ABSTRACT

The time delay between a signal and its received echo is a useful parameter. Acoustic Source Localization, Speech Recognition and its enhancement are some of the applications of TDE Methods. There are various methods to implement in MATLAB for time delay estimation. The reason for choosing the MATLAB as the analysis and simulation tool is that it has more flexible choices to support the simulation and is easy to do modification or data recording. The various TDE methods are Cross-Correlation (CC) method and Phase Transform (PHAT) methods falling under the Generalized Cross-Correlation Method and the Average Square Difference Function (ASDF). Various TDE methods are compared through simulations and measurements. Their simulation results are compared in terms of computation complexity, hardware implementation, precision and accuracy.

Keywords: TDE, GCC, CC, PHAT, ASDF

I. INTRODUCTION

During last few years, the calculation of time delay [1] between a signal and its echo received at a microphone in the presence of noise is used in a number of applications like radar communications, microphone array processing and speech recognition.

The received signal at a microphone can be expressed as:

\[ R_1(t) = s(t) + n_1(t), \quad (1) \]
\[ R_2(t) = s(t-d) + n_2(t), \quad (2) \]

Where \( R_1(t) \) and \( R_2(t) \) are the output signals and \( s(t) \) is the source signal, \( n_1(t) \) and \( n_2(t) \) are the two noise signals, \( d \) is the time delay between two signal. The signal and noise are assumed to be uncorrelated having zero-mean and Gaussian distribution. There are many algorithms to calculate the time delay (d). The cross-correlation (CC) is a very important method for the calculation of time delay. The CC method cross-correlates the microphone output and takes the time argument to find the peak in the output as the estimated time delay. Another method for the calculation of time delay is phase transform (PHAT). It is also called generalized cross correlation (GCC). In this method, we improve the peak detection and the time delay estimation. We use various filter or weighting functions after the cross correlation. The ASDF-based estimator gives perfect estimation in the absence of noise while the direct correlation does not.

II. TDE METHODS

In this paper we describe various methods (CC, PHAT AND ASDF) of time delay estimation. They are explained below:

Cross-correlation: In this method we calculate the cross correlation between two noisy signals. Then locate the maximum peak in the output which gives the estimated time delay. The CC method is expressed as:

\[ R_{r1r2}(\tau) = E[r_1(t)r_2(t-\tau)] \quad (3) \]
\[ D_{cc} = \arg \max (R_{r1r2}(\tau)) \quad (4) \]

Phase transform (PHAT) method: It is a method to improve the estimation of time delay. It is also called generalized cross correlation (GCC). It avoids the spreading of the peak of the correlation function. The mathematical modal can be given as:

\[ R_{r1r2}(\tau) = \int_{-\infty}^{\infty} \Psi(f) G_{r1r2}(f) \exp(j2\pi f \tau) df \quad (5) \]
\[ \Psi(f) = \frac{1}{G_{r1r2}(f)} \quad (6) \]
\[ D_{p} = \arg \max [R_{r1r2}(\tau)] \quad (7) \]

Where \( G_{r1r2}(f) \) is the cross-spectrum of the received signal, \( \Psi(f) \) is the PHAT weighting function. The simulation is carried out in simulated noisy environment. Here we take Gaussian noise for simulation and add noise to it.

Average Square Difference Function (ASDF) method: The ASDF-based estimator gives perfect estimation in the absence of noise while the direct co-relation does not [6]. The ASDF method is based on finding the position...
of the minimum error square between the two received noisy signals and considering this position value as the estimated time delay [6]. From (8), it is apparent that ASDF requires no multiplication, which is the most significant practical advantage over the other methods

\[ R_{ASDF}(\tau) = \frac{1}{N} \sum_{n=0}^{N-1} [r_1(n) - r_2(n + \tau)] \] (8)

\[ D_{ASDF} = \arg \min \{ R_{ASDF}(\tau) \} \] (9)

The two signals are plotted in figure(1). The Gaussian signal has zero mean and variance equal to one is shown in figure(2).

Then choose the time delay as 1000T and 500T seconds, where T is source signal sampling period. The cross correlation result using CC and PHAT method are given in figure (3). The x-coordinate denotes the time lag and y-coordinate denotes the resulted cross-correlations. In figure (3) it is clear that peak occurs at the actual time delay i.e. 1000T.

The cross-correlation using CC and PHAT in a simulated noisy environment with a delay of 1000T is shown in figure (3) and 500T in figure (4). The signal-to-noise ratio (SNR) is 16.3480dB and 16.337dB, T is the source signal sampling period; the sampling frequency of signal in ‘matlab’ is 7418Hz in MATLAB, T is 1.3481x10^-4 s. Among the x-coordinate is the time-lag, and the y-coordinate is the resulted cross-correlations. From Figure(3) and figure(4), it is observed that the peak occurs at the actual time delay. Again we took the value 1000T and 500T for the actual time-delay, using ASDF (Average square difference function) algorithm, the SNR is calculated as 16.3628dB and 16.3751dB respectively, where the insignificant SNR value difference is due to that the Gaussian noise is generated randomly in MATLAB. Along the y-coordinate is the error square of the two difference noisy signal instead of their cross-correlations. It is apparent from figure(5) and figure(6), that the time lag corresponding to the minimum error is the same as the actual time delay.

