Crossing Gate Detection Using Audio Frequency Pitch Tracking

Caleb Rascon[1*], L. Oyuki Rojas-Perez[2], Victoria Eugenia Vazquez-Meza[2], Aldrich A. Cabrera-Ponce[3] and Jose Martinez-Carranza[2*]
[1]Instituto de Investigaciones en Matematicas Avanzadas y en Sistemas, Universidad Nacional Autonoma de Mexico.
[2]Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE), Puebla, Mexico.
[3]Benemérita Universidad Autónoma de Puebla (BUAP), Puebla, Mexico.

ABSTRACT

Autonomous Drone Racing (ADR) has garnered significant interest in the aerial robotics community. Early solutions used classical computer vision algorithms for gate detection, while more recent approaches employed visual Simultaneous Localisation and Mapping (SLAM). The latest advancements have showcased solutions capable of winning races against world champions. However, these rely primarily on visual data from onboard cameras, whereas humans complement visual sensing with auditory perception. Motivated by the benefits of auditory perception, this study investigates the use of audio signal processing to detect when a drone crosses a gate during a race. This detection addresses the blind spot issue, where the gate disappears from the visual sensor's view after crossing. Initial results indicate the feasibility of using audio signals to identify gate crossings, based on sound changes caused by drone propellers. This is a first effort to explore the broader potential of auditory perception in autonomous drone racing.

1 INTRODUCTION

Autonomous Drone Racing (ADR) is a scientific challenge that has garnered significant interest within the aerial robotics community. Initially, solutions relied on classical computer vision algorithms for gate detection to guide the drone's flight towards the gate [1]. Subsequent competitions incorporated more sophisticated techniques such as visual Simultaneous Localisation and Mapping (SLAM) [2]. Recently, impressive results were achieved in [3], demonstrating a comprehensive solution that won several races against two world champions.

Despite these advancements, the best solutions to date rely primarily on visual data captured by onboard cameras. While impressive, humans do not rely solely on visual sensing to navigate through the flying arena. Our sense of surroundings is complemented by auditory perception, which, in

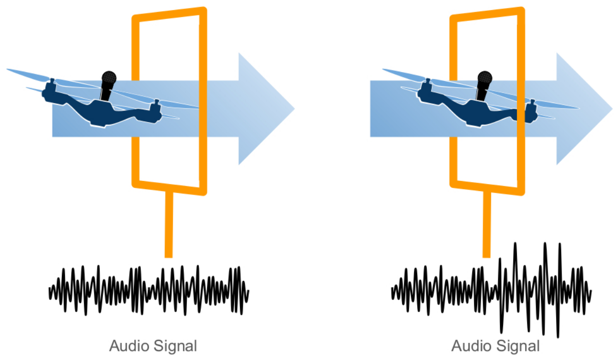


Figure 1

the context of a drone race with several adversaries, is used to infer when other drones fly by. Furthermore, trained pilots use auditory perception to detect malfunctioning motors or propellers.

Motivated by the advantages of auditory perception, we present a preliminary study on how audio signal processing could be used for a specific task in autonomous drone racing: detecting when the drone crosses the gate during the race. Detecting gate crossings can be highly beneficial for designing navigation policies for autonomous drone racers. If a policy relies on vision for gate detection, the gate will eventually move out of the visual sensor's view once the drone crosses it. However, this can also occur if the drone navigates off the planned trajectory, flying past the gate but not through it.

The moment the drone flies towards the gate and crosses it—when the gate disappears from the visual sensor's field of view—is what we call the blind spot. Given the importance of detecting this blind spot, we explore the use of audio signals to detect when the drone crosses the gate. This is based on the observation that there is a slight change in sound due to the wind produced by the drone's propellers hitting the gate structure.

Our initial results indicate that this detection is possible. We acknowledge that this detection might be specific to our experimental setup, including the type of drone and gate used. Nevertheless, it is intriguing to see that audio signals can be utilised in this context. Further research could explore the potential of auditory perception in autonomous drone racing

*Email address(es): carranza@inaoep.mx, caleb@unam.mx

2 RELATED WORK

Drone racing is a challenging benchmark to gauge progress in complex perception, planning, and control algorithms, as drones must be able to perceive and interpret the scene, plan, and act within milliseconds [2, 4, 5]. In a racing setting, the quadrotor must fly through a sequence of gates in the minimum time without collisions. Two strategies are proposed to solve this challenge.

