A GCC Time Delay Estimation Algorithm Based on Wavelet Transform

Liyan Zhang[1] and Li Jun Zhang[2]

[1]Dalian Jiaotong University, Dalian, Liaoning Province, China
[2]Grace International Group,Division R&D ,Montreal, Canada

Summary
This paper introduces a new time delay estimation(TDE) algorithm, which applies the discrete wavelet transform and generalized cross-correlation(GCC). This algorithm first handles separately the signals received by the two microphones with wavelet transform, then under each scale, applying separately GCC to estimate the time delay in the wavelet threshold. At last, getting the final result by transforming the estimated time delay to the time field. In order to testify the stability and feasibility of the new algorithm, this paper illustrates the simulation result under different circumstances (including different signal to noise ratio and different reverberation circumstances).

Key words:
Time delay estimation. Generalized Cross Correlator. Wavelet transform.

1. Introduction

In the various speech process system based on microphone array, time delay estimation is the key problem. For example, in the teleconferencing system [1], the basic idea is to ascertain the direction and distance of the object according to the estimated time delay among different channel signals. So the accuracy of the time delay decides the precision of the speaker orientation. Another example is in the system of microphone array speech enhancement, many algorithms need to compensate the data received by the microphone with time delay, that is to say, keeping various channel signals isochronous is the precondition of the latter process, such as the microphone array speech enhancement based on wavelet transform [2].
The problem of TDE is always a very active research problem in the signal process field, and the correlative methods are more and more. There are three methods disposing these problems: GCC [3,4], cross-power-spectrum (CSP) [5]and the least-mean-square(LMS) [6,7] adaptive filter. These three methods are all for the ideal models, i.e. no noise and reverberation. For the method of CSP, only when the noise of the two input signals is not correlative, can get relatively better result. But in practice, the species of the noise is not predictable and often correlative. For the algorithm of LMS, some literature also carries out adaptive filter time delay estimation in the wavelet field, but this will lead more calculation job. The three methods mentioned above, the simplest and the most basic method is GCC. But GCC is very sensitive to the noise, and with the increase of the noise, the performance of the time delay estimation declines greatly.. Then this paper puts forward the algorithm of GCC based on wavelet transform, i.e. carrying out the GCC in the wavelet field, instead of the traditional GCC in the time field.

2. Signal Model

The speech signal arrival time delay estimation of the different microphone is very important to the problem of speech source estimation. For this problem, the simplest model is assuming there are only two microphones. Given two microphones M_0, M_1, the distance between adjacent microphone is d , the position of the voice resource is (x, y), we illustrate the situation with the Fig1.

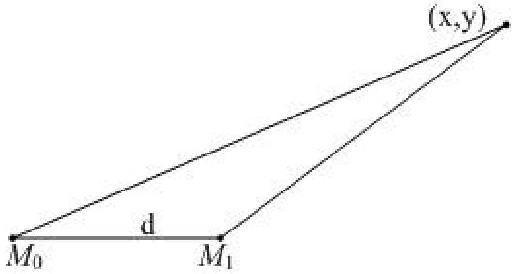

Fig 1 Signal Model

So, the signal arriving one microphone is used as the reference signal $x_1(n)$, and the original pure speech signal is . The signal $x_2(n)$ received by the second microphone is the delay of the first signal, and the delayed time is D . At the same time, the two signals are both interfered by the noise and $v_2(n)$. Then the two signals arrived to the two microphones can be expressed as follows:

$$x_1(n) = s(n) + v_1(n) \quad (1a)$$

Manuscript received June 5, 2010
Manuscript revised June 20, 2010

$$x_2(n) = s(n-D) + v_2(n) \tag{1b}$$

The most basic and prevalent method for the time delay estimation is GCC. This algorithm assumes the noise $v_1(n)$, $v_2(n)$ and the original signal $s(n)$ are all balanced random signals with zero mean value. So, the time delay of the two signals is the peak value of function $r_{x_1x_2}(n)$, which is the cross correlated function of $x_1(n)$ and $x_2(n)$. The estimated time delay can be expressed as follows:

$$\begin{aligned} d_0 &= \arg\max_{\tau} E\{x_1(n+D)x_2^*(n)\} \\ &= \arg\max_{\tau} r_{x_1x_2}(n) \end{aligned} \tag{2}$$

There are two steps when applying GCC algorithm to estimate the time delay, first calculating the cross correlative function of the two signals, and then detecting the peak value of the cross correlative function, at last can get the time delay d_0.

