TIME DELAY ESTIMATION VIA MINIMUM-PHASE AND ALL-PASS COMPONENT PROCESSING

Saeed Mosayyebpour, Hannan Lohrasbipeydeh, Morteza Esmaeili and T. Aaron Gulliver

Dept. of Electrical and Computer Engineering
University of Victoria, Victoria, BC Canada
{saeedm, lohrasbi, emorteza, agullive}@ece.uvic.ca

ABSTRACT

Reverberation is a major problem for time delay estimation (TDE) in enclosed environments. In this paper, a robust TDE method based on the generalized cross-correlation (GCC) is proposed. An efficient preprocessing technique to calculate the all-pass component is introduced to improve the performance of GCC-based techniques in reverberant conditions. Performance results are given which demonstrate that the proposed approach provides better performance for a wide range of microphone locations and reverberation times. Results in real acoustic environments confirm the effectiveness of the proposed TDE method. Compared to other TDE methods, our solution has low computational complexity and can be employed in real-time applications.

Index Terms— TDE, all-pass response, minimum phase response, GCC, TDOA

1. INTRODUCTION

Time delay estimation (TDE) methods estimate the relative time difference of arrival (TDOA) between spatially separated microphones. This can be used for passive localization of the dominant speaker in applications such as locating and tracking acoustic sources in radar and sonar, camera pointing in teleconferencing, microphone array beam steering, and speech enhancement. The two major sources of degradation for this estimation are noise and reverberation. While the noise only problem has been addressed in the literature, current TDE methods can perform poorly, especially in high reverberation conditions.

To deal with the problem of reverberation, most approaches exploit the redundant information from multiple sensor pairs. However, TDE between two microphones using information from only the two microphones is still a very challenging problem. Among the TDE approaches based on information from two microphones, the most popular is the generalized cross-correlation (GCC) method [1]-[4]. The phase transform GCC (PHAT) [4] is the most robust technique among the GCC methods that perform well under noisy conditions, but they typically fail in the presence of reverberation. Several GCC-based TDE methods have been proposed to make the PHAT method more robust to reverberation such as the cepstral prefiltering technique [5], but TDE remains an unsolved problem in highly reverberant rooms.

A blind channel identification method based on eigenvalue decomposition was proposed in [6] to deal with the problem of reverberation. However, this method is ineffective with information from only two microphones, which is a serious problem in real situations. The poor performance is due to the zeros of the two channel responses being close, especially in high reverberation conditions, which leads to an ill-conditioned system that is difficult to identify. Recently, a TDE method has been proposed [7] based on adaptive inverse filtering [8]. This technique blindly estimates the inverse filters of the two channels separately by using reverberation time estimation [9], and then the TDOA between the two channels is calculated based on the estimated filters. This method performs well in reverberant environments but it has high computational complexity, and can only be used when the input signal is speech.

From the above discussion, it is clear that TDE in reverberant environments using information from only two microphones is a practical but challenging problem, particularly if the computational complexity must be kept low and the input signals are arbitrary. In this paper, we propose a robust GCC-based TDE method for reverberant environments. The effects of reverberation is mitigated using the all-pass component. An important feature of the proposed method is that it can be used in real-time applications and does not require significant data to estimate the TDOA. Conversely, the most effective method in reverberant environments [7] requires at least 20 s of input speech data to accurately estimate the TDOA. As a result, it cannot be used in practical real-time applications. The main advantages of the proposed approach over other TDE methods are: a) only two microphones are required to estimate the TDOA; b) the computational complexity is low and it can be used in real-applications; c) it is robust to reverberant conditions; and d) it can be employed with arbitrary input signals.

The remainder of this paper is organized as follows. Sec-

tion 2 presents the proposed GCC-based TDE method. Some performance results are given in Section 3, and the conclusions are given in Section 4.

