Comparison of different processing for DOA estimation of an Unmanned Aerial Vehicle with few sensors

Nathan Itare, Jean-Hugh Thomas, Kosai Raoof

QUIET DRONES
Second International e-Symposium
on
UAV/UAS Noise
27th to 30th June 2022

Comparison of different processing for DOA estimation of an Unmanned Aerial Vehicle with few sensors

N. Itare: nathan.itare@univ-lemans.fr
Jean-Hugh Thomas: jean-hugh.thomas@univ-lemans.fr
Kosai Raoof: kosai.raoof@univ-lemans.fr
Laboratoire d'Acoustique de l'Université du Mans (LAUM), UMR 6613, Institut d'Acoustique-Graduate School (IA-GS), CNRS, Le Mans Université, 72085, France

Summary

During the last years, many methods have been developed and studied in the field of source localization. A particular area of interest is the tracking of UAVs (Unmanned Aerial Vehicles) because of the numerous threats that can appear near sensitive sites. Delay and Sum Beamforming (DSB) is one of the methods that can be very useful to face this challenge. Indeed, this technique has a good robustness to noise which makes it an interesting tool. Recently, some processes have been studied to enhance the performance of DSB by taking into account the signature of UAVs. Signals obtained from a microphone antenna can be filtered according to the signature of an UAV before beamforming. Beamforming can also be performed from the measured signals, then the harmonic signature can be considered using the time-frequency representation of the focused signal. A pitch tracking algorithm can provide the fundamental frequency of the signals for consideration of the UAV's signature. Another interesting approach is the Steered Response Power (SRP) which can perform well in noisy environment. The use of generalized cross correlation with different spectral weightings provides a wide range of options. This study aims at comparing the performance of beamforming with time-frequency representation and SRP-PHAT on an experimental measurement with a UAV in flight.

1. Introduction

Airspace traffic has to be regulated because of the increase of users, particularly unmanned aerial vehicle (UAV) flyers. The amount of recreational flyers is nowadays more important and UAVs can be extremely useful in other activities like deliveries or aerial imagery. These flies can

be dangerous near airplane traffic or airports. Privacy is also an issue near sensitive sites or more generally when flyers use UAV for surveillance. Several surveys have been conducted about the threats and the different methods that can be used to counter them [1], [2]. Among these methods, the use of acoustical methods enables to estimate the direction of arrival (DOA) of an UAV thanks to the sound emitted during the flight. Beamforming is often used because of its robustness to noise and its real time implementation. Considering the UAV's acoustic signature enables to enhance the performance of beamforming. This can be made using the time-frequency representation of the temporal beamformer's output [3]. Thus, energy is calculated with time-frequency bins associated with the harmonic signature of the drone. Other algorithms based on the time difference of arrival (TDOA) can estimate the DOA of the drone [4]–[7]. These algorithms exploit the generalized cross correlation (GCC) between pairs of microphone with specific weightings (SCOT [8], PHAT-β [9], …) . GCC with a phase transform also called GCC-PHAT is known to be robust to reverberation. The Steered Response Power with Phase Transformation (SRP-PHAT) makes use of GCC-PHAT to estimate the source DOA estimation. The aim of this study is to evaluate the performance of beamforming with time-frequency representation and SRP-PHAT on an experimental measure.

The work is presented as follows: Part 2 describes the two DOA estimation approaches used, Part 3 tackles the comparison between the two algorithms on the experimental case, and Part 4 concludes and gives some perspectives.

2. DOA estimation algorithms

2.1 Steered Response Power – Phase Transformation

Given a microphone antenna with N microphones, the signal $p_n(t)$ received by the nth microphone can be described with Equation (1). In this equation, $s(t)$ is the signal emitted by the source, Ω_s is the direction of the source, $h_n(\Omega_s, t)$ is the impulse response between the source and the nth microphone, and $b_n(t)$ is the noise due to the environment or the microphone system.

