Time Delay Estimation via Minimum Entropy

Jacob Benesty, *Senior Member, IEEE*, Yiteng Huang, *Member, IEEE*, and Jingdong Chen, *Member, IEEE*

***Abstract*—Time delay estimation (TDE) is a basic technique for numerous applications where there is a need to localize and track a radiating source. The most important TDE algorithms for two sensors are based on the generalized cross-correlation (GCC) method. These algorithms perform reasonably well when reverberation or noise is not too high. In an earlier study by the authors, a more sophisticated approach was proposed. It employs more sensors and takes advantage of their delay redundancy to improve the precision of the time difference of arrival (TDOA) estimate between the first two sensors. The approach is based on the multichannel cross-correlation coefficient (MCCC) and was found more robust to noise and reverberation. In this letter, we show that this approach can also be developed on a basis of joint entropy. For Gaussian signals, we show that, in the search of the TDOA estimate, maximizing MCCC is equivalent to minimizing joint entropy. However, with the generalization of the idea to non-Gaussian signals (e.g., speech), the joint entropy-based new TDE algorithm manifests a potential to outperform the MCCC-based method.**

***Index Terms*—Acoustic source localization, cross-correlation coefficient, joint entropy, Laplace distribution, time delay estimation (TDE).**

I. Introduction

THE aim of time delay estimation (TDE) is to measure the relative time difference of arrival (TDOA) among spatially separated sensors. This technique is widely used in radars and sonars for localizing radiating sources. Nowadays, the same technique is used in room acoustics for localization and tracking of talkers for applications such as speech enhancement [1], automatic camera tracking for video-conferencing [2], [3], and microphone array beam steering [4].

Many techniques exist for TDE, but the most popular and most useful algorithms in practice are based on the generalized cross-correlation (GCC) method proposed by Knapp and Carter [5]. The delay estimate between two sensors is obtained as the time-lag that maximizes the cross-correlation between filtered versions of the received signals. This method is well studied, and it performs fairly well in moderately noisy and non-reverberant environments [6], [7]. However, this method tends to break down when reverberation or noise is high and/or noise is not Gaussian. In alpha-stable distributed noise environments, fractional lower order statistics (FLOSs) [8] were found more robust than the GCC for TDE [9]. However, for Gaussian or non-impulsive noise, the FLOS-TDE method is nothing better than the phase-transform GCC algorithm. Alternatively, when more than two microphones are available, the TDOA measurements between different microphone pairs are not independent. Therefore, it is possible to generalize the GCC technique in such a way that all the redundant information can be fully taken into account for achieving an optimal TDE performance in adverse environments. This idea was developed into a multichannel TDE algorithm based on multichannel cross-correlation coefficient (MCCC) in [10] and [11]. It was found that the algorithm's robustness to noise and reverberation gets better as the number of microphones increases.

While the MCCC-based TDE performs well in the presence of noise and reverberation, the MCCC is by no means the only choice for developing the concept of multichannel TDE. MCCC is a second-order-statistics (SOS) measure of dependence among multiple random variables and is ideal for Gaussian source signals. However, for non-Gaussian source signals, MCCC is not sufficient, and higher order statistics (HOS) have more to say about their dependence.

The concept of entropy, which is a statistical (apparently HOS) measure of randomness or uncertainty of a random variable, was introduced by Shannon in the context of communication theory [12]. As it will be demonstrated later, minimizing the entropy is, in fact, equivalent to maximizing the MCCC for TDE if the source signal is Gaussian. While using MCCC for TDE implies that we deal with Gaussian signals, using joint entropy can certainly allow us to go beyond this constraint. In this letter, we show how to use the concept of minimum entropy in TDE.

This letter is organized as follows. In Section II, we present the basic concepts of entropy and joint entropy from the information theory. Section III describes how MCCC and entropy are used in TDE and explains why maximizing MCCC and minimizing entropy are equivalent for TDE for a Gaussian source. In Section IV, we show how a minimum-entropy-based TDE algorithm is derived for speech signals that are assumed to follow Laplace distributions. Simulations are presented in Section V. Finally, this letter is concluded in Section VI.

II. Entropy

In this section, we briefly describe the principles of entropy.