The first strategy uses external sensors, such as a motion capture system, and assumes a controlled environment with privileged knowledge. This means that the control system has full access to the drone's state (position and orientation) and the number and state of the gates, which makes it possible to implement a time-optimal planner and a flight controller based on MPC [6, 7]. Nevertheless, this is a partial solution to drone racing because the perception task is removed to complete the racetrack at high speed.

The second strategy uses only onboard sensors for state estimation and gate identification [2, 8, 9, 10], the latter employing techniques based on Deep Learning used for object detection, such as a Single Shot Detector (SSD) [11, 12] or YOLO [13]. However, systems based on visual information face a challenge when the drone crosses the gate because the gate disappears from the field of view, making it impossible for the system to determine if the drone is in the middle of the gate or outside. This issue is referred to as a blind spot. Some solutions have utilised different sensors to overcome this problem, such as a camera to see the gate's base [8, 9], a LIDAR to measure changes between the drone and the ground [1], an IMU to estimate the drone's displacement [14, 15], or visual temporal information to identify the entrance and exit of the gate [16, 12, 17].

This paper proposes a novel approach based on acoustics to assist in blind spot detection. By incorporating audio sensors, the system can detect subtle changes in sound that indicate the drone's position relative to the gate, providing an additional layer of information that visual sensors might miss. Auditory perception offers several opportunities. Audio sensors have introduced auxiliary features that can be useful for drone navigation. In [18], the authors provide an extensive analysis of works that have reported using microphones to capture audio signals and the applications where drones are involved. For example, avoid collisions [19], sound source localisation [20, 21], detection of propeller anomalies [22], intruder detection [23, 24], classification [25, 26, 27], and drone proximity [28]. This motivates the analysis of noise production and mitigation [29].

3 PROPOSED METHODOLOGY

When a drone is passing through a gate, its height sensor will briefly sense the gate, which may cause the drone to assume that the floor has suddenly "jumped" up which, in turn, may cause its flight control framework (that has the task of maintaining a given altitude) to incorrectly fly upward. To avoid this, it is common for drones' flight control frameworks to ignore high-variance changes in the height sensor values. Additionally, since the height sensor sampling frequency and/or the frequency of the feedback loop are usually relatively low (¡ 5 Hz.), this issue may not even be contemplated.

In any case, it is frequently observed in flight control frameworks that the drone's altitude is briefly "corrected" to counter this sudden change in height. During this momentary correction, the drone's motors can be heard being briefly engaged. In Figure 2, it can be seen how the pitch (reported in Hertz in the vertical axis) changes through time (horizontal axis). It can also be seen how the pitch of the drone's motors presents sudden but major changes while crossing the gate (this moment is presented as a vertical black line[1]).

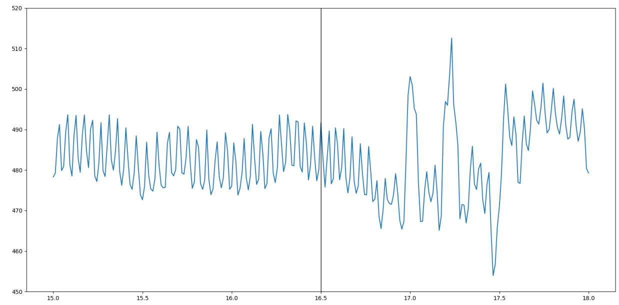


Figure 2: Drone's motor pitch through time. Black vertical line: a rough estimation of when the drone crossed the gate.

In this work, the pitch of an audio signal refers to its highest-magnitude frequency. The pitch reported in Figure 2 was estimated by first carrying out the Short-Time Fourier Transform (STFT, with a FFT window length of N samples and hop length of H samples) of L_w seconds of the audio recording of the drone's noise. Since the pitch is the main focus of the proposed technique, it is important that its estimation is not limited by the frequency binning of the discrete Fast Fourier Transform (FFT). Thus, the maximum value of each FFT window of the resulting STFT'ed signal was refined using the popular quadratically-interpolated FFT technique (QIFFT) [30]. This process finds a closer value of the true peak, between frequency bins, by extrapolating a parabola using the frequency bin of the maximum value of the FFT window and its two neighbouring frequency bins. The result is similar to using a maximum-likelihood estimator, since a parabola is a Gaussian window in the decibel scale.