![Figure 1: The speech signal and the delayed signal with delay of 500T and 1000T respectively](image1.png)

![Figure 2: Gaussian noise signal with a zero mean and unit variance](image2.png)

![Figure 3: Cross-correlation using CC and PHAT in a simulated noisy environment with delay of 1000T.](image3.png)

![Figure 4: Cross-correlation using CC and PHAT in a simulated noisy environment with delay of 500T.](image4.png)

![Figure 5: ASDF output with Gaussian noise with the delay of 500T](image5.png)

![Figure 6: ASDF output with Gaussian noise with the delay of 1000T](image6.png)
Next we carried the simulation in actual environment, where the additive noises are no longer uncorrelated with zero mean. The actual noise is taken from the real environment; the noise of the working vacuum cleaner, radio broadcasting and mobile phone ring tones can be considered as actual noise in practical condition. The actual noise signal recorded is plotted in Figure (7).

![Figure (7): Actual noise signal.](image)

The signal-to-noise ratio (SNR) is 16.3480dB and 16.337dB and we chose the time delay as 1000T and 500T seconds respectively, T is the source signal sampling period; the sampling frequency of signal 'mtlb' is 7418Hz in MATLAB, T is 1.3481x10-4s again. From Figure (8) and Figure (9), it is observed that the peak occurs at the actual time delay. Again we took the value 1000T and 500T for the actual time-delay, using ASDF algorithm, the SNR is calculated as 16.3628dB and 16.3751dB respectively, where the insignificant SNR value difference is due to that the Gaussian noise is generated randomly in MATLAB. Along the y-coordinate is the error square of the two difference noisy signal instead of their cross-correlations. It is apparent from figure 10 and 11 that the time lag corresponding to the minimum error is the same as the actual time delay.

![Figure (8): Cross-correlations using CC and PHAT with actual noise with the delay of 1000T.](image)

![Figure (9): Cross-correlations using CC and PHAT with actual noise with the delay of 500T.](image)

Generally, room reverberation is considered as the main problem for TDE. Moreover, acoustic background noise [3] may further decrease the performance of time-delay estimators. The performance of TDE is always affected by the reverberation in a room. The problem becomes more challenging once room reverberations rise. In a highly reverberant room [4], all the known TDE methods [6] become unreliable and even fail. Few early studies have investigated the TDE problem in the presence of a few correlated additive echoes. However, the results obtained cannot be used to predict the effects of reverberation on the TDE performance since reverberation consists in the superposition of a very large number of closely spaced echoes, indeed, which are characterized by temporal and spatial correlation. In particular, the quantitative behavior of the estimator variance for reverberation can be explained naturally in terms of an equivalent signal-to-noise ratio (SNR), which treats the reverberant energy at the microphone output as undesirable noise. Namely, the high level of reverberation causes the low value of SNR. These problems in TDE can be combined into a situation where the SNR is low. There is degradation of performance at low SNR. This is often evidenced as the threshold phenomena in a plot of the variance of the estimates as a function of the SNR in TDE. These thresholds divide the SNR range into several non-overlapping regions. At high SNR, this is the ambiguity-free mode of operation where differential delay estimation is subjected only to local errors. For very low SNR values, observations are dominated by noise and are essentially unhelpful for TDE.

### III. RELATION BETWEEN THE SNR
LEVEL AND THE TIME DELAY ESTIMATION

In this section, the time delay is estimated under various SNR levels. The noise type is Gaussian with zero mean and variance is equal to one and then actual noise condition is taken. To study the performance of the time delay estimation, the following experiments have been set. The actual time delay value is set to 1000T and 500T, we calculate the time delay using CC, PHAT and ASDF in different SNR situations. The various SNR level is obtained by altering the noise power. The results are plotted in Figure(12) and figure(13). The x-coordinate presents the various SNR values, while the y-coordinate presents the estimated time delay.

The plots show that the estimated time delay becomes incorrect when the SNR exceeds a certain threshold. In Gaussian noise case, SNR thresholds are about -8dB, -10dB and -9.5dB for CC, PHAT and ASDF, respectively. In actual noise case SNR thresholds change to about -3dB, -7.5dB and -12dB for CC, PHAT and ASDF, respectively.

Figure (12): The relations between SNR and estimated time delay using CC, PHAT and ASDF with Gaussian noise.

Figure (13): The relation between SNR and estimated time delay using CC, PHAT and ASDF with actual noise.

IV. CONCLUSION

All the methods described here in the simulated noisy environment as well as in the actual noisy environment, ASDF (Average square difference function) should be adopted due to its simple computational steps, easy detection of delay and small SNR threshold. In the actual noisy environment, PHAT (Phase Transform) seems to be the best choice because of its perfect performance in sharpening the correlation at the correct time delay and its small SNR threshold. The Cross correlation method is not a robust approach for Time Delay Estimation when the noises are added to the signals. When the SNR is below a specified threshold, the performance of these methods degrades. When the SNR drops under a minimal level it is very difficult to estimate the exact value of time delay. So the future work should be focused on it and to develop algorithms for the source localization in Acoustic environment.

REFERENCES