$$p_n(t) = s(t) * h_n(\Omega_s, t) + b_n(t). \quad (1)$$

The generalized cross correlation between microphones n and m is given by:

$$R_{nm}(\tau) = \int_{-\infty}^{+\infty} \Phi_{nm}(\mathrm{f})\, G_{nm}(f) e^{2\pi j f \tau} df, \quad (2)$$

with $\Phi_{nm}(\mathrm{f})$ a weighting of the cross-spectrum $G_{nm}(f)$ between microphones n and m. The weighting corresponding to the phase transformation is given by: $\Phi_{nm}(\mathrm{f}) = \frac{1}{|G_{nm}(f)|}$. SRP-PHAT is the extension of GCC-PHAT with multiple microphones, it is computed with:

$$P(\Omega) = \sum_{n=1}^{N} \sum_{m=1}^{N} R_{nm}(\tau_{nm}(\Omega)), \quad (3)$$

where $\tau_{nm}(\Omega)$ is the TDOA between microphones n and m. The TDOA can be calculated using a plane or spherical wave model given the direction Ω. The direction Ω which maximizes $P(\Omega)$ gives the estimation of the source DOA.

2.2 Two beamforming-based processing

2.2.1 Temporal Delay and Sum Beamforming

Delay and Sum Beamforming (DSB) enables to focus a signal in a direction Ω by calculating the TDOA between the antenna's microphones and a reference microphone. The delays can also be calculated with a plane or spherical wave model. Each signal is delayed with its own TDOA and then all the delayed signals are summed. The DOA estimate is given by the direction Ω which maximizes the energy of the focused signal.

2.2.2 Beamforming adapted to the acoustic signature

The time-frequency representation (TFR) of the focused signal in the direction Ω enables to consider the acoustic signature of the source. This TFR can be obtained with the Short Time Fourier Transform (STFT). An interesting property of UAVs' signatures is their harmonic characteristics. By using a fundamental frequency detection algorithm, it is possible to automatically select time-frequency bins from the TFR corresponding to chosen harmonics linked to the spectral properties of the signals produced by the UAV. The DOA estimation is therefore given by the direction Ω that maximizes the energy of these selected bins instead of considering all the spectral content. The algorithm used here for the fundamental frequency detection is the Spectral Harmonic Correlation (SHC) [10]. The choice of this frequency is very important to obtain a good estimate of the DOA. A procedure has been implemented for selecting the fundamental frequency among those given by SHC, which is detailed in part 3.2.

3. Comparison on an experimental trajectory

3.1 Array geometry and measurement set-up

The measurements were realized with an array of 10 BSWA Technology MPA 416 1/4 in. microphones (20 Hz-20 kHz) in the arrangement shown in Figure 1. Three different inter-microphonic spacings were used to give the [220,5 3430] Hz array bandwidth:

$$\begin{aligned} \|x_1\| = \|x_4\| = \|x_7\| &= 0.05\ m, \\ \|x_2\| = \|x_5\| = \|x_8\| &= 0.2\ \ m, \\ \|x_3\| = \|x_6\| = \|x_9\| &= 1.1\ \ m. \end{aligned} \tag{1}$$

A PXI-1036 chassis from National instruments is used with a laptop for the recording of the acoustic signals with a sampling frequency of 20 kHz. A source is located from its direction $\Omega = (\varphi, \theta)$, which is respectively the azimuth and the elevation, and with its distance r to the reference microphone (placed at the origin). During the recording, the drone starts with a circular trajectory followed by a smaller one with both a height to the ground around 14 m. The drone used is a Phantom IV from DJI.

3.2 TFR parameters

According to the study of the acoustic signature of the drone used [11], [12], two types of harmonics are present. Weak harmonics are produced by the rotation of the rotors and strong harmonics are produced by the aerodynamic phenomenon. The rotor frequency f_0 is related to the blade passing frequency f_{bp} with the number of blades N_b, $f_{bp} = N_b \times f_0$. Since there are two blades, weak harmonics (of f_0) are odd and strong harmonics are even. The spectrogram of the reference microphone is presented in Figure 2. Because of the noise present during the recording, weak harmonics are not visible in the spectrogram and f_{bp} is detected by SHC instead

of f_0 (f_{bp} detections are visible with red points in Figure 2). In order to select the time-frequency bins

