HAL open science

Acoustic localization and identification of drones with a disturbance source

Valentin Baron, Simon Bouley, Matthieu Muschinowski, Jerome I. Mars, Barbara Nicolas

To cite this version:

Valentin Baron, Simon Bouley, Matthieu Muschinowski, Jerome I. Mars, Barbara Nicolas. Acoustic localization and identification of drones with a disturbance source. Forum Acusticum 2020, Dec 2020, Lyon (virtual), France. pp.3149-3154, 10.48465/fa.2020.0402 . hal-03233664

HAL Id: hal-03233664

https://hal.science/hal-03233664v1

Submitted on 26 May 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ACOUSTIC LOCALIZATION AND IDENTIFICATION OF DRONES WITH A DISTURBANCE SOURCE

Valentin Baron[1,2,3] **Simon Bouley**[1] **Matthieu Muschinowski**[3]
Jérôme Mars[3] **Barbara Nicolas**[2]
[1] MicrodB, 28 Chemin du Petit Bois, 69130 Ecully, France
[2] Univ Lyon, INSA-Lyon, Université Claude Bernard Lyon 1, UJM Saint-Etienne, CNRS, Inserm, CREATIS, UMR 5220, U1206, F-69100, Lyon, France
[3] Univ. Grenoble Alpes, CNRS, Grenoble-INP*, GIPSA-Lab, 38000 Grenoble, France
valentin.baron@microdb.fr, barbara.nicolas@creatis.insa-lyon.fr

ABSTRACT

Due to their low cost and their easy use, drones are increasingly used as tools to threaten valuable assets in both military and civil domains. Thus systems have been developed to localize and/or identify them before their neutralization. They can rely on radio frequency, radar signal, or even video recording. However, these modalities suffer from limitations like near-field detection, or weather conditions. Besides, drones do radiate acoustic noise during their flight due to their propellers. Consequently, the proposed work takes the advantage of the acoustic waves emitted by drones to localize them through array processing techniques, and to identify them as drones thanks to machine learning.

As multiple acoustic sources can be present within the nearby environment, this work focuses on localization and identification of a drone in the presence of a disturbing acoustic source. An experiment has been performed with a drone flying above a 81-microphones acoustic array of 60 cm diameter while a motionless loudspeaker emits sounds at different volumes. Signals recorded on the array are cut into successive sequences of 200 ms to allow an "instantaneous" identification and tracking the trajectories of the different sources. The high resolution MUSIC algorithm is performed on each sequence to estimate the angular directions of arrival of the sources, using local maxima of the localization map. Then focused signals are recovered for each source, and classified as *drone* or *other noise* thanks to a machine learning model (support vector machine, SVM) learned previously on similar data. Results show that drone localization is accurate, even for drone noise volume lower than the sounds played by the loudspeaker. The limit from which localization is not valid anymore depends on the sound played. The *drone* class identification is very robust with at least 84 % of correct identification rate for all loudspeaker volumes tested. The *other noise* class performs worse with incorrect identifications that arise when the loudspeaker volume decreases, with rates that depend on the considered sound.

*Institute of Engineering Univ. Grenoble Alpes.

1. INTRODUCTION

Drones are known threats for both military and civil assets. They can carry explosives or drop illegal objects, and they are easy to buy and to use. Solutions are then developed to localize and neutralize them [1]. Radar is one of the principal solution, but for small quadcopters they suffer from a lack of reflected signal [2]. Radio-frequency based solutions are very promising as they allow to both detect the drone and its pilot. Nevertheless, autonomous drones are already under development to avoid this kind of detection [3]. Optronic solutions that rely on video are also investigated, but they can quickly have some limitations or be difficult to set at night or with tough weather conditions [4, 5]. In this study, the acoustic modality is chosen. Quadcopters do produce noise while they fly [6, 7], and this noise increases with the engine regime [8]. It becomes logical to study the possibility to localize a drone from the acoustic waves it generates. Moreover, the typical acoustic signature of a drone can be exploited to identify it.