Let x be a random variable with a density $p(x)$. (In this letter, we choose not to distinguish random variables and their realizations.) The entropy is defined as [13]

$$\begin{aligned} H(x) &= -\int p(x)\ln p(x)dx \\ &= -E\left\{\ln p(x)\right\} \end{aligned} \tag{1}$$

where $E\{\cdot\}$ denotes mathematical expectation. The entropy (in the continuous case) is a measure of the structure contained in the density p [14].

Manuscript received June 21, 2006; revised August 15, 2006. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Vesa Valimaki.

J. Benesty is with the Université du Québec, INRS-EMT, Montréal, QC H5A 1K6, Canada (e-mail: benesty@emt.inrs.ca).

Y. Huang and J. Chen are with Bell Laboratories, Lucent Technologies, Murray Hill, NJ 07974 USA (e-mail: arden@research.bell-labs.com; jingdong@ research.bell-labs.com).

Digital Object Identifier 10.1109/LSP.2006.884038

Let us now consider N random variables

$$\mathbf{x} = [x_1 \quad x_2 \quad \cdots \quad x_N]^T$$

with joint density $p(\mathbf{x})$; the corresponding joint entropy is

$$H(\mathbf{x}) = -\int p(\mathbf{x}) \ln p(\mathbf{x}) d\mathbf{x} \tag{2}$$

where $[\cdot]^T$ denotes a vector/matrix transpose.

Entropy of a Multivariate Gaussian Distribution: Let $x_1, x_2, \cdots, x_N$ have a multivariate normal distribution with mean 0 and covariance matrix

$$\mathbf{R} = E\{\mathbf{x}\mathbf{x}^T\} = \begin{bmatrix} \sigma_{x_1}^2 & r_{x_1x_2} & \cdots & r_{x_1x_N} \\ r_{x_1x_2} & \sigma_{x_2}^2 & \cdots & r_{x_2x_N} \\ \vdots & \vdots & \ddots & \vdots \\ r_{x_1x_N} & r_{x_2x_N} & \cdots & \sigma_{x_N}^2 \end{bmatrix}. \tag{3}$$

The probability density function (pdf) of $x_1, x_2, \cdots, x_N$ is then given by

$$p(\mathbf{x}) = \frac{1}{(\sqrt{2\pi})^N [\det(\mathbf{R})]^{1/2}} e^{-\frac{1}{2}\mathbf{x}^T\mathbf{R}^{-1}\mathbf{x}}. \tag{4}$$

By substituting (4) into (2), we can now compute the joint entropy

$$\begin{aligned} H(\mathbf{x}) &= \frac{1}{2}\int p(\mathbf{x})\mathbf{x}^T\mathbf{R}^{-1}\mathbf{x}d\mathbf{x} + \ln\left\{(\sqrt{2\pi})^N [\det(\mathbf{R})]^{1/2}\right\} \\ &= \frac{1}{2}E\left\{\mathbf{x}^T\mathbf{R}^{-1}\mathbf{x}\right\} + \frac{1}{2}\ln\left\{(2\pi)^N \det(\mathbf{R})\right\} \\ &= \frac{1}{2}\mathrm{tr}\left\{E[\mathbf{R}^{-1}\mathbf{x}\mathbf{x}^T]\right\} + \frac{1}{2}\ln\left\{(2\pi)^N \det(\mathbf{R})\right\} \\ &= \frac{N}{2} + \frac{1}{2}\ln\left\{(2\pi)^N \det(\mathbf{R})\right\} \\ &= \frac{1}{2}\ln\left\{(2\pi e)^N \det(\mathbf{R})\right\}. \end{aligned} \tag{5}$$

