LAPLACE ENTROPY AND ITS APPLICATION TO TIME DELAY ESTIMATION FOR SPEECH SIGNALS

Yiteng (Arden) Huang[†]*, Jacob Benesty*[‡]*, and Jingdong Chen*[†]

[†] Bell Laboratories, Lucent Technologies
600 Mountain Avenue
Murray Hill, New Jersey 07974, USA
{arden, jingdong}@research.bell-labs.com

[‡] Université du Québec, INRS-EMT
800 de la Gauchetière Ouest, Suite 6900
Montréal, Québec, H5A 1K6, Canada
benesty@emt.inrs.ca

ABSTRACT

Time delay estimation (TDE) is a basic technique for numerous applications where there is a need to localize and track a radiating source. It is particularly challenging in the presence of noise and reverberation, and when the source signal is speech which is inherently nonstationary and random. 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 of 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 TDOA (time difference of arrival) 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 paper, 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. But with the generalization of the idea to non-Gaussian speech signals, the joint entropy based new multichannel TDE algorithm manifests a potential to outperform the MCCC-based method. Since there is no rigorous mathematical formula for speech entropy, we use the assumption that speech can be plausibly modeled by a Laplace distribution and develop a practical approximation of Laplace entropy for TDE of speech signals. The performance of the proposed new algorithm is investigated via simulations.

Index Terms— Time delay estimation, multichannel cross-correlation coefficient, entropy, Laplace distribution

1. 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], and microphone array beam steering [3].

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 [4]. 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 [5]. However, this method tends to break down when reverberation or noise is high. 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 [6], [7]. 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. But 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. 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 paper, we show how to use the concept of minimum entropy in TDE.

Speech is a complicated random process and there is no rigorous mathematical formula for its entropy. In our study, we employ the assumption that speech can be fairly well modeled by a Laplace distribution and we then try to use Laplace entropy in developing the new idea of minimum entropy for TDE of speech signals. But computing Laplace entropy is not straightforward. An approximation will be derived and its viability will be justified via simulations.

2. ENTROPY

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

Let x be a random variable with a density $p(x)$. (In this paper, we choose not to distinguish random variables and their realizations.) The entropy is defined as [8]:

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

where $E\{\cdot\}$ denotes mathematical expectation. The entropy (in the continuous case) is a measure of the structure contained in the density p [9].

Let us now consider N random variables

$$\mathbf{x} = \left[\begin{array}{cccc} x_1 & x_2 & \cdots & x_N \end{array} \right]^T,$$

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.

2.1. 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

$$\begin{aligned} \mathbf{R} &= E\left\{\mathbf{x}\mathbf{x}^T\right\} \\ &= \left[\begin{array}{cccc} \sigma_{x_1}^2 & r_{x_1x_2} & \cdots & r_{x_1x_N} \\ r_{x_1x_2} & \sigma_{x_2}^2 & \cdots & r_{x_2x_N} \\ \vdots & \vdots & \ddots & \vdots \\ r_{x_1x_N} & r_{x_2x_N} & \cdots & \sigma_{x_N}^2 \end{array} \right]. \end{aligned} \tag{3}$$

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

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

where $\det(\cdot)$ denotes the determinant of the involved matrix. By substituting (4) into (2), we can now compute the joint entropy,

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

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

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

3. APPLICATION TO TIME DELAY ESTIMATION

3.1. 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:

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

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 paper, 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. But presumably, 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.

3.2. 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 high level of noise and reverberation.

Consider the following vector:

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

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 $H\left[\mathbf{x}(k,m)\right]$, 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 multichannel cross-correlation coefficient (MCCC) among the N random variables $x_1, x_2, \cdots, x_N$, as [10], [11], [6], [7],

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

We can show that, $0 \leq \rho_{\mathbf{x}}^2(m) \leq 1$ [7]. 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 [6] and [7], it was shown that the MCCC can be used to estimate the relative delay:

$$\hat{\tau}_{\mathrm{c}} = \arg\max_m \rho_{\mathbf{x}}^2(m). \tag{13}$$

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

4. 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 [12], [13]. 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 to be obvious. Also note that since the noise is assumed to be Gaussian, the signal x_n cannot be exactly modeled by a Laplace distribution. But 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

