New Considerations for Accumulated ρ-Cross Power Spectrum Phase with Coherence Time Delay Estimation

Radu-Sebastian Marinescu[12], Andi Buzo[2], Horia Cucu[2], Corneliu Burileanu[2]

[1]Rohde & Schwarz Topex, Bucharest, Romania

[2]Univeristy POLITEHNICA of Bucharest, Romania

radu-sebastian.marinescu@rohde-schwarz.com, {andi.buzo, horia.cucu}@upb.ro, cburileanu@messnet.pub.ro

Abstract—**Time delay estimation (TDE) has many applications in a variety of digital signal applications. The main issues today are more or less application-dependent, because every specific utilization scenario has its own demands (related to accuracy, computational load, etc.) As a solution for this topic, in this paper we continue the evaluation of the recently proposed *accumulated ρ-cross power spectrum with coherence* TDE method. The experimental results confirm that the method is faster and more accurate than the previous separated variants. Another key finding is that the TDE based on accumulation of cross-power spectrum is at least two times more accurate as the TDE based on time domain averaging.**

Keywords-component: Time Delay Estimation; Accumulated ρ-Cross Power Spectrum with Coherence

I. Introduction

As technology evolved more and more applications demanded a solution for time delay estimation. For echo canceling, acoustics, radar and sonar localization, seismic and medical processing, pattern detection and speech enhancement, scientists are still looking for better solutions. The variety of time delay estimation (TDE) applications, implementation aspects and proper constraints inhibit the design of a unique solution. Instead, various approaches have been developed based on application specific aspects.

The numerous proposed methods are based mainly on the *generalized cross-correlation* (GCC), *least mean square* (LMS) adaptive filtering and *adaptive eigenvalue decomposition* (EVD). Each category has its advantages making it optimal for specific applications. The large family of adaptive filtering methods [1-7] achieve very high accuracy, but despite the variety of optimized variants the adaptation time it is too long in some applications. A faster solution proven to be efficient in audio applications from reverberant environment is represented by EVD [8].

But, the most popular TDE methods, which do not need any adaptation time, are based on the generalized cross-correlation, proposed in 1976 by Knapp and Carter [9]. They have also presented a particular GCC weighting function named Cps-m. Based on this work, over the time multiple variations of the GCC weighting function were proposed: ROTH and SCOT [10], Eckart [11], Phase Transform (PHAT) or Cross-power Spectrum Phase (CSP) [12, 13], Wiener [14], HT (ML) [15], *accumulated CSP* (*acc-CSP*) [16], *ρ-CSPC* [17], HB [18]. For the majority of them a review and a compering based on root mean square deviation of the estimated delay and mean value was presented in [19]. In [20] we proposed two new methods *acc-ρCSPC* and *acc-ρCSP*, which benefit of the higher accuracy which characterized the *ρ-CSPC* [17] and the lower computational load and robustness of the *acc-CSP* method [16]. *Acc-ρCSPC* and *acc-ρCSP* outperforms previous methods on computing time, because of the accumulation of cross-power spectrum phase in frequency domain. This leads to only one Inverse Discrete Fourier Transform (IDFT) for any number of accumulated frames used. Also, in [20] it is shown that the first method *(acc-ρCSPC)* generally has a higher accuracy than *(acc-ρCSP),* but in specific conditions the second method achieves practically the same accuracy as the *acc-ρCSPC* at a lower computationally load.

In this work we continue to evaluate *acc-ρCSPC* over previous methods. We show that for multiple frames estimations of the time delay, results based on accumulating cross-power spectrum in frequency have in generally at least twice the accuracy compared with a the normal results obtained by a time averaging.

This paper is organized as follows. The presentation in of the TDE problem and recently proposed solutions are included in section 2. In section 3 we provide the experiments and discussion about the results. Finally, the conclusions are reserved for section 4.

II. Time Delay Estimation and Evaluated Method

In several applications we are confronting with two (or sometimes more) signals $y_1(t)$ and $y_2(t)$, which are delayed and noisy versions of the same source signal $x(t)$. The time delay estimation tries to find the relative delay between these signals. By the time, there was proposed a large variety of approaches for TDE, but the most widely used methods are based on the cross-correlation between the two signals. In [9] it was introduced the so-called generalized cross-correlation (GCC) which adds a filtering function:

$$R_{y_1 y_2}^{g}(t) = \int_{-\infty}^{\infty} \Psi(f) \cdot G_{y_1 y_2}(f) \cdot e^{j2\pi ft} df \qquad (1)$$

where $\Psi(f)$ represents a general frequency weighting function and Gy_1y_2 is the cross-power spectrum. The introduction of the weighting function takes advantage of some characteristics of the source and noise, emphasizing different spectral information [13]. Thus, the value that maximizes the general cross-correlation function represents the estimated time delay.