Localization and identification of acoustic sources remain uncoupled in the literature. Localization is given using different algorithms and array shapes [9–13], and identification is often proposed using a single microphone, combined with machine learning techniques [14–18]. This article proposes to perform both together, in order to benefit from array processing gain before the identification step. Using an array, the goal is first to localize acoustic sources in the nearby environment, then to spatially filter each of them to enhance their signal-to-noise ratio, and finally to identify these focused audio signals in definite classes.

The proposed system has been developed and validated for simple cases with one acoustic source around it [19]. The purpose of this study is to assess its robustness in the presence of a disturbing source. Experimental data have been collected with a drone flying while a loudspeaker was playing different sounds. Acoustic localization is performed, and focused signals are recovered in the estimated sources directions. The identification is then carried out, using a machine learning model trained on drone-labeled data and sounds of the loudspeaker, recorded alone previously. This identification step tries to separate two classes: the drone class, against the other noises class. The aim is

then to show the impact of the difference level between the loudspeaker sound and the drone on the localization or the identification step. It allows to set an operating range for the considered system in this configuration.

The next section describes the method used to localize and identify the acoustic sources around the array. Then section 3 details the experimental data available while the results are discussed in section 4.

2. METHODS

The drone localization and identification constitutes two successive steps. It starts with an array processing method to localize potential acoustic sources around the microphone array. This localization allows to focus over these sources, by applying a spatial filter in the direction of each of them. The audio signals obtained are then identified using a machine learning algorithm learned on similar data. To track the sources along time, these described steps are undertaken on successive sequences of the recorded signals, each of them during 200 ms.

2.1 Localization

The array processing algorithm selected to localize the sources is the multiple signal classification method (MUSIC) [20]. Due to the sparsity of the acoustic sources searched for, and the weak correlation expected between them, this algorithm is adapted as it allows to significantly increase the resolution compared to conventional beamforming techniques, and to avoid sidelobes that can mask some of the sources if they are not loud enough. Another advantage is its low complexity, which allows to compute quickly its outputs. As two sources are expected in this configuration, a fixed number of eigenvalues of two is kept all along the processed cases. An automatic procedure to determine the number of eigenvalues to select must be developed to reduce observed errors with the current choice.

The angular directions of arrival are searched for using a plane wave propagation model. To validate the localization, the estimated azimuth angles are compared to the ones obtained by 12 infrared Vicon cameras. Installed within a motion capture room, they produce a localization ground truth with which the acoustic one should be aligned. Figure 1 shows an example of the comparison between the azimuth angle found by the acoustic array in gray circles (○), and the one given by the Vicon cameras in green for the drone (-) and in blue for the loudspeaker (-).

For each sequence of 200 ms, the direction of the two first maxima of the computed MUSIC map are determined, and a delay-and-sum in each of these directions is performed to spatially filter the temporal signals, and obtain two focused sequences. As both drone and loudspeaker are present all along the time, these two position estimations should lie close to the ground truth given by the Vicon cameras.

In order to evaluate the localization, upper and lower angular thresholds are set around the Vicon camera system outputs for both the drone and the loudspeaker. For each

