



A Cooperative Time-Frequency Approach to Detect, Recognize and Track Drones with Audio Sensor Networks

Claudio Santo Malavenda and Carlo Malavenda¹, Marco Martalò²

¹Claudio S. and Carlo Malavenda, Freelance, Rome

²Marco Martalò, University of Parma, Parma
ITALY

Email: claudosanto.malavenda@gmail.com

ABSTRACT

This work aims at proposing a detection, recognition and tracking solution for Unmanned Aircraft Systems (UAS) with a wireless audio sensor network. According to technology trends applicable to UAS (smaller, cheaper and cooperative), we propose a distributed surveillance solution with the same technology approach of the 'attacker' one. In particular, since the drone causes a variation of the surrounding acoustic environment, we investigate the use of an audio sensor network. More precisely, a three-phase algorithm is employed to detect the presence of audio energy in the monitored environment, recognize a particular audio signature and then cooperate with a multiple node approach to track the drone. We show preliminary performance of the proposed approach by relying on experimentally acquired audio signals. We also discuss the future work to improve the actual implementation.

1.0 INTRODUCTION

Unmanned Aerial Systems (UAS) took huge popularity in recent years and their proliferation has become always more persistent in military conflict as in everyday life. The wide spectrum of UAS variety and offer in military and civil market allows to choose the best fitting UAS for each particular situation and mission with very accessible costs. In particular, mini and micro drones are easily available on the market at low prices. Moreover, a payload capability of up to 2 kg makes them a possible threat that must be contrasted to avoid their use by criminal people and organizations to realize terroristic attack or demonstrative operations.

This work proposes an audio-based detection, recognition and tracking method for UAS. The goal is to investigate the achievable results in the tracking and recognition of such commercial small drones (Class I UAS), combining commercial off-the-shelf solutions and state-of-the-art algorithms. In particular, the key idea of the investigated scheme is to analyse, by relying on a deployed audio sensor network, the UAS sound signature with a layered algorithm, thus making its computational complexity and energy consumption as limited as possible.

2.0 REFERENCE SCENARIO

The low cost of mini & micro drones (1 keuro and below), as well as the easiness of procurement, reduces to zero the barriers that a terrorist may find to use this technology. In addition, the open source technology, often used to design some components of the UAS system, delimiters a very small gap of knowledge to fill from whoever aims to design a dedicated payload. A flying platform so easily procured and personalized could finally access sensible targets. According to this trend, squadrons of coordinated drones could soon become a resource accessible to anyone. The countermeasure of such menace cannot be a centralized. As of today, we see asymmetric solution applicable to Forward Operating Bases or quite desert scenarios with large drone detector installation. However, the potential pervading presence of such menace in urban

scenarios, suggest to adopt a symmetric countermeasure, i.e., distributed, small and cheap.

In particular, the constant evolution of wireless sensor networks in last years, as well as the node miniaturization and low cost, could suggest the most suitable solution for urban environment or complex scenario, where civil presence is possible and fixed military installation may be not suitable.

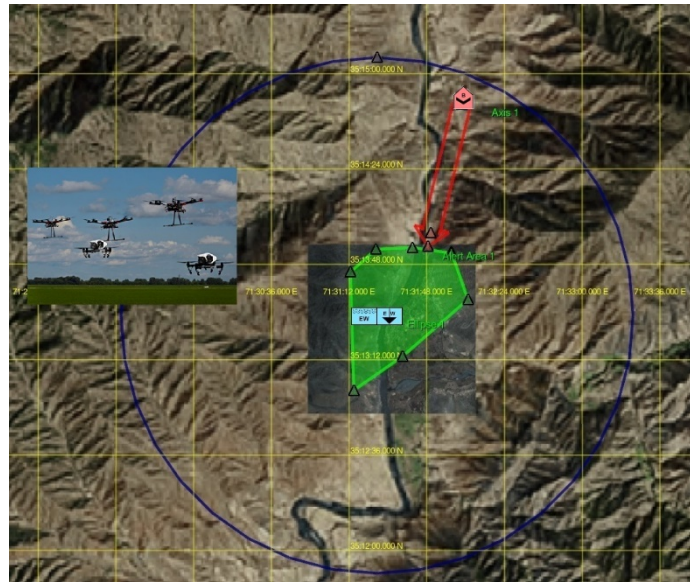


Figure 2-1: Reference scenario.

As depicted in Fig. 2-1, this work will take as reference a scenario a large number of cheap audio sensors, each one capable to recognize the drone's audio signature and, when a match is found, locate the malicious source, by collaborating with other nodes. The focus of the research is to detect, recognize and track with audio array a single drone or a small swarm of drones, i.e. assimilable to a single large unit, in an open field scenario.

3.0 PROPOSED APPROACH

3.1 The three layers

Since the goal of the proposed method is three-fold (i.e., detect, recognize, and track a malicious drone), our approach uses three algorithm layers stacked. A general description of the pursued research methodology is given in Fig. 3-1. We rely on the acquisition of audio signals that can accurately characterize the presence of an UAS. The first layer addresses the detection of the presence of an UAS. In this phase, the audio sensor takes few samples from the environment, e.g., every second, in order to reveal anomalies in the energy of the audio perceived from the environment.

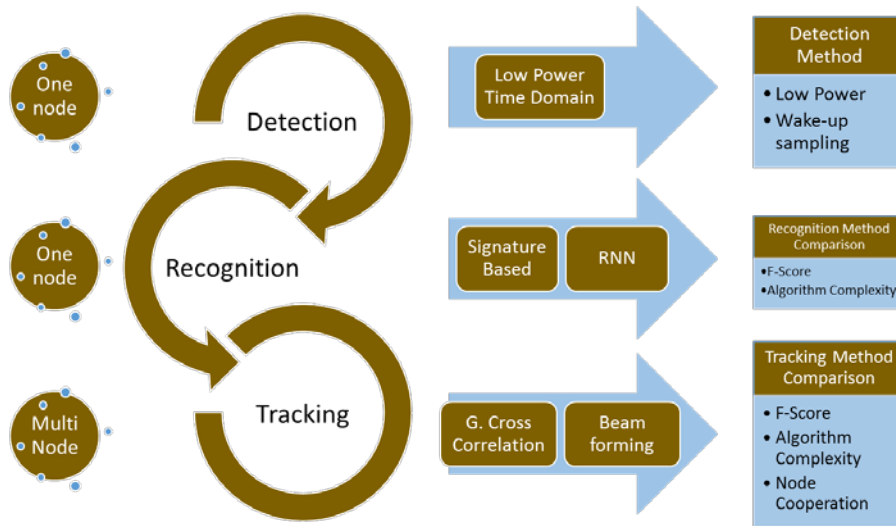


Figure 3-1: Research methodology.

When a match is detected in this layer, the second one implements the recognition with continuous sampling for a short amount of time, e.g., for about 240ms [3]. The goal of the recognition phase is to distinguish if the anomalous sound is associated with an UAS flying over the audio sensor network and, eventually, its type. We will investigate two main methodologies: the one proposed by some of the authors in [1] and the Recurrent Neural Network (RNN) [6]. A positive matching in the second recognition layer will enable the tracking algorithm of the third phase. In this phase, the alerted node (the one with a positive match in the recognition phase) wake up neighbours' node to perform a beam forming tracking. This is the most power consuming phase, due both to the communication link (between the audio sensor network and its remote-control centre) to maintain and to the continuous sampling phase of the audio sensor.

The adopted layered strategy should optimize computational power and battery need. In fact, the first layer, which is always active, performs low-complexity and consumption mathematical computation. On the other hand, more powerful computations, i.e., fine-grained signature recognition and audio tracking, are performed only in the presence of an anomalous audio.

3.2 Markov Model

It is possible to model the evolution of the proposed algorithm with a Markov chain. In particular it is possible to establish the Markov chain model with the associated Stochastic Transition Matrix displayed as the following P matrix.

$$P = \begin{pmatrix} A & B_1 & 0 & 0 \\ B_2 & 0 & C_1 & 0 \\ C_2 & 0 & D_2 & D_1 \\ 0 & 0 & D_3 & E \end{pmatrix} \quad (1.)$$

This generates the following graph that describes the possible transition among the algorithm states.

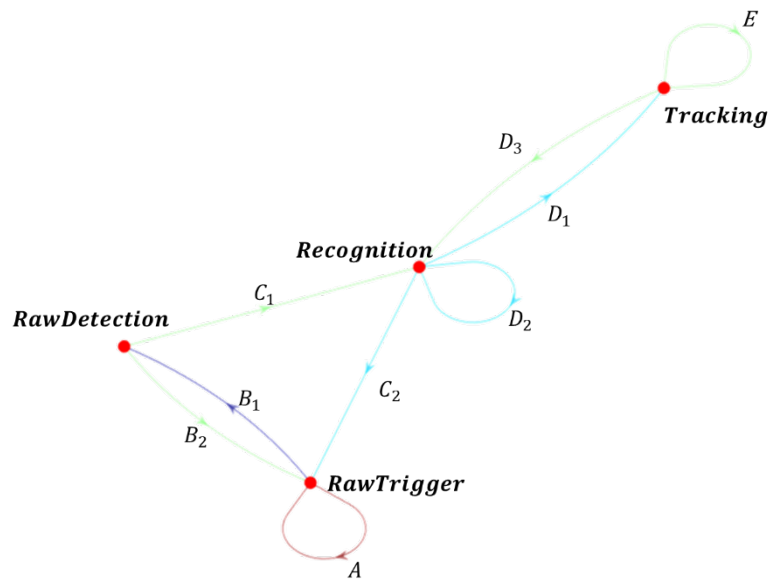


Figure 3-2: Graphic view of the Markov chain model associated to the proposed algorithm.

where

- A is the probability to detect no alarm in Raw Trigger*
- B₁ is the probability to detect an anomaly in Raw Trigger*
- B₂ is the probability to false detection from Raw Trigger*
- C₁ is probability to detect an anomaly in Raw Detetcion*
- C₂ is probability to NOT recognize a drone in 'Recognition'*
- D₁ is the probability to recognize a drone in 'Recognition'*
- D₂ is the probability to keep recognizing a drone with no other collaborating nodes available*
- D₃ is probability to lost tracking*
- E is the probability to keep the tracking*

Having empiric values or estimates of the relevant probabilities in Fig. 3-2, it possible to evaluate the evolution of the algorithm and finally obtain the probability to recognize and track an unattended drone. To this end, an experimental evaluation of all phases of the algorithm is needed. In this paper, we will present the state of the research with the implementation and testing reached in each state of the algorithm and some indications of future work

4.0 LEVEL 1 ALGORITHM – DETECTION

4.1 Event Modelling

In order to start to model the relevant signature, we use some measurement performed in [8] and summarized in Fig. 4-1. The test performs some sound level measurement of different drones: three measurements (called Test 1) are performed leaving the drone flying in a fixed position and moving the meter at 2ft, 10ft and 20ft, two (called test 2) are performed moving the drone in the z axis and taxing measurement from below when it

flies at 25ft and 100ft.

