

Robust time delay estimation based on non-Gaussian impulsive acoustic channel

Danilo S. Pena, Arthur D. L. Lima, Vicente A. de Sousa Jr., Luiz F. Silveira and Allan M. Martins

Abstract—The aim of this letter is to propose a new robust method for time delay estimation over impulsive noise and investigate its practical implications. The method uses a non-linear data transformation associated with the generalized cross-correlation technique. Hence, simulations and experiments show better performance than traditional methods without increasing the computational cost. Our practical experiments indicate the possibility of obtaining a correlated impulsive noise, in which the proposed method is still presenting accurate estimations.

Index Terms—direction-of-arrival estimation, acoustic noise, acoustic signal processing.

I. INTRODUCTION

The time delay estimation (TDE) for speech and audio signals is crucial to many applications and some emerging technologies, for instance, to the smart speakers, robots, integrated media, and assistive technologies. Thus, research conception of such technologies must consider realistic acoustic channel models. Notably, the channels subject to impulsive noise are more accurately characterized as non-Gaussian processes [1], [2], although many TDE methods assume implicitly or explicitly that the signal observations are Gaussian distributed. This assumption is found in generalized cross-correlation (GCC) techniques, and their different frequency domain weighting such as phase transform (GCC-PHAT) [3], Roth weighting (GCC-ROTH), and smoothed coherence factor weighting (GCC-SCOT). As a result, these classical methods experience severe degradation in the spatial resolution [4] when the Gaussian assumption does not work.

A non-Gaussian channel could be modeled with a symmetric α -stable (S α S) distribution. The S α S has heavier tails than the Gaussian distribution, giving a much better approximation to real-world audio signals [1], [2]. Although an approach based on fractional lower order statistics (FLOS) available in α -stable noises is proposed [4], this method relies on the estimating of the characteristic exponent of the model.

In this letter, we develop a straightforward and robust solution for TDE in acoustic channels subject to non-Gaussian impulsiveness noise. Particularly, we propose the use of a non-linear data-based transformation [5], which we call non-linear transformation (NLT) method, to allow the use of classical TDE methods even in channels where noise has unbounded

variance. Our main contribution is a method based on the traditional GCC technique, enabling its uses in non-gaussian impulsive noise. Moreover, the performance of the proposed architecture is evaluated through simulation and experimental data, comparing the proposed method to classical and robust techniques. Our proposed method significantly increases the estimation accuracy in the presence of non-Gaussian S α S noise. Finally, we also discuss the influence of the noise correlation between the received signals from our real measurements.

This paper is organized as follows. In Section II, we describe the signal model and the state-of-the-art solution, presenting its benefits and limitations. Our method is proposed in Section III. In Section IV, the main results are presented and discussed, comparing the performance of the proposed and the reference TDE algorithms by simulations. In Section V, we present our final remarks.

II. PROBLEM FORMULATION

Assuming microphone receivers in the acoustic field of a single speech source, the signal received in the presence of additive noise, in the instant n , can be expressed as

$$\begin{aligned} x_1(n) &= s(n) + u_1(n) \\ x_2(n) &= as(n - \tau) + u_2(n), \end{aligned} \quad (1)$$

where $s(n)$ represents the speaker's signal, τ represents the time delay, $u_1(n)$ and $u_2(n)$ are background noise sources, a is a random complex-valued gain. We assume that $u_1(n)$, and $u_2(n)$ are S α S uncorrelated processes.

As previously mentioned, we consider the FLOS as our reference case regarding a robust method to impulsive noise. The fractional lower-order covariance, which mitigates the effects of heavy-tailed noise model, is defined by [4] for an α -stable process as:

$$R_{x_1x_2}(\tau) = E[x_1(n)^{\langle A \rangle} x_2(n + \tau)^{\langle B \rangle}], \quad (2)$$

for $0 \leq A, B < \alpha/2$. Consequently, it is necessary a parametrization of the method. Due to the lack of finite variance, we use covariation [6] (instead of covariance):

$$R_{x_1x_2}(\tau) = \frac{E[x_1(n)x_2(n + \tau)^{\langle p-1 \rangle}]}{E[|x_2(n + \tau)|^p]}, \quad (3)$$

for any $1 \leq p < \alpha$. For $1 < \alpha < 2$, the α -stable distributions have finite first-order moments and all the fractional moments of order p . It is worth mentioning that the covariation is not linear regarding the second variable x_2 , and it is difficult to calculate analytically. Its estimation can be obtained by [6]