2. THE PROPOSED TDE METHOD

An input signal recorded in an enclosed space $x[n]$ can be modeled as the convolution of a source signal $s[n]$ with a room impulse response (RIR) $h[n]$

$$x[n] = s[n] * h[n] = \sum_{k=0}^{N-1} s[k]h[n-k], \quad (1)$$

where N is the length of the RIR and $*$ denotes convolution. The RIR can be assumed to be time invariant, and for typical applications the noise is below perceptible levels and so can be ignored. The RIR has non-minimum phase because of the late energy and thus it can be represented by a minimum phase component $h_{min}[n]$ and an all-pass component $h_{all}[n]$ as follows

$$h[n] = h_{min}[n] * h_{all}[n]. \quad (2)$$

In the frequency domain, (2) can be written as

$$H(f) = H_{min}(f)H_{all}(f), \quad (3)$$

where $H(f)$, $H_{min}(f)$ and $H_{all}(f)$ are the fast Fourier transforms (FFTs) of $h[n]$, $h_{min}[n]$, and $h_{all}[n]$, respectively. The minimum phase component contains no poles or zeros outside the unit circle and has modulus $|H(f)| = |H_{min}(f)|$.

In general, the effects of reverberation on the minimum phase and all-pass components of the RIR are fundamentally different. Fig. 1(a) shows two RIRs between a common source position and two distinct microphone locations in a room synthesized using the image method [10]. The minimum phase and all-pass components for each RIR are shown in Figs. 1(b) and 1(c), respectively. It is clear from this figure that each minimum phase component consists of a main positive peak at the origin followed by several secondary peaks of smaller amplitudes whose envelope decays quite rapidly, and the energy is concentrated near the origin [11]. On the other hand, the all-pass response decays slower than the original RIR and the position of the first dominant positive peak corresponds to the delay of the direct path signal. Thus the all-pass component of the RIR provides location information which is useful for TDE applications. This figure also shows that the reverberation intensity, especially the early reverberation, is greatly attenuated in the all-pass component compared to the original RIR. Thus the all-pass component of the signal not only preserves the direct path delay information but also decreases the reverberation effects.

The decomposition of a signal into its minimum phase and all-pass components can be carried out using homomorphic filtering [11]. Fig. 2 shows the procedure to decompose a received input signal $x[n]$ into its minimum phase $x_{min}[n]$ and all-pass $x_{all}[n]$ components. The input signal sequence is first zero-padded and the cepstrum sequence $c_x[n]$ is determined by calculating the FFT of $x[n]$ to get $X(f)$, taking the complex logarithm of $X(f)$, and then calculating the inverse fast Fourier transform (IFFT). The complex cepstrum of the minimum phase sequence $c_x^{min}[n]$ is obtained by multiplying $c_x[n]$ with $2u[n] - \delta[n]$, where $u[n]$ and $\delta[n]$ are the unit step and Dirac delta functions, respectively. Taking the FFT and then the exponential of $c_x^{min}[n]$ gives the minimum phase component in the frequency domain, $X_{min}(f)$. The IFFT of $X_{min}(f)$ is the minimum phase component $x_{min}[n]$. Finally, the all-pass component in the frequency domain $X_{all}(f)$ is obtained by dividing $X(f)$ by $X_{min}(f)$. The IFFT of $X_{all}(f)$ is the all-pass component $x_{all}[n]$.

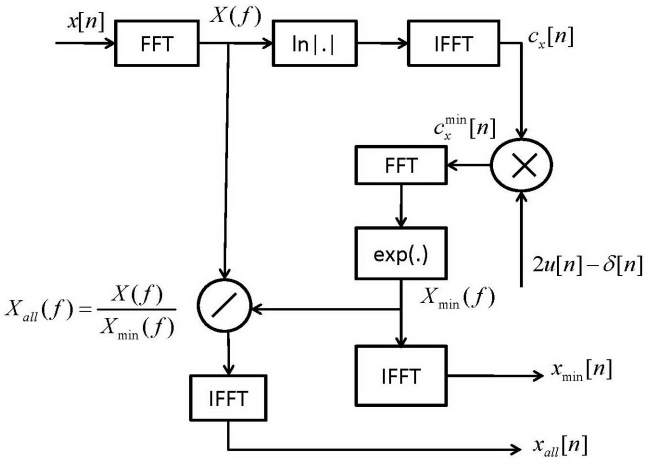

Fig. 2. Minimum phase and all-pass component decomposition using homomorphic filtering.

A block diagram of the proposed method for TDOA estimation between two spatially separated microphones is shown in Fig. 3. The input signal is first segmented with a Hamming window of size 128 ms with 50% overlapping. Then the all-pass component of each segment for each microphone is calculated to reduce the reverberation. Finally, the PHAT method is used to estimate the TDOA between the two microphones.