3. GCC Based on the Wavelet Transform

3.1 Wavelet Transform[8]

As one useful mathematics tool, wavelet transform is widely applied in the signal process field. For one signal $x(n)$, its discrete wavelet transform can be defined as follows:

$$C_x(j,k) = 2^{-\frac{j}{2}} \sum_{n=-\infty}^{+\infty} x(n)\psi^*(2^{-j}n-k) \tag{3}$$

For the discrete signal $x(n-m)$, its discrete wavelet transform is as follows:

$$C_{x'}(j,k) = 2^{-\frac{j}{2}} \sum_{n=-\infty}^{+\infty} x(n-m)\psi^*(2^{-j}n-k) \tag{4}$$

If assume $h = n - m$, then we can get:

$$\begin{aligned} C_{x'}(j,k) &= 2^{-\frac{j}{2}} \sum_{h=-\infty}^{+\infty} x(h)\psi^*(2^{-j}h-(k-2^{-j}h)) \\ &= C_{x'}(j,k-2^{-j}h) \end{aligned} \tag{5}$$

From function (4) and (5) we can conclude: to the wavelet coefficient of discrete signals if d_j represents the point number under the jth scale in the wavelet field, then the real time delay point number in the time field is $2^j d_j$, this characteristics is also based on the theory of wavelet transform GCC.

3.2 GCC based on the Wavelet Transform

In order to estimate the time delay, this paper puts forward the GCC algorithm, which is based on wavelet transform. The process of this algorithm is divided into three parts, first, to the speech signals frames collected by the two microphones, using wavelet transform, and the transform is fulfilled under several scales. Second, to the two wavelet coefficients corresponding to each frame, carrying out GCC estimation, then get the delay under each scale in the wavelet field. At last, transform the delay of wavelet field into the delay of time field, and assuming the time field delay is d_0, and d_j is the delay under the jth scale in the wavelet field. The corresponding relationship between d_0 and d_j can be expressed as follows according to the wavelet transform characteristics:

$$d_0 = 2^j \cdot d_j \tag{6}$$

Thus, under each scale, can all get the corresponding time field time delay, and then from the average value of these delay, can get the last delay needed.

4. The Simulation Result

In the simulation experiment, this paper choose sym6 wavelet as the mother function of the wavelet, and the frame length is 80 sampling points, besides compared the two time delay estimation algorithms put forwarded in this paper: GCC based on wavelet transform and LMS adaptive filter based on wavelet transform. In order to compare the performance of the system, defining one variable as MSD, which indicates the deviation degree between the estimated time delay and the real time delay. The smaller the value, the more exact is the time delay estimation. In addition, the size of the room for the experiment is 5m x 8m x4m, and the reflection model applied is Image model[9.10]. The reverberation time T of the room is donated separately as 0.45 second and 0.05 second.(see Fig2. and Fig3.)

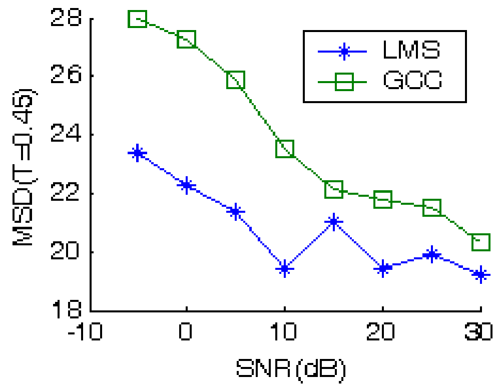

Fig2.Comparison of the MSD result between the LMS and the GCC(When the reverberation time T is 0.45 second)

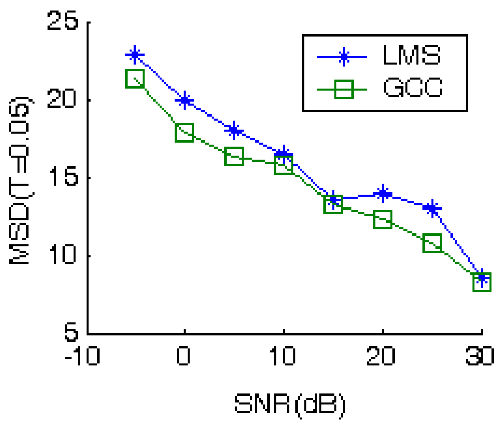

Fig2. Comparison of the MSD result between the LMS and the GCC(When the reverberation time T is 0.45 second)

