Audio Engineering Society

Convention e-Brief 308

Presented at the 142nd Convention
2017 May 20–23, Berlin, Germany

# Feature selection for real-time acoustic drone detection using genetic algorithms

Joaquín García-Gómez[1], Marta Bautista-Durán[1], Roberto Gil-Pita[1], and Manuel Rosa-Zurera[1]

[1]*Signal Theory and Communications Department, University of Alcalá, Alcalá de Henares, Spain*

Correspondence should be addressed to Joaquín García-Gómez (joaquin.garciagomez@edu.uah.es)

## ABSTRACT

Drones are taking off in a big way, but people sometimes use them in order to invade the privacy of others or to bypass the security systems, making their detection an actual issue. The objective of the proposed system is to design real-time acoustic drone detectors, able to distinguish them from objects that can be acoustically similar. A set of features related to the propeller sounds have been extracted, and genetic algorithms have been used to select the best subset. The classification error achieved with 30 features is below 13%, making feasible the real-time implementation of the proposed system.

## 1 Introduction

Drones are booming nowadays because of the advantages they provide to the society. However, people sometimes go beyond the ethical boundaries using them in order to invade the privacy of others or to bypass the security systems [1]. For this reason it is important to develop a system capable of detecting the presence of drones in particular environments where they can be used for malicious purposes, such as households, public buildings, or restricted-access areas.

This problem can be approached using different data sources, like radar information, radio frequency, video, or even audio signals. All of them have some drawbacks [2]. Radar system, which is the main way to detect large aircrafts, fails when the size of the object is as small as a quadcopter. Video has the disadvantage of being highly dependent on the weather and the moment of the day. For instance, it is difficult to detect drones in foggy periods or even at night. In this type of situations, other systems have to be applied to overcome these constraints.

Some works in the literature have treated the problem of detecting drones. In [2], a detector is developed based on video and audio information. That system is proposed for some specific environments where background image need to be static. Taking into account just audio information, an acoustic array is used for drone detection and tracking in [3]. In [4] a single microphone is used to determinate the position of a drone in a short range, but the work has been done only using a model of drone. Most of the sounds used in these studies were recorded in calm places where noise is not present, thus without taking into account usual background noises presented in cities.

In this work we have studied drones detection as a problem of sound event detection because audio recording and processing implies low cost, good coverage and low privacy intrusion compared to other sources. In order to develop a more general system, several types of drones have been studied in real environments. That way, background noise has been taken into account. To achieve the best solution for the detection problem, different classifiers and features have been tested.

## 2 Methods

In this section the steps we have followed to develop the system are detailed, including the feature extraction and the feature selection process.

Before being evaluated, audio needs to be processed in order to accommodate the different files. The first step we have implemented in the algorithm is to resample all audios to the same sampling frequency (8 kHz). This value is chosen because the energy of the signals under study is concentrated at low frequencies. Then, the files are splitted into frames of one second. Each frame, in turn, is divided into subframes of 64 ms with an overlap of 50%.

The proposed algorithm has to reach a binary decision in each frame, indicating whether there is a drone close to the system or not. In order to do that, we have taken into account different observation periods ($T_{obs}$). This parameter represents how many seconds from the past are evaluated to decide in the current moment.

### 2.1 Feature Extraction

The sound produced by a propeller is basically a pseudo-harmonic noise, whose frequency is related to the number of propeller blades and to the rotatory frequency. The total sound produced by a drone is mainly caused by the combination of the sounds produced by each propeller, in which the rotatory frequencies are not always exactly the same. Therefore, fundamental frequency related features are interesting to analyze and distinguish this kind of noise from other noises present in the acoustic environment.

There are several audio measurements related to the harmonic component that could exhibit a good discrimination capability for the problem at hand, which have been included in the present study: the Mel-Frequency Cepstral Coefficients (MFCCs), the Delta Mel-Frequency Cepstral Coefficients (ΔMFCCs), the fundamental frequency (Pitch), the zero crossing rate (ZCR), the harmonic noise rate (HNR), the ratio of non-harmonic time frames (RUF), the short time energy (STE), the energy entropy (EE), the spectral rolloff (SR), the spectral centroid (SC) and the spectral flux (SF). This features have been calculated as in [5].

