World Scientific Research Journal (WSRJ) ISSN: 2472-3703
www.wsr-j.org

Research on Delay Estimation Based on Speech Enhancement

Mingzhu Yan, Chunmao Tang

Research Center of Analysis and Control for Complex Systems, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

Abstract: The generalized cross-correlation algorithm is the most commonly used in the time difference of arrival (TDOA) estimation method. Aiming at the weak anti-noise performance of the traditional generalized cross-correlation algorithm, an algorithm based on LMS-PHAT TDOA estimation is proposed, and it's also verified by simulation experiments. With several common weighted generalized cross-correlation algorithm for comparison. The results show that the LMS-PHAT algorithm can also get the correct delay estimation in larger noise.
Keywords: TDOA, GCC, LMS-PHAT

1. INTRODUCTION

A microphone array is a group of microphones that are arranged in a specific way so that the spatial information can be accurately obtained [1]. The purpose of the noise suppression algorithm is to estimate the target speech signal from the observed signal that is affected by additive noise [2, 3]. There are already many single-microphone based technologies for this problem. However, the main problem with all single-microphone noise suppression algorithms is that they distort the speech signal. Despite the improvement in speech quality, but the intelligibility is reduced. However, with a microphone array, noise can be suppressed without affecting the voice signal.

Sound source localization technology is mainly divided into three categories [4, 5]: 1) controllable beamforming technology based on maximum output power; 2) estimation based on high resolution spectrum; and 3) positioning technology based on time difference of arrival (TDOA). Which TDOA technology because of the small amount of computation, suitable for real-time processing, is the most common application of positioning technology.

2. TIME DELAY ESTIMATION ALGORITHM

Time delay estimation (TDE) is the key technique in the field of speech enhancement and sound source localization. The so-called time delay refers to the time difference caused by the different transmission distances between the same-source signals received by different sensors in the sensor array. TDE technology is the use of parameter estimation and signal processing theory and methods. At present, the time delay estimation mainly uses the method of generalized cross-correlation.

The generalized cross-correlation (GCC) algorithm proposed by Knapp and Cater is the most common TDOA estimation method [6]. The TDOA estimate between the microphones can be equivalent to the maximum time interval at which the cross-correlation function between the filtered signals output by the microphones.

$$\hat{\tau}^{GCC} = arg \max_{\tau} r_{y_1 y_2}^{GCC}(p) \tag{1}$$

In the formula

$$\begin{aligned} r_{y_1 y_2}^{GCC}(p) &= F^{-1}\left[\psi_{y_1 y_2}(f)\right] \\ &= \int_{-\infty}^{\infty} \psi_{y_1 y_2}(f) e^{j2\pi f p}\, df \\ &= \int_{-\infty}^{\infty} \vartheta(f) \Phi_{y_1 y_2}(f) e^{j2\pi f p}\, df \end{aligned} \tag{2}$$

is the GCC function; $F^{-1}[.]$represents the discrete-time fourier inverse transform;

$$\Phi_{y_1 y_2}(f) = E[Y_1[f] Y_2(f)] \tag{3}$$

is the cross spectrum, and

$$Y_n(f) = \sum_k y_n(k) e^{-j2\pi f k}, \mathrm{n} = 1,2 \tag{4}$$

$\vartheta(f)$ is the frequency-domain weighting function

$$\psi_{y_1 y_2}(f) = \vartheta(f) \Phi_{y_1 y_2}(f) \tag{5}$$

is the generalized cross spectrum.

The frequency domain weighting function $\vartheta(f)$ has many different choices, corresponding to different GCC methods. We will analyze several of the common time delay estimation algorithm in the following.

In this simulation experiment, a purer speech signal is used to construct a speech signal with delay of 0.1s. White gaussian noise and two signals received by the microphone. We take the general cross-correlation like Roth, Scot, Phat, which is the commonly used generalized cross-correlation weighting function, and compare the merits of performance in different signal to noise ratio. When SNR is 20dB, the simulation results shows as figure 1.