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

It is easy to show that the corresponding entropy is [8],

$$H(x) = 1 + \ln\left(\sqrt{2}\,\sigma_x\right). \tag{15}$$

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 [14], [15]:

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

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,

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

The joint entropy is:

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

with

$$\theta = \mathbf{x}^T\mathbf{R}^{-1}\mathbf{x}. \tag{19}$$

The two quantities $E\{\ln(\theta/2)\}$ and $E\left\{\ln K_P\left(\sqrt{2\theta}\right)\right\}$ 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:

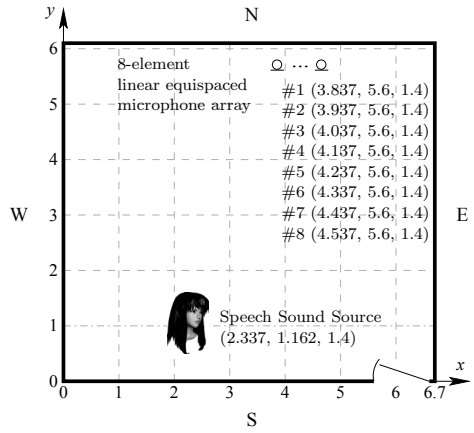

Fig. 1. Floor plan of the varechoic chamber at Bell Labs (coordinate values measured in meters).

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

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

with

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

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.

5. 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 simulations were carried out using the impulse responses measured in a real, reverberant environment: the varechoic chamber at Bell Labs [16]. The chamber is a rectangular room (6.7 m $\times$ 6.1 m $\times$ 2.9 m) with 368 electronically controlled panels that vary the acoustic absorption of the walls, floor, and ceiling [17]. Therefore the level of room reverberation is well controlled by the percentage of open panels. Three panel configurations were investigated: 75%, 30%, and 0% open panels. Their average T_{60} reverberation times are approximately 310 ms, 380 ms (moderately reverberant), and 580 ms (highly reverberant), respectively. The original impulse responses were measured at 8 kHz and had 4096 samples. For our simulations, they are truncated to 300 samples. The source is a female speech signal of 20 seconds in length sampled at 8 kHz. Eight microphones are employed. The positions of the sound source and microphones are shown in Fig. 1. The true TDOA between the first two microphones is $\tau = 1$ sample. The clean speech is convolved with the measured impulse responses to generate the microphone outputs. The additive noise is white and Gaussian. The frame size is 2000 samples.

Table 1. Simulation summary statistics in terms of the percentage of successful time delay estimates using the previously developed MCCC-based and the proposed entropy-based algorithms. The impulse responses were measured in the varechoic chamber at Bell Labs under three different room acoustics setups. The additive noise is Gaussian at -5 dB SNR.

TDE Methods	Percent Successful Estimates (%)			
	$N = 2$	$N = 4$	$N = 6$	$N = 8$
	$T_{60} = 310$ ms			
MCCC	52.50	73.75	81.25	83.75
Entropy	60.00	76.25	81.25	86.25
	$T_{60} = 380$ ms			
MCCC	48.75	70.00	73.75	78.75
Entropy	52.50	75.00	72.50	83.75
	$T_{60} = 580$ ms			
MCCC	33.75	60.00	67.50	67.50
Entropy	47.50	66.25	72.50	81.25

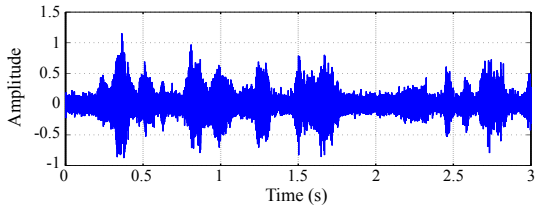

Fig. 2. The first 3 s waveform of the first microphone output for $T_{60} = 310$ ms and SNR $= -5$ dB.

