

Politecnico di Torino

Politecnico di Torino

Ingegenria del Cinema e dei Mezzi di Comunicazione

A.a. 2025/2026

Sessione di laurea Settembre 2026

Psychoacoustic Assessment of Drone Noise through Acoustic Measurements, Subjective Evaluation and Spatial Sound Reproduction

Relatori:

Arianna Astolfi

Candidati:

Francesco Grillo

Sommario

The increasing use of unmanned aerial vehicles (UAVs), commonly referred to as drones, in civil, industrial and recreational applications has raised growing interest in their environmental noise impact. Although drone noise can be evaluated through conventional acoustic descriptors, such as A-weighted sound pressure levels, these parameters alone may not fully explain the subjective response elicited by drone sounds. In particular, the complex spectral and temporal characteristics of drone noise may influence its perceived loudness, sharpness, roughness, fluctuation and tonality, as well as the overall annoyance experienced by listeners. This thesis investigates the psychoacoustic characteristics and subjective perception of drone noise through an experimental approach combining objective acoustic measurements, psychoacoustic analysis and listening tests. Experimental measurements are conducted in an outdoor flight field, where different drone types and flight manoeuvres are evaluated under controlled experimental conditions. The considered manoeuvres include vertical take-off, landing and horizontal cruise. During the experiments, acoustic measurements are simultaneously acquired using a sound level meter, while spatial audio recordings are performed using a Zylia microphone system. The subjective response of participants is collected through a structured questionnaire designed to evaluate overall annoyance and the perceived characteristics of the drone sound. In addition to an annoyance scale from 0 to 10, participants assess perceptual attributes related to loudness, fluctuation, roughness, sharpness and tonality using semantic differential scales. The subjective evaluations are subsequently compared with conventional acoustic descriptors obtained from the sound level meter measurements and with psychoacoustic parameters extracted from the recorded signals. A further objective of the research is to investigate the relationship between physical and perceptual characteristics of drone noise. The analysis considers whether conventional acoustic indicators and psychoacoustic metrics can explain differences in perceived annoyance and in the subjective characterization of the sounds produced by different drones and flight manoeuvres. Finally, the spatial characteristics of drone noise are considered through the use of third-order Ambisonics recordings obtained with the Zylia microphone and their reproduction in the Audio Space Lab of Politecnico di Torino. This approach provides a framework for investigating drone noise not only as a conventional acoustic source, but also as a spatially evolving sound source whose perception may depend on its movement and acoustic characteristics. The results are intended to contribute to a better understanding of the relationship between measurable acoustic and psychoacoustic characteristics of drone noise and human perception,

providing a methodological framework for the assessment of UAV noise and for future studies concerning its spatial reproduction and perceptual impact.

Indice

Elenco delle figure v

1 Introduction 1
- 1.1 Motivation . . . 1
- 1.2 Aim and Research Questions . . . 2
- 1.3 Experimental and Analytical Approach . . . 2
- 1.4 Contribution and Scope . . . 3
- 1.5 Thesis Organisation . . . 3

2 Literature Review 4
- 2.1 UAVs and the Emerging Noise Problem . . . 4
- 2.2 Acoustic Characteristics of Drone Noise . . . 5
  - 2.2.1 Tonal Components and Blade Passing Frequency . . . 6
  - 2.2.2 Broadband Noise . . . 6
  - 2.2.3 Temporal Characteristics and Modulation . . . 7
  - 2.2.4 Acoustic Directivity . . . 7
  - 2.2.5 Influence of Flight Conditions . . . 8
- 2.3 Acoustic Measurement and Characterization of UAV Noise . . . 9
  - 2.3.1 Laboratory Measurements . . . 9
  - 2.3.2 Field Measurements . . . 9
  - 2.3.3 Single-Channel and Multichannel Measurements . . . 10
  - 2.3.4 Measurement of Dynamic Flight Events . . . 11
- 2.4 From Conventional Acoustic Metrics to Psychoacoustic Assessment 11
  - 2.4.1 Perceived Noise Level . . . 12
  - 2.4.2 Sound Quality Metrics . . . 12
  - 2.4.3 Relationship with Subjective Annoyance . . . 12
- 2.5 Psychoacoustic Metrics for Drone Noise . . . 13
  - 2.5.1 Loudness . . . 14
  - 2.5.2 Sharpness . . . 14
  - 2.5.3 Tonality . . . 15
  - 2.5.4 Roughness . . . 15

2.5.5 Fluctuation Strength . . . 15
2.5.6 Psychoacoustic Annoyance . . . 16
2.6 Human Response and Noise Annoyance . . . 16
2.6.1 Influence of Acoustic Context . . . 17
2.6.2 Non-Acoustic Factors . . . 17
2.7 Listening Tests for Drone Noise Assessment . . . 18
2.7.1 Rating Scales . . . 18
2.7.2 Participants and Experimental Design . . . 18
2.7.3 Reproduction of Drone-Noise Stimuli . . . 19
2.8 Binaural Recording and Headphone Reproduction . . . 19
2.8.1 Spatial Hearing and Ear-Specific Signals . . . 19
2.8.2 Direct Binaural Acquisition . . . 20
2.8.3 Headphone Transfer and Playback Level . . . 20
2.8.4 Externalization and Head Movement . . . 21
2.8.5 Outdoor Exposure and Indoor Listening . . . 21
2.9 Recent Developments in Psychoacoustic Prediction . . . 22
2.9.1 Association, Explanatory Modelling and Prediction . . . 22
2.10 Research Gap and Study Positioning . . . 23

3 Materials and Methods 26
3.1 Study Design . . . 26
3.2 Experimental Site and Flight Protocol . . . 27
3.2.1 UAVs and Manoeuvres . . . 27
3.3 Acquisition and Reproduction Equipment . . . 30
3.3.1 Conventional and Multichannel Acquisition . . . 30
3.3.2 Binaural Acquisition . . . 30
3.3.3 Indoor Playback . . . 30
3.4 Participants and Questionnaire . . . 31
3.4.1 Response Scales . . . 32
3.5 Outdoor and Indoor Test Procedures . . . 32
3.6 Data Structure and Measurement Correspondence . . . 33
3.7 Analysis Datasets and Units of Observation . . . 35

4 Signal Processing and Analysis Methods 36
4.1 Processing Workflow and Traceability . . . 36
4.1.1 Preliminary Ambisonics Processing . . . 36
4.2 Conventional Acoustic and Psychoacoustic Descriptors . . . 37
4.2.1 ArtemiS Analysis and Channel Treatment . . . 37
4.3 Independent Repetitions and Descriptive Variability . . . 39
4.4 Background-Noise Quality Control . . . 39
4.5 Subjective Descriptions and Graphical Comparisons . . . 40

4.6 Paired Wilcoxon Signed-Rank Tests . . . 40
4.7 Systematic Correlation Analysis . . . 41
4.7.1 Subjective–Subjective Matrices . . . 41
4.7.2 Objective–Objective Matrices . . . 42
4.7.3 Objective–Subjective Matrices . . . 42
4.7.4 Individual Responses: Supplementary Analysis . . . 42
4.7.5 Interpretive Limits of Correlation Probabilities . . . 43
4.8 Simple Regression Models . . . 43
4.8.1 Model Families . . . 43
4.8.2 Fit Measures and Diagnostics . . . 44
4.9 Computational Traceability . . . 45

5 Results 46
5.1 Dataset Coverage and Traceability . . . 46
5.2 Acoustic Characterisation . . . 46
5.2.1 Conventional Field Descriptors . . . 46
5.2.2 Psychoacoustic Descriptors and Independent Events . . . 47
5.2.3 Background-Noise Quality Control . . . 51
5.3 Subjective Response Distributions . . . 53
5.3.1 Separate Outdoor and Indoor Samples . . . 53
5.3.2 Matched Outdoor–Indoor Distributions . . . 53
5.3.3 Verbal Annoyance . . . 53
5.4 Paired Wilcoxon Signed-Rank Tests . . . 60
5.5 Systematic Correlation Results . . . 62
5.5.1 Subjective–Subjective Relationships . . . 62
5.5.2 Objective–Objective Relationships . . . 67
5.5.3 Objective Descriptors and Mean Responses . . . 67
5.5.4 Individual-Response and Sample-Composition Sensitivity . . . 72
5.6 Simple Regression Models . . . 72
5.6.1 Predicting Mean Annoyance from One Objective Descriptor 72
5.6.2 Predicting the Corresponding Subjective Attribute . . . 73
5.6.3 Does Roughness Add Information to Loudness? . . . 73
5.6.4 Diagnostics and Evidential Boundaries . . . 76

6 Discussion 77
6.1 What the Study Establishes . . . 77
6.2 UAV, Manoeuvre and Event Variability . . . 77
6.3 Outdoor Exposure and Indoor Binaural Listening . . . 78
6.4 Interpreting the Paired Wilcoxon Evidence . . . 79
6.5 Correlations in the Physical Context of Drone Sound . . . 79
6.5.1 Loudness and the Common Magnitude Component . . . 79

6.5.2 Roughness and Fluctuation Strength 80
6.5.3 Sharpness and Tonality 80
6.5.4 Means and Individual Responses 81
6.6 What the Simple Models Add 81
6.7 Limitations and Priorities for Further Work 82
6.8 Synthesis 83

7 Conclusions 84
7.1 Main Findings 84
7.2 Contribution and Applicability 85
7.3 Future Development 85

A Outdoor Subjective Evaluation Questionnaire 87

B Recording Correspondence and Objective Variability 93
B.1 HDF–XL2 Correspondence 93
B.2 Objective Event Summaries 94
B.3 Preliminary Ambisonics Branch 96

C Subjective Descriptive Tables 97

D Supplementary Correlation Results 101
D.1 Individual-Response Associations 101
D.2 Complete Numerical Archive 105

E Complete Simple Regression Results 106
E.1 Diagnostic Qualification 107

Bibliografia 109

Elenco delle figure

3.1 Overview of the outdoor flight field. 28

3.2 ZYLIA ZM-1 array and adjacent NTi Audio XL2 measurement setup. The documented ZYLIA height was 1.60 m. This photograph does not document the position of the wearable binaural headset. 28

3.3 Programmed ascent, descent and horizontal-cruise geometry. Speeds are programmed values, not telemetry estimates. The microphone marker represents the ZYLIA height of 1.60 m; it does not establish the binaural headset position. 29

5.1 Objective event distributions: ISO N5, ISO N10, HM loudness and sharpness. Each point is a distinct flight event; the numbers above the groups are event counts. Boxes show interpolated quartiles and median, with whiskers to the most extreme points within 1.5 IQR. A singleton is shown only as a point. TO: take-off; LA: landing; HC: horizontal cruise. 49

5.2 Objective event distributions: roughness, fluctuation strength and tonality. The same 28-event sample and grouping conventions as Figure 5.1 apply. With one to three events per condition, the points carry more information than the apparent smoothness or width of a box. 50

5.3 Minimum left/right total-to-background margin in each one-third-octave band of the position-matched quality-control series. A margin below 3 dB in either channel prevents a reliable combined corrected value under the adopted rule. 52

5.4 Outdoor subjective distributions, numerical annoyance, loudness and fluctuation. All available respondents are retained. Boxes show median and quartiles; whiskers extend to the most extreme values within 1.5 IQR, and all individual responses are plotted. Three manoeuvres are adjacent within each UAV. Axis limits include the full response scale and graphical margins. 54

5.5 Outdoor subjective distributions, roughness, sharpness and tonality. All available respondents are retained. Boxes show median and quartiles; whiskers extend to the most extreme values within 1.5 IQR, and all individual responses are plotted. Three manoeuvres are adjacent within each UAV. Axis limits include the full response scale and graphical margins. 55

5.6 Indoor subjective distributions, numerical annoyance, loudness and fluctuation. All available respondents are retained. Boxes show median and quartiles; whiskers extend to the most extreme values within 1.5 IQR, and all individual responses are plotted. Three manoeuvres are adjacent within each UAV. Axis limits include the full response scale and graphical margins. 56

5.7 Indoor subjective distributions, roughness, sharpness and tonality. All available respondents are retained. Boxes show median and quartiles; whiskers extend to the most extreme values within 1.5 IQR, and all individual responses are plotted. Three manoeuvres are adjacent within each UAV. Axis limits include the full response scale and graphical margins. 57

5.8 Outdoor–indoor comparison restricted to the same participants for each UAV–manoeuvre condition. Orange denotes outdoor and blue indoor. There are 125 pairs from eleven returning listeners; pair counts are 7, 10 and 9 for Agras landing, take-off and cruise, and 11 elsewhere. Boxplot and point conventions match the separate-sample figures. 58

5.9 Outdoor–indoor comparison restricted to the same participants for each UAV–manoeuvre condition. Orange denotes outdoor and blue indoor. There are 125 pairs from eleven returning listeners; pair counts are 7, 10 and 9 for Agras landing, take-off and cruise, and 11 elsewhere. Boxplot and point conventions match the separate-sample figures. 59

5.10 Verbal annoyance by UAV and manoeuvre, using the full sample in each environment. Each bar totals 100% of the available responses for that condition; categories are ordered from not at all to extremely annoyed. The original Italian wording is reproduced in Appendix A. 63

5.11 Verbal annoyance in the matched sample. Each adjacent Out/In pair uses the same respondents for one UAV–manoeuvre condition. Denominators are seven for Agras landing, ten for Agras take-off, nine for Agras cruise and eleven for all other conditions. 64

5.12 Outdoor subjective–subjective matrix of twelve condition means. Cells contain Spearman rho and the nominal two-sided p-value. Means use all available respondents per condition; they are not 156 independent acoustic observations. 65
5.13 Indoor subjective–subjective matrix of twelve condition means. Cells contain Spearman rho and the nominal two-sided p-value. Means use all available respondents per condition; they are not 156 independent acoustic observations. 66
5.14 Objective–objective Spearman matrix using all 28 independent HDF events. Unequal repetition counts are retained. Cells show rho and nominal p; the observations are grouped within twelve conditions and four UAVs. 68
5.15 Selected objective–objective associations across all 28 events. Colours identify UAVs and marker shapes identify manoeuvres. Each point is one physical event. Coefficients and nominal probabilities are the corresponding cells of the extended objective matrix; the grouping shows why a pooled association need not describe a within-UAV causal relationship. 69
5.16 Outdoor objective–subjective Spearman matrix. One row of the analysis dataset corresponds to one of the twelve evaluated recordings and its condition-mean responses. Additional flight repetitions are excluded from this cross-domain analysis. Cells give rho and nominal p. 70
5.17 Indoor objective–subjective Spearman matrix. One row of the analysis dataset corresponds to one of the twelve evaluated recordings and its condition-mean responses. Additional flight repetitions are excluded from this cross-domain analysis. Cells give rho and nominal p. 71
5.18 HM loudness versus mean numerical annoyance and mean subjective loudness, separately outdoors and indoors. Each point is one evaluated UAV–manoeuvre condition; colour identifies UAV and marker identifies manoeuvre. Vertical bars show the sample standard deviation of the ratings in each condition, not uncertainty of the fitted mean. Lines are OLS fits over the observed predictor range. Reported R-squared and p-values concern in-sample linear models, not external predictive validation. 74

D.1 Extended Spearman matrix of objective HDF and XL2 descriptors across 28 physical events. Each off-diagonal cell displays the coefficient and nominal p-value. 102

D.2 Outdoor XL2–subjective matrix of twelve condition means. Cells give Spearman rho and the nominal two-sided p-value. 103
D.3 Indoor XL2–subjective matrix of twelve condition means. Cells give Spearman rho and the nominal two-sided p-value. 104

Capitolo 1

Introduction

1.1 Motivation

The acoustic impact of unmanned aerial vehicles (UAVs) concerns both the magnitude of sound exposure and the way that sound is perceived. Multirotor aircraft produce tonal components, broadband contributions and temporal fluctuations that change with operating condition and source–receiver geometry. Their use near people therefore raises questions that cannot be addressed by describing an aircraft with a single sound level alone. Reviews of UAV acoustics and perception identify loudness, spectral balance, tonality and modulation as relevant features, while also highlighting the influence of the listening context [1, 2].