All authors are with the Graduate Program in Electrical and Computer Engineering (PPgEEC), Federal University of Rio Grande do Norte (UFRN), Natal, 59078-970, Brazil, e-mail: {danilo,arthurlima,allan,lfelipe}@dca.ufrn.br, vicente.sousa@ufrn.edu.br.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. The proof of concept simulations provided by this paper was supported by High Performance Computing Center at UFRN (NPAD/UFRN).

Digital Object Identifier: 10.14209/jcis.2020.9

$$\hat{R}_{x_1 x_2}(\tau) = \frac{\sum_{i=1}^N x_1(n) |x_2(n - \tau)|^{p-1} \text{sign}(x_2(n - \tau))}{\sum_{i=1}^N |x_2(n - \tau)|^p}. \quad (4)$$

In order to ensure the existence of second-order statistics, the generalized FLOS (GFLOS) has been shown as a smaller variance than the FLOS method [7]. It is defined as

$$R_{x_1 x_2}(\tau) = E[g(x_1(n))g(x_2(n - \tau))], \quad (5)$$

where $g(\cdot)$ is a class of non-linear transform functions. This function may be the characteristic exponent like the FLOS, or other functions as the logarithm, sign, or arctangent.

Although the maximum value of $R_{x_1 x_2}(\tau)$ corresponds to the robust estimated TDE for a particular function, it may not be possible to use the fast algorithms from Fourier convolution computing or it may require some parametrization, depending on the function chosen. The failure of the method to use optimized algorithms makes it an inappropriate method in the case with speed constraints.

III. PROPOSED METHOD

In the acoustic signals, external noises may consist of impulsive electrical noises, human-made audio noises, and uncalibrated sensors. In our proposed method, named non-linear transform (NLT), the TDE is estimated by a nonlinear transformation followed by the GCC method.

Based on the GFLOC method [7], we compute the covariance using a set of functions from a class called sigmoid functions. Those functions present essential characteristics, such as monotonically increasing, and symmetry. We compared the modulus, algebraic, gudermannian, and hyperbolic tangent functions, with the hyperbolic tangent reaching a lower RMSE than other functions.

Therefore, the non-linear data transformation of the i th signal received is a generalized sigmoid function, wherein a practical case, we use the hyperbolic tangent as a shifted and scaled version of the sigmoid, described by

$$y_i(n) = \frac{e^{x_i(n)} - e^{-x_i(n)}}{e^{x_i(n)} + e^{-x_i(n)}} = \tanh(x_i(n)). \quad (6)$$

In the NLT, we use the following covariance:

$$\hat{R}_{y_1 y_2}(\tau) = \mathcal{F}^{-1}(Y_1(k)Y_2^*(k)\phi(k)), \quad (7)$$

where $\phi(\cdot)$ is a weighting function, and $Y_i(k)$ are the frequency domain of the transformed signals $y_i(n)$. It may implement any frequency domain weighting of the cross-correlation according to a specific problem, producing the variations as NLT-PHAT, NLT-ROTH, and NLT-SCOT. Finally, the delay between the channels can be estimated by

$$\hat{\tau}_{NLT} = \arg \max_{\tau_0} (R_{y_1 y_2}(\tau)). \quad (8)$$

IV. RESULTS AND DISCUSSION

Proof-of-concept computer simulations evaluated the NLT-PHAT against GCC-PHAT, and FLOS-PHAT methods, as shown in Figures 1 and 2. We configured two simulation scenarios concerning the α -stable noise. We modeled the noise using the direct integration method for the S α S model [8] with $\alpha = 1.9$ and $\alpha = 1.3$ for low and high impulsiveness, respectively. Numerical results are provided by 10000 Monte Carlo runs.

