applied sciences

MDPI

Article

On-Board Outdoor Acoustic Measurement of Multirotor Unmanned Air Vehicles: Effect of Flight Conditions on Emission and Perception

Nicola Gravina [1,*], Massimiliano Masullo [1,*], Luigi Maffei [1] and Juan Miguel Navarro [2]

[1] Department of Architecture and Industrial Design, Università degli Studi della Campania "Luigi Vanvitelli", 81031 Aversa, Italy; luigi.maffei@unicampania.it
[2] Escuela Politécnica Superior, Universidad Católica San Antonio de Murcia, 30107 Murcia, Spain; jmnavarro@ucam.edu
* Correspondence: nicola.gravina@unicampania.it (N.G.); massimiliano.masullo@unicampania.it (M.M.)

Featured Application

The in situ psychoacoustic (sound quality metric) characterisation of commercial multirotor Unmanned Air Vehicles' noise emission, to support noise-annoyance assessment and the acoustic acceptability of Urban Air Mobility operations.

Abstract

Urban Air Mobility development depends on overcoming several challenges. Noise emission is one of the most critical, making Unmanned Air Vehicles' acoustic characterisation a scientific and regulatory priority. Through an on-board outdoor measurement campaign on three commercial multirotor drones (DJI Mavic 3M, Mavic 3 Classic, and DJI Matrice 4T) during hovering, this study analyses and shows the effects that the most important flight variables: altitude, propeller design and payload, have on the overall noise emission, its directivity and the psychoacoustic annoyance, more specifically through the descriptor developed by Boucher. The results showed that different altitudes (10, 15 and 20 m) do not affect noise emissions. Moreover, overall sound pressure levels may fail to discriminate drones' noise annoyance. In fact, despite the lower sound pressure levels, for the specific platform and propeller configuration tested (DJI Mavic 3M with manufacturer low-noise blades), the low-noise-propeller configuration showed higher psychoacoustic annoyance than the standard-propeller one, driven primarily by higher roughness (and, to a lesser extent, tonality). Additional payloads (50–150 g) did not increase psychoacoustic annoyance monotonically. The heaviest Matrice 4T showed the most stable psychoacoustic profile, with psychoacoustic annoyance remaining below its unloaded baseline at all payload levels, while the lightest Mavic 3 Classic showed the largest swing. Moreover, considering the drone's structural symmetry, the angular asymmetries revealed by the directivity analysis suggest hovering-stabilisation dynamics attributions. Overall sound pressure levels alone proved an unreliable predictor of drones' noise annoyance across all tested flight conditions.

Keywords: urban air mobility; on-board recordings; payload; propeller design; psychoacoustic annoyance; sound quality metrics

Check for updates

Academic Editor: Suchao Xie

Received: 31 July 2026
Revised: 19 September 2026
Accepted: 22 September 2026
Published: 25 September 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

The use of drones has expanded rapidly across environmental monitoring, logistics, surveillance, and urban service sectors. In Europe, the Urban Air Mobility (UAM) market

is projected to reach approximately €4.2 billion by 2030, representing nearly one-third of global market share [1]. Achieving this potential, however, requires public acceptance of drone operations in anthropized environments—a challenge in which acoustic impact plays a central role.

Drone noise-induced annoyance is emerging as one of the primary barriers to large-scale UAM development [2]. In this context, acoustic acceptability should be regarded as an integral dimension of the broader transition towards sustainable mobility, alongside reductions in carbon emissions and improvements in energy efficiency [3].

Communities consistently perceive Unmanned Air Vehicle (UAV) noise as more annoying than conventional transportation sources at equivalent sound pressure levels [4], and several studies report that drones are perceived as considerably more annoying than commercial aircraft at equal A-weighted levels [5]. The distinctive character of drone noise—high-frequency broadband content, strong tonal components at blade pass frequencies (BPF) and harmonics, and amplitude modulation—has been identified as the principal driver of this elevated annoyance response [6].

These considerations have prompted growing scientific and regulatory attention toward structured acoustic characterisation of UAVs. In 2022, the European Union Aviation Safety Agency (EASA) published the first worldwide harmonised guidelines for the noise measurement of unmanned aircraft systems below 600 kg [7], covering multicopter configurations and providing voluntary but standardised field and hovering measurement procedures. Building comparable, reproducible acoustic datasets across drone platforms is therefore both a scientific and a regulatory priority.

Despite this emerging interest, on-board campaigns that jointly characterise how multiple flight-condition variables (altitude, propeller design, and payload) affect both the acoustic emission and its directivity within a single, coherent dataset remain scarce: the existing studies have typically examined these factors in isolation, or under controlled laboratory conditions that do not reproduce the real flight dynamics just described. This study intends to address this gap through an on-board outdoor acoustic measurement campaign designed to jointly characterise these effects on both the acoustic and psychoacoustic emission of multirotor UAVs.

Although realistic flight operations can be replicated in anechoic chambers to obtain highly accurate measurements, such controlled environments cannot reproduce the conditions that shape a drone's acoustic emission in the field. Environmental disturbances continuously act on the airframe, and the flight controller compensates for them in real time by adjusting rotor speed and thrust distribution with direct consequences on the emitted acoustic signature. The availability and size of anechoic facilities are, in any case, limiting factors, and several field measurement techniques have therefore been proposed to capture drone noise under these real operating conditions. The main approaches rely on ground-based measurements [8,9] using microphone arrays mounted on a reflective plate, from which near-drone measurements can be inferred through back-propagation processes. However, field measurements must contend with environmental conditions and background noise, making it challenging to obtain objective and repeatable results.

On-board, embedded acoustic measurement systems directly address this need: by recording at the source while the drone is airborne in an open field, they capture the noise it actually emits as it responds, moment by moment, to wind and the other environmental factors that affect its flight stability, rather than the noise it would emit under idealised, undisturbed conditions—while also removing the need to invert propagation effects and minimising ambient interference.

Recent studies have highlighted that the operating mode of multi-rotor UAVs (hover vs. cruise) significantly affects their overall sound pressure level (OSPL). Measure-

ments performed in controlled and semi-controlled environments indicate that differences in propeller type and rotor configuration can lead to variations of several decibels between hovering and forward flight, attributable to unsteady aerodynamic interactions between rotor wakes and the drone structure rather than solely rotor speed [10]. This emphasises that realistic operational measurements are essential to capture the full complexity of drone noise.

The aeroacoustic performance of drone propellers is a central topic in UAV noise research. The noise generated by rotating propellers combines tonal components (at BPF and harmonics, linked to periodic aerodynamic loading) and broadband noise (due to turbulence ingestion and trailing-edge scattering). Low-noise propeller designs, incorporating features such as swept-back tips, increased blade count, optimised chord distribution, and bio-inspired serrations, have been shown to achieve significant noise reductions, primarily in frequency bands above 1 kHz, without necessarily compromising aerodynamic efficiency [11]. Integrated evaluation methodologies combining aeroacoustics and psychoacoustic analysis confirm that modifications reducing tonal content can disproportionately improve perceived annoyance, even for modest reductions in overall sound pressure level [12].

Regarding the relationship between payload and acoustic emission, field studies have confirmed a strong positive correlation between UAV take-off weight (TOW) and the magnitude of the acoustic emission at constant flight speed [13]: drones are noisier the heavier their payload or the faster they fly. In parallel, research on acoustic-based payload estimation has demonstrated that changes in motor and blade speed induced by different payload weights produce measurable shifts in the acoustic emission profile, with even simple pitch-based metrics achieving reliable discrimination between loaded and unloaded conditions [14].

When a drone carries additional payload, its flight controller should compensate for the increased total weight by commanding higher rotor speeds, which in turn should raise both the BPF and the overall aerodynamic loading noise. It is expected that acoustic changes should tend to upward frequency shifts in tonal peaks and increase OSPL by modifying the spectral balance between tonal and broadband components. At the psychoacoustic level, increased rotor speed affects the perception of several acoustic characteristics: loudness increases with the overall level, sharpness may increase due to the spectral centroid shift, and tonality may vary, depending on whether the new BPF falls in a perceptually salient region.

Field studies on UAV noise directivity have shown that both tonal components associated with BPFs and broadband noise due to aerodynamic turbulence contribute to complex angular emission patterns, which are sensitive to orientation, payload, and operating conditions [13]. These results underline the importance of on-board and in situ measurements, particularly when assessing urban noise impact.

Because propeller design and payload alter both the tonal and spectral content of drone noise, as discussed above, their effects can be expected to extend beyond overall sound pressure level to influence perceived annoyance itself. The assessment of drone noise annoyance increasingly relies on psychoacoustic metrics, which model the subjective perception of sound beyond simple energy-based measures. Parameters such as loudness, sharpness, tonality, roughness, and fluctuation strength have all been identified as relevant predictors of annoyance to UAV noise [15]. Laboratory listening experiments have shown that tonality, sharpness, and loudness contribute most significantly to annoyance models for drone noise [16,17]. The prediction of perceived annoyance (PA) from psychoacoustic metrics, often computed as a composite sound quality metric (SQM), offers a more perceptually relevant characterisation than SPL alone and is increasingly used in comparative studies across drone fleets [18]. Zwicker's general PA model was calibrated on synthesised narrow-

band and broadband stimuli and does not explicitly capture the tonal character prominent in UAV noise. To address this gap, Boucher et al. [19] developed a PA-tailored model for Urban Air Mobility vehicle noise: its weighting coefficients and activation thresholds were derived by fitting perceived-annoyance ratings from human listeners in the Test of UAM Sound Quality (TUSQ), a listening experiment based on 141 stimuli (136 UAM + 5 reference sound conditions).

An interdisciplinary approach integrating pure technical acoustics and human perception is emerging as the most comprehensive path toward understanding the noise effects of UAVs [2], and the availability of well-characterised acoustic datasets is a prerequisite for this line of research.

Building on this interdisciplinary perspective, this study concentrates on a single, representative platform category: multirotor UAVs in the sub-2 kg, middle-sized class. Under the EASA Open Category referenced above [7], platforms in this weight range fall within the lightest-touch classes (C0–C2, up to 4 kg), which face substantially fewer certification, licensing, and operational restrictions than heavier UAS, making them the platforms most frequently deployed for everyday urban and peri-urban aerial services, and therefore the most representative of the drone noise the public is actually exposed to in these environments. Within this category, three commercial models covering distinct use cases (precision agriculture/mapping, consumer/prosumer imaging, and public-safety operations) were selected, allowing the effects of propeller design, payload, and directivity to be examined consistently across a single, operationally relevant drone class. Because even platforms within the same weight class can differ substantially in airframe design, propulsion architecture, and propeller integration, comparing multiple commercial models under identical conditions provides a direct empirical basis for assessing how much acoustic and psychoacoustic emission varies within the class.

