DOI: 10.61782/fa.2025.0072

ACOUSTIC DETECTION OF DRONES: A MACHINE LEARNING AND PSYCHOACOUSTIC APPROACH

Diana Tejera-Berengue[1*] **Gonzalo Corral-García**[1] **Fangfang Zhu-Zhou**[1]
Manuel Rosa-Zurera[1*] **Roberto Gil-Pita**[1] **Manuel Utrilla-Manso**[1]
[1] Department of Signal Theory and Communications, University of Alcalá, Spain

ABSTRACT

This study presents a detailed analysis of how acoustic signals generated by drones are perceived by humans compared to their detection by advanced microphone-based systems. The research includes various types of drones and employs a psychoacoustic model to evaluate the perceptual sound power of these signals. This model is used to estimate the maximum distance at which UAV noise remains audible, as a function of frequency. Additionally, a detection system incorporating machine learning techniques and the YAMNet neural network is implemented to investigate how drone acoustic signals are influenced by factors such as distance, frequency, and surface reflections during propagation. The results show that even the most basic acoustic detection systems significantly outperform human hearing in identifying drones. These findings underscore the effectiveness and versatility of such systems, highlighting their potential as valuable tools for enhancing security and surveillance in real-world scenarios.

Keywords: UAV, drone audibility, psyschoacoustic model, distance, masking threshold, comparison

1. INTRODUCTION

The acoustic detection and localization of drones might initially seem ineffective due to the low signal-to-noise ratio (SNR) typically associated with the sound emitted by small unmanned aerial vehicles (UAVs) at the receiver. Additionally, it is sometimes assumed that an acoustic surveillance system cannot outperform the human auditory system. In response to these assumptions, this study investigates the limits of the human auditory system's ability to detect drones, using MPEG Psychoacoustic Model 1 [1], originally developed for audio encoding. The goal is to determine the maximum distance at which a drone remains perceptible to the human ear and to compare this threshold with the detection capabilities of existing automated systems.

In the field of environmental acoustics, several works have evaluated the perceptual impact of drone noise using psychoacoustic sound quality metrics (SQMs), such as loudness, sharpness, and fluctuation strength. For instance, the study in [2] analyzes the perceived annoyance of UAV noise under varying altitudes and flight conditions, concluding that heavier drones flying at low altitudes generate greater discomfort, with loudness emerging as the dominant factor. Similarly, the authors of [3] compare the sounds of drones, helicopters, and lawnmowers, again identifying loudness as the main predictor of annoyance. In a related study [4], multirotor drones are reported to be more annoying than civil aircraft, with sharpness and fluctuation strength playing a key role in perceived disturbance.

However, there is still a lack of systematic analysis regarding the maximum distance at which drone sound remains perceptible to humans. This paper aims to address that gap by estimating the distance beyond which drone sound becomes fully masked by environmental noise and, therefore, no longer causes audible annoyance, or can be detected by humans beings. The analysis is carried out using the MPEG Psychoacoustic Model 1, which enables

****Corresponding author***:
diana.tejera@uah.es; manuel.rosa@uah.es.

the calculation of individual and global masking thresholds, to determine the audibility of signals emitted at a given distance from the listener and affected by propagation losses. This study also reinforces the relevance of acoustic localization and detection systems, demonstrating their ability to detect drones even at distances beyond human auditory capabilities.

To contextualize the comparison, a state-of-the-art acoustic detection system is used as a benchmark. Many modern detection systems rely on machine learning (ML) and deep learning (DL) methods applied to features such as the short-time Fourier transform (STFT) or Mel-frequency cepstral coefficients (MFCC) extracted from drone-specific acoustic patterns. These techniques have shown high classification accuracy and effective range under certain conditions [5–7]. Among these, the YAMNet neural network, used in a previous study by the authors [8], serves here as a reference to evaluate the performance of an automatic system. Rather than providing a detailed assessment of detection systems, this paper focuses on quantifying the audibility of drone noise using a perceptual model, and comparing it with the response of an automatic classifier.

This paper is structured as follows. Section 2 presents the theoretical background of the psychoacoustic model and describes the acoustic characteristics of sound produced by drones. Section 3 analyzes the audibility of a drone at varying distances and compares these findings with the performance of the YAMNet-based detection system. Finally, Section 4 presents the main conclusions of the study.

