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

Radu-Sebastian Marinescu[*,#], Andi Buzo[*], Horia Cucu[*], Corneliu Burileanu[*]
[#]Reasearch & Development, Rohde & Schwarz Topex
[*]SpeeD Laboratory, University Politehnica of Bucharest
Bucharest, Romania
radu-sebastian.marinescu@rohde-schwarz.com, {andi.buzo, horia.cucu}@upb.ro, cburileanu@messnet.pub.ro

***Abstract*—Time delay estimation (TDE) remains an important research issue because of its several approaches and large field of digital signal applications. 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 twice as accurate as the TDE based on time domain averaging.**

Keywords-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] achieves 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, initially proposed in 1976 by Knapp and Carter [9]. They have also presented a particular GCC weighting function named Cps-m. Based on this work, gradually, over 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 comparison, based on the root mean square deviation of the estimated delay and mean value, were presented in [19]. In [20], we proposed two new methods: *acc-ρCSPC* and *acc-ρCSP*, which benefit from 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* outperform previous methods in computation 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 computational 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 general, at least twice the accuracy compared to the normal results obtained by time averaging.

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

II. TIME DELAY ESTIMATION AND EVALUATED METHOD

In several applications, we are confronted with two (or sometimes more) signals, $y_1(t)$ and $y_2(t)$, delayed and noisy versions of the same source signal $x(t)$. The time delay estimation tries to find the relative delay between these signals. Over the years, a large variety of approaches was proposed for TDE, but the most popular methods are based on the cross-correlation between the two signals.

In 1976, Knapp and Carter introduced, in [9], the so-called generalized cross-correlation (GCC), which adds a filtering function:

$$R_{y_1y_2}^{g}(t) = \int_{-\infty}^{\infty} \Psi(f) \cdot G_{y_1y_2}(f) \cdot e^{j2\pi ft} df \qquad (1)$$

 ISBN: 978-1-61208-262-2

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. 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:

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

The TDE with CSP uses an analysis window, which is usually split in several smaller frames. Then the CSP is computed for every frame. The final result is then calculated as the average of all frame estimates. This last operation is done in time domain. Due to the fact that response time is important in almost all TDE applications, the focus should be on two factors: processing time and window length. The use of a larger frame leads to a higher accuracy rate for correct CSP estimation, but the downside is the increasing computing time.

As a solution to the above circumstances, the *acc-CSP* method was proposed by Matassoni and Svaizer [16]. The importance of the method is that it estimates time delay by averaging CSP over all frames, in the frequency domain. This way, it is shown that the processing time decreases. This is explained, because the cross-power spectrum phase is accumulated over multiple frames, remaining only one Inverse Fast Fourier Transform (IFFT) to be computed after the last accumulation. In frequency domain it can be expressed as follows:

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

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. The previous methods compute the TDE as the average of all partial estimated delays of each frame from the analysis window. This way, for K frames, the number of total Fast Fourier Transform (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 is faster because it does not calculate any partial TDEs. Because the cross-power spectrum averaging is computed in frequency domain, only one estimate will result, for any number of frames. This way, only one IFFT is needed for the final estimation and 2xK FFTs for time to frequency transformations. This leads to a total number of 2xK + 1 FFT for the *accCSP* method, which is less than the 3xK FFT needed by previous methods.

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 also shown that the computing time decreases for *acc-CSP* compared with CSP, at the cost of accuracy degradation. Separate from this, an accuracy improvement for the CSP method was proposed by Shen and Liu [17], with a modified GCC weighting function $\Psi(f)$ as in 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 signals (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

Combining *acc-CSP* and *ρ-CSPC* methods, we proposed the new *accumulated ρ-Cross Power Spectrum Phase with Coherence* (*acc-ρCSPC*) 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 both methods. Its effectiveness was proven by experimental results from [20], which showed a better accuracy even for low signal-to-noise ratios (SNR).

The new approach, summarized by (6), leads to faster computations compared to previous methods. This is due to the fact that it uses the accumulating scheme, which can also provide better results in unfavorable conditions for smaller frame sizes. Beside this, emphasis of speech regions from the spectrum is achieved by the whitening parameter (ρ), which reduces, at the same time, the impact of noise outside

 ISBN: 978-1-61208-262-2

the speech region. For parts of the signal with small energy, the addition of the minimum coherence function limits the effect of a very small denominator.