A. TDE based on Cross-Power Spectrum Phase

A popular derivation of GCC is represented by the CSP, because this method does not require any a priori knowledge of noise or source, making this approach independent of the input waveform characteristics, unless signals are strictly narrowband [13]. It has a large area of applications and it was shown to be an efficient technique for time delay estimation [12, 13, 16]. The weighting function Ψ_g for CSP is computed as follows:

$$\Psi(f) = 1/\left|G_{y_1y_2}(f)\right| \tag{2}$$

In general, the time delay between two signals is estimated on an analysis window, which is split into several small frames. TDE is obtained by calculating GCC at each frame and by calculating the average over all frames. This last operation is made in the time domain. Because the response time is important in almost all TDE applications, the attention is focused on next two factors: processing time and window length. The use of a larger frame leads to a higher accuracy rate for correct estimation, but the downside is the increasing computing time. Taking into account all these circumstances, *accCSP* method [16] proposes an alternative way. This method estimates time delay by averaging CSP over all frames in the frequency domain. This way the processing time decreases because the cross-power spectrum phase is accumulated over multiple frames and, consequently, only one IFFT needs to be computed (after the last accumulation). In the frequency domain it can be expressed as follows:

$$G_{CSP}(f) = \sum_{k=1}^{K} \frac{G_{y_1y_2,k}(f)}{\left|G_{y_1y_2,k}(f)\right|}, \tag{3}$$

where K represents the number of accumulated frames.

The *acc-CSP* method proposes the *accumulation scheme* of cross power spectrum in frequency domain, increasing the computation speed. This is possible because the previous methods compute delay estimation for all of the frames from the analysis window. The final result is then obtained by the average of all previous estimated delays. In this way, for K frames, the number of total FFT operation is equal to 3xK, because two FFT are used to transform the signals from time to frequency domain, and then one IFFT is used on the cross power spectrum to return in the time domain, for each frame. Instead, the accumulation scheme needs only one IFFT, so the numbers of FFT is reduced to 2xK + 1.

Beside the reduced computational complexity, the *acc-CSP* method enhances the estimation by intrinsic integration for fixed delay during the analysis window.

In [16] it was shown that the computing time decreases for *acc-CSP* compared with CSP, at the cost of accuracy degradation. Separated from this, an accuracy improvement for the CSP method was proposed in [17]. The modified GCC weighting function $\Psi(f)$ has the following expression:

$$\Psi_{y_1y_2}(f) = \frac{1}{\left|G_{y_1y_2}(f)\right|^{\rho} + \min\left[\gamma^2_{y_1y_2}(f)\right]} \tag{4}$$

where $\gamma^2_{y_1y_2}(f)$ is the signal's coherence function:

$$\gamma^2_{y_1y_2}(f) = \frac{\left|G_{y_1y_2}(f)\right|^2}{G_{y_1}(f)\cdot G_{y_2}(f)} \tag{5}$$

The tuning parameter ρ (with values between 0 and 1) is a whitening parameter, which discards the non-speech portion of the CSP (below 200Hz) [17],[21]. To reduce errors for relatively small energy signals, the minimum of the coherence function was added in (5).

B. Accumulated ρ-Cross-Power Spectrum Phase Methods

By combining *accCSP* and *ρ-CSPC* methods, we provide a new one: *accumulated ρ-Cross Power Spectrum Phase with Coherence* (*acc-ρCSPC*) in was previously proposed in [20], defined as follows:

$$G_{acc-\rho CSPC}(f) = \sum_{k=1}^{K} \frac{G_{y_1y_2,k}(f)}{\left|G_{y_1y_2,k}(f)\right|^{\rho} + \min\left[\gamma^2_{y_1y_2,k}(f)\right]} \tag{6}$$

This way it is possible to take advantage of the strong points of each method. The experimental results from [20] show a better accuracy even for low signal-to-noise ratios (SNR). This approach leads to faster computational speed, comparing with previous methods, because it uses the accumulating scheme, which can provide better results in low SNR conditions for smaller frame sizes. Beside this, speech regions are emphasized from the spectrum by the whitening parameter (ρ), which reduces the impact of noise outside the speech region as well. For parts of the signal with small energy, the addition of the minimum coherence function limits the effect of a very small denominator.