2. THEORETICAL BACKGROUND

This section introduces the main concepts about psychoacoustics and signal characteristics relevant to the perceptual analysis of UAV sound.

2.1 Psychoacoustic Model

A psychoacoustic model is a quantitative framework that approximates the behavior of the human auditory system [9]. It computes the global masking threshold, which defines the sound pressure level below which acoustic signals are no longer perceived by the human ear. This threshold varies dynamically across both frequency and level, depending on the energy distribution in adjacent spectral components [10].

Human auditory perception is primarily determined by the threshold in quiet and by masking effects produced

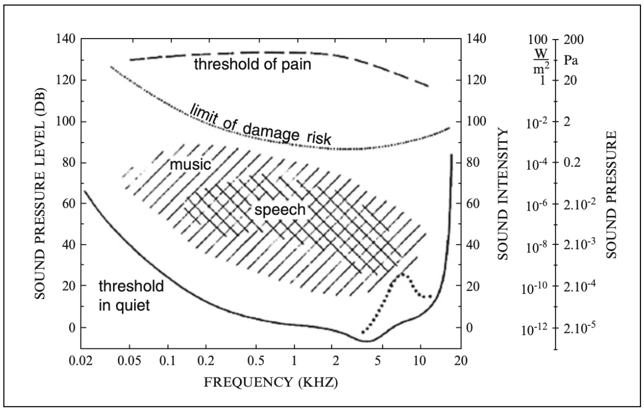


Figure 1: Hearing range [11].

by other sounds. The threshold in quiet refers to the minimum sound pressure level required to detect a pure tone in the absence of any background noise. This threshold is illustrated in Fig. 1, which shows the human hearing range using Sound Pressure Level (SPL) curves. As shown, human hearing ranges from approximately 20 Hz to nearly 20 kHz, with frequency-dependent thresholds for both hearing and pain. The threshold in quiet can be approximated by the frequency-dependent function provided in Eqn. (1).

$$T_Q(f) = 3.64\left(\frac{f}{1000}\right)^{-0.8} - 6.5\cdot\exp\left[-0.6\left(\frac{f}{1000}-3.3\right)^2\right] + 10^{-3}\left(\frac{f}{1000}\right)^4 \tag{1}$$

Masking effects at a given frequency depend solely on the signal energy within a limited bandwidth surrounding that frequency [12]. Due to the structure of the cochlea, which converts frequency information into spatial displacement along the basilar membrane, each position on the membrane responds to a specific range of frequencies. As a result, the peripheral auditory system functions as a spectral analyzer and can be modeled as a bank of partially overlapping band-pass filters. The audible range, approximately from 20 Hz to 16 kHz, is divided into 24 non-overlapping critical bands, each corresponding to a unit known as the Bark. Given the nonlinear distribution of resonant frequencies along the membrane, the bandwidths of these critical bands are not uniform and vary with frequency [9]. An analytical expression derived from experimental data is used to approximate the critical bandwidth

Δf as a function of the center frequency f_c:

$$\Delta f = 25 + 75\left(1 + 1.4\left(\frac{f_c}{1000}\right)^2\right)^{0.69} \quad (2)$$

The concept of critical bandwidth around a masker defines the frequency region in which masking effects are most pronounced. Consequently, the perception of a sound depends not only on its own frequency and intensity, but also on the presence of nearby spectral components. Masking phenomena are typically classified as either simultaneous or non-simultaneous. Simultaneous masking occurs when a sound becomes inaudible due to the concurrent presence of another sound with nearby frequencies, which raises the auditory threshold in that band. This can occur under several conditions: when narrowband noise masks tones within the same critical band, when tones mask other tones, or when narrowband noise masks other narrowband noise components within the same critical band. In contrast, non-simultaneous masking refers to the phenomenon where the audibility of a sound is affected by a masker that occurs immediately before or after it in time, due to the temporal dynamics of auditory masking.

In this work, perceptual analysis is based on Model 1 of the ISO/IEC MPEG-1 Audio Standard [1, 13]. This model divides the audio signal into subbands that approximate critical bands and estimates the masking threshold, to be used to quantize each subband signal according to the audibility of quantization noise [12].