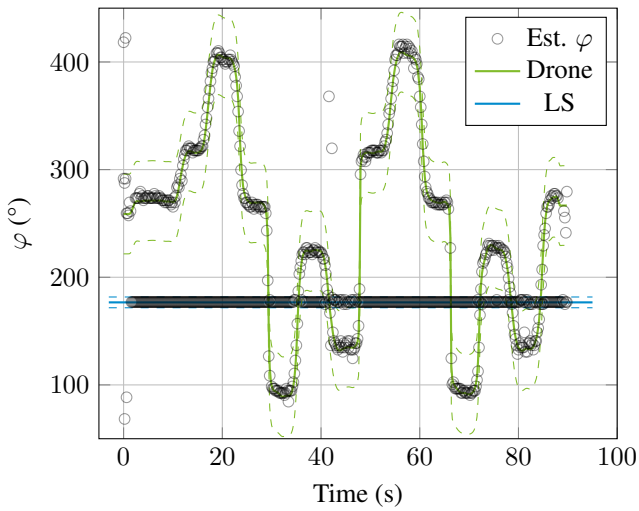

Figure 1. Estimated azimuth angles of the acoustic sources present around the array, given by ○. The Vicon camera system output is given by the green line - for the drone, and by the blue line at constant azimuth angle 180° - for the loudspeaker (LS). Dashed lines represent the thresholds set to determine the sequence labels. The angles are unwrapped to clearly see them.

position estimation ○, if it is contained within the drone or the loudspeaker thresholds, it is labeled as *drone* (D) or *other noise* (ON), respectively. For a position that falls outside both, it is considered as *Unknown* (U). According to these labels, the correct localization rate (CLR) can be computed for both D and ON classes as the number of sequences that are labeled within them over the total number of sequences available during the considered flights.

2.2 Identification

Once the focused signals are recovered for each source, a bench of 96 features is computed over this temporal signal, producing a so-called observation (a vector of dimension [96, 1] containing these features). They describe shape, statistics, and entropy of the signals [21], in three different domains: temporal, spectral, and cepstral. A linear support vector machine (SVM) model [22] is then learned using these features. This model has been deliberately chosen to be simple, in order to deal with an easy classification task, and to be able to easily interpret its outputs. The coefficient C, which controls the capacity of the model to select support vectors that are in the wrong side of the hyperplan, is set to 1. A cross validation of the learning data is performed to check their balance and it shows that for the drone and the loudspeaker sounds are separable if they do not emit simultaneously. All these computations are made within a dedicated Python package [23].

From the labeled observations obtained after localization step, the correct identification rate (CIR) defines the ratio between the correctly identified observations, over the total number of labeled observations for each class D vs ON.

3. EXPERIMENTAL DATA

Experimental data have been recorded within a motion capture room in order to be able to compare the estimated angular position of the drone using acoustic array of microphones, with a ground truth given by a system composed of 12 infrared Vicon cameras [24]. The array is composed of 81 MEMS microphones, with a diameter of 60 cm. It has already been used in other scientific studies ([25] for instance), and it is represented in Figure 2 with the drone.

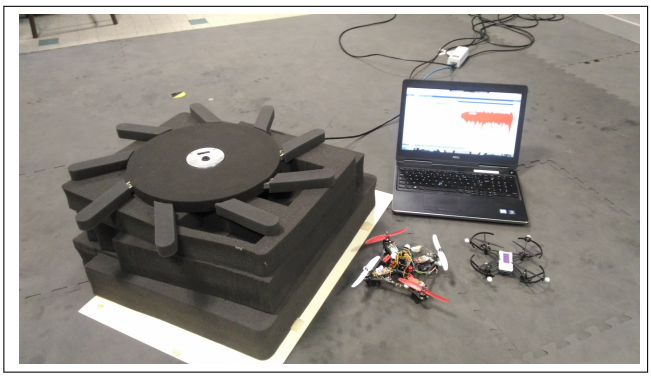

Figure 2. Array used to record acoustic data, and drone used to produce data (left one).

Experimental data includes drone flights alone, sounds played by the loudspeaker alone, and both of them together. The sounds have been kept sufficiently different from a drone noise in order to focus on classification issues due to the difference level between the drone noise and the loudspeaker sounds: it means that the model should easily separate D and ON classes. Two sounds were selected: a white noise (WN), and a monochromatic signal at 1 kHz (S1k). As a consequence, the processing frequency of the localization algorithm was set at 1 kHz, to process the harmonic signal. It is important to notice that the drone harmonics are not present at this specific frequency. Its blade passing frequency and its harmonics are multiples of 400 Hz, so this frequency choice is sub-optimal in terms of drone localization. Table 1 sums up the different number of observations available in each case.