The main contribution of this study lies in the integrated experimental assessment of how relevant flight and configuration variables—altitude, propeller design, and payload—affect not only the acoustic emission of commercial multirotor UAVs but also their directivity and psychoacoustic annoyance. The on-board measurement approach enables these effects to be investigated under actual outdoor hovering conditions and across different UAV configurations. Analysing the drones' noise emission at the source in situ, the present work investigates the following research questions:

RQ1. *How much do acoustic and psychoacoustic metrics differ among middle-sized class drone models as flight altitude, propeller types and payload weight change?*

RQ2. *How much does the signature of drones' noise emission differ among middle-sized class drone models as flight altitude, propeller types and payload weight change?*

RQ3. *Which angular patterns characterise the noise emission of the selected drone models?*

RQ4. *How much does UAM Psychoacoustics Annoyance model of drone noise emissions differ as the flight altitude, propeller types and payload weight change?*

2. Methodology

On-board measurements were carried out through two different measurement setups. The simplest measurement setup consists of positioning a single microphone under the drone's body at a distance sufficient to be considered just beyond the far-field condition ($kr > 1$) relative to the Blade Pass Frequency (BPF). While the $kr > 1$ criterion is satisfied at the BPF, it should be noted that a multirotor UAV is a spatially distributed source; the potential influence of rotor–airframe interactions and near-field contributions on the measured spectra is discussed in detail in Section 4, where their implications for the psychoacoustic descriptors are explicitly addressed.

This setup allows the acquisition of the noise emitted in real flight conditions by the drone's on-board noise sources, minimising as much as possible the ambient noise. The second measurement setup consists of an arc-shaped support structure with microphones positioned at multiple angular positions to provide data on the drone's emission pattern in both the sagittal and frontal planes.

Preliminary measurement sessions were conducted in both configurations to assess the sensitivity of the acoustic and psychoacoustic descriptors to flight altitude. Since the microphone was mounted 50 cm below the drone's fuselage, the source-receiver distance remained fixed across the tested altitudes; any difference observed across altitudes can be attributed to the change in height above ground. On this basis, a representative reference height of 10 m was selected for the subsequent measurement campaign.

Sessions took place outdoors on the campus of the Universidad Católica San Antonio de Murcia, Spain (UCAM), in an area free from buildings or significant obstacles near the flight zone, with the ground consisting mainly of soft, absorptive terrain (G = 1, for the CNOSSOS-EU ground classification [20]).

The measurement layout (Figure 1), in terms of flight altitude and distance from reflective surfaces, was defined in accordance with the EASA methodological recommendations for unmanned aerial vehicle noise measurement [7], to ensure comparable and reproducible experimental conditions.



Figure 1. In situ single-mic recording configuration.

For each drone and condition, hovering recordings of at least 8 min were obtained to ensure repeatability and to identify anomalous acquisitions caused by external factors (i.e., wind gusts, transient ambient noise). From each recording, a continuous 3 min segment of undisturbed hovering was selected for psychoacoustic analysis.

This yields an audio dataset primarily representative of drone noise emission, minimising external influences and enabling high-quality psychoacoustic analysis at the source.

During each session, key environmental parameters (i.e., air temperature, relative humidity, wind speed, and background noise levels) were monitored to verify operational stability during acquisitions.

2.1. Drone Models and Measurement Setup

Three commercial multirotor drones (DJI Co., Ltd., Shenzhen, China) were selected for the measurement campaign, representing a range of size classes, rotor configurations, and intended applications within the medium-sized UAV category. The selection covers three distinct use-case categories recognised in the UAM ecosystem: precision agriculture and mapping (DJI Mavic 3M), consumer/prosumer aerial imaging (DJI Mavic 3 Classic), and public-safety/emergency-response operations (DJI Matrice 4T). All platforms are quadrotor configurations with foldable arm designs, operating in the sub-2 kg class, which corresponds to the most commercially prevalent category in urban and peri-urban environments. Their main characteristics are summarised in Table 1.

Table 1. Main characteristics of the drone models tested in this campaign.

Drones [-]	Propellers [-]	Weights [g]	Rotor Config. [-]	Airdrop Weight [g]	Baseline Weight [g]
Dji Mavic 3m	Standard Low noise	951 *	Quadrotor	123	1114
Dji Mavic 3 Classic	Standard	895 *	Quadrotor	150	1085
Dji Matrice 4t	Standard	1219 *	Quadrotor	50	1309

* Standard aircraft weight (including battery, propellers, and a microSD card). The actual product weight may vary due to differences in materials across production batches and external factors. (±5–20 g).

The primary acoustic mic is a Rode Wireless Go system (RØDE Microphones (Freedman Electronics Pty Ltd.), Sydney, Australia), selected for its light weight (Table 2) and compact dimensions, making it suitable for on-board drone installation. The system employs a miniaturised omnidirectional condenser microphone capsule, with digital wireless transmission via the proprietary RØDE Series IV protocol (128-bit AES encryption, 2.4 GHz band). The received signal is recorded in real time using Audacity 3.4.7 audio software at a 48 kHz sampling rate and 24-bit resolution, ensuring a sufficient bandwidth and dynamic range for subsequent acoustic and psychoacoustic processing.

Table 2. Measurement system components and mass for the two recording configurations.

Setup Types	Components	Weights [g]
Single mic	Rode Go Attachment system	32 8
Array mic	Rode Go 4 × Lavalier Go II 4 × Rode Pro	32 4 × 12 4 × 36
	Attachment system	16

Microphone signal calibration was conducted at the recording studio of the Universidad Católica San Antonio de Murcia, the reference instrument used for calibration was a Rion NL-05 Class 1 integrating sound level meter [21]. Calibration was performed by exposing both the microphone capsule and the reference SLM simultaneously to a broadband white noise source at discrete SPL steps of 5 dB from 60 to 90 dB SPL. The correction factors derived from this comparison and applied in Artemis Suite were +3.5 dB for the Rode Wireless Go (Gen 1)internal capsule used in the single-microphone configuration, and +10.5 dB for the Lavalier GO II external capsule connected to the Rode Wireless GO II (Gen 3) systems used in the arc array. Both systems feature a declared frequency range of 50 Hz–20 kHz and an equivalent input noise of approximately 22 dB(A). This calibration procedure was repeated periodically to verify system stability over time.

In the primary measurement configuration (Figure 2a), the microphone (Rode Go) was enclosed in a windscreen and mounted at the end of a 0.50 m arm attached to the drone's airdrop system. The attachment system consisted of a vertical element made of 1.6 mm iron wire, connected to a cord-based suspension that held the microphone in place. The arm is composed of elements with varying stiffness (partially flexible).

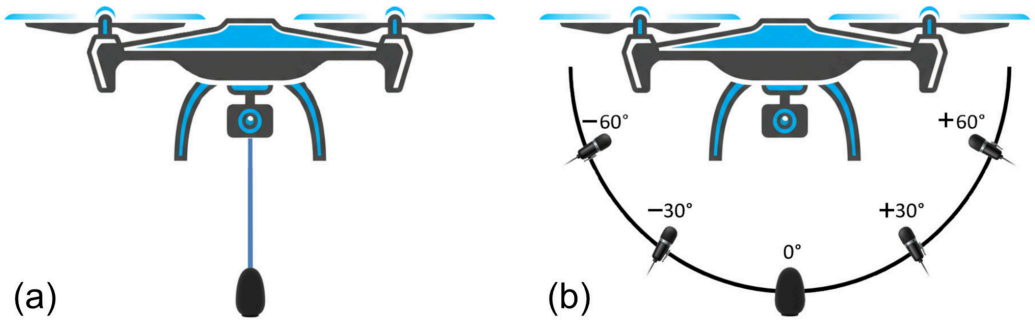


Figure 2. Single-microphone configuration on the left (**a**) and arc-shaped multi-microphone configuration on the right (**b**).

As an evolution of the single-microphone setup, the second configuration employs an arc-shaped support structure (Figure 2b), built from 1.6 mm iron wire and stabilised by cord-based tensioners, in addition to the single-microphone attachment described above. Four additional Lavalier microphones (Lavalier Go II) are mounted on this structure at 30° and 60° angular positions relative to the vertical axis. The Lavalier microphones are connected to two Rode Wireless GO II (Gen3) systems, providing five audio tracks (0°, ±30°, ±60°), corresponding to the main directions of the drone's acoustic field. The arc can be rotated, enabling the capture of directivity in both the sagittal plane (front-back) and the frontal plane (left-right) of the drone. These angular positions allow the acquisition of off-axis acoustic signals, where aerodynamic contributions from the rotors and rotor–airframe interaction effects produce significant variations in the angular distribution of the sound field.

To simulate different payload conditions, calibrated laboratory weights were attached to the drone's airdrop system in the single-microphone configuration (Figure 2a). Three individual weight units were used—50 g, 60 g, and two units of 20 g—combined to obtain three incremental payload levels relative to the baseline weight reported in Table 1: +50 g (single 50 g unit), +100 g (60 g unit combined with the two 20 g units), and +150 g (all four units combined). This approach allowed the acoustic effect of incremental, precisely known payload increases to be isolated while keeping the attachment point and system configuration otherwise unchanged.

2.2. Psychoacoustic Descriptors and Noise Annoyance

The drone's emission noise annoyance was evaluated by using a more specific version of the Zwicker psychoacoustic annoyance index [22] based on the UAM-specific model developed by Boucher et al. [19]. This version includes the following descriptors extracted from the recordings using the software Artemis Suite 14 [23]:

- Loudness (N, soneGF), following ISO 532-1, representing the overall loudness level of stationary sound. N_5 is the 5th percentile of the time-varying loudness signal [24].
- Sharpness (S, acum), following DIN 45692, quantifying the prevalence of high-frequency content [25].
- Tonality (T, tuHMS), following ECMA-74 (17th) [26]/ECMA-418-2 (2nd) [27], measuring the prominence of tonal components relative to the masking noise, computed following the hearing model (HMS) method.