To apply Model 1, the following steps are carried out:

1. Computation of the FFT and the SPL in each subband.
2. Determination of the absolute threshold of hearing.
3. Identification of tonal and non-tonal components in the input signal.
4. Decimation of invalid tonal and non-tonal maskers.
5. Calculation of individual masking thresholds.
6. Determination of the global masking threshold.
7. Estimation of the minimum masking threshold in each subband.
8. Calculation of the Signal-to-Mask Ratio (SMR).

In the final step, the SMR is computed as the difference between the signal level and the corresponding masking threshold. This value indicates, for each subband or frequency bin, whether a given sound is perceptible or masked by other components. If the SMR is low (i.e., negative in dB), the human ear cannot perceive the sound, even though it is physically present in the signal.

2.2 UAV signal Characterization

The acoustic signal generated by a drone originates primarily from the rotation of its propellers and motors, resulting in a sound with harmonic content when the rotational speed remains constant. However, during flight, this speed continuously varies in response to aerodynamic forces and control adjustments, leading to a non-stationary signal in the time domain. Consequently, frequency-domain analysis is more suitable, as the signal exhibits a spectral pattern characterized by peaks and valleys (as shown in Fig. 2), with energy concentrated in specific frequency bands associated with the rotational speeds of different components.

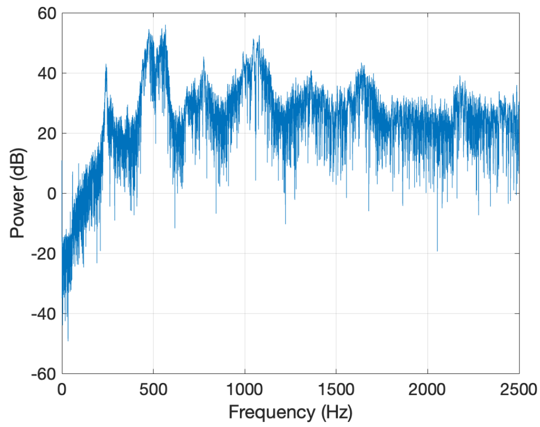


Figure 2: Power Spectrum of Hobbyking FPV250 drone.

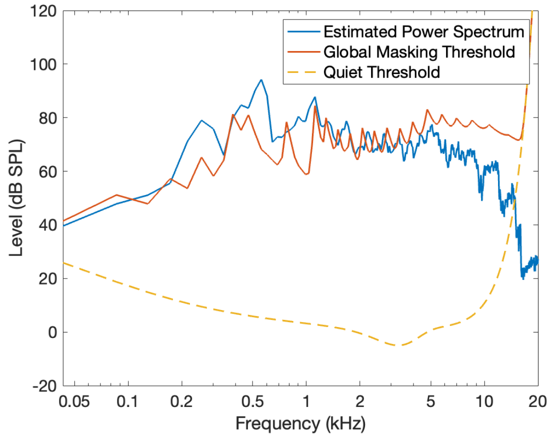
In particular, drone acoustic signals contain tonal components in the low-frequency range, corresponding to the blade pass frequency (BPF) and its harmonics [14]. These tonal components are linked to variations in rotor speed and are clearly reflected in Fig. 2. According to the study by Torija et al. [15], drones exhibit an average tonality of 0.36 tonality units (tu), indicating a moderately prominent tonal component in their acoustic signature. This characteristic has the potential to influence auditory masking effects and, therefore, human perception. These spectral characteristics highlight the suitability of drone noise for perceptual analysis based on auditory masking principles, which will be explored in the next section.

3. RESULTS

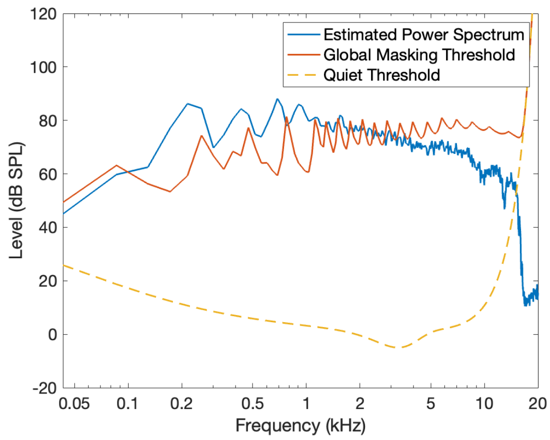
This section presents the results of the psychoacoustic analysis applied to UAV acoustic signals using the MPEG Psychoacoustic Model 1. The goal is to determine the maximum distance at which drone noise remains perceptible to the human ear, considering both ideal conditions and more realistic scenarios with propagation losses and environmental noise. In addition, results are also compared against a state-of-the-art detection system to highlight the limitations of human hearing.