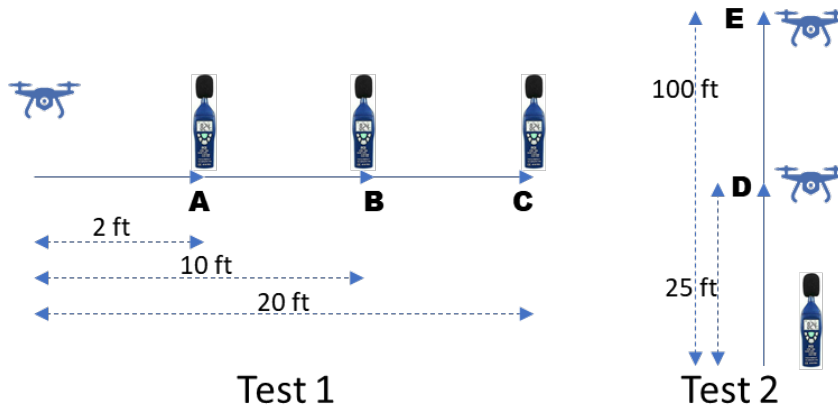


Figure 4-1: Measurement Campaign

Assuming to have the sound meter at a constant height of one meter, it is possible to transform the Cartesian coordinates of the Test geometry into the following spherical points of measure, expressed as radius in meters and phi in degree:

$$(r ; \varphi)$$

$$A = (1,17 ; 58,63)$$

$$B = (3,21 ; 18,16)$$

$$C = (6,18 ; 9,32)$$

$$D = (7,64 ; 85,43)$$

$$E = (30,49 ; 88,85)$$

Following figure resumes values of sound pressure registered with a sound level meter for the following drones: DJI Mavic 2 Pro; DJI Mavic AIR; DJI Mavic Mini; DJI Spark.

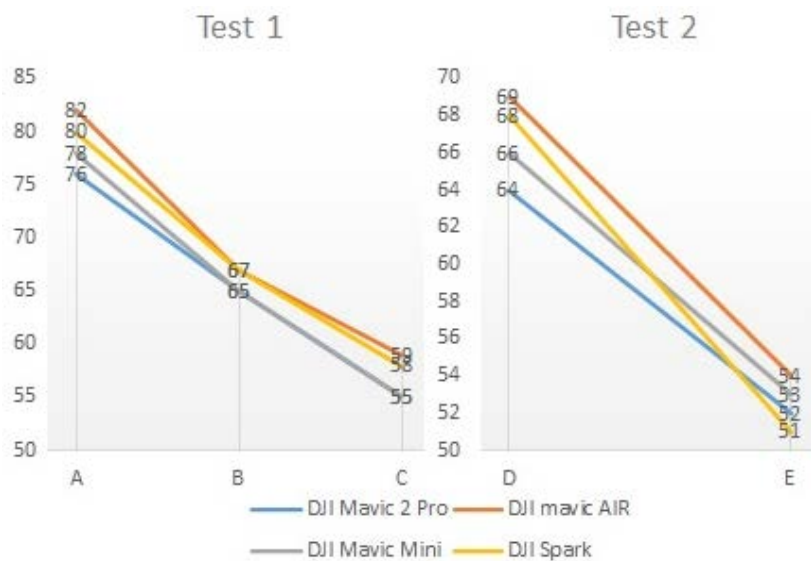


Figure 4-2: Sound pressure: comparison of drones

The quite high noise level of the drone rotors allows us orient to implement the first level of the detection algorithm with an energy-based approach, focusing the detection in ‘anomalies’ that the presence of such audio signal would introduce in the environment.

4.2 Set Up Time

This first stage of the audio processing is based on a low power technique that cyclically power on and off the audio circuits to sample the signal and detect “anomalies” in the audio level.

The first step to put in place this method is to have a stable signal: when the audio circuit is triggered on, the output signal can bounce for a certain period. In order to avoid that the ADC samples an unstable signal, it is possible to measure the power absorption. We will assume that a constant current drain will be linked to a stable output signal. Such set-up time T_{on} is the minimum time that the sampling firmware has to wait between the cyclic sensor power-on and the start of the sampling.

4.3 RAW Sensing - Triggering

According to [2], this approach is based on per-frame processing, where a frame corresponds to a sequence of consecutive discrete-time samples. Denoting as K the number of samples per frame, the average per frame SNR can be defined as follows:

$$SNR \triangleq \frac{E_{voice}}{E_{noise}} = \frac{\frac{\sum_{i=1}^k |s_i|^2}{k}}{\frac{\sum_{i=1}^k |n_i|^2}{k}} = \frac{\sum_{i=1}^k |s_i|^2}{\sum_{i=1}^k |n_i|^2} \quad (2.)$$

where

$$x_k = \begin{cases} s_k + n_k & \text{in the presence of an atypical signal} \\ n_k & \text{in the absence of any atypical signal} \end{cases}$$

The node performs an audio sampling sequence every wake-up time. This allows to perform an energy-based detection with a minimum set of samples and a minimum set of nodes wake up. Obviously, the wake-up time should be such that relevant events are not missed by the audio sensor network. The anomalies detection is performed establishing a dynamic threshold value, according to recorded past samples, wrapped by an upper and a lower boundary value. Every T_{active} wake-up cycle a “ l ” number of samples (c_i) are stored in ADC buffer with a sampling period of T_{sample} .

All c_i samples from a T_{active} phase are then merged into a single average s_j sample with the following:

$$s_j = \frac{\sum_{i=0}^l c_i}{l} \quad (3.)$$

Each s_j sample is then stored in a circular buffer memory with a length of “ k ” samples. After an initial number of ‘ k ’ T_{active} periods, the memory buffer is full of s_j and is possible to calculate $\overline{s_j^k}$ as their mean value with

$$\overline{s_j^k} = \frac{\sum_{l=-k}^j s_l}{k} \quad (4.)$$

This will be the floating threshold value.

For the detection of anomalies, we add a fixed confidence bandwidth denoted as Δ so that the confidence interval is equal to $[\overline{s_j^k} + \Delta, \overline{s_j^k} - \Delta]$. When a s_j sample occur inside this interval, the Raw sensing phase continue and the value is used to update the sample buffer. When the value is measured out of the confidence interval, an anomaly detection processing starts.

4.4 RAW Sensing – Anomaly Detection

The anomaly detection phase has the purposes to filter false alarm finding out an anomalous behaviour in the digital values of the sampled signal. In this paper, we will propose two different approaches.

- The first one, called below ‘trigger counter’, will activate the Level 2 phase when the raw sensing triggering detect an alarm an ‘ α ’ consecutive number of times.
- Alternatively, the second one called ‘trigger follower’ will interrupt the raw sensing and starts a continuous sampling to detect the anomaly. This case introduces a continuous optimization and variability of §4.2 algorithm parameters, e.g., with a Particle Swarm Optimization (PSO) technique.

It is important to notice that, according to those methods, this first level of detection is engaged with any type of audio noise detected above the floor noise. We assume that in the case of a swarm we well have a larger amount of energy to detect and it will be even more responsive.

4.4.1 Trigger Counter

This first approach allows to perform the detection while continuing the wake-up scheduling performed into the raw sensing phase. The main benefit is to save battery energy, while maintaining a sufficient level of detection of relevant events.

The following figure shows s_j values of the sampled signal, values from the interval $[\overline{s_j^k} + \Delta, \overline{s_j^k} - \Delta]$ and in red signal values detected as anomalous. The trigger-counter detects anomalies when the signal has a considerable variation.

This is the case when a flying drone enters the detection range of the audio sensor.

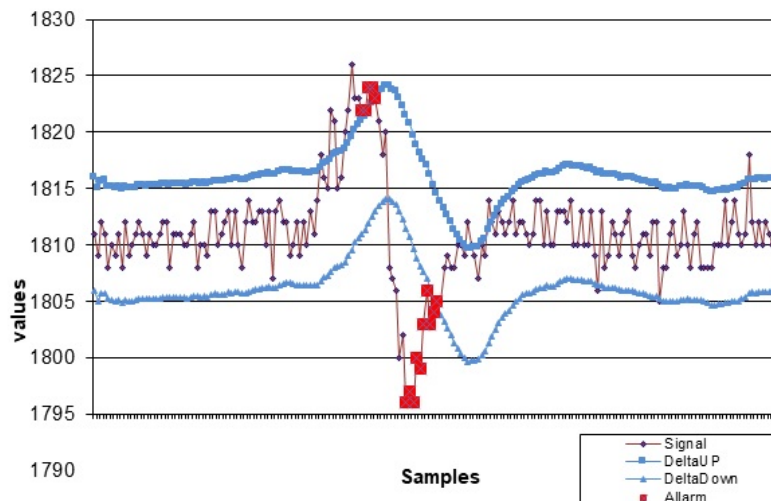


Figure 4-3: Algorithm Representation in the trigger counter phase.

As reported in the figure, the algorithm detects an anomaly counting how many times the sampled signal outranges the confidence level. According to the reference scenario and surrounding noise, it could be very difficult to use the same set of parameters for all type of drone detection, but it would be the best choice for a battery powered device with a long service-time.

Table 4-1: Trigger Counter Parameter Values.

Parameter	Value
K	24
Δ	5
A	4

A pseudo-code implementation of this algorithm is reported below.

```

for i = (K+1):signal-length
    for j= (i-K):i
        delta=delta+y(j);
    end
    delta=delta/K;
    threshup(i)=y(i)+delta;
    threshdown(i)=y(i)-delta;
    if (y(i)>threshup(i)) || (y(i)<threshdown(i))
        countA=countA+1;
        if countA>A
            allarm
        end
    else
        countA=0;
    end
end
end
    
```

4.4.2 Trigger Follower

This adaptation of the previous algorithm performs a continuous updating of parameters [k, Δ , a] using a Particle Swarm Optimization (PSO) approach.

We defining the following “position” vector as:

$$\vec{x}_i = [k, \Delta]; \quad (4)$$

Where l identifies the number of different audio sensors, and setting their initial values as in Table 4-1.

Assuming to use the following law to update location and speed values:

$$\begin{cases} x_i \in \mathbb{R}^2; & \vec{x}_i(t) = \vec{x}_i(t-1) + \vec{v}_i(t) \\ v_i \in \mathbb{R}^2; & \vec{v}_i(t) = \vec{I}(t-1) + \vec{M}(t-1) + \vec{C}(t-1) \end{cases} \quad (5.)$$

In which the values of inertia, memory and cooperation are expressed by the following ones:

$$\begin{cases} \vec{I}(t-1) = \varphi_{ic} * \vec{v}_i(t-1) \\ \vec{M}(t-1) = \varphi_1 rand(\vec{p}_i - \vec{x}_i(t-1)) \\ \vec{C}(t-1) = \varphi_2 rand(\vec{p}_g - \vec{x}_i(t-1)) \end{cases} \quad (6.)$$

where values of φ are adopted as ones in following

Table 4-2: Trigger Follower φ Parameter Values

Parameter	Value
φ_{ic}	0.6
φ_1	2
φ_2	2

We will also assume a limited speed value of $v_{max}^k = 12$ and $v_{max}^\Delta = 7$ which in the following examples will be used as empirically found values.