Dynamic flight adds a further difficulty. An ascending, descending or passing UAV changes its relationship to a stationary observer over time. The duration and acoustic trajectory of an event can affect an evaluation, and successive executions of a nominally identical manoeuvre need not yield identical recordings. A meaningful assessment must therefore connect the physical event, the acoustic descriptors and the response of the listener without treating these as interchangeable levels of observation.

Listening experiments offer a way to investigate these relationships. Outdoor evaluation preserves the live setting, while recorded stimuli allow the same selected material to be presented subsequently indoors. Binaural acquisition retains two ear-related signals and provides a practical basis for headphone presentation. Nevertheless, changing from a live event to playback also changes the visual scene, room, session and listening perspective. Agreement or disagreement between the two sets of ratings must be interpreted in relation to this complete change of context.

1.2 Aim and Research Questions

The aim of this thesis is to characterise the acoustic and perceptual responses associated with four UAV platforms during three dynamic manoeuvres, and to relate outdoor judgements to subsequent binaural indoor judgements and objective descriptors of the evaluated events. The work addresses four questions:

1. How do the measured acoustic and psychoacoustic descriptors vary across UAVs, manoeuvres and the available independent repetitions?

2. How are numerical annoyance, verbal annoyance and five subjective sound attributes distributed outdoors and indoors, and what differences occur when the same participants are compared for the same condition?

3. Which associations emerge among subjective ratings, among objective metrics, and between the objective descriptors and the responses to the evaluated events?

4. Can the most relevant associations be expressed through simple regression equations, and does combining loudness with roughness provide useful additional information within the observed dataset?

The fourth question concerns transparent empirical models. It does not assume that a model fitted to four aircraft has already been validated for new aircraft or operational settings.

1.3 Experimental and Analytical Approach

Four dataset-labelled platforms—Agras 25, Matrice 350, Matrice 4E and Mavic Mini 5—were assessed during take-off, landing and horizontal cruise. The outdoor campaign took place near Turin on 1 July 2026. Field recordings included a conventional sound-level-meter chain and a separately acquired binaural chain. The indoor sessions took place in September 2026 using the selected binaural recordings and a fixed headphone playback configuration.

Fourteen participants supplied outdoor responses and thirteen completed the indoor experiment. Eleven took part in both, yielding 125 complete condition-specific pairs. The wider acoustic archive contains 28 distinct flight events, with unequal numbers of repetitions across conditions. Only the first evaluated event in each of the twelve UAV–manoeuvre conditions was associated with subjective responses; the other events inform acoustic variability and objective–objective correlations.

The analysis begins with distributions and measurement traceability. Paired Wilcoxon signed-rank tests then compare outdoor and indoor numerical ratings

within each condition, reporting medians and original two-sided probabilities. Systematic correlation matrices examine the three domains of association, with condition means as the principal subjective summaries and individual-response associations as supplementary information. Finally, simple and two-predictor regressions express selected relationships in interpretable equations.

1.4 Contribution and Scope

The contribution is a connected case study that follows the same set of evaluated UAV events from field acquisition to subjective assessment, objective characterisation and simple explanatory modelling. The traceable HDF–XL2 map, explicit separation of evaluated events from independent repetitions, and separation of full samples from paired samples are central to this contribution.

The study provides evidence about the tested recordings and listeners. It does not establish causal effects of individual sound qualities, long-term community annoyance, general equivalence of outdoor exposure and binaural playback, or prediction accuracy for aircraft outside the dataset. These boundaries are particularly important when strong condition-level correlations are observed in a small collection of related flight conditions.

1.5 Thesis Organisation

Chapter 2 reviews UAV sound generation, psychoacoustic attributes, subjective response and binaural reproduction. Chapter 3 describes the campaign, participants, equipment and correspondence between measurements. Chapter 4 specifies signal processing, background checks and statistical methods. Chapter 5 presents the distributions, paired comparisons, correlations and final regression results. Chapter 6 interprets these findings in their acoustic and methodological context. Chapter 7 answers the research questions and identifies the next steps supported by the study. Appendices reproduce the outdoor questionnaire and provide detailed numerical summaries and traceability information.

Capitolo 2

Literature Review

2.1 UAVs and the Emerging Noise Problem

Unmanned Aerial Vehicles (UAVs), commonly referred to as drones, have undergone rapid technological development in recent years. Improvements in electric propulsion, battery technology, autonomous navigation and control systems have progressively extended their use beyond recreational applications. UAVs are currently considered for a broad range of civil and commercial activities, including aerial photography, environmental monitoring, infrastructure inspection, surveillance, agriculture, emergency services and parcel delivery[1, 2].

The increasing integration of UAVs into everyday activities, particularly within urban and peri-urban environments, introduces several challenges related to safety, regulation, airspace management, privacy and public acceptance. Among these factors, noise represents one of the main environmental concerns associated with the large-scale deployment of UAV operations [2]. Unlike conventional aircraft, which generally operate at relatively high altitudes except in the vicinity of airports, many UAV applications require operations at considerably lower altitudes and potentially in close proximity to residential areas and exposed communities.

Although small electric UAVs generally produce lower absolute sound levels than conventional aircraft, their acoustic impact cannot be evaluated only in terms of emitted sound pressure level. UAVs generate an unconventional acoustic signature characterized by spectral and temporal features that differ considerably from those associated with conventional transportation noise sources [1, 2]. Consequently, the perceptual impact of drone noise may be disproportionate to its absolute sound level.

Previous studies have shown that drone noise can produce greater perceived annoyance than conventional transportation noise at comparable acoustic exposure levels. In particular, comparisons between UAV and road-vehicle noise have shown

that conventional A-weighted metrics alone may not fully account for the observed differences in subjective response [3]. This indicates that characteristics beyond overall sound energy contribute to the perception of UAV noise.

The distinctive tonal components and high-frequency broadband content typically associated with multirotor UAVs have been identified as possible contributors to their increased perceptual salience and annoyance [1]. Furthermore, the repetitive or persistent nature of certain drone operations, together with their relatively low operating altitude, may increase their noticeability within an existing acoustic environment.

The acoustic context in which a UAV operates also plays an important role. Torija et al. demonstrated that the perceptual impact of the same hovering UAV noise depends strongly on the characteristics of the underlying urban soundscape [3]. When the background environment was already dominated by road traffic, the introduction of UAV noise produced comparatively small changes in perceived loudness, annoyance and pleasantness. Conversely, in quieter urban soundscapes, the same UAV noise resulted in substantially larger perceptual changes.

These observations highlight an important limitation of assessing UAV noise exclusively through conventional energetic descriptors. While metrics such as A-weighted equivalent sound pressure level provide essential information on acoustic exposure, they do not explicitly describe several spectral and temporal characteristics that influence human perception. This has motivated a growing body of research focused on the psychoacoustic characterization of UAV noise and on the relationship between objective acoustic parameters and subjective response.

Understanding the acoustic mechanisms responsible for these perceptual effects therefore represents a necessary first step towards a comprehensive assessment of drone noise. The following section examines the principal sound-generation mechanisms and the spectral, temporal and directional characteristics of multirotor UAV noise.

2.2 Acoustic Characteristics of Drone Noise

The acoustic signature of a multirotor UAV results from the interaction of several aerodynamic and electromechanical sound-generation mechanisms. Unlike conventional fixed-wing aircraft, multirotor UAVs rely on multiple electrically driven propellers operating simultaneously to generate lift, control attitude, and produce translational motion. Consequently, changes in operating condition and flight manoeuvre are associated with variations in rotor operating state and, therefore, with changes in the emitted sound [1].

The dominant acoustic contributions of small multirotor UAVs are generally associated with the rotating propellers and their electric motors. Rotor noise

contains both tonal and broadband components produced by different physical mechanisms, resulting in a distinctive spectral and temporal signature [4, 5].

2.2.1 Tonal Components and Blade Passing Frequency

One of the most distinctive characteristics of multirotor UAV noise is the presence of pronounced tonal components associated with the periodic motion of the rotor blades. A fundamental quantity for describing this phenomenon is the blade passing frequency (BPF), which depends on the rotor rotational speed and on the number of blades.

If the rotational speed is expressed in revolutions per minute (RPM), the rotational frequency can be written as

$$f_{\mathrm{rot}} = \frac{N}{60}, \tag{2.1}$$

where N is the rotational speed in RPM. The corresponding blade passing frequency is

$$f_{\mathrm{BPF}} = N_b f_{\mathrm{rot}} = \frac{N_b N}{60}, \tag{2.2}$$

where N_b represents the number of blades.

Periodic aerodynamic loading and blade-thickness effects generate tonal radiation at the BPF and at its harmonics. For small UAV propellers, the first two or three harmonics of the BPF can constitute prominent components of the measured spectrum [4].

The acoustic signature becomes more complex in a multirotor configuration, as several rotors operate simultaneously. Small differences in rotational speed can produce closely spaced tonal components, while aerodynamic interactions among adjacent rotors, the vehicle body, and the surrounding flow can further modify the spectral structure. As a consequence, the radiated sound cannot generally be represented as the simple harmonic spectrum of an isolated propeller.

The presence of strong tonal components is particularly relevant from a perceptual perspective. Tonality has repeatedly been identified as one of the characteristics that distinguish drone noise from more conventional environmental noise sources and may contribute to its perceptual salience and annoyance [1].

2.2.2 Broadband Noise

In addition to discrete tonal components, UAV noise contains broadband contributions associated with aerodynamic phenomena and with the propulsion system.

Turbulent inflow and interactions between the blades and turbulent flow can contribute to broadband or quasi-tonal sound, while additional broadband components may originate from the electric propulsion system [4, 5].

The spectra of multirotor UAVs therefore typically contain low- and mid-frequency tonal and harmonic components associated with rotor operation, together with broadband energy extending towards higher frequencies [1].

This combination is particularly relevant for human perception. High-frequency spectral content can increase the perceptual prominence of the source and is not explicitly represented by an overall sound pressure level alone. Metrics designed to characterize spectral balance and tonal content, such as sharpness and tonality, can therefore provide complementary information when evaluating drone noise.

2.2.3 Temporal Characteristics and Modulation

Drone noise is inherently time-varying. Even during nominally stationary operations, small variations in rotor speed and aerodynamic loading can produce changes in both amplitude and frequency. These effects become more pronounced during manoeuvring flight, when the propulsion system continuously modifies the operating condition of individual rotors in order to control aircraft attitude and trajectory [1].

Temporal variations are relevant from a psychoacoustic perspective because the auditory system is sensitive not only to the overall magnitude of a sound but also to its modulation characteristics. Slow amplitude variations can be associated with fluctuation strength, whereas more rapid variations are related to roughness. Both quantities have consequently been considered in previous investigations of UAV sound quality [6, 5].

Rotor interactions may also produce beating phenomena and complex modulation patterns. Torija and Nicholls found fluctuation strength to be associated with perceived responses in their experimental investigation and suggested that the perceptual effect of rotor interaction may contribute to this relationship [6].

2.2.4 Acoustic Directivity

Multirotor UAVs cannot generally be treated as perfectly omnidirectional sound sources. Experimental investigations have demonstrated a pronounced dependence of the radiated sound on the relative direction between the aircraft and the receiver, particularly in the vertical plane [1, 5].

The radiation pattern is also frequency dependent. The combined effects of individual propeller directivity, interactions among multiple rotors, and scattering from the aircraft structure result in a complex spatial distribution of the emitted sound. Consequently, during a dynamic flight event, a stationary listener is exposed

not only to changes in source–receiver distance but also to different regions of the UAV radiation pattern.

Alkmim et al. investigated drone-noise directivity through a hemispherical microphone array and represented the measured sound field by means of a spherical harmonic decomposition [5]. Their results demonstrate the importance of spatially resolved measurements for the characterization of UAV noise and establish a relevant connection between drone directivity, spherical harmonic representations, and sound-quality analysis.

2.2.5 Influence of Flight Conditions

The acoustic emission of a multirotor UAV depends strongly on the operating condition. Drone model, payload, rotor speed, aircraft attitude, and flight manoeuvre can modify both the magnitude and the spectral characteristics of the radiated noise [1].

During hover, the propulsion system operates under approximately stationary conditions in order to balance the aircraft weight. Forward flight requires changes in aircraft attitude and differences in the rotational speeds of the rotors. The resulting modification of aerodynamic loading and rotor interaction can alter the emitted acoustic signature.

Vertical manoeuvres introduce additional differences. During climb, the increased thrust demand is associated with changes in rotor operating conditions. During descent, unsteady aerodynamic conditions may cause variations in the rotational speeds of the individual rotors, producing fluctuating sound-pressure levels and time-varying tonal components [1].

The importance of studying realistic operating conditions has also been highlighted by field-measurement investigations. Ramos-Romero et al. showed that UAV acoustic characteristics depend on vehicle configuration, operating condition, and environmental factors, and emphasized the relevance of on-field measurements for characterizing UAV operations under realistic conditions [7].

