# Source localization and identification with a compact array of digital mems microphones

Aro Ramamonjy, Eric Bavu, Alexandre Garcia, Sébastien Hengy

## To cite this version:

Aro Ramamonjy, Eric Bavu, Alexandre Garcia, Sébastien Hengy. Source localization and identification with a compact array of digital mems microphones. 25th International Congress on Sound and Vibration (ICSV25), Jul 2018, Hiroshima, Japan. hal-02088346

HAL Id: hal-02088346
https://hal.science/hal-02088346

Submitted on 2 Apr 2019

25th International Congress on Sound and Vibration
8-12 July 2018 HIROSHIMA CALLING
ICSV25

# SOURCE LOCALIZATION AND IDENTIFICATION WITH A COMPACT ARRAY OF DIGITAL MEMS MICROPHONES

Aro Ramamonjy, Eric Bavu, Alexandre Garcia
*Laboratoire de Mécanique des Structures et des Systèmes Couplés, CNAM (LMSSC), Paris, France*
*email: aroramamonjy@gmail.com*

Sébastien Hengy
*French-German Research Institute of Saint-Louis (ISL), Saint-Louis, France*

A compact microphone array was developed for source localization and identification. This planar array consists of an arrangement of 32 digital MEMS microphones, concentrated in an aperture of fewer than 10 centimeters, and connected to a computer by Ethernet (AVB protocol). 3D direction of arrival (DOA) localization is performed using the pressure and the particle velocity estimated at the center of the array. The pressure is estimated by averaging the signals of multiple microphones. We compare high order pressure finite differences to the Phase and Amplitude Gradient Estimation (PAGE) method for particle velocity estimation. This paper also aims at presenting a method for UAV detection using the developed sensor and supervised binary classification.

## 1. Introduction and global approach

The use of unmanned aerial vehicles (UAV) for both civil and military applications is emerging, and the surveillance of these devices is becoming a major concern.

A network of compact microphone arrays (CMA) is used to detect and localize a potential target, and the 3D DOA of this potential target is transfered to an optical system for a multi-modal audio-video tracking and identification.

The video counterpart of the proposed acoustic system consists in an active imaging system which was developed by the French-German Research Institute of Saint-Louis (ISL). This system can give a clear image of a drone flying hundreds of meters away (see Fig. 1, right). This system can detect a drone at a distance up to 1.5 km, but it has a restricted viewing angle, so it has to be oriented towards the target before being able to trigger video tracking and identification. The developed CMA aims at achieving this task in real time.

The present paper focuses on the localization and identification tasks to be achieved by one CMA of the surveillance network. The CMA is composed of a microphone array of 32 digital MEMS microphones arranged in the 2D plane (see Section 2), and connected to a computer substation, which performs the signal processing tasks presented in Fig. 1.

First, spatial filtering is achieved using differential beamforming to focus the array on four principal directions [1] in order to enhance the initial detection without altering the drone sound signature representation. Then, an initial source detection is performed on these four directions (Section 4). The sources are then localized (Section 3). Localization is performed with an estimate of the pressure and the particle velocity at the center of the CMA. The localized sources are enhanced by DOA informed spatial filtering [1], and identification is performed on the enhanced source signals (Section 4).