The angle estimation accuracy is assessed by the root-mean-square error (RMSE) in degrees and the probability of resolution (PR) versus the signal-to-noise ratio (SNR). However, the infinite variance of non-Gaussian S α S processes prevents the signal-to-noise ratio to be used as a measurement of signal quality. In this letter, we use the geometric signal-to-noise ratio (GSNR) [9] instead of the SNR. The GSNR is given by

$$\text{GSNR} = \frac{1}{2C_g} \left(\frac{A}{S_0} \right)^2, \quad (9)$$

where the normalization constant $C_g = e^{C_e} \approx 1.78$ is the exponential of the Euler constant (C_e), used to ensure that GSNR corresponds to SNR when the channel is Gaussian ($\alpha = 2$); S_0 is the geometric power of a S α S random variable; and A is the root-mean-square value of the audio signal.

The PR is a performance measure based on asymptotic analysis able to evaluate the resolution capability and its statistical performance. It is known that the direction of arrival methods is given by the positions of the spectral peaks and the TDOA methods by their cross-correlation peaks. However, the PR allows us to evaluate the direction of arrival (and time delay estimation) accuracy statistically with respect to SNR. Therefore, it is useful when we are comparing two methods or evaluating methods in relation to critical conditions.

Due to comparisons with the most common GCC method, the GCC-PHAT, we use the same phase transform (PHAT) weighting function in the robust methods such as FLOS-PHAT and NLT-PHAT, given by [2]

$$\phi_{PHAT}(k) = \frac{1}{|Y_1(k)Y_2^*(k)|}. \quad (10)$$

We evaluate the FLOS-PHAT with different values of parameter p based on scenarios with different impulsiveness level ($\alpha = 1.9$ and $\alpha = 1.3$). Then, we adopt the parameter $A = B = 0.15$ in FLOS-PHAT method, since it achieves its best performance [2]. Our results corroborate with the best performance of FLOS-PHAT against the classic GCC-PHAT. However, FLOS-PHAT requires previous knowledge of the noise. On the other hand, our proposed method (NLT-PHAT) does not require any parameter, providing better performance than others over impulsive noise even for low GSNR values.

According to our previous work [1], the noise characterization is presented as an impulsive noise in outdoor and indoor scenarios using fitting for S α S distributions. The outdoor environment experiences a noise with severe impulsiveness ($\alpha = 1.3$), and the indoor environment with less impulsiveness

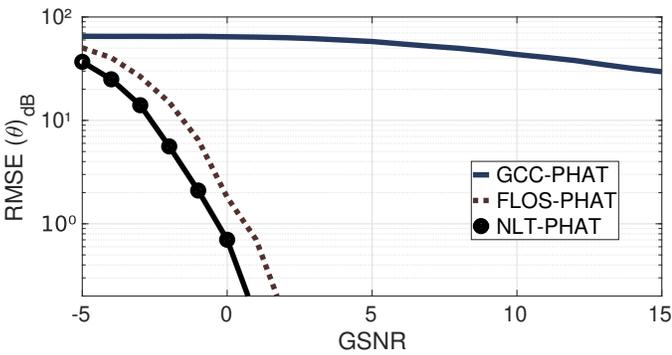


Fig. 1. Comparison between TDE methods subject to α -stable noise with $\alpha = 1.3$ (high impulsiveness).

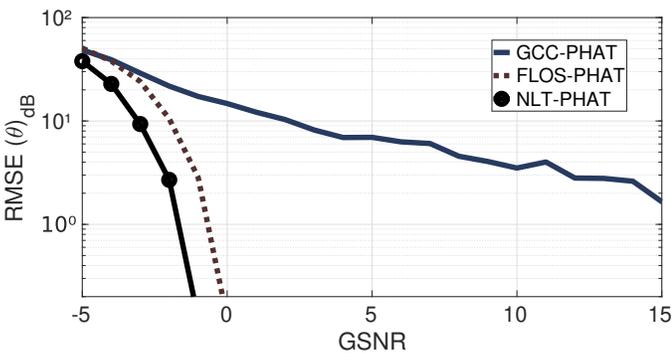


Fig. 2. Comparison between TDE methods subject to α -stable noise with $\alpha = 1.9$ (low impulsiveness)

($\alpha = 1.9$). Thus, we also present a performance evaluation of the proposed NLT-PHAT with the signal measurements.