More recent investigations of manoeuvring drone noise have further shown that changes in aircraft operation and orientation can substantially modify the tonal content of the emitted sound. In particular, variations in rotor speed during different flight phases can produce corresponding variations in blade-passing-frequency components [8].

These findings indicate that measurements performed exclusively under static or hovering conditions cannot necessarily describe the complete acoustic behaviour of a UAV during realistic operation. This aspect is particularly relevant when take-off, landing, and horizontal flight are considered as distinct acoustic events.

2.3 Acoustic Measurement and Characterization of UAV Noise

The characterization of UAV noise presents several experimental challenges. The sound emitted by a multirotor aircraft depends not only on the vehicle itself, but also on its operating condition, source–receiver geometry, acoustic environment, and atmospheric conditions. Consequently, the measurement methodology adopted in an experiment can substantially affect the information that can be extracted from the acquired signals.

Previous UAV-noise studies have employed a wide variety of measurement configurations, ranging from highly controlled laboratory experiments to outdoor field measurements under realistic flight conditions [1, 7].

2.3.1 Laboratory Measurements

Laboratory measurements provide a high degree of experimental control. Anechoic and semi-anechoic environments can minimize unwanted reflections and external background noise, allowing the acoustic characteristics of the source to be investigated under repeatable conditions. Such configurations are particularly useful for analysing individual sound generation mechanisms, the influence of rotor speed, and source directivity.

Several studies have used fixed or hovering UAVs in combination with multiple microphones distributed at different angular positions. These configurations allow the radiation pattern of the source to be estimated while maintaining well-controlled source and receiver positions [1].

Laboratory measurements nevertheless involve an important limitation: the operational and propagation conditions may differ from those encountered during actual outdoor flight. Fixing the aircraft, constraining its motion, or reproducing only selected operating states may prevent some aerodynamic and dynamic effects from being represented accurately.

This produces a general trade-off between experimental control and ecological validity. Laboratory experiments provide greater control and repeatability, whereas field measurements provide greater realism but are more exposed to environmental variability [1].

2.3.2 Field Measurements

Outdoor field measurements allow UAV noise to be acquired during realistic flight operations. This is particularly important when investigating dynamic events such as take-off, landing, or flyover, for which the source position, orientation, rotor operating state, and source–receiver distance vary continuously.

Ramos-Romero et al. proposed an on-field framework for the acoustic characterization of multirotor small UAVs based on measurements obtained during realistic flyover operations [7]. Their work considered conventional noise metrics, frequency-domain characteristics, and directivity, while also monitoring relevant operational and environmental parameters.

Field measurements, however, introduce sources of uncertainty that are largely absent under laboratory conditions. Background noise, wind, ground effects, atmospheric propagation, aircraft trajectory, and variations in operating conditions may influence the recorded signal. For this reason, environmental conditions and background-noise levels should be documented when interpreting outdoor measurements.

The influence of wind is particularly relevant because the interaction between atmospheric flow and rotating UAV components can modify acoustic emission. In addition, wind affects sound propagation between source and receiver. Ramos-Romero et al. therefore included wind speed and direction among the environmental parameters monitored during their field campaign [7].

2.3.3 Single-Channel and Multichannel Measurements

The complexity of the measurement system determines the type of acoustic information that can be obtained. A single sound-level measurement position can provide quantities such as sound-pressure-level time histories, maximum levels, equivalent levels, or sound-exposure levels. These descriptors are useful for quantifying acoustic exposure but provide limited information about the spatial structure of the sound field.

Using multiple independent microphones at different locations extends the measurement to several source–receiver geometries and can provide information about directivity. More advanced phase-synchronized microphone arrays can additionally be employed for source localization, beamforming, sound-field reconstruction, or spatial analysis [1].

This distinction is particularly relevant for UAVs because of their directional radiation characteristics. Measurements at a single position cannot completely describe a source whose spectral content and radiation strength vary with observation angle.

Alkmim et al., for example, employed a hemispherical microphone array to capture the sound field around a drone and subsequently represented the radiation through spherical harmonics [5]. Such approaches provide information that cannot be obtained from conventional single-channel level measurements alone.

2.3.4 Measurement of Dynamic Flight Events

Dynamic flight events introduce an additional challenge because several variables evolve simultaneously. During a flyover or manoeuvre, the source–receiver distance changes continuously, the radiation angle varies, and the propulsion system adjusts to maintain the required trajectory.

For this reason, metrics that integrate the complete acoustic event, such as sound exposure level, can be useful for characterizing flyovers, while time histories preserve information about the temporal development of the event. In contrast, equivalent continuous levels are commonly used for more stationary operating conditions such as hover [1].

The analysis of complete flight events is especially important when psychoacoustic perception is considered. A listener experiences the temporal evolution of the drone sound rather than a single stationary acoustic value. The measurement methodology should therefore retain sufficient temporal, spectral, and, when relevant, spatial information to support the subsequent perceptual analysis.

2.4 From Conventional Acoustic Metrics to Psychoacoustic Assessment

Traditional environmental-noise assessment is largely based on physical descriptors of acoustic exposure. Metrics derived from sound pressure, including A-weighted equivalent sound pressure level and sound exposure level, provide established and reproducible methods for quantifying the magnitude of an acoustic event.

These metrics remain essential for the objective characterization of UAV noise. However, increasing evidence suggests that they may not be sufficient to explain the complete subjective response to this particular source [1, 6].

Comparisons between UAV noise and conventional transportation noise have shown that sounds producing comparable values of conventional acoustic metrics may nevertheless result in different annoyance responses. Such differences indicate that subjective evaluation depends not only on the overall acoustic energy but also on specific spectral and temporal characteristics of the sound.

Schäffer et al. concluded that differences in annoyance at equivalent A-weighted exposure levels cannot be fully represented by conventional metrics such as L_{AE} or L_{Aeq} [1]. This limitation is particularly important for UAVs because their acoustic signatures contain strong tonal components, high-frequency content, and temporal modulations.

Torija and Nicholls similarly noted that A-weighted quantities do not explicitly account for several sound characteristics that influence the reported annoyance of

drone noise [6]. Their work therefore investigated both conventional noise metrics and sound quality metrics in relation to subjective responses.

2.4.1 Perceived Noise Level

Perceived Noise Level (PNL) represents one approach for incorporating the frequency-dependent response of human hearing into the assessment of aircraft-related noise. Compared with a simple broadband sound-pressure level, PNL places greater emphasis on frequency regions that contribute more strongly to perceived noisiness.

In the subjective experiment conducted by Torija and Nicholls, PNL showed a strong association with perceived annoyance and perceived loudness. However, the authors also found that additional psychoacoustic variables were required to characterize other perceptual aspects of drone sound [6].

This illustrates a central issue in UAV-noise assessment: metrics related to overall perceived magnitude can explain an important proportion of the subjective response, but the distinctive spectral and temporal features of the source may require additional descriptors.

2.4.2 Sound Quality Metrics

Psychoacoustic or sound quality metrics aim to quantify specific auditory sensations rather than only the physical magnitude of an acoustic signal. Commonly investigated descriptors include loudness, sharpness, tonality, roughness, and fluctuation strength [1, 6, 5].

These metrics provide complementary information because each represents a different perceptual dimension. Loudness is primarily associated with perceived sound magnitude, sharpness with the relative contribution of high-frequency components, tonality with the prominence of tonal structures, and roughness and fluctuation strength with different ranges of temporal modulation.

Their relevance to drone noise follows directly from the acoustic characteristics discussed in Section 2.2. For example, the high-frequency components of UAV spectra can influence sharpness, BPF-related components contribute to tonal perception, and rotor interaction can produce amplitude variations associated with fluctuation strength or roughness.

Consequently, psychoacoustic metrics provide a conceptual link between physical acoustic measurements and subjective listener response.

2.4.3 Relationship with Subjective Annoyance

Experimental studies have investigated which objective descriptors are most strongly associated with perceived drone-noise annoyance. Torija and Nicholls reported

statistically significant relationships between perceived annoyance and several quantities, with PNL, loudness, and sharpness showing particularly strong correlations in their dataset [6].

Their regression analysis identified PNL and sharpness as the principal predictors of perceived annoyance for the drone sounds considered, while PNL and fluctuation strength were associated with perceived loudness. Sharpness, roughness, and Aures tonality contributed to the modelling of perceived pitch [6].

More recent investigations have reached broadly compatible conclusions while also highlighting that the relative importance of individual psychoacoustic metrics may depend on the aircraft and operating condition. König et al. investigated annoyance in relation to technical, operational, and psychoacoustic variables and found that reductions in loudness, sharpness, tonality, and roughness or fluctuation strength were associated with improvements in perceived annoyance [9].

The use of psychoacoustic descriptors does not imply that conventional acoustic levels become irrelevant. Rather, current research indicates that the two approaches should be considered complementary. Physical quantities describe the acoustic exposure, while psychoacoustic quantities can help explain how specific properties of that exposure are perceived by a listener.

This combined perspective is particularly appropriate for UAV noise, where a direct relationship between physical emission and human response cannot always be represented through a single conventional noise metric.

2.5 Psychoacoustic Metrics for Drone Noise

Psychoacoustic metrics attempt to describe specific auditory sensations associated with a sound rather than only its physical magnitude. This is particularly relevant for drone noise, whose perception is influenced by a combination of level, spectral content, tonal components, and temporal modulation.

Among the descriptors most frequently considered in UAV-noise studies are loudness, sharpness, tonality, roughness, and fluctuation strength. These metrics provide complementary information and can be related to different features of the acoustic signature discussed in Section 2.2 [1, 6, 5].

A computed psychoacoustic descriptor and a questionnaire response with a similar name remain different variables. For example, calculated loudness is a signal-derived quantity, whereas perceived loudness in a questionnaire is a participant's rating on the specified response scale. Their association is an empirical question, not an identity. The same distinction applies to sharpness, roughness, fluctuation strength and tonality.

Comparisons also require a consistent definition of the computational model, channel treatment, analysed interval and temporal summary. A mean, a maximum

and an exceedance percentile of a time-varying descriptor represent different aspects of an event and should not be used interchangeably. The exact implementations used here are specified in the analysis-methods chapter.

2.5.1 Loudness

Loudness represents the perceived magnitude of a sound. Unlike sound pressure level, which quantifies the physical amplitude of pressure fluctuations, loudness attempts to account for the frequency-dependent response of the human auditory system.

Loudness is commonly expressed in sones. A larger loudness value indicates that the corresponding sound is perceived as louder by a listener. Zwicker loudness can be calculated according to ISO 532-1 [5].

The relevance of loudness to UAV-noise assessment has been demonstrated in several subjective investigations. Hui et al. reported a strong relationship between annoyance ratings and loudness-related metrics in their listening experiment [4]. Similarly, Torija and Nicholls found loudness to be strongly correlated with perceived annoyance and perceived loudness [6].

More recent work by König et al. also identified loudness as one of the most relevant psychoacoustic descriptors for predicting drone-noise annoyance, although the authors emphasized that the relative importance of individual metrics depends on the specific metric implementation and experimental condition [9].

2.5.2 Sharpness

Sharpness describes the perceptual contribution of high-frequency energy to the overall sound. A sound with a greater proportion of energy at high frequencies is generally perceived as sharper or more piercing.

Sharpness is commonly expressed in acum. Its calculation is based on a frequency-weighted distribution of specific loudness, with greater weight assigned to higher critical-band rates [5, 10].

This descriptor is particularly relevant for multirotor UAVs because their spectra often contain substantial high-frequency broadband content. Consequently, two sounds with similar overall level may differ in sharpness and may therefore produce different perceptual responses.

Torija and Nicholls found sharpness to be one of the principal variables associated with perceived annoyance in their experiment. Their regression model identified PNL and sharpness as the main predictors of perceived annoyance for the tested drone sounds [6].

2.5.3 Tonality

Tonality describes the perceptual prominence of discrete tonal components within a more complex sound. It is therefore directly related to the tonal structures generated at the blade passing frequency and its harmonics in multirotor UAV noise.

Different computational models of tonality exist, and the numerical value obtained can depend significantly on the adopted method. Alkmim et al., for example, considered a hearing-model-based tonality metric in their psychoacoustic analysis of drone radiation [5].

The relevance of tonality to annoyance is not necessarily identical across all UAV types and operating conditions. König et al. showed that different tonality formulations may provide different levels of predictive performance and emphasized that the particular definition or standard used for a psychoacoustic metric should be reported explicitly [9].

Nevertheless, tonal content remains an important characteristic of drone noise because prominent blade-passing-frequency components contribute to its distinctive and recognizable acoustic signature.

2.5.4 Roughness

Roughness describes the auditory sensation generated by relatively rapid temporal variations in a sound. It is commonly expressed in asper and is associated with amplitude-modulation components at higher modulation frequencies than those primarily described by fluctuation strength.

Rough sounds may be perceived as more aggressive or unpleasant even when their overall loudness does not change significantly [5].

For UAV noise, roughness can be influenced by interactions among the tonal components generated by multiple rotors, variations in rotational speed, and non-stationary operating conditions. These effects may become particularly relevant during dynamic manoeuvres.

Torija and Nicholls found roughness to contribute to the prediction of perceived pitch in their dataset, while König et al. observed relationships between roughness and annoyance under specific experimental conditions [6, 9].

2.5.5 Fluctuation Strength

Fluctuation strength describes the perception of relatively slow amplitude variations that can be individually distinguished by the auditory system. It is commonly expressed in vacil and is particularly sensitive to low-frequency modulation.

The sensation reaches high sensitivity for modulation frequencies in the region of a few hertz. In the formulation discussed by Vaiuso et al., the analysis covers

low modulation frequencies and gives particular relevance to modulation around 4 Hz [10].

Fluctuation strength is relevant to multirotor noise because small differences in rotor speed and interactions among multiple propellers can generate beating and slowly varying amplitude patterns.

Torija and Nicholls reported statistically significant associations between fluctuation strength and subjective responses and identified it as one of the predictors of perceived loudness in their regression analysis [6].

2.5.6 Psychoacoustic Annoyance

Psychoacoustic Annoyance (PA) is a combined metric intended to represent the overall unpleasantness of a sound by integrating several auditory sensations rather than considering them independently.

The Zwicker–Fastl formulation combines loudness with contributions related to sharpness, fluctuation strength, and roughness. A commonly reported form can be expressed as

$$PA = N_5 \left(1 + \sqrt{w_S^2 + w_{FR}^2}\right), \tag{2.3}$$

where N_5 is the loudness exceeded for 5% of the analysed time and w_S and w_{FR} represent terms associated respectively with sharpness and with fluctuation strength/roughness [10].

