EURASIP Journal on
Audio, Speech, and Music Processing
a SpringerOpen Journal

RESEARCH Open Access

# Robust time delay estimation for speech signals using information theory: A comparison study

Fei Wen* and Qun Wan

**Abstract**

Time delay estimation (TDE) is a fundamental subsystem for a speaker localization and tracking system. Most of the traditional TDE methods are based on second-order statistics (SOS) under Gaussian assumption for the source. This article resolves the TDE problem using two information-theoretic measures, joint entropy and mutual information (MI), which can be considered to indirectly include higher order statistics (HOS). The TDE solutions using the two measures are presented for both Gaussian and Laplacian models. We show that, for stationary signals, the two measures are equivalent for TDE. However, for non-stationary signals (e.g., noisy speech signals), maximizing MI gives more consistent estimate than minimizing joint entropy. Moreover, an existing idea of using modified MI to embed information about reverberation is generalized to the multiple microphones case. From the experimental results for speech signals, this scheme with Gaussian model shows the most robust performance in various noisy and reverberant environments.

## Introduction

Time delay estimation (TDE) is a basic problem in modern signal processing and it has found extensive applications such as localizing and tracking radiating sources in radar and sonar. Nowadays, the same technique is used to localize and track acoustic sources in room environments. For example, in automatic camera tracking for video conferencing [1,2], the location of the current speaker is required for the camera to turn toward them; in speech enhancement [3,4] using a steerable microphone array, the speaker location is required for noise cancellation.

TDE for speech signals in adverse acoustic environments with strong noise and reverberation levels has long been a challenging problem. Among the traditional methods for TDE, the most popular one is the generalized cross-correlation (GCC) method proposed by Knapp and Carter [5]. The relative delay is estimated by maximizing the cross-correlation between filtered versions of the received signals. It has been shown in [6,7] that, the GCC method performs fairly well in moderately noisy and lightly reverberant environments. However, it degrades dramatically when noise or reverberation is high. In an attempt to deal better with noise and reverberation, an effective approach was introduced based on multichannel cross-correlation coefficient (MCCC) [8], which performs well in combating both noise and reverberation by taking advantage of the redundant information from multiple sensor pairs. It is found that the approach's robustness gets better as the number of sensors increases.

As a second-order statistics (SOS) measure of the dependence among multiple random variables, the MCCC is ideal for Gaussian signals. However, for non-Gaussian source signals, higher order statistics (HOS) have more to say about their dependence. More recently, the two information-theoretic concepts of joint entropy and mutual information (MI), which can be considered as higher order statistics [9], are used to develop new TDE estimators [10,11]. In [10], the Laplacian is employed to model the speech source, and the relative delay is estimated via minimizing the joint entropy of the multiple microphone output signals. In [11], based on characterizing the speech source as Gaussian, the MI measure is used for TDE, however, the method is restricted to the two microphone case.

Analysing further the work of [10,11], in this article, we present a framework that treats the TDE problem from an information theory point-of-view. Since the two information-theoretic measures have the freedom of selecting a specific distribution model for the source

* Correspondence: wenfei1@hotmail.com
Department of Electronic Engineering, University of Electronic Science and Technology of China, No. 4, Section 2, North Jianshe Road, Chengdu, China

signal, the solutions based on minimizing the joint entropy and maximizing the MI of the multichannel output signals are derived for both Gaussian and Laplacian models. From the experimental results, the Gaussian, compared to the Laplacian, is a better model for the small frames of noisy speech signals used for TDE. Moreover, we show that the two measures are equivalent for TDE when the source signal is stationary. However, for non-stationary signals, maximizing the MI gives more stable and consistent estimate of the relative delay than minimizing the joint entropy.

In addition, in order to combat reverberation more effectively, the MI of multichannel outputs is modified to embed information about reverberation, which helps to improve the estimator's robustness against reverberation. The proposed scheme is verified by simulations in various noisy and reverberant environments.

This paper is organized as follows. 'Signal model' section describes the signal model used throughout this article. 'TDE based on information theory' section presents the joint entropy and MI based methods for both Gaussian and Laplacian models. 'Modified MI of multichannel outputs' section details how to modify the MI based estimator to be more robust against reverberation for multiple microphones. Simulations are presented in 'Simulations' section. 'Conclusion' section summarizes the conclusions of the article.

## Signal model

In an attempt to estimate only one time delay, two sensors are enough. However, it has been shown in [8,10] that employing more than two sensors can significantly improve the estimator's robustness against noise and reverberation by taking advantage of the available redundant information. Consider that we have a linear microphone array consisting of $N$ microphones positioned in an acoustical enclosure. When the reverberation is ignored, the received signals from a single far-field source can be denoted as

$$x_n(k) = \lambda_n s[k - t - \varphi_n(\tau)] + \omega_n(k) \tag{1}$$

for $n = 1,2,...N$, where $\lambda_n$ are the attenuation factors, $t$ is the propagation time from the source $s(k)$ to microphone 1 (without loss of generality, microphone 1 is selected as the reference point), the noise term $\omega_n(k)$ is assumed to be white Gaussian with zero mean and uncorrelated with the source signal and the noise signals at other microphones, $\phi_n(\tau)$ is the relative delay between microphones 1 and $n$ (with $\phi_1(\tau) = 0$ and $\phi_2(\tau) = \tau$). Since we consider only linear equispaced arrays and the far-field case, the function $\phi_n(\tau)$ solely depends on the delay $\tau$

$$\varphi_n(\tau) = (n-1)\tau. \tag{2}$$

In other scenarios with linear but non-equispaced or non-linear arrays, the mathematical formulation of $\phi_n(\tau)$ can be obtained depending on the array geometry. In addition, we assume that the sampling rate was sufficiently high such that the value of $\varphi_n(\tau)$ can be treated as integer.