For a faster computation for applications where relatively small energy signals are not encountered, the minimum coherence function can be omitted from (6). This leads to *accumulated ρ-Cross Power Spectrum Phase* (*acc-ρCSP*) proposed in [20], as follows:

$$G_{acc-\rho CSP}(f) = \sum_{k=1}^{K} \frac{G_{y_1 y_2,k}(f)}{\left|G_{y_1 y_2,k}(f)\right|^{\rho}} \quad (7)$$

This method is faster than *acc-ρCSPC* because of the omission of the coherence calculation in equation (7). The experimental results from [20] confirm the utility of this method for proper ρ and above conditions.

III. Experimental Results and Discussions

A. Experimental Setup

In this work, we present a further evaluation and discussions for our recently proposed *acc-ρCSPC* method in [20]. The *acc-ρCSPC* and previous *CSP*, *accCSP* and *ρCSPC* methods were implemented in Matlab. The Noizeus [22] database corpus was used as the main database experiments. It contains 30 sentences (produced by three male and female speakers at 8 kHz) corrupted by 8 different real-world noises (suburban train, 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).

All possible combinations of noise types were used, which results in a number of $C_2^8 = 28$ variants and all 4 different SNR levels. The whitening parameter ρ was set to 0.73 after the calibration as in [20]. The frame overlap factor was chosen 25%.

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. This metric is the most relevant one for applications where the exact delay estimation is required.

B. Further evaluations for acc-ρCSPC method

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

TABLE I. Accuracy Comparison

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

The N/A cells from Table I correspond to the following cases. Firstly, for the proposed *acc-ρCSPC* method the accumulation of the cross power spectrum over multiple frames is used, so accuracy results are not available for a final averaging in time domain. Secondly, *CSP* and *ρCSPC* use the averaging in time and, for them, the accumulation scheme cannot be used.

Table I shows that *acc-ρCSPC* outperforms the previous methods on the accuracy tests. We observe that the accuracy for *acc-ρCSPC* increases with the number of frames, while for the others it decreases. 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. This way, it is maintaining correlation between the frames. On the other hand, the accuracy of 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.

The dependency of *acc-ρCSPC* accuracy over the SNR levels and delay variations is presented in Fig. 1. For this experiment, the frame size was set to 512 samples, resulting in a frame of 64 ms for an 8 kHz sampling frequency. The average accuracy was computed for all 16800 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 a frame size of 64 ms) at 15 dB SNR. This is an important aspect because most of the GCC methods provide reasonable results for delays up to 60-70% of the frame size.

Also, Fig. 1 shows that, for delays longer than 50% of the frame size, the influence of the SNR level affects much more the method's accuracy. For delays up to 50% of the frame size, the difference between accuracies on various levels of SNR remains almost the same.

A comparison between proposed *acc-ρCSPC* and previous *acc-CSP* methods is presented in Fig. 2. For this evaluation the same configuration as for the above experiment was used.