The use of a combined psychoacoustic index is appealing for UAV-noise assessment because annoyance is unlikely to depend on a single acoustic feature. However, PA should not be regarded as equivalent to the annoyance rating reported directly by a participant. The former is a model-based objective descriptor derived from an acoustic signal, whereas the latter is a subjective response collected through a perceptual experiment.

This distinction is particularly important in the present work, where objective acoustic and psychoacoustic descriptors are intended to be compared with subjective ratings rather than treated as interchangeable.

2.6 Human Response and Noise Annoyance

Noise annoyance is one of the principal subjective outcomes considered in environmental noise research. It represents a negative reaction to an acoustic exposure and may include sensations of disturbance, irritation, dissatisfaction, or nuisance.

For UAV noise, annoyance is particularly relevant because the acoustic signature is relatively unfamiliar and because future operations may expose people to repeated

low-altitude events in environments that are not currently affected by conventional aircraft noise [2].

Experimental evidence suggests that perceived annoyance cannot be predicted only from absolute sound level. Comparisons with road traffic and aircraft noise have repeatedly shown that UAV sounds may be judged as more annoying at comparable values of conventional acoustic exposure metrics [1, 3].

This difference has been attributed to several characteristics of UAV noise, including tonal components, high-frequency content, temporal modulation, and the noticeability of the source within the surrounding soundscape. Consequently, annoyance should be interpreted as the result of both acoustic and non-acoustic factors.

2.6.1 Influence of Acoustic Context

The surrounding acoustic environment can substantially affect the perception of drone noise. Torija et al. investigated the effect of introducing a hovering UAV into different urban soundscapes and found that the same UAV sound produced different perceptual responses depending on the background environment [3].

In soundscapes dominated by road traffic noise, the addition of the UAV produced relatively smaller perceptual changes. In quieter environments, the same drone sound became more salient and produced larger increases in reported annoyance.

This suggests that the perceptual impact of a drone is partly related to its emergence from the acoustic background. The relationship between source level and background level, as well as the spectral contrast between the two, can influence how noticeable and intrusive a UAV event becomes.

2.6.2 Non-Acoustic Factors

Subjective annoyance may also vary according to individual and contextual factors. Previous environmental-noise research has identified variables such as noise sensitivity, attitude towards the source, familiarity, age, and situational context as possible moderators of subjective response [2].

The importance of these variables is not necessarily identical across studies. König et al., for example, did not observe substantial differences in the fundamental relationships between annoyance and psychoacoustic metrics across the sociocultural groups considered in their large-scale user study [9].

Zhang et al., on the other hand, found that age and familiarity with the target noise could influence annoyance ratings in their laboratory experiments, while other investigated individual variables produced smaller or non-significant differences [11].

These findings indicate that subjective response should not automatically be interpreted as a direct measurement of a physical acoustic quantity. Participant characteristics and experimental context should therefore be documented and, where appropriate, considered during the statistical analysis.

2.7 Listening Tests for Drone Noise Assessment

Listening tests provide a controlled method for investigating the relationship between acoustic stimuli and subjective perception. In a typical experiment, participants are exposed to a set of recorded, synthesized, or reproduced sounds and are asked to rate one or more perceptual attributes.

For UAV research, listening tests have been used to evaluate annoyance, perceived loudness, pitch, distraction, pleasantness, and other perceptual responses under controlled conditions [4, 6, 12].

2.7.1 Rating Scales

The choice of rating scale is an important methodological component of a listening experiment. Noise-annoyance studies have employed verbal category scales, numerical category scales, and continuous or quasi-continuous rating methods.

Standardized questions are particularly useful because they improve the comparability of results among studies. The work of Fields et al. was developed specifically to improve international comparability in community-noise annoyance surveys and recommended standardized general noise-reaction questions [13].

For laboratory listening tests, however, several alternative scales remain in use. Zhang et al. compared different rating-scale formats and reported a very high correlation between mean annoyance values obtained with a 0–100 linear scale and a 0–10 category scale [11].

These results suggest that multiple rating formats can provide consistent relative evaluations, provided that the scale and experimental procedure are clearly defined.

2.7.2 Participants and Experimental Design

Participant selection represents another important methodological issue. Listening-test results may be affected by hearing ability, familiarity with the sound source, age, experience in acoustics, and other individual characteristics.

Zhang et al. emphasize that listeners effectively act as the measurement system in subjective experiments and therefore recommend careful consideration of participant characteristics and sample size [11].

Experimental design can also introduce bias through presentation order. Randomization or counterbalancing is commonly used to reduce serial-position, learning, or fatigue effects when several stimuli are evaluated.

The systematic review by Schäffer et al. identified lack of playback randomization and inadequate treatment of repeated participant observations as methodological limitations in some previous drone-noise studies [1].

2.7.3 Reproduction of Drone-Noise Stimuli

UAV listening studies have used different acquisition and reproduction configurations. These choices determine which features of the original acoustic event can be presented to participants. The systematic review by Schäffer et al. identified simplified spatial presentation as a limitation in some earlier experiments [1].

Hui et al. used spherical-array recordings and third-order encoding for reproduction through a 16-channel loudspeaker system [4]. This provides an example of sound-field reproduction in UAV research. Binaural headphone presentation instead targets ear-specific signals, making the recording perspective and headphone reproduction chain central to the experimental description. The following section develops this approach because it was used in the indoor listening tests of the present study.

2.8 Binaural Recording and Headphone Reproduction

Binaural recording and reproduction provide a framework for retaining information associated with the two ears during the acquisition and presentation of an acoustic event. This is relevant to UAV listening experiments because a moving aircraft produces changes in direction as well as in level, spectrum and temporal structure. In the present study, binaural headphone playback forms the basis of the indoor experiment; its opportunities and limitations therefore require explicit consideration.

2.8.1 Spatial Hearing and Ear-Specific Signals

Spatial hearing uses differences in arrival time and level between the ears, together with direction-dependent spectral filtering by the external ears and head. These cues vary with source position. Binaural reproduction seeks to retain such ear-specific information rather than merely distribute a common signal between two channels [14].

Head-related transfer functions describe the filtering between a source and an ear for a specified geometry. Wightman and Kistler demonstrated a headphone-simulation procedure based on measured free-field-to-eardrum and headphone-to-eardrum transfer functions. Their work establishes the importance of considering both the desired ear signal and the transfer introduced by the reproduction system [15].

2.8.2 Direct Binaural Acquisition

Direct binaural acquisition records two channels at ear-related positions. The resulting material contains the acoustic event as captured from that recording perspective. In the present campaign, the acquisition equipment has been identified as a HEAD acoustics BHS II binaural headset connected to a SQuadriga III recorder. The system description, installation and available technical specifications are reported in the experimental methods chapter.

This distinction resolves an important issue in describing the experiment: the binaural material was acquired separately from the spherical-array recordings. It was not obtained by decoding the Ambisonics files. Moreover, a wearable binaural headset should not be described as a dummy-head microphone. The acquisition geometry must be documented using information about the actual recording arrangement, rather than inferred from the file format or the manufacturer's description of possible applications.

The two-channel representation also needs to be distinguished from a complete spatial sound-field reconstruction. For this study, the relevant experimental object is the selected recorded event presented over headphones. Claims about arbitrary changes in listening position or orientation would require a reproduction procedure beyond the documented playback of those recordings.

2.8.3 Headphone Transfer and Playback Level

The headphone transfer function forms part of the reproduction chain. Wightman and Kistler's synthesis method explicitly included measurement of this transfer, illustrating why a binaural signal cannot be assessed independently of the system used to deliver it to the ears [15].

For the present experiment, Sennheiser HD 560S headphones were used for indoor listening. Their technical specifications belong in the equipment table in the methods chapter. A manufacturer's frequency range or sensitivity rating describes the product under specified measurement conditions; it does not establish the sound pressure level delivered to an individual participant during the test. Consequently, equipment specifications and the documented playback settings must be reported separately.

The present work also distinguishes the acoustic segments submitted to objective analysis from the reproduction chain through which participants heard them. The confirmed use of the same selected segments supports stimulus traceability. It does not, by itself, constitute an independent verification of the sound pressure at each listener's ears.

2.8.4 Externalization and Head Movement

Externalization describes hearing a sound as originating outside the head. In controlled experiments, Brimijoin et al. found that consistency between head movement and the apparent source position affected externalization. Headphone signals that remained fixed relative to the world were more frequently externalized than signals that moved with the listener's head [14].

Accordingly, binaural playback should not automatically be described as perceptually identical to field exposure. The present experiment did not include a dedicated localization or externalization assessment. Its outcomes concern annoyance and the rated perceptual attributes, so the study cannot establish spatial fidelity from those ratings alone.

2.8.5 Outdoor Exposure and Indoor Listening

Outdoor exposure and indoor playback differ in more than their technical means of delivering sound. The earlier literature on UAV soundscapes shows that acoustic context can influence reported responses [3]. In the present design, recording and reproduction are accompanied by a change of setting and session. Any observed outdoor–indoor difference must therefore be interpreted in relation to the complete experimental conditions.

Comparing the same participants across environments provides a direct within-listener comparison for each drone and manoeuvre. This comparison addresses whether their ratings change between the two test conditions; it does not isolate the causal contribution of the headphones, the recording system or the surrounding context. This distinction guides the paired analyses described in the methods and results chapters.

Ambisonics remains relevant to the broader literature on UAV auralization, as illustrated by the spherical-array and loudspeaker-reproduction study of Hui et al.[4]. However, it is not the reproduction method evaluated by the participants in the present indoor test. The preliminary Ambisonics processing is therefore documented separately from the main binaural experimental chain.

2.9 Recent Developments in Psychoacoustic Prediction

Beyond identifying correlations between individual metrics and subjective response, recent research has increasingly investigated predictive models of UAV-noise annoyance.

Such models attempt to estimate subjective response from combinations of acoustic, psychoacoustic, technical, operational, and contextual variables. This approach is relevant for the future development of perception-informed UAV design and noise-management strategies.

König et al. analysed the relationship between annoyance and a range of psychoacoustic, technical, and operational parameters. Their findings showed that loudness was among the strongest predictors in the considered datasets, while sharpness, tonality, roughness, and fluctuation strength also contributed depending on the experimental configuration and metric implementation [9].

Vaiuso et al. investigated a different modelling approach based on artificial neural networks. Their work combined acoustic measurements, psychoacoustic descriptors, flight information, and drone characteristics with the objective of predicting Psychoacoustic Annoyance [10].

More recently, Lotinga et al. developed prediction models for short-term annoyance responses to UAS sounds using data obtained in immersive acoustic scenes. Their methodology combined non-parametric machine-learning methods for metric selection with regression-based modelling while accounting for the clustered structure of repeated subjective observations [16].

Their results further indicate that annoyance prediction can involve not only sound intensity and sound-quality characteristics but also ambient acoustic context and the number of UAS events.

These developments reinforce the idea that UAV-noise annoyance is a multidimensional phenomenon and that predictive models may benefit from combining several objective descriptors rather than relying on a single sound-level metric.

2.9.1 Association, Explanatory Modelling and Prediction

The modelling literature motivates a distinction between three research questions. A correlation describes association between two variables. A multiple regression estimates an outcome from several predictors considered together. Predictive validation asks whether that model remains useful for observations not used to fit it. These questions should remain separate when interpreting the analyses of the present dataset.

Combining descriptors may be useful when they capture complementary features. For instance, a proposed loudness–roughness model would ask whether

temporal-modulation information adds to a measure of perceived magnitude. Such a combination must be fitted and evaluated explicitly; it cannot be claimed on the basis of two separate correlations. Similarly, a model using a participant's roughness rating is different from one using roughness calculated from an acoustic recording.

The regression analyses reported in this thesis are exploratory. Their fit to the available conditions does not establish predictive performance for new aircraft, manoeuvres or listeners. The number of distinct acoustic conditions, dependence among predictors and repeated ratings from the same participants constrain the interpretation. These considerations motivate a clear separation between objective-only models, models using subjective ratings and analyses of paired outdoor–indoor responses.

2.10 Research Gap and Study Positioning

The literature reviewed in the previous sections demonstrates that substantial progress has been made in the acoustic and psychoacoustic characterization of UAV noise. Previous studies have identified the importance of tonal components, high-frequency content, modulation, source directivity, operating condition, and acoustic context in determining the perceptual response to drone sound.

Subjective experiments have also demonstrated that conventional acoustic metrics alone may not fully explain perceived annoyance and that psychoacoustic descriptors such as loudness, sharpness, tonality, roughness, and fluctuation strength can provide additional information.

Nevertheless, several methodological limitations and open questions remain.

First, a substantial portion of the available acoustic literature is based on stationary or controlled operating conditions. Laboratory measurements provide high repeatability and detailed acoustic information but cannot always reproduce the aerodynamic and operational complexity of real flight. Field measurements improve realism but introduce environmental variability and are more difficult to control [1, 7].

Second, dynamic manoeuvres remain less extensively characterized than stationary hover conditions. The acoustic signature of a drone changes with rotor operating state, orientation, trajectory, and source–receiver geometry. These effects are particularly relevant during take-off, landing, and horizontal flight [1, 8].

Third, previous psychoacoustic studies have employed a wide variety of reproduction techniques. While some experiments have used realistic three-dimensional reproduction systems, other studies have relied on mono or simplified playback conditions. The systematic review by Schäffer et al. identified the spatial simplification of acoustic stimuli as a potential limitation to ecological validity [1].

At the same time, the study by Hui et al. demonstrates that spherical microphone-array recordings and third-order three-dimensional reproduction can be successfully integrated into a UAV psychoacoustic experiment [4].

More recent work has also begun to incorporate immersive acoustic scenes into predictive models of UAV annoyance [16]. Consequently, immersive reproduction should not be considered completely unexplored. Rather, an open methodological question concerns how realistic spatial recording and reproduction can be combined systematically with dynamic field manoeuvres, objective acoustic measurements, and subjective psychoacoustic evaluation.

The present work addresses these questions by connecting field measurements, objective acoustic characterization, outdoor subjective assessment and binaural indoor listening. Four UAV platforms and three manoeuvres—take-off, landing and horizontal flight—define the main experimental conditions.

The analysis follows four complementary directions. First, the recorded manoeuvres are characterized using acoustic and psychoacoustic descriptors, with the available independent flight repetitions used to describe between-event variability. Repeated flights and repeated listener responses represent different sources of variation and are kept separate. The available number of recorded events is reported for each condition; it is not assumed that all conditions contain three usable repetitions.