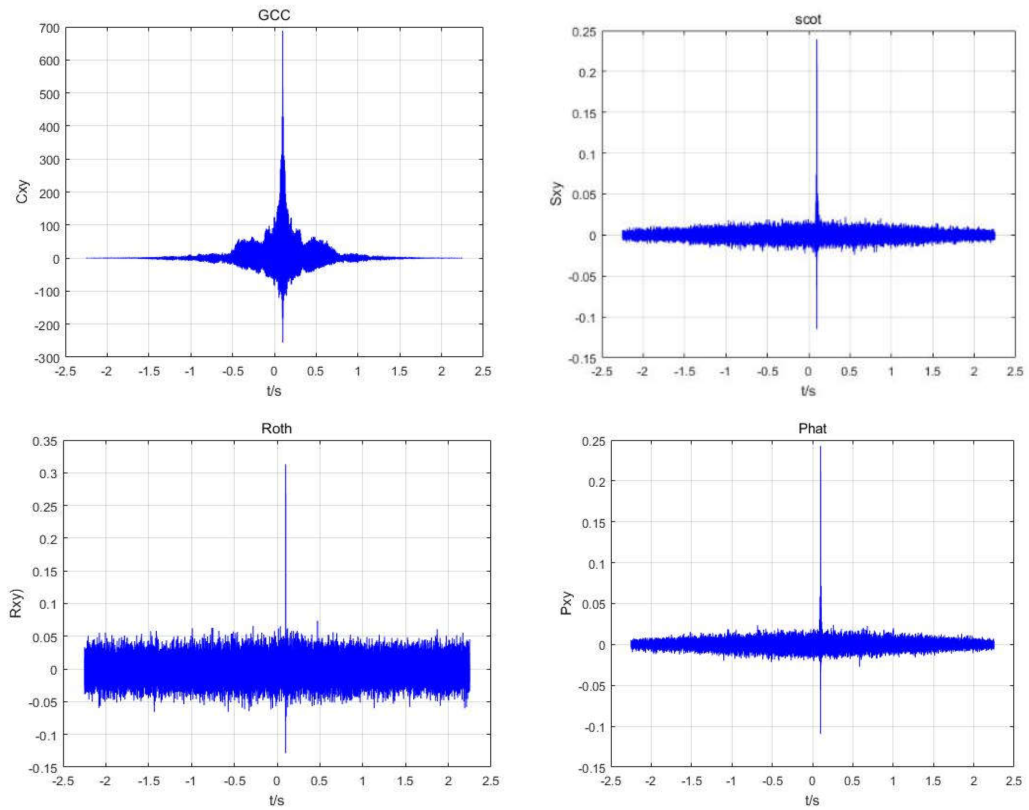


Figure 1 cross-correlation diagram when SNR is 20Db

When SNR is 0dB, the simulation results shows as figure 2.