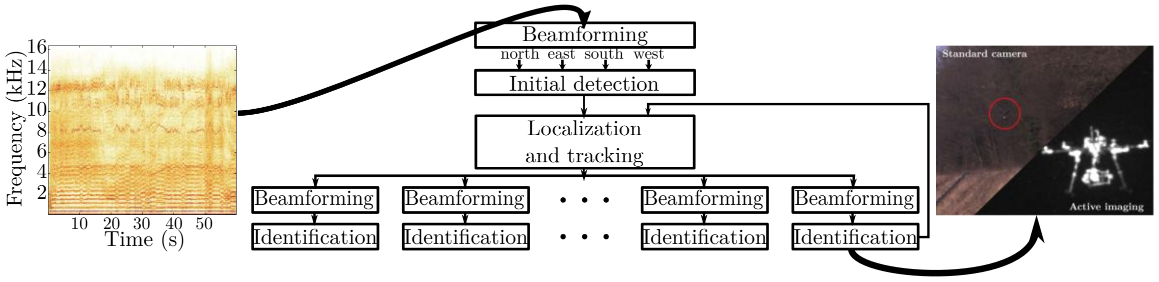


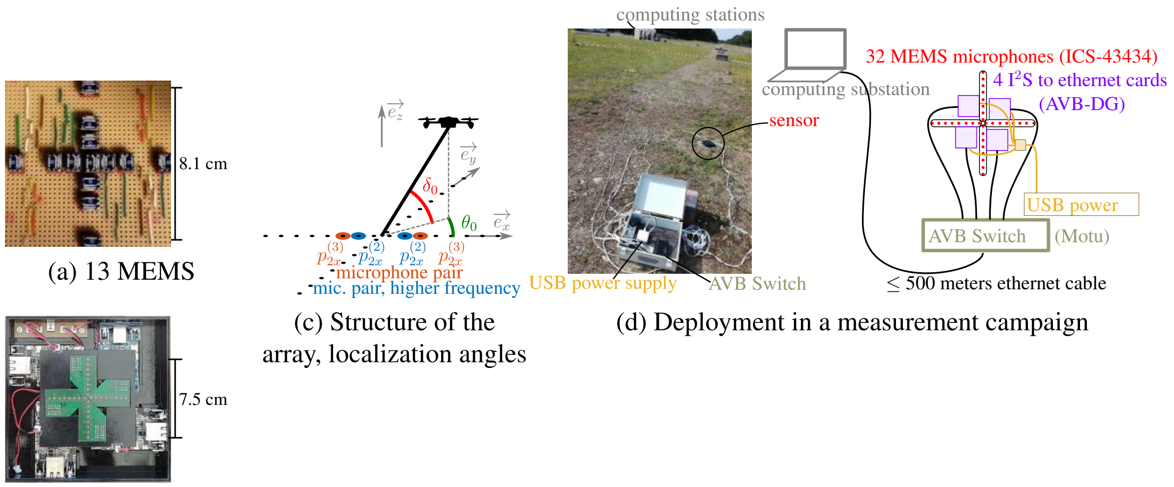
FIGURE 1 – Global approach

A microphone recording of a drone is presented on Fig. 1 (left). It shows a lot of strong harmonic components between 200 Hz and 5 kHz, which can be useful for source localization. 3D DOA estimation with a fewer than 10 degrees error between these two frequencies and source detection with a low false negative rate would give a good initialization to the video tracking and identification.

## 2. The microphone array

### 2.1 Structure

The Fig. 2 shows the last two prototypes of the developed CMAs. Both consist in two orthogonal lines of MEMS microphones which are placed in the horizontal plane. Multiple microphone pairs are used to estimate the pressure and the particle velocity components on two orthogonal axis at the center of the CMA, i.e. at the crossing of the two lines of microphones. Different spacings between the microphones are used either separately to measure the acoustic field at different frequencies (in this case decreasing spacing are used for increasing frequencies, see Fig. 2c), or together to obtain a more accurate estimate (higher order estimations). The use of logarithmic spacing between the microphones (Fig. 2a) allows to perform localization in log scaled frequency bands with a limited number of microphones, while the use of linear spacing (Fig. 2b) makes possible to use classical beamforming algorithms conceived for linear arrays.



(a) 13 MEMS

(b) 32 MEMS

(c) Structure of the array, localization angles

(d) Deployment in a measurement campaign

FIGURE 2 – Developed compact microphone arrays

### 2.2 Technology

The CMA relies on the digital MEMS microphones technology. More and more acoustic arrays use this type of microphones. Their advantages rely on their small size, low cost, and integrated system-on-chip packaging and digitization. In addition, we can now find MEMS microphones that have very consistent audio performances and low background noise. These advantages make it possible today to deal relatively easily with the development of large acoustic networks, and the densification and miniaturization of acoustic antennas.

The last prototype (see Fig. 2b) has 4 branches of 8 digital I$^2$S MEMS microphones (models : Invensense ICS-43434). The elements that make the connection between the microphones and the computer, located at 500 meters of cable further, are shown in Fig. 2d. Each block of 8 MEMS is connected via a custom designed electronic chip, to an I$^2$S to Ethernet (AVB protocol) card (AVB-DG). The 32 signals from the four 8-channels acquisition cards are then gathered with an AVB switch and transmitted to the computer with an Ethernet cable.