Our experimental analysis is based on a setup with a circular array using four microphones. We collected 240000 signal samples in the indoor and outdoor scenarios. The spacing between the microphones is 5.75 cm and sampling frequency is 48 kHz. The audio source is a speech signal with 1.5 meters and 20 degrees from the first microphone. All experiments are obtained by hand and subject to limited accuracy on the order of centimeters.

Table I shows the RMSE and PR for our experimental data. We separated the evaluation in two signal windows: first, using all 240000 samples (scenarios Indoor 1 and Outdoor 1); after that, a window with 100000 samples where the speech signal (scenarios Indoor 2 and Outdoor 2) occurs. The second case was obtained manually, and the same window was used in all methods. Although the lower sampling frequency implies lower resolution, causing an increase of lowest achievable RMSE ($RMSE_{min} = 4.7$), we use the PR with $\xi = 6^\circ$ to ensure a proper evaluation. The loss of performance in the full-sized window occurs due to the absence of source signal in some regions of the window.

Analyzing the results from Table I, the FLOS-PHAT algorithm exhibits lower RMSE than the classic GCC-PHAT solution, although the FLOS-PHAT has lower resolution. On the other hand, the NLT-PHAT has the lower RMSE and higher probability of resolution in comparison to other methods. Using our proposed approach, one can reach the minimum

TABLE I
RMSE AND PROBABILITY OF RESOLUTION OF THE METHODS.

Scenario	GCC-PHAT		FLOS-PHAT		NLT-PHAT	
	RMSE	PR	RMSE	PR	RMSE	PR
Indoor 1	33.2	61%	21.3	43%	26.6	73%
Indoor 2	7.7	89%	7.7	89%	7.7	89%
Outdoor 1	24.7	52%	18.9	39%	15.8	70%
Outdoor 2	11.9	89%	9.7	89%	4.7	100%

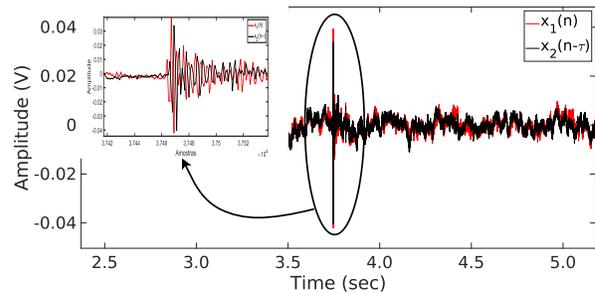


Fig. 3. Correlated impulsive noise.

RMSE and maximum PR in the outdoor scenario with an appropriate window.

A careful observation of the algorithms is performed to investigate the reason for similar performance among the methods. The signal analyzed suggests that the impulsive noise is occasionally correlated in the channels because it is produced by an acoustic source, as illustrated in Figure 3. In this case, the methods have similar performance due to the inappropriate model considered, in which the noise is uncorrelated in each signal received. Thus, the performance of non-robust estimators increases in the presence of the correlated impulsive noise. Nevertheless, these results reveal a better performance of the robust methods even for the case of an occasional presence of the correlated impulsive noise.

V. CONCLUSION

The proposed method exhibits the same low computational cost than the classical solutions presented in the literature. The solution shows some advantages compared to the robust approaches for TDE such as enhanced performance with impulsive noise (even when correlated), computational feasibility (with no need of parametrization), and the availability to use second-order statistics even in unbounded variance scenarios. Our method would be a beneficial complementary approach in the current day attempts to solve the position location issue in the non-Gaussian noise scenarios. It is a feasible solution in the trade-off between model complexity and performance.

REFERENCES

- [1] D. Pena, C. Lima, M. Dória, L. Pena, A. Martins, and V. Sousa, "Acoustic Impulsive Noise Based on Non-Gaussian Models: An Experimental Evaluation," *Sensors*, vol. 19, no. 12, p. 2827, jun 2019, doi: 10.3390/s19122827.
- [2] P. G. Georgiou, P. Tsakalides, and C. Kyriakakis, "Alpha-stable modeling of noise and robust time-delay estimation in the presence of impulsive noise," *IEEE Transactions on Multimedia*, vol. 1, no. 3, pp. 291–301, sep 1999, doi: 10.1109/6046.784467.