A faster method is obtained if the minimum coherence function is eliminated from (6). This operation can be used in some cases where fast response is very important and where relatively small energy signals are not usually encountered. For these proposes, in [20], the *accumulated ρ-Cross Power Spectrum Phase* (*acc-ρCSP*) was defined as follows:

$$G_{acc-\rho CSP}(f)=\sum_{k=1}^{K}\frac{G_{y_1y_2,k}(f)}{\left|G_{y_1y_2,k}(f)\right|^{\rho}} \tag{7}$$

The experimental results from [20] confirm the utility of this method for in determined conditions and for proper value of ρ.

III. EXPERIMENTAL RESULTS AND DISCUSSIONS

A. Experimental Setup

In this work, we present further evaluation and discussions about *acc-ρCSPC* method recently proposed in [20]. For consistency to previous work we use almost the same experimental configuration. The *acc-ρCSPC*, *CSP*, *accCSP* and *ρCSPC* methods are implemented in Matlab. The Noizeus corpus [22] was used as the main experimental database. It contains 30 sentences, produced by three male and female speakers and sampled at 8 kHz. The sentences are corrupted using 8 different real-word noises (suburban train noise, babble, car, exhibition hall, restaurant, street, airport and train-station noise) from the AURORA database [23] at 4 different SNRs (0, 5, 10 and 15 dB).

We cover all possible combinations of noise types, resulting in of C_2^8=28 combinations. All 4 different SNR levels are used for signal pairs that are going to be aligned.

The whitening parameter ρ was set to 0.73 and the overlap factor to 25%. Since the testing conditions are the same as in [20], we do not need to perform another calibration stage.

The metric used in these experiments is the accuracy, which is defined as the ratio between the number of perfectly estimated delays and the total number of estimations performed.

B. Further evaluations for acc-ρCSPC method

In Table I we show the accuracy improvement brought by *acc-ρCSPC* compared to *CSP* and *ρCSPC*. The frame size was set to 1024 samples and we artificially introduced 5 delay values (5, 10, 25, 50 and 100 ms). Taking into count the combinations described in the previous subsection, the total number of test pairs is 28x30x4x5=16800. The evaluation was performed for 4 and 8 frames.

The proposed method uses the accumulation of the cross power spectrum over multiple frames, so accuracy results are not available for a final averaging in time domain. Instead, *CSP* and *ρCSPC* use the averaging in time and for them we cannot use the accumulation scheme.

Table I shows that *acc-ρCSPC* outperforms the previous methods from the accuracy point of view. We observe that the accuracy for *acc-ρCSPC* increases with the number of frames. The differences between *averaging in time* and *accumulated cross power spectrum in frequency domain* result from the fact that in frequency domain, the accumulation keeps the spectral information over multiple frames. In this way it maintains the correlation between the frames.

TABLE I. ACCURACY COMPARETION

No. Frames	Estimated scheme	Method accuracy [%]		
		CSP	*ρCSPC*	*acc-ρCSPC*
4	Average in time	23.0	34.0	N/A
	Accumulation scheme	N/A	N/A	92.8
8	Average in time	12.5	24.2	N/A
	Accumulation scheme	N/A	N/A	99.9

On the other hand, the accuracy for averaging in time domain decreases if more frames are used. This is explained because with the increasing number of frames, the probability of a false estimation is also increasing.

All TDE methods are affected by the SNR levels and delay variations. Fig. 1 characterizes *acc-ρCSPC* accuracy by these factors. For this case, the frame size was set to 512 samples, resulting in a frame of 64 ms. The average accuracy was computed for all sentences combinations. Fig 1 shows that the method can achieve a high accuracy rate of more than 90%, even for delays of 78% of the frame size (50 ms delay for 64 ms frame size) at 15 dB SNR. This is an important aspect because most of the GCC methods provide reasonable results for delays up to 60-70% from the frame size.

In addition, Fig. 1 shows that for delays longer than 50% of the frame size, the influence of the SNR level has stronger influence over the accuracy. For delays up to 50% of the frame size, the difference between accuracies on various levels of SNR remains almost the same.

The configuration from previous evaluation (Fig. 1) is used to compare the proposed *acc-ρCSPC* and previous *acc-CSP* methods. Fig. 2 represents the average accuracy for all 16800 sentence pairs and confirms the effectiveness of the new *acc-ρCSPC.* The notable difference between the methods is due to the proposed combination of previous *ρCSPC* and *acc-CSP* techniques. Fig. 2 shows that for a delay of 78% of the frame size (50/64 ms) the average accuracy of *acc-ρCSPC* is two times higher than the one of *acc-CSP*.