	Learning	Test
Drone alone	1378	0
WN (L1 / L2 / L3)	150 / 150 / 149	460 / 462 / 467
S1k (L1 / L2 / L3)	150 / 150 / 150	474 / 468 / 449

Table 1. Number of observations values available. In the learning data, drone and loudspeaker sounds are played separately, whereas in the test data, the loudspeaker plays while the drone is flying. Each observation is a set of 96 features computed over 200 ms of temporal focused signal.

One observation is a vector of 96 features computed over a sequence of 200 ms of temporal signal. L1, L2, and L3 represents the three different volume levels used for the loudspeaker, from lower one to maximum one. They allow to decrease the level difference between the drone noise and the loudspeaker sounds, in order to disturb both the localization and identification steps. During the learning phase, the drone and the loudspeaker are recorded separately. On the contrary, during the test phase, sounds are played while the drone is flying in the room simultaneously.

Thanks to the use of a controlled trajectory for the drone that can be repeated between learning and test phases, two difference levels between the drone and the loudspeaker sound can be defined when the sources are recorded separately: at first the ratio between the measured pressure amplitude at 1 kHz for the drone over the one of the sound played by the loudspeaker (ΔL_{1k}), and secondly the ratio between the mean pressure amplitude of the drone power spectral density over the whole frequency band [0 Hz; 25.6 kHz], over the same mean pressure amplitude for the loudspeaker sound (ΔL_{mean}). While ΔL_{1k} represent the ratio available for the localization as it is performed only at 1 kHz, ΔL_{mean} represents the level difference that is really at stake during the identification step as observations are computed for full temporal focused signals, in which all frequencies are present.

4. RESULTS

4.1 Localization

The correct localization rates (CLR) for every sound and loudspeaker volume is given in Table 2.

CLR	Drone	Loudspeaker
L1 (WN / S1k)	93 % / 93 %	81 % / 86 %
L2 (WN / S1k)	63 % / 90 %	94 % / 88 %
L3 (WN / S1k)	26 % / 96 %	97 % / 92 %

Table 2. CLR computed over all the observations of the trajectories, as a function of loudspeaker sounds levels.

For the WN, the higher the loudspeaker volume level, the lower the drones are correctly localized, and inversely for the sounds. In the S1k case, the drone is correctly localized for every sound volume, and the loudspeaker is better localized when it is emitting louder. The WN disturbs more drone localization with only 26 % of correct localization for its higher volume. Finally, the loudspeaker position is estimated more accurately than the drone one, which can be expected as it does not move during time.

To analyze these results deeper, the number of observations labeled in each category (D, ON, or U) is given as a function of ΔL_{1k}. In Figure 3, the S1k case is considered. It is clear from the distribution obtained that both D and ON classes are accurately localized for this signal, with only a few observations labeled in U class, even for highly negative values of ΔL_{1k}.

Then the WN case is given in Figure 4. The analysis is there a lot different, with ON observations that are correctly localized, but D ones cannot be accurately recovered for all ΔL_{1k}. Consequently, for low ΔL_{1k}, numerous U