 ISBN: 978-1-61208-262-2

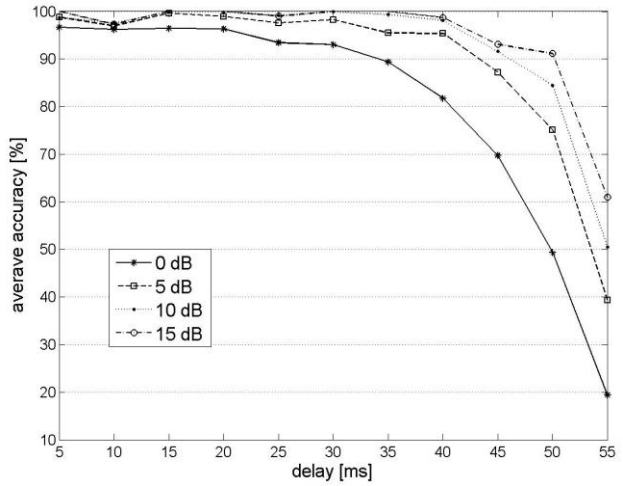

Figure 1. The influence of SNR and delay over the *acc-ρCSPC* accuracy

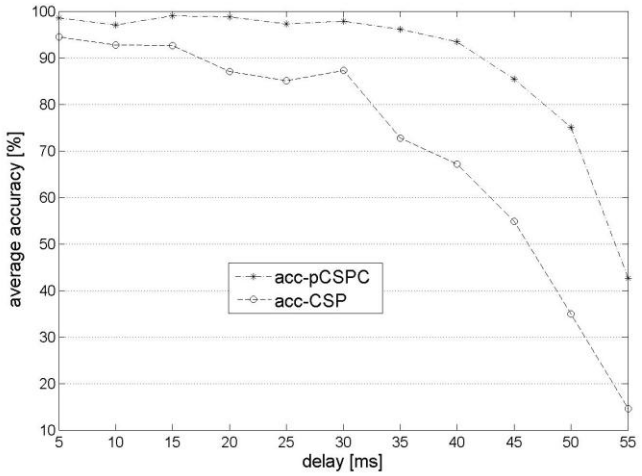

Figure 2. Comparison between acc-CSP and *acc-ρCSPC*

The results from Fig. 2 confirm the effectiveness of the *acc-ρCSPC* method. The important difference between the methods is due to the proposed mixture of previous *ρCSPC* and *acc-CSP* techniques. In Fig. 2 we observe that, for a delay of 78% from the frame size (50/64 ms), the average accuracy of *acc-ρCSPC* is twice as for *acc-CSP*.

The nonmonotonic characteristics from Fig. 1 and Fig. 2 are explained by the use of the limited 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 from this work propose the acc-*ρCSPC* for TDE applications where the accuracy of estimation and the response time are important demands.

IV. CONCLUSIONS AND FUTURE WORK

In this paper, we continued the evaluation of the proposed TDE method, *accumulated ρ-Cross Power Spectrum Phase with Coherence* (*acc-ρCSPC*). The experiments were performed on the standard *Noizeus* database. The obtained results showed that the new combination of *acc-ρCSPC*, based on previous *acc-CSP* and *ρCSPC*, offers higher accuracy rate and faster computational speed.

The accumulating cross power spectrum phase, which is performed in the frequency domain, leads to an accuracy that is more than twice than the accuracy obtained with the previous methods, which average the final results in the time domain. Also, the accuracy of the *acc-ρCSPC* increases with the number of frames, while for previous methods that use the average in time it decreases as the number of frames grows.

The *acc-ρCSPC* outperforms other methods, with an accuracy rate over 90%, even for delays that are longer than 75% of the frame size. Its' accuracy remains almost stable by the SNR variations for delays which are smaller than 50% of the frame size.

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

Future work will involve development of *acc-ρCSPC* and acc-*ρCSP* applications for the VoIP environment. Specific analysis will also involve methods characterization for different system implementation.

REFERENCES

[1] B. Widrow and S.D. Strearns, "Adaptive signal processing", Penitence-Hall, ISBN 0130040290, USA, 1985.

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

[3] A.W.H. Khong and 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, Nov. 2006, pp. 1375-1379.

[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, Mar. 2008, pp. 241-245.

[5] D. Hongyang and 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, Mar. 2008, pp. 233-236.

[6] D. Hongyang and 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, Apr. 2009, pp. 1329-1332.

[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, Apr. 2011, pp. 14-19.

 ISBN: 978-1-61208-262-2

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

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

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

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

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

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

[14] V. Zetterberg, M.I. Pettersson, and 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, Sep. 2005, pp. 2356 – 2361.

[15] K.W. Wilson and 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, Nov. 2006, pp. 2156-2164.

[16] M. Matassoni and P. Svaizer, "Efficient Time Delay Estimation Based on Cross-Power Spectrum Phase", European Signal Processing Conference (EUSIPCO), Florence - Italy, Sep 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, Oct. 2009, pp. 1286 – 1291.

[18] Y. Sun and 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, Aug. 2010, pp. 35-38.

[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, Apr. 2011, pp. 1 – 11.

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

[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, Apr. 1996, pp. 2510-2527.

[22] NOIZEUS: A noisy speech corpus http://www.utdallas.edu/~loizou/speech/noizeus/ [retrieved: Feb, 2013]

[23] AURORA database, http://www.elda.org/article52.html [retrieved: Feb, 2013]

 ISBN: 978-1-61208-262-2