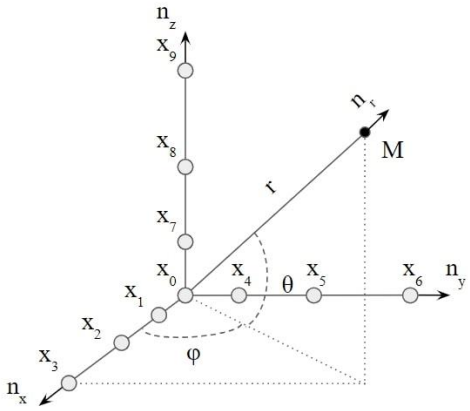

Figure 1: Microphone antenna used for the experimental measurements.

corresponding to the drone's signature in the TFR, the bandwidth is defined as dependent of its centre frequency as for bandpass filters. All the bins selected are in bands centred on $f_{bp} * i$ with bandwidth $\Delta f = \frac{f_{bp} \times i}{Q}$, where $i = 1, \dots, N_h$, N_h is the number of harmonics chosen, and Q, the quality factor. The signal is cut into portions and for each portion, f_{bp} is detected by taking the maximum of the SHC calculated in a given frequency range. To verify that this f_{bp} gives a good DOA estimation, a validation procedure is added to the TFR. If the DOA estimate is higher or lower of 10° than the previous one in azimuth or in elevation, then another f_{bp} is chosen and another DOA is estimated. This condition enables to select the appropriate frequency content by favouring the continuity of the drone trajectory. The frequency range for the SHC calculation is extended and other DOAs are estimated if no previous DOA estimate meets the conditions.

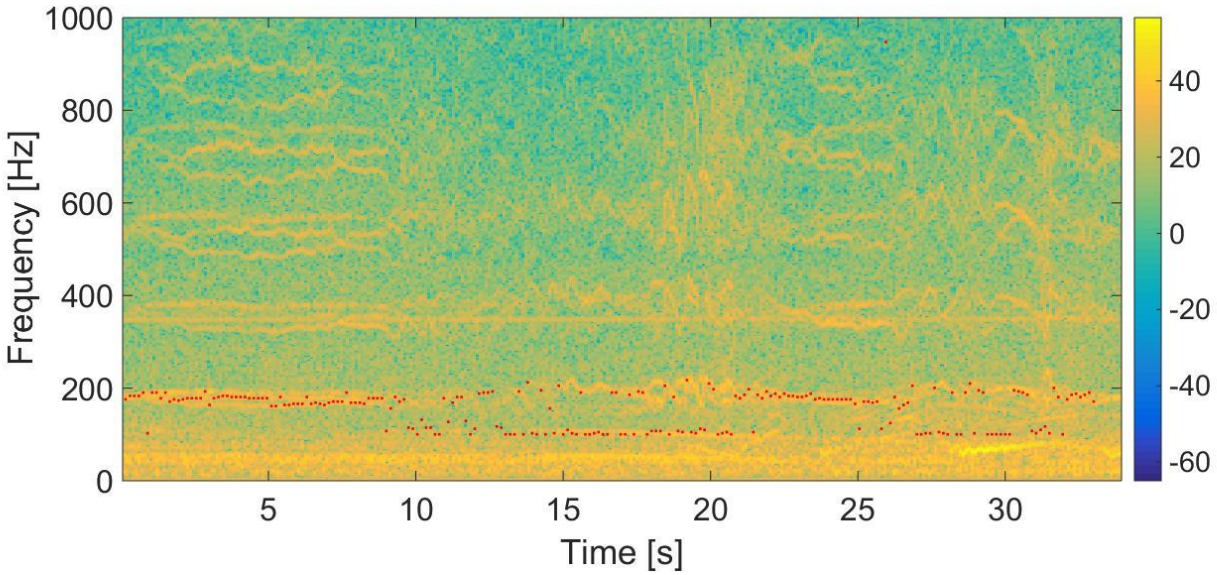

Figure 2: Spectrogram of the reference microphone, red points are detections of f_{bp} using SHC.