- Roughness (R, asper), following ECMA-418-2 (2nd) [27], capturing fast amplitude modulations (20–300 Hz range), associated with harshness and buzzing quality.
- Fluctuation Strength (FS, vacil) (1/1 bark), sensitive to slower modulations (<20 Hz), linked to pulsating character [28].

As specified in the following formula:

$$PA = N_5\left(1 + \sqrt{w_S^2 + w_{FR}^2 + w_T^2}\right) \quad (1)$$

where $w_S = (S - 1.75) \cdot 0.25 \cdot log_{10}(N_5 + 10)$ for S > 1.75 acum (zero otherwise) is a sharpness-dependent weight, following Zwicker's model [22], $w_{FR} = \frac{2.18}{N_5^{0.4}} \cdot (0.4 \cdot FS + 0.6R)$ is the roughness-fluctuation term (following Zwicker's structure), and $w_T = 3.2 \cdot \frac{T}{N_5}$ is the UAM-specific tonality term, with T in tuHMS and N_5 in soneGF.

Throughout this study, the value computed via Equation (1) from on-board recordings is referred to as the on-board Psychoacoustic Annoyance Index (PAI) to distinguish it from the absolute annoyance values perceived by a ground-based receiver.

The sharpness values measured in this study ranged between 1.74 and 2.16 acum, placing most tested conditions at or just above the 1.75 acum activation threshold; w_S is therefore computed directly from the measured S of each condition, as specified by the model. The practical implications of this near-threshold positioning are discussed in the Section 4.

In addition, because hovering flight is not perfectly stationary over the duration of a recording, as it is subject to both environmental effects and the flight controller's own stabilisation corrections, a single time-averaged value may not fully capture how each descriptor varied during a given recording. For this reason, additionally, the 5th and 95th percentile values of sharpness, roughness, tonality, and fluctuation strength were extracted in Artemis Suite for every recorded condition. Since these descriptors determine the weight terms w_S, w_{FR}, and w_T in the PA formula, their percentile range translates directly into a plausible range for PA itself. N_5 is unaffected by this step, as it is already defined as a 5th-percentile value. These ranges are used in discussion and limitations paragraphs to compare the inter-configuration PA differences.

3. In Situ Measurements

3.1. Altitude Effect

The choice of measurement altitude directly affects the signal-to-noise ratio of on-board recordings and the relationship between measured levels and perceived emission at community-relevant distances. A minimum flight height is required to satisfy the far-field condition and to minimise ground reflection contributions; conversely, excessively high altitudes may introduce larger wind-induced variability. The EASA guidelines [7] provide reference geometrical configurations for field measurements, which informed the selection of the altitude range tested here. Table 3 reports the key acoustic and psychoacoustic descriptors for all drone configurations (no added payload, single microphone).

The data confirm that the variation in overall sound pressure levels and loudness across the three altitudes remains within approximately 3–4 soneGF for all models, with no monotonic trend consistently observed across all drones. This limited sensitivity to altitude in the 10–20 m range supports the selection of 10 m above ground level (AGL) as the reference measurement height, providing a practical balance between signal quality and operational safety margin.

The inter-model comparison highlights substantial differences in psychoacoustic character despite similar overall sound pressure levels. The 3M ST has the highest loudness

(N_5 = 69.9 soneGF) and the lowest roughness (R = 0.320 asper) among all tested drones, with a relatively low tonality (T = 0.248 tuHMS). The 3M LN presents a distinct profile: lower loudness (N_5 = 64.5 soneGF) but markedly elevated roughness (R = 1.090 asper) and tonality (T = 0.358 tuHMS). The 3 Classic combines the lowest loudness (N_5 = 64.0 soneGF) with the lowest tonality (T = 0.192 tuHMS) in the entire dataset, together with a low roughness (R = 0.396 asper). The Matrice 4T occupies an intermediate position: sharpness is the lowest among all models (S = 1.75 acum), while roughness (R = 0.855 asper) and tonality (T = 0.298 tuHMS) are substantially above those of the 3M ST and 3 Classic.

Table 3. Acoustic and psychoacoustic descriptors by flight altitude (10, 15, and 20 m above ground level), no added payload, single-microphone configuration.

Drone	Height [m]	OSPL [dB]	$OSPL_A$ [dBA]	N_5 [soneGF]
3M ST	10	84.7	84.0	69.9
	15	86.5	84.8	73.6
	20	85.6	84.6	73.5
3M LN	10	83.9	82.4	64.5
	15	83.7	83.0	66.0
	20	83.9	83.2	63.0
3 Classic	10	83.3	82.8	64.0
	15	83.1	82.7	62.3
	20	83.2	82.7	63.9
Matrice 4T	10	84.1	82.6	65.0
	15	83.4	82.1	64.6
	20	83.2	81.6	63.5

3.2. Influence of the Propeller Type

Propeller design is one of the primary causes of drone noise character. Standard propellers optimise aerodynamic efficiency (thrust-to-power ratio) at the expense of acoustic output, generating strong tonal peaks at the Blade Pass Frequency (BPF = RPM/60 × N~blades~) and its harmonics, alongside broadband turbulence noise. Low-noise propeller designs address this trade-off through various strategies: swept-back and tapered blade tips to reduce tip vortex noise, increased blade count to distribute aerodynamic loading, optimised pitch distribution, and bio-inspired trailing-edge geometries [11]. Studies of low-noise UAV propellers have demonstrated significant acoustic improvements—particularly in frequency bands above 1 kHz—while maintaining adequate thrust performance [11]. At the psychoacoustic level, modifications that reduce tonal content tend to lower both the tonality and sharpness components of perceived annoyance, which are known significant predictors of UAV noise annoyance [16]. The DJI Mavic 3M was the only model in this campaign equipped with two propeller variants: the manufacturer's standard blades and the DJI low-noise propeller option. All other experimental conditions were kept identical (hovering at 10 m AGL, single-microphone configuration, no additional payload). Table 4 reports the overall sound pressure levels (OSPL, $OSPL_A$) and loudness (N_5) for the 3M with standard propellers (ST) and 3M with low-noise propellers (LN) at the reference altitude of 10 m.

Despite producing nearly identical broadband SPL (Δ = +0.8 dB) and a modest difference in A-weighted level (Δ = +1.6 dBA), the two propeller configurations differ markedly in their psychoacoustic profiles. The LN blades yield lower loudness (N_5: 64.5 vs. 69.9 soneGF, Δ = 5.4 soneGF) but substantially higher roughness (1.090 vs. 0.320 asper) and tonality (0.358 vs. 0.248 tuHMS).

Table 4. Overall sound pressure levels (OSPL, $OSPL_A$) and loudness (N_5) for the DJI Mavic 3M with standard propellers (ST) and low-noise propellers (LN) at the reference altitude of 10 m.

Parameter	3M ST	3M LN	Δ (ST − LN)
OSPL [dB]	84.7	83.9	+0.8
$OSPL_A$ [dBA]	84.0	82.4	+1.6
N_5 [soneGF]	69.9	64.5	+5.4

For this specific platform and propeller configuration (DJI Mavic 3M with manufacturer-designated low-noise blades), this result suggests that the 'low-noise' label refers primarily to overall level reduction and does not necessarily reflect the full psychoacoustic profile.

Whether this finding generalises to other drone platforms or low-noise propeller designs remains an open question that future studies should address by testing multiple blade geometries across different airframes and operating points.

The explicit tonality term w_T in the Boucher et al. UAM PA model [19] proves particularly consequential here, as it captures the elevated tonal character of the LN configuration that would be partially overlooked by a purely loudness- or sharpness-based index.

3.3. Influence of Added Payload Weights

Payload was added incrementally using calibrated weights attached to the drone in a mechanically stable position. The tested levels were +50 g, +100 g, and +150 g relative to the unloaded baseline. All recordings were performed at 10 m AGL in the single-microphone configuration.

Table 5 reports the percentage weight increment for each drone under each payload condition, relative to the take-off weight of the unloaded aircraft. The baseline weight includes the drone airframe, battery, airdrop system, and the on-board acoustic acquisition system.

Table 5. Weight gain, acoustic descriptors and loudness for all drone models under each payload condition.

Drone	Payload [-]	Weight Gain * [%]	OSPL [dB]	$OSPL_A$ [dBA]	N_5 [soneGF]
3M ST	Baseline	17.1	84.7	84.0	69.9
	+50 g	22.4	83.7	82.6	64.4
	+100 g	27.7	84.2	83.1	64.3
	+150 g	32.9	85.7	85.0	71.7
3M LN	Baseline	17.1	83.9	82.4	64.5
	+50 g	22.4	84.5	84.0	65.9
	+100 g	27.7	85.6	84.1	69.9
	+150 g	32.9	86.2	84.8	73.6
3 Classic	Baseline	21.2	83.3	82.8	64.0
	+50 g	26.8	83.6	83.6	64.0
	+100 g	32.4	85.1	85.1	73.0
	+150 g	38	85.0	85.0	69.5
Matrice 4T	Baseline	7.4	84.1	82.6	65.0
	+50 g	11.5	83.3	81.2	59.8
	+100 g	15.6	84.0	82.3	64.7
	+150 g	19.7	84.3	82.9	67.6

* Weight gain [%] = (baseline weight—standard aircraft weight)/standard aircraft weight × 100, where standard aircraft weight is as reported in Table 1. The non-zero baseline value reflects the mass of the on-board acoustic acquisition system alone, with no added payload.

Notably, the three platforms respond very differently to the same absolute payload addition: for the Matrice 4T—which has the highest baseline weight (1309 g)—adding 150 g represents only a +19.7% increase over baseline, whereas the same addition corresponds to a +38.0% increase for the lighter DJI Mavic 3 Classic (baseline 1085 g). This differential loading ratio is consistent with the observed acoustic response: the Matrice 4T exhibits the most stable psychoacoustic profile across payload conditions, while the Mavic 3 Classic and 3M ST show more pronounced variations in roughness and tonality.

Table 5 presents weight gains, acoustic descriptors and loudness for all drone models under baseline and added-payload conditions at 10 m AGL. Baseline is the starting weight of the drone plus the recording system.