However, the model described by (1) does not include the effect of reverberation in real room acoustic environments. In order to describe the TDE problem in a room environment where each microphone often receives a large number of echoes due to reflections of the wavefront from objects and room boundaries, we can use a more realistic reverberation model which models the received signals as [12]

$$x_n(k) = h_n * s(k) + \omega_n(k) \tag{3}$$

where $h_n$ denotes the reverberant impulse response between the source and the $n$th microphone and the symbol * denotes convolution. In this model, $j_n$ contains not only the effect of the direct path delay but also that of other reflected path delays. The size of $j_n$ is generally a function of the reverberation time.

## TDE based on information theory

Most of the traditional TDE algorithms are proposed based on a SOS criterion. Since the sensor output signals are random variables, it makes more sense to take into account the probability density functions (pdfs) in quantifying the dependence among those multiple random variables by employing a HOS criterion.

### Entropy and MI

In general, the entropy is a measure of uncertainty of a random variable. Shannon, using an axiomatic approach [13], defined entropy of a random variable $x$ with a pdf $f(x)$ as

$$H[x] = -\int f(x)\ln f(x)\mathrm{d}x. \tag{4}$$

Let us now consider $N$ random variables

$$\mathrm{X} = [x_1\ x_2\ \ldots\ x_N]^{\mathrm{T}} \tag{5}$$

with joint density $f(\mathbf{x})$, where $[\cdot]^{\mathrm{T}}$ denotes a vector/matrix transpose. The corresponding joint entropy of the $N$ random variables can be considered to be the entropy of the single vector-valued random variable $\mathbf{x}$

$$H[\mathrm{X}] = -\int f(\mathrm{X})\ln f(\mathrm{X})\mathrm{d}\mathrm{X}. \tag{6}$$

The MI is an information-theoretic measure of the information that one random variable contains about another random variable. If we consider two variables $x_1$

and $x_2$, then the MI $I(x_1, x_2)$ is the Kullback-Leibler (KL) divergence between the joint density $f(x_1, x_2)$ and the factorized marginal density $f(x_1)$ and $I(x_2)$ [9], i.e.,

$$I(x_1, x_2) = \iint f(x_1, x_2) \ln \frac{f(x_1, x_2)}{f(x_1) f(x_2)} \mathrm{d}x_1 \mathrm{d}x_2. \tag{7}$$

When multiple random variables are concerned, we use the *total correlation* [14], which is one of several generalizations of the MI in probability theory and in particular in information theory, to express the amount of dependency existing among the variables. The multivariate MI of $\mathbf{x}$ can be formulated as

$$\begin{aligned} I(\mathrm{X}) &= \int_{\mathrm{X}} f(\mathrm{X}) \ln \frac{f(\mathrm{X})}{\prod_{n=1}^{N} f(x_n)} \mathrm{dX} \\ &= \sum_{n=1}^{N} H[x_n] - H[\mathrm{X}]. \end{aligned} \tag{8}$$

According to (1), we consider the following parameterized vector:

$$\mathrm{X}(k, m) = [x_1(k)\ x_2[k + \varphi_2(m)]\ \ldots\ x_N[k + \varphi_N(m)]]^{\mathrm{T}}. \tag{9}$$

Obviously, when we determine the correct delay $m = \tau$, the signal components at different microphones will be synchronized, and the information that one microphone signal has about the others will be maximum. In this case, the entropy and MI of x($k$, $m$) will reach minimum and maximum, respectively. Thus, the relative delay can be estimated by minimizing the entropy or maximizing the MI

$$\hat{\tau}_{\mathrm{e}} = \arg\min_{m} H(\mathrm{X}(k, m)) \tag{10}$$

$$\hat{\tau}_{\mathrm{MI}} = \arg\max_{m} I(\mathrm{X}(k, m)). \tag{11}$$

In order to apply the two measures, the joint density and marginal distributions of the multichannel output signals are required. Since the information-theoretic concepts have the advantage of freely source model selection, other potential density such as Laplacian can be tried as in this article or [10].

#### Gaussian signals

A Gaussian random variable $x$ with mean zero and variance $\sigma_x^2$ has a pdf given by

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma_x} \mathrm{e}^{-\frac{1}{2}x^2/\sigma_x^2}. \tag{12}$$

The resulting entropy is

$$H(x) = \frac{1}{2} \ln\{2\pi e \sigma_x^2\} \tag{13}$$

Let that $x_1$, $x_2$,...., $x_N$ follow a multivariate Gaussian distribution with mean 0 and covariance matrix

$$\mathbf{R} = E\{\mathrm{X}\mathrm{X}^{\mathrm{T}}\} = \begin{bmatrix} \sigma_{x_1}^2 & r_{x_1x_2} & \cdots & r_{x_1x_N} \\ r_{x_1x_2} & \sigma_{x_2}^2 & \cdots & r_{x_2x_N} \\ \vdots & \vdots & \ddots & \vdots \\ r_{x_1x_N} & r_{x_2x_N} & \cdots & \sigma_{x_N}^2 \end{bmatrix}. \tag{14}$$

The joint pdf of $x_1$, $x_2$,...., $x_N$ is

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{N/2} \det(\mathbf{R})^{1/2}} \mathrm{e}^{-\frac{1}{2}\mathbf{x}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{x}}. \tag{15}$$