3.3 Results

The DOA estimation is computed with classical DSB, DSB with TFR and SRP-PHAT on signal portions of 3000 points. DSB is performed with 2048 points both for classical and TFR, using a plane wave model with a resolution of 4° in azimuth and 2° in elevation for the spherical search grid with a radius of 1 m. SHC is computed with 8192 points and 5 harmonics. For the energy

calculation with the TFR, 5 harmonics are chosen and Q=8. SRP-PHAT is performed with 2048 points for the cross-spectrum calculation with the same search grid as for DSB. Figure 3 shows the evolution of azimuth and elevation during time in comparison with data provided by a GPS embedded in the drone. The drone has an incertitude around 3 meters which can explain the constant bias between the GPS data and the results. It is visible that selecting frequencies from the TFR enables to enhance the performance of beamforming both in azimuth and elevation. SRP-PHAT gives also better results than classical DSB but some fluctuations are still present. Table 1 presents the mean errors and standard deviations for the three approaches. Beamforming with TFR and SRP-PHAT give results very close in azimuth but the TFR approach performs better in elevation.

Table 1: Comparison of mean errors and standard deviations for classical DSB, DSB with TFR, and SRP-PHAT

	Azimuth (°)		Elevation (°)	
	μ	σ	μ	σ
Classical DSB	63,7	42.7	29,1	21,4
DSB with TFR	53,3	14,5	8,9	4,9
SRP-PHAT	53	20,3	14,9	9,6

