta_j \leq 1$, which is independent of the frequency and the angle of incidence of the acoustic rays. Points in the room are referenced with a rectangular coordinate system $Oxyz$ with origin in one corner of the room and axes parallel to the walls (the positive $z$-axis represents the upward vertical direction). The dimensions of the room along the $x$, $y$ and $z$-axes are denoted by $L_x$, $L_y$, and $L_z$, respectively. In this room, an omnidirectional point acoustic source radiates an audio signal which is monitored by two ideal point receivers (microphones) whose directivity patterns need not be omnidirectional. The position vectors of the source and the microphones, which remain fixed during a given simulation, are denoted by $\mathbf{r}_s$, $\mathbf{r}_{m,1}$, and $\mathbf{r}_{m,2}$, respectively. Except for the presence of the source and the microphones, the room is assumed to be empty.

Under the assumption of a linear and time-invariant acoustic medium, the microphone output signals, $x_i(t)$, can be expressed in the form

$$\begin{aligned} x_1(t) &= [h_1 * s](t) + n_1(t), \qquad 0 \leq t \leq T \\ x_2(t) &= [h_2 * s](t) + n_2(t) \end{aligned} \tag{6}$$

where $s(t)$ is the source signal, $*$ denotes the convolution operation, $h_i(t)$ is the impulse response of the acoustic transmission channel between the source and the $i$th microphone and $n_i(t)$ is an additive noise component. Here, $h_i(t)$ completely characterizes the reverberation effects in the room, while $n_i(t)$ is used to model external (i.e., uncorrelated) interferences other than reverberation. Thus, the same statistical assumptions as in Section II are made on $s(t)$ and $n_i(t)$. Since we are concerned with digital implementations of the MLTDE method, it is further assumed that the microphone output signals $x_i(t)$ in (6) are passed through identical low-pass filters with a common cutoff frequency $f_{\max} = 5000$ Hz and then sampled synchronously at the Nyquist rate, i.e., $f_s = 2f_{\max} = 10$ kHz (the sampling interval is $T_s = 1/f_s = 10^{-4}s$).

To generate the low-pass sampled versions of the acoustic impulse responses $h_i(t)$ in (6), an improved version of Allen and Berkley's implementation of the image model technique [8] is used. This version incorporates Peterson's modification [10] along with new features that allow for the simulation of microphones with arbitrary, although frequency independent, directivity patterns. In practice, directional microphones can be used to attenuate a significant portion of the unwanted echos. In this work, the following cardioid pattern is used for both microphones: $f(\theta) = (1 + \cos\theta)/2$, where $\theta$ is the angle of arrival of an incident ray, as measured from the direction of maximum response. Although $h_i(t)$ in (6) theoretically extends to infinity, the sampled impulse responses used in the simulations are truncated to about 6000 samples $(0.6s)$. Even for the worst-case scenario considered (i.e., strongest reverberation), the truncated tail of the response is approximately 50 dB below the main peak corresponding to the direct path signal.

The sampled version of the source signal $s(t)$ in (6) is obtained by passing a Gaussian white noise sequence with zero-mean and unit variance through a band-pass linear-phase FIR filter (Park–McClellan, 132 taps) with unitary magnitude response and lower and upper cut-off frequencies (3 dB) denoted by $f_l$ and $f_u$, respectively. The resulting source signal is then convolved with a pair of synthetic room impulse responses. Finally, independent Gaussian white noise, properly scaled to obtain the desired value of in-band SNR in the absence of reverberation, is added to each channel.

After discarding the initial transients, the microphone signals are partitioned into contiguous frames of $K = 2048$ samples for processing (integration time $T = 204.8$ ms). For each frame, a time-delay estimate is obtained via a digital (software) implementation of the ML equations (2)–(4). Based on *a priori* knowledge of the signal and noise spectra, $\psi_{ML}(f)$ (4) is set to one for $f_l \leq f \leq f_u$ (signal passband) and to zero otherwise. A sampled version of $R_{ML}(\tau)$ (3), corresponding to the lag values $\tau = kT_s$, $|k| = 0, 1, ..., K/2$, is calculated using conventional processing techniques based on the fast Fourier transform (FFT). To obtain the ML estimate of the time delay, $R_{ML}(kT_s)$ is first maximized over an *a priori* search grid defined by $|k| \leq k_{\max}$. Here, $k_{\max}$ is the smallest integer not less than $d_{12}/(cT_s)$, where $d_{12}$ is the distance between the microphones and $c \approx 340$ m/s is the speed of sound in air. Finally, quadratic interpolation is used to refine the estimate so obtained [5].