As a first step, the analysis was performed on an in-flight recording of the Hobbyking FPV250 quadcopter. This UAV was selected as a representative example of a small and lightweight drone, typically classified in the C0 or C1 category according to current UAV classification standards. To complement this analysis, the DJI Phantom 3 was also evaluated, representing a larger and more robust drone classified within the C2 category. These two UAVs differ significantly in weight, physical dimensions, and structural design, providing a broader perspective on how human auditory perception responds to UAVs with different spectral and operational characteristics. Although some variations are expected between drones, most quadcopters share common tonal and non-tonal features in their acoustic signatures. Therefore, the perceptual results presented here are expected to generalize well to similar UAV configurations.

The initial analysis was conducted at a reference distance of 1 meter, without considering attenuation or background noise. Fig. 3 shows the estimated power spectrum, the global masking threshold calculated using the MPEG model, and the absolute threshold of hearing, for both UAV models. In both cases, the drone signals exhibit prominent energy in the low-frequency range, especially around the BPF and its harmonics. These tonal components result in masking thresholds that closely follow the spectrum shape, particularly where tonal maskers dominate. However, it can be observed that above approximately 1 kHz, the power spectrum begins to fall below the global masking threshold due to the progressive decline in signal energy with increasing frequency, combined with the accumulation of masking contributions from lower-frequency components. As a result, the perceptual significance of high-frequency components is reduced, and only the lower-frequency bands of the signal remain clearly audible. As observed in Fig. 3, both UAVs exhibit comparable spectral distributions and global masking thresholds under ideal conditions, with the signal remaining well above the masking threshold for frequencies below approximately 2 kHz. This confirms that both drones are clearly audible in quiet scenarios, although high-frequency information appears to be imperceptible. Given the similarity of their acoustic profiles, only the results of one of the two UAVs are presented in this paper for the subsequent analyses involving distance-based attenuation and ambient noise.



(a) Hobbyking FPV250



(b) DJI Phantom 3

Figure 3: Estimated power spectrum and masking threshold without considering noise or attenuation.

To evaluate auditory perception as a function of distance, an attenuation model was applied to estimate the frequency-dependent degradation of the acoustic signal over propagation. For this purpose, the standardized models defined in ISO 9613-1 [16] and ISO 9613-2 [17] were used, which account for effects such as atmospheric absorption or geometric spreading to simulate realistic propagation losses across different frequency bands. Based on this model, the next scenario considers the effect of distance alone, without introducing ambient noise.

As shown in Fig. 4, which corresponds to the DJI Phantom 3 at a distance of 100 meters, the signal spectrum shows a noticeable overall attenuation, particularly at high frequencies. This behaviour is expected due to the frequency-dependent nature of atmospheric absorption. However, in the frequency range where the signal was previously audible—primarily below 1 kHz—the spectral shape remains largely preserved, but with a lower dependence on the energy level. As a result, the masking threshold maintains the shape of the signal spectrum, and the signal-to-mask ratio (SMR) remains positive across most of the low-frequency bands. This indicates that, in the absence of ambient noise, increasing distance alone has the effect of reducing the signal power, but the SMR in the dominant low-frequency range remains almost unchanged.

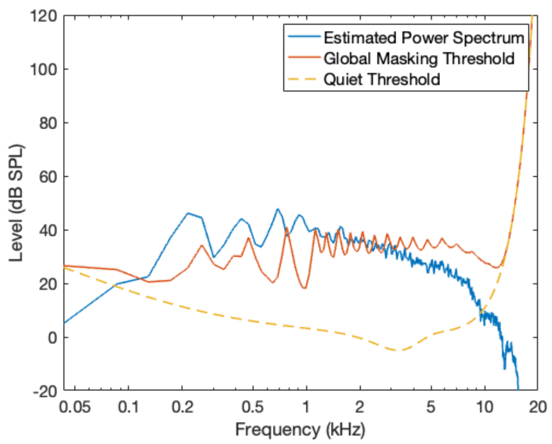


Figure 4: Estimated power spectrum and masking threshold at 100 meters without noise.