For the 3M ST, the OSPL was rather insensitive to payload (84.7, 83.7, 84.2, 85.7 dB), and the psychoacoustic response resulted non-monotonic. The 5th percentile loudness oscillates between 64.3 and 71.7 soneGF across payload conditions (lowest at +100 g, highest at +150 g), while roughness varies non-monotonically (0.320–0.750 asper, peaking at +50 g). This non-monotonic behaviour tracks the $OSPL_A$, which dips by ~1.4 dB(A) at +50 g and rises again at +150 g. The measured blade-passing frequency (BPF) increases with payload (173 → 204 Hz for the 3M ST; Table 6), so its harmonics shift toward the more sensitive region of the equal-loudness contours rather than away from it. The loudness reduction at +50 g is therefore driven by the lower overall level, not by a spectral shift into a less sensitive band.

Table 6. Blade-passing frequency (BPF, Hz) for each drone under the baseline and added-payload conditions during hovering at 10 m AGL.

Drone	Baseline [Hz]	+50 g [Hz]	+100 g [Hz]	+150 g [Hz]
DJI Mavic 3M ST	173	183	193	204
DJI Mavic 3M LN	173	173	183	183
DJI Mavic 3 Classic	173	183	193	204
DJI Matrice 4T	140	140	140	140

For the 3M LN, the payload effect is more regular: loudness increases monotonically with load (N_5: 64.5 → 73.6 soneGF), and roughness declines steadily from its elevated baseline (R: 1.090 → 0.912 asper); tonality, by contrast, rises from its baseline and then plateaus (T: 0.358 → ~0.53 tuHMS). The higher rotor speed demanded by added payload appears to redistribute aerodynamic loading in a way that partially attenuates the modulation components responsible for the elevated roughness of the unloaded condition.

The 3 Classic shows little response at +50 g (N_5: unchanged), the sharpness rises modestly (S: 1.96 → 2.08 acum), the roughness nearly doubles (R: 0.396 to 0.729 asper), then a marked jump at +100 g (N_5: 64.0 → 73.0 soneGF, S: 1.96 → 2.10 acum), followed by partial recovery at +150 g. The large roughness increase at +50 g is notable given the unchanged loudness, suggesting a spectral redistribution rather than a global level increase. These non-linearities are consistent with the relatively high fractional loading imposed on this lightweight platform.

The Matrice 4T presents the most stable psychoacoustic profile across all payload conditions. Loudness varies between 59.8–67.6 soneGF, sharpness remains essentially constant (1.75–1.78 acum), and roughness drops from the baseline (0.855 asper) to a stable value around 0.577–0.608 asper for all payload conditions. This stability is consistent with the comparatively modest fractional load increase on the heavier platform and reflects a robust motor-control response that limits rotor speed variation.

3.4. Changes of Spectral Features

All recorded audio files were analysed using the Artemis Suite software for the extraction of time-frequency representations (spectrograms). Narrowband spectra were computed via 8192-point FFTs on 48 kHz sample rate recordings (see description in Section 2.1), yielding a frequency resolution sufficient to resolve the blade-passing frequency and its low-order harmonics for all tested configurations. During controlled hovering operations, the noise spectra of multirotor drones typically display clear tonal components at deterministic, low-frequency narrowband peaks linked to the propeller rotation rate, with prominent harmonic series extending into the mid-frequency range [13]. The spectral structure depends on the number of blades, the number of rotors, and their relative phasing. The interaction of these BPF harmonic series with the broadband noise floor defines the tonality metric and contributes to roughness through amplitude modulation effects.

The fundamental BPF was extracted for every drone and payload condition (Table 6). Two patterns are relevant to the results. First, for the two Mavic-3-class airframes (3M ST and 3 Classic, which share the same propulsion platform), the BPF rises with the payload from 173 Hz at the baseline to 204 Hz at +150 g, reflecting the higher rotor speed commanded by the flight controller to carry the added weight. Second, the 3M LN shares the same baseline BPF as the 3M ST (173 Hz): the low-noise blades do not reduce loudness by rotating more slowly, so their acoustic benefit is a property of the blade design itself. Their smaller BPF increase under payload (up to 183 Hz vs. 204 Hz for the standard blades) further suggests a higher thrust per revolution, consistent with low-noise blade designs that enlarge the blade area [11]. Finally, the Matrice 4T maintains a constant, lower BPF (140 Hz) across all payload levels: its heavier platform absorbs the added 50–150 g with a negligible change in rotor speed, which explains its uniquely stable psychoacoustic profile under payload.

Figure 3 shows the narrowband spectra (SPL vs. frequency) of the four tested configurations—3M ST, 3M LN, 3 Classic, and Matrice 4T—at the three tested altitudes (10, 15, and 20 m AGL). Across all four configurations, the 10 m curve consistently sits slightly above the 15 m and 20 m curves in the low-frequency rising portion of the spectrum (below ~150–200 Hz), consistent with the shorter distance to the ground and with a possible interaction; all three curves converge closely from the first harmonic peak onward, with matching peak positions at the blade-passing frequency (BPF) and its harmonics, confirming that tonal content is governed by rotor speed rather than by the tested altitude range. The 3 Classic and 3M LN (Figure 3b,c) show the closest overlap between altitudes across the full spectrum, while the 3M ST and, more markedly, the Matrice 4T (Figure 3a,d) exhibit a visible low-frequency offset and, for the Matrice 4T, additional divergence at several mid-spectrum peaks. These spectral observations reinforce the limited altitude sensitivity already quantified in Table 3: the residual differences across altitudes are small enough that they do not bias the between-configuration comparisons presented in the remainder of this study. This confirms the rationale for selecting 10 m as the reference height, already discussed in Section 3.1.

Figure 4 presents the narrowband spectra (SPL vs. frequency) of the four configurations under the baseline and added-payload conditions. All four exhibit the characteristic multirotor structure: a low-frequency tonal region dominated by the blade-passing frequency (BPF) and its harmonic series, superimposed on a broadband floor that decays above roughly 3–4 kHz. The BPF fundamental lies near 170–200 Hz for the three Mavic-3-class airframes (3M ST, 3M LN and 3 Classic) and lower, near 140 Hz, for the heavier Matrice 4T, whose harmonics are correspondingly more closely spaced—in agreement with the measured BPF values (Table 6).

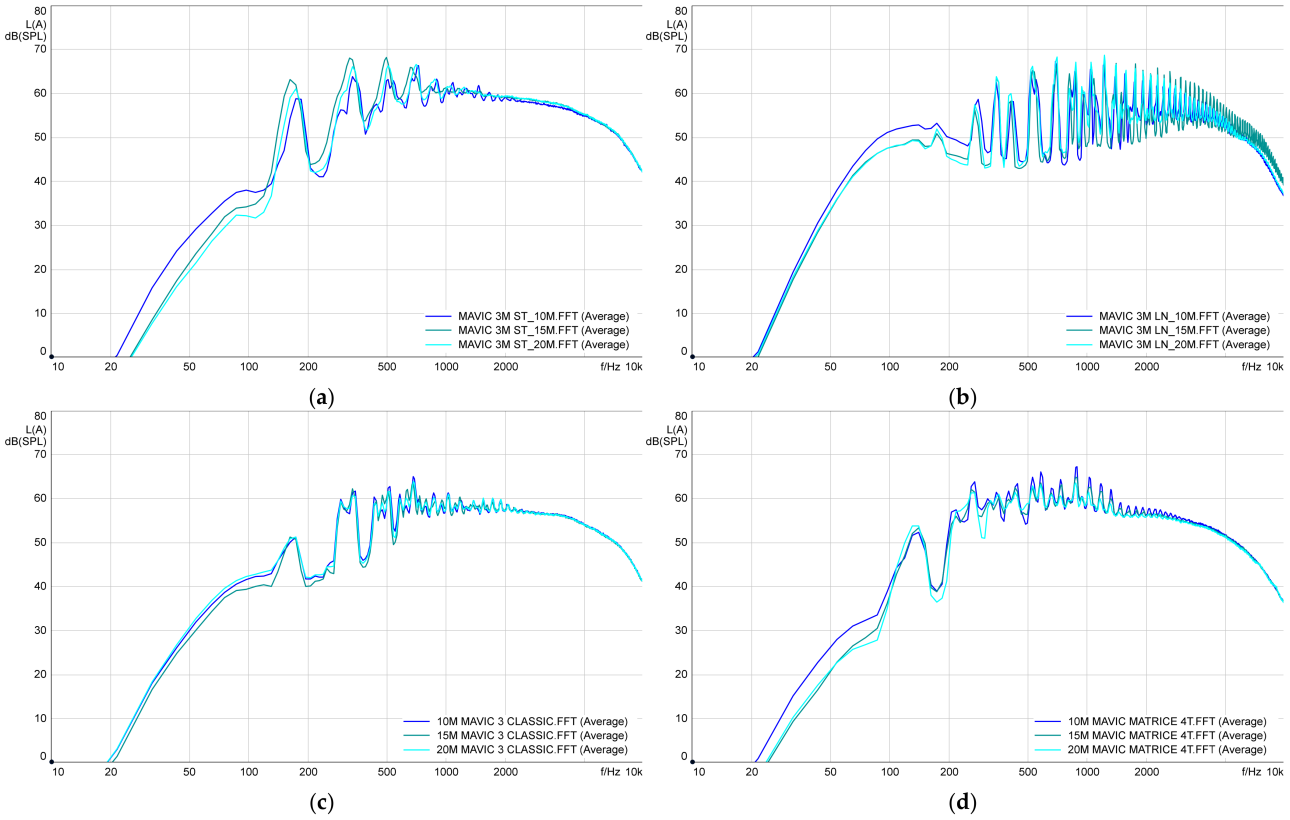


Figure 3. Spectrum of the tested drones at different heights: DJI Mavic 3M ST (**a**); DJI Mavic 3M LN (**b**); DJI Mavic 3 Classic (**c**); DJI Matrice 4T (**d**).