By substituting (15) into (6), the entropy of x can be obtained as [10]

$$H(\mathrm{X}) = \frac{1}{2} \ln\left[(2\pi e)^N \det(\mathbf{R})\right]. \tag{16}$$

Accordingly, the MI of the jointly Gaussian distributed random vector x can be formulated as [11]

$$I(\mathrm{X}) = -\frac{1}{2} \ln\left[\frac{\det(\mathbf{R})}{\prod_{n=1}^{N} \sigma_{x_n}^2}\right]. \tag{17}$$

In practice, with $K$ observations of x, we firstly estimate the covariance matrix

$$\mathbf{R}(m) = E\{\mathrm{X}(k, m)\mathrm{X}^{\mathrm{T}}(k, m)\}. \tag{18}$$

Then, we compute the entropy $H$(x($k$, $m$)) (or the MI $I$ (x($k$, $m$))) for different $m$ and choose the one that minimizes the entropy (or maximizes the MI) to be the optimal estimate of the relative delay.

It can be easily checked that maximizing the MI for Gaussian signals (17) is, indeed, equivalent to maximizing the squared MCCC among the $N$ random variables, which is defined as [8]

$$\rho^2(m) = 1 - \frac{\det[\mathbf{R}(m)]}{\prod_{n=1}^{N} \sigma_{x_n}^2}. \tag{19}$$

Furthermore, note that, the time shift independent variance $\sigma_{x_n}^2$ are constant if the signals are stationary and the data sample length $K$ is sufficiently large (ideally $K \to \infty$). In this case, it is obvious that, minimizing the entropy (16) is equivalent to maximizing the MI (17) or MCCC (19) for TDE. However, for non-stationary signals, the entropy (16) is affected by the variance change. These findings will be verified by simulations later.

#### Laplacian signals

The univariate Laplacian distribution with mean zero and variance $\sigma_x^2$ is given by

$$f(x) = \frac{\sqrt{2}}{2\sigma_x} e^{-\sqrt{2}|x|/\sigma_x}. \quad (20)$$

The corresponding entropy is

$$H(x) = 1 + \frac{1}{2}\ln\{\sqrt{2}\sigma_x\} \quad (21)$$

Suppose that the elements of the random vector x have a multivariate Laplacian distribution with mean **0** and covariance matrix **R**. The joint density is given by [15]

$$f(\mathrm{X}) = 2(2\pi)^{-N/2}\det(\mathbf{R})^{-1/2}(\mathrm{X}^T\mathbf{R}^{-1}\mathrm{X}/2)^{P/2}B_P(\sqrt{2\mathrm{X}^T\mathbf{R}^{-1}\mathrm{X}}) \quad (22)$$

where $P = 1\text{-}N/2$ and $B_P(\cdot)$ is the modified Bessel function of the second kind.

The joint entropy can be obtained as [10]

$$H(\mathrm{X}) = \frac{1}{2}\ln\left[\frac{(2\pi)^N\det(\mathbf{R})}{4}\right] - \frac{P}{2}E\{\ln(\beta/2)\} - E\left\{\ln B_P(\sqrt{2\beta})\right\} \quad (23)$$

with

$$\beta = \mathrm{X}^{\mathrm{T}}\mathbf{R}^{-1}\mathrm{X}. \quad (24)$$

By substituting (21) and (23) into (8), the MI is given by

$$I(\mathrm{X}) = -\frac{1}{2}\ln\left[\frac{\pi^N\det(\mathbf{R})}{4e^{2N}\prod_{n=1}^{N}\sigma_{x_n}^2}\right] + \frac{P}{2}E\{\ln(\beta/2)\} + E\left\{\ln B_P(\sqrt{2\beta})\right\} \quad (25)$$

When the entropy (23) or MI (25) is applied to TDE, we use a numerical way to estimate $E\{\ln(\beta/2))\}$ and $E\{\ln B_P(\sqrt{2\beta})\}$ from observed data since they do not seem to have a closed form. Suppose that we have $K$ samples for each element of the observation vector x($k$, $m$), we replace ensemble averages by time averages

$$E\{\ln(\beta/2)\} \approx \frac{1}{K}\sum_{k'=0}^{K-1}\ln[\beta(k-k',m)/2] \quad (26)$$

$$E\left\{\ln B_P(\sqrt{2\beta})\right\} \approx \frac{1}{K}\sum_{k'=0}^{K-1}\ln B_P[\sqrt{2\beta(k-k',m)}] \quad (27)$$

with

$$\beta(k-k',m) = \mathrm{X}^{\mathrm{T}}(k-k',m)\mathbf{R}^{-1}(m)\mathrm{X}(k-k',m). \quad (28)$$

In practice, we estimate the covariance matrix $\mathbf{R}(m)$ firstly. Afterwards, (26) and (27) can be estimated immediately. Then, the entropy (23) or MI (25) can be computed to estimate the relative delay.

It has been shown that the Laplacian distribution is the best model for speech samples during voice activity intervals compared to the Gaussian, generalized Gaussian and gamma distribution [16], which has been taken into account for the estimation of entropy for speech signals in [10]. However, since the noise is typically Gaussian, assuming a Laplacian distribution for the noisy microphone array outputs is questionable, particularly for low SNR conditions.

In addition, similar to the solutions for Gaussian signal, the MI (25) is insensitive to variance change of the sensor outputs compared to the entropy (23).

## Modified MI of multichannel outputs

It is shown in [11] that the estimator searching the relative delay between two microphone signals by directly maximizing the MI suffers from the same limitations of GCC, and it is not robust enough in reverberant acoustic environments.