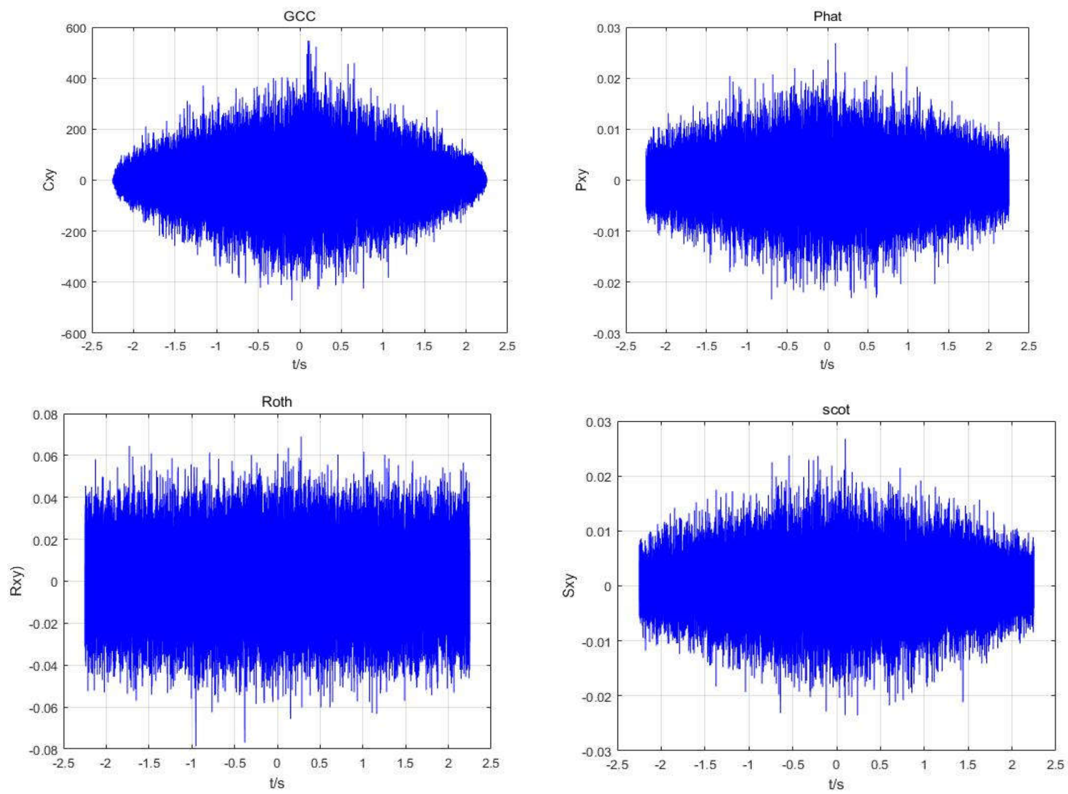


Figure 2 cross-correlation diagram when SNR is 0dB

Simulation results show that the correct delay estimation can be obtained by PHAT, Roth and Scot weights, when the signal-to-noise ratio is greater than 10 dB. When the signal to noise ratio is too small, indicating that the noise is larger, the effect of generalized correlation law will be worse, which also shows the scope of application of generalized correlation law. Overall, PHAT, Scot weighted noise immunity is stronger than Roth weighted noise immunity. Considering the impact of noise on the estimation of delay, the algorithm improvement starts with noise reduction.

3. LMS-PHAT TIME DELAY ESTIMATION ALGORITHM

In signal processing, it is common practice for a signal to be contaminated with additive noise to pass the signal through a filter that requires that the noise be suppressed while leaving the signal relatively unchanging. The implementation steps of LMS algorithm is following.

1) Set the initial value of the filter W (k)

$$\mathbf{W}(0) = 0, 0 < \mu < \frac{1}{\lambda_{max}} \tag{6}$$

2) Calculate the actual output of the filter estimate

$$\mathrm{y(k)} = \boldsymbol{W}^T(k)\boldsymbol{X}(k) \tag{7}$$

3) Calculate estimated error

$$e(k) = d(k) - y(k) \tag{8}$$

4) Calculate the filter coefficient at time k + 1

$$\boldsymbol{W}(k+1) = \boldsymbol{W}(k) + \mu e(k)\boldsymbol{X}(k) \tag{9}$$

5) Increase k to k + 1 and repeat steps 2) to 4).

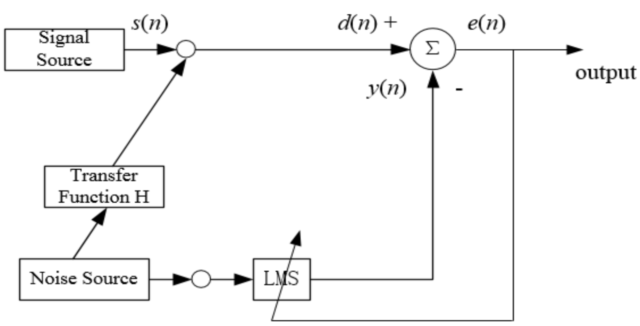
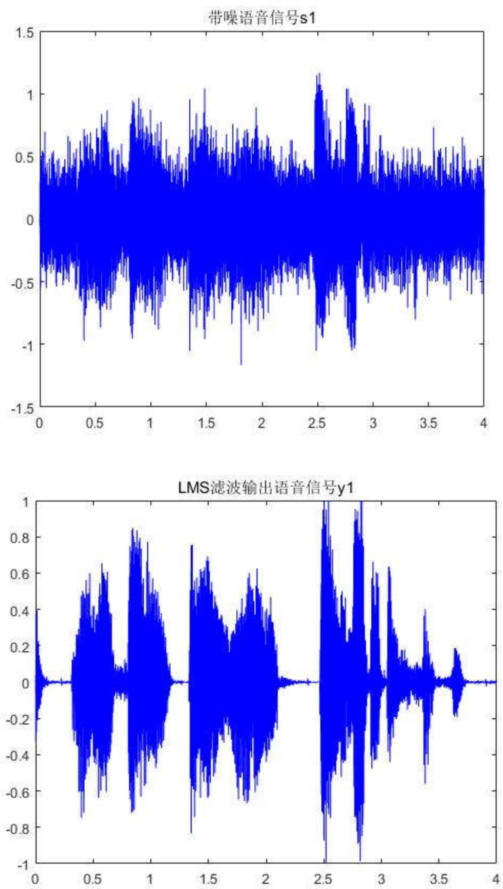