Second, outdoor and indoor responses are described separately, preserving the distinction between numerical annoyance, verbal annoyance categories and the five subjective sound-quality attributes. Distributional summaries provide the basis for comparing drones and manoeuvres without reducing each condition to a mean alone.

Third, the outdoor–indoor comparison uses matched responses from participants who completed both tests. Pairing is defined by participant, drone and manoeuvre. Condition-specific two-sided Wilcoxon signed-rank tests complement the descriptive comparisons; the mixed-effects analyses address broader patterns across the repeated-observation design. Their hypotheses and scope are described in the methods chapter rather than treated as interchangeable analyses.

Fourth, objective–subjective associations are examined using clearly identified acoustic events and corresponding subjective summaries. Objective descriptors for the selected stimuli were calculated on the segments used for indoor listening. Analyses involving other recorded repetitions must retain their event identity and must not imply that listeners independently evaluated every archived recording. The exploratory regressions are distinguished from validated predictive models, as discussed in Section 2.9.1.

The indoor experiment used separately acquired binaural recordings and headphone playback. Although spherical-array recordings were converted to a third-order

Ambisonics representation during preliminary processing, wind contamination prevented their use in the final listening test. That branch documents an acquisition and processing limitation; it is not an experimental comparison between Ambisonics and binaural reproduction.

The contribution of the study is therefore a coordinated assessment of drone and manoeuvre differences, the consistency of subjective responses across outdoor exposure and indoor binaural playback, and the relationships between recorded sound characteristics and listener ratings. Recording quality, limited repetition counts, sample size and differences between the two test settings remain part of the interpretation. Neither non-significant differences nor strong correlations alone establish perceptual equivalence between field exposure and reproduction.

Capitolo 3

Materials and Methods

3.1 Study Design

The study combined outdoor acoustic measurements and subjective evaluations with a subsequent binaural headphone experiment. Four UAVs and three manoeuvres defined twelve conditions. Outdoor testing took place on 1 July 2026 at Tetti Neirotti, near Turin. Indoor sessions took place on 3, 4, 8 and 10 September 2026.

The outdoor sample comprised fourteen participants, the indoor sample thirteen, and the matched sample eleven participants who completed both experiments. Separate outdoor and indoor descriptions used the available responses in each environment. Within-participant comparisons used only complete outdoor–indoor pairs for the same participant, UAV and manoeuvre.

Acoustic repetitions were distinct physical flight events. They were not additional subjective trials: participants evaluated one event per condition in each environment. Different channels and multiple exports of a recording were not counted as independent repetitions.

Tabella 3.1: Experimental design and available subjective observations.

Dataset	Participants	Evaluations or pairs
Outdoor	14	156 evaluations
Indoor	13	156 evaluations
Matched outdoor–indoor	11	125 pairs
Indoor-only	2	24 evaluations included in indoor total

3.2 Experimental Site and Flight Protocol

The outdoor site was an open flight field with a natural grass surface. Occasional road traffic and aircraft overflights contributed to the environmental background. A nearby hangar provided a sheltered preparation area. Background recordings were acquired separately from UAV operations.

The ZYLIA ZM-1 microphone was positioned at 1.60 m above ground. The NTi Audio XL2 was positioned adjacent to it. The HEAD acoustics operator wore the binaural headset next to the researcher, approximately 1 m from the ZYLIA and sound-level-meter setup. This separation documents proximity, rather than identical microphone positions. The 1.60 m height refers to ZYLIA and is not assigned to the binaural headset.

The field and conventional measurement setup are documented in Figures 3.1 and 3.2.

3.2.1 UAVs and Manoeuvres

The dataset identifies the four platforms as Agras 25, Matrice 350, Matrice 4E and Mavic Mini 5. These names are retained consistently across the recordings, questionnaires and analyses.

The vertical ascent extended from 10 to 30 m above ground, and descent from 30 to 10 m. Horizontal cruise followed a straight trajectory at 30 m altitude, extending from 10 m before to 10 m beyond the vertical axis of the measurement position. Programmed speeds were 1 m/s vertically and 2 m/s horizontally. These were programmed values rather than independently verified telemetry.

The labels take-off and landing identify the analysed ascent and descent phases. They do not imply that the retained acoustic segments necessarily included ground departure or touchdown.

Tabella 3.2: Programmed flight protocol. Altitudes refer to height above ground.

Condition	Altitude	Speed	Trajectory
Take-off	10–30 m	1 m/s	Vertical ascent
Landing	30–10 m	1 m/s	Vertical descent
Horizontal cruise	30 m	2 m/s	−10 to +10 m

Figure 3.3 summarises the programmed trajectories.



Figura 3.1: Overview of the outdoor flight field.



Figura 3.2: ZYLIA ZM-1 array and adjacent NTi Audio XL2 measurement setup. The documented ZYLIA height was 1.60 m. This photograph does not document the position of the wearable binaural headset.

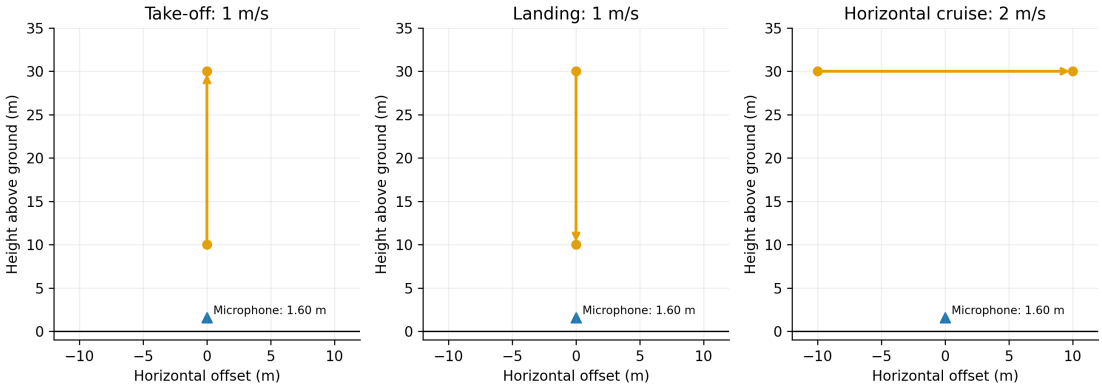


Figura 3.3: Programmed ascent, descent and horizontal-cruise geometry. Speeds are programmed values, not telemetry estimates. The microphone marker represents the ZYLIA height of 1.60 m; it does not establish the binaural headset position.

3.3 Acquisition and Reproduction Equipment

Three acquisition streams served different purposes. The NTi Audio XL2 provided conventional acoustic measurements; the ZYLIA ZM-1 provided multichannel recordings for preliminary spatial processing; and the HEAD acoustics system provided the binaural recordings used for psychoacoustic analysis and indoor playback.

3.3.1 Conventional and Multichannel Acquisition

The retained configuration reports a microphone sensitivity of 43.9 mV/Pa, a 20–120 dB input range and broadband logging at 0.1 s intervals. Broadband levels and one-third-octave spectra were retained with their measurement identifiers and integration intervals.

The ZYLIA recordings contained nineteen channels sampled at 48 kHz and 24 bit. The array orientation remained fixed, with its reference marker directed towards the documentation camera. Conversion to sixteen-channel third-order Ambisonics was completed, but wind contamination prevented use of this material in the final listening experiment.

3.3.2 Binaural Acquisition

Binaural acquisition was performed by the HEAD acoustics operator using a BHS II headset connected to a SQuadriga III recorder. BHS II is a wearable binaural headset, not a dummy-head microphone. The analysed HDF recordings contain separate left and right pressure channels at 48 kHz.

The BHS II data sheet specifies a recording frequency range of 20 Hz–20 kHz, a typical equivalent noise level of 27 dB(A), and a maximum sound pressure level of 130 dB SPL. The SQuadriga III data sheet specifies 24-bit analogue-to-digital conversion and support for sampling rates up to 96 kHz. These are equipment specifications; the sampling frequency of the analysed recordings was 48 kHz [17, 18].

3.3.3 Indoor Playback

The indoor stimuli were reproduced from HDF files using ArtemiS Companion in Automatic Playback Mode. The Sennheiser HD 560S headphones were connected to a HEAD acoustics BPU binaural playback unit. The HD 560S are open-back circumaural dynamic headphones.

HEAD acoustics describes the BPU as a USB device for binaural headphone reproduction, with equalisation filters for the headphones supplied in its bundles

and playback-software support [19]. In the configuration used for this study, the playback chain was set up to reproduce the recorded field levels. Volume and equalisation settings were fixed and could not be changed by the researcher or participants during the tests. No individual level adjustment or per-stimulus normalisation was introduced. This describes the study configuration, rather than a general claim that every BPU configuration prevents software adjustment.

Most indoor tests took place in the Audio Space Lab or offices within the DENERG, DIMEAS and DIATI departments at Politecnico di Torino. The indoor condition therefore did not denote a single acoustically identical room. A participant-by-room allocation and room-noise measurements were not available for the present analysis, so room effects were not estimated separately.

Tabella 3.3: Acquisition and reproduction chains and their roles in this study. Manufacturer capabilities are distinguished from the documented experimental use.

Equipment	Role	Documented configuration or specification
ZYLIA ZM-1	Preliminary spatial recording	Nineteen channels; 1.60 m above ground; not the source of final headphone stimuli
NTi Audio XL2	Conventional field descriptors	Adjacent to ZYLIA; event-level acoustic quantities linked through measurement IDs
HEAD BHS II	Binaural acquisition	Wearable two-channel headset; manufacturer recording range 20 Hz–20 kHz
SQuadriga III	Binaural recording frontend	Analysed HDF sampling rate 48 kHz; manufacturer's ADC specification 24 bit
HEAD BPU	Headphone playback	Fixed study playback settings; no participant volume adjustment
Sennheiser HD 560S	Indoor listening	Open-back circumaural dynamic headphones, connected to the BPU

3.4 Participants and Questionnaire

The fourteen outdoor participants comprised twelve men and two women, aged 22–56 years, with a mean age of 33.36 years and a standard deviation of 11.78 years. Seven reported professional acoustics experience, three reported an interest without professional expertise, and four reported no expertise. None reported hearing difficulties. Hearing status was self-reported rather than established through audiometry.

Consent to participation and research use of the data was obtained through a separate Google Form.

Eleven outdoor participants returned indoors, where two additional participants joined the experiment. The indoor sessions included five listeners on 3 September, two on 4 September, one on 8 September and five on 10 September. The two participants missing from the registration form are included in the 10 September total. Demographic information was not recollected for returning listeners. Equivalent preliminary demographic and hearing information was not available for the two indoor-only listeners.

3.4.1 Response Scales

The questionnaire collected numerical annoyance on a 0–10 scale, a separate verbal annoyance judgement, and five subjective attributes: perceived loudness, fluctuation, roughness, sharpness and tonality. The attribute scales ranged from -3 to $+3$. These signed codes represent response categories and do not imply negative physical acoustic quantities. The questionnaire's intensity item is referred to as subjective loudness in the analyses.

Verbal annoyance was retained as an ordered categorical response. It was not substituted for numerical annoyance or converted into a continuous physical descriptor.

The outdoor questionnaire is reproduced in Appendix A. Verbal categories were Per niente, Poco, Moderatamente, Molto and Estremamente, displayed in increasing order in the analyses. The numerical-annoyance endpoints were Per niente (0) and Estremamente (10). The five attribute endpoints ranged from Per niente to Estremamente, referring respectively to forte, fluttuante, ruvido, acuto and fischiante. These original Italian terms are retained in the appendix.

The indoor dataset contains the same seven response variables and numerical coding. The available documentation does not establish word-for-word equivalence of all indoor and outdoor instructions; the questionnaire reproduced in the appendix is the outdoor version.

3.5 Outdoor and Indoor Test Procedures

Outdoors, participants listened to and observed the first evaluated event for each condition and completed the questionnaire afterwards. Further recordings, where available, did not generate additional subjective responses. The UAV order was Mavic Mini 5, Matrice 4E, Matrice 350 and Agras 25. For each UAV, the order was take-off, landing and horizontal cruise. The main sequence was not randomised.

Indoors, participants first completed three familiarisation trials using recordings outside the twelve-stimulus experimental set. These practice stimuli were presented in random order and their ratings were excluded from analysis. The twelve main stimuli then followed the same fixed UAV and manoeuvre sequence as outdoors.

Each main stimulus was presented once without replay. There were no scheduled rest breaks within the main sequence, although questionnaire completion separated successive presentations. The two indoor-only listeners contributed to indoor descriptions but not to paired comparisons.

3.6 Data Structure and Measurement Correspondence

Participant codes, UAV names and manoeuvre labels were harmonised in the research workbook. Pairing used participant, UAV and manoeuvre, with indoor and outdoor scores placed in separate columns. Missing responses were not imputed.

The researcher confirmed that the outdoor-rated event is the first available HDF recording within each UAV–manoeuvre condition: the unsuffixed file is selected when present, otherwise the lowest available numerical suffix is selected (1, then 2, then 3). This identification rule was checked against the recording sequence, rather than inferred from file modification dates. The selected files and their use are listed in Table 3.4. Position-matched files labelled `posfonometro` form a separate quality-control series and are not substituted for the rated-event recordings.

The researcher additionally specified chronological pairing within each UAV–manoeuvre series: ascending HDF numbering corresponds to ascending XL2 acquisition order. For Matrice 350, Matrice 4E and Mavic Mini 5, XL2 entries labelled *Fabrizio near the microphone* belong to the separate `posfonometro` HDF series, rather than the ordinary rated-event series. For Agras, the operator remained near the microphone throughout, so this label does not separate a control series. The first valid ordinary event is therefore paired with the first available ordinary HDF; discarded XL2 acquisitions are excluded.

The resulting selected-event XL2 identifiers are 1/5/8 for Mavic Mini 5, 11/16/19 for Matrice 4E, 24/28/33 for Matrice 350 and 40/41/42 for Agras, in take-off/landing/cruise order. This map is based on the researcher's identification rule, not an independent timestamp verification. The remaining non-consecutive links were explicitly identified: the second Mavic Mini 5 take-off corresponds to XL2 recording 3, and the second Matrice 4E horizontal flight to XL2 recording 21. All 28 available HDF events therefore have an assigned XL2 counterpart; the complete mapping is provided in Appendix B. Same-event pairing also does not imply identical microphone position or segment duration.