For each set of simulation parameters, 500 consecutive frames are processed, resulting in 500 independent time-delay estimates. Following [5], estimates for which the absolute error exceeds $T_c/2$, where $T_c$ is the signal correlation time, are identified as anomalies. Here, $T_c$ is defined as the width of the main lobe of the signal autocorrelation function (taken between the -3 dB points). In practice, anomalies can be detected by using a tracking algorithm, provided their probability of occurrence is sufficiently low. Finally, the 500 time-delay estimates are used to calculate the following statistical performance measures: the percentage of anomalous estimates and the sample bias and standard deviation of the nonanomalous estimates.

## IV. Results and Interpretation

A first series of experiments was conducted using the following parameter values:

- Room dimension: $L_x = 10.0$ m, $L_y = 6.6$ m, and $L_z = 3.0$ m.
- Reflection coefficients: $\beta_j \equiv \beta$ varying between 0 and 1.
- Source position: $\mathbf{r}_s = (2.4835, 2.0, 1.8)$ m.
- Microphone positions: $\mathbf{r}_{m,1} = (6.5, 2.8, 1.8)$ m, $\mathbf{r}_{m,2} = (6.5, 3.8, 1.8)$ m.
- Microphone orientation: (–1, 0, 0) (pointing toward negative x-axis)
- True delay between direct path signals: $= -9T_s$.
- Signal bandwidth: $f_l = 450$ Hz, $f_u = 3375$ Hz ($\sim$ telephone transmission bandwidth).
- In-band SNR: 30 dB.

The results obtained are presented in Figs. 1–3, where the percentage of anomalous time-delay estimates and the sample bias and standard deviation of the nonanomalous estimates, respectively, are plotted as a function of Eyring's reverberation time. The latter is given here by [11]

$$T_R = -13.82/[c(L_x^{-1} + L_y^{-1} + L_z^{-1}) \ln \beta]. \tag{7}$$

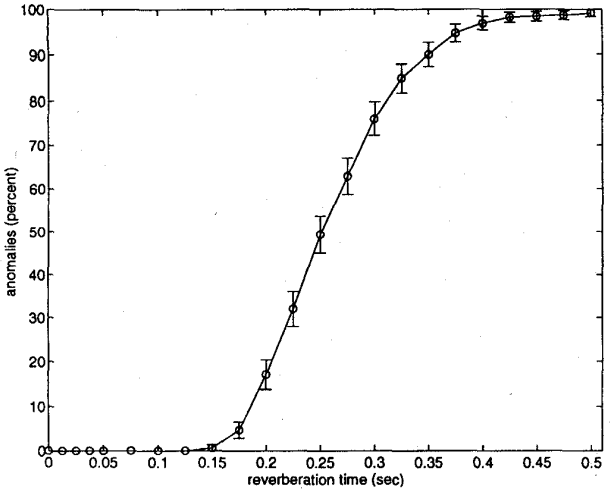


Fig. 1. Percentage of anomalous time-delay estimates versus reverberation time $T_R$.

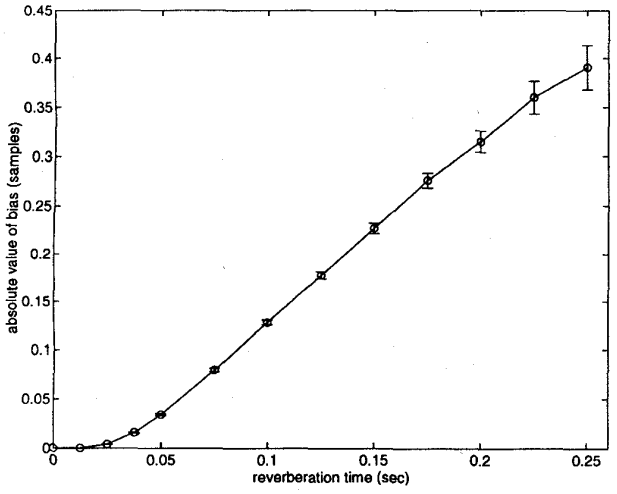


Fig. 2. Bias of nonanomalous time-delay estimates versus reverberation time $T_R$.