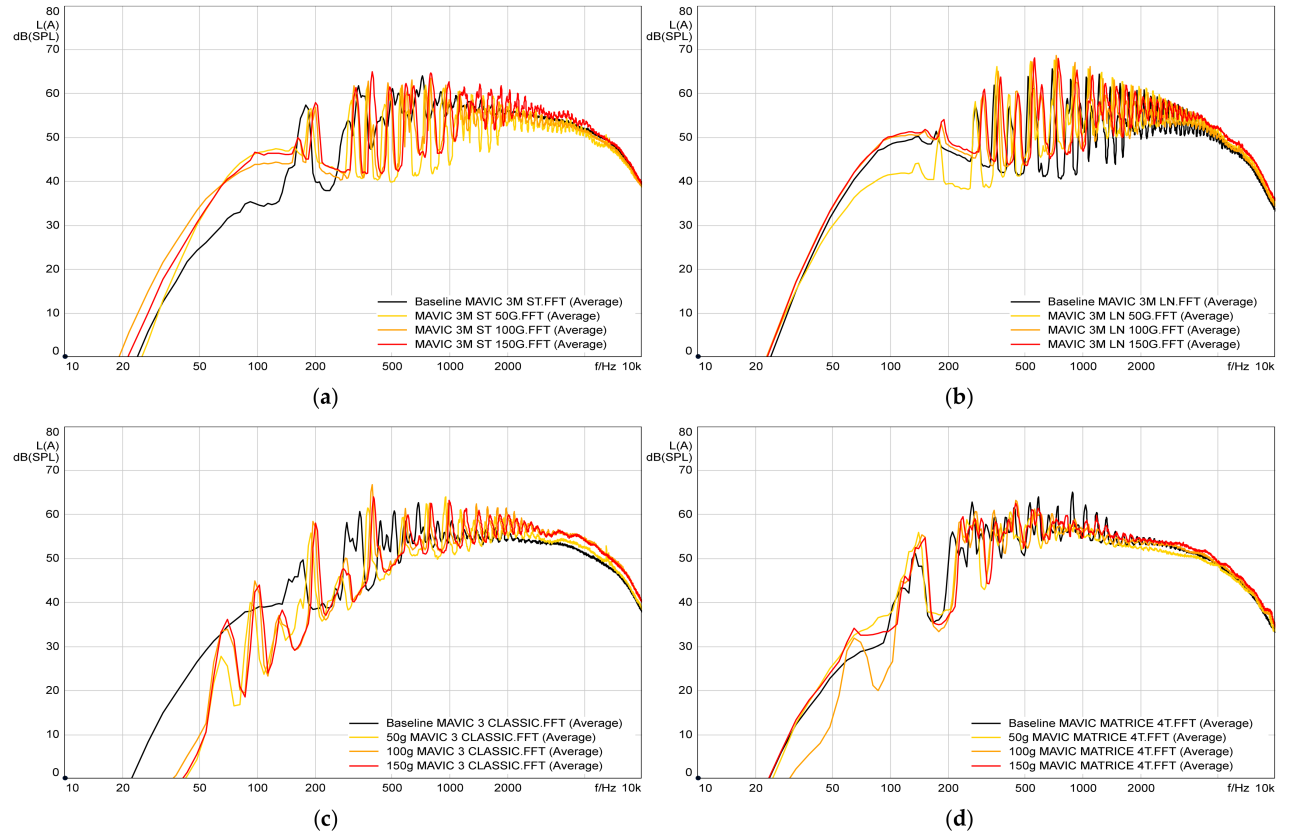


Figure 4. Spectrum of the tested drones with different payloads: DJI Mavic 3M ST (**a**); DJI Mavic 3M LN (**b**); DJI Mavic 3 Classic (**c**); DJI Matrice 4T (**d**).

The payload conditions reveal configuration-dependent behaviour consistent with the psychoacoustic results. For the 3M ST and the 3 Classic (Figure 4a,c), the harmonic comb shifts slightly upward in frequency with increasing load, reflecting the higher rotor speed required to carry the added weight (BPF rising from 173 to 204 Hz), while the mid- and high-frequency broadband region changes little. The 3M LN (Figure 4b) displays the sharpest and most deeply modulated comb structure of the whole set—strong tonal peaks separated by deep inter-harmonic troughs—visually corroborating its markedly higher roughness and tonality relative to the 3M ST, despite the two sharing essentially the same fundamental frequency. The Matrice 4T (Figure 4d) shows the opposite extreme: its four payload curves nearly coincide across the entire band, confirming the uniquely stable psychoacoustic profile reported above and its constant BPF (140 Hz) under load.

3.5. Directivity Angle

In idealised models, the acoustic radiation of a multirotor UAV is often assumed to be symmetric with respect to its transversal plane. The present dataset shows small but measurable differences between mirror-image angular positions (+30°/−30°, +60°/−60°). Mean angular asymmetry between mirror-image positions ranges from 0.2 to 4.2 dB in OSPL (0.1–2.9 dBA) across drones and planes (Table 7), with clear platform-level differences: the Matrice 4T shows by far the largest and most plane-dependent asymmetry (~4.0 dB longitudinal vs. ~2.0 dB transversal), while the 3M LN shows the smallest and most uniform (0.25–1.0 dB), with the 3M ST and 3 Classic falling in between. Notably, the 3M LN's high spatial symmetry contrasts with its elevated roughness (Table 8), showing that its psychoacoustic complexity is not accompanied by a correspondingly complex directivity. Because the absolute heading of each drone relative to the arc rig was not strictly controlled across measurement sessions, these differences cannot be attributed to a fixed physical asymmetry of a specific rotor or side of the airframe. They are more plausibly related to the active attitude corrections performed by the flight controller to maintain hovering stability, combined with session-to-session variability in drone heading—notably, the sign of the asymmetry is not consistent across drones and planes (e.g., it reverses between the longitudinal and transversal plane for the 3 Classic), supporting an operational rather than structural origin. Accordingly, the following analysis focuses on the magnitude of the angular asymmetry rather than on attributing it to a specific side of the platform.

Table 7. Mean angular asymmetry (dB) by drone and measurement planes.

Drone	Plane	Angle [° Left/° Right]	OSPL [dB]	$OSPL_A$ [dBA]
3M ST	Longitudinal	+60°/+30°/0°/−30°/−60°	84.1/87.0/86.9/85.5/85.6	82.4/85.2/85.1/84.8/84.2
	Transversal	+60°/+30°/0°/−30°/−60°	83.5/87.2/86.9/85.4/83.8	80.7/84.5/85.1/83.5/82.0
3M LN	Longitudinal	+60°/+30°/0°/−30°/−60°	83.1/85.0/87.2/84.8/83.4	81.2/82.9/84.5/84.0/82.0
	Transversal	+60°/+30°/0°/−30°/−60°	84.2/84.4/87.2/84.2/82.4	81.1/83.2/84.5/83.1/80.3
3 Classic	Longitudinal	+60°/+30°/0°/−30°/−60°	85.4/87.0/87.1/85.2/83.3	83.0/84.9/87.0/84.0/81.8
	Transversal	+60°/+30°/0°/−30°/−60°	85.0/85.4/87.1/88.1/86.3	82.4/84.4/87.0/85.7/83.8
Matrice 4T	Longitudinal	+60°/+30°/0°/−30°/−60°	83.0/83.2/84.2/87.4/86.9	80.3/81.2/82.8/84.1/82.9
	Transversal	+60°/+30°/0°/−30°/−60°	85.5/84.5/84.2/86.4/87.6	81.7/81.9/82.8/83.7/83.1

The complete angular dataset also allows off-axis levels to be compared directly against each configuration's own 0° on-axis reference recorded in the same session (Table 7), rather than against the single-microphone baseline recorded separately with different acquisition hardware (Table 3). Three configurations (3M LN, 3M ST, 3 Classic) show a broadly consistent pattern, with off-axis levels mostly at or below the 0° reference—most clearly for the 3M LN (2.2–4.8 dB below at all eight angles), indicating a directivity lobe

peaked at 0°. The Matrice 4T departs from this pattern, with off-axis OSPL exceeding the 0° reference at six of eight angles (up to +3.4 dB), consistent with its larger angular asymmetry above and plausibly related to its larger airframe increasing rotor–airframe interaction effects. Given the substantially greater mass of the arc-shaped microphone array relative to the single-microphone baseline (see Section 6), these directivity patterns should be interpreted as characteristic of the loaded (array-mic) configuration rather than as a direct proxy for each drone's unloaded emission pattern.

Table 8. Psychoacoustic descriptors at 10 m above ground level (all drones, no added payload).

Drone	N_5 [soneGF]	S [Acum]	R [Asper]	T [tuHMS]
3M ST	69.9	2.00	0.320	0.248
3M LN	64.5	1.78	1.090	0.358
3 Classic	64.0	1.96	0.396	0.192
Matrice 4T	65.0	1.75	0.855	0.298

4. Application of UAM-PA Model

As the PAI computed here is derived from recordings acquired at 50 cm below the fuselage, it serves as a comparative index of the emission noise annoyance among drone configurations and does not directly represent the annoyance experienced at a ground-based receiver. Translating on-board measurements to receiver-position annoyance estimates would require propagation modelling (geometric spreading, atmospheric absorption, ground effect), which falls outside the scope of the present study.

This section reports PAI values computed via the Boucher et al. [19] UAM PA model across all tested conditions, applying the model in full, including the sharpness-dependent weight w_S; PA is therefore determined by the combined contributions of N_5, w_S (**sharpness**), w_{FR} (roughness-fluctuation), and w_T (tonality).

First, Sharpness values measured across this dataset range from 1.7 to 2.2 acum, placing most tested conditions at or just above the model's 1.75 acum activation threshold. This range sits at the upper edge of the TUSQ calibration set underlying the Boucher et al. [19] UAM PA model, in which S remained below 1.75 acum for 138 of the 141 stimuli tested; consequently, w_S is small relative to w_{FR} for most configurations. The 3M ST configuration is the notable exception: its S (≈2.0 acum, consistently across altitude and payload) sits further above threshold, making w_S a non-negligible contribution for this configuration, and in some of its recording conditions, it is even the largest contributor to its PAI (see Sections 3.2 and 4.3). The implications of this threshold-adjacent positioning for the reliability of w_S are discussed further in Limitations.

Second, elevated high-frequency content measured close to the drone is consistent with a specific, well-documented electromechanical noise source: measurements on small quadcopter Brushless Direct Current motors have identified discrete acoustic tones near 5000 Hz, produced by electromagnetic force ripple exciting a structural resonance mode of the rotor, with installation in the airframe further increasing radiated levels between 2 and 4 kHz relative to the isolated motor [29]. The present configuration, with the microphone mounted 50 cm below the fuselage close to the motors themselves, is particularly susceptible to this electromechanical contribution, which is a plausible explanation for why the measured sharpness values sit persistently close to the model's activation threshold.