Under the two reverberation circumstances mentioned below, to the input signals whose signal to noise ratio is –5dB and 30dB, working out the MSD and the curve separately as follows (curve GCC represents the GCC algorithm based on wavelet transform and the curve LMS represents the LMS adaptive filter algorithm based on the wavelet transform). From the Fig2. and Fig3., we can concludes that, under different signal to noise ratio and different degree of reflection circumstances, compared to the LMS adaptive filter algorithm based on wavelet transform, the algorithm put forward in this paper can get more small MSD, i.e. can get more accurate time delay estimation.

5. Conclusion

This paper put forward one new GCC algorithm based on wavelet transform, and after emulation experiment, testifying this algorithm is more effective on time delay estimation than the LMS adaptive filter based on wavelet transform.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant No.60172073, No.60372082 and Trans-Century Training Program Foundation for the Talents by the Ministry of Education of China.

References

[1] H. Wang, P. Chu. Voice source localization for automatic camera pointing system in videoconferencing. IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 187-190, 1997.

[2] X. Ma, F. Yin and X. Lu, A method for microphone array speech enhancement based on wavelet transform, Journal of Dalian University of technology, 43(4):511-515, Jul.2003

[3] C. Knapp, G.. Carter. The generalized correlation method for estimation of time delay. IEEE Transactions on Acoustics, Speech, and Signal Processing [see also IEEE Trans. on Signal Processing], 24(4): 320 -327,1976.

[4] C. H. Knapp, G. C. Carter, the generalized correlation method for estimation of time delay, IEEE Trans. ASSP-24(4): 320-327, 1976.

[5] M. Omologo, P. Svaizer. Use of the crosspower-spectrum phase in acoustic event location. IEEE Transactions on Speech and Audio Processing, 5(3): 288-292, 1997.

[6] M. Omologo, P. Svaizer, Use of the cross-power -spectrum phase in acoustic event localization, IEEE Trans. on Speech and Audio Processing, 5(3):288- 292, 1997.

[7] F.A. Reed, P.L. Feintuch, N. J. Bershad, Time delay estimation using the LMS adaptive filter-static behavior, IEEE Trans. ASSP, 29(1):561-571, 1981.

[8] S.G. Mallat, A theory for multiresolution signal decomposition: the wavelet representation, IEEE Trans. on Pattern Analysis and Machine Intelligence, 11(7), pp.674 -693 , 1989.

[9] J. B. Allen and D.A.Berkley, Image method for efficiently simulating small-room acoustics, J. Acoust. Soc. Amer., 65(4): 943-950, 1979.

[10] P.M.Peterson, Simulating the response of multiple microphones to a single acoustic source in a reverberant room, J. Acoust. Soc. Amer.,80(5): 1527-1529, 1986.

Liyan Zhang received the B.S and M.S. degree in electronic information engineering from Harbin Engineering University, P.R. China, in 1998 and 2000,respectively. And received the Ph.D. Degree in School of Electronic and Information Engineering, Dalian University of Technology, in P. R. China in 2009. She worked as assistant professor at School of Electronic and Information, Dalian Jiaotong University, Dalian,China, where her research interests include the theories and application of speech/array signal processing.She published 9 journal papers and about 15 international conference papers. She was a member of ACM. And CIC of China.

Li Jun Zhang received the B.S degree in mathematics from Jilin University, P.R. China, in 1993,the M.A.Sc and Ph.D. Degree in computer engineering from Ecole Polytechnique of Montreal, University of Montreal,Montreal,Canada in 2007 and 2008,respectively. And received the Degree in School of Electronic and Information Engineering, Dalian University of Technology, in P. R. China in 2009. She works as R&D Project Manager at Geninov inc and Telecom projects management,and techical coordination.She filed 3 US patents application,published 3 book chapter, and 9 journal and conference papers.She was a student member of IEEE and ACM, a member of Polytechnique Alumni Association.