The vertical bar superimposed on each data point in the figures represents the 95% confidence interval for that measurement.

In Fig. 1, a sudden increase in the percentage of anomalies occurs around $T_R = 0.15$ s (i.e., $\beta = 0.6$). This behavior, similar to the threshold effect observed in the single-path scenario when SNR $\approx \mathrm{SNR}_{th}$, is due to the presence of erroneous peaks in the ML cross correlator output (3). In this case, however, the peaks are not caused by the background noise; they result from the correlation existing between echoes received on different channels. As $T_R$ increases and the echoes become stronger, the number and the amplitudes of these erroneous peaks increase, eventually making the ML estimator totally unreliable. Setting the acceptable level of anomalous estimates to 10%, Fig. 1 indicates that the MLTDE method cannot be used reliably when $T_R > 0.18$ s. Such values of $T_R$ are not uncommon in teleconference applications.

Fig. 2 shows a deterioration in the absolute bias as $T_R$ increases from 0 to 0.25 s. According to our experience, this deterioration is due to the presence of strong (initial) echos whose time differences of arrival at the two microphones are close to that of the direct path signals. The net effect of these echos is an apparent modification in the shape of the ML cross correlator output for lag values close to the true delay. This behavior is very difficult to predict analytically, for the precise value of the bias depends on the shape of the signal autocorrelation function and the specific delay structure of the received echoes, which in turns depends in a very sensitive way on the problem geometry (source and microphone positions, room shape and dimensions, etc.).

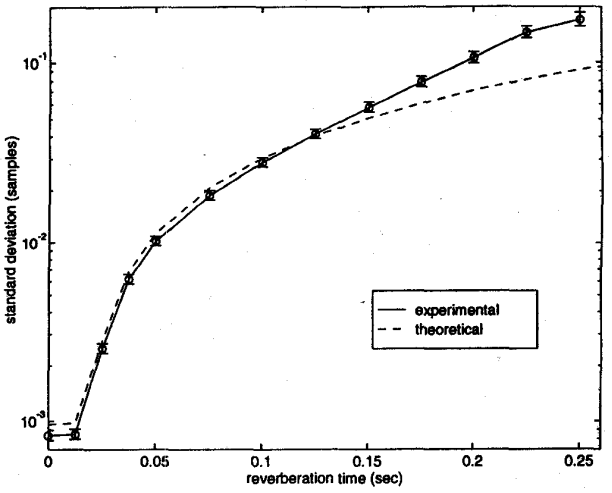


Fig. 3. Standard deviation of nonanomalous time-delay estimates versus reverberation time $T_R$ ($\sigma_{eq}$ given by dashed line).

Fig. 3 shows a severe deterioration of the standard deviation of the estimator as $T_r$ increases. Our experience indicates that, unlike the case of bias, this effect is practically independent of the initial echo structure. Rather, it can be attributed to secondary echoes corresponding to the tails of the room impulse responses, whose effect on the ML cross correlator output for lag values close to the true delay is similar to that of uncorrelated additive noise. To support this view, let us introduce an equivalent SNR as follows:

$$\mathrm{SNR}_{eq} = \frac{\int |H_i(f;0)|^2 S(f) df}{\int [N_i(f) + |H_i(f;\beta) - H_i(f;0)|^2 S(f)] df} \tag{8}$$

where $H_i(f;\beta)$ is the Fourier transform of $h_i(t)$ for a given value of reflection coefficient $\beta$. $\mathrm{SNR}_{eq}$ measures the ratio of the direct-path signal power to the total interference power, which includes both the background noise and the reverberation powers. In the simulations, $\mathrm{SNR}_{eq}$ is calculated using the synthetic impulse responses $h_i(t)$ and is almost identical for both microphones. We saw in Section II that for single-path models, the variance of the ML estimator in the small error regime is closely predicted by the CRLB (5) which, for flat signal and noise spectra, is a function of the in-band SNR. In Fig. 3, we have plotted an equivalent CRLB, denoted $\sigma^2_{eq}(\beta)$ and obtained from (5) upon replacement of the in-band SNR by $\mathrm{SNR}_{eq}$ (8). For $T_R \leq 0.15$, the simulation results closely match this curve, indicating that in the small error regime, the effect of reverberation on the variance of the ML estimator is somewhat equivalent to that of uncorrelated noise.