In order to face the limitation in a real device of the size of the memory buffer we model the k parameter to be limited to the maximum value for 24 as a reference dimension for a buffer array in an embedded device.

As fitness functions we will assume:

$$\begin{cases} F(k_i) = \left| \left(\frac{\sigma_i^{std}}{s_i} * K_MAXBUFFERSIZE \right) - k_{i-1} \right| \\ F(\Delta_i) = |(s_i - \bar{s}_i) - \sigma_i^{std}| \end{cases} \quad (7.)$$

Where s_i is the value of the signal at the current sampling cycle and σ_i^{std} represents the value of the standard deviation calculated on a number of samples buffered in the signal indicated by the value of k_{i-1}

At each evaluation cycle, the optimum positioning values occupied by the particle-node will be evaluated by the following:

$$\vec{p}_i = \begin{cases} p_i^k = MIN[F(k_{(i-k_i)}); \dots; F(k_i)] \\ p_i^\Delta = MIN[F(\Delta_{(i-k_i)}); \dots; F(\Delta_i)] \end{cases} \quad (8.)$$

Following figures reports values obtained for simulation in this phase.

In particular, Fig 4-4 shows values of the reference signal with dynamic parameters.

We can observe that using a PSO algorithm to optimize parameters in real time we have introduced a dynamic responsiveness to signal changes and an inertia of reaction. In fact, in this case we have a value of k and Δ variable and optimized at each new sample calculated by the PSO algorithm.

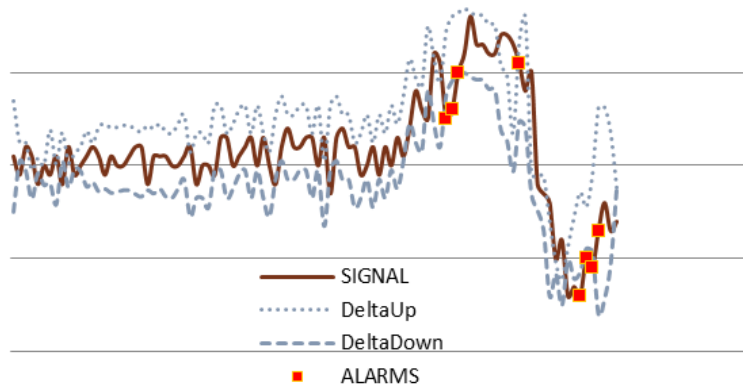


Figure 4-4: PSO Application to Trigger Follower A.

Due to the adaptability of this method, we also notice fewer alarms that are produced against the same event. The random parameters with which velocity are calculated introduce in each iteration of the multiplier values of memory and speed cooperation components that cannot change the result much in terms of revealed alarms. In this regard, you can see in Fig. 4-5 a representation of a second solution obtained from the PSO algorithm using the same input values, but applying different random values to each iteration to evaluate velocities.

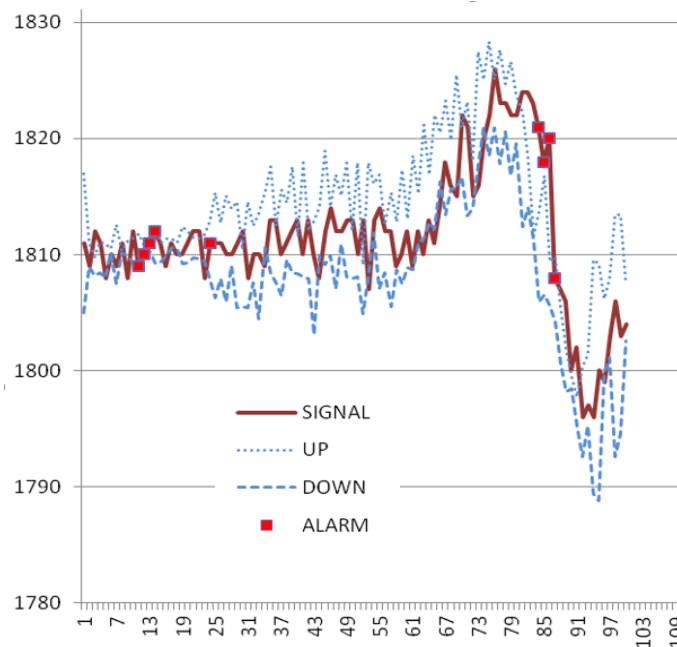


Figure 4-5 PSO Application to Trigger Follower B.

5.0 LEVEL 2 ALGORITHM – RECOGNITION

If the energy-based first level of the algorithm finds a match, the following recognition phase is starting its computing.

This investigation focuses on recognition of the type of the anomalous audio signal and, if it corresponds to a

drone, it tries to identify the specific type of UAS. The recognition phase will investigate the applicability of two main algorithms. The first one is an audio recognition algorithm developed for a military tactical wireless sensor network. The study will investigate the frequency detection implemented by [2] to the detection of drones. Performance achieved can be measured in terms of F-score and algorithm coding complexity. Such results can be compared with customization of commonly known algorithm from literature such as RNN.

In this work, we preliminary investigate the spectral characteristics of the audio signals to be identified; future work under development will feed the detection phase with empirical gained dataset to test the algorithm performance.

5.1 SIGNATURE-BASED

This recognition algorithm is based on the matching of a frequency signature of the drone.

The scope of the algorithm is to compute a signature in the ‘learning phase’ using audio samples captured from drones. The learning phase occurs offline when the recognition algorithm is not running. When the signature is available, the algorithm matches the real-time data with signature previously stored.

Using audio files available online, it is possible to recognize that different drones may have different audio spectra. In particular using the audio sampling campaign in [8], we have compared the spectral signatures using MATLAB. In particular, test has been conducted with following drones:

- Mavic mini
- Mavic 2 Pro
- Mavic AIR
- DJI Spark

The following Fig 5-1 compares the FFT of above drones in cases A, C and D described in Fig. 4-1 plus the take-off phase of such drones.

From the comparison, it is possible to see main differences from one model to another and also from take-off phase from a flying phase. In fact, this first phase it seems richer in frequencies. According to the use case, we will assume that the drone to be detected is jet flying and approaching the surveyed area. As investigated into [1], in order to calculate the signature, we filter the FFT calculated from all samples captured by the same type of drone, adopt a normalization of values and then we get the signature by performing the mean of all normalized FFT.

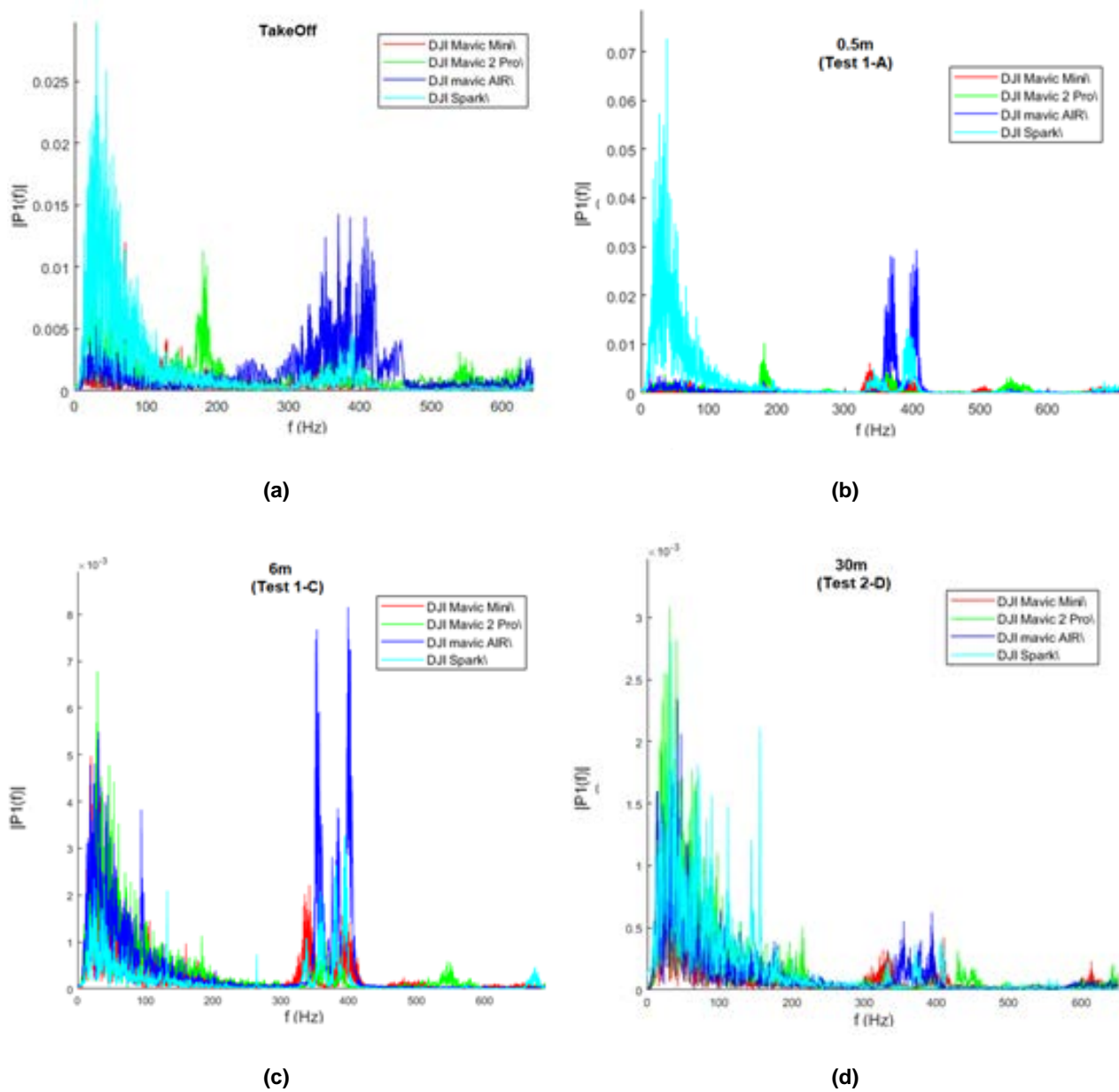


Figure 5-1: Test comparison of drones' spectrum captured at different distances (a) Take off, (b) distance 0,5m, (c) distance 6m, (d) distance 30m.

Using this method for all drones we get the results in following Fig. 5-2.