Consider that the relative delay between the two signals $x_1(k)$ and $x_2(k)$ is $\tau$. In the absence of reverberation, only a single delay is present between the two signals. Thus, the information contained in a sample $l$ of $x_1(k)$ is only dependent on the information contained in the sample $l$ - $\tau$ of $x_2(k)$. When reverberation is present, then, the information contained in a sample $l$ of $x_1(k)$ is also contained in neighboring samples of the sample $l$ - $\tau$ of $x_2(k)$. In this scenario, the MI is not representative enough in the presence of reverberation. Thus, in order to better estimate the information conveyed by the two signals, the modified MI that consider jointly $Q$ neighboring samples can be formulated as [11]

$$\begin{aligned} &I_Q(x_1(k), x_2(k)) \\ &= H[x_1(k)] + H[x_1(k+1)] + \cdots + H[x_1(k+Q)] \\ &\quad + H[x_2(k)] + H[x_2(k+1)] + \cdots + H[x_2(k+Q)] \\ &\quad - H[x_1(k), \cdots, x_1(k+Q), x_2(k), \cdots, x_2(k+Q)] \end{aligned} \quad (29)$$

When the condition of using multiple sensors is concerned, the modified MI of x($k$, $m$) can be formulated as

$$I_Q(\mathrm{X}(k,m)) = I(\mathrm{X}_Q(k,m)) \quad (30)$$

with

$$\begin{aligned} \mathrm{X}_Q(k,m) = [&x_1(k)\ x_1(k+1)\ \cdots\ x_1(k+Q)\ x_2[k+\varphi_2(m)] \\ &x_2[k+\varphi_2(m)+1]\ \cdots\ x_2[k+\varphi_2(m)+Q]\ \cdots \\ &x_N[k+\varphi_N(m)]\ x_N[k+\varphi_N(m)+1]\ \cdots\ x_N[k+\varphi_N(m)+Q]]^{\mathrm{T}} \end{aligned} \quad (31)$$

The length of $\mathrm{x}_Q$ is $N(Q + 1)$. We call $Q$ the order of the system. Accordingly, with the $K$ data samples, we compute the MI $I(\mathrm{x}_Q(k, m))$ for different $m$ and choose the one that maximizes the MI to be a good estimation of the relative delay

$$\hat{\tau}_Q = \arg\max_m I(X_Q(k, m)). \quad (32)$$

## Simulations

In this section, we conduct experiments for speech signals to evaluate the estimators using both simulated and real impulse responses in reverberant room environments. A real female speech signal is convolved with the room impulse responses to generate microphone signals. The microphone signals are partitioned into non-overlapping frames with a frame size of 600 samples. In addition, mutually independent zero-mean white Gaussian noise is introduced to each microphone signal to control the SNR.

For each set of experimental conditions, the 100 frames are processed to generate 100 estimates. The TDE performance is evaluated in terms of the root mean-squared error (RMSE) of the estimates.

### Simulated reverberant channels

The image model technology [17,18] is used to simulate real reverberant acoustic environments of a room with room dimensions of [8 6.5 3] $m$. A linear equispaced microphone array of six omni-directional receivers with inter-element spacing of 10 cm is considered. Two reverberation conditions are simulated for different reverberation time $T_{60}$, which is defined as the time for the sound to decay to a level 60 dB below its original level. The two reverberation times are approximately 200 and 500 ms, respectively. The results are averaged