3. PERFORMANCE RESULTS

In this section, we evaluate the performance of the proposed TDE method in different reverberant environments. All results were obtained for a $[5 \times 4 \times 6]$ m rectangular room assuming omnidirectional microphones. Experiments were first conducted with speech utterances from four male and four female speakers (with an average duration of 4 s for each utterance), using the TIMIT database and sampled at 16 kHz. The proposed method (prop) is compared with CC [1], SCOT [2], PHAT [4], and ML [3] methods. Ten RIRs with different microphone-speaker positions having reverberation times in

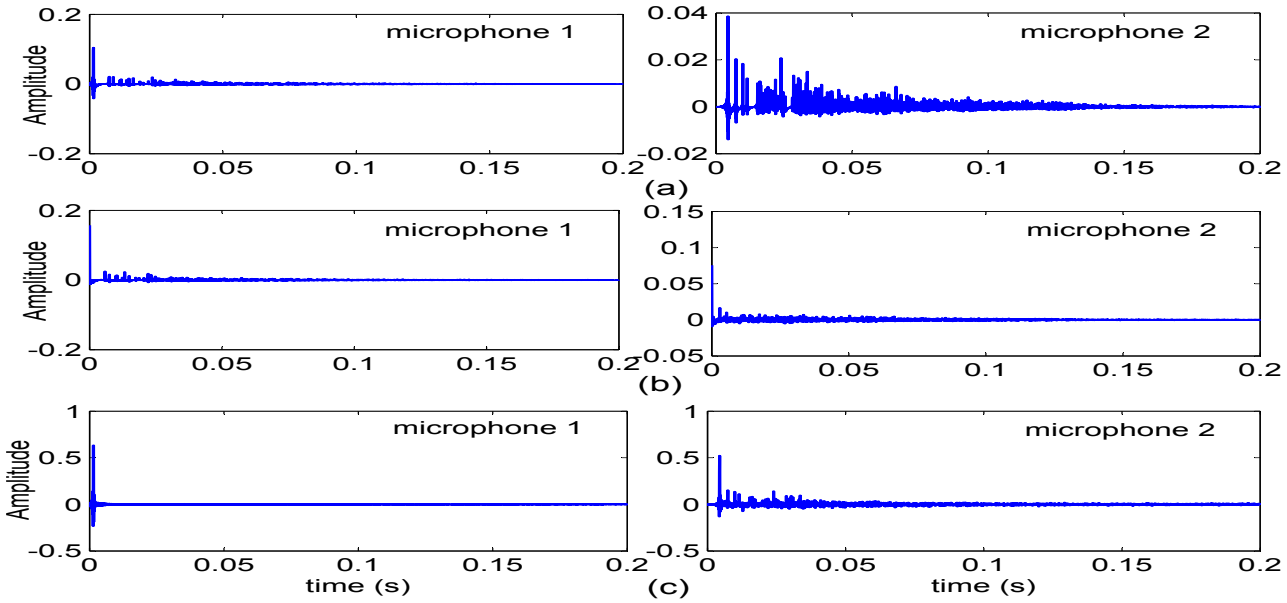

Fig. 1. Minimum-phase and all-pass components of the room impulse responses for two spatially separated microphones: (a) RIR, (b) minimum-phase component, and (c) all-pass component.

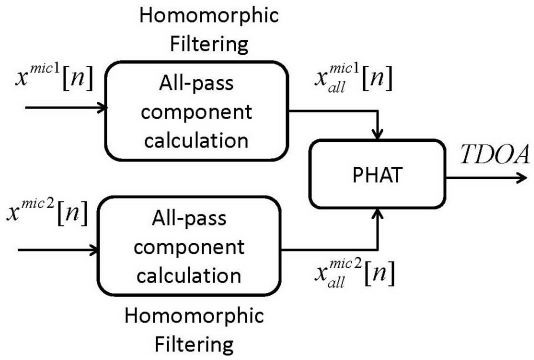

Fig. 3. Block diagram of the proposed method for estimating the TDOA between two spatially separated microphones.

the range 200 ms to 1200 ms were generated using the image method [10].

An acceptable estimate is defined as one that satisfies

$$|TDOA| \leq f_s \frac{R}{c}, \tag{4}$$

where R is the distance between microphones, f_s is the sampling rate, and $c = 340$ m/s is the velocity of sound. The TDOA estimates were obtained for the 45 possible pairs of RIRs using the four methods. The room mean square error (RMSE) of these estimates is shown in Fig. 4 (upper plot). From the figure, it is clear that our proposed method (prop) using the all-pass calculation has the best performance in different reverberant environments. This shows that using the all-pass component decreases the effects of reverberation compared to the other techniques.