The entropy for any of the random variables x_n, $n = 1, 2, \cdots, N$ is

$$H(x_n) = \frac{1}{2}\ln\left\{2\pi e \sigma_{x_n}^2\right\}. \tag{6}$$

III. Application to Time Delay Estimation

A. Signal Model

Suppose that we have an array, which consists of N microphones whose outputs are denoted as $x_n(k)$, for $n = 1, 2, \cdots, N$, and with k being the time index. Without loss of generality, we select microphone 1 as the reference point and consider that the propagation of the signal from a far-field source to the array is modeled as

$$x_n(k) = \alpha_n s\left[k - t - f_n(\tau)\right] + w_n(k) \tag{7}$$

where α_n, $n = 1, 2, \cdots, N$ are the attenuation factors due to propagation effects, t is the propagation time from the unknown source $s(k)$ to microphone 1, $w_n(k)$ is an additive noise signal at the nth microphone, τ is the relative delay between microphones 1 and 2, and $f_n(\tau)$ is the relative delay between microphones 1 and n [with $f_1(\tau) = 0$ and $f_2(\tau) = \tau$]. In this letter, we are considering only linear equispaced arrays and the far-field case (i.e., plane wave propagation), in which the function f_n depends on a sole delay τ

$$f_n(\tau) = (n-1)\tau. \tag{8}$$

In other scenarios, f_n probably involves two or three TDOAs and also depends on the microphone array geometry. Presumably, though, the exact mathematical relation of the relative TDOAs is accessible. In addition, the sampling rate needs to be chosen high enough for sufficient resolution such that the values of $f_n(\tau)$'s are all treated as integers.

It is further assumed that $w_n(k)$ is a zero-mean Gaussian random process that is uncorrelated with $s(k)$ and the noise signals at other microphones. It is also assumed that $s(k)$ is reasonably broadband.

B. Minimum Entropy for a Gaussian Source

We are interested in estimating only one time delay (τ) from multiple sensors. Obviously, two sensors are enough to estimate τ. However, the redundant information that is available when more than two sensors are used will help to improve the estimator, especially in the presence of a high level of noise and reverberation.

Consider the following vector:

$$\mathbf{x}(k,m) = [x_1(k) \;\; x_2\left[k + f_2(m)\right] \;\; \ldots \;\; x_N\left[k + f_N(m)\right]]^T.$$

We can check that for $m = \tau$, all the signals $x_n[k + f_n(\tau)]$, $n = 1, 2, \cdots, N$ are aligned. This observation is essential because it already gives an idea on how to find τ. The covariance matrix corresponding to the signal $\mathbf{x}(k, m)$ is

$$\mathbf{R}(m) = E\left\{\mathbf{x}(k,m)\mathbf{x}^T(k,m)\right\}. \tag{9}$$

Therefore, the joint entropy for Gaussian signals is

$$H\left[\mathbf{x}(k,m)\right] = \frac{1}{2}\ln\left\{(2\pi e)^N \det\left[\mathbf{R}(m)\right]\right\}. \tag{10}$$

We argue that the value of m that gives the minimum of $H[\mathbf{x}(k,m)]$, for different m, corresponds to the time delay between microphones 1 and 2. Hence, the solution to our problem is

$$\hat{\tau}_{\mathrm{e}} = \arg\min_m H\left[\mathbf{x}(k,m)\right] \tag{11}$$

where $m \in [-\tau_{\max}, \tau_{\max}]$, and $\tau_{\max}$ is the maximum possible delay.

Let us see now why minimum entropy makes sense for TDE. We define the squared MCCC among the N random variables $x_1, x_2, \cdots, x_N$ as [10], [11], [15], [16]

$$\rho_{\mathbf{x}}^2(m) = 1 - \frac{\det\left[\mathbf{R}(m)\right]}{\prod_{n=1}^N \sigma_{x_n}^2}. \tag{12}$$

We can show that $0 \le \rho_{\mathbf{x}}^2(m) \le 1$ [11]. If two or more random variables are perfectly correlated, then $\rho_{\mathbf{x}}^2 = 1$. If all the processes are completely uncorrelated, then $\rho_{\mathbf{x}}^2 = 0$. In [10] and

[11], it was shown that the MCCC can be used to estimate the relative delay

$$\hat{\tau}_{\rm c} = \arg\max_{m} \rho^2_{\mathbf{x}}(m). \quad (13)$$

It is clear from (10) through (13) that minimizing the entropy or maximizing the MCCC is equivalent for Gaussian signals, so that $\hat{\tau}_{\rm e} = \hat{\tau}_{\rm c}$.

IV. Application to Speech Signals

In room acoustics environments, the sources of interest are speech signals. It is well known that speech samples are well modeled by a Laplace distribution [17], [18]. In this scenario, it makes more sense to take this into account for the estimation of the entropy. However, as it will be seen in the rest of this section, this estimation is far from obvious. Also note that since the noise is assumed to be Gaussian, the signal x_n cannot be exactly modeled by a Laplace distribution. However, we believe that this approximation is plausible and will rely on simulations to justify its viability.