We note however that $\mathrm{SNR}_{eq}$ can not be used to accurately predict the onset of threshold in Fig. 1 or Fig. 3. Indeed, assuming that the reverberant energy is equivalent to uncorrelated white noise and using the mathematical expressions given in [6] (bandpass case), we find that the boundary between small and large estimation errors should occur at around $\mathrm{SNR}_{eq} = -7.5$ dB, which corresponds here to $T_R = 0.36$ s. In Fig. 1 and Fig. 3, the onset of threshold occurs at the much smaller value of $T_R = 0.15$ s. This indicates that in the case of reverberation, the probability of making a large estimation error (i.e.,

selecting the wrong peak of the ML cross correlator) is substantially higher than it is for uncorrelated white noise with equivalent power.

Additional experiments were made using different values of the simulation parameters. In particular, when the source and microphone positions are varied, the same qualitative behavior as above is generally observed (although the shape of the bias curve for large values of $T_R$ may change). Only two significant exceptions have been noticed. The first, which is not very practical, involves a perfectly symmetrical echo structure in which the direct path signal and its echoes reach the two microphones simultaneously. Here, the total signal components received by the two microphones (i.e., direct path signal plus attenuated echoes) are identical and thus, the performance of MLTDE improves as $T_R$ increases. In the second case, the source or one of the microphones is very close to a wall. As a result, the first few echoes are relatively strong and the deterioration of MLTDE performance occurs for much smaller values of $T_R$. This situation can usually be avoided in practice.

Finally, the conclusions of this study remain also valid when different values of the source signal bandwidth are used, provided that the condition $WT \gg 1$ is satisfied.

## REFERENCES

[1] J. L. Flanagan, D. A. Berkley, G. W. Elko, J. E. West, and M. M. Sondhi, " Autodirective microphone systems," *Acustica*, vol. 73, pp. 58–71, 1991.

[2] H. F. Silverman and S. E. Kirtman, "A two-stage algorithm for determining talker location from linear microphone arrays," *Comput., Speech and Language*, vol. 6, pp. 129–152, 1992.

[3] M. Omologo, "Acoustic event localization using a cross power-spectrum phase based technique," in *Proc. IEEE Int. Conf. on Acoust., Speech, Signal Processing*, Adelaide, Australia, Apr. 1994, vol. 2, pp 273–276.

[4] C. H. Knapp and G. C. Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-24, pp. 320–327, Aug. 1976.

[5] J. P. Ianniello, "Time-delay estimation via cross-correlation in the presence of large estimation errors," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-30, pp. 998–1003, Dec. 1982.

[6] E. Weinstein and A. J. Weiss, "Fundamental limitations in passive time-delay estimation—Part II: Wide-band systems," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-32, pp. 1064–1078, Dec. 1982.

[7] Y. T. Chan, "Time-delay estimation in the presence of multipath propagation," in *Adaptive Methods in Underwater Acoustics*. H. G. Urban Ed. Norwell, MA: Reidel, 1985, pp. 197–206.

[8] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Amer.*, vol. 65, pp. 943–950, Apr. 1979.

[9] M. Tanaka and Y. Kaneda, "Performance of sound source direction estimation methods under reverberant conditions," *J. Acoust. Soc. Jap. (E)*, vol. 14, pp. 291–292, 1993.

[10] P. M. Peterson, "Simulating the response of multiple microphones to a single acoustic source in a reverberant room," *J. Acoust. Soc. Amer.*, vol. 80, pp. 1527–1529, 1986.

[11] H. Kuttruff, H. *Room Acoustics*. London, U.K.: Elsevier, 1991, 3rd edit.