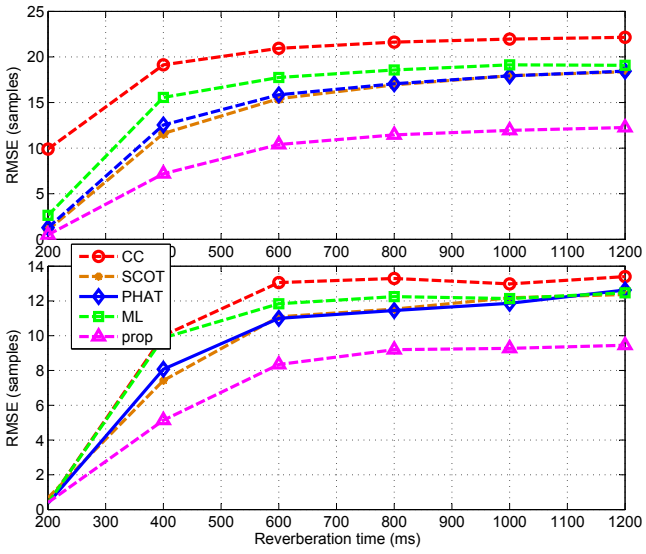

Fig. 4. TDOA estimation RMSE for 8 speech utterances (upper plot) and for a white Gaussian signal (bottom plot) in different reverberant environments using the CC [1], SCOT [2], PHAT [4], ML [3], and proposed (prop) methods.

To evaluate the proposed method with an input signal other than speech, computer-generated white Gaussian noise is employed as the input signal. This signal is convolved with the same 10 RIRs discussed above, and the averages of the TDOA estimates over the 45 possible pairs of RIRs using the

four methods are shown in Fig. 4 (bottom plot). This figure indicates that the all-pass component again has better performance than the other GCC methods. Therefore, the proposed method is effective in reverberant environments with an input signal that is not speech.

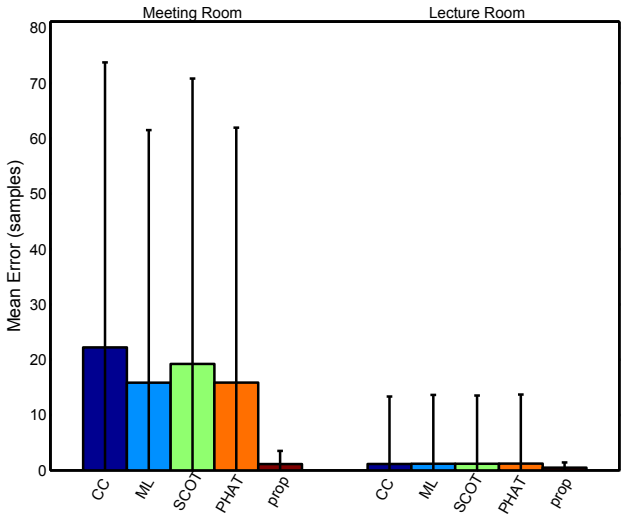

Fig. 5. Average TDOA estimation error for the CC [1], SCOT [2], PHAT [4], ML [3], and proposed (prop) methods in a real meeting room with RT60 $= 0.67$ s and a real lecture room with RT60 $= 1.23$ s. The error bars denote the standard deviation.

To evaluate the performance of the proposed method in a real environment, two different binaural RIRs from the Aachen Impulse Response (AIR) database [13]: 1) meeting room, RT60 $= 0.67$ s, and 2) lecture room, RT60 $= 1.23$ s were used. The RIRs were measured without a dummy head using only the left channel. Five microphones in different locations were used for the meeting room and six microphones for the lecture room [13]. As before, 8 clean utterances (4 female and 4 male speakers), were convolved with the measured RIRs to obtain the reverberant speech signals. The TDOA between each pair of microphones was estimated using the four conventional GCC methods and our method. The average estimation error for these methods is shown in Fig. 5 for the two rooms. The error bars denote the standard deviation. It is clear from the figure that our method has the best performance in both rooms. Thus, the proposed method is better able to deal with reverberation in real environments.