The univariate Laplace distribution with mean zero and variance σ_x^2 is given by

$$p(x) = \frac{\sqrt{2}}{2\sigma_x} e^{-\sqrt{2}|x|/\sigma_x}. \quad (14)$$

It is easy to show that the corresponding entropy is [13]

$$H(x) = 1 + \ln(\sqrt{2}\sigma_x). \quad (15)$$

Let $x_1, x_2, \cdots, x_N$ have a multivariate Laplace distribution with mean $\mathbf{0}$ and covariance matrix $\mathbf{R}$. The pdf of $x_1, x_2, \cdots, x_N$ is [19], [20]

$$p(\mathbf{x}) = 2(2\pi)^{-N/2}\left[\det(\mathbf{R})\right]^{-1/2}\left(\mathbf{x}^T\mathbf{R}^{-1}\mathbf{x}/2\right)^{P/2} \times K_P\left(\sqrt{2\mathbf{x}^T\mathbf{R}^{-1}\mathbf{x}}\right) \quad (16)$$

where $P = (2-N)/2$ and $K_P(\cdot)$ is the modified Bessel function of the third kind (also called the modified Bessel function of the second kind) given by

$$K_P(a) = \frac{1}{2}\left(\frac{a}{2}\right)^P \int_0^\infty z^{-P-1}\exp\left(-z-\frac{a^2}{4z}\right)dz, \; a > 0. \quad (17)$$

The joint entropy is

$$H(\mathbf{x}) = \frac{1}{2}\ln\left[\frac{(2\pi)^N}{4}\det(\mathbf{R})\right] - \frac{P}{2}E\left\{\ln(\theta/2)\right\} - E\left\{\ln K_P(\sqrt{2\theta})\right\} \quad (18)$$

with

$$\theta = \mathbf{x}^T\mathbf{R}^{-1}\mathbf{x}. \quad (19)$$

The two quantities $E\{\ln(\theta/2)\}$ and $E\{\ln K_P(\sqrt{2\theta})\}$ do not seem to have a closed form. So we need to find a numerical way to estimate them. One possibility to do this is the following. Assume that all processes are ergodic; in this case, we can replace ensemble averages by time averages. If we have K samples for each element of the observation vector $\mathbf{x}(k,m)$, we propose to use the following estimators:

$$E\{\ln(\theta/2)\} \approx \frac{1}{K}\sum_{k'=0}^{K-1}\ln\left[\theta(k-k',m)/2\right] \quad (20)$$

$$E\left\{\ln K_P(\sqrt{2\theta})\right\} \approx \frac{1}{K}\sum_{k'=0}^{K-1}\ln K_P\left[\sqrt{2\theta(k-k',m)}\right] \quad (21)$$

with

$$\theta(k-k',m) = \mathbf{x}^T(k-k',m)\mathbf{R}^{-1}(m)\mathbf{x}(k-k',m). \quad (22)$$

In practice, we first estimate $\mathbf{R}(m)$ with the K observations of $\mathbf{x}(k,m)$. When the covariance matrix is estimated, we use the same data to estimate (20) and (21). We then compute the entropy H with (18) for different m, and the one that minimizes H will be a good estimate of the relative delay τ.

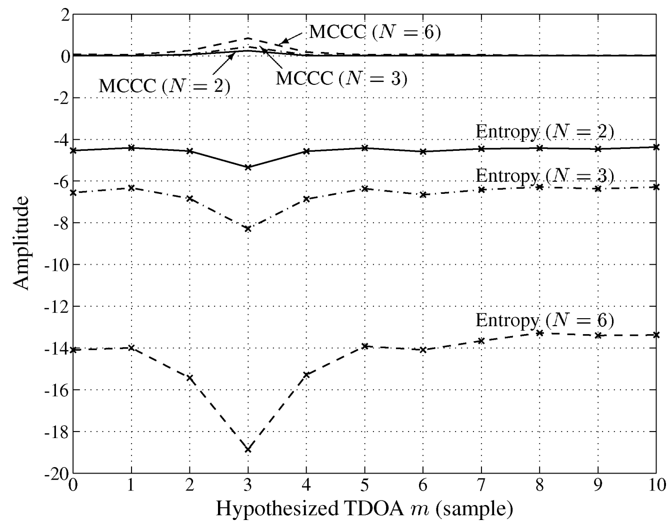

Fig. 1. Comparison of performance between the multichannel TDE algorithms based on MCCC and entropy in an open space. The curves with an x-mark at each data point correspond to the entropy-based algorithm. The true TDOA between the first two microphones is $\tau = 3$ samples.