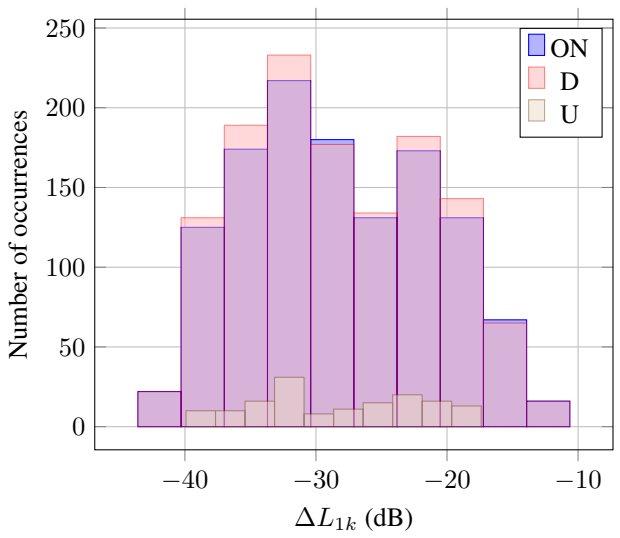

Figure 3. Number of labels given after the labeling step, as a function of the difference level at 1 kHz between the drone and the loudspeaker sound ΔL_{1k}. The three labels type are D, ON, or U. Monochromatic signal at 1 kHz case.

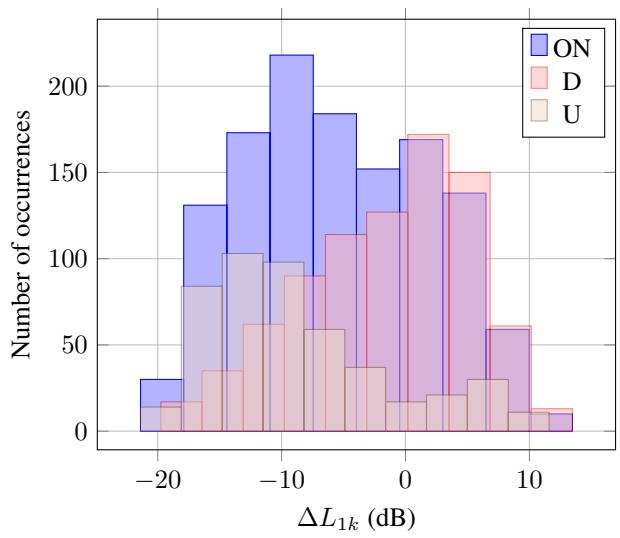

Figure 4. Number of labels given after the labeling step, as a function of the difference level at 1 kHz between the drone and the loudspeaker sound ΔL_{1k}. The three label types are D, ON, or U. White noise case.

labels are obtained. The limit between the two distributions is around ΔL = -8 dB.

To conclude, the system is able to localize a drone in the presence of the S1k sound without any problem. Concerning the WN sound, the localization is accurately recovered for the drone, up to a loudspeaker that emits 8 dB higher than the drone at 1 kHz. This difference is due to the fixed number of eigenvalues kept during the computation of the MUSIC map. While two eigenvalues successfully describe the signal subspace of a drone and a sinus signal at 1 kHz, they are not sufficient anymore for two wideband signals together.

This conclusion is valid within this particular indoor configuration, and must not be understood as a fixed limit for the system in all the other configurations. Indeed, strong reflections are present within this room, so they can add difficulties to the localization step. Moreover, it is reminded here that the 1 kHz frequency chosen for the array processing is sub-optimal for the drone detection (with levels that can vary more than 10 dB if an harmonic frequency of the drone noise is chosen).

4.2 Identification

From the labeled observations obtained after the localization, the correct identification rate for each of them is given in Table 3.

These results show that drone identification is very accurate, as for every volume or played sound it enables to reach at least 84 % of correct identification. Moreover, in the WN case it is easier to identify the drone. Concerning the ON class identification, the volume differences play a strong role. Indeed, for lower volumes, both sound cases are not recognized at all. Concerning the middle volume, S1k is identified correctly whereas WN is just partially identified. For higher volumes, both sounds are correctly identified.

CIR	D class	ON class
L1 (WN / S1k)	100 % / 100 %	1 % / 3 %
L2 (WN / S1k)	100 % / 97 %	42 % / 98 %
L3 (WN / S1k)	100 % / 84 %	91 % / 100 %

Table 3. CIR computed over all the labeled observations of the trajectories, as a function of loudspeaker sounds levels.