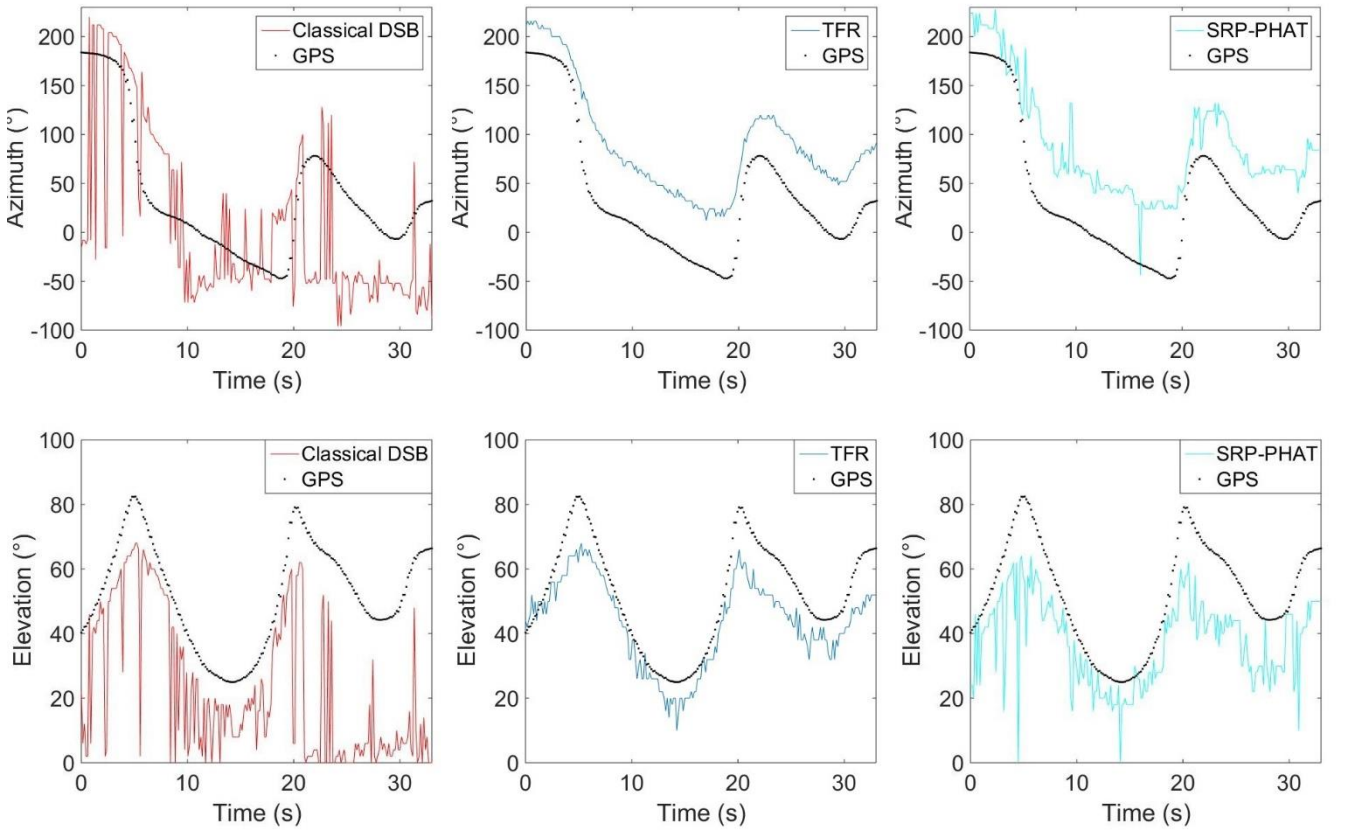

Figure 3: Evolution of azimuth and elevation in time for classical DSB, DSB with TFR, and SRP-PHAT. The GPS trajectory is shown in black dotted lines.

Figure 4 shows an example of an energy map in the azimuth/elevation plane of a portion of signals where classical DSB is very noisy and gives a poor estimate of the DOA in comparison with the TFR approach and SRP-PHAT. Because all the frequency content is selected with classical DSB, the signal to noise ratio (SNR) is very low. The TFR enables to enhance the SNR and therefore clarifies the energy map to find the DOA associated to the source. SRP-PHAT gives also a close estimation thanks to the GCC and weighting.

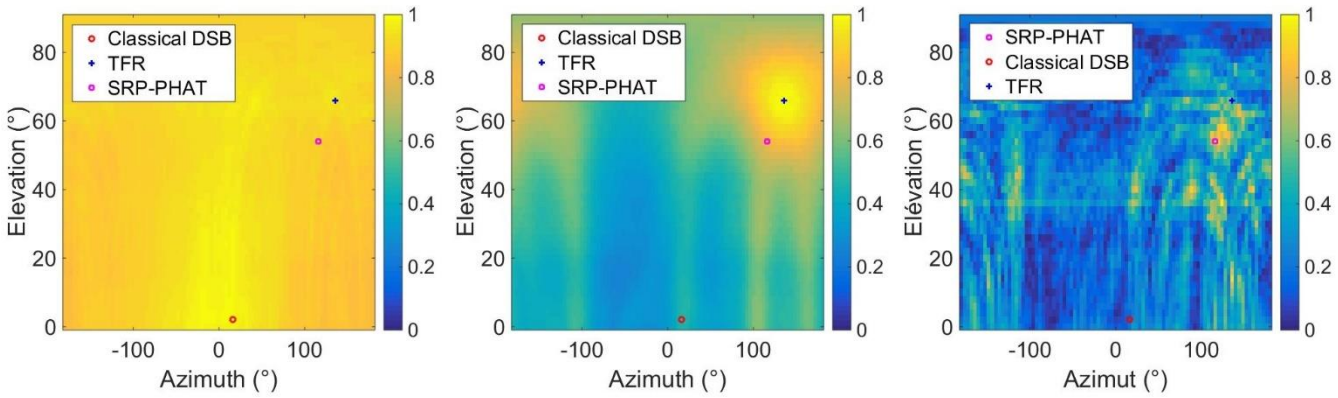

Figure 4: Energy maps in the azimuth/elevation plane for the same signal portion with classical DSB, DSB with TFR, SRP-PHAT (from left to right). The maximum of each cartography gives the DOA estimate for the considered method. For comparison the three DOA estimates are shown on each map

Two successive portions of signals have been chosen to demonstrate the importance of choosing the right f_{bp}. Figure 5 presents the energy maps associated to both portions for classical DSB and SRP-PHAT. The first portion shows accurate DOA estimates for the three methods. Indeed, despite a map with a lot of energy everywhere, classical DSB gives a good estimate of the DOA (a). SRP-PHAT gives a localized energy maximum near the true DOA for both portions [(c) (d)].