**Table 1:** Features computed and statistics applied to them.

| Feature | Statistics | No. features |
|---|---|---|
| MFCCs | mean, std | 50 |
| ΔMFCCs | mean, std | 50 |
| Pitch | mean, std | 2 |
| HNR | mean, std | 2 |
| RUF | - | 1 |
| STE | mean, std | 2 |
| EE | mean, std, max,<br>max/median | 4 |
| ZCR | mean, std, max/mean | 3 |
| SR | mean, std, median | 3 |
| SC | mean, std | 2 |
| SF | mean, std | 2 |
| HA | mean, std | 2 |
| HF | mean, std | 2 |
| HPS Pitch | mean, std, mode | 3 |

Furthermore, more features related to the fundamental frequency have been added. That is the case of the amplitude and frequency of the first harmonic of the signal (Harmonic Amplitude, HA; Harmonic Frequency, HF), and the pitch based on Harmonic Product Spectrum method (HPS Pitch).

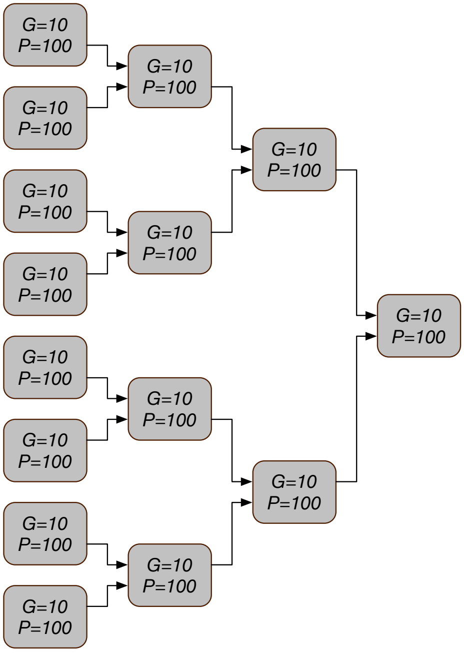
Different statistics have been applied to these measurements, as it is shown in Table 1.

### 2.2 Feature Selection

Once the features have been extracted, we have implemented a selection process due to the large amount of features extracted. It has been done by means of Evolutionary Algorithms (EAs), which are based on natural selection laws. The objective is to search the solution whose probability of error is lower.

First, a population of $P$ individuals is generated. These individuals have to be restricted to a set of selected features, so they are modified in order to fulfill the requirements. Then, the Least Squares Linear Detector (LSLD) [6] is applied with the features selected in the individuals, getting the Mean Square Error for each of them. The population is ranked according to this value. Subsequently, the best 10% of the individuals are kept, while the remaining 90% is deleted. These percentage

**Figure 1:** Elimination tournament with *R = 4* rounds, *G = 10* generations per round and *P = 100* individuals per AE.



is regenerated by random crossovers between the best subset. After that, we applied mutations to all the population except the best solution. The process is repeated since the restriction step until $G$ generations. After that, the best solution of the last iteration will be the solution in that EA.

We have used an elimination tournament algorithm of small EAs because the convergency of the solution is better respect to running an unique algorithm several times. At the beginning, we have implemented 32 small EAs with $P = 100$ individuals and $G = 10$ generations (first round). The best solutions of the first round are paired, generating a second round with half of the populations (16). The process is repeated $R$ rounds until we get the best solution. In our case we have implemented 6 rounds. Figure 1 shows an example of the algorithm with 4 rounds.

## 3 Results

In this section, the properties of the database are explained as well as the different experiments carried out to obtain the performance of the system.

**Table 2:** Summarized properties of the database.

| Parameters | Value |
|---|---|
| Total duration | 3425.9 s |
| Drone sound duration | 1812.8 s |
| Percentage of drone presence | 52.9% |
| Sampling frequency | 8000 Hz |
| Number of fragments | 36 |
| Minimum audio length | 5.8 s |
| Minimum audio length | 316.3 s |

### 3.1 Database

The database used in this work has been elaborated with *YouTube* audios, where different models of drones have been considered. Drones are recorded in motion as well as in a static position, and at different distances from the microphone. Some models of drones are included in this database, such as *Cheerson CX10*, *DJI Phantom 3* or *Eachine Racer 250*. Similar no-drone sounds are included too (mowers, cut-off wheels, combine harvesters, motorbikes, shavers, etc.). The purpose of including these similar audios to the database is to make it more challenging. The properties of the database are summarized in Table 2.