Figure 3 MIC1 and MIC2 received and LMS filtering schematic



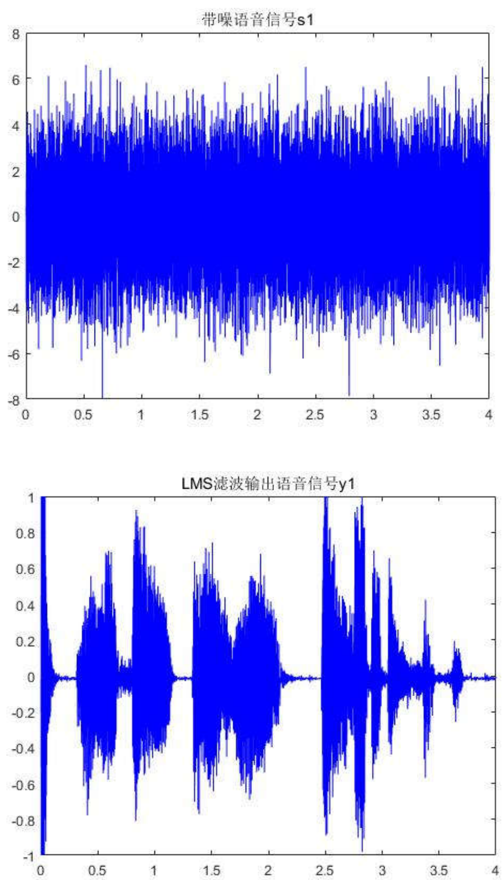


Figure 4 a) Initial SNR: SNR1 = 0B; SNR after denoising : SNR2 = 15.5097dB;

Figure 4 b) Initial SNR: SNR1=-20dB; SNR after denoising:SNR2=-2.0817 dB;

The estimated information of TDOA is expressed by the phase of the cross-spectrum rather than the amplitude. Therefore, the amplitude can be simply discarded and only the phase remains. By setting

$$\vartheta(f) = \frac{1}{|\Phi_{y_1 y_2}(f)|} \tag{10}$$

get the phase change. In this case, the generalized cross-correlation spectrum is

$$\psi_{y_1 y_2}^{PHAT}(f) = e^{-j2\pi f\tau} \tag{11}$$

It depends only on TDOA τ. By substituting (7) into (2), the ideal GCC function can be obtained.

$$r_{y_1 y_2}^{PHAT} = \int_{-\infty}^{\infty} e^{-j2\pi f(p-\tau)} df = \begin{cases} \infty, \mathrm{p} = \tau \\ 0, \mathrm{others} \end{cases} \tag{12}$$

Therefore. The PHAT method generally performs a better TDOA estimation of the speech sound source than the CC method and the SCOT method.

Combined with LMS in the low signal to noise ratio which has better noise suppression. The speech signal is subjected to LMS-based adaptive filtering, and then the output signal is subjected to Fast Fourier Transform (FFT) to PHAT the frequency domain signal. Determine the peak position corresponding to the delay value.

In this simulation experiment, a purer speech signal is also used to construct a speech signal with delay of 0.1s, add Gaussian white noise, simulate the two signals received by the microphone, and observe the peak position and noise suppression ability.