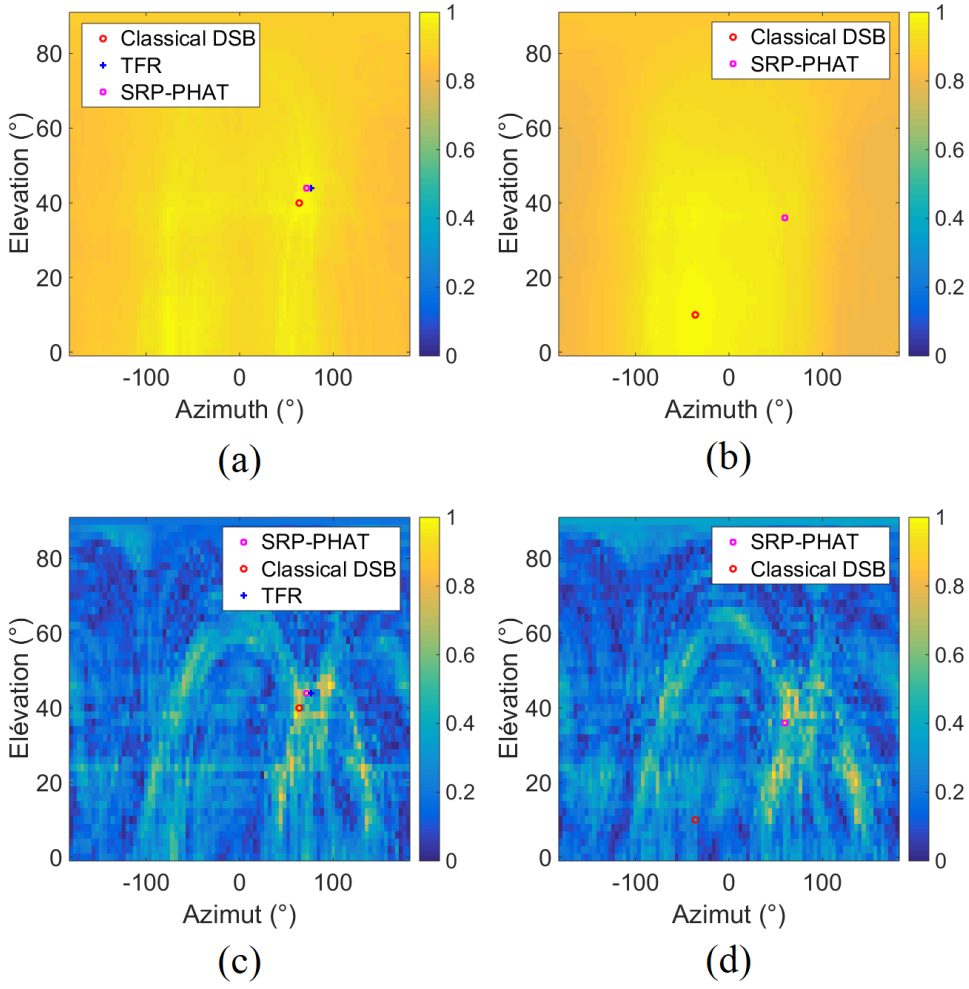

Figure 5: Energy maps in the azimuth/elevation plane for: the first chosen signal portion with classical DSB (a) and SRP-PHAT (c); the next signal portion with classical DSB (b) and SRP-PHAT (d).

The second portion gives a good DOA estimate for SRP-PHAT (d) but not for classical DSB (b). Figure 6 shows energy maps in the azimuth/elevation plane as well as SHC calculated between 100 Hz and 250 Hz for both signal portions using the TFR approach. This makes it possible to see which blade passing frequency is chosen for the computation of the energy contributed by its harmonics. The first portion [Fig. 6 (a)] gives a clearer map than classical DSB [fig. 5 (a)] with a good estimate of the DOA (a) using the max of SHC for f_{bp} (d). The second portion gives a poor estimate of the DOA (b) using the max of SHC for f_{bp} (e). Taking another f_{bp} in the SHC (f) enables to change the content selected for the energy calculation and thus gives a good estimate of the DOA (c). In this case, the third maximum (156 Hz) provided by SHC (e) is selected given the cartography (c). The TFR is very interesting in its ability to choose different DOA estimates depending on the blade passing frequency chosen. A good example is between 28 s and 32 s in the trajectory where a low frequency content is present associated to a car acceleration (see Figure 2). During this time, classical DSB is unable to provide a good localization (Figure 3) while the TFR approach performs well choosing the appropriate frequency content in the energy calculation. SRP-PHAT is also more effective than classical DSB but with more fluctuations than the TFR approach.

