A SEMBLANCE BASED TDOA ALGORITHM FOR SOUND SOURCE LOCALIZATION

Aldeia, G. S. I., Crispim, A. E., Barreto, G., Alves, K., Ferraira, H., Nose-Filho, K.

guilherme.aldeia, alex.enrique, naziozeno.g, kaleb.lucas, hfsantos@aluno.ufabc.edu.br
kenji.nose@ufabc.edu.br

UFABC

Signals and Systems Laboratory

Abstract

We propose a new time difference delay of arrival technique based on the semblance multichannel coherency function for the problem of sound source localization. The proposed algorithm was tested on recordings from an Unmanned Aerial Vehicle equipped with an array of 8 microphones. Our results shown that the semblance method has proven to have a good performance, obtaining good results regardless of the ego noise even in cases where the signal-to-noise ratio was very low.

Introduction

The problem of estimating the Direction of Arrival (DOA) of a propagating wave recently has been of great interest in search and rescue scenarios [3]. For the sound source localization (SSL) problem, one of the main techniques employed is based on the time difference of arrival (TDOA), i.e., the delay that the sound arrives at several microphones disposed in different locations. However, in applications involving Unmanned Aerial Vehicle (UAV) the main problem arises from the ego noise and the fact that the sound source location and the microphones can be in movement [3].

A semblance based TDOA algorithm

The proposed algorithm is based on correcting the time-delay that the propagating wave arrives in each of the 8-channel microphones. Given a source at direction $\mathbf{k_d} \in \mathcal{R}^3$, that point towards a source parametrized by azimuth $\Theta_d \in [-\pi, \pi]$ and elevation $\Phi_d \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$, for each value of Θ_d and Φ_d in an equally spaced grid with a distance given by Δ, we correct the time-delay for each microphone and compute the semblance for each pair (θ, ϕ), obtaining a 2-dimensional array.

The time delay of a microphone at location $\mathbf{m_i}$ and a reference point at the origin $\mathbf{0} = [\mathbf{0}, \mathbf{0}, \mathbf{0}]^\mathbf{T}$ is given by:

$$\tau(\mathbf{m_i}) = -\frac{\mathbf{k_d} \cdot \mathbf{m_i}}{v}, \tag{1}$$

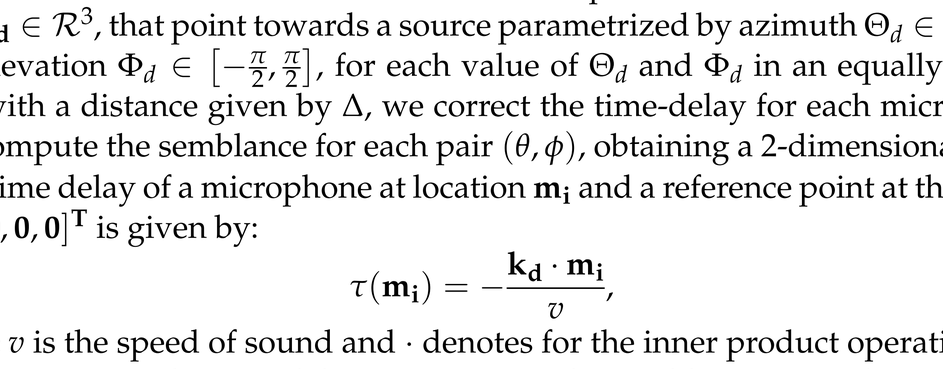
where v is the speed of sound and $\cdot$ denotes for the inner product operation.

After correcting the time delay, we compute the semblance, given by:

$$Z_d = \frac{\sum_n \left|\sum_k \hat{s}_k(n)\right|^2}{N_r \sum_n \sum_k \left|\hat{s}_k(n)\right|^2}, \tag{2}$$

where k denotes the microphones, n denotes the time samples, N_r is the number of sensors and $\hat{s}_k(n) = s_k(n - \tau_k)$ is the signal at the time sample n of the k-th microphone after correcting the delay, τ_k, for a given Θ and Φ.

The semblance measures the level of similarity between the signals [2]. The assumptions made are: the sound propagates with a plane wave-front, arrives on the n microphones in different times, and the direction that maximizes the semblance value may be the sound source direction.



(a) Original signal per channel **(b)** Aligned signal per channel

Figure 1: Frame of an 8-channel audio signal in a noiseless scenario before and after the alignment performed by selecting Θ and Φ with highest semblance value.