The proposed technique is as follows:

Every L_h seconds, the pitch information ($[p_1, p_2, \ldots, p_K]$) is estimated from each of the K FFT windows of the current L_w seconds of captured audio. Then, it is established if the estimated pitch of the last FFT

[1] This is a rough estimation, based on manually checking the video of the drone's flight.

Algorithm 1 Pseudo-code for proposed technique.

K := number of FFT windows from STFT
$\bar{p}$:= mean pitch of the last L_w window
σ_p := pitch variance of the last L_w window
for every L_h seconds **do**
  X := STFT of current L_w seconds
  X_k := kth FFT window of X
  for $k = 1$ **to** K **do**
    $p_k \leftarrow max(QIFFT(|X_k|))$
  end for
  if $|p_K - \bar{p}| > M_\sigma \sigma_p$ **then**
    DRONE IS CROSSING GATE
  end if
  $\bar{p} \leftarrow \left[\sum_{k=1}^{K} p_k\right] / K$
  $\sigma_p \leftarrow \sqrt{\left[\sum_{k=1}^{K} (p_k - \bar{p})\right] / K}$
end for

window (p_K) has suddenly considerably changed from past pitches. This is the case if the difference between p_K and the mean pitch of the last L_w seconds ($\bar{p}$) is greater than its pitch variance (σ_p), multiplied by a sensitivity factor (M_σ). Finally, $\bar{p}$ and σ_p are re-calculated with the pitch information of the current L_w seconds to be used for the next capture window.

An important virtue of the proposed technique is that its response time can be calibrated by establishing the value of L_h. The lower limit of this value is the processing time of Algorithm 1. This time, in turn, depends on the processing time of carrying out the STFT and QIFFT, both of which are quite efficient. However, to allow for real-time operation, the reduction of the value of L_h should be accompanied with a reduction of L_w, since the STFT operation depends on it, as well as the $k \in [1, K]$ loop in Algorithm 1. Reducing the value of L_w provides less past pitch information, making the pitch change less predictable. The values of L_h and L_w presented in the next section provide a good balance between response time and predictable behaviour.

It may be tempting to change Algorithm 1 so that, instead of calculating the STFT of the whole L_w audio window every L_h seconds, we only calculate the FFT of the last FFT window. Doing so would reduce the processing time considerably. However, this would be equivalent on carrying out the STFT with no overlap between FFT windows ($H = 0$ samples), which would result in a highly varying pitch-over-time signal. Carrying out the STFT of the whole L_w audio window results in a much smoother pitch transition through time, which simplifies detecting sudden pitch changes. Moreover, the LibROSA [31] implementation of both the STFT and QIFFT was used (via the *piptrack* function), which provided very low response times (~ 0.002 seconds to process $L_w = 1.5$ seconds of audio in hardware of moderate computing power). Having established all of this, it would be of interest (and left for future work) to further lower the response time of Algorithm 1 by only updating $\bar{p}$ and σ_p using only the current FFT window to model a smooth pitch-over-time signal.

Figure 3: Drone and hardware employed in our experimental framework. We used a wireless microphone transmitting to the GCS, an Alienware 15 R3 laptop.

It is important to mention that this technique is aimed to be used once a gate crossing is expected in the near future of the planned flight; no further flight adjustments should be carried out once this technique is triggered. This is because a change in pitch can occur mid-flight for reasons other than a gate crossing. As a reminder, the advantage of this technique is its potentially low response time compared to using other light-based sensors, which (as previously described) is highly configurable given different scenarios.

4 Evaluation and Results

For our experiments, we used the Bebop 2.0 Power Edition drone from Parrot. We attached a Bluetooth microphone to this drone, which transmitted to the Ground Control Station (GCS) - an Alienware 15 R3 laptop equipped with a GeForce GTX 1070 Graphics Processing Unit. The microphone weighs 80g and can transmit over a distance of up to 20 metres, see Fig. 3.

Several recordings were captured for testing, where the drone was given a linear route that passes through a gate. A microphone was connected to the onboard computer to record the audio. Additionally, a video recording was also captured during the flight to provide an estimate of when the drone passed through the gate.