## 3. Sound source angular localization

A real time, time domain DOA estimation algorithm was developed, which is based on estimates of the pressure $p_0$ and the 2 horizontal components $v_{0x}$ and $v_{0y}$ of the particle velocity at the center of the CMA, the CMA being placed horizontally on the floor. Every 85 ms, the estimated time samples of the normalized velocity and pressure $v_{0x}\rho_0 c_0, v_{0y}\rho_0 c_0, p_0$ (where $c_0$ is the celerity of the waves in the air) are plotted on the $(O, v_{0x}\rho_0 c_0, v_{0y}\rho_0 c_0, p_0)$ space, and a line that crosses zero is fitted from this data by using the RANSAC [4] algorithm. The localization angles $\theta_0$ and $\delta_0$ are estimated from the coefficients $X, Y, P$ (representing $v_{0x}\rho_0 c_0, v_{0y}\rho_0 c_0, p_0$ respectively) of the obtained leading vector :

$$\begin{cases} \theta_0 = & \text{atan2}\left\{-(Y/P), -(X/P)\right\} \\ \delta_0 = & \text{arccos}\left(\sqrt{(X/P)^2 + (Y/P)^2}\right) \end{cases} \tag{1}$$

with atan2 being the four quadrant arctangent function. The reason for an elevation estimate without measuring the $v_{0z}$ component with vertically placed microphones pairs is a simplification of the CMA design as well as a compensation of the floor effects by placing all the microphones at the same height in a 2D plane. $v_{0z}$ is implicitly inferred from $v_{0x}$, $v_{0y}$ and the air characteristic impedance $\rho_0 c_0$, under the assumptions that the CMA is placed on the floor and the sources are at positive elevation angles.

### 3.1 Central pressure estimation

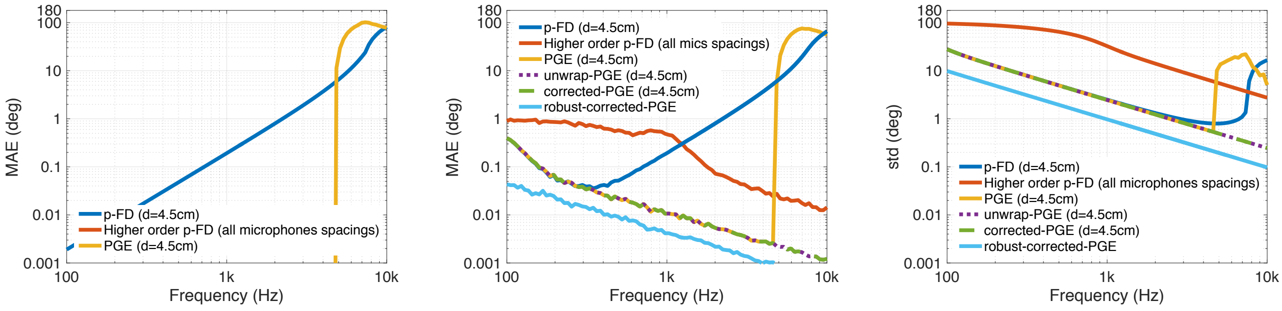
With the 32 MEMS sensor (see Fig. 2b), instead of directly measuring the central pressure by placing a microphone at the center of the probe, we estimate this quantity by averaging the signals of the four microphones which are at ±0.25 cm on the $\overrightarrow{e_x}$ axis and ±0.25 cm on the $\overrightarrow{e_y}$ axis. This simplifies the CMA design and allows uncorrelated noise reduction (6 dB), with an acceptable bias error on the pressure estimation (maximum error <0.5 dB at 10kHz for a spacing of 0.5 cm). Techniques can be used to reduce this bias at the price of noise amplification (or less noise reduction). These techniques involve using higher order accuracy pressure finite sums using multiple microphone spacings, and summing only the signals of the microphones that are on the axis that is estimated to be the most orthogonal to projection on the horizontal plane of the source's DOA.