Third, a related and likely concurrent but less dominant contribution may arise from near-field acoustic effects more generally: prior UAV noise measurement studies have documented amplification of high-frequency bands and self-noise effects attributable to microphone mounting and rotor-induced airflow, applying dedicated windscreens or correction filters to mitigate them [10,16]. Unlike the electromechanical resonance discussed

above, however, this mechanism alone would not straightforwardly explain the relative stability of sharpness observed here across markedly different rotor speeds (Table 6); the present dataset cannot fully disentangle the two mechanisms without dedicated spectral analysis targeting the 4.5–5.5 kHz band associated with the motor resonance.

Taken together, these considerations indicate that the sharpness values measured in this on-board dataset should be interpreted with caution, sitting as they do close to the 1.75 acum activation threshold of the model, in a region only sparsely represented in the TUSQ calibration set. Consistent with the approach described in Section 2.2, w_S is therefore computed directly from the measured S of each condition, and all PAI values reported in this study include its contribution alongside the roughness/fluctuation-strength and tonality terms; the reliability implications of this near-threshold positioning are discussed further in the 'Limitations' section.

Table 8 presents the complete set of psychoacoustic descriptors for all four drone configurations at the reference altitude of 10 m above ground level, enabling a direct inter-model comparison; the corresponding PAI values are reported in Table 9 and computed using the adapted Boucher et al. [19] UAM PA model.

Table 9. Psychoacoustic Annoyance Index (PAI, Boucher et al. UAM PA model [19]) computed with and without the sharpness term w_S at 10 m and averaged between heights (10 m, 15 m, 20 m).

Drone	PAI (10 m) [-]	PAI (10 m) (w_S = 0) [-]	Mean PAI [-]	Mean PAI (w_S = 0) [-]
3M ST	79.9	75.4	84.6	80.8
3M LN	82.1	82.0	82.9	82.7
3 Classic	73.0	70.4	71.5	69.9
Matrice 4T	78.8	78.8	77.1	77.1

4.1. PAI Across Drones

The PAI values at 10 m above ground level lead to the following ranking of the expected drone noise annoyance: $PAI_{3M\ LN} > PAI_{3M\ ST} > PAI_{Matrice\ 4T} > PAI_{3\ Classic}$—driven primarily by roughness (3M LN) and by sharpness engaging a non-negligible w_S term for the 3M ST (Section 4.3), which is enough to offset the Matrice 4T's higher roughness and combined weight factor. This illustrates that PAI reflects the product of loudness and the combined weight factor, not the weight factor alone: the Matrice 4T's lower loudness (65.0 vs. 69.9 soneGF) is enough to offset its larger weight factor.

The 3 Classic achieves the lowest PAI (72.97) despite a moderate w_S (0.098, from S = 1.96 acum) comparable in magnitude to its w_{FR} (0.100), because it has the lowest loudness (64.0 soneGF) among all models, keeping the overall product $N_5 \times$ factor lowest in the dataset.

Table 9 reports PAI values computed with and without the sharpness term w_S. Removing w_S affects the ranking in two cases: at 10 m, the 3M ST drops from second to third place behind the Matrice 4T; in the altitude-averaged comparison, the 3M ST and 3M LN ranking reverses entirely. These results confirm that rankings involving the 3M ST should be interpreted with caution, as they depend materially on a sharpness term that sits close to the model activation threshold.

4.2. PAI Across Heights

Table 10 reports the PAI values computed with the adapted UAM PA model [19] for all configurations at 10 m, 15 m, and 20 m above ground level, together with a mean PAI value across the three altitudes.

The mean PAI was not computed as a simple arithmetic average of the three altitude-specific PAI values, but rather by averaging the N_5 component and the multiplicative factor

$\left(1+\sqrt{w_S^2+w_{FR}^2+w_T^2}\right)$ separately across the three altitudes, and then combining these averaged components via the model formula. This approach is more consistent with the structure of the Boucher et al. [19] UAM PA model, as it averages the perceptual components before applying the non-linear combination, rather than averaging the final outputs of three distinct conditions. The two methods yield numerically near-identical results on this dataset (differences < 0.1 PAI units), confirming that altitude-induced variability in the individual psychoacoustic components is limited.

Table 10. Psychoacoustic Annoyance Index (PAI, Boucher et al. UAM PA model [19]) by drone and altitude. Mean PA is computed using the average of N_5 and the average multiplicative factor across the three altitudes.

Drone	PAI (10 m) [-]	PAI (15 m) [-]	PAI (20 m) [-]	Mean PAI [-]
3M ST	79.9	88.0	86.1	84.6
3M LN	82.1	86.5	80.1	82.9
3 Classic	73.0	69.5	72.2	71.5
Matrice 4T	78.8	78.2	74.3	77.1

The ranking of the mean PAI shows that $PAI_{3M\ ST}$ > $PAI_{3M\ LN}$ > $PAI_{Matrice\ 4T}$ > $PAI_{3\ Classic}$—a reversal of the ranking observed at the 10 m reference height alone, where the $PAI_{3M\ LN}$ was the highest. Averaged across altitudes, the 3M ST's sharpness-related weight term remains elevated at every altitude tested, enough to outweigh the 3M LN's advantage from lower loudness. Because the on-board microphone's distance to the source is constant, regardless of the flight height, these altitude-to-altitude differences do not reflect sound propagation effects, consistent with the limited, non-monotonic altitude sensitivity already reported in Table 3. The $PAI_{3M\ ST}$ shows the largest spread across altitudes (~8.1 PAI units) and the $PAI_{3\ Classic}$ the most stable (~3.5 units); while the $PAI_{Matrice\ 4T}$ shows a decreasing trend (78.8 → 74.3); none of these patterns should be read as a systematic altitude effect, since the broader dataset shows no consistent monotonic trend across drones.

4.3. PAI Across Propellers

The DJI Mavic 3M was the only model in this campaign available in two propeller variants: the manufacturer's standard blades and the DJI low-noise option. All other experimental conditions were identical (hovering at 10 m AGL, single microphone, no added payload).

At the 10 m reference height, despite nearly identical broadband emission, the 3M LN produces a PAI approximately 2.2 units higher than the 3M ST (82.1 vs. 79.9)—a considerably narrower gap than a comparison based on roughness and tonality alone would suggest: the w_{FR} term is about 3.5× higher for the LN (0.271 vs. 0.078) and the w_T term about 1.6× higher (0.018 vs. 0.011). What narrows the gap is sharpness: the 3M ST's S = 2.00 acum, just above the 1.75 acum threshold, yields a w_S (0.119) comparable to its own w_{FR}, while the 3M LN's S = 1.78 acum yields a much smaller w_S (0.014). Averaged across altitudes, this effect is large enough to reverse the ranking entirely (Section 4.2), since the 3M ST's sharpness term stays elevated and consistent across all three heights while its roughness and tonality—always the lowest in the dataset—do not. This illustrates that a PAI model based on loudness and sharpness alone would not resolve the 3M ST vs. 3M LN comparison consistently across conditions.

4.4. PAI Across Payloads

Table 11 reports the full set of psychoacoustic descriptors and PAI (Boucher et al. UAM PA model [19]) for all drone models under baseline and added-payload conditions at 10 m AGL.

Table 11. Psychoacoustic descriptors (N_5, S, R, T, FS) and PAI (Boucher et al. UAM PA model [19]) for all drone models under baseline and added-payload conditions at 10 m AGL.

Drone	Payload [-]	N5 [soneGF]	S [Acum]	R [Asper]	T [tuHMS]	PAI [-]
3M ST	Baseline	69.9	2.00	0.320	0.248	79.9
	+50 g	64.4	1.95	0.750	0.315	77.9
	+100 g	64.3	1.95	0.356	0.159	72.6
	+150 g	71.7	1.94	0.608	0.336	84.0
3M LN	Baseline	64.5	1.78	1.090	0.358	82.1
	+50 g	65.9	1.85	0.993	0.519	82.4
	+100 g	69.9	1.74	0.918	0.534	85.4
	+150 g	73.6	1.78	0.912	0.529	89.6
3 Classic	Baseline	64.0	1.96	0.396	0.192	73.0
	+50 g	64.0	2.08	0.729	0.410	79.3
	+100 g	73.0	2.10	0.677	0.490	90.1
	+150 g	69.5	2.10	0.688	0.364	85.9
Matrice 4T	Baseline	65.0	1.75	0.855	0.298	78.8
	+50 g	59.8	1.77	0.608	0.119	69.2
	+100 g	64.7	1.76	0.581	0.130	74.1
	+150 g	67.6	1.78	0.577	0.110	77.2

Payload effects on PAI differ markedly by platform (Table 11, Figure 5). The $PAI_{3M\ ST}$ shows a non-monotonic pattern, dipping to a minimum at +100 g (72.6) before rising sharply to 84.0 at +150 g—its highest value, ~4.1 units above the baseline. The $PAI_{3M\ LN}$ increases monotonically across all payload conditions (82.1 → 89.6), driven primarily by rising loudness. The $PAI_{3\ Classic}$ shows the largest swing in the dataset, jumping ~17.1 units to a maximum of 90.1 at +100 g—the highest PAI recorded anywhere in this study—before a partial recovery at +150 g. The $PAI_{Matrice\ 4T}$ stands apart: it is the only configuration for which PAI never exceeds its unloaded baseline, dipping at +50 g and recovering only approximately linearly thereafter—a pattern consistent with its high absolute weight (which limits the fractional demand on the motor) and with its baseline roughness being attenuated, rather than amplified, by the increased rotor speed under payload, mirrored in its stable BPF (140 Hz unchanged across all payload conditions).

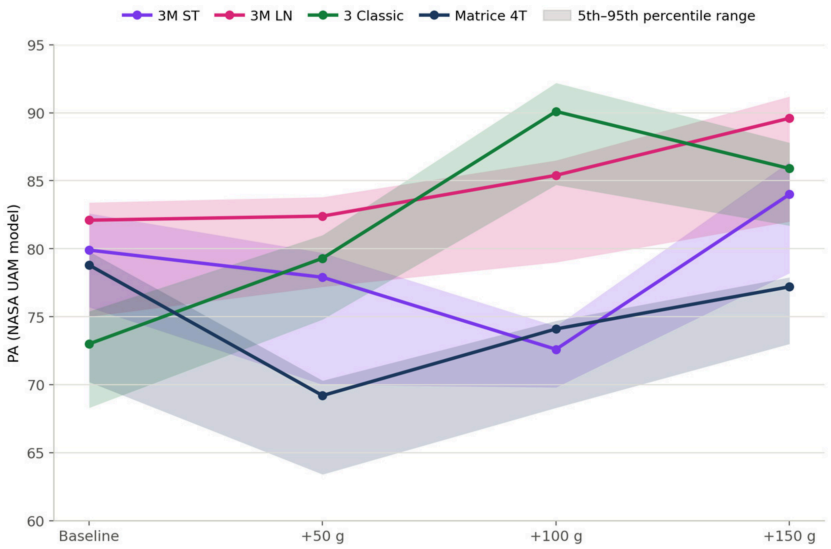


Figure 5. Psychoacoustic Annoyance Index (PAI, Boucher et al. UAM PA model [19]) as a function of added payload, by drone/propeller configuration. Dotted horizontal lines mark each configuration's unloaded baseline PAI.