In our tests, we used the following configuration that allows for real-time operation, with a relatively low response time:

In Figure 5, three runs are shown using the proposed technique to detect the crossing of the gate. Similar to Figure 2, a vertical black line shows a rough estimation of the moment when the drone is crossing the gate. To facilitate the visualisation of the results, an orange line reports the gate crossing

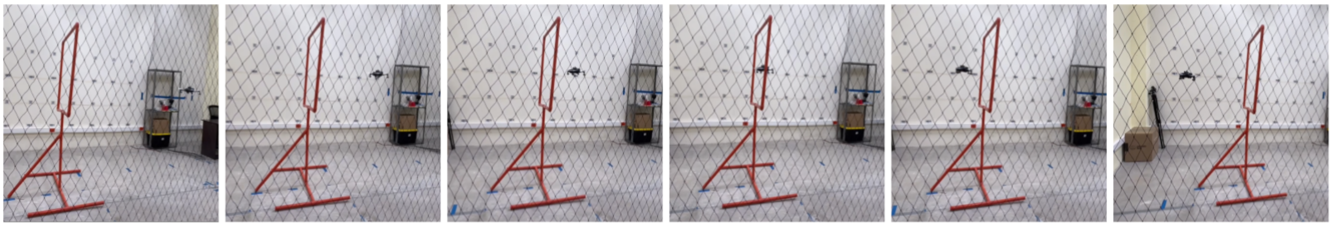


Figure 4: From left to right: image shots illustrating one of the runs in the lab where the drone flew through a square gate with a width of 1m and 2m in height. The drone has a microphone attached on board transmitting the audio signal to the GCS.

L_w	1.5 seconds
L_h	0.1 seconds
M_σ	2
N	2048 samples
H	512 samples
f_s	48000 samples per second
K	$L_w f_s / H \rightarrow 140$ FFT windows

Table 1: Configuration values used in tests.

detection results in the form of a peak.

As it can be seen, a peak is present when the pitch has changed dramatically from the value of its predecessors, clearly marking the moment when the drone's motors were briefly adjusted. This, in turn, can be assumed as being the moment in which the drone is crossing the gate.

It can also be seen how the technique picks up other pitch changes after the gate crossing, probably due to the drone's motor changing speed during the drone's landing procedure. This recalls what was mentioned earlier of how this technique should only be used when the gate crossing is eminent in the drone's flight plan.

5 Conclusion

In this work, we have presented initial results on the use of audio signal processing to detect when a drone flies through a gate in the context of autonomous drone racing. We used an off-the-shelf Bluetooth microphone that transmits the audio signal to the Ground Control Station (GCS). As shown in our experiments, our methodology enables the detection of spikes in the audio signal triggered by the drone's propellers when it crosses the gate. This sound is similar to that produced by the propellers during take-off or when switching to forward motion after hovering, and likewise when stopping and returning to hovering. However, it is notable that this similar sound is produced when crossing the gate.

We argue that when crossing the gate, the drone detects the change in height from the floor to the horizontal bar of the gate with its altimeter sensor, and its internal controller attempts to maintain horizontal flight to keep moving forward. This slight compensation produces the change in sound that can be detected with our methodology.

Although detecting the gate crossing could be done with the altimeter alone, audio processing could act as a complementary measurement, hence our interest in investigating this approach.

Aware of the specificity of our experimental scenario, in our future work, we plan to conduct additional experiments with various gate types and different drones. Furthermore, we will assess whether the changes in the audio signal are caused by the gate's aerodynamics or result from the Bebop's internal control system, where the altitude sensor detects the base of the gate and adjusts the drone's signals to maintain the required altitude.

Acknowledgements

The second, third, and fourth authors are thankful to the National Council of Humanities, Science and Technology of Mexico (CONAHCYT) for their scholarship number 924254, 1026123, and 802791, respectively.

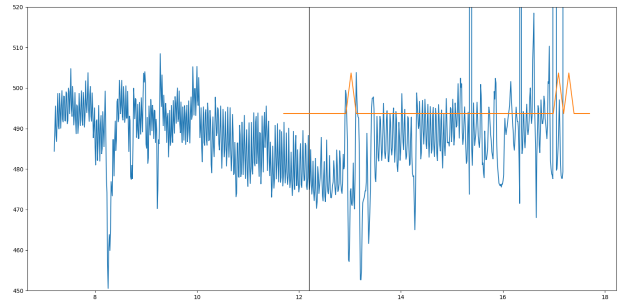
References

[1] Sunggoo Jung, Sungwook Cho, Dasol Lee, Hanseob Lee, and David Hyunchul Shim. A direct visual servoing-based framework for the 2016 iros autonomous drone racing challenge. *Journal of Field Robotics*, 35(1):146–166, 2018.