To assess the impact of environmental noise on auditory perception, an additional analysis was conducted by introducing Additive White Gaussian Noise (AWGN) at a level of 40 dB SPL. Fig. 5 shows the resulting power spectrum and global masking threshold for the DJI Phantom 3 at a distance of 100 meters under these conditions. Compared to the previous case, the masking threshold is significantly elevated across the entire spectrum, particularly in the mid and high-frequency bands due to the effect of noise.

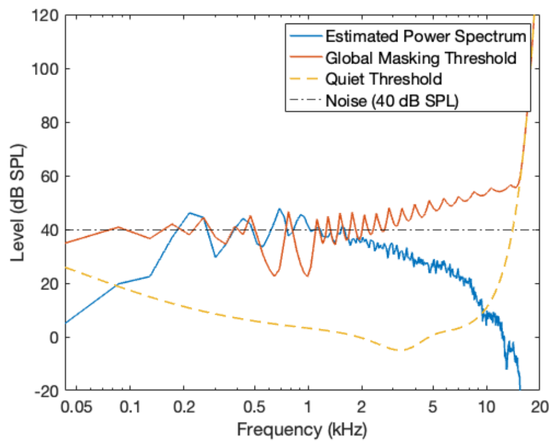


Figure 5: Estimated power spectrum and masking threshold at 100 meters with 40 dB SPL noise.

In this scenario, the estimated power spectrum of the drone signal falls below the masking threshold in most frequency regions, including parts of the low-frequency range. This is mainly due to the combined effects of attenuation and background noise, which increases the masking contributions and reduces the SMR. While a few low-frequency components—particularly some between approximately 100-1000 Hz—may remain marginally audible, their perceptibility is highly uncertain and depends on the local contribution of noise and the local masking level.

Fig. 6 shows the average Signal to Mask Ratio computed for the first 13 Bark bands, plotted at the center frequency of each band. This analysis was performed to assess how far the drone signal remains perceptible under varying propagation conditions. Two horizontal reference lines are included at 0 dB and 3 dB SMR. While an SMR above 0 dB indicates that the signal's spectral energy surpasses the global masking threshold, suggesting potential audibility, a margin of at least 3 dB is commonly adopted to ensure robust perceptual detection. The figure includes SMR curves obtained for multiple distances ranging from 1 meter to 500 meters for the DJI Phantom 3. The analysis is limited to the first 13 Bark bands (approximately up to 2 kHz), since previous results have shown that signal components beyond this frequency range are completely masked and thus perceptually irrelevant.

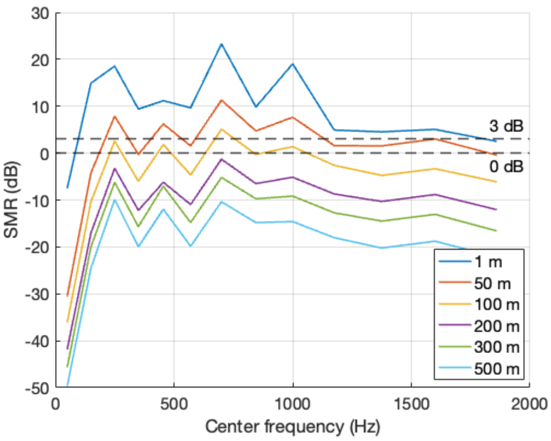


Figure 6: SMR degradation with distance across the first 13 Bark bands.

As illustrated, the drone signal is clearly audible at 1 meter, with all SMR values in the analyzed frequency range exceeding the 3 dB threshold. At 50 meters, the SMR remains above 0 dB in all bark bands, and although several bands still exceed the 3 dB margin, the perceptibility of the signal begins to decrease. At 100 meters, only a few low-frequency bands remain above the 0 dB threshold, while the majority of the spectrum falls below it, indicating that the signal is becoming increasingly masked. Beyond this distance, the SMR continues to decrease, confirming that drone audibility deteriorates rapidly due to the combined effects of frequency-dependent attenuation and ambient noise. Consequently, auditory detection by a human observer becomes extremely limited, or even entirely unfeasible, at distances exceeding 100 meters under typical outdoor conditions.