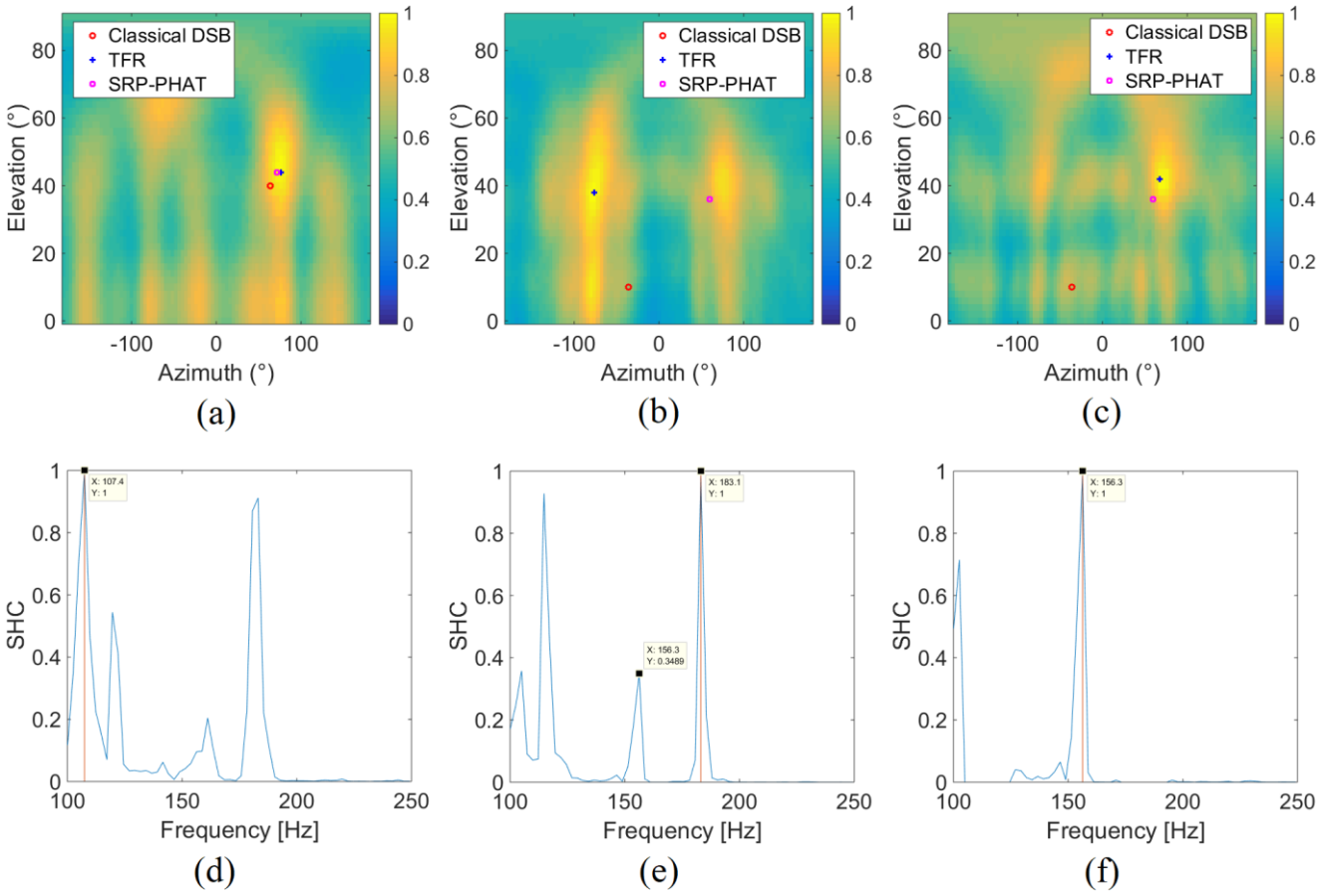

Figure 6: First signal portion chosen: energy map in the azimuth/elevation plane for the TFR approach (a) and SHC associated (d). Second signal portion: energy map in the azimuth/elevation plane for the TFR approach and for the first f_{bp} estimate (b) with the SHC associated (e), for another (right) f_{bp} estimate (c) with the SHC associated (f).