### 3.2 Particle velocity estimation

The central pressure and the particle velocity $v_{0i}$ on the $i, i = \{x, y, z\}$ axis are linked by the Euler equation $v_{0i} = -\frac{1}{\rho_0}\int_0^t g_{0i} d\tau$, with $\rho_0$ the air density and $g_{0i}$ the $i$ component of the pressure gradient at the center of the CMA.

In this part, we present different potential approaches to estimate these components. The Fig. 3 compares these approaches. The localization of planar sine waves was repeated in the frequency domain for combinations of 1000 random draws of 30 dB signal noise noise and random phase applied to each microphone, 64 azimuth angles $\theta_i, i = \{1...64\}$ equally distributed between $-\pi$ and $\pi - {}^{\pi}/_{32}$, and 15 elevation angles $\delta_j, j = \{1...15\}$ equally distributed between 0 and ${}^{\pi}/_{4}$ degrees. For each input

DOA $(\theta_i, \delta_j)$, we define the average estimated azimuth $\widetilde{\theta_{i,j,\text{mean}}}(\theta_i, \delta_j)$ calculated from the 1000 noise draws. Figs. 3a and 3b represent the mean absolute error (MAE) defined as $\underset{\text{all }\theta_i,\delta_j}{\text{mean}} |\widetilde{\theta_{i,j,\text{mean}}}(\theta_i, \delta_j) - \theta_i|$. Fig. 3c represents the standard deviation to the estimated DOAs.



(a) MAE, SNR = $+\infty$ dB (b) MAE, SNR = 30 dB (c) STD, SNR = 30 dB

FIGURE 3 – Localization errors for different SNRs

**Pressure finite differences gradient estimation (p-FD)** One can obtain an estimate $\widetilde{g_{0i,\text{p-FD}}}$ of $g_{0i}$ with finite differences of pressure measurements from microphones that measures $p_{2i}$ and $p_{1i}$ at the positions $+d/2$ and $-d/2$ on the $i$ axis ($d$ being the distance between the 2 microphones) :

$$\widetilde{g_{0i,\text{p-FD}}} = \frac{p_{2i} - p_{1i}}{d} = g_{0i} \times \frac{\sin\left(k\frac{d}{2}A_i\right)}{k\frac{d}{2}A_i} \tag{2}$$

with the term highlighted in red being a bias term which depends on the wavenumber $k$, the microphone spacing $d$ and the ambisonic coefficient $A_i = -[\cos\theta_0\cos\delta_0, \sin\theta_0\cos\delta_0, \sin\delta_0]^T$ which contains the source direction information. A too small microphone spacing $d$ increases the sensitivity to noise and calibration errors (see Fig. 3c), while a too large microphone spacing increases the influence of the bias term at high frequencies (see Fig. 3a). A solution is to use microphones spacings that decreases for increasing frequencies, by using multiple multiple microphones pairs. In our case of a 2D CMA, this results in a CMA that contains multiple microphones pairs on the $x$ and $y$ axis, forming two orthogonal lines of microphones, see Fig. 2.

**Higher order pressure finite differences gradient estimation** The pressure finite difference error can be reduced by using higher order pressure finite differences [2]. The Fig. 3a shows that without noise the resulting azimuth error is very low when using higher order pressure finite differences with the 8 available microphone spacings. But the increase in estimation accuracy is achieved at the cost of noise amplification, that causes a high angle estimation standard deviation (see Fig. 3c) and a resulting high mean absolute error (3b) if we do not average multiple estimations.

**Phase differences pressure gradient estimation (PGE)** The pressure finite difference error can be suppressed using the Phase And Gradient Estimation (PAGE) method [3]. It consists in replacing pressure differences by pressure amplitude and pressure phase differences. The pressure difference bias error is suppressed with the PAGE method. Since we assume that the sources are in the far field, pressure amplitude differences can be neglected, and we can consider an estimate $\widetilde{g_{0i,\text{PGE}}}$ of $g_{0i,\text{PGE}}$ with Phase differences (only) based Pressure Gradient Estimation (PGE) :

$$\widetilde{g_{0i,\text{PGE}}} = j\frac{\text{phase}(p_{2i}) - \text{phase}(p_{1i})}{d}p_0 = -jkA_i p(x=0) + \text{phase ambiguity} \tag{3}$$