### 3.2 Discussion

The measure chosen to evaluate the accuracy of the system is the Probability of error ($P_e$) since the database is balanced. In other words, the percentage of the database with drone sounds is 50%. Different experiments have been carried out in order to compare which features are the best for the problem and which observation period shows best results. As we have mentioned before, LSLD has been applied in the experiments.

The first step is to prepare the data, so different $T_{obs}$ are used ($T_{obs}$ = 1, 2, 5, 7, 10). The second step is to fix the number of features which will be selected with the genetic algorithm ($N_f$ = 10, 20, 30, 40, 50). The percentage of error obtained in each different situation is displayed in Table 3.

The best result achieves 12.8% of error, obtained with 2 seconds of observation and 40 features. But due to the result of 12.9% is quite similar and it is got with 30 features, which implies low cost in computational terms, this will be the result selected in this paper and the discussion of the results below are referenced to these situation ($T_{obs}$=2s, $N_f$=30).

**Table 3:** Probability of error in function of observation period and number of features selected.

| Error (%) | $N_f$ | | | | |
|---|---|---|---|---|---|
| $T_{obs}$ | 10 | 20 | 30 | 40 | 50 |
| 1 s | 17.2 | 15.2 | 15.6 | 16.5 | 15.7 |
| 2 s | 17.6 | 14.0 | **12.9** | **12.8** | 16.3 |
| 5 s | 15.4 | 13.6 | 14.5 | 14.7 | 15.2 |
| 7 s | 16.1 | 14.6 | 16.1 | 14.4 | 16.0 |
| 10 s | 20.1 | 14.8 | 15.6 | 13.1 | 14.1 |

By analyzing this result, the error of 12.9% is splitted into False Positives (FP) and False Negatives (FN), considering FP when the system detect drone and it is not present, and FN when the system do not detect the drone when it is present. Table 4 shows the no-drone audios which are included in the database. The most problematic sound corresponds to building work due to the acoustic similarity between some construction tools and drones. Fire siren is another difficult sound because its spectrum is quite similar to drones one. However, sounds like helicopter, shaver or mower, which can be confused with drones, are correctly classified by the system.

In Table 5 FP and FN are shown in function of the model of drone. Six different models are included in this database. The best result corresponds to *DJI Phantom*, which is not detected in just 46 of 1219 seconds.

**Table 4:** False Positive results in no-drone sounds.

| Description | Total Dur. (s) | FP (s) | Contribution to the total FP (%) |
|---|---|---|---|
| Plane | 127.8 | 1.0 | 0.5 |
| Helicopter | 123.6 | 5.0 | 2.5 |
| Shaver | 248.6 | 14.0 | 7.1 |
| Building work | 316.3 | 66.0 | 33.7 |
| Digger | 147.5 | 0.0 | 0.0 |
| Motorbike | 150.0 | 17.0 | 8.7 |
| Mower | 268.5 | 21.0 | 10.7 |
| F1 car | 18.2 | 5.0 | 2.5 |
| Cut-off wheel | 21.8 | 9.0 | 4.6 |
| Fire siren | 135.3 | 56.0 | 28.6 |
| Drag racer | 55.5 | 2.0 | 1.0 |
| Total | 1613.1 | 196.0 | 100.0 |

**Table 5:** False Negative results in drone sounds.

| Model | Total Dur. (s) | FN (s) | Contribution to the total FN (%) |
|---|---|---|---|
| DJI Phantom | 1218.9 | 46.0 | 17.6 |
| UDI 817 | 16.0 | 16.0 | 6.1 |
| Parrot AR | 103.3 | 46.0 | 17.6 |
| Cheerson CX10 | 71.5 | 39.0 | 14.9 |
| Eachine Racer 250 | 149.1 | 26.0 | 10.0 |
| Rest of models | 253.9 | 88.0 | 33.7 |
| Total | 1812.8 | 261.0 | 100.0 |

Related to the most selected features, Table 6 shows a ranking corresponding to $T_{obs} = 2s$ and $N_f = 30$. The importance of the MFCCs is reflected in the results, since 20 of the 30 best measurements correspond to them and the five first features are MFCCs with a percentage of appearance of 100%. The standard deviation of the EE is another remarkable feature, as well as the standard deviation of the pitch, the mean of the SC, the mean of HF or the mean of STE.