While the SMR-based analysis illustrates the limitations of human hearing, the following evaluation examines the capabilities of an automatic system under the same acoustic conditions. To compare the human auditory perception with a state-of-the-art acoustic detection system, the YAMNet neural network was evaluated under the same conditions used in the perceptual analysis, that is, with an ambient noise level of 40 dB SPL and propagation-induced attenuation applied according to outdoor transmission models. This deep learning architecture, specifically designed for audio event classification, was previously analyzed by the authors in [8], and the results are reused here to enable a direct comparison between human and machine-based detection performance.

Fig. 7 shows the Receiver Operating Characteristic (ROC) curves obtained for distances ranging from 1 to 500 meters. These curves represent the probability of detection (P_D) vs. the probability of false alarm (P_{FA}), serving as a standard performance metric in detection theory. In general, a larger area under the ROC curve indicates a more reliable detector. As shown, YAMNet maintains high detection performance even at 500 meters, significantly outperforming human auditory capabilities under the same acoustic conditions. These findings reinforce the importance of automated systems for reliable UAV detection over long distances.

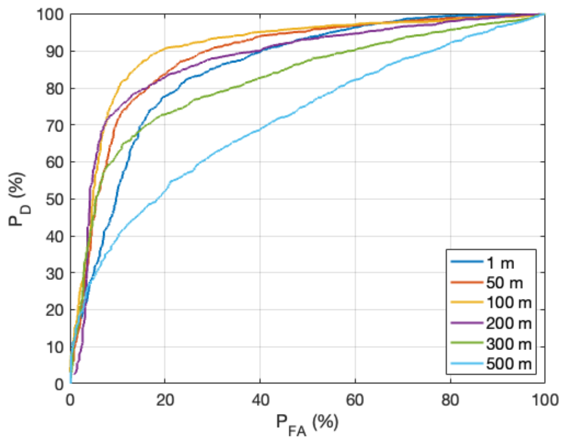


Figure 7: ROC curve for the YAMNet-based detector, at different distances [8].

As demonstrated in the previous analysis, human auditory perception of drone noise significantly deteriorates beyond 100 meters, mainly due to propagation losses and ambient noise. In contrast, state-of-the-art acoustic systems, such as the YAMNet-based model evaluated, can reliably detect drones at distances up to 500 meters. This highlights the critical importance of automated detection technologies, especially in surveillance and security applications, where relying on human hearing alone would severely limit detection capabilities and operational effectiveness. Furthermore, in line with findings from the literature on psychoacoustic sound quality metrics, it can be inferred that in order to avoid any perceptual disturbance to the population, drones should ideally operate at distances of at least 100 meters from inhabited areas, ensuring that their acoustic signature remains below the threshold of audibility and therefore does not cause annoyance.

4. CONCLUSION

In this work, the audibility of UAV noise was evaluated using the MPEG Psychoacoustic Model 1, with the objective of estimating the maximum distance at which a drone can be perceived by a human listener. Under ideal conditions, the acoustic signature of the drone remains clearly audible up to several hundred meters. However, when more realistic scenarios are considered, including propagation-induced attenuation and a moderate ambient noise level of 40 dB SPL, audibility declines significantly, becoming nearly null beyond 100 meters.

The analysis of the global masking threshold and the power spectrum revealed that the perceptible content of the drone signal is mainly concentrated in low-frequency components, particularly below 1 kHz. By calculating the Signal-to-Mask Ratio (SMR) across Bark bands and examining its variation with distance, it was observed that beyond 100 meters, the signal becomes largely masked by environmental conditions, resulting imperceptible to the human auditory system. This threshold naturally depends on ambient noise levels, and more adverse conditions would further reduce detectability.

In contrast, an automatic detection system based on the YAMNet neural network demonstrated robust performance under the same acoustic conditions, achieving reliable drone detection at distances up to 500 meters. These results highlight both the limitations of human auditory perception and the practical viability of acoustic detection systems for surveillance, security, and environmental monitoring applications, even in scenarios characterized by low signal-to-noise ratios.

Finally, the findings are consistent with prior research on psychoacoustic annoyance and support the recommendation that small UAVs should operate at least 100 meters away from populated areas under these conditions. At this range, the acoustic signature remains below the threshold of audibility, minimizing both detectability and potential disturbance.