V. Simulations

In this section, we will evaluate the performance of the proposed entropy-based multichannel TDE algorithm by simulation. A comparison to the MCCC-based method is presented.

The first experiment was carried out in an open space. The source is a female speech signal of 512 samples in length. The addition noise is Gaussian and the signal-to-noise radio (SNR) is 10 dB. The attenuation factors are randomly selected from the range 0.5 to 1. The true TDOA between the first two microphones is three samples. Three linear, equispaced arrays of two, three, and six microphones were investigated. Fig. 1 visualizes the results. It is clear that both MCCC- and entropy-based multichannel TDE algorithms perform pretty well in such an environment with moderate noise and no reverberation. As more microphones are employed, both algorithms demonstrate better robustness to noise while the valleys of the entropy curves are in general sharper than the peaks of their MCCC counterparts.

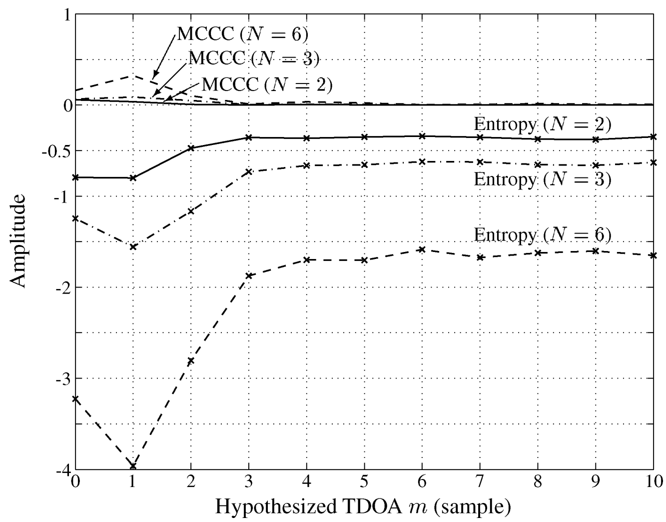

Fig. 2. Comparison of performance between the multichannel TDE algorithms based on MCCC and entropy in the varechoic chamber at Bell Labs. The curves with an x-mark at each data point correspond to the entropy-based algorithm. The true TDOA between the first two microphones is $\tau = 1$ sample. Note that the MCCC has its peak at a wrong position $(m = 0)$ when only two microphones are used.

In the second experiment, we studied the performance of these TDE algorithms in a real, reverberant environment. The channel impulse responses were measured in the Varechoic chamber at Bell Labs [21]. The chamber is a rectangular room with 368 electronically controlled panels that vary the acoustic absorption of the walls, floor, and ceiling [22]. Therefore, the level of room reverberation is well controlled by the percentage of open panels. In this experiment, 30% of the panels are open, which leads to a reverberation time of approximately 380 ms. The original impulse response were measured at 8 kHz and had 4096 samples. For this experiment, they are truncated to 512 samples. Again, the source is a female speech signal, the addition noise is Gaussian, and the SNR is 10 dB. The true TDOA between the first two microphones is one sample. The results are presented in Fig. 2. We see that when only two microphones are used, the MCCC has its peak at a wrong position $m = 0$ while the entropy produces a correct TDOA estimate. As more microphones are employed, while both algorithms work fine, the entropy gives much better-defined extrema.

VI. Conclusions

TDE is a challenging problem in adverse environments with strong noise and considerable reverberation. In this letter, the concept of minimum entropy is introduced, and a novel entropy-based multichannel TDE algorithm is developed. It is explained that minimizing joint entropy is equivalent to maximizing MCCC for Gaussian sources. However, for non-Gaussian sources, entropy is a more comprehensive measure of statistical dependence than MCCC. Simulations show that the proposed minimum-entropy-based TDE algorithm is much more robust to noise in general and reverberation in particular than the MCCC-based TDE approach.