## 4 Summary

The obtaining of an error below 13% demostrates the capability of the system proposed to be implemented into a real-time and real-environment system.

Some ongoing work could be related to an increase in the database, including more different drone sounds, and even more no-drone sounds. Furthermore, some different classifiers could be tested then, because with the current database more complex classifiers achieved worse results. Because of that, they are not included in this work. A more complex preprocessing could be applied to the data in order to clean the background noise which appears in audios of real-environments.

## 5 Acknowledgments

This work has been funded by the Spanish Ministry of Economy and Competitiveness (under project TEC2015-67387-C4-4-R, funds Spain/FEDER) and by the University of Alcalá (under project CCG2016/EXP-033) under an FPI scholarship.

**Table 6:** Summary of the most selected features.

| No. | Measure | Statistic | Occ. (%) |
|---|---|---|---|
| 1 | MFCC 4 | Mean | 100.0 |
| 2 | MFCC 5 | Mean | 100.0 |
| 3 | MFCC 10 | Mean | 100.0 |
| 4 | MFCC 11 | Mean | 100.0 |
| 5 | MFCC 12 | Mean | 100.0 |
| 6 | EE | Std | 100.0 |
| 7 | MFFC 14 | Mean | 97.2 |
| 8 | ΔMFCC 1 | Mean | 97.2 |
| 9 | Pitch | Std | 97.2 |
| 10 | SC | Mean | 97.2 |
| 11 | HF | Mean | 97.2 |
| 12 | MFCC 3 | Median | 94.4 |
| 13 | STE | Mean | 91.7 |
| 14 | MFCC 19 | Mean | 83.3 |
| 15 | HNR | Std | 80.6 |
| 16 | MFCC 6 | Mean | 66.7 |
| 17 | MFCC 19 | Std | 61.1 |
| 18 | HNR | Mean | 58.3 |
| 19 | ΔMFCC 3 | Std | 50.0 |
| 20 | ΔMFCC 14 | Std | 50.0 |
| 21 | EE | Max | 50.0 |
| 22 | MFCC 2 | Mean | 44.4 |
| 23 | ΔMFCC 14 | Mean | 38.9 |
| 24 | MFCC 12 | Std | 33.3 |
| 25 | SR | Std | 33.3 |
| 26 | MFCC 20 | Std | 30.6 |
| 27 | MFCC 25 | Std | 30.6 |
| 28 | STE | Std | 30.6 |
| 29 | ΔMFCC 10 | Std | 27.8 |
| 30 | MFCC 17 | Mean | 25.0 |

## 6 References

[1] Altawy, R., Youssef, A. M. (2016). Security, Privacy, and Safety Aspects of Civilian Drones: A Survey. ACM Transactions on Cyber-Physical Systems, 1(2), 7.

[2] Ganti, S. R., Kim, Y. (2016, June). Implementation of detection and tracking mechanism for small UAS. In Unmanned Aircraft Systems (ICUAS), 2016 International Conference on (pp. 1254-1260). IEEE.

[3] Case, E. E., Zelnio, A. M., Rigling, B. D. (2008, July). Low-cost acoustic array for small UAV detection and tracking. In Aerospace and Electronics Conference, 2008. NAECON 2008. IEEE National (pp. 110-113). IEEE.

[4] King, J. M., Faruque, I. (2016). Small Unmanned Aerial Vehicle Passive Range Estimation from a Single Microphone. In AIAA Atmospheric Flight Mechanics Conference (p. 3545).

[5] J. García-Gómez, M. Bautista-Durán, R. Gil-Pita, I. Mohino-Herranz and M. Rosa-Zurera, *Violence Detection in Real Environments for Smart Cities*, in Ubiquitous Computing and Ambient Intelligence: 10th International Conference, UCAmI, (Springer International Publishing, Spain, 2016), Part II 10, pp. 482–494.

[6] Van Trees, H. L. (2004). Detection, estimation, and modulation theory. John Wiley Sons.