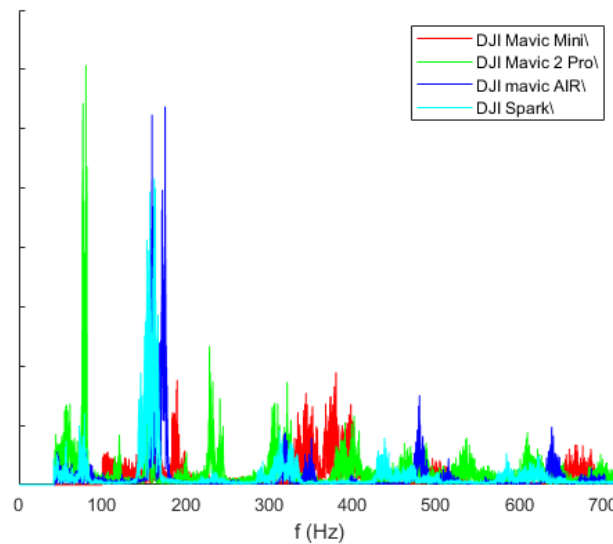


Figure 5-2: Signature Comparison.

According to technical data available from the drones (battery of 11,55V to 15,4V according to the model) and assuming a typical value of 900 rpm per volt for the brushless motors (which are the drone’s main noise source), we can assume that fundamental audio content will be within a maximum frequency of about $F_{max} = 900 * 15 / 60 = 225$ Hz.

Resizing the signature in this band with a 6th order Butterworth filter high pass with a cut off frequency of 10Hz, it is possible to focus the audio recognition algorithm to find out signatures drawn in Fig. 5-3.

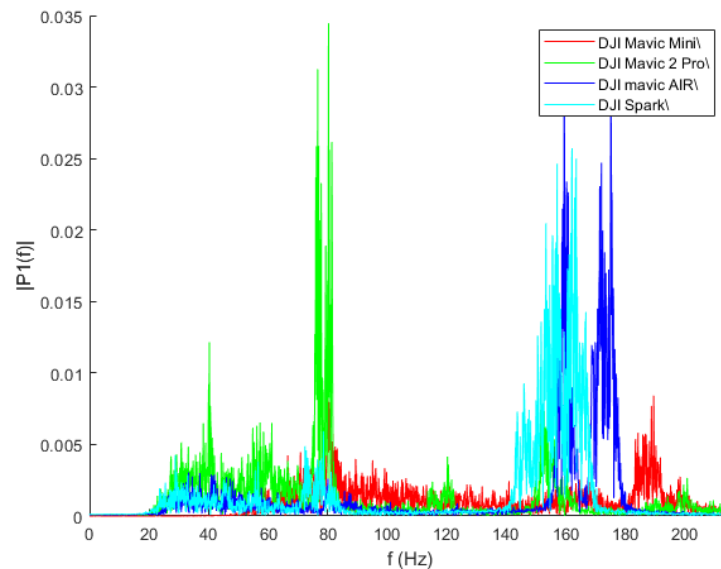


Figure 5-3: Fundamental Signature.

Once modelled this signature, it is possible to use it to recognize the drones. The proposal is to adopt the

same recognition algorithm discussed in [2]. Experimentation on a physical device, such as [9], to on-field assess the algorithm performance is an interesting future activity.

The experiments refer to open field recordings, but we assume an intrinsic resilience of the methodology even for low SNR values. Note that this level of detection is activated only if a 'threat' has already been recognized, i.e. a considerable variation from the background noise has already been recorded.

5.2 RNN

The drone recognition in the phase 2 of the algorithm can be also performed and enhanced by leveraging state-of-the-art techniques, e.g., coming from the realm of artificial intelligence and Machine Learning (ML). In particular, the use of such techniques is gaining momentum in the research community for the task of sound event detection inside audio segments. Even if the application of ML-inspired techniques goes beyond the scope of this manuscript, we now describe a few possible approaches that can be incorporated in the detection/recognition system.

The most promising ML-inspired approaches for target recognition can be the so-called RNN and convolutional RNN (CRNN), see, e.g., [9] and references therein. These approaches are of particular interest due to the fact that they can easily exhibit temporal dynamic behaviour, thus making them attractive for the audio recognition of UAS. In fact, in our reference scenario, the temporal behaviour of the audio segment is a key aspect to be taken into account.

The key idea beyond these approaches is to extract proper features that are representative of the audio to be detected. The meaningful features can be obtained, as already discussed in the previous section, in the frequency domain with fine-grained computation of the signal spectrum and its key characteristics (e.g., relevant frequency peaks). The spectral features are used two-fold.

- First, in a preliminary offline stage they are extracted from a sufficiently large set and employed to properly train the neural network with all possible labels.
- Then, in the online stage the spectral features are extracted in real-time from the acquired audio signals and given at the input of the RNN to obtain an estimate of the presence (or not) of the UAS sound.

Since the frequency domain-based computation is common with the proposed signature-based detection algorithm, we expect to improve the recognition capabilities of the system. This is due to the fact that the system has a sort of a priori information (the training phase) that is properly exploited by the RNN. The performance of RNN-based schemes can be easily (and in a standardized manner) assessed in terms of F-score and error rate, which are well-known indicators of the effectiveness of the sound recognition. However, one should note the use of neural network poses a higher computational complexity than simple spectral comparison, thus highlighting an inherent trade-off between the two approaches.

6.0 LEVEL 3 ALGORITHM – TRACKING

The third level of detection will activate the cooperative tracking of the drone. In this phase, the alarmed sensor node will activate neighbour ones in order to start the continuous sampling and the cooperative tracking. In our testing scenario, cooperative nodes will be linked to a command and control station capable to manage all incoming waveforms. Future investigation will compare a beamforming technique from multi audio sensors with a generalized cross correlation algorithm [5], as depicted in Fig 3.

Note that the such operations allows for a passive tracking of the moving target. The capability of fine tracking depends on the used array of microphones. In particular, various arrays of microphones should be

put in the monitored area. First, a single array is responsible to locally track the target. Moreover, by exchanging information among various arrays displaced at different relevant points, one can track the UAV in large areas. This communication exchange will be the focus of future research.

6.1 Tracking – G. Cross-Correlation

This phase will assume that two or more audio sensor array has been capable to recognize the drone and each one is capable to evaluate the Direction Of Arrival (DOA) of the sound beam.

6.1.1 Localization – 2 beams

Referring to Fig. 6-1 , assuming

$$L = y \tan \theta_1 + y \tan \theta_2$$

one can estimate the target coordinates by solving the following system of equations

$$(x, y) = \begin{cases} y = \frac{L}{(\tan \theta_1 + \tan \theta_2)} \\ x = y * \tan \theta_1 \end{cases} \quad (9.)$$

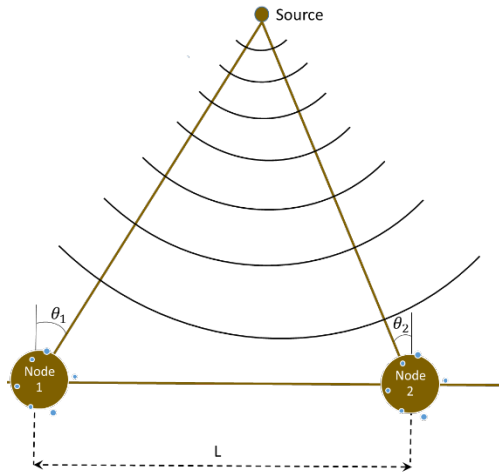


Figure 6-1: Generalized cross-correlation method.

In this phase, the main performance metrics will be the localization error and the algorithm complexity. Experimentation of this algorithm will be analysed in future work. Paragraph below introduce a discussion for a possible implementation.

6.1.2 Localization Multibeam

Assuming the notation adopted in following figure, we aim to calculate the position vector $\mathbf{R} = (x_e, y_e)^T$.

It is possible to express all Direction Of Arrival (DOA_i) in the following system

$$\begin{cases} DOA_1 = f_1(\mathbf{R}) + n_1 \\ \vdots \\ DOA_N = f_N(\mathbf{R}) + n_N \end{cases} \quad (10.)$$

Where

$$f_i(\mathbf{R}) = DOA_{i_{TRUE}} = \tan^{-1} \left(\frac{x_E - x_i}{y_E - y_i} \right) \quad (11.)$$

And n_i is the measurement of the error.

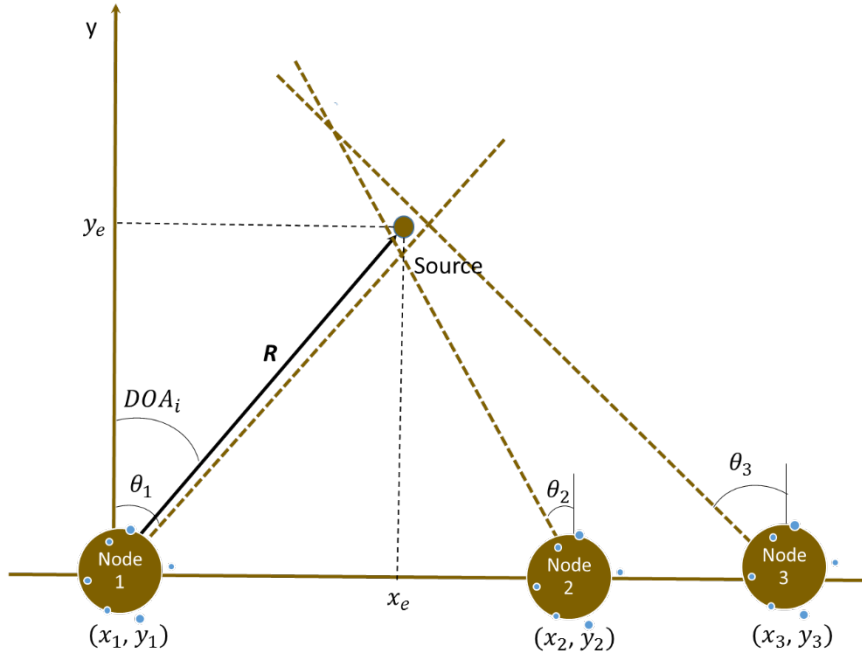


Figure 6-2: Reference system for cross-correlation multibeam.