Objective psychoacoustic metrics for the selected indoor stimuli were calculated on the same segments presented indoors. All available flight repetitions are retained for acoustic variability and the objective–objective correlations. Only the selected rated-event recording from each condition enters objective–subjective correlations;

Tabella 3.4: Evaluated HDF events and corresponding XL2 report identifiers. Duration is the edited HDF interval shown in the repetition export; XL2 integration intervals can differ. TO: take-off; LA: landing; HC: horizontal cruise.

UAV	Phase	HDF basename	XL2	HDF (s)
Agras 25	TO	`MatriceAgras_decollo`	40	34.22
Agras 25	LA	`MatriceAgras_atterraggio`	41	23.24
Agras 25	HC	`MatriceAgras_voloorizzontale`	42	38.24
Matrice 350	TO	`Matrice350_decollo`	24	17.89
Matrice 350	LA	`Matrice350_atterraggio`	28	23.62
Matrice 350	HC	`Matrice350_voloorizzontale`	33	26.93
Matrice 4E	TO	`Matrice4E_decollo`	11	23.68
Matrice 4E	LA	`Matrice4E_atterraggio`	16	22.08
Matrice 4E	HC	`Matrice4E_voloorizzontale`	19	30.08
Mavic Mini 5	TO	`MavicMini5_decollo1`	1	17.80
Mavic Mini 5	LA	`MavicMini5_atterraggio`	5	20.27
Mavic Mini 5	HC	`MavicMini5_voloorizzontale`	8	32.63

no subjective rating is assigned to an additional unrated repetition. Processing and statistical procedures are specified in Chapter 4.

3.7 Analysis Datasets and Units of Observation

The study required different datasets for different questions. Keeping these units explicit avoids counting an acoustic event repeatedly simply because several listeners evaluated it.

Tabella 3.5: Data scope for the final analyses.

Analysis	Unit and size	Inclusion rule
Objective variability and objective correlations	28 physical events	All available independent HDF recordings and assigned XL2 reports
Subjective distributions	156 ratings per environment	All available participants in each environment
Paired Wilcoxon	7–11 paired participants per condition	Same participant and same UAV–manoeuvre, both environments
Mean-rating correlations and regressions	12 conditions per environment	First evaluated recording; mean of available ratings for that condition
Individual-response associations	156 ratings per environment	Participant-centred supplementary description, preserving participant and stimulus identifiers

For outdoor condition means, the number of respondents varies because some responses are missing. Indoor means use thirteen responses per condition. Each condition receives one row and equal weight in the mean-based correlations and regressions. The matched-sample sensitivity analyses use the same complete respondent pairs in each condition; they do not replace the full-sample descriptions. No missing acoustic repetition or subjective response was generated artificially.

Capitolo 4

Signal Processing and Analysis Methods

4.1 Processing Workflow and Traceability

The analysis retained separate identifiers for the selected headphone stimuli, independent repeated flights, position-matched background-control recordings and XL2 measurement reports. A shared UAV–manoeuvre label establishes condition correspondence but does not establish an identical recording interval, microphone position or physical event.

The binaural HDF files formed the main processing branch. They supplied both the selected indoor stimuli and the signals analysed in ArtemiS SUITE 18.1. HEAD acoustics had edited the files to include the beginning and end of each relevant manoeuvre. Within a condition, differences in duration therefore describe distinct complete events rather than arbitrary additional silence. No retrospective shortening was imposed solely to equalise duration.

4.1.1 Preliminary Ambisonics Processing

The ZYLIA nineteen-channel recordings were converted to sixteen-channel third-order Ambisonics using ZYLIA Ambisonics Converter. Wind contamination reduced their suitability for perceptual testing, and the converted files were not used for the final listening experiment. The binaural stimuli were separately acquired and were not decoded from these Ambisonics recordings.

The documented conversion workflow is retained in Appendix B.3.

4.2 Conventional Acoustic and Psychoacoustic Descriptors

XL2 descriptors included equivalent levels, maximum levels, sound exposure and statistical levels, with the exact percentile labels retained from each export. The final cross-domain analysis used L_{Aeq}, L_{AFmax} and sound exposure level (SEL). Other exported percentile levels were retained in the source archive but were not added as interchangeable loudness predictors.

Sound exposure was interpreted together with event duration:

$$L_{\mathrm{AE}} = L_{\mathrm{Aeq},T} + 10\log_{10}(T/T_0), \qquad T_0 = 1\ \mathrm{s}. \tag{4.1}$$

Full-event exposure and equivalent level answer different questions. A longer event can have greater exposure without having a greater equivalent level.

4.2.1 ArtemiS Analysis and Channel Treatment

The selected analyses comprised Hearing-Model loudness, roughness, fluctuation strength and tonality, together with Aures sharpness. ISO 532-1 loudness was retained as an additional descriptor. The Hearing-Model analyses used the ECMA-418-2 implementations selected in the saved project.

Tabella 4.1: Treatment of the objective psychoacoustic outputs.

Descriptor	Exported channels	Condition-level treatment
Loudness, Hearing Model	Combined binaural	Retained directly
Roughness, Hearing Model	Combined binaural	Retained directly
Fluctuation, Hearing Model	Combined binaural	Retained directly
ISO loudness	Left and Right	Arithmetic channel mean
Aures sharpness	Left and Right	Arithmetic channel mean
Tonality, Hearing Model	Left and Right	Arithmetic channel mean

An arithmetic left–right mean is a derived descriptor, not a native binaural model output. Channels were kept separately in the repetition archive and were never counted as separate flights. Native units and the exact scalar labels were retained. The software's Default scalar was not automatically described as a temporal arithmetic mean.

The repetition exports contain ISO loudness N5 and N10. These percentile quantities retain their exact source labels; N5 and N10 describe exceedance fractions of 5% and 10%, respectively. They were not relabelled as N95.

Tabella 4.2: Objective metric dictionary for the final dataset. HM denotes Hearing Model. Scalar labels follow the exported Single Values Tables; "Default" is not redefined as an arithmetic time mean.

Descriptor	Implementation shown in project	Unit	Scalar and channels
N5, N10	ISO 532-1:2017 loudness	soneGF	5% and 10% exceedance values; mean of L/R outputs
HM loudness	ECMA-418-2 Hearing Model	soneHMS	Default overall loudness; combined binaural
Sharpness	Aures; DIN 45631/A1:2010 loudness input	acum	Default S; arithmetic mean of L/R outputs
Roughness	ECMA-418-2 Hearing Model	asper	Default R; combined binaural
Fluctuation strength	ECMA-418-2 Hearing Model	vacilHMS	Default F; combined binaural
Tonality	ECMA-418-2 Hearing Model	tuHMS	Default T; arithmetic mean of L/R outputs
L_{Aeq}	XL2 A-weighted equivalent level	dB	Entire XL2 measurement interval
L_{AFmax}	XL2 A-weighted, Fast maximum	dB	Maximum over XL2 interval
SEL (L_{AE})	XL2 A-weighted sound exposure	dB	Reference duration 1 s

4.3 Independent Repetitions and Descriptive Variability

The HDF archive contains 28 independent manoeuvres with unequal coverage. For each condition, event-level values were retained individually. Means and sample standard deviations were calculated only where at least two events were available. A single event does not provide a within-condition estimate of dispersion.

Tabella 4.3: Available independent HDF manoeuvres.

UAV	Landing	Take-off	Horizontal cruise
Agras 25	1	1	2
Matrice 350	3	3	2
Matrice 4E	3	3	2
Mavic Mini 5	3	2	3

Objective boxplots display the individual observations and the number of events. Where there are only one or two events, individual points provide the primary representation and any interpolated quartiles require caution. These analyses describe between-event variability; they do not establish measurement-system precision or full conformity with UNI ISO 5725.

4.4 Background-Noise Quality Control

The dedicated HDF background was inspected separately from the drone recordings. Candidate intervals beginning at 259 s were compared, and 259–274 s was retained as the longest stable candidate before the increase observed in longer intervals. Both the full recording and selected interval were retained for comparison.

Position-matched UAV recordings were compared with this background. The files labelled `posfonometro` supplied the corresponding recordings for Matrice 350, Matrice 4E and Mavic Mini 5; the supplied Agras files were already position-matched. These recordings were distinguished from headphone-playback files.

The saved spectral configuration used unweighted one-third-octave analysis, a Hanning window, 50% overlap and spectrum size 32768. Thirty-one bands from 20 Hz to 20 kHz were retained, separately for the left and right channels.

For each band, the measured total-to-background margin was

$$\Delta L_k = L_{\text{total},k} - L_{\text{background},k}. \tag{4.2}$$

Margins of at least 10 dB were retained without correction. Margins between 3 and 10 dB were treated using energetic subtraction:

$$L_{\text{source},k} = 10 \log_{10} \left(10^{L_{\text{total},k}/10} - 10^{L_{\text{background},k}/10} \right). \tag{4.3}$$

Bands below 3 dB were marked unreliable. Both channels had to satisfy the usability criterion before a combined result was retained.

These operational rules depend on a representative background and sufficiently uncorrelated additive contributions. The calculation corrects band-energy estimates; it does not denoise the waveform or generate new listening stimuli. Psychoacoustic metrics were not recalculated from the corrected bands.

4.5 Subjective Descriptions and Graphical Comparisons

Separate outdoor and indoor descriptions retained their respective samples. Paired plots used only complete responses from the same participants for the same condition. Numerical annoyance and the five attribute scores were represented by distributions, including medians, quartiles and individual observations. Where means were reported, the corresponding standard deviation was also reported.

Within each UAV, the three manoeuvres were placed next to one another. Paired outdoor–indoor boxes were adjacent within each manoeuvre, with consistent environment colours across outcomes. Plot limits were expanded to include all observations and whiskers. Numerical-annoyance plots retained the 0–10 response range and attribute plots the -3 to $+3$ range, with graphical margins where necessary.

Verbal annoyance was represented by category counts or percentages, retaining the original order of categories. Its separate-environment plots used all available responses, while paired comparisons used the matched sample. Denominators were stated explicitly.

The plots are reported in Chapter 5; numerical summaries are provided in Appendix C. Verbal categories and scale endpoints were checked against the outdoor questionnaire and the response-workbook headers.

4.6 Paired Wilcoxon Signed-Rank Tests

Two-sided Wilcoxon signed-rank tests compared indoor and outdoor numerical ratings separately for each UAV–manoeuvre condition. Six outcomes were analysed: numerical annoyance, loudness, fluctuation, roughness, sharpness and tonality. Twelve conditions therefore generated 72 tests.

For participant i in condition c, the paired difference was defined as

$$d_{ic} = y_{ic,\mathrm{Indoor}} - y_{ic,\mathrm{Outdoor}}. \tag{4.4}$$

Each participant contributed at most one difference to a given condition-specific test. The 125 pairs were not pooled as 125 independent participants. Agras landing,

take-off and cruise contained 7, 10 and 9 complete pairs, respectively; all other conditions contained 11.

The signed-rank test uses the signs and ranks of paired differences. Under a symmetric location-shift interpretation, it assesses whether these differences are centred on zero. It is not calculated by subtracting the two sample medians. Separate indoor and outdoor medians were nevertheless reported to describe each paired sample. Zero differences reduce the number of observations contributing non-zero ranks. [20].

The reported W values retain the Jamovi convention used in the exports, corresponding to the positive-rank sum for Indoor minus Outdoor. Original two-sided Jamovi p values were retained, with values below 0.05 shown in bold. Independently reconstructed asymptotic values with continuity correction matched the exported values at their displayed precision; they are not presented as exact small-sample probabilities.

4.7 Systematic Correlation Analysis

The final analysis followed the sequence subjective–subjective, objective–objective and objective–subjective. Spearman's rank correlation was the principal coefficient because the interest concerned monotonic ordering across conditions and the response variables were bounded ratings. Pearson coefficients were also calculated as supplementary linear summaries and are retained in the numerical archive. The coefficients are not interchangeable: a strong rank association does not imply that a straight line provides an equally strong fit.

For two variables with observations x_c and y_c, Spearman's coefficient is the Pearson correlation of their ranks, with average ranks for ties. Rows with missing values in either variable are excluded pairwise. The main mean-based matrices contain twelve complete condition pairs, while the full-event objective matrix contains 28 events. The matrices show both the coefficient and its two-sided nominal p value.

4.7.1 Subjective–Subjective Matrices

Outdoor and indoor matrices were calculated separately from condition means of numerical annoyance and the five subjective attributes. The verbal annoyance item remained categorical and was described separately. The mean of each attribute was calculated across the available participants in that condition, before correlating the twelve means. A supplementary matched-sample calculation used only the complete outdoor–indoor respondent pairs to examine sensitivity to sample composition.

4.7.2 Objective–Objective Matrices

All 28 independent recorded manoeuvres entered the primary objective matrix. The seven psychoacoustic descriptors were N5, N10, Hearing-Model loudness, sharpness, roughness, fluctuation strength and tonality. A second matrix included the assigned XL2 equivalent level, maximum level and SEL. The HDF–XL2 mapping identifies the same physical events, but the instruments occupied different positions and their integration intervals could differ. Correlation between their descriptors is therefore an event-level comparison, not proof that the two measurement chains are interchangeable.

All available repetitions were retained without duplicating missing events. Consequently, conditions with three events have more weight than conditions with one event in the pooled objective matrix. UAV and manoeuvre labels were retained in scatterplots to make the physical grouping visible. A twelve-selected-event calculation was retained as a supplementary sensitivity analysis.

4.7.3 Objective–Subjective Matrices

Only the first evaluated HDF event within each condition supplied the psychoacoustic predictors. The extra acoustic repetitions were not assigned copies of subjective means. Each environment contributed twelve rows pairing objective event descriptors with the corresponding mean ratings. A further matrix related the three XL2 field descriptors to the same twelve response means.

The outdoor response describes the live event, including its context; the indoor response describes playback of its selected binaural recording. An objective descriptor associated with that event does not capture all visual, contextual or individual factors influencing either response.

4.7.4 Individual Responses: Supplementary Analysis

The individual-response analysis retained all 156 ratings in each environment. To remove differences in respondents' overall use of the scale, each variable was centred within participant, then a common linear correlation was calculated from the centred pairs:

$$r_{\text{within}} = \operatorname{cor}\left(x_{ic} - \overline{x}_i,\ y_{ic} - \overline{y}_i\right). \tag{4.5}$$

This is a participant-centred linear coefficient, not a Spearman coefficient or a model predicting each participant's response. Objective values repeat across participants because the same stimulus was evaluated; those copies do not create new independent acoustic observations.