Without noise and while $d\lambda < 1$, PGE method offers a very small error, which globally (except when phase ambiguity occurs) decreases with the source distance. Phase ambiguities can cause very large errors (see the yellow line in Fig. 3a). These ambiguities can be suppressed by phase unwrapping (see the unwrap-PGE method on Fig. 3a), provided that phase unwrapping is feasible. In the presence of noise, phase unwrapping can be replaced by replacing the $i$-th pressure gradient component $\widetilde{g_{0i,\mathrm{PGE}}}^{(k)}$ estimated with the sensor spacing number $k, k = \{1...8\}$ by $\widetilde{g_{0i,\mathrm{PGE}}}^{(k)} -$ round $\left\{\frac{d_k}{2\pi}\left(\widetilde{g_{0i,\mathrm{PGE}}}^{(k)} - \widetilde{g_{0i,\mathrm{PGE}}}^{(1)}\right)\right\}$ where $d_k$ is the $k$-th sensor spacing, $\widetilde{g_{0i,\mathrm{PGE}}}^{(1)}$ the estimate obtained with the smallest microphone spacing. The effect of this *corrected*-PGE estimation is to shift towards higher frequencies the appearance of phase ambiguities (see the green line in Fig. 3b). Finally, a more robust to noise PGE estimation (*robust-corrected*-PGE estimation, see Fig. 3c) can be obtained by averaging the PGE estimations obtained with multiple large spacings.

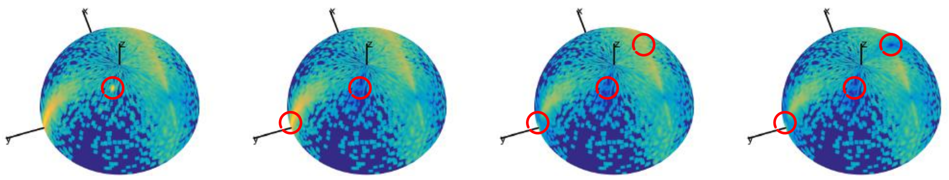
## 3.3 Discussion

Experimental measurements using a previous CMA prototype and an associated localization algorithm were conducted. The results [5] show a mean absolute error of 5 degrees, which is a good first estimate of the source direction for the orientation of the imaging system developed by ISL.

The observed noise is filtered by using the RANSAC algorithm, and its effect is reduced by using frequency dependent microphones spacings. At each frequency, a strategy is to use the order 1 estimation with the largest spacing that gives an acceptable maximum error (say 3 degrees). We use the largest microphone spacing for the lowest frequency and for increasing frequencies until the maximum error reaches the fixed limit. We repeat the same procedure for higher frequencies with smaller microphones spacings, until no smaller microphone spacing is available. This results in a high frequency limit of the sensor bandwidth, which is extended with the use of higher order pressure differences for frequencies above this limit. PGE algorithm may be a good alternative which would need a smaller number of microphones spacings, provided that we can remove the phase ambiguities with real microphones signals.

## 3.4 Towards multiple sources localization

Multiple sources localization is currently under study. Our current strategy, based on [6], is to perform single source localization on multiple time-frequency zones, and to count the occurrences of each found directions on a localization histogram (see Fig. 4). At each iteration, we consider as a source direction candidate the direction associated with the highest peak on the localization histogram and then suppress an estimate of the contribution of this potentially detected source to the histogram localization, to prepare the next iteration.



(a) Iteration 1 (b) Iteration 2 (c) Iteration 3 (d) Iteration 4

FIGURE 4 – Example of a localization histogram.

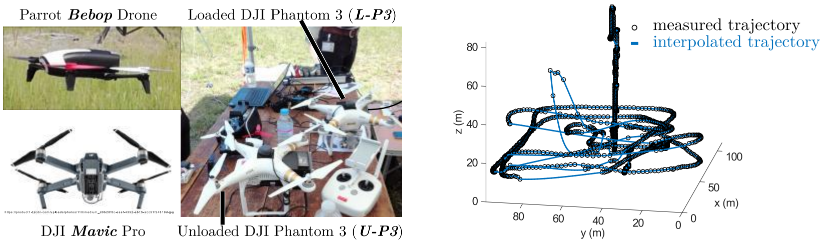
Kalman or particle filtering could be applied to reject some outliers candidates over time. Then, the remaining potential sources DOAs can then be used to selectively beamform on each of these, ending with spatially filtered signals which could facilitate sources identification (see Fig. 1). In this regard, differential beamforming and minimum variance distortionless response (MVDR) beamforming were compared in [1].