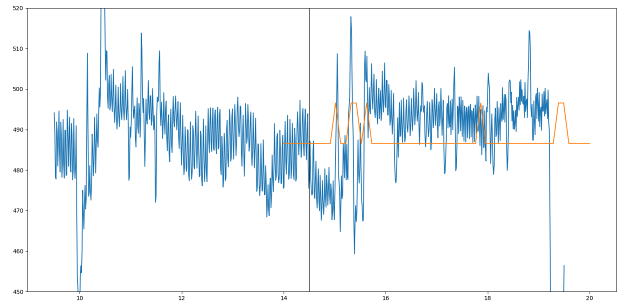
[2] Hyungpil Moon, Jose Martinez-Carranza, Titus Cieslewski, Matthias Faessler, Davide Falanga, Alessandro Simovic, Davide Scaramuzza, Shuo Li, Michael Ozo, Christophe De Wagter, et al. Challenges and implemented technologies used in autonomous drone racing. *Intelligent Service Robotics*, 12(2):137–148, 2019.

[3] Elia Kaufmann, Leonard Bauersfeld, Antonio Loquercio, Matthias Müller, Vladlen Koltun, and Davide Scaramuzza. Champion-level drone racing using deep reinforcement learning. *Nature*, 620(7976):982–987, 2023.

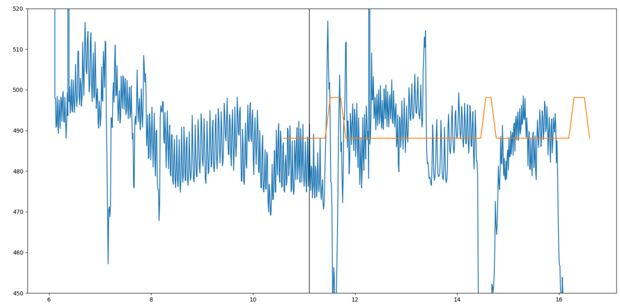
[4] L Oyuki Rojas-Perez and José Martínez-Carranza. Onboard processing for autonomous drone racing: An overview. *Integration*, 80:46–59, 2021.



(a) Run 1.



(b) Run 2.



(c) Run 3.

Figure 5: Detection results (orange line) plotted over the pitch-over-time signal.

[5] Drew Hanover, Antonio Loquercio, Leonard Bauersfeld, Angel Romero, Robert Penicka, Yunlong Song, Giovanni Cioffi, Elia Kaufmann, and Davide Scaramuzza. Autonomous drone racing: A survey. *IEEE Transactions on Robotics*, 2024.

[6] Philipp Foehn, Dario Brescianini, Elia Kaufmann, Titus Cieslewski, Mathias Gehrig, Manasi Muglikar, and Davide Scaramuzza. Alphapilot: Autonomous drone racing. *Autonomous Robots*, 46(1):307–320, 2022.

[7] Angel Romero, Robert Penicka, and Davide Scaramuzza. Time-optimal online replanning for agile quadrotor flight. *IEEE Robotics and Automation Letters*, 7(3):7730–7737, 2022.

[8] Sunggoo Jung, Sunyou Hwang, Heemin Shin, David Hyunchul Shim, et al. Perception, guidance, and navigation for indoor autonomous drone racing using deep learning. *IEEE Robotics and Automation Letters*, 3(3):2539–2544, 2018.

[9] Sunggoo Jung, Hanseob Lee, Sunyou Hwang, and David Hyunchul Shim. Real time embedded system framework for autonomous drone racing using deep learning techniques. In *2018 AIAA Information Systems-AIAA Infotech@ Aerospace*, page 2138. 2018.

[10] Nitin J Sanket, Chahat Singh, Kanishka Ganguly, Cornelia Fermuller, and Yiannis Aloimonos. Gapflyt: Active vision based minimalist structure-less gap detection for quadrotor flight. *IEEE Robotics and Automation Letters*, 2018.

[11] Aldrich A Cabrera-Ponce, L Oyuki Rojas-Perez, Jesus Ariel Carrasco-Ochoa, Jose Francisco Martinez-Trinidad, and Jose Martinez-Carranza. Gate detection for micro aerial vehicles using a single shot detector. *IEEE Latin America Transactions*, 17(12):2045–2052, 2019.

[12] L Oyuki Rojas-Perez and Jose Martinez-Carranza. Where are the gates: discovering effective waypoints for autonomous drone racing. In *Ibero-American Conference on Artificial Intelligence*, pages 353–365. Springer, 2022.