To confirm these results, histograms of ΔL_{mean} are given for both signal cases and for the ON class, grouped by correct and incorrect identification. Figure 5 represents it for the S1k case. It shows a strong separation between both distributions, which reinforces the results given in the Table 3: in this case, identification is accurate up to a clear threshold for ΔL_{mean} above which it does not work anymore. This threshold is defined at ΔL_{mean} = 0 dB.

Then the WN case is represented in Figure 6. The separation is not as clear as or S1k, explaining the increments observed as a function of the volumes in the results shown in Table 3 for this case. Nevertheless, a threshold can also be pointed out in this case, for ΔL_{mean} around 2 dB.

To conclude, identification is very robust concerning the D class. Almost all the D observations are correctly identified as such, with a lower CIR obtained at 84 %, even with negative ΔL_{mean} values (meaning a loudspeaker higher in volume than the drone). This is certainly due to the low number of drone observations that are kept after localization in the WN case, so they are not taken into account during identification studying only the ones that emit louder than the loudspeaker. Concerning the S1k, the signal structure is so different between these two that drone identification is not disturbed by it when it is present at the

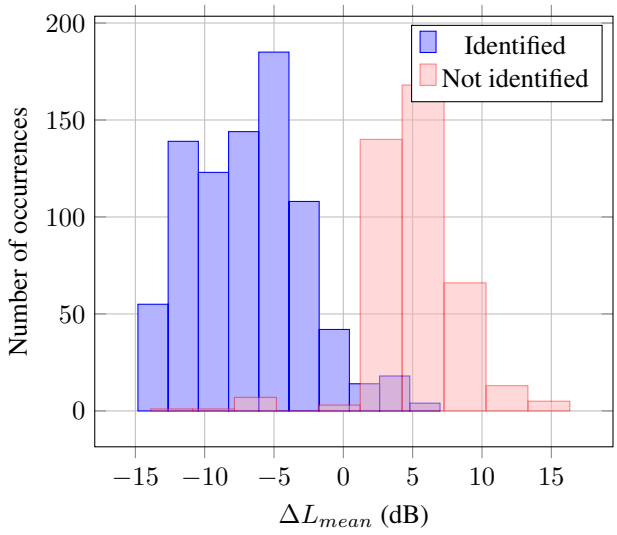

Figure 5. Histogram of the correct and incorrect identifications of ON class as a function of the difference level between the drone and the loudspeaker sound ΔL_{mean}. Monochromatic signal at 1 kHz case.

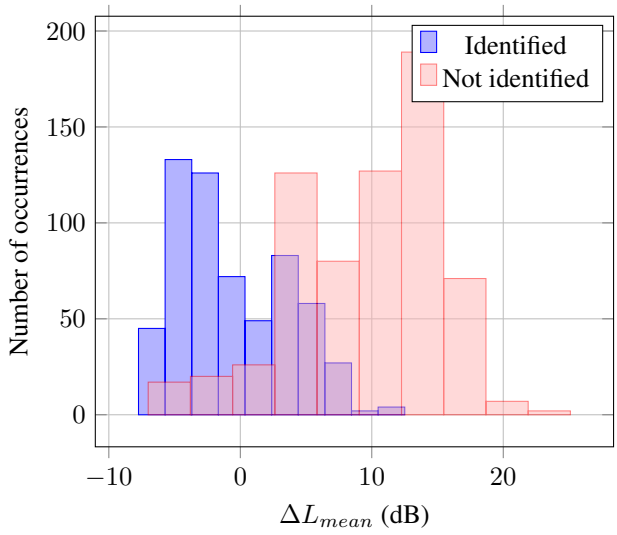

Figure 6. Histogram of the correct and incorrect identifications of ON class as a function of the difference level between the drone and the loudspeaker sound ΔL_{mean}. White noise case.

same time.