A first evaluation $\mathbf{R}_0 = [x_{0E}, y_{0E}]^T$ is possible dividing:

$$d_{0i} = \sqrt{(x_{0E} - x_i)^2 + (y_{0E} - y_i)^2} \quad (12.)$$

with

$$\begin{cases} d_{0i} \cos(DOA_{0i}) = y_{0E} - y_i \\ d_{0i} \sin(DOA_{0i}) = x_{0E} - x_i \end{cases} \quad (13.)$$

Assuming N different beam measures it is possible to derive the following matrix system:

$$\begin{bmatrix} \cos(DOA_{01}) & -\sin(DOA_{01}) \\ \vdots & \vdots \\ \cos(DOA_{0N}) & -\sin(DOA_{0N}) \end{bmatrix} \cdot \begin{bmatrix} x_{0E} \\ y_{0E} \end{bmatrix} = \begin{bmatrix} \cos(DOA_{01}) x_1 - \sin(DOA_{01}) y_1 \\ \vdots \\ \cos(DOA_{0N}) x_N - \sin(DOA_{0N}) y_N \end{bmatrix} =$$

$$\mathbf{A} \cdot \mathbf{R}_0 = \mathbf{c} \quad (14.)$$

Since the system is overdetermined it can be inverted in the following:

$$\mathbf{R}_0 = (\mathbf{A}^T \cdot \mathbf{A})^{-1} \cdot \mathbf{A}^T \cdot \mathbf{c} \quad (15.)$$

So, the initial \mathbf{R}_0 evaluation can be refined with the following iterative process:

Calculating the vector having as extremes the coordinates of the audio sensor and the \mathbf{R}_0

Assuming

$$\begin{cases} r_{x_i} = x_{0E} - x_i \\ r_{y_i} = y_{0E} - y_i \end{cases} \quad (16.)$$

$\forall i = 1, \dots, N.$

Evaluation of following Matrix:

$$\mathbf{G} = \begin{bmatrix} \frac{r_{y_1}}{r_{x_1}^2 + r_{y_1}^2} & -\frac{r_{x_1}}{r_{x_1}^2 + r_{y_1}^2} \\ \vdots & \vdots \\ \frac{r_{y_N}}{r_{x_N}^2 + r_{y_N}^2} & -\frac{r_{x_N}}{r_{x_N}^2 + r_{y_N}^2} \end{bmatrix} \quad (17.)$$

Evaluation of the difference between DOA measured and DOA expected:

$$\boldsymbol{\phi}_R = \begin{bmatrix} \phi_{R_1} \\ \vdots \\ \phi_{R_N} \end{bmatrix} = \begin{bmatrix} \phi_{0_1} - \hat{\phi}_{0_1} \\ \vdots \\ \phi_{0_N} - \hat{\phi}_{0_N} \end{bmatrix} \quad (18.)$$

Where

$$\begin{cases} \phi_{0_i} = DOA_{0i} , \forall i = 1, \dots, N \\ \hat{\phi}_{0_i} = f_i(\mathbf{R}_0) = \tan^{-1} \left(\frac{r_{x_i}}{r_{y_i}} \right) , \forall i = 1, \dots, N \end{cases}$$

Estimates refinement:

$$\mathbf{R} = \mathbf{R}_0 + [(\mathbf{G}^T \cdot \mathbf{N}^{-1} \cdot \mathbf{G})^{-1} \cdot \mathbf{G}^T \cdot \mathbf{N}^{-1}] \cdot \boldsymbol{\phi}_R \quad (19.)$$

Assuming $\mathbf{R}_0 = \mathbf{R}$ it is possible to restart from point 1 with an iterative cycle of the procedure.

Starting from point 5 it is also possible to evaluate the uncertainty ellipsoid, linked to the probability error P_e , which axis can be represented as:

$$\begin{cases} a == 2\sqrt{k\lambda_{max}} \\ b == 2\sqrt{k\lambda_{min}} \end{cases} \quad (20.)$$

Where λ_{max} and λ_{min} are the maximum and minimum eigenvalue of \mathbf{P} and k is:

$$k = -2 \ln(1 - P_e) \quad (21.)$$

In order to get the ellipsoid in the x,y system we have to translate values on \mathbf{R} and rotate with a spectral decomposition of \mathbf{P} [7][8].

6.1.3 Kalman Filtering

Finally, having the algorithm to deal with measured DOA, i.e. affected by measurement error, also positioning obtained will be affected by error. In order to smooth such an effect, it is possible to filter them with a Kalman filtering. In fact, every measure has a different error variability and the updating of the positioning obtained by previous cycle has not a periodic timing. The Kalman filtering is in fact the most suitable to adapt its parameters to the variability of read measure

6.2 Tracking – Beamforming

In order to track the UAS movement, it is important to provide the audio sensor network with beamforming capabilities to identify, for each microphone, the direction from which the audio comes from. Even if this is a future feature for our system, we now briefly discuss state-of-the art solutions that can cope with our problem of interest.

In the past, various binaural beamforming techniques have been proposed in the literature which exploit the acoustic transfer function between each microphone and loudspeaker. All of them can exhibit interesting performance in terms of audio directivity. However, the key issue of such approaches is that they usually require the a priori knowledge of the auditory of listener, which is not suitable in the considered scenario.

To this end, most of the researcher tries to use other information coming from other external sensors. Our future work will be devoted in investigating proper beamforming algorithms that can allow the use of a reduced number of audio sensor in the tracking network.

7.0 CONCLUSIONS

In this work, we have proposed a methodology based on a three-layer approach to detect, recognize and track small drones with audio sensors networks. The idea is to use small, cheap and cooperating nodes to defend an area from unattended UAS. The defence method is, in this case, symmetric to the attack one. In fact, according to the drone's technology trend they are becoming even more smaller, cheaper and capable to cooperate. Dually, the defence is then implemented in small, cheap and collaborating nodes. The research is still ongoing and this paper has been focused on the methodology, the raw sensing and the recognition phases. We also discuss most promising approach for the audio tracking. In particular, starting from the use case and a measurement campaign performed over several drones and publicly available, we adjust the level one algorithm to detect anomalies on the raw sound level. This level uses the audio-sensor node in a very

low power consumption schema, waking up the node periodically for a sampling campaign of the environment. The raw sensing algorithm search for anomalies and trigger a further stage of 'detection' capable to follow the anomaly and discriminate if it is a sporadic event or a real audio anomaly. This first level of detection triggers a further level of recognition. We experiment the audio sampled from various drones in order to model the relevant signature capable to recognize them. In particular, according to the use case, we process audio recording of flying drones in order to evaluate the signature. In our analysis, it was possible to recognize a relevant low-frequency bandwidth in the audio spectra of the drones. According to needed authorization from owner company, it could be relevant to test such a signature with the algorithm developed for the UGS sensors [11]. This step of testing could fill the need to characterize the hit/miss rate of the algorithm.

The outcome of this work may help the implementation in term of complexity and efficacy on a target platform, such as [9] that may be chosen for implementation on a commercial system.

8.0 REFERENCES

- [1] Martalò, M., Ferrari, G., Malavenda, C. S. (2013). Low-Complexity Hybrid Time-Frequency Audio Signal Pattern Detection. *IEEE Sensors Journal*, 13(2), 501-509
- [2] Marco Martalò, Gianluigi Ferrari, Claudio Santo Malavenda, “Wireless Sensor Networks and Audio Signal Processing for Homeland Security”; chapter contribution in “Effective Surveillance for Homeland Security: Balancing Technology and Social Issues”, CRC Press, Taylor and Francis Group, accepted, 2012, published 06/2013. ISBN: 9781439883242
- [3] Sungho Jeon¹, Jong-Woo Shin, Young-Jun Lee, Woong-Hee Kim, YoungHyouon Kwon, and Hae-Yong Yang, “Empirical Study of Drone Sound Detection in Real-Life Environment with Deep Neural Networks”, The Affiliated Institute of ETRI, (2017)
- [4] Raya Islam, Sam Kelly, “Small UAV Noise Analysis Design of Experiment”, Duke University, (2017)
- [5] B.V.D. Broeck, A. Bertrand, P. Karsmakers, B. Vanrumste, H. Van hamme, M. Moonen, “Time-Domain Generalized Cross Correlation Phase Transform Sound Source Localization For Small Microphone Arrays”, 2012
- [6] T.Makino, H. Liaoy, Y. Assaelz, B. Shillingfordz, B. Garciay, O. Bragay, O. Siohany, “Recurrent Neural Network Transducer For Audio-Visual Speech Recognition”, 2019
- [7] T.M. Apostol, “Calcolo Vol 2 – Geometria”, Bollati Boringhieri, 2003
- [8] D. J. Torrieri, “Statistical Theory of Passive Location Systems”, *IEEE Transactions on Aerospace and Electronic Systems*, Vol.20, No. 2, March 1984.
- [9] “DJI Mavic Mini Propeller Noise Comparison”, you tube video, last accessed 25/06/2020
- [10] Leonardo S.p.A., HALO Products, <https://www.leonardocompany.com/it/products/halo>
- [11] Leonardo, The UGS system
https://www.leonardocompany.com/documents/20142/3163330/body_Border_Control_LQ_mm08482_.pdf/5fa30e5c-b031-25da-1876-44fd7b8f9316?version=1.0&t=1542838371075
- [12] Kyoungjin Noh and Joon-Hyuk Chang, “Joint Optimization of Deep Neural Network-Based Dereverberation and Beamforming for Sound Event Detection in Multi-Channel Environments,” *MDPI Sensors*, vol. 20, no. 7, 2020.

ACKNOWLEDGEMENTS

The contribution realization of this paper has been possible on the basis of a former industrial research conducted by most of the authors with the support of Selex E.S., now Leonardo S.p.A, Rome, Italy.

AUTHOR/SPEAKER BIOGRAPHIES

Claudio Santo Malavenda: PhD, MBA. He covered several roles in STMicroelectronics, Selex ES (actual Leonardo) and Elettronica dealing with Project and Contract management. His works activity deals first series products management and marketing, Wireless Sensor Network, C4I systems for fire control in