5. ACKNOWLEDGMENTS

This work was funded by the Spanish Ministry of Science and Innovation under Project PID2021-129043OB-I00 (funded by MCIN/AEI/10.13039/501100011033/FEDER, EU).

6. REFERENCES

[1] K. Brandenburg and G. Stoll, "Iso/mpeg-1 audio: A generic standard for coding of high-quality digital audio," *Journal of the Audio Engineering Society*, vol. 42, no. 10, pp. 780–792, 1994.

[2] C. J. Hui, M. J. Kingan, Y. Hioka, G. Schmid, G. Dodd, K. N. Dirks, S. Edlin, S. Mascarenhas, and Y.-M. Shim, "Quantification of the psychoacoustic effect of noise from small unmanned aerial vehicles," *International Journal of Environmental Research and Public Health*, vol. 18, no. 17, p. 8893, 2021.

[3] N. Sieben, R. Aalmoes, and V. Vos, "A deeper understanding in the psychoacoustic characterization of drone noise," in *Forum Acusticum*, 2023.

[4] D. Y. Gwak, D. Han, and S. Lee, "Sound quality factors influencing annoyance from hovering uav," *Journal of sound and vibration*, vol. 489, p. 115651, 2020.

[5] J. Yousaf, H. Zia, M. Alhalabi, M. Yaghi, T. Basmaji, E. A. Shehhi, A. Gad, M. Alkhedher, and M. Ghazal, "Drone and controller detection and localization: Trends and challenges," *Applied Sciences*, vol. 12, no. 24, p. 12612, 2022.

[6] M. H. Rahman, M. A. S. Sejan, M. A. Aziz, R. Tabassum, J.-I. Baik, and H.-K. Song, "A comprehensive survey of unmanned aerial vehicles detection and classification using machine learning approach: Challenges, solutions, and future directions," *Remote Sensing*, vol. 16, no. 5, p. 879, 2024.

[7] V. Semenyuk, I. Kurmashev, A. Lupidi, D. Alyoshin, L. Kurmasheva, and A. Cantelli-Forti, "Advance and refinement: The evolution of uav detection and classification technologies," *arXiv preprint arXiv:2409.05985*, 2024.

[8] D. Tejera-Berengue, F. Zhu-Zhou, M. Utrilla-Manso, R. Gil-Pita, and M. Rosa-Zurera, "Analysis of distance and environmental impact on uav acoustic detection," *Electronics*, vol. 13, no. 3, p. 643, 2024.

[9] Y. Lin, W. H. Abdulla, *et al.*, "Audio watermark," *Audio Watermark A Comprehensive Foundation Using MATLAB*, 2015.

[10] R. Arelhi and D. Campbell, "A matlab simulink implementation of speech masking based on the mpeg psychoacoustic model 1," in *3rd International Symp. on Image and Signal Processing and Analysis, 2003. ISPA 2003*, vol. 1, pp. 543–548, IEEE, 2003.

[11] E. Zwicker and H. Fastl, *Psychoacoustics: Facts and models*, vol. 22. Springer Science & Business Media, 2013.

[12] D. Pan, "A tutorial on mpeg/audio compression," *IEEE multimedia*, vol. 2, no. 2, pp. 60–74, 1995.

[13] I. O. for Standardization/International Electrotechnical Commission *et al.*, "Information technology - coding of moving pictures and associated audio for digital storage media at up to about 1.5 mbit/s: Part 3: Audio," *ISO/IEC 11172-3*, 1993.

[14] D. Y. Gwak, D. Han, and S. Lee, "Sound quality factors influencing annoyance from hovering uav," *Journal of Sound and Vibration*, vol. 489, p. 115651, 2020.

[15] A. J. Torija, R. H. Self, and J. L. Lawrence, "Psychoacoustic characterisation of a small fixed-pitch quadcopter," in *Inter-Noise and Noise-Con Congress and Conference Proc.*, vol. 259, pp. 1884–1894, Institute of Noise Control Engineering, 2019.

[16] I. O. for Standardization, "Acoustics: Attenuation of sound during propagation outdoors," *ISO Standard 9613-1*, 1993.

[17] I. O. for Standardization, "Acoustics – attenuation of sound during propagation outdoors: Part 2: General method of calculation," *ISO Standard 9613-2*, 1996.