[13] L Oyuki Rojas-Perez and Jose Martinez-Carranza. Towards autonomous drone racing without gpu using an oak-d smart camera. *Sensors*, 21(22):7436, 2021.

[14] J Arturo Cocoma-Ortega, L Oyuki Rojas-Perez, Aldrich A Cabrera-Ponce, and J Martinez-Carranza. Overcoming the blind spot in cnn-based gate detection for autonomous drone racing. In *2019 Workshop on Research, Education and Development of Unmanned Aerial Systems (RED UAS)*, pages 253–259. IEEE, 2019.

[15] Shuo Li, Michal MOI Ozo, Christophe De Wagter, Guido CHE de Croon, et al. Autonomous drone race: A computationally efficient vision-based navigation and control strategy. *Robotics and Autonomous Systems*, page 103621, 2020.

[16] L Oyuki Rojas-Perez and Jose Martinez-Carranza. Deeppilot: A cnn for autonomous drone racing. *Sensors*, 20(16):4524, 2020.

[17] L Oyuki Rojas-Perez and Jose Martinez-Carranza. Effective training to improve deeppilot. *AI Communications*, (Preprint):1–18, 2023.

[18] Jose Martinez-Carranza and Caleb Rascon. A review on auditory perception for unmanned aerial vehicles. *Sensors*, 20(24):7276, 2020.

[19] Brendan Harvey and Siu O'Young. Acoustic detection of a fixed-wing uav. *Drones*, 2(1):4, 2018.

[20] Koutarou Furukawa, Keita Okutani, Kohei Nagira, Takuma Otsuka, Katsutoshi Itoyama, Kazuhiro Nakadai, and Hiroshi G Okuno. Noise correlation matrix estimation for improving sound source localization by multirotor uav. In *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3943–3948. IEEE, 2013.

[21] Takuma Ohata, Keisuke Nakamura, Takeshi Mizumoto, Tezuka Taiki, and Kazuhiro Nakadai. Improvement in outdoor sound source detection using a quadrotor-embedded microphone array. In *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1902–1907. IEEE, 2014.

[22] Sai Srinadhu Katta, Kide Vuojärvi, Sivaprasad Nandyala, Ulla-Maria Kovalainen, and Lauren Baddeley. Real-world on-board uav audio data set for propeller anomalies. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 146–150. IEEE, 2022.

[23] Oscar Ruiz-Espitia, Jose Martinez-Carranza, and Caleb Rascon. Aira-uas: an evaluation corpus for audio processing in unmanned aerial system. In *2018 International Conference on Unmanned Aircraft Systems (ICUAS)*, pages 836–845. IEEE, 2018.

[24] Aldrich A Cabrera-Ponce, J Martinez-Carranza, and Caleb Rascon. Detection of nearby uavs using a multi-microphone array on board a uav. *International Journal of Micro Air Vehicles*, 12:1756829320925748, 2020.

[25] Matthew Ritchie, Francesco Fioranelli, Hervé Borrion, and Hugh Griffiths. Multistatic micro-doppler radar feature extraction for classification of unloaded/loaded micro-drones. *IET Radar, Sonar & Navigation*, 11(1):116–124, 2017.

[26] JJM De Wit, RIA Harmanny, and P Molchanov. Radar micro-doppler feature extraction using the singular value decomposition. In *2014 International Radar Conference*, pages 1–6. IEEE, 2014.

[27] Sara Al-Emadi, Abdulla Al-Ali, Amr Mohammad, and Abdulaziz Al-Ali. Audio based drone detection and identification using deep learning. In *2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC)*, pages 459–464. IEEE, 2019.

[28] Qian Zhao, Jason Hughes, and D Lyons. Drone proximity detection via air disturbance analysis. In *Unmanned Systems Technology XXII*, volume 11425, pages 141–149. SPIE, 2020.

[29] Caleb Rascon and Jose Martinez-Carranza. A review of noise production and mitigation in uavs. *Machine Learning for Complex and Unmanned Systems*, pages 220–235, 2024.

[30] Julius O Smith. Parshl: A program for the analysis/synthesis of inharmonic sounds based on a sinusoidal representation. *Stanford Music Department Technical Report, STAN-M-43*, 1987.

[31] Brian McFee, Colin Raffel, Dawen Liang, Daniel PW Ellis, Matt McVicar, Eric Battenberg, and Oriol Nieto. librosa: Audio and music signal analysis in python. In *Proceedings of the 14th python in science conference*, volume 8, 2015.