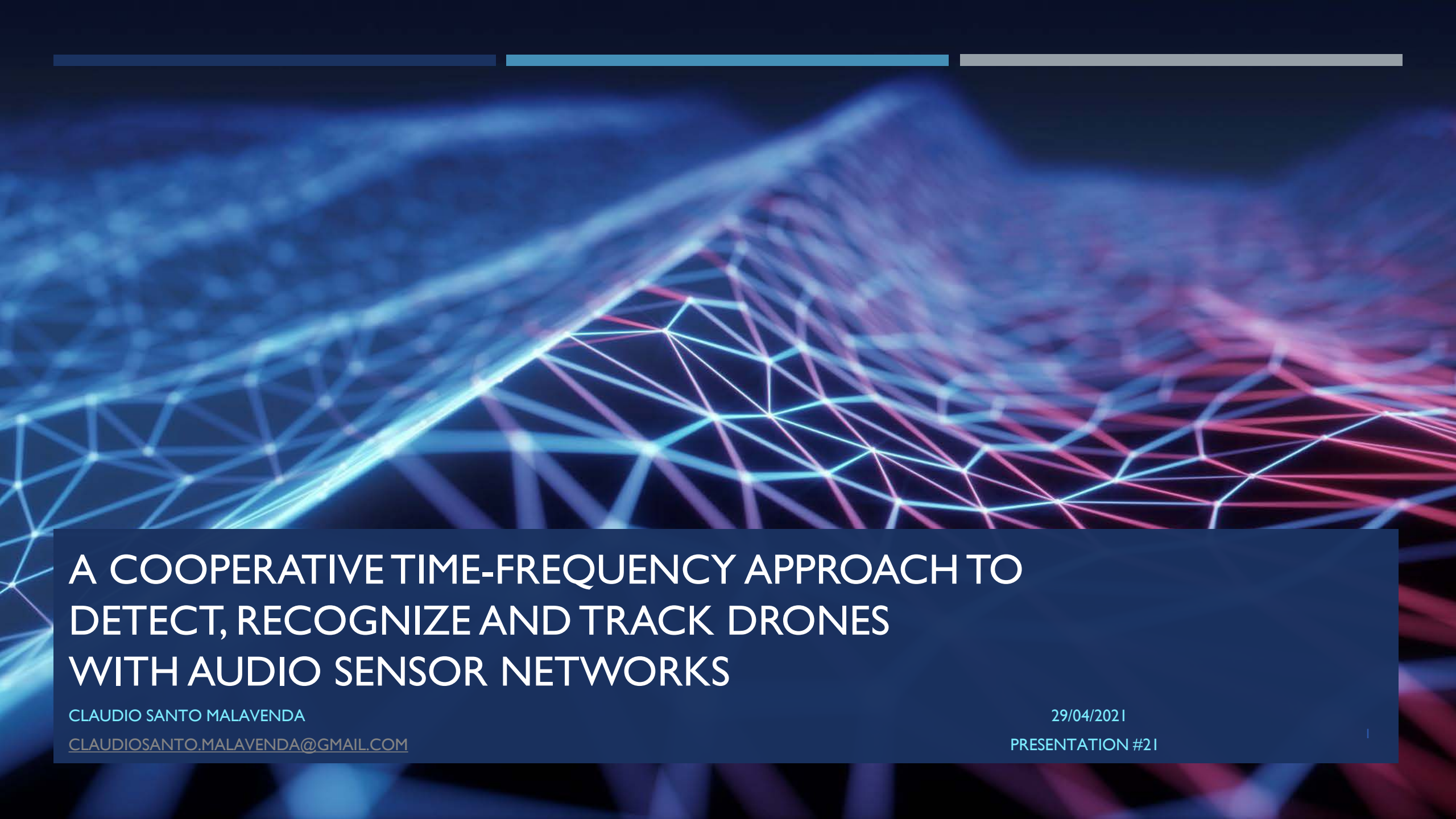


network centric environment and Cyber Electromagnetic Activities.

Carlo Malavenda: Carlo, after obtaining the scientific high school diploma, is a student involved in the first level Master in Sound and Entertainment Engineering from Tor Vergata, University of Rome.

Marco Martalò: received in 2009 the Ph.D. in Information Technologies from the University of Parma, Italy, where he actually is a Post-Doctoral Researcher. Between 2012 and 2017, he also was an Assistant Professor at E-Campus University, Italy. His activities are focused on communication and signal processing for wireless systems.





A COOPERATIVE TIME-FREQUENCY APPROACH TO DETECT, RECOGNIZE AND TRACK DRONES WITH AUDIO SENSOR NETWORKS

CLAUDIO SANTO MALAVENDA

CLAUDIOSANTO.MALAVENDA@GMAIL.COM

29/04/2021

PRESENTATION #21

AGENDA

- SCENARIO
- THE APPROACH
- LEVEL I - DETECTION
- LEVEL II - RECOGNITION
- LEVEL III – TRACKING
- CONCLUSIONS

A FEW WORDS ON THE SPEAKER

Claudio Santo Malavenda: PhD, MBA. He covered several roles in STMicroelectronics, Selex ES (actual Leonardo) and Elettronica dealing with Project and Contract management. His works activity deals first series products management and marketing, Wireless Sensor Network, C4I systems for fire control in network centric environment and Cyber Electromagnetic Activities.

He actually supports companies as a freelance with innovative projects for process optimization and innovative product release.



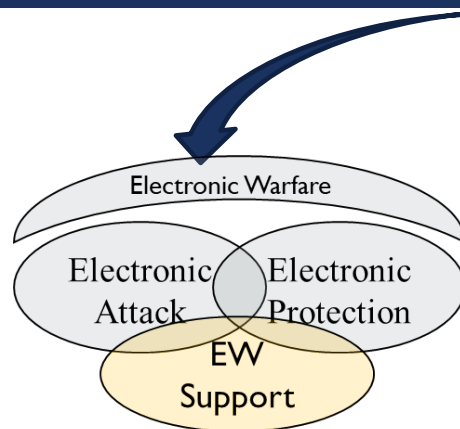
SCENARIO – TECHNOLOGY POSITIONING

**Mechanical Wave - audio
(not electromagnetic wave)**

During this RESEARCH

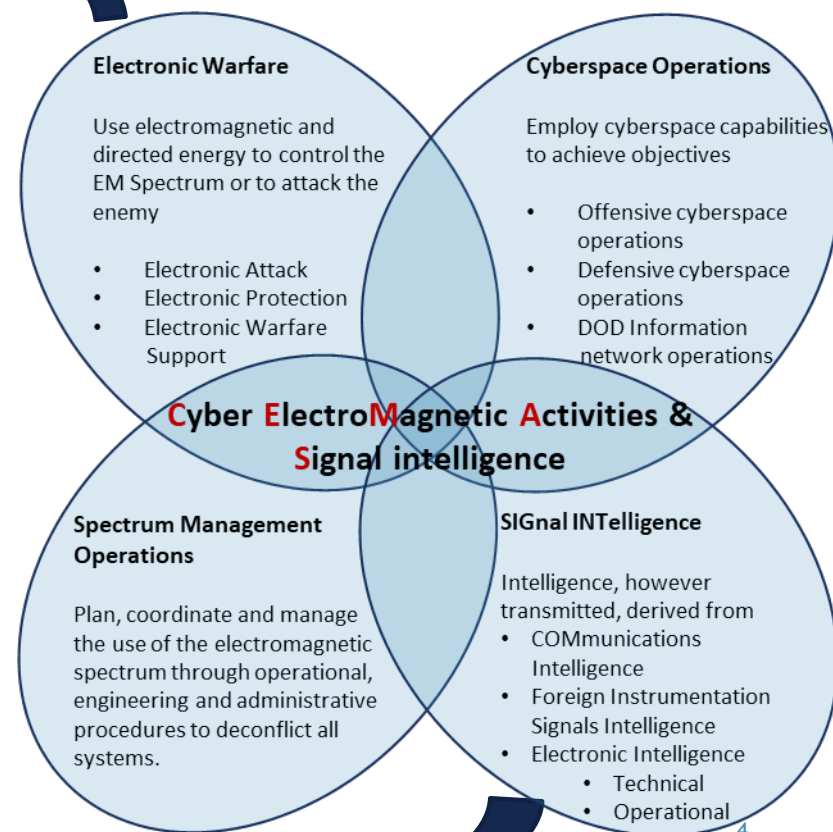
OPELINT is concerned with operationally relevant information such as the **location, movement, employment, tactics, and activity** of foreign noncommunications emitters and their associated weapon systems

TECHELINT is concerned with the technical aspects of foreign noncommunications emitters such as **signal characteristics, modes, functions, associations, capabilities, limitations, vulnerabilities, and technology levels**

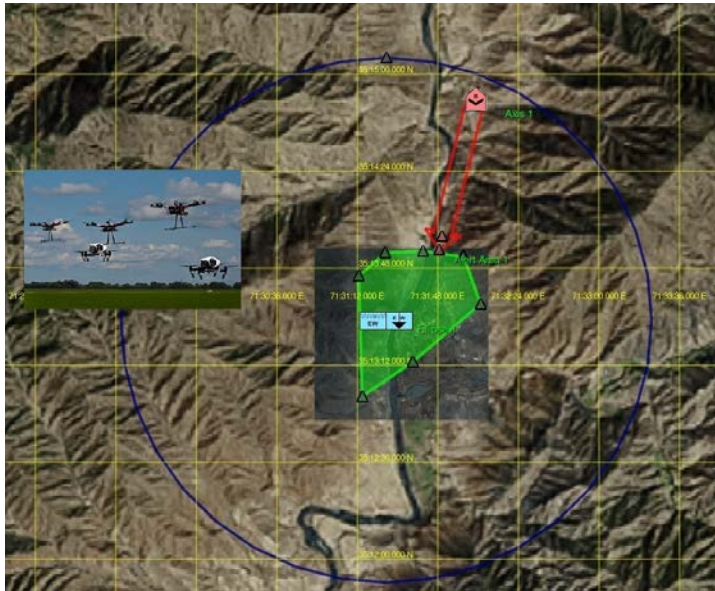


When USED

Actions tasked by, or under direct control of, an operational commander to search for, **intercept, identify, and locate or localize sources** of intentional and unintentional radiated electromagnetic energy for the purpose of immediate threat recognition, targeting, planning, and conduct of future operations



SCENARIO



A large number of cheap audio sensors, each one capable to recognize the drone's audio signature and, when a match is found, locate the malicious source, by collaborating with other nodes.

The focus of the research is to detect, recognize and track with audio array a single drone or a small swarm of drones, i.e. assimilable to a single large unit, in an open field scenario.

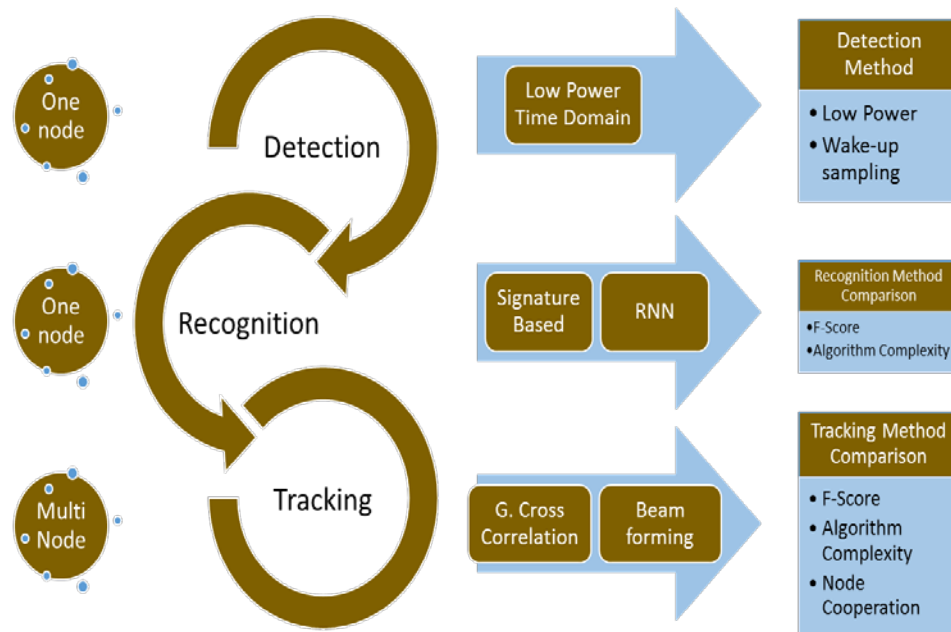
Symmetric warfare defence:

Drones: smaller, cheaper and capable to cooperate.

