Time Delay Estimation Using Weighted CPSP Function

Hongseok Kwon, Siho Kim, and Keunsung Bae

School of Electronic and Electrical Engineering
Kyungpook National University, Korea
hskwon@mir.knu.ac.kr

Abstract

A new TDE technique using the CPSP function weighted on each band is proposed. Observed and theoretical CPSP spectra are uniformly divided into several bands, and time delay having maximum cross correlation coefficient between observed and theoretical CPSP functions on each band is searched. Then the observed CPSP function is weighted according to the phase linearity on each band. A weighted CPSP function is constructed by summing all the weighted CPSP functions on each band. The time-lag that maximizes the weighted CPSP function is then determined as the finally estimated time delay. The simulation results show that the proposed method has slightly better performance than other methods in the noisy and reverberant environment.

1. Introduction

Microphone array is widely used for speech enhancement with beamforming, talker localization in videoconferencing, acoustic surveillance system and hands-free speech recognition, etc. Since the success of these applications is largely relied on the accurate time delay estimation (TDE) for a given microphone pair, TDE techniques are important. The generalized cross correlation (GCC) function is one of the most popular techniques to estimate time delay [1]. Time delay estimated by GCC function is obtained as the time-lag that maximizes the cross correlation function between filtered delay estimated by GCC function is obtained as the time-lag that maximizes the GCC function as given in eq. (4). Finally time delay is then obtained by picking up the time-lag that maximizes the weighted CPSP function.

This paper is organized as follows. Section 2 describes the TDE technique based on GCC function. Section 3 explains the proposed method using the weighted CPSP function. Simulation results are presented with our discussions in section 4, and finally the conclusion is given in section 5.

2. TDE Based on GCC Function

The received signal \( x_i(n) \) at the \( i \)-th microphone in the microphone pair can be modeled as the source signal convolved with room impulse response \( h_i(n) \) and contaminated by background noise \( v_i(n) \). It is denoted in eq. (1). It can be written in the spectral domain by \( N \)-point DFT as shown in eq. (2). The GCC function \( \hat{c}(m) \) is then obtained by taking the inverse DFT of the cross power spectrum \( C(k) \) filtered by a prefilter \( \Psi(k) \) as defined in eq. (3). Finally estimated time delay \( \hat{m} \) corresponds to the time-lag which maximizes the GCC function as given in eq. (4).

\[
x_i(n) = s(n) * h_i(n) + v_i(n), \quad i = 1,2
\]

\[
X_j(k) = S(k) * H_j(k) + V_j(k), \quad k = 0,\ldots,N-1
\]

\[
\hat{c}(m) = \sum_{k=0}^{N-1} \Psi(k) \cdot X_1(k) \cdot X_2^*(k)e^ {-\frac{2\pi mk}{N}}
\]

\[
\hat{m} = \left\{ \begin{array}{l}
\arg \max \hat{c}(m), \quad \arg \max \hat{c}(m) < \frac{N}{2}
\end{array} \right. \arg \max \hat{c}(m) \geq \frac{N}{2}
\]
3. TDE Using CPSP Function Weighted on Each Band

The TDE using a CPSP function finds the time-lag with maximum value of a CPSP function that is calculated for all spectra. When some part of a spectrum shows linear phase characteristics and other parts show random phase ones by noise or reverberation, it may fail to estimate exact time delay due to this random phase. Therefore, in this paper, CPSP spectra are divided into some bands, and then TDE technique using CPSP functions weighted according to the phase linearity on each band is proposed.

When an observed CPSP function on a band is highly correlated with a theoretical CPSP function that has linear phase by time delay, it has high cross correlation in that time delay. If it has random phase, all CPSP functions on the band has small amplitude. Therefore cross correlation coefficient between observed and theoretical CPSP functions can be used to weight the observed CPSP function on each band.