We found that when the signal-to-noise ratio (SNR) was fairly high (> 0 dB), both MCCC and entropy-based TDE algorithms can accurately estimate τ with no errors. In order to show their difference in performance, we choose to present in Table 1 a set of results at a fairly low SNR of -5 dB. As shown in Fig. 2 with the output of the first microphone for $T_{60} = 310$ ms, at such a low SNR many weak speech tails are covered by the additive noise such that TDE is challenging as one can imagine. The results are visualized in Fig. 3. We can clearly see a performance degradation as reverberation time increases for both methods. But using more microphones, the robustness against room reverberation is significantly improved, which is particularly true for the entropy-based algorithm. Between the two studied algorithms, the entropy-based algorithm performs in general comparably to or better than the MCCC-based method (only occasionally worse). The advantage is more obvious for small numbers of used microphones, e.g., $N = 2$.

6. CONCLUSIONS

Time delay estimation is a challenging problem in adverse environments with strong noise and considerable reverberation. In this paper, 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 multichannel cross-correlation coefficient (MCCC) for Gaussian sources. But 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 reverberation than the MCCC-based TDE approach.

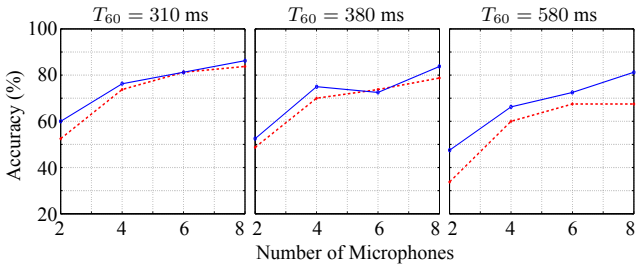

Fig. 3. Comparison of the percentage of successful TDEs between the MCCC (dotted line) and entropy (solid line) algorithms for SNR $= -5$ dB .

7. REFERENCES

[1] J. Benesty, S. Makino, and J. Chen, eds., *Speech Enhancement*. Springer-Verlag, Berlin, 2005.

[2] 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., Kluwer Academic Publishers, Boston, MA, chap. 11, pp. 239–259, 2000.

[3] M. Brandstein and D. Ward, *Microphone Arrays*. Berlin: Springer, 2001.

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

[5] B. Champagne, S. Bédard, and A. Stéphenne, "Performance of time-delay estimation in presence of room reverberation," *IEEE Trans. Speech Audio Processing*, vol. 4, pp. 148–152, Mar. 1996.

[6] J. Chen, J. Benesty, and Y. Huang, "Robust time delay estimation exploiting redundancy among multiple microphones," *IEEE Trans. Speech Audio Processing*, vol. 11, pp. 549–557, Nov. 2003.

[7] J. Benesty, J. Chen, and Y. Huang, "Time-delay estimation via linear interpolation and cross-correlation," *IEEE Trans. Speech Audio Processing*, vol. 12, pp. 509–519, Sept. 2004.

[8] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & Sons, Inc., NY, 1991.

[9] I. Kojadinovic, "On the use of mutual information in data analysis: an overview," in *International Symposium on Applied Stochastic Models and Data Analysis*, 2005.

[10] H. Gish and D. Cochran, "Generalized coherence," in *Proc. IEEE ICASSP*, vol. 5, 1988, pp. 2745–2748.

[11] D. Cochran, H. Gish, and D. Sinno, "A geometric approach to multichannel signal detection," *IEEE Trans. Signal Processing*, vol. 43, pp. 2049–2057, Sept. 1995.

[12] L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals*. Englewood Cliffs, NJ: Prentice-Hall, 1978.

[13] S. Gazor and W. Zhang, "Speech probability distribution," *IEEE Signal Processing Lett.*, vol. 10, no. 7, pp. 204–207, July 2003.

[14] S. Kotz, T. J. Kozubowski, and K. Podgórski, "An asymmetric multivariate Laplace distribution," Technical Report No. 367, Department of Statistics and Applied Probability, University of California at Santa Barbara, 2000.

[15] T. Eltoft, T. Kim, and T.-W. Lee, "On the multivariate Laplace distribution," *IEEE Signal Processing Letters*, vol. 13, pp. 300–303, May 2006.

[16] A. Härmä, "Acoustic measurement data from the varechoic chamber," Technical Memorandum, Agere Systems, Nov. 2001.

[17] 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 Symposium*, 1994, pp. 343–346.