Dually, the target defence is then implemented in small, cheap and collaborating nodes

Class	Category	Normal Employment	Normal Operating Altitude	Normal Mission Radius	Civil Category (UK CAA)	Example Platform
Class I <150 kg	MICRO < 2 kg	Tactical Platoon, Section, Individual (single operator)	Up to 200ft AGL	5 km (Line of Sight (LOS))	Weight Classification Group (WCG) 1 Small Unmanned Aircraft (<20 kg)	Black Widow
	MINI 2-20 ¹³ kg	Tactical Sub-Unit (manual launch)	Up to 3000ft AGL	25 km (LOS)		Scan Eagle, Skylark, Raven, DH3
	SMALL > 20 kg	Tactical Unit (employs launch system)	Up to 5000ft AGL	50 km (LOS)	WCG 2 Light Unmanned Aircraft (20><150 kg)	Luna, Hermes 90
Class II 150-600kg	TACTICAL	Tactical Formation	Up to 10,000ft AGL	200 km (LOS)	WCG 3 UAV (>150 kg)	Sperwer, Iview 250, Aerostar, Watchkeeper
Class III >600 kg	Medium Altitude, Long Endurance (MALE) ¹⁴	Operational/Theatre	Up to 45,000ft AGL	Unlimited (BLOS)		Reaper, Heron, Hermes 900
	High Altitude, Long Endurance (HALE)	Strategic/National	Up to 65,000ft AGL	Unlimited (BLOS)		Global Hawk
	Strike/Combat	Strategic/National	Up to 65,000ft AGL	Unlimited (BLOS)		

THE APPROACH



DETECTION

this phase detects an anomaly in the audio spectrum

RECOGNITION

this phase assumes that we have an anomaly approaching and discriminates if it is assimilable to a drone

TRACKING

this phase assumes there is at least a drone in the detection area and tries to track its movement

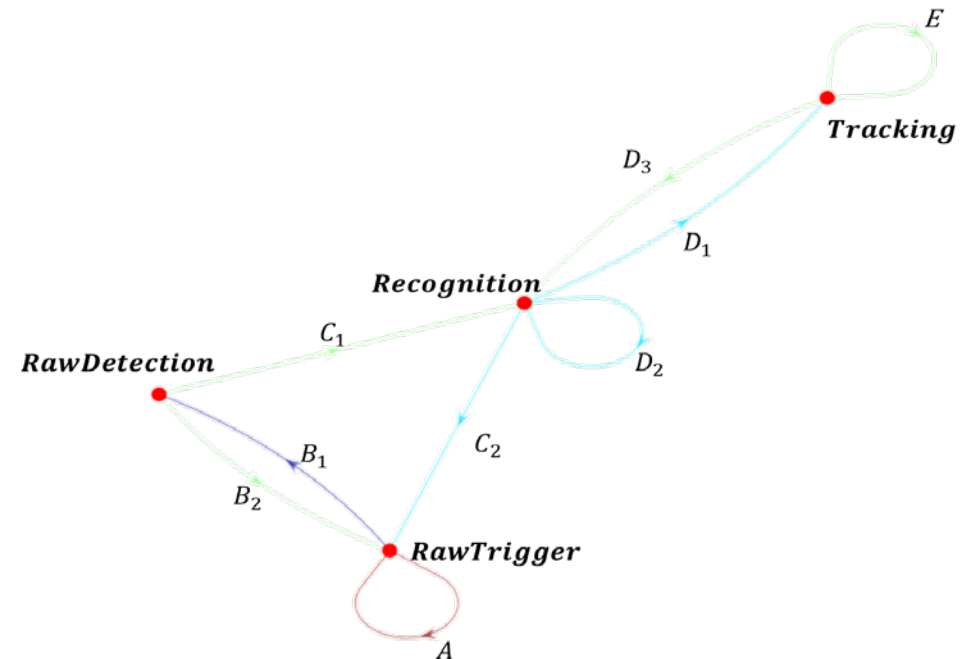
The research is still ongoing: focus on the methodology, raw sensing and recognition phases

THE APPROACH - MODELLING

A first level of modelling this approach is possible through the Markov chains

Associating each model phase to a state and evaluating the probability to change the state.

- A* the probability to detect no – alarm in Raw Trigger
- B₁* the probability to detect an anomaly in Raw Trigger
- B₂* the probability to false detection from Raw Trigger
- C₁* probability to detect an anomaly in Raw Detetcion
- C₂* probability to NOT recognize a drone in 'Recognition'
- D₁* the probability to recognize a drone in 'Recognition'
- D₂* the probability to keep recognizing a drone with no other collaborating nodes available
- D₃* probability to lost tracking
- E* the probability to keep the tracking

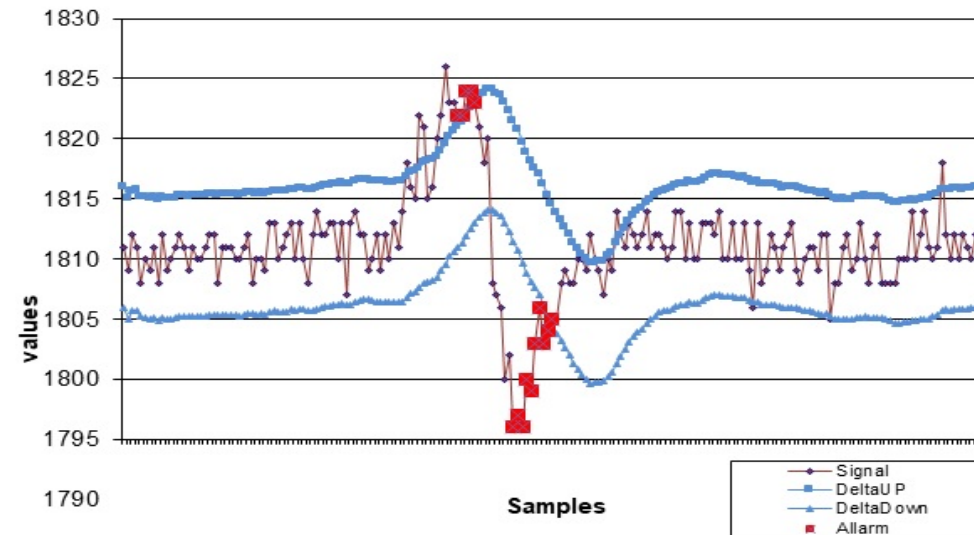


LEVEL I – DETECTION (1/2)

- Raw sensing with trigger counter

```
for i = (K+1):signal-length
    for j= (i-K):i
        delta=delta+y(j);
    end
    delta=delta/K;
    threshup(i)=y(i)+delta;
    threshdown(i)=y(i)-delta;
    if (y(i)>threshup(i)) || (y(i)<threshdown(i))
        countA=countA+1;
        if countA>A
            allarm
        end
    else
        countA=0;
    end
end
```

- Low computation / Low energy (suitable for embedded device)
- Non dynamic Parameter $[k, \Delta, a]$

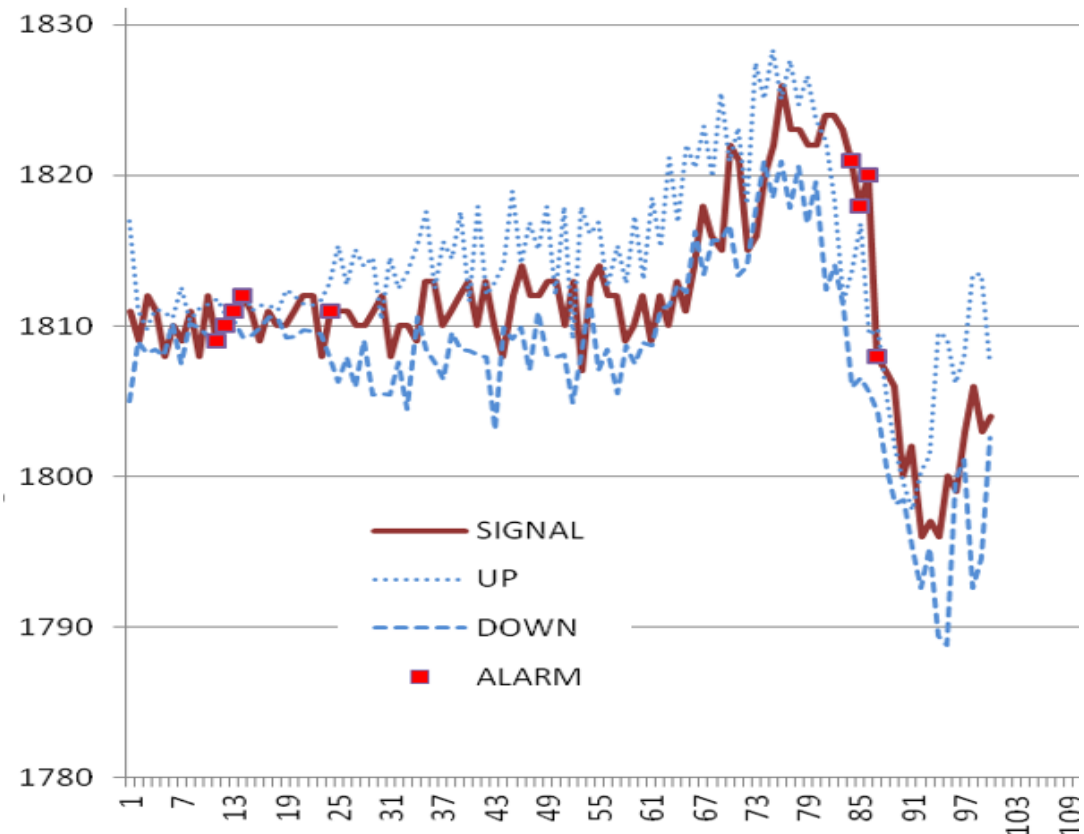


LEVEL I – DETECTION (2/2)

- Particle Swarm Optimization (PSO)
 - continuous updating of parameters [\mathbf{k} , Δ , \mathbf{a}]
- Higher computational need
 - At each cycle it optimizes a fitness function

Calculating the optimum values for the algorithm with:

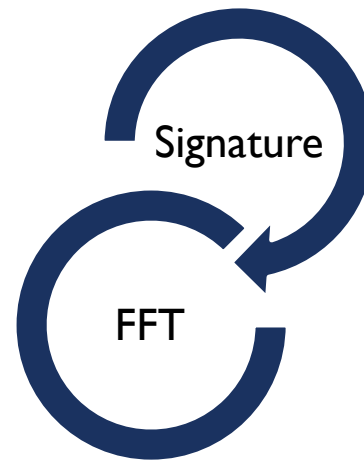
$$\vec{p}_i = \begin{cases} p_i^k = \text{MIN}[F(k_{(i-k_i)}); \dots; F(k_i)] \\ p_i^\Delta = \text{MIN}[F(\Delta_{(i-k_i)}); \dots; F(\Delta_i)] \end{cases}$$



LEVEL 2 – RECOGNITION (1/2)

■ SIGNATURE-BASED

- Produce signature with algebraic calculation from FFT
- Tested with audio track from
 - Mavic mini
 - Mavic 2 Pro
 - Mavic AIR
 - DJI Spark