Eqs. from (6) to (8) represent l-th observed CPSP spectrum, l-th theoretical CPSP spectrum with linear phase by d sample delay, and l-th IDFT coefficient to get m-th value of CPSP function on b-th band, respectively.

\[
\psi(k) = \frac{1}{X_1(k) \cdot X_2^*(k)}
\]  

(5)

\[
C(b, l) = C(b \cdot L + l), \quad 0 < b < B, \quad 0 < l < L
\]  

(6)

\[
\tilde{G}(b, l, d) = e^{\frac{-2\pi (b \cdot L + l) d}{N}}, -D \leq d \leq D
\]  

(7)

\[
\tilde{W}(b, l, m) = e^{\frac{2\pi (b \cdot L + l) m}{N}}, -M \leq m \leq M
\]  

(8)

where \(D\) and \(M\) are the available time delay for a given microphone distance and an interval for calculation of cross correlation coefficient, respectively. \(B\) and \(L\) denote the number of bands up to half sampling rate and the number of frequency bins on each band, respectively. \(M\) is set to be equal to the available time delay \(D\), because the CPSP function is meaningful only in this interval. \(m\)-th value of observed CPSP function on \(b\)-th band is calculated by eq. (9), and \(m\)-th value of theoretical CPSP function with linear phase by \(d\) sample delay on \(b\)-th band is computed by eq. (10).

\[
c(b, m) = \sum_{l=0}^{L-1} \text{Re}\left\{\tilde{C}(b, l) \cdot \tilde{W}(b, l, m)\right\}
\]  

(9)

\[
g(b, d, m) = \sum_{l=0}^{L-1} \text{Re}\left\{\tilde{G}(b, l, d) \cdot \tilde{W}(b, l, m)\right\} = g(b, 0, m - d)
\]  

(10)

Cross correlation coefficient \(\gamma(b, d)\) that represents degree of correlation between observed CPSP function and theoretical CPSP function with linear phase by \(d\) sample delay on \(b\)-th band. It can be obtained by eq. (11), where \(\sigma_c(b), \sigma_g(b, d), \text{ and } \sigma_{cg}(b, d)\) are standard deviation of observed CPSP function, that of theoretical CPSP function with linear phase by \(d\) sample delay, and cross correlation on \(b\)-th band, respectively.

\[
\gamma(b, d) = \frac{\sigma_{cg}(b, d)}{\sigma_c(b) \cdot \sigma_g(b, d)}
\]  

(11)

When observed CPSP function on \(b\)-th band is correlated with theoretical CPSP function with linear phase by \(d\) sample delay on \(b\)-th band, cross correlation coefficient has maximum value in the \(d\) sample delay. Therefore time delay \(\hat{d}(b)\) with maximum cross correlation coefficient is found as shown eq. (12), which is used to obtain the final weighting factor on \(b\)-th band.

\[
\hat{d}(b) = \arg \max_d \gamma(b, d)
\]  

(12)

An example of CPSP functions depending on the phase linearity is shown in fig. 1, which has random phase on a low band and linear-like phase on a high band. If the CPSP spectrum on \(b\)-th band has linear phase, then the CPSP function has large amplitude in the interval near the delay corresponding to the linear phase. If it has random phase, all samples of CPSP function has small amplitude. Therefore standard deviation calculated in the \(M\) interval of observed CPSP function on the band with linear-like phase becomes large, which results in large standard deviation ratio \(\rho(b, d)\), the ratio between standard deviations of observed and theoretical CPSP functions as given in eq. (13). In the same manner, the standard deviation ratio on the band with random phase becomes small. This means that standard deviation ratio is related to the phase linearity of observed CPSP spectrum on the band. So final weighting factor \(f(b, d)\) is defined as the multiplication of standard deviation ratio by cross correlation coefficient like eq. (14), which is used to weight the observed CPSP function on \(b\)-th band. That is, the final weighting factor emphasizes the CPSP spectra on the band having linear-like phase, but attenuates the band having random phase.

![Figure 1: Example of CPSP functions depending on the phase linearity](image-url)
\[
\rho(b, d) = \frac{\sigma_g(b)}{\sigma_g(b, d)}
\]  
(13)

\[
\gamma(b, d) = \frac{\sigma_g(b)}{\sigma_g(b, d)}
\]  
(14)

Observed CPSP function on \(b\)-th band is weighted by a weighting factor in \(\hat{d}(b)\), and then a weighted CPSP function is obtained by summing all the weighted CPSP functions as shown in eq. (15). Finally the time delay is estimated by finding the time-lag, in which the weighted CPSP function is maximized.

\[
\hat{c}(m) = \sum_{b=0}^{B-1} f(b, \hat{d}(b)) \cdot c(b, m)
\]  
(15)

Figs. 2 and 3 show the block diagrams of the procedure explained so far. Fig. 2 explains the procedure to calculate weighting factor on \(b\)-th band, and the procedure to estimate the final time delay is shown in Fig. 3.