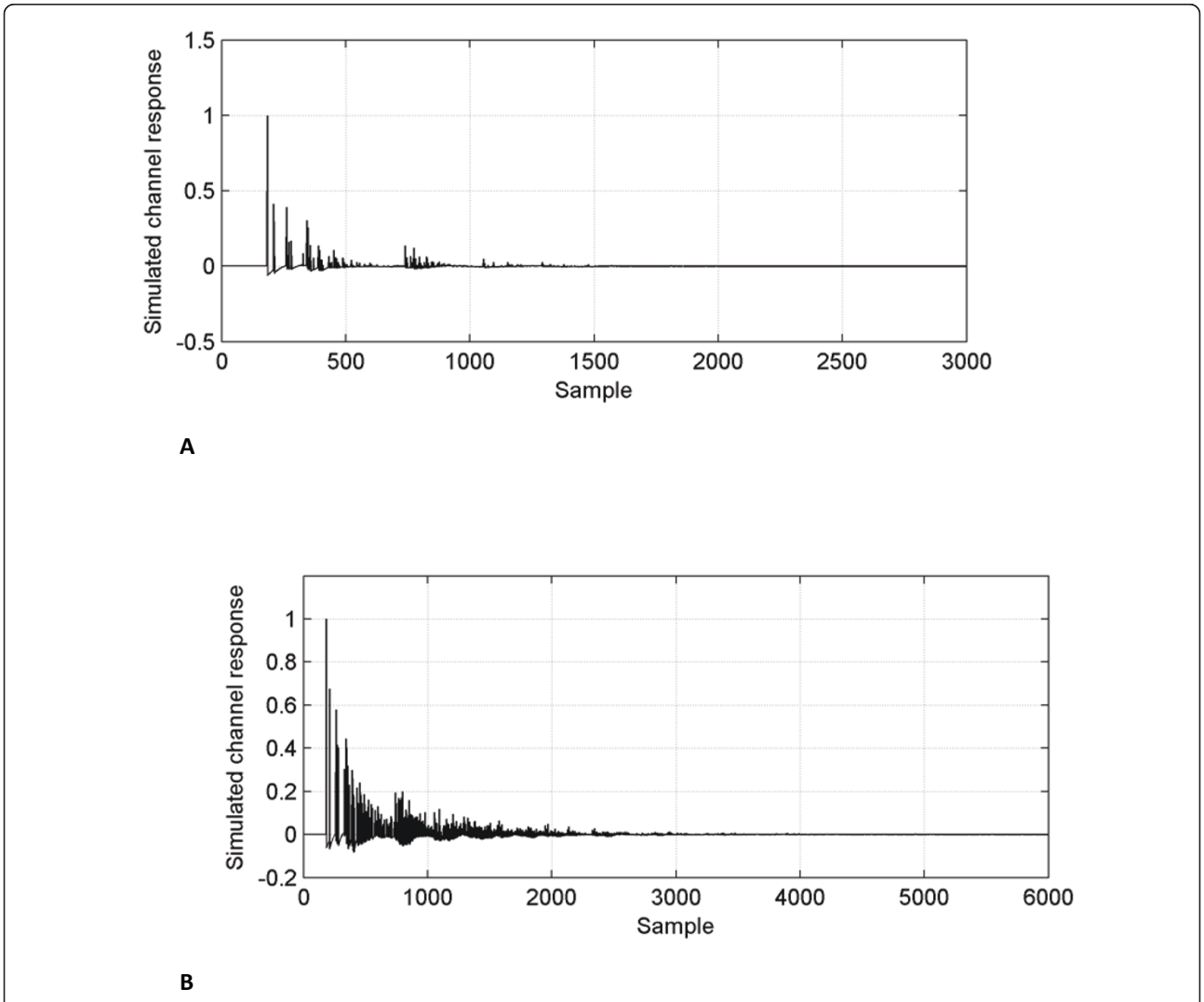
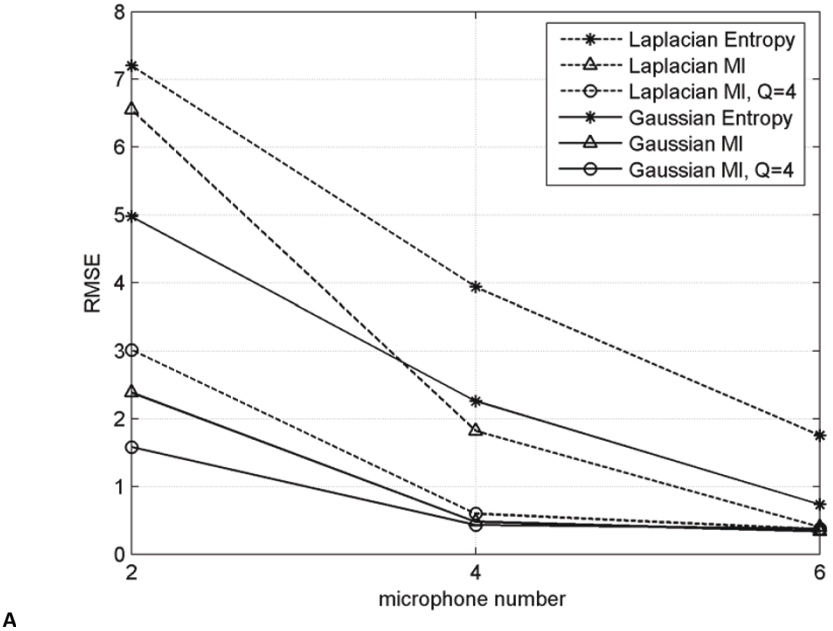


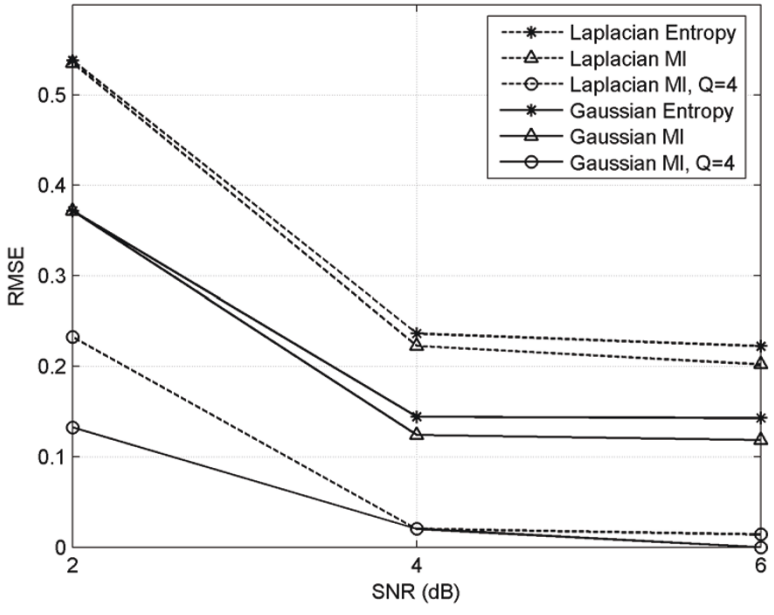

**Figure 1 Examples of simulated channel responses between the source and the first microphone for two reverberation conditions**. **(a)** $T_{60}$ = 200 ms and **(b)** $T_{60}$ = 500 ms.

over twenty random displacements and rotations of the relative geometry between the source and the array inside the room. Figure 1 shows two examples of the simulated channel responses between the source and the first microphone for the two reverberation conditions.