Then, identification of the loudspeaker sounds shows two different behaviors, depending on the sound played. For the S1k sound, identification works very well if the loudspeaker emits louder than the drone, and it does not work anymore when the drone is noisier. For the WN case, this identification performance decreases more progressively. So the disturbance sound nature impacts on its identification, with a clearer separation between correct and incorrect identification when the played sound is completely different than the drone one. As this disturbance is due to the signal residuals that come from a source when focusing towards the other, it seems logical than similar signal are less easily separable than highly different ones.

5. CONCLUSION

This article studies the robustness of the acoustic localization and identification of a drone in the presence of another acoustic source. The system, composed of a 60 cm diameter acoustic array with 81 MEMS microphones, shows that an accurate localization of the drone is obtained every 200 ms when the disturbance is monochromatic, whatever its volume. However, when the disturbance is wideband, this accuracy decreases when its volume increases. Then during identification, the drone is correctly identified each time it is localized and for every sound and volume, whereas the disturbance sound is not always correctly classified. For the monochromatic one a clear separation arises: when it emits louder than the drone it is recognized, otherwise it is not. For the wideband one this separation is more fuzzy and the identification can be reached even when the drone is noisier. These results show the strong dependence between the machine learning models and the data that are used to build them. Different kinds of sound provoke different analysis, above all in the case studied here in which they are present simultaneously with the drone. As a consequence, an important issue in drone identification concerns the creation of the *other noise* class database, in order to correctly sample it to yield results that can be generalized when drone signals are disturbed.

To increase the capacities of the proposed system, the localization must be improved by selecting a number of eigenvalues more adapted to the measured data, and different strategies must be compared to focus the temporal signals in order to attenuate the residuals coming from other angular directions. With these two additions, the system should cope with lots of experimental configurations that could occur in real-life settings.

6. ACKNOWLEDGMENTS

This work was supported by the Equipex ROBOTEX (ANR-10-EQPX-44-01) and it was performed within the framework of the LABEX CeLyA (ANR-10-LABX-0060) of Université de Lyon, within the program "Investissements d'Avenir" (ANR-16-IDEX-0005) operated by the French National Research Agency (ANR).

7. REFERENCES

[1] B. Taha and A. Shoufan, "Machine learning-based drone detection and classification: state-of-the-art in research," *IEEE Access*, vol. 7, pp. 138669–138682, 2019.

[2] D. Shin, D. Jung, D. Kim, J. Ham, and S. Park, "A Distributed FMCW Radar System Based on Fiber-Optic Links for Small Drone Detection," *IEEE Transactions on Instrumentation and Measurement*, vol. 66, pp. 340–347, Feb. 2017.

[3] F. Caballero, L. Merino, J. Ferruz, and A. Ollero, "Vision-Based Odometry and SLAM for Medium and High Altitude Flying UAVs," *Journal of Intelligent and Robotic Systems*, vol. 54, pp. 137–161, Mar. 2009.

[4] T. Multerer, A. Ganis, U. Prechtel, E. Miralles, A. Meusling, J. Mietzner, M. Vossiek, M. Loghi, and V. Ziegler, "Low-cost jamming system against small drones using a 3D MIMO radar based tracking," in *2017 European Radar Conference (EURAD)*, pp. 299–302, Oct. 2017.

[5] A. Rozantsev, V. Lepetit, and P. Fua, "Detecting Flying Objects Using a Single Moving Camera," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, pp. 879–892, May 2017.

[6] P. Candeloro, T. Pagliaroli, D. Ragni, and S. Di Francesco, "Small-scale rotor aeroacoustics for drone propulsion: a review of noise sources and control strategies," Oct. 2019.

[7] R. Serré, H. Fournier, and JM. Moschetta, "A design methodology for quiet and long endurance MAV rotors," *International Journal of Micro Air Vehicles*, vol. 11, pp. 1–14, May 2019.