References

[1] J. Benesty, S. Makino, and J. Chen, Eds., *Speech Enhancement*. Berlin, Germany: Springer-Verlag, 2005.
[2] H. Wang and P. Chu, "Voice source localization for automatic camera pointing system in videoconferencing," in *Proc. IEEE ASSP Workshop Applications Signal Processing Audio Acoustics*, 1997.
[3] Y. Huang, J. Benesty, and G. W. Elko, "Microphone arrays for video camera steering," in *Acoustic Signal Processing for Telecommunication*, S. L. Gay and J. Benesty, Eds. Norwell, MA: Kluwer, 2000, ch. 11, pp. 239–259.
[4] M. Brandstein and D. Ward, *Microphone Arrays*. Berlin, Germany: Springer, 2001.
[5] C. H. Knapp and G. C. Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-24, no. 4, pp. 320–332, Aug. 1976.
[6] J. P. Ianniello, "Time delay estimation via cross-correlation in the presence of large estimation errors," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-30, no. 6, pp. 998–1003, Dec. 1982.
[7] B. Champagne, S. Bédard, and A. Stéphenne, "Performance of time-delay estimation in presence of room reverberation," *IEEE Trans. Speech Audio Process.*, vol. 4, no. 2, pp. 148–152, Mar. 1996.
[8] M. Shao and C. L. Nikias, "Signal processing with fractional lower order moments: stable processes and their applications," *Proc. IEEE*, vol. 81, no. 7, pp. 986–1010, Jul. 1993.
[9] P. G. Georgiou, P. Tsakalides, and C. Kyriakakis, "Alpha-stable modeling of noise and robust time-delay estimation in the presence of impulsive noise," *IEEE Trans. Multimedia*, vol. 1, no. 3, pp. 291–301, Sep. 1999.
[10] J. Chen, J. Benesty, and Y. Huang, "Robust time delay estimation exploiting redundancy among multiple microphones," *IEEE Trans. Speech Audio Process.*, vol. 11, no. 6, pp. 549–557, Nov. 2003.
[11] J. Benesty, J. Chen, and Y. Huang, "Time-delay estimation via linear interpolation and cross-correlation," *IEEE Trans. Speech Audio Process.*, vol. 12, no. 5, pp. 509–519, Sep. 2004.
[12] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, 1948, pp. 379–423, 623–656.
[13] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
[14] I. Kojadinovic, "On the use of mutual information in data analysis: an overview," in *Proc. Int. Symp. Applied Stochastic Models Data Analysis*, 2005.
[15] H. Gish and D. Cochran, "Generalized coherence," in *Proc. IEEE ICASSP*, 1988, vol. 5, pp. 2745–2748.
[16] D. Cochran, H. Gish, and D. Sinno, "A geometric approach to multichannel signal detection," *IEEE Trans. Signal Process.*, vol. 43, no. 9, pp. 2049–2057, Sep. 1995.
[17] L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals*. Englewood Cliffs, NJ: Prentice-Hall, 1978.
[18] S. Gazor and W. Zhang, "Speech probability distribution," *IEEE Signal Process. Lett.*, vol. 10, no. 7, pp. 204–207, Jul. 2003.
[19] S. Kotz, T. J. Kozubowski, and K. Podgórski, An Asymmetric Multivariate Laplace Distribution Dept. Statist. Appl. Probab., Univ. California at Santa Barbara, Tech. Rep. No. 367, 2000.
[20] T. Eltoft, T. Kim, and T.-W. Lee, "On the multivariate Laplace distribution," *IEEE Signal Process. Lett.*, vol. 13, no. 5, pp. 300–303, May 2006.
[21] A. Härmä, Acoustic Measurement Data From the Varechoic Chamber Tech. Memo., Agere Systems, Nov. 2001.
[22] W. C. Ward, G. W. Elko, R. A. Kubli, and W. C. McDougald, "The new Varechoic chamber at AT&T Bell Labs," in *Proc. Wallance Clement Sabine Centennial Symp.*, 1994, pp. 343–346.