# 4. Sound source detection and identification

Source sound detection and identification can be performed using machine learning. The principle is to use binary classification to estimate the presence or absence of a drone sound in a sound mixture. Both initial detection and final identification are binary classification tasks. Initial detection is a background process whose objective is to fastly (fewer than 1 second) detect the potential presence of a drone with a low false negative rate and low computational resources. If a detection threshold is exceeded, sources localization and beamforming are triggered, and the spatially filtered sources signals are fed to a second binary classifier for a final identification, which can eventually be more computationally demanding, and be performed on a longer term (more than 1 second). Results on experiments with short term initial detection using the JRip [7] classifier from the WEKA library are presented here. Longer term final identification using deep neural networks is currently under study.

## 4.1 Measurement campaign

A 3 days measurement campaign was conducted with 4 flying drones (see Fig. 5a) in a countryside (Baldersheim, France) (see Fig. 2d) with ambient noises including birds, insects, people speaking, detonations and fire shots noises. The recorded drones were a Parrot *Bebop* drone, a loaded DJI phantom 3 (*L-P3*), an unloaded DJI phantom 3 (*U-P3*) and a DJI Mavic Pro drone. A whole variety of drones trajectories, flight phases and drones-to-CMA distances were observed. The sound were recorded with the last CMA prototype, both in the presence and in the absence of a flying drone. A GPS-RTK system was used to measure the trajectory of the drones in the coordinate system of the CMA. This trajectory can be used as a ground truth trajectory for localization experiments, or to use the drone distance as a parameter for the drone detection experiments.



(a) Used drones

(b) Example of a measured trajectory using GPS-RTK

| dist. (m) | [0-50] | [50-100] | [100-200] | [200-400] |
|---|---|---|---|---|
| Bebop | 04 :29 | 02 :57 | 0 | 0 |
| Mavic | 02 :15 | 05 :15 | 03 :34 | 0 |
| L-P3 | 05 :25 | 14 :17 | 01 :15 | 00 :07 |
| U-P3 | 07 :46 | 27 :45 | 05 :29 | 01 :17 |

(c) Drones flight time

FIGURE 5 – Measurement campaign

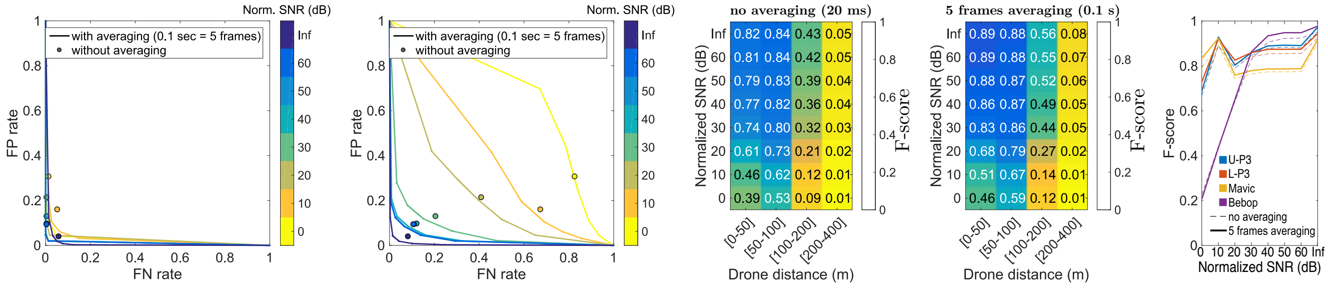
## 4.2 Database construction

The recorded sounds (noted as "*Baldersheim* sounds") are randomly mixed with sounds from the DCASE 2016 residential sounds data base [8] (noted as "DCASE sounds"), because preliminary detection tests has shown that detection with noise corrupted test data is facilitated when using noise corrupted training data. Different Normalized SNR from 0 to 60 dB are used in the training data, the Normalized SNR being the relative global level between Baldersheim sounds in the absence of drone, and the global level of DCASE 2016 sounds. The 2/3 first samples of both Baldersheim and DCASE sounds are dedicated to the training database, while the 1/3 last samples are dedicated to the test database. When doing a classification exercise, as much positive (label 1 : Baldersheim with flying drones + DCASE mixtures) and negative (label 0 : Baldersheim without drones + DCASE mixtures) are used, and we ensure that the training data has as much examples for the 4 available drones, and, if possible, as much data corresponding to drones flying from the distances [0 to 50 meters], [50 to 100 meters], [100 to 200 meters], [200 to 400] meters.