[8] O. Gur and A. Rosen, "Design of Quiet Propeller for an Electric Mini Unmanned Air Vehicle," *Journal of Propulsion and Power*, vol. 25, pp. 717–728, May 2009.

[9] A. Ramamonjy, E. Bavu, A. Garcia, and S. Hengy, "Source localization and identification with a compact array of digital MEMS microphones," in *ICSV25*, (Hiroshima), p. 8, July 2018.

[10] X. Chang, C. Yang, J. Wu, X. Shi, and Z. Shi, "A Surveillance System for Drone Localization and Tracking Using Acoustic Arrays," in *2018 IEEE 10th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, pp. 573–577, July 2018.

[11] G. Herold, A. Kujawski, C. Strumpfel, S. Huschbeck, M. de Haag, and E. Sarradj, "Detection and separate tracking of swarm quadcopter drones using microphone array measurements," p. 19, 2020.

[12] T. Pham and N. Srour, "TTCP AG-6: acoustic detection and tracking of UAVs," in *Unattended/Unmanned Ground, Ocean, and Air Sensor Technologies and Applications VI*, vol. 5417, pp. 24–31, Sept. 2004.

[13] T. Blanchard, JH. Thomas, and K. Raoof, "Acoustic localization estimation of an Unmanned Aerial Vehicle using microphone array," in *Inter-Noise*, (Madrid), Sept. 2019.

[14] A. Bernardini, F. Mangiatordi, E. Pallotti, and L. Capodiferro, "Drone detection by acoustic signature identification," *IS&T International Symposium on Electronic Imaging*, vol. 2017, pp. 60–64, Jan. 2017.

[15] L. Shi, I. Ahmad, Y. He, and K. Chang, "Hidden markov model based drone sound recognition using MFCC technique in practical noisy environments," *Journal of Communications and Networks*, vol. 20, pp. 509–518, Oct. 2018.

[16] M. Nijim and N. Mantrawadi, "Drone classification and identification system by phenome analysis using data mining techniques," in *2016 IEEE Symposium on Technologies for Homeland Security (HST)*, pp. 1–5, May 2016.

[17] Y. Seo, B. Jang, and S. Im, "Drone detection using convolutional neural networks with acoustic STFT features," in *2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pp. 1–6, Nov. 2018.

[18] U. Ganapathi and M. Sabarimalai Manikandan, "Convolutional neural network based sound recognition methods for detecting presence of amateur drones in unauthorized zones," in *Machine Learning, Image Processing, Network Security and Data Sciences*, pp. 229–244, Springer, June 2020.

[19] V. Baron, S. Bouley, M. Muschinowski, J. Mars, and B. Nicolas, "Drone localization and identification using an acoustic array and supervised learning," in *Artificial Intelligence and Machine Learning in Defense Applications*, vol. 11169, International Society for Optics and Photonics, Sept. 2019.

[20] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. AP-34, pp. 276–280, Mar. 1986.

[21] M. Malfante, M. Dalla Mura, J.P. Metaxian, J. Mars, O. Macedo, and A. Inza, "Machine Learning for Volcano-Seismic Signals: Challenges and Perspectives," *IEEE Signal Processing Magazine*, vol. 35, pp. 20–30, Mar. 2018.

[22] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, pp. 273–297, Sept. 1995.

[23] M. Malfante, J. Mars, and M. Dalla Mura, "Automatic analysis architecture." `https://zenodo.org/record/1216028`, 2019. Accessed: 16/03/2020.

[24] P. Merriaux, Y. Dupuis, R. Boutteau, P. Vasseur, and X. Savatier, "A Study of Vicon System Positioning Performance," *Sensors*, vol. 17, p. 1591, July 2017.

[25] J. Antoni, T. Le Magueresse, Q. Leclère, and P. Simard, "Sparse acoustical holography from iterated Bayesian focusing," *Journal of Sound and Vibration*, vol. 446, pp. 289–325, Apr. 2019.