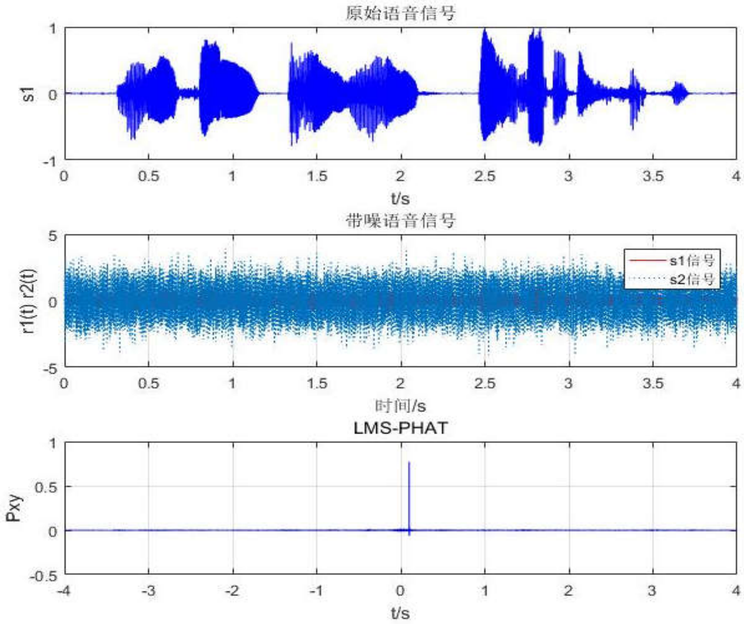


Figure 5 a) SNR1 = -20dB; LMS-PHAT cross-correlation function

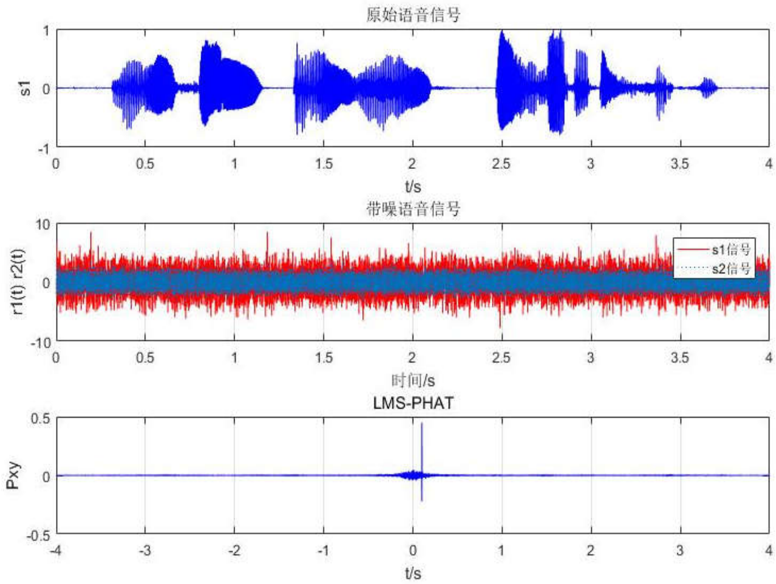


Figure 5 b) SNR1 = 0dB; LMS-PHAT cross-correlation function

From the above simulation results we can see. When SNR1 = -20dB; delay1 = 0.1; when SNR1 = 0dB; delay2 = 0.1; the simulation result shows that using the LMS-PHAT algorithm, the accurate delay value can still be obtained under the condition of noisy.

4. CONCLUSION

The time-delay-based microphone array sound source localization is a system with high accuracy and low computational complexity, which is easy to implement. It can achieve good results in the real-time location system. There are many factors that affect positioning accuracy. Such as noise, reverb and so on. In this paper, the commonly used time-delay estimation algorithms are compared, and a time-delay estimation algorithm based on LMS-PHAT-GCC is proposed. The proposed algorithm can suppress the noise effectively and can obtain the correct time when the noise is greater Delay value, improve the positioning accuracy. Due to the fact that there is reverberation in the actual environment, we should choose the appropriate algorithm to achieve more accurate positioning after considering the reverberation and noise.

ACKNOWLEDGEMENTS

This work is supported by Chongqing Research and Innovation Project of Graduate Students Grants CYS16158.

REFERENCES

[1] H.Krin and M.Viberg, “Two decades of array signal processing research: the parametric approach,” IEEE Signal Process. Mag., vol. 13, pp.67-94, July 1996.
[2] J.Benesty, J.Chen, eds., Speech Enhancement. Berlin, Germany: Springer-Verlag, 2005.
[3] P.Loizou, Speech Enhancement: Theory and Practice. Boca Raton, FL: CRC Press, 2007.
[4] Pourmohammad A, Ahadi S M. Real time high accuracy 3D PHAT based sound source localization using a simple 4-microphone arrangement [J]. IEEE Systems Journal, 2012, 6(3): 455-468.
[5] Windrow B, Fa-LongLuo. Microphone Arrays for Hearing Aids: An Overview [J]. Speech Communication, 2003, 39(1-2): 139-146.
[6] C.H.Knapp and G.C.Carter, “The generalized correlation method for estimation of time delay,” IEEE Trans.Acoust., Speech, SignalProcess., vol.24, pp.320-327, Aug.176.