The non-monotonic characteristics from Fig. 1 and Fig. 2 are due to the small size of the database. For larger databases with much more signal combinations we expect the accuracy characteristic to change to a monotonic shape. But, even with the actual obtained characteristics it is easy to conclude about the performance of the methods.

The results suggest that the acc-*ρCSPC* is suitable for TDE applications where the accuracy of estimation and the response time are important demands.

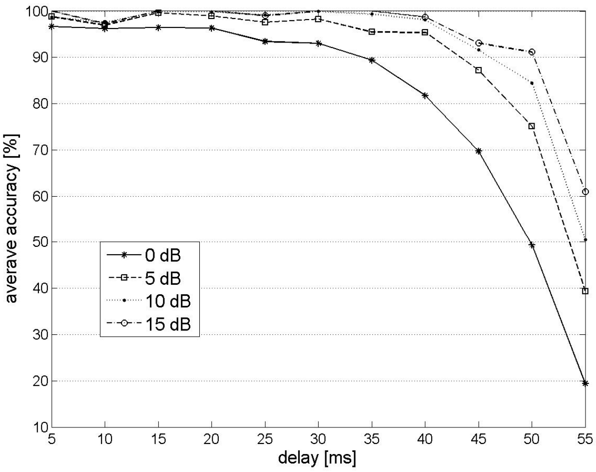

Figure 1. The influence of SNR and delay over the *acc-ρCSPC* accuracy

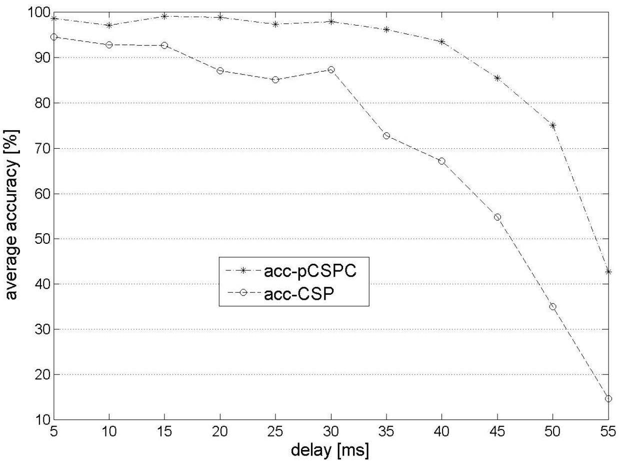

Figure 2. Comparison between acc-CSP and *acc-ρCSPC*

IV. Conclusions

In this paper we continue the evaluation of the newly proposed TDE method: *acc-ρCSPC*. The experiments were performed on the standard *Noizeus* database. The experimental results showed that *acc-ρCSPC* combination based on previous *acc-CSP* and *ρCSPC* offers higher accuracy rate and higher computational speed.

The accumulating cross power spectrum phase, which is performed in the frequency domain, obtained an accuracy which is twice higher than the accuracy of the previous methods. Moreover, the accuracy of the *acc-ρCSPC* increases with the number of frames, which was not true for the previously proposed methods.

The *acc-ρCSPC* accuracy is greater than 90% even for delays that are longer than 75% of the frame size. Its accuracy is most of the times invariant to SNR variations for delays which are smaller than 50% of the frame size.

The results from this work suggest that the acc-*ρCSPC* is suitable in TDE applications where the accuracy of estimation and the response time are important demands. It can be efficiently implemented to provide solutions for realigning noisy signals in applications such as speech enhancement, echo canceling, seismic and medical processing, radar and sonar localization, and pattern detection.

References

[1] B. Widrow, S.D. Strearns, "Adaptive signal processing", Penitence-Hall, ISBN 0130040290, USA

[2] S.N. Lin, S.J. Chern, "A new adaptive constrained LMS time delay estimation algorithm", Signal Processing, Volume 71, Issue 1, pp. 29-44, November 1998.

[3] A.W.H. Khong, P.A. Naylor, "Efficient Use Of Sparse Adaptive Filters", In Proceedings of ACSSC '06, Fortieth Asilomar Conference on Signals, Systems and Computers, article ID 10.1109/ACSSC.2006.354982, pp.1375-1379, 2006.