In the first experiment, the entropy, MI and modified MI based estimators for both Gaussian and Laplacian

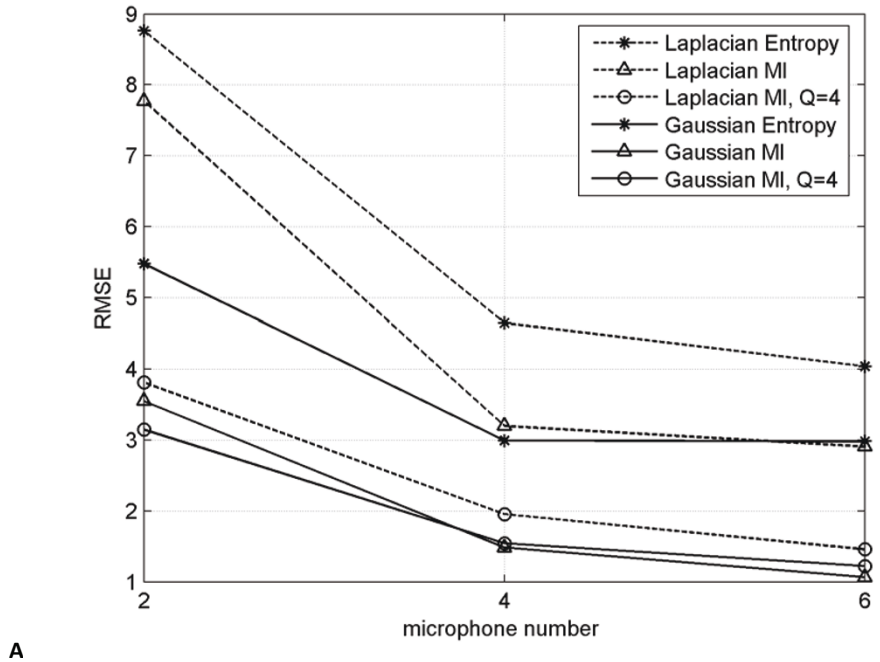



A

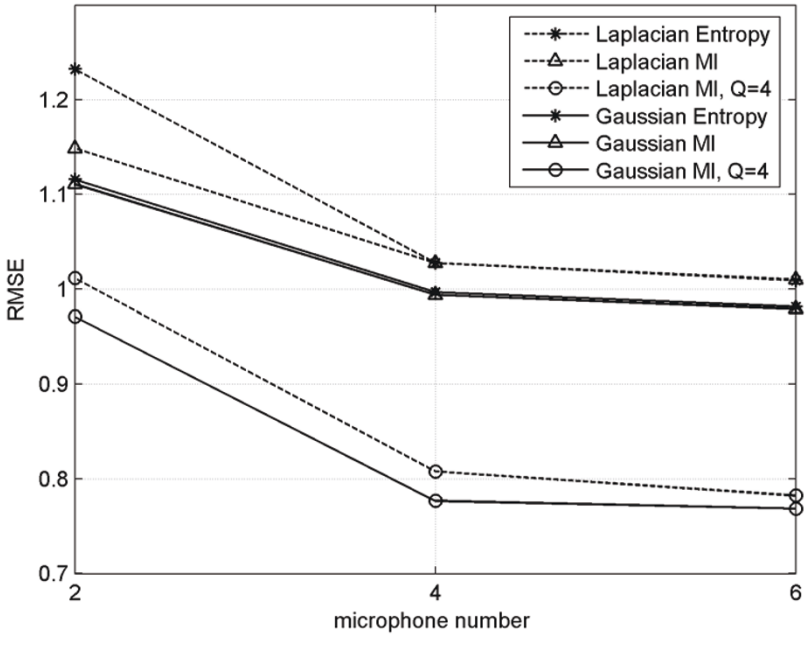



B

**Figure 2 RMSE versus different number of microphones for the two noise conditions**. **(a)** SNR = -5 dB, **(b)** SNR = 25 dB in the moderately reverberant environments where $T_{60}$ = 200 ms.




**A**




**B**

**Figure 3 RMSE versus different number of microphones for the two noise conditions**. **(a)** SNR = -5 dB, **(b)** SNR = 25 dB in the moderately reverberant environments where $T_{60}$ = 500 ms.

models are compared in two different noise conditions with SNR = -5 and 25 dB, respectively. Figures 2 and 3 depict the relationship between the estimate RMSE and the number of microphones for the two reverberation conditions, respectively. The system order of the modified MI based method is chosen to be $Q = 4$.

As clearly shown in Figures 2 and 3, all the estimators deteriorate as noise or reverberation time increases. For

example, for two microphones, the RMSE of each approach for SNR = -5 dB is at least more than six times that for SNR = 25 dB in the moderate reverberation condition with $T_{60}$ = 200 ms. Meanwhile, when the number of microphones is fixed and in the same noise conditions, each approach shows much higher RMSE in the highly reverberant environment compared to the moderately reverberant environment. However, for the same noise and reverberation conditions, the RMSE drops evidently as the number of microphones increases for all the algorithms, particularly in the high noise condition. This indicates that better performance can be achieved by employing more microphones.

Moreover, it can be seen that the entropy and MI measures have comparable performance in the low noise condition with SNR = 25 dB. But in the high noise condition with SNR = -5 dB, the MI based approaches performs distinctly better than the entropy based ones. That can be interpreted as the MI, compared to entropy, is insensitive to the variance change caused by the non-stationary of the noise corrupted speech signals.

In addition, each of the three measures with the Gaussian model exhibits a better performance compared to Laplacian, especially for the high noise condition. This can be explained as follows. The speech samples during voice activity intervals are Laplacian random variables [16] and the noise is typically Gaussian. Thus, the noisy microphone output, which is a mixture of Laplacian and Gaussian random variables, cannot be well modeled by Laplacian, particularly when the noise is high. Moreover, it has been shown that, the joint distribution of two samples of speech with 0.1 ms distance looks very like Gaussian [16]. That is the case of this article, where the sampling period is approximately 0.1 ms.