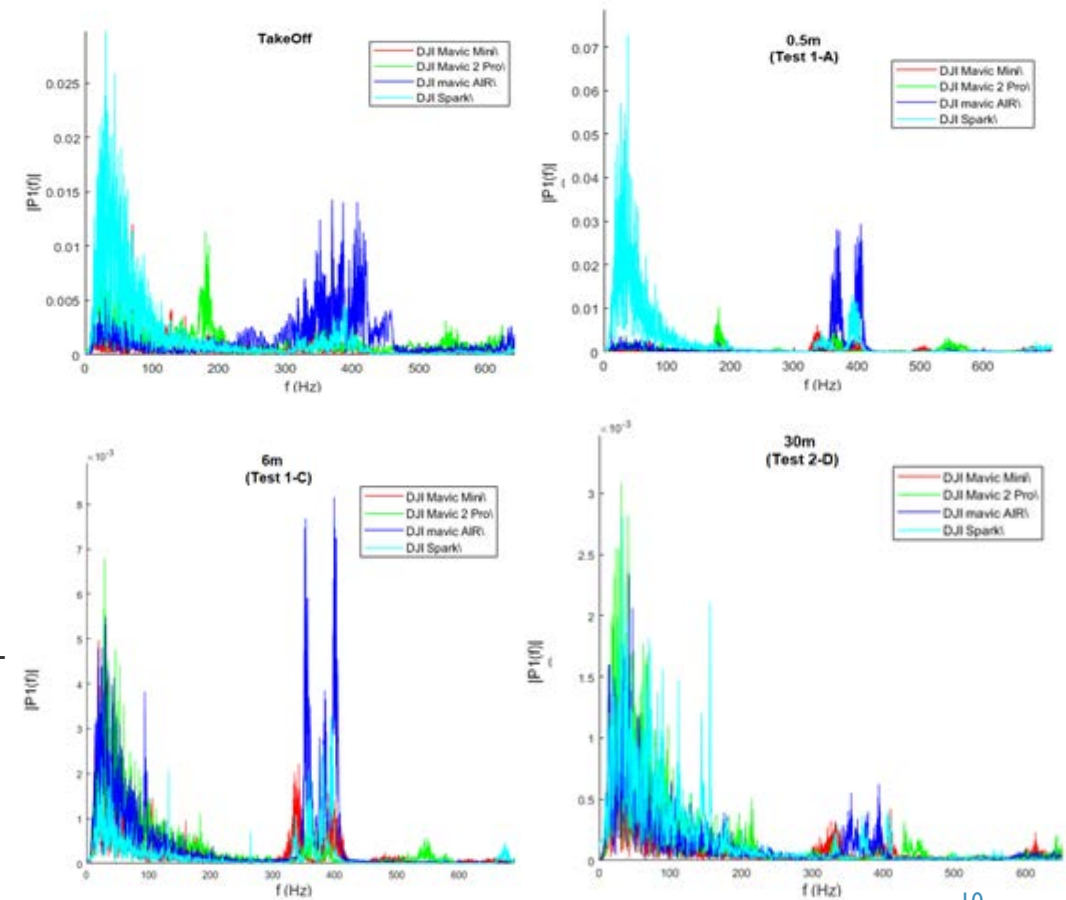
MATCH PRE-RECORDED
SIGNATURE WITH CURRENT FFT

■ RNN

produce signature with a neural network

Offline stage: use record to produce signature

Online stage: use signature to detect incoming audio



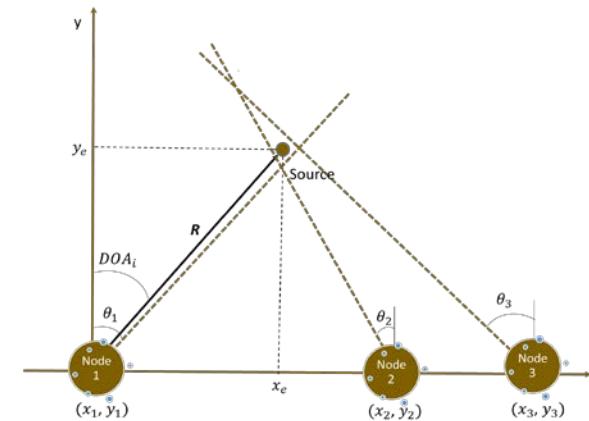
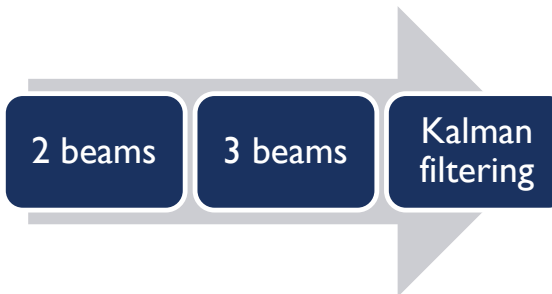
LEVEL 3 – LOCALIZATION

- MULTI SENSOR: NODES COOPERATE

- G-Cross reference

- 2 BEMS / 2 Nodes cooperating
- 3 BEAMS / 3 Nodes Cooperates

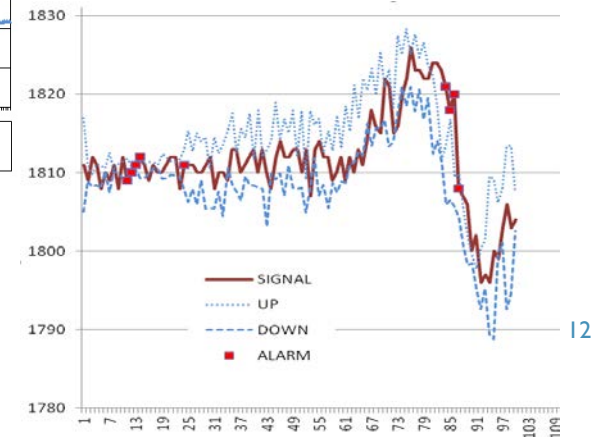
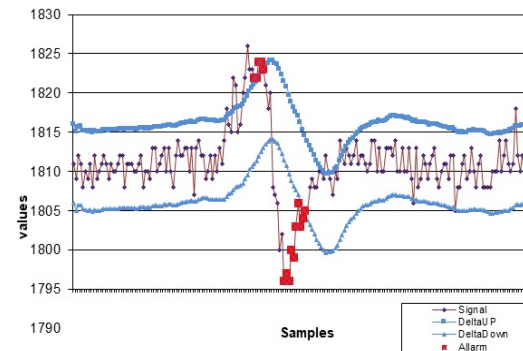
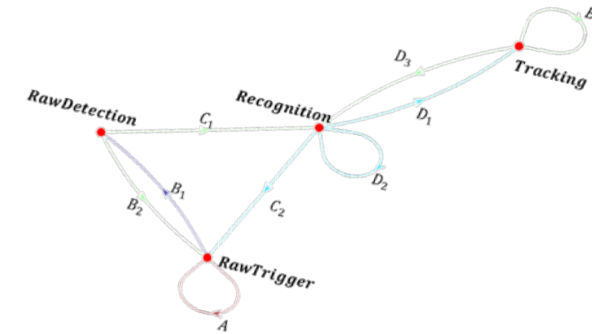
- Beamforming



Reference system for cross-correlation multibeam

SIMULATION RESULTS (1/2)

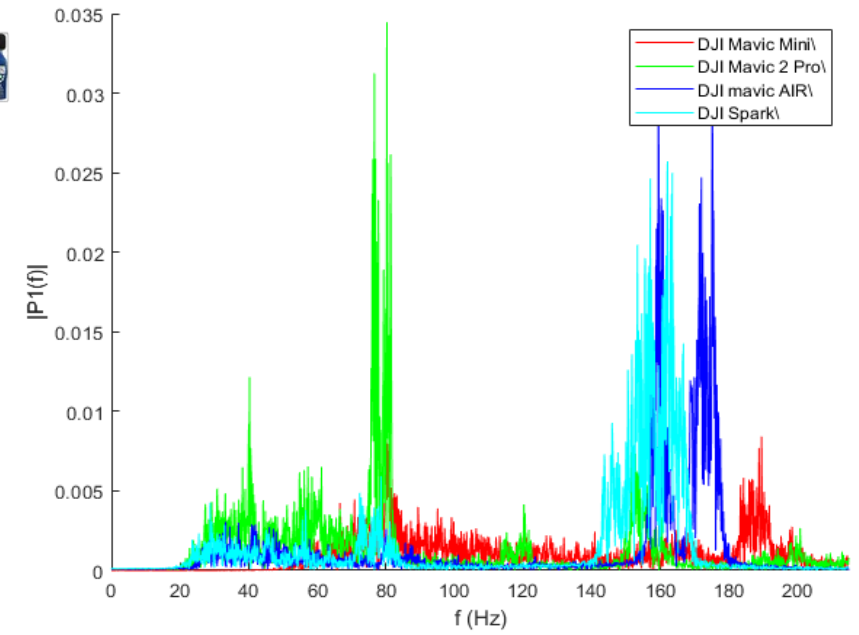
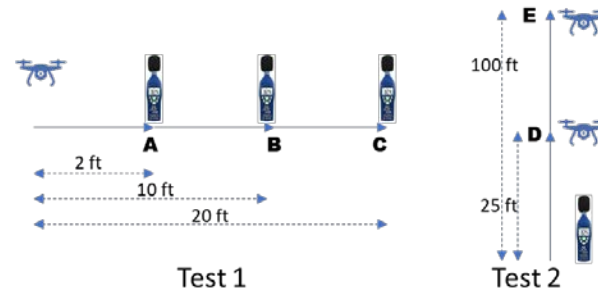
- Markov Modelling phase
 - According to empirical values it is possible to assume that all nodes will stay most of the time in the Raw Trigger phase, i.e. the most energy saving modes for the detecting node
- LEVEL I detection
 - According to MATLAB simulation of this phase
 - Trigger counter method
 - Most suitable to detect a particular audio profile (fixed parameters)
 - Generates more false alarm
 - Has a low computational need
 - Trigger follower method
 - Has a higher adaptability to detect audio anomalies
 - Uses a larger amount of computational energy
 - Generates a lower amount of false alarm
 - Random contributes in calculation of velocities produce non-deterministic alarms



SIMULATION RESULTS (2/2)

- LEVEL II recognition

- Tested five recording points for each drone
- It is possible to discriminate drones with their energy content in the [20Hz, 200Hz] window.
- This value matches typical rpm values and input voltage of brushless motors used in commercial drones.
- It could also be possible to build a ‘library’ of detectable drones matching those values with expected audio fundamentals



Fundamental Signature

CONCLUSION

- Symmetric warfare defence for Electronic Warfare (mechanical wave case)
 - Defense method dual to the attack one: small, cheap and capable to cooperate
- A multi level algorithm can keep low the amount of energy used in battery powered nodes (for long time surveillance) and use most sophisticated techniques only when necessary
- Suitable for several HW targets from small embedded sensors to more advanced ones

- Low frequencies can lead the characterization of different drones
- Literature on commercial drones can be used to build up a raw library of expected fundamental signature according to a specific type of drone

- Smart sensor networks can be used as a enabling technology for audio localization

QUESTIONS