Exploratory uncertainty intervals were obtained from 2,000 crossed bootstrap resamples of participant and stimulus identifiers, with replacement in each dimension

and re-centring within each resample. Percentiles 2.5 and 97.5 of the resulting coefficients define the reported interval. No ordinary pooled-observation p value was assigned to these coefficients. Stimuli remain grouped within four UAVs, so even these intervals should not be interpreted as a complete adjustment for every level of experimental dependence. This analysis is supplementary to the simpler condition-mean matrices.

4.7.5 Interpretive Limits of Correlation Probabilities

The exported Spearman probabilities use SciPy's asymptotic approximation. The documentation cautions about its accuracy in small samples [21]. Here the nominal probabilities are supplied for transparent reporting, not treated as exact small-sample evidence. The twelve conditions share four UAV platforms and largely the same listeners; the 28 events are also grouped by condition. Ordinary correlation tests do not fully represent these dependencies.

No multiplicity correction was applied to the exploratory correlation matrices. The complete matrices, coefficient magnitudes and directions were inspected together, rather than retaining only cells below 0.05. A low p value does not indicate a larger physical effect, causal influence or independent contribution after accounting for another descriptor. The paired-sample and selected-event sensitivity results are retained in the numerical supplement.

4.8 Simple Regression Models

The correlation results motivated a set of transparent ordinary least-squares (OLS) models. They were fitted separately outdoors and indoors to the twelve condition means, with an intercept and unstandardised predictors. The basic model was

$$\overline{y}_c = a + bx_c + \varepsilon_c, \tag{4.6}$$

where $\overline{y}_c$ is the mean rating for condition c, x_c is its objective or subjective predictor, and a and b minimise the sum of squared residuals. For a single predictor,

$$b = \frac{\sum_c (x_c - \overline{x})(\overline{y}_c - \overline{y})}{\sum_c (x_c - \overline{x})^2}, \qquad a = \overline{y} - b\overline{x}. \tag{4.7}$$

The fitted slopes have the units of the rating scale per unit of predictor. They are not correlation coefficients and were not taken directly from a correlation matrix.

4.8.1 Model Families

The same predefined set of 17 candidate fits was evaluated within each environment, giving 34 fits in total:

1. Eight single-predictor models of numerical annoyance using HM loudness, N5, N10, SEL, roughness, fluctuation strength, sharpness or tonality.

2. Five models relating an objective descriptor to its corresponding subjective attribute: HM loudness, sharpness, roughness, fluctuation strength and tonality.

3. Two models of annoyance using subjective loudness or subjective roughness alone.

4. Two two-predictor models of annoyance: HM loudness plus objective roughness, and subjective loudness plus subjective roughness.

The two-predictor models ask whether roughness adds information beyond loudness. The subjective models describe relationships between ratings and are not substitutes for objective-only models when ratings are unavailable. The candidate set was motivated by the observed correlations and is therefore exploratory. All candidates are reported, including weak fits and coefficients with nominal $p \geq 0.05$.

4.8.2 Fit Measures and Diagnostics

The coefficient of determination was calculated as

$$R^2 = 1 - \frac{\sum_c (\overline{y}_c - \widehat{y}_c)^2}{\sum_c (\overline{y}_c - \overline{y})^2}. \tag{4.8}$$

Adjusted R^2 accounts for the number of fitted predictors k:

$$R^2_{\text{adj}} = 1 - (1 - R^2)\frac{n-1}{n-k-1}, \qquad n = 12. \tag{4.9}$$

Adding a predictor cannot reduce ordinary R^2, so its incremental usefulness was assessed using adjusted R^2 and the conditional coefficient of the added predictor. The two-sided coefficient tests and overall model F tests are ordinary, unadjusted OLS probabilities. For a one-predictor model with an intercept, the slope test and overall model test yield the same probability.

Residual normality, heteroscedasticity and influence were examined using Shapiro–Wilk, Breusch–Pagan, Cook's distance and studentised residual summaries. Diagnostic probabilities were not used as proof that assumptions held. With twelve observations, bounded responses and common participants, ordinary OLS intervals and probabilities remain provisional descriptions under the working model.

No stepwise selection, machine-learning algorithm, cross-validation or leave-one-UAV-out procedure was used in these final models. R^2 and the recorded fitting errors refer to the data used to estimate the equations. The models are simple

empirical summaries of the measured conditions; their performance on new drones, recordings or participants was not established. Predictions should remain within the observed predictor range and should not be interpreted as unconstrained physical laws for bounded rating scales.

4.9 Computational Traceability

Descriptive plots, paired tests and correlation exploration were developed from the Jamovi-ready research workbook and Jamovi exports. The final numerical matrices and simple regressions were calculated with Python using pandas and NumPy for data handling and least-squares calculations, with SciPy distributions and diagnostic tests. The numerical supplement retains the accepted model inputs, coefficient tables, fit summaries, diagnostic outputs and full correlation results, as well as the 28-event mapping and 72 paired-test results. The thesis figures and tables were generated from those outputs; rounding is applied only for display. Earlier exploratory model variants are not mixed with the final simple-regression results.

Capitolo 5

Results

5.1 Dataset Coverage and Traceability

The final dataset contains twelve evaluated UAV–manoeuvre conditions. Outdoor and indoor samples each contain 156 ratings, but their participant composition differs: fourteen people contributed outdoors and thirteen indoors. The paired analysis contains eleven people and 125 complete condition pairs. Agras landing, take-off and cruise have seven, ten and nine pairs, respectively; each of the other nine conditions has eleven. Missing responses were not replaced by condition means.

The objective archive contains 28 distinct flight events. There are no triplicate recordings for every condition: Agras landing and take-off each have one event, while several other conditions have two. Figure-based distributions and numerical dispersion estimates must be read with these counts in mind. Table 3.4 identifies the twelve evaluated HDF events; Appendix B lists all 28 HDF–XL2 associations, including the two explicitly resolved non-consecutive links.

5.2 Acoustic Characterisation

5.2.1 Conventional Field Descriptors

Table 5.1 reports equivalent level, Fast maximum level, SEL and XL2 integration duration for the twelve corresponding field events. Agras has the highest equivalent levels in this selected set, from 68.0 to 71.9 dB. The Matrice 350 values range from 60.2 to 65.7 dB; the Matrice 4E values from 50.3 to 60.0 dB; and the Mini 5 values from 53.1 to 55.0 dB. The relative positions of the two smaller platforms therefore depend on manoeuvre.

The equivalent level, maximum level and exposure level describe different aspects of an event. For example, Mini 5 take-off has an equivalent level of 54.3 dB and

a maximum of 70.6 dB, illustrating why a transient maximum should not be treated as the sustained level. Exposure also incorporates duration. HDF and XL2 durations are reported separately because their edited and instrument-integrated intervals are not necessarily identical.

Tabella 5.1: Conventional acoustic descriptors of the twelve corresponding field events. These are XL2 quantities, not measurements at the headphone listener's ears.

UAV	Phase	L_{Aeq} (dB)	L_{AFmax} (dB)	SEL (dB)	XL2 (s)
Agras 25	TO	69.7	76.3	85.2	35
Agras 25	LA	71.9	78.7	85.7	24
Agras 25	HC	68.0	76.7	83.9	39
Matrice 350	TO	60.2	69.8	73.0	18
Matrice 350	LA	65.7	72.6	79.8	25
Matrice 350	HC	61.9	73.1	77.9	40
Matrice 4E	TO	50.3	62.7	65.8	35
Matrice 4E	LA	60.0	69.4	73.3	21
Matrice 4E	HC	56.6	68.1	70.0	22
Mavic Mini 5	TO	54.3	70.6	66.9	18
Mavic Mini 5	LA	53.1	69.2	65.9	19
Mavic Mini 5	HC	55.0	69.6	68.5	22

5.2.2 Psychoacoustic Descriptors and Independent Events

Table 5.2 presents the descriptors of the twelve evaluated recordings. Figures 5.1 and 5.2 include every available independent event, grouping the three manoeuvres within each UAV. Individual points and event counts accompany the boxes. The detailed event means, sample standard deviations and ranges are given in Appendix B.2.

HM loudness is high for the selected Agras events and low for the Mini 5 events. Other descriptors do not follow a perfectly identical order. This is relevant when evaluating whether several metrics supply complementary information or mainly reflect the same contrast between aircraft.

The repetitions show that a condition label does not imply a fixed acoustic descriptor. For example, Matrice 4E take-off roughness spans approximately 0.359–0.623 asper across its three events. Mini 5 horizontal-flight HM loudness is approximately 2.96, 3.54 and 2.87 soneHMS. The two Agras horizontal events have HM loudness of 8.09 and 8.68 soneHMS, while their roughness values are approximately 0.640 and 0.514 asper. A higher loudness in one repetition is therefore not necessarily accompanied by a higher roughness.

Tabella 5.2: Psychoacoustic descriptors of the twelve evaluated recordings. Units and channel treatment are defined in Table 4.2.

UAV	Phase	N5	N10	HM	Sharp.	Rough.	Fluct.	Tonal.
Agras 25	TO	34.05	32.40	9.13	2.76	0.800	0.200	0.087
Agras 25	LA	34.90	33.90	9.93	2.69	0.376	0.202	0.103
Agras 25	HC	32.10	30.80	8.09	2.60	0.640	0.191	0.056
Matrice 350	TO	18.40	17.00	6.28	2.40	0.716	0.166	0.114
Matrice 350	LA	23.90	22.95	7.68	2.54	0.223	0.152	0.129
Matrice 350	HC	23.30	22.25	6.50	2.17	0.222	0.142	0.152
Matrice 4E	TO	11.00	10.25	4.60	2.15	0.359	0.154	0.153
Matrice 4E	LA	13.30	12.95	5.35	2.17	0.150	0.134	0.137
Matrice 4E	HC	12.80	12.20	4.47	1.95	0.612	0.109	0.188
Mavic Mini 5	TO	7.06	6.66	3.30	2.08	0.134	0.055	0.229
Mavic Mini 5	LA	8.79	8.55	3.95	2.28	0.038	0.093	0.092
Mavic Mini 5	HC	6.50	6.32	2.96	1.75	0.096	0.021	0.164

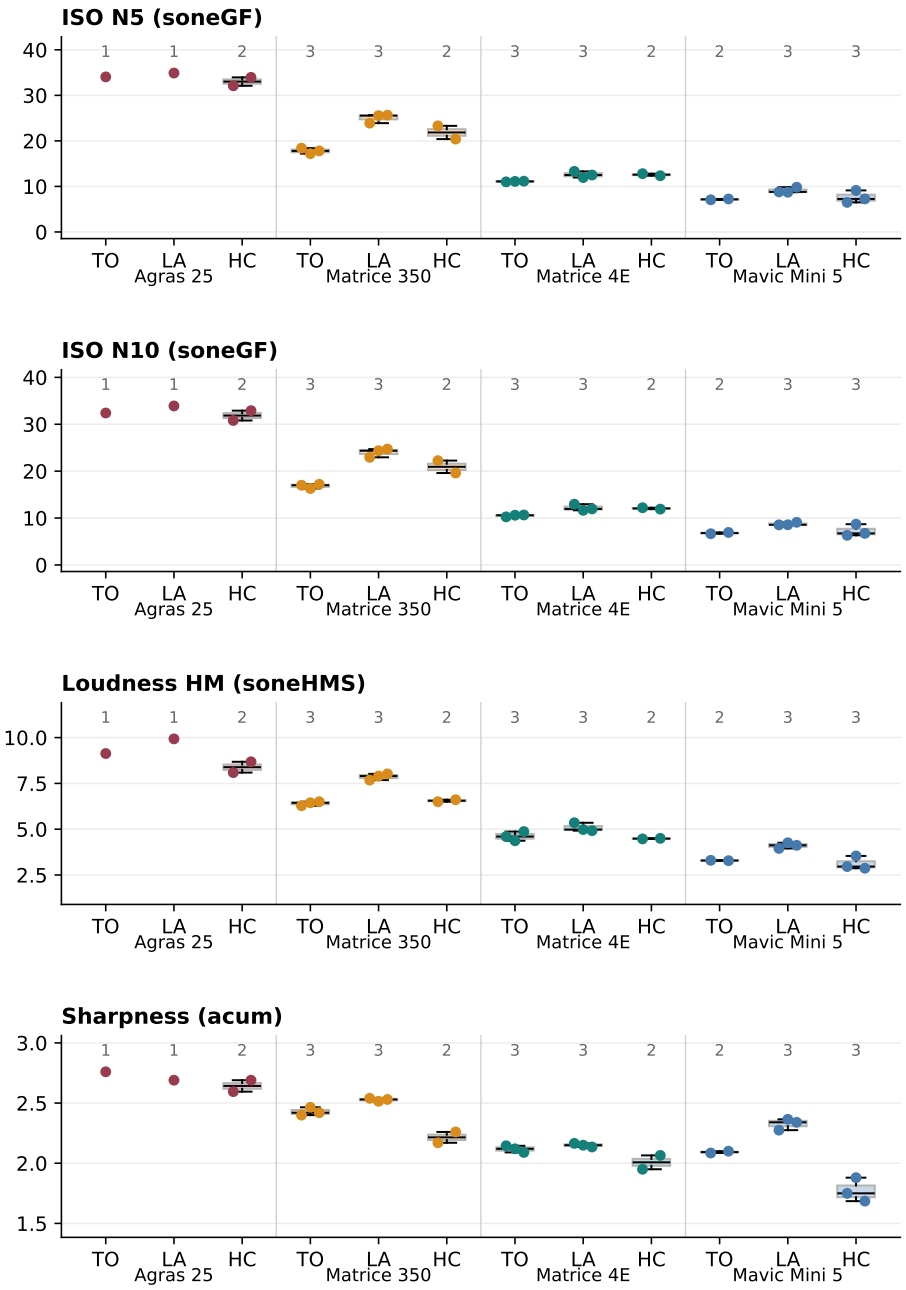


Figura 5.1: Objective event distributions: ISO N5, ISO N10, HM loudness and sharpness. Each point is a distinct flight event; the numbers above the groups are event counts. Boxes show interpolated quartiles and median, with whiskers to the most extreme points within 1.5 IQR. A singleton is shown only as a point. TO: take-off; LA: landing; HC: horizontal cruise.