**Figure 2:** Procedure to calculate the weighting factor on \(b\)-th band

**Figure 3:** Procedure to estimate the final time delay in the proposed method

The proposed method was compared with the conventional GCC based TDE methods in the simulated 7m×4m×2.75m rectangular room as shown in fig. 4. Sources with 10, 40, 70 degrees were located in 3m distance from the center of a microphone pair. Room impulse responses at 20kHz sampling rate were computed by the image method [5], and a microphone is assumed to have omni-directional beampattern. Room reverberation time \(T_{60}\) was varied from 0s to 0.3s at 0.1s intervals for each source position and the corresponding reflection coefficient \(\beta\) was calculated by Eyring’s formula with eq. (16). The signal-to-noise ratio (SNR) was varied from 0dB to 40dB at 5dB intervals by adding white, zero-mean, Gaussian noise to the reverberant speech signal, which is made by convolving the original speech signal with the room impulse response. A frame size of 2048-sample-length was set by a rectangular window and moved by 1024 samples, and 4096-point FFT was done. A cubic spline interpolation was employed to increase the resolution of the GCC function.

\[
\beta = e^{-\frac{c(L_x + L_y + L_z)T_{60}}{L_xL_yL_z}}
\]  
(16)

where \(L_x\), \(L_y\), \(L_z\) are the room dimension and \(c\) is the sound velocity. The performance of the following three methods was investigated in our experiments.

CC: method using CC function with no pre-emphasis

CPSP: method using CPSP function with pre-emphasis

WCPSP: proposed method with pre-emphasis

Pre-emphasis was performed with eq. (17) to decrease the harmful effect by spectral leakage occurred due to framing by rectangular window, which may results in TDE errors. In CC method, however, pre-emphasis was not carried out because pre-emphasis attenuates the spectra in the low frequency bands, which results in reduction of signal energy and poor performance. The number of bands in the proposed method was empirically set to be 2.

\[
y_i(n) = x_i(n) - 0.95x_i(n - 1), \quad i = 1, 2
\]  
(17)

Fig. 5 shows the average anomalies of TDE for noises added 20 times at 70 degree, where anomalies means that the error between the angles corresponding to the estimated and real time delays is greater than the ±5 degree. Lower anomalies means the estimated time delay is more accurate. From fig. 5(a), corresponding to anechoic environment, CC
has the best performance in lower SNR and WCPSP is similar to CPSP. However the performance of the CC is dramatically degraded in reverberant environment as shown in figs. from 5(b) to 5(d). As you can see from these figs., WCPSP shows the best performance than other methods in all SNRs and it has maximum 7% improvement in high SNR. The experiments for other angles showed consistent results with those of 70 degrees. It has been shown that though the performance of the proposed method has slightly low in noisy and anechoic environment, but it shows better performance in the real situation where both noisy and reverberant environment. Especially, it shows better performance than the CPSP for both noisy and reverberant environment.

Figure 5: Comparison of the average anomalies of TDE depending on SNR at 70 degree; (a) Reverberation time: 0s (b) Reverberation time: 0.1s (c) Reverberation time: 0.2s (d) Reverberation time: 0.3s

5. Conclusion
To reduce the TDE error occurred by similar peaks in real and spurious time delays, the method using the CPSP functions weighed on each band has been proposed. Observed and theoretical CPSP spectra are divided into several bands, and then time delay with maximum cross correlation coefficient between observed and theoretical CPSP functions on each band is found. A weighted CPSP function is constructed by summing all the CPSP functions weighted according to the phase linearity on each band. The simulation results showed that the performance of the proposed method was superior to that of other methods in noisy and reverberant environment.

6. Acknowledgements
This work was partially supported by grant No. R01-2003-000-10242-0 from the Basic Research Program of the Korea Science & Engineering Foundation.

7. References