As discussed in the introduction, the proposed GCC-based method can be used in real-time TDOA applications with limited input data. Nevertheless, in order to compare the computational efficiency of the proposed method with a recently developed robust TDE method [7], TDOA estimation is performed with the same input size of 4 s for a RIR with a reverberation time of 200 ms. The average CPU time for 100 runs was determined using a laptop computer with an Intel Core 2 Duo processor T9300 at 2.5 GHz with 4 GB of RAM. The results for the proposed and PHAT methods, and the technique in [7], are shown in Table 1. It is clear that the computational complexity of the method in [7] is high. Comparing the PHAT method with the proposed method indicates that the preprocessing including the all-pass component calculation has a minimal effect on the computational complexity.

Table 1. Average CPU time required for the proposed (prop) and PHAT [4] methods, and the technique in [7].

Method	Average CPU Time
[7]	17.17
PHAT	0.19
prop	0.67

4. CONCLUSIONS

In this paper, a robust GCC-based TDE method was proposed for reverberant environments. Preprocessing is used to mitigate the effects of reverberation. It was shown that using the all-pass component of the input signal can significantly decrease the effects of reverberation, in particular early reverberation. Performance results were presented which show the effectiveness of this preprocessing in improving TDE performance in different reverberant environments. In addition, the proposed method can be employed with an input signal that is not speech, making it suitable for a wider variety of signal processing applications. The performance was also evaluated in real recorded environments to show that the proposed method is robust in real conditions. The proposed solution also has low computational complexity.

5. REFERENCES

[1] G. C. Carter, Coherence and Time Delay Estimation: An Applied Tutorial for Research, Development, Test, and Evaluation Engineers, IEEE Press, Piscataway, NJ, 1993.

[2] G. C. Carter, A. H. Nuttall, and P. G. Cable, "The smoothed coherence transform," *Proc. IEEE,* vol. 61, no. 10, pp. 1497–1498, 1973.

[3] C. H. Knapp and G. C. Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans. Acoustics, Speech, and Signal Process.,* vol. 24, no. 4, pp. 320–327, Aug. 1976.

[4] M. Omologo and P. Svaizer, "Use of the crosspower-spectrum phase in acoustic event location," *IEEE Trans. Speech Audio Process.,* vol. 5, no. 3, pp. 288–292, May 1997.

[5] A. Stéphenne and B. Champagne, "A new cepstral prefiltering technique for time delay estimation under reverberant conditions," *Sig. Proc.*, vol. 59, pp. 253–266, 1997.

[6] J. Benesty, "Adaptive eigenvalue decomposition algorithm for passive acoustic source localization," *J. Acoust. Soc. Amer.,* vol. 107, pp. 384–391, Jan. 2000.

[7] S. Mosayyebpour, A. Sayyadiyan, E. Soltan Mohammadi, A. Shahbazi, and A. Keshavarz, "Time delay estimation using one microphone inverse filtering in highly reverberant room," *Proc. IEEE Int. Conf. on Signal Acquisition and Process.*, pp. 140–144, Feb. 2010.

[8] S. Mosayyebpour, H. Sheikhzadeh, T. A. Gulliver, and M. Esmaeili, "Single-microphone LP residual skewness-based approach for inverse filtering of room impulse response," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 20, no. 5, pp. 1617–1632, July 2012.

[9] A. Keshavarz, S. Mosayyebpour, M. Biguesh, A. Gulliver, and M. Esmaeili "Speech-model based accurate blind reverberation time estimation using an LPC filter," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 20, no. 6, pp. 1884–1893, Aug. 2012.

[10] J. B. Allen, and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoustical Soc. America*, vol. 65, no. 4, pp. 943–950, 1979.

[11] A. V. Oppenheim and R. W. Schafer, *Digital Signal Processing*, Prentice Hall, Englewood Cliffs, NJ, 1975.

[12] S. Mosayyebpour, M. Esmaeili, and T. A. Gulliver, "Single-microphone early and late reverberation suppression in noisy speech," *IEEE Trans. Audio, Speech, Lang. Process.*,vol. 21, no. 2, pp. 322–335, Feb. 2013.

[13] M. Jeub, M. Schäfer, and P. Vary, "A binaural room impulse response database for the evaluation of dereverberation algorithms," in *Proc. Int. Conf. Digital Signal Process.*, pp. 1–5, July 2009.