In general, for the same number of microphones and the same noise and reverberation conditions, the modified MI based algorithms with an order of $Q$ = 4 obviously performs better than their entropy based and MI based counterparts, which is demonstrated by their distinct lower RMSE in most cases.

### Real reverberant channels

In this subsection, we repeat the first experiment using real measured room impulse responses from the Multichannel Acoustic Reverberation Database at York (MARDY) to evaluate the algorithms. The database comprises a collection of room impulse responses measured with a linear array for various source-array separations in a varechoic room. The collected data are available at http://www.commsp.ee.ic.ac.uk/sap/. Figure 4 shows one of the recorded channel responses. The reverberation time of the used channel responses is approximately 447 ms.

Figure 5 presents the relationship between the estimate RMSE and the number of microphones for two noise conditions with SNR = -5 dB and SNR = 25 dB, respectively. The modified MI based algorithms distinctly performs better than other algorithms except for the six microphones case with SNR = 25 dB. Moreover, while the Gaussian model shows better performance than the Laplacian model in the low SNR condition with SNR = -5 dB, both the models in general give comparable performance in the high SNR condition with SNR = 25 dB.

## Conclusions

In this article, the TDE problem is viewed from an information theory point. It is revealed that, maximizing the MI for TDE gives more consistent results compared to minimizing the joint entropy since it is insensitive to

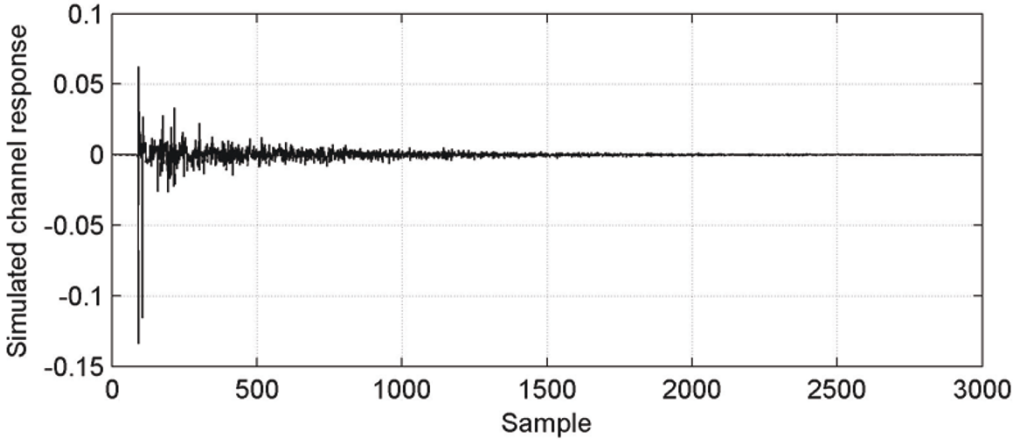
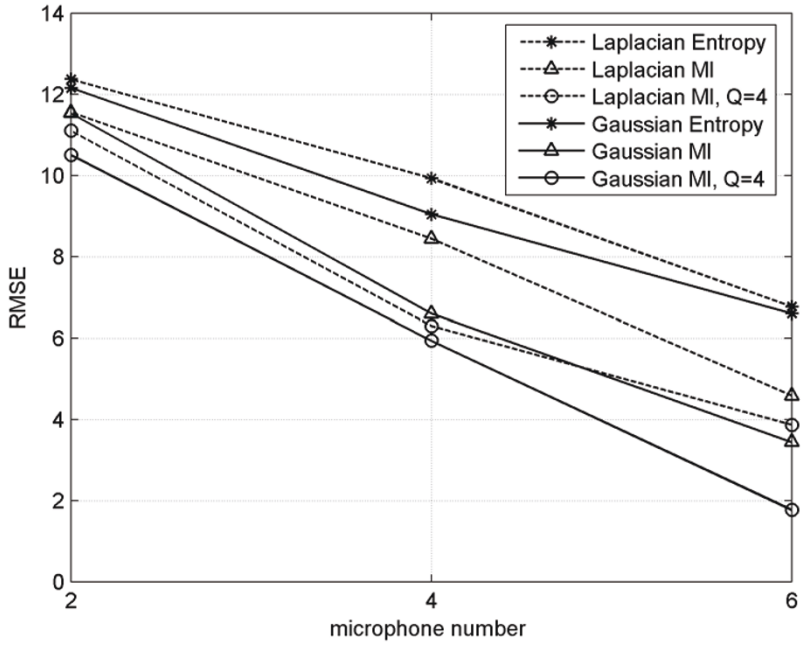
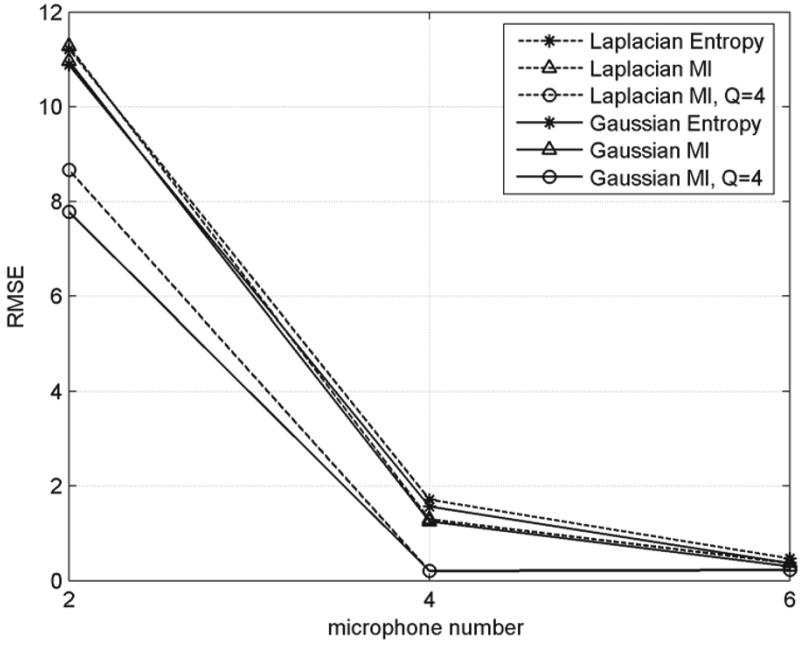