[4] R.A. Dyba, "Parallel Structures for Fast Estimation of Echo Path Pure Delay and Their Applications to Sparse Echo Cancellers", In Proceedings of CISS 2008, 42nd Annual Conference on Information Sciences and Systems, article ID 10.1109/CISS.2008.4558529, pp.241-245, 2008

[5] D. Hongyang, R.A. Dyba, "Efficient Partial Update Algorithm Based on Coefficient Block for Sparse Impulse Response Identification", In Proceedings of CISS 2008, 42nd Annual Conference on Information Sciences and Systems, article ID 10.1109/CISS.2008.4558527, pp.233-236, 2008.

[6] D. Hongyang, R.A. Dyba, "Partial Update PNLMS Algorithm for Network Echo Cancellation", In Proceedings of ICASSP 2009, IEEE International Conference on Acoustics, Speech and Signal Processing, article ID 10.1109/ICASSP.2009.4959837, pp.1329-1332, 2009.

[7] K. Sakhnov, E. Verteletskaya and B. Simak, "Partial Update Algorithms and Echo Delay Estimation," Communications – Scientific Journal of the University of Zilina, Zilina – Slovakia, vol. 13, no. 2, p. 14-19, 2011.

[8] J. Benesty, "Adaptive eigenvalue decomposition algorithm for passive acoustic source localization", J. Acoust. Soc. Am. Volume 107, Issue 1, pp. 384-391, 2000.

[9] C. Knapp, G.C. Carter, "The Generalized Correlation Method for Estimation of Time Delay", IEEE Transactions on Acoustics, Speech and Signal Processing, vol. 24, issue 4, pp.320-327, 1976

[10] D.H. Youn, N. Ahmed, G.C. Carter, "On the Roth and SCOTH Algorithms: Time-Domain Implementations", In Proceedings of the IEEE, vol. 71, issue 4, pp.536-538, 1983.

[11] Q. Tianshuang, W. Hongyu, "An Eckart-weighted adaptive time delay estimation method", IEEE Transactions on Signal Processing, vol. 44, issue 9, pp.2332-2335, 1996.

[12] M. Omologo, P. Svaizer, "Acoustic event localization using a crosspower-spectrum phase based technique", Proceedings of ICASSP, Australia, pp. 273-276, 1994.

[13] M. Omologo, P.Svaizer, "Use of the crosspower-spectrum phase in acoustic event location", IEEE Transactions on Speech Audio Process, pp. 288-292, 1997.

[14] V. Zetterberg, M.I. Pettersson, I. Claesson, "Comparison Between Whitened Generalized Crosscorrelation and Adaptive Filter for Time Delay Estimation", In Proceedings of TS/IEEE, OCEANS, vol. 3, article ID 10.1109/OCEANS.2005.1640117, 2005.

[15] K.W. Wilson, T. Darrell, "Learning a Precedence Effect-Like Weighting Function for the Generalized Cross-Correlation

Framework", IEEE Transactions on Audio, Speech, and Language Processing, vol. 14, issue 6, pp.2156-2164, 2006.

[16] M. Matassoni and P. Svaizer, "Efficient Time Delay Estimation Based on Cross-Power Spectrum Phase", European Signal Processing Conference (EUSIPCO), Florence - Italy, 2006.

[17] M. Shean and H. Liu, "A Modified Cross Power-Spectrum Phase Method Based on Microphone Array for Acoustic Source Localization," IEEE International Conference on System, Man and Cybernetics, San Antonio, TX, USA, pp 1286 – 1291, 2009.

[18] Y. Sun, T. Qiu, "The SCOT Weighted Adaptive Time Delay Estimation Algorithm Based on Minimum Dispersion Criterion", In Proceedings of the ICICIP Conference on Intelligent Control and Information Processing, pp.35-38, 2010.

[19] K. Sakhnov, E. Verteletskaya and B. Simak, "Echo Delay Estimation Using Algorithms Based on Cross-correlation," Journal of Convergence Information Technology, Volume 6, Number 4, pp. 1 – 11, April 2011.

[20] R.S. Marinescu, A. Buzo, H. Cucu, C. Burileanu, Fast Accurate Time Delay Estimation Based on Enhanced Accumulated Cross-Power Spectrum Phase, submitted to ICASSP 2013.

[21] D.V. Rabinkin, R.J. Renomeron, A. Dahl, J.C. French, J.L. Flanagan and M.H. Bianchi, "A DSP Implementation of Source Location Using Microphone Arrays", The Journal of the Acoustical Society of America, Volume 99, Issue 4, pp. 2510-2527, April 1996.

[22] NOIZEUS: A noisy speech corpus http://www.utdallas.edu/~loizou/speech/noizeus/

[23] AURORA database, http://www.elda.org/article52.html