### 4.3 Classification

We use the JRip classifier implemented in the WEKA library together with 13 MFCC [9] coefficients (calculated from a bank of mel scaled bands from 200 to 8000 Hz) and the spectral roll-off, flatness, entropy, irregularity and brightness [9], calculated from 20 ms audio frames. We selected this set of features by using an evolutionary algorithm from a larger set of features. The JRip classifier was used because it provided good classification performances with a small amount of training data and no classifier tuning. Drone presence predictions are made for each audio frame, and are averaged on 5 consecutive frames (0.1 s) chunks, thus merging 5 consecutive drone presence binary probabilities into 1 absolute drone presence probability on which a detection threshold is fixed to obtain a cost-sensitive classifier.

The Fig. 6a represents the false negative (FN) VS false positive (FP) plot using varying detection thresholds on the averaged predictions, for several SNR values for the L-P3 drone. The same plot for the Parrot Bebop drone is plotted on Fig. 6b. We can see that to obtain a decreasing amount of FN rates we have to accept an increasing amount of FP rates. For initial detection we want to chose a rather small detection threshold at the cost of a rather high FP rate. For all the drones except the Parrot Bebop, we obtain a strong *L*-shaped FP rate VS FN rate curve for all SNR values (see Fig. 6a). This means that a low FN rate can be obtained along with a relatively low FP rate, except for the Parrot Bebop drone. This exception may be explained by lack of data and/or non adapted audio features. The Bebop sound signature was quite different from the others, and we collected less recordings for this drone, see Fig. 5c. Even if the training time was the same for each drone (2 minutes), the diversity of recorded sounds may be smaller for a drone that has flied for a smaller total period of time, because the randomly selected samples used as training data arise from very close time samples from the audio recordings, and chances are higher for the Parrot Bebop drone that the same audio samples are trained multiple times, mixed with different DCASE data.



| Normalized SNR (dB) | [0-50] | [50-100] | [100-200] | [200-400] |
|---|---|---|---|---|
| Inf | 0.82 | 0.84 | 0.43 | 0.05 |
| 60 | 0.81 | 0.84 | 0.42 | 0.05 |
| 50 | 0.79 | 0.83 | 0.39 | 0.04 |
| 40 | 0.77 | 0.82 | 0.36 | 0.04 |
| 30 | 0.74 | 0.80 | 0.32 | 0.03 |
| 20 | 0.61 | 0.73 | 0.21 | 0.02 |
| 10 | 0.46 | 0.62 | 0.12 | 0.01 |
| 0 | 0.39 | 0.53 | 0.09 | 0.01 |

| Normalized SNR (dB) | [0-50] | [50-100] | [100-200] | [200-400] |
|---|---|---|---|---|
| Inf | 0.89 | 0.88 | 0.56 | 0.08 |
| 60 | 0.89 | 0.88 | 0.55 | 0.07 |
| 50 | 0.88 | 0.87 | 0.52 | 0.06 |
| 40 | 0.86 | 0.87 | 0.49 | 0.05 |
| 30 | 0.83 | 0.86 | 0.44 | 0.05 |
| 20 | 0.68 | 0.79 | 0.27 | 0.02 |
| 10 | 0.51 | 0.67 | 0.14 | 0.01 |
| 0 | 0.46 | 0.59 | 0.12 | 0.01 |

(a) Varying detection threshold, L-P3

(b) Varying detection threshold, Bebop

(c) F-score, L-P3, without averaging

(d) F-score, L-P3, 5 frames averaging

(e) Global F-score for all distances, all drones

FIGURE 6 – Detection scores (without spatial filtering)