- [3] C. Knapp and G. Carter, "The generalized correlation method for estimation of time delay," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 24, no. 4, pp. 320–327, aug 1976, doi: 10.1109/TASSP.1976.1162830.
- [4] C. Nikias, "Joint estimation of time delay and frequency delay in impulsive noise using fractional lower order statistics," *IEEE Transactions on Signal Processing*, vol. 44, no. 11, pp. 2669–2687, 1996, doi: 10.1109/78.542175.
- [5] C. H. Lim, J. P. Lie, C. M. S. See, and A. M. Zoubir, "Modified modulus transformation for high resolution direction finding," in *European Signal Processing Conference*, 2009, pp. 1126–1130.
- [6] M. Shao and C. L. Nikias, "Signal Processing with Fractional Lower Order Moments: Stable Processes and Their Applications," *Proceedings of the IEEE*, vol. 81, no. 7, pp. 986–1010, jul 1993, doi: 10.1109/5.231338.
- [7] J.-L. Jiang and D.-F. Zha, "Generalized Fractional Lower-Order Spectrum of Alpha Stable Distribution Process," in *2008 4th International Conference on Wireless Communications, Networking and Mobile Computing*, IEEE, oct 2008, pp. 1–4, doi: 10.1109/WiCom.2008.477.
- [8] J. P. Nolan, "Numerical calculation of stable densities and distribution functions," *Communications in Statistics. Stochastic Models*, vol. 13, no. 4, pp. 759–774, jan 1997, doi: 10.1080/15326349708807450.
- [9] J. G. Gonzalez, J. L. Paredes, and G. R. Arce, "Zero-order statistics: A mathematical framework for the processing and characterization of very impulsive signals," *IEEE Transactions on Signal Processing*, vol. 54, no. 10, pp. 3839–3851, 2006, doi: 10.1109/TSP.2006.880306.



Luiz Felipe De Q. Silveira received his Bachelor Degree in Electrical Engineering from Universidade Federal da Paraíba (UFPB), Brazil, 2000, his Master Degree (2002) and Ph.D. (2006) in Electrical Engineering, from Universidade Federal de Campina Grande (UFCG), Brazil. Since 2010, he is Professor at the Department of Computing Engineering and Automation of the Universidade Federal do Rio Grande do Norte (UFRN), Brazil. In 2006 - 2010, he was also Professor in Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Norte (IFRN). He has experience in the Electrical Engineering area, with emphasis in digital signal processing, acting mainly in the following subjects: information theoretic learning, channel coding, dynamic spectrum access, industrial wireless sensor networks, energy saving in data processing and telecommunication systems.



Danilo de S. Pena received his BSc. degree in 2012 and his M.Sc. degree in 2014, where he worked on signal processing and embedded systems projects. He obtained his Ph.D. degree in signal processing in 2019, where his research involves statistical signal processing, optimization, and array signal processing. He developed a novelty method for source localization based on acoustic signals over severe conditions at the Federal University of Rio Grande do Norte (UFRN), Brazil.



Arthur D. L. Lima received the B.S. degree in computer engineering and the M.S. degree in electrical and computer engineering from the Universidade Federal do Rio Grande do Norte (UFRN), in 2009 and 2014, respectively, where he is currently pursuing the Ph.D. degree in electrical and computer engineering. His research interests include cyclostationary signal processing, data processing, parallel software efficiency, parallel software scalability, and communication systems.



Allan De M. Martins received the M.S. and Ph.D. degrees in electrical engineering. He held a post-doctoral position with the Computational Neuro-Engineering Laboratory (CNEL), Florida University, and also with the Geneva Observatory, Switzerland. He is currently an Associate Professor with the Federal University of Rio Grande do Norte (UFRN). He has experience in the areas of digital signal and image processing, machine learning, control systems, and related areas.



Vicente A. de Sousa Jr. received his BSc, MSc and Ph.D. degrees in Electrical Engineer from the Federal University of Ceará (UFC) in 2001, 2002 and 2009, respectively. Between 2001 and 2006, He developed solutions to UMTS/WLAN interworking for UFC and Ericsson of Brazil. Between 2006 and 2010, Sousa contributed to WIMAX standardization and Nokia's product as a researcher at the Institute of Technological Development (INDT). Sousa is now a professor at the Federal University of Rio Grande do Norte (UFRN), Brazil.