The most evident pattern in Figure 6 is a convergence in the roughness-fluctuation weight w_{FR} across platforms as payload increases: starting from the widest split in the dataset (0.078–0.271 at baseline), w_{FR} declines steadily for the two configurations with the highest starting values (3M LN, Matrice 4T) while ending higher than baseline for the two with the lowest (3M ST, 3 Classic), narrowing the across-drone spread to roughly a third of its baseline value by +150 g. The sharpness-driven weight w_S follows a distinct pattern: it remains comparatively flat or negligible for the other three configurations but rises sharply for the 3 Classic (0.098 → 0.168), approaching the magnitude of its own w_{FR} by +100 g.

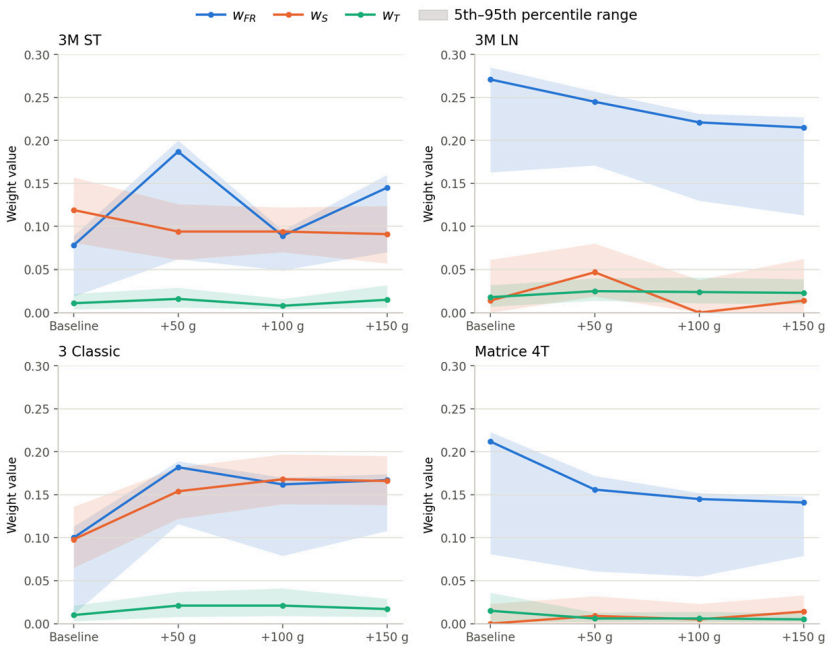


Figure 6. Model weight components of the Boucher et al. [19] UAM PA model across payload conditions, by drone configuration. Shaded bands show the 5th–95th intra-recording percentile range around each mean.

5. Discussion

Taken together, the four main questions addressed in this study converge on a single overarching finding: energy-based acoustic metrics alone are insufficient to characterise, rank, or predict the perceived annoyance of multirotor UAVs, and a psychoacoustic annoyance framework is necessary rather than optional for UAM noise assessment.

Different flight altitudes, between 10 and 20 m, showed a limited effect on acoustic descriptors, supporting 10 m as a more convenient and stable reference height.

On the other hand, in some cases, psychoacoustic descriptors may play an important role in the definition of PAI value. In particular, comparing the two 3M drones with different propeller designs, it can be observed that, while at 10 m the 3M-LN shows the highest PAI, as the altitude increases, changes in loudness, roughness and tonality make the PAI of the 3M-ST higher than the 3M-LN (Sections 4.2 and 4.3). These differences observed in the PAI values are, however, comparable to those due to recording-to-recording variability. This dependence on the specific recording condition underscores the practical relevance of a full multi-descriptor PAI model over any single-metric or single-recording comparison.

The standard and low-noise propellers of the DJI Mavic 3M differ in their nominal geometry: the low-noise propeller has a smaller diameter (21.8 cm vs. 23.88 cm) and a larger pitch (14.7 cm vs. 13.46 cm). Additionally, DJI indicates that the low-noise propellers undergo dynamic balancing and feature softer blade-edge materials. Both variants share the same baseline BPF (173 Hz, Table 6), indicating that the low-noise blades do not re-

duce broadband noise by rotating more slowly. Under added payload, the 3M LN shows a smaller BPF increase than the 3M ST (up to 183 Hz vs. 204 Hz at +150 g), consistent with a higher thrust per revolution attributable to the larger pitch. The observed changes in roughness and tonality may depend on these geometric and material differences and their interaction with blade loading and operating point: the low-noise propeller appears to reduce the broadband floor while leaving the harmonic components at comparable absolute levels, increasing the tonal-to-broadband contrast—as visually confirmed by the more deeply modulated harmonic comb structure observed for the 3M LN in Figure 4b, already noted in Section 3.4. Since rotor speed, blade loading, and thrust per rotor were not independently measured in this campaign, this interpretation is presented as plausible and not as a demonstrated causal mechanism. Further measurements involving different low-noise propeller designs and UAV platforms are required to determine whether the observed psychoacoustic behaviour can be generalised. An added payload did not increase annoyance monotonically for any platform; its effect depended on the fractional weight increase relative to each airframe's baseline mass rather than on the absolute added weight, with the heaviest platform (Matrice 4T) remaining at or below its own baseline at every tested payload level and the lightest platform (3 Classic) showing the largest swing (17.1 PAI units).

The completed directivity analysis added a further, non-obvious layer to this picture: the Matrice 4T, the most payload-stable platform, was simultaneously the platform with the largest and least symmetric angular directivity (up to 4.2 dB), while the 3M LN, the configuration with the highest roughness and PAI among the unloaded baselines, showed the most spatially symmetric and on-axis-peaked emission pattern in the dataset, indicating that perceived annoyance and spatial directivity are governed by largely independent mechanisms, a distinction relevant to UAM noise-impact models that often assume a simplified, uniform directivity from a single on-axis measurement.

These findings should be read alongside the limitations discussed below, particularly regarding the single-recording basis of each condition and the calibration uncertainty of the sharpness term. Within these limits, the dataset generated in this campaign—spanning multiple drone platforms, propeller types, payload levels, and emission angles under controlled, comparable field conditions—provides a reference resource for psychoacoustic model validation and for informing noise-aware UAM operational and regulatory guidelines: the systematic variation of operational parameters enables the disentanglement of individual acoustic factors, a prerequisite for building reliable drone noise prediction and annoyance assessment tools for UAM planning. Future work should prioritise repeated independent hovering trials to quantify test-retest reliability, a mass-matched directivity configuration, and direct comparison of on-board and ground-based sharpness measurements to establish whether sharpness values this close to the model's activation threshold reflect a stable near-field acoustic property or measurement-level noise that could shift w_S materially.

Future work should also extend the present protocol to heavier cargo drone platforms and fixed-wing UAVs, which represent an increasingly relevant segment of the UAM ecosystem and whose acoustic and psychoacoustic emission profiles remain largely uncharacterised.

The psychoacoustic profiles documented in this study could also serve as a reference basis for future drone detection and classification studies in mixed urban acoustic environments. Future work should also consider the acoustic characterisation of drone flows and UAM airways, where the simultaneous operation of multiple UAVs may produce cumulative noise effects that differ substantially from those of individual platforms.

6. Limitations and Future Studies

Three aspects of the present protocol should be considered when interpreting the results. The first concerns measurement geometry: the acoustic acquisition system is mounted 0.50 m below the drone's fuselage, so the source-receiver distance stays constant throughout each hover—unlike ground-based EASA-compliant measurements [7], which require the aircraft to repeatedly enter and exit a defined geometric window across at least six independent runs per condition. This EASA multiple-run protocol is designed around a receiver fixed on the ground and does not directly transfer to a configuration where the receiver is co-located with the source: with the microphone mounted on the drone itself, there is no fixed geometric window for the aircraft to enter or exit, since the source-receiver distance does not change, regardless of how the drone is flown. Detaching and re-attaching the microphone array for repeated take-off–hover–landing cycles was not attempted within this campaign, and adapting an equivalent multiple-run scheme to an on-board configuration is left for future work.

A related consideration applies to the directivity configuration: the arc-shaped multi-microphone array adds approximately 208 g relative to the single-microphone system (Table 5), corresponding to a mass increase of 39.0%, 44.5%, and 24.4% over baseline for the 3M, 3 Classic, and Matrice 4T respectively—larger than any payload condition tested elsewhere (up to 38% at +150 g, Table 5) and, since payload measurably affects BPF and psychoacoustic descriptors (Section 3.3), a genuine confound: the directivity measurements (Table 7) were therefore not obtained under mass-matched conditions, and the angular asymmetries discussed cannot be fully disentangled from this additional structural mass.

A second consideration follows from the first: each condition is represented by a single continuous hovering recording (3 min, extracted from a session of at least 8 min) rather than repeated independent trials, so the reported descriptors reflect within-recording temporal behaviour rather than hover-to-hover reproducibility. This matters beyond a formal caveat: the angular asymmetries reported in the directivity analysis are attributed to active attitude-correction dynamics during hovering rather than structural asymmetry, and the same source of variability could, in principle, also affect other single-hover comparisons (e.g., propeller type, payload level) without being independently quantified. The 5th/95th percentile range of each descriptor offers a first check against this intra-recording spread: the Roughness-driven propeller comparison holds up well (non-overlapping ranges for the 3M ST and 3M LN), but the Tonality and Sharpness contributions are more sensitive to it—for several conditions, the model's 1.75 acum sharpness threshold falls within the intra-recording spread itself, meaning w_S should be read as one plausible value along a range that may include zero, rather than as a stable quantity. Future work should prioritise repeated independent hovering trials to establish hover-to-hover variance directly, ideally with confidence-interval reporting as required by EASA guidelines [7] for ground-based measurements.