4. Conclusions

This study compares three approaches for estimating the DOA of an UAV. The first uses beamforming, the second, beamforming with a time-frequency representation and the last concerns the steered response power with a phase transformation. The performance of these approaches is evaluated on an experimental trajectory where the drone is flying in circles. Results

show that estimations with the TFR and SRP-PHAT are better than classical DSB. However, TFR approach gives better estimations thanks to frequency content selection. The detection of the blade passing frequency thanks to the SHC is an important step in the process and a bad choice can result in a poor estimate of the DOA. A procedure to avoid this case is presented and enables to better follow the trajectory of the drone. In the presence of strong noise or perturbing sources, classical DSB is not able to estimate the DOAs but SRP-PHAT and the TFR approach still give correct DOAs. To enhance the performance of SRP-PHAT, it could be interesting to compute different weightings as SCOT, PHAT-β, or others.

References

[1] T. Humphreys, « Statement on the security threat posed by unmanned aerial systems and possible countermeasures », *Overs. Manag. Effic. Subcomm. Homel. Secur. Comm. Wash. DC US House*, vol. 1, p. 9, 2015.

[2] R. Altawy et A. M. Youssef, « Security, Privacy, and Safety Aspects of Civilian Drones: A Survey », *ACM Trans. Cyber-Phys. Syst.*, vol. 1, n° 2, p. 1-25, févr. 2017, doi: 10.1145/3001836.

[3] N. Itare, T. Blanchard, J.-H. Thomas, et K. Raoof, « Tracking of an Unmanned Aerial Vehicle with few sensors using time-frequency representation », in *Forum Acusticum*, Lyon, France, déc. 2020, p. 3143-3147. doi: 10.48465/fa.2020.0965.

[4] A. Brutti, M. Omologo, et P. Svaizer, « Comparison Between Different Sound Source Localization Techniques Based on a Real Data Collection », in *2008 Hands-Free Speech Communication and Microphone Arrays*, Trento, Italy, mai 2008, p. 69-72. doi: 10.1109/HSCMA.2008.4538690.

[5] J. Wang, X. Qian, Z. Pan, M. Zhang, et H. Li, *GCC-PHAT with Speech-oriented Attention for Robotic Sound Source Localization*. 2021. doi: 10.1109/ICRA48506.2021.9561885.

[6] A. Johansson, G. Cook, et S. Nordholm, « Acoustic Direction of Arrival Estimation, a Comparison Between Root-MUSIC and SRP-PHAT », *2004 IEEE Reg. 10 Conf. TENCON 2004*, p. 629--632, 2004.

[7] J. H. DiBiase, H. F. Silverman, et M. Brandstein, *Robust Localization in Reverberant rooms*, from Microphone arrays: signal processing, Techniques and Applications. Berlin: Springer, 2001.

[8] G. C. Carter, A. H. Nuttal, et P. G. Cable, « The smoothed coherence transform », in *Proceedings of the IEEE*, oct. 1973, vol. 61, p. 1497--1498. doi: 10.1109/PROC.1973.9300.

[9] K. Donohue, A. Agrinsoni, et J. Hannemann, *Audio signal delay estimation using partial whitening*. 2007, p. 471. doi: 10.1109/SECON.2007.342946.

[10] S. A. Zahorian et H. Hu, « A spectral/temporal method for robust fundamental frequency tracking », *J. Acoust. Soc. Am.*, vol. 123, n° 6, p. 4559-4571, juin 2008, doi: 10.1121/1.2916590.

[11] T. Blanchard, « Caractérisation de drones en vue de leur localisation et de leur suivi à partir d'une antenne de microphones (Characterization of drones for their localization and their tracking from a microphone array) », PhD thesis, Le Mans Université, 2019.

[12] T. Blanchard, « Acoustic localization and tracking of a multi-rotor unmanned aerial vehicle using an array with few microphones », *J Acoust Soc Am*, p. 12, 2020, doi: https://doi-org.doc-elec.univ-lemans.fr/10.1121/10.0001930.