The F-score, being the harmonic mean of the precision (ratio of the number of true positive predictions and the number of positive predictions) and the recall (ratio of the number of true positive predictions and the number of positive examples), is a measure of the global performance of a classifier. This score is globally decreasing for increasing distances and decreasing SNR values (see Fig. 6c for the L-P3 drone) : it is harder to detect a drone when it is far from the CMA or in loud ambient noise. The F-score increases when frame averaging is applied (see Fig. 6d for the loaded DJI Phantom drone). These trends are also observable for the 3 other tested drones, see Fig. 6e. The global F-scores are above 0.6 even for 0 dB SNR. Averaging on a increasing period of time increases the F-score, but this increase in F-score becomes progressively negligible for increasing averaging time : the global F-score (all drones, test SNR between 0 and 60 dB) being [0.834, 0.852, 0.850, 0.853, 0.854, 0.859, 0.866, 0.873] for averaging on [1, 2, 5, 13, 25, 50, 113, 250] frames ([0.02, 0.04, 0.1, 0.26, 0.5, 1, 2.226, 5] seconds averaging). This justifies the choice for an averaging time of 0.1 seconds.

## 5. Conclusions and future work

A prototype of a new compact microphone array for acoustic source localization and identification has been presented, along with a new localization technique, which uses the RANSAC algorithm in the time domain in order to estimate the source direction from estimates of the pressure and 2 components of particle velocity at the center of the sensor. Different techniques were compared to estimate these acoustic quantities. Central pressure is estimated by using pressure finite sums. Pressure finite differences are used together with frequency-dependent microphone spacings to estimate the particle velocity. Extension of the obtained bandwidth is obtained by the use of higher order pressure finite differences at very high frequencies. Pressure gradient estimation may be an alternative to pressure finite differences for a use with less microphones, provided that phase unwrapping can be performed with real microphones signals.

Multiple drones acoustic signatures were recorded, and their detection were performed by using supervised binary classification. A relatively high F-score was obtained by using the JRip classifier from a selected set of acoustic features. The F-score is decreasing for increasing background noise and for increasing drone-sensor distance. In this regard, beamforming techniques could be used to facilitate source identification, provided that it does not alter the source's acoustic signature.

A final identification on a longer period of time is under study. Two approaches are developed : the construction of higher level features from statistics and operations on descriptors observed in multiple consecutive frames, and the analysis of a spectrogram-like image using deep neural networks.

## Acknowledgments

This work is financially supported by the French Ministry of Defense - Direction Générale de l'Armement (DGA)

## REFERENCES

1. Ramamonjy, A., Bavu, E., Garcia, A., Hengy, S., A distributed network of compact microphone arrays for drone detection and tracking, *The Journal of the Acoustical Society of America*, **141** (5), 3651, (2017).
2. Fornberg, B., Generation of finite difference formulas on arbitrarily spaced grids, *Mathematics of Computation*, **51**, 699, (1988).
3. Thomas, DC., Christensen, BY., Gee, KL., Phase and amplitude gradient method for the estimation of acoustic vector quantities, *The Journal of the Acoustical Society of America* **137** (6), 3366–3376, (2014).
4. Fischler, M. A., Bolles, R. C., Random sample consensus : a paradigm for model fitting with applications to image analysis and automated cartography, *Readings in computer vision*, 726–740, (1987).
5. Ramamonjy, A., Bavu, E., Garcia, A., Hengy, S., Détection, classification et suivi de trajectoire de sources acoustiques par captation pression-vitesse sur capteurs MEMS numériques, *Actes du 13ème Congrès Français d'Acoustique*, 1083–1089 (2016).
6. Delikaris-Manias, S., Pavlidi, D., Pulkki, V., Mouchtaris, A., 3D localization of multiple audio sources utilizing 2D DOA histograms, *24th European Signal Processing Conference (EUSIPCO 2016)*, 1473–1477, (2016).
7. Cohen, W. W., Fast effective rule induction, *Machine Learning Proceedings 1995*, 115–123, (1995).
8. Mesaros, A., Heittola, Toni., Virtanen, T., Tut database for acoustic scene classification and sound event detection, *24th European Signal Processing Conference (EUSIPCO 2016)*, 1128–1132, (2016).
9. Peeters, G., A large set of audio features for sound description (similarity and classification) in the CUIDADO project, (2004).