A third limitation concerns the sharpness-dependent weight of the psychoacoustic annoyance model. The sharpness values measured throughout this study ($S \approx 1.7$–2.2 acum, Tables 8 and 11) sit within, or just above, the model's 1.75 acum activation threshold—a region represented by very few stimuli in the original TUSQ calibration set (only 3 of 141 stimuli exceeded 1.75 acum [19]). The resulting w_S values therefore carry more calibration uncertainty than the roughness- and tonality-driven terms, most consequential for the 3M ST configuration, whose PAI ranking relative to the Matrice 4T and, when averaged across altitude, relative to the 3M LN, depends materially on this term. Part of this near-threshold sharpness may itself be a measurement-geometry effect—from motor-induced electromechanical noise near the source [29] or near-field microphone effects [10,16]—rather than a genuine property of the emitted sound; a direct on-board vs. ground-based com-

parison under matched flight conditions, as recommended above, would help disentangle the two and clarify how reliably w_S can be applied to on-board recordings of this kind. Additionally, a quantitative propagation of the calibration uncertainty to the individual psychoacoustic descriptors—in particular sharpness, roughness, and tonality—could not be performed within the present campaign, as it would require a dedicated frequency-response characterisation of the full measurement chain under controlled conditions; this is identified as a further priority for future work.

A fourth consideration concerns the scope of flight conditions tested: this study characterised only the hovering condition. Hovering alone is not representative of a UAV's full acoustic signature in operation, since forward-flight/pass-by conditions have been shown to differ measurably from hovering in both overall sound pressure level and psychoacoustic character, with differences of up to several decibels attributable to unsteady aerodynamic interactions between rotor wakes and the airframe rather than to rotor speed alone [10]. Future work should therefore extend the present protocol beyond hovering to include flyover and other representative flight manoeuvres and should expand the number of drone platforms and the range of payload conditions tested toward a broader, standardised on-board noise and psychoacoustic annoyance database for UAM applications.

7. Conclusions

This study shows that overall sound pressure level alone cannot predict UAV noise annoyance: propeller design, payload, and directivity each shape psychoacoustic annoyance in configuration-specific, non-monotonic ways. A full multi-descriptor PAI framework is therefore needed for UAM noise assessment, and the dataset generated here offers a reference basis for future model validation.

More broadly, the on-board hovering recordings collected here across multiple drone platforms, propeller types, and payload conditions represent an initial contribution toward a larger, standardised database of drone acoustic and psychoacoustic annoyance data; a resource still largely missing from the UAM literature and needed to support future model validation and noise-aware regulatory guidelines.

Author Contributions: Conceptualisation, M.M. and J.M.N.; data curation, N.G. and M.M.; investigation, N.G.; methodology, N.G., J.M.N. and M.M.; project administration, M.M., J.M.N. and L.M.; resources, N.G., M.M., J.M.N.; software, N.G.; supervision, J.M.N. and M.M.; validation, N.G. and M.M.; writing—original draft, N.G. and M.M.; writing—review and editing, J.M.N. and L.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Ministero dell'Università e della Ricerca: DM 118/23 PNRR (M4C1—Inv. 3.4 TDA)—VANVITELLI.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy restrictions.

Acknowledgments: The authors gratefully acknowledge the Protección Civil Mazarrón (DJI Matrice 4T and DJI Mavic 3 Classic) and the UCAM Drone Hub (recording site and DJI Mavic 3M) for their support during the measurement campaign.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. European Union Aviation Safety Agency (EASA). *Study on the Societal Acceptance of Urban Air Mobility in Europe*; EASA: Cologne, Germany, 2021.
2. Rinaldi, M.; Maghsoodi, S.; Primatesta, S. On the assessment of drone noise for sustainable urban air mobility operations. *Eng. Proc.* **2026**, *133*, 43. [CrossRef]
3. Maketo, L.; Ashworth, P. Social acceptance of green hydrogen in European Union and the United Kingdom: A systematic review. *Renew. Sustain. Energy Rev.* **2025**, *218*, 115827. [CrossRef]
4. Torija, A.J.; Clark, C. A psychoacoustic approach to building knowledge about human response to noise of unmanned aerial vehicles. *Int. J. Environ. Res. Public Health* **2021**, *18*, 682. [CrossRef] [PubMed]
5. Gwak, D.Y.; Han, D.; Lee, S. Sound quality factors influencing annoyance from hovering UAV. *J. Sound Vib.* **2020**, *489*, 115651. [CrossRef]
6. Kawai, C.; Jäggi, J.; Georgiou, F.; Meister, J.; Pieren, R.; Schäffer, B. Short-term noise annoyance towards drones and other transportation noise sources: A laboratory study. *J. Acoust. Soc. Am.* **2024**, *156*, 2578–2595, Erratum in *J. Acoust. Soc. Am.* **2025**, *157*, 288. [CrossRef] [PubMed]
7. European Union Aviation Safety Agency (EASA). Guidelines on Noise Measurement of UAS Lighter Than 600 kg. EASA, Cologne, 2022/2023. Available online: https://www.easa.europa.eu (accessed on 1 July 2026).
8. Alexander, W.N.; Whelchel, J. Flyover noise of multi-rotor sUAS. In Proceedings of the INTER-NOISE and NOISE-CON Congress and Conference Proceedings (InterNoise19), Madrid, Spain, 16–19 June 2019; Volume 259, pp. 2548–2558.
9. Cabell, R.; McSwain, R.G.; Grosveld, F. Measured noise from small unmanned aerial vehicles. In Proceedings of the NOISE-CON 2016, Providence, RI, USA, 13–15 June 2016; NASA/NF1676L-23558.
10. Meister, J.; Jäggi, J.; Van Renterghem, T.; Pieren, R. Measured differences in the sound emission of multi-rotor unmanned aerial systems during hover and cruise. *J. Acoust. Soc. Am.* **2026**, *159*, 2398–2415. [CrossRef] [PubMed]
11. Noda, R.; Hirose, M.; Nakata, T.; Liu, H. Low-noise propeller design with enlarged blade area for drones. *J. Robot. Mechatron.* **2025**, *37*, 799–806. [CrossRef]
12. Sun, J.; Yonezawa, K.; Shima, E.; Liu, H. Integrated evaluation of the aeroacoustics and psychoacoustics of a single propeller. *Int. J. Environ. Res. Public Health* **2023**, *20*, 1955. [CrossRef] [PubMed]
13. Ramos-Romero, C.; Green, N.; Torija, A.J.; Asensio, C. On-field noise measurements and acoustic characterisation of multi-rotor small unmanned aerial systems. *Aerosp. Sci. Technol.* **2023**, *141*, 108537. [CrossRef]
14. Ibrahim, O.A.; Sciancalepore, S.; Di Pietro, R. Noise2Weight: On detecting payload weight from drones acoustic emissions. *Future Gener. Comput. Syst.* **2022**, *134*, 319–333. [CrossRef]
15. Boucher, M.; Rafaelof, M.; Begault, D.; Christian, A.; Krishnamurthy, S.; Rizzi, S. A psychoacoustic test for urban air mobility vehicle sound quality. *SAE Int. J. Adv. Curr. Pract. Mobil.* **2024**, *6*, 972–985. [CrossRef]
16. König, R.; Babetto, L.; Gerlach, A.; Fels, J.; Stumpf, E. Prediction of perceived annoyance caused by an electric drone noise through its technical, operational, and psychoacoustic parameters. *J. Acoust. Soc. Am.* **2024**, *156*, 1929–1941. [CrossRef] [PubMed]
17. Torija Martinez, A.J.; Li, Z. Metrics for assessing the perception of drone noise. In Proceedings of the Forum Acusticum 2020, Lyon, France, 7–11 December 2020; pp. 3163–3168.
18. Vaiuso, A.; Righi, M.; Coretti, O.; Apicella, M. Drone acoustic analysis for predicting psychoacoustic annoyance via artificial neural networks. *arXiv* **2024**, arXiv:2410.22208.
19. Boucher, M.A.; Christian, A.W.; Krishnamurthy, S.; Tracy, T.; Begault, D.R.; Shepherd, K.; Rizzi, S.A. *Toward a Psychoacoustic Annoyance Model for Urban Air Mobility Vehicle Noise*; NASA/TM-20240003202; NASA Langley Research Center: Hampton, VA, USA, 2024.
20. Kephalopoulos, S.; Paviotti, M.; Anfosso-Lédée, F. *Common Noise Assessment Methods in Europe (CNOSSOS-EU)*; EUR 25379 EN.; Publications Office of the European Union: Luxembourg, 2012. [CrossRef] [PubMed]
21. *IEC Standard No. 61672-1:2013*; Electroacoustics—Sound Level Meters—Part 1: Specifications. International Electrotechnical Commission: Geneva, Switzerland, 2013.
22. Fastl, H.; Zwicker, E. *Psychoacoustics: Facts and Models*, 3rd ed.; Springer: Berlin/Heidelberg, Germany, 2007.
23. HEAD acoustics GmbH. *ArtemiS SUITE, version 14*; HEAD acoustics GmbH: Herzogenrath, Germany, 2022.
24. *ISO 532-1:2017*; Acoustics—Methods for Calculating Loudness—Part 1: Zwicker Method. International Organization for Standardization: Geneva, Switzerland, 2017.
25. *DIN 45692:2009*; Measurement Technique for the Simulation of the Auditory Sensation of Sharpness. Deutsches Institut für Normung: Berlin, Germany, 2009.
26. Ecma International. *Measurement of Airborne Noise Emitted by Information Technology and Telecommunications Equipment*, 17th ed.; ECMA-74; Ecma International: Geneva, Switzerland, 2019.
27. Ecma International. *Psychoacoustic Metrics for ITT Equipment–Part 2: Models Based on Human Perception*, 2nd ed.; ECMA-418-2; Ecma International: Geneva, Switzerland, 2022.

28. Sottek, R.; Genuit, K. Models of signal processing in human hearing. *AEU-Int. J. Electron. Commun.* **2005**, *59*, 157–165. [CrossRef]
29. Henderson, B.S.; Huff, D.L.; Cluts, J.D.; Ruggeri, C.R. Electric motor noise from small quadcopters: Part II—Source characteristics. In Proceedings of the 2018 AIAA/CEAS Aeroacoustics Conference (AIAA Aviation Forum), Atlanta, GA, USA, 25–29 June 2018.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.