This approach can be improved by dividing the audio in several frames and applying the algorithm for each frame, then combining the semblance values for each frame with a method called *Max pooling* — by picking the highest value for each pair of angles (θ, ϕ) in all panels.. This enhances the signal-to-noise (SNR) in the frames that the sound source to be localized is active.

Methodology

We used three clean speech audio files (recorded with the drone in a fixed position) and a file with pure ego noise (with all motors at a speed of 70 rotations per second), combined in different SNR levels. As a measure of performance, we compute the great circle distance. All the files were provided by [1], with their respective correct azimuth (θ) and elevation (ϕ) angles.

We performed a *gridsearch* to tune the algorithm before obtaining the results. The best values for the parameters obtained are those who presented the smallest mean error for all audios. The parameters to be tuned are **overlap**, Δ and **frameSize**.

To contrast the obtained results with another robust technique, we choose the Generalized Cross Correlation PHase-Transform method, namely GCC-PHAT [3] due to similarities and moderate time consumption.

Results

The *gridsearch* values for the parameters are $\Delta = 10$, frameSize=0.064s, and 20% overlap. The GGC-PHAT parameters are: 10°, a default FFT window of 0.064s, and two different pooling methods (Max and Sum). The results are presented by the curves of Figure 2.

Our method outperforms the GCC-PHAT for audios 1 and 2. The global approach has an intermediate performance, between the GCC-PHAT with the Max and Sum pooling methods. Even with the acceptable margin of error defined as $< 10^\circ$, in most of the cases the predicted azimuth and elevation angles returned by the algorithm presented an error of $\approx 2^\circ$.

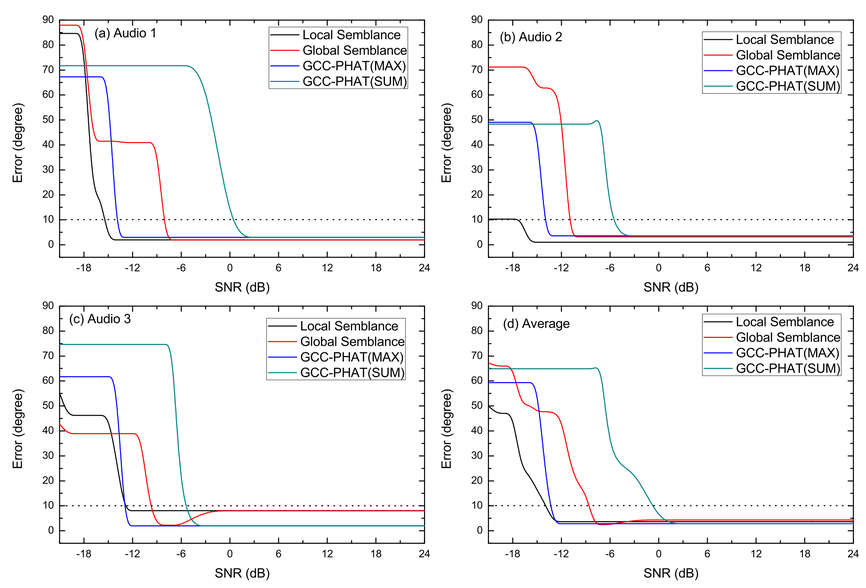


Figure 2: Comparison between our method and the best proposed method in [3]. The dashed lines denotes a threshold of 10° for the error.

Conclusions

In this paper we present a new TDOA technique based on the semblance multichannel coherency function for the problem of SSL. The obtained results showed that the method presents a good performance, making able to retrieve the source location in cases where the SNR was of -16 dB.

Acknowledgment

The authors would like to thank the Federal University of ABC for funding.

References

[1] IEEE Technical Commitee for Audio and Acoustic Signal Processing and IEEE Autonomous System Initiative. *2019 IEEE Signal Processing Cup: Search and Rescue with Drone-Embedded Souns Source Localization*, 2019.

[2] N. S. Neidell and M. T. Taner. Semblance and other coherency measures for multichannel data. *GEOPHYSICS*, 36(3):482–497, 1971.

[3] M. Strauss, P. Mordel, V. Miguet, and A. Deleforge. DREGON: Dataset and Methods for UAV-Embedded Sound Source Localization. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2018)*, pages 5735–5742, Madrid, Spain, Oct. 2018. IEEE.