**Figure 4 One of the recorded channel responses of MARDY, $T_{60}$ = 447 ms.**




**A**




**B**

**Figure 5 RMSE versus different number of microphones for the two noise conditions**. **(a)** SNR = -5 dB, **(b)** SNR = 25 dB using the real measured room impulse responses of MARDY, $T_{60}$ = 447 ms.

the variance change of sensor outputs. Moreover, an existing idea of using modified MI to embed information about reverberation is generalized to the multiple microphones case. The effectiveness of the proposed scheme is verified by simulations for speech signals in different reverberant environments. Simulation results also demonstrate that the Gaussian distribution models the small segments of noise speech signals better than the Laplacian distribution for TDE.

### List of Abbreviations

GCC: generalized cross-correlation; HOS: higher order statistics; MCCC: multichannel cross-correlation coefficient; MI: mutual information; pdfs: probability density functions; RMSE: root mean-squared error; SOS: second-order statistics; TDE: time delay estimation.

### Acknowledgements

This work was supported by the National Natural Science Foundation of China (60772146), the National High Technology Research and Development Program of China (2008AA12Z306), the Key Project of Chinese Ministry of Education (109139), and Open Research Foundation of Chongqing Key Laboratory of Signal and Information Processing (CQKLS&IP).

### Competing interests

The authors declare that they have no competing interests.

Received: 19 February 2011 Accepted: 29 July 2011
Published: 29 July 2011

### References

1. H Wang, P Chu, Voice source localization for automatic camera pointing system in videoconferencing, in *Proceedings of IEEE ASSP Workshop on Applications of Signal Processing Audio Acoustics* (1997)
2. Y Huang, J Benesty, GW Elko, Microphone arrays for video camera steering. in *Acoustic Signal Processing for Telecommunication*, ed. by SL Gay, J Benesty, Kluwer, Norwell, MA pp. 239–259 (2000)
3. M Brandstein, D Ward, in *Microphone Arrays* (Springer, Berlin, Germany, 2001)
4. J Benesty, S Makino, J Chen, in *Speech Enhancement* (Springer-Verlag, Berlin, Germany, 2005)
5. CH Knapp, GC Carter, The generalized correlation method for estimation of time delay. IEEE Trans Acoust Speech Signal Process. **24**(4), 320–327 (1976). doi:10.1109/TASSP.1976.1162830
6. JP Ianniello, Time delay estimation via cross-correlation in the presence of large estimation errors. IEEE Trans Acoust Speech Signal Process. **30**(6), 998–1003 (1982). doi:10.1109/TASSP.1982.1163992
7. B Champagne, S Bédard, A Stéphenne, Performance of time-delay estimation in presence of room reverberation. IEEE Trans Speech Audio Process. **4**(2), 148–152 (1996). doi:10.1109/89.486067
8. J Chen, J Benesty, Y Huang, Robust time delay estimation exploiting redundancy among multiple microphones. IEEE Trans Speech Audio Process. **11**(6), 549–557 (2003). doi:10.1109/TSA.2003.818025
9. TM Cover, JA Thomas, in *Elements of Information Theory*. (Wiley, New York, 1991)
10. J Benesty, J Chen, Y Huang, Time delay estimation via minimum entropy. IEEE Signal Process Lett. **14**(3), 157–160 (2007)
11. F Talantzis, AG Constantinides, LC Polymenakos, Estimation of direction of arrival using information theory. IEEE Signal Process Lett. **12**(8), 561–564 (2005)
12. J Chen, Y Huang, J Benesty, "Time delay estimation in room acoustic environments: an overview. EURASIP J Appl Signal Process. **2006**, 1–19 (2006)
13. CE Shannon, A mathematical theory of communication. Bell Sys Tech J. **27**, 379–423 (1948)
14. S Watanabe, Information theoretical analysis of multivariate correlation. IBM J Res Dev. **4**(1), 66–82 (1960)
15. T Eltoft, T Kim, TW Lee, On the multivariate Laplace distribution. IEEE Signal Process Lett. **13**(5), 300–303 (2006)
16. S Gazor, G Zhang, Speech probability distribution. IEEE Signal Process Lett. **10**(7), 204–207 (2003). doi:10.1109/LSP.2003.813679
17. JB Allen, DA Berkley, Image method for efficiently simulating small-room acoustics. J Acoust Soc Am. **65**(4), 943–950 (1979). doi:10.1121/1.382599
18. MR Schroeder, New method for measuring reverberation. J Acoust Soc Am. **37**, 409–412 (1965). doi:10.1121/1.1909343

doi:10.1186/1687-4722-2011-3
**Cite this article as:** Wen and Wan: **Robust time delay estimation for speech signals using information theory: A comparison study.** *EURASIP Journal on Audio, Speech, and Music Processing* 2011 **2011**:3.