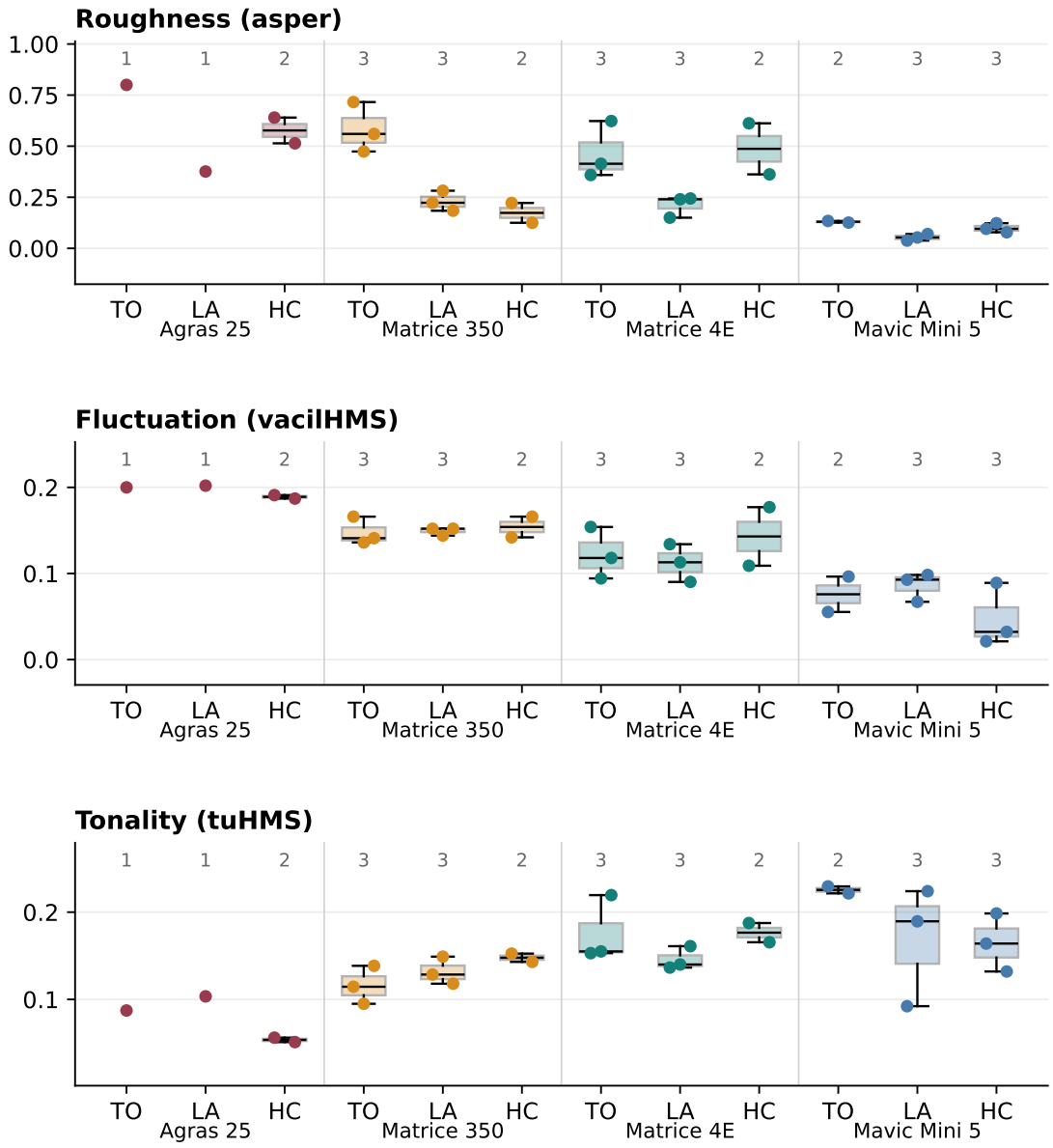


Figura 5.2: Objective event distributions: roughness, fluctuation strength and tonality. The same 28-event sample and grouping conventions as Figure 5.1 apply. With one to three events per condition, the points carry more information than the apparent smoothness or width of a box.

5.2.3 Background-Noise Quality Control

The position-matched recordings were compared with the selected 259–274 s background interval in 31 one-third-octave bands. Table 5.3 separates bands with at least 10 dB margin in both channels, bands requiring energetic correction, and bands below the 3 dB usability threshold in either channel. Figure 5.3 shows the smaller of the two channel margins.

Of the 372 condition–band combinations, 340 have a usable result under the two-channel criterion and 32 do not. All 31 bands are usable for each Agras and Matrice 350 condition. The Mini 5 conditions have 23, 26 and 19 usable bands for take-off, landing and cruise; the corresponding Matrice 4E counts are 31, 25 and 30. Residual background is therefore a more material issue for some quieter recordings than for the higher-level events.

These checks concern a dedicated position-matched series. They neither establish the same signal-to-background margin at every binaural recording position nor remove background from the evaluated waveform. The psychoacoustic descriptors and listening responses analysed below retain their original signal content.

Tabella 5.3: Bandwise background margin in the position-matched quality-control recordings. Classification uses the smaller of the two channel margins; each row contains 31 bands.

UAV	Manoeuvre	$\geq$ 10 dB	3–10 dB	Below 3 dB
Matrice 4E	Landing	21	4	6
Matrice 4E	Take-off	20	11	0
Matrice 4E	Horizontal cruise	21	9	1
Matrice 350	Landing	27	4	0
Matrice 350	Take-off	21	10	0
Matrice 350	Horizontal cruise	24	7	0
Agras 25	Landing	26	5	0
Agras 25	Take-off	26	5	0
Agras 25	Horizontal cruise	26	5	0
Mavic Mini 5	Landing	15	11	5
Mavic Mini 5	Take-off	13	10	8
Mavic Mini 5	Horizontal cruise	13	6	12

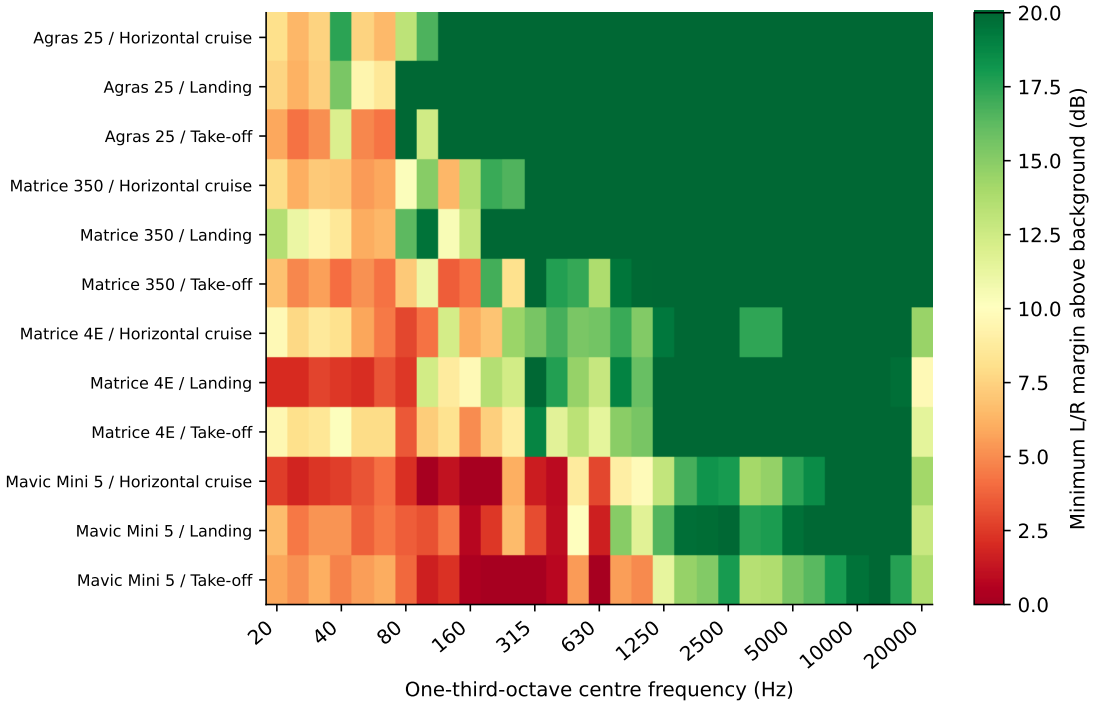


Figura 5.3: Minimum left/right total-to-background margin in each one-third-octave band of the position-matched quality-control series. A margin below 3 dB in either channel prevents a reliable combined corrected value under the adopted rule.

5.3 Subjective Response Distributions

5.3.1 Separate Outdoor and Indoor Samples

Figures 5.4–5.7 show numerical annoyance and the five subjective attributes using all available responses in each environment. Tables in Appendix C give the associated means with sample standard deviations, medians and quartiles. Outdoor and indoor boxes in this subsection do not have identical participant composition.

The annoyance distributions generally place Agras and Matrice 350 above the smaller platforms, while the amount of overlap and spread depends on the manoeuvre. Individual ratings remain dispersed within conditions, so the separation between condition means should not be mistaken for agreement among all listeners. Subjective loudness broadly follows the high-annoyance conditions. Fluctuation and roughness also vary across platforms, whereas sharpness and tonality do not consistently reproduce the annoyance ordering.

The signed attribute codes denote positions on the questionnaire's -3 to $+3$ response scale. A negative sharpness or tonality score is a low coded rating, not a negative physical descriptor. Means on this scale are summaries of category codes, whereas the boxplots retain the discrete observations.

5.3.2 Matched Outdoor–Indoor Distributions

Figures 5.8 and 5.9 restrict both environments to the same complete participant–condition pairs. The three manoeuvres remain adjacent within each UAV, and outdoor and indoor boxes are adjacent within each manoeuvre. This construction makes the environment comparison visible without introducing lines between unrelated UAV categories.

The boxes show marginal distributions of the paired samples; they do not themselves display every person's change. Two conditions can have equal outdoor and indoor medians even when several participants change their ratings, and conversely a change in a median does not identify its distribution of signed ranks. The paired tests below therefore use the original participant-level differences.

5.3.3 Verbal Annoyance

The verbal response is shown as ordered category percentages in Figures 5.10 and 5.11. This representation preserves the distinction between "not at all", "slightly", "moderately", "very" and "extremely" annoyed. It complements numerical annoyance by showing how responses occupy categories rather than replacing the 0–10 scale with a second arithmetic score.

The separate-sample panels use all 156 ratings per environment, with the percentage denominator defined within each condition. The matched panel uses

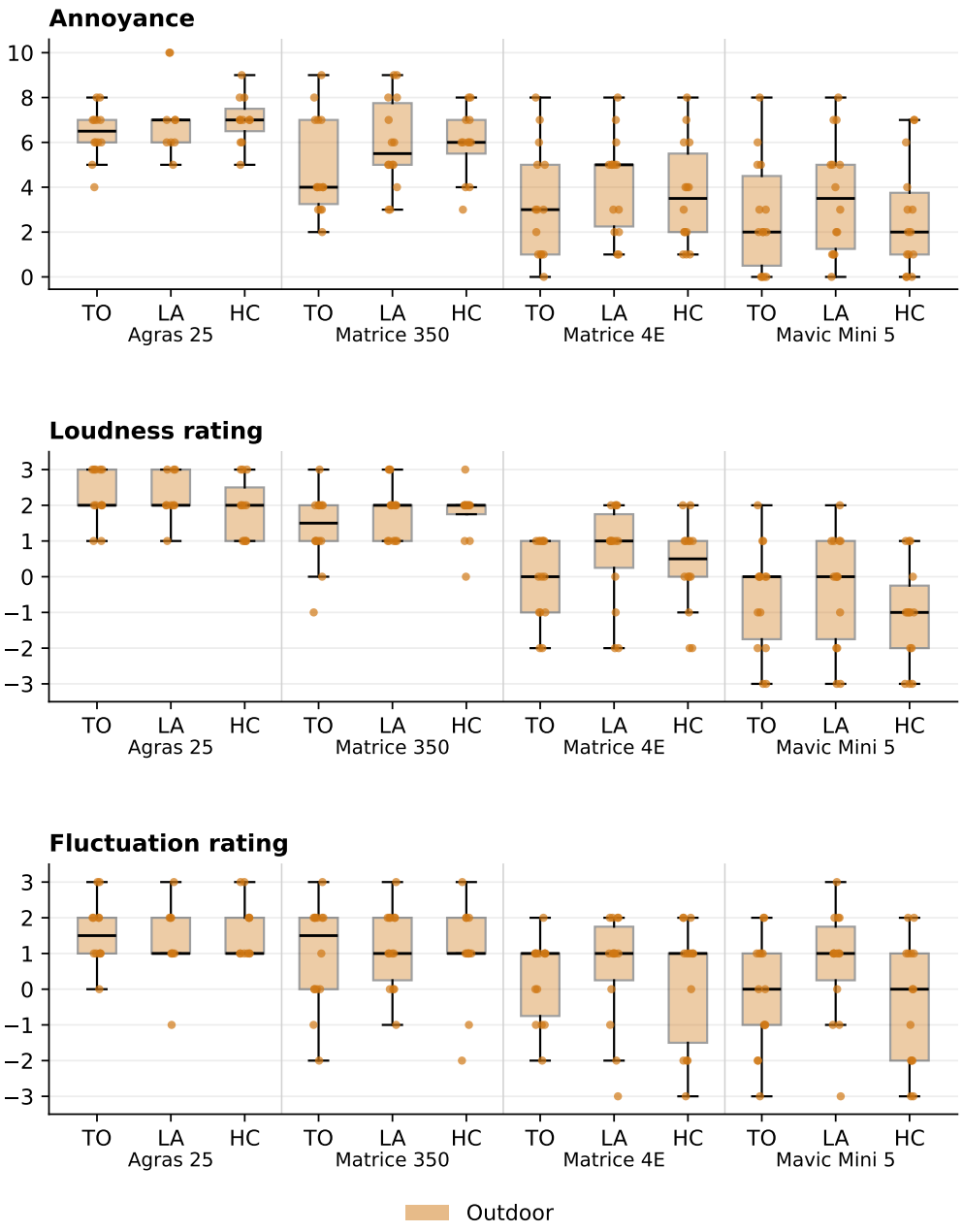


Figura 5.4: Outdoor subjective distributions, numerical annoyance, loudness and fluctuation. All available respondents are retained. Boxes show median and quartiles; whiskers extend to the most extreme values within 1.5 IQR, and all individual responses are plotted. Three manoeuvres are adjacent within each UAV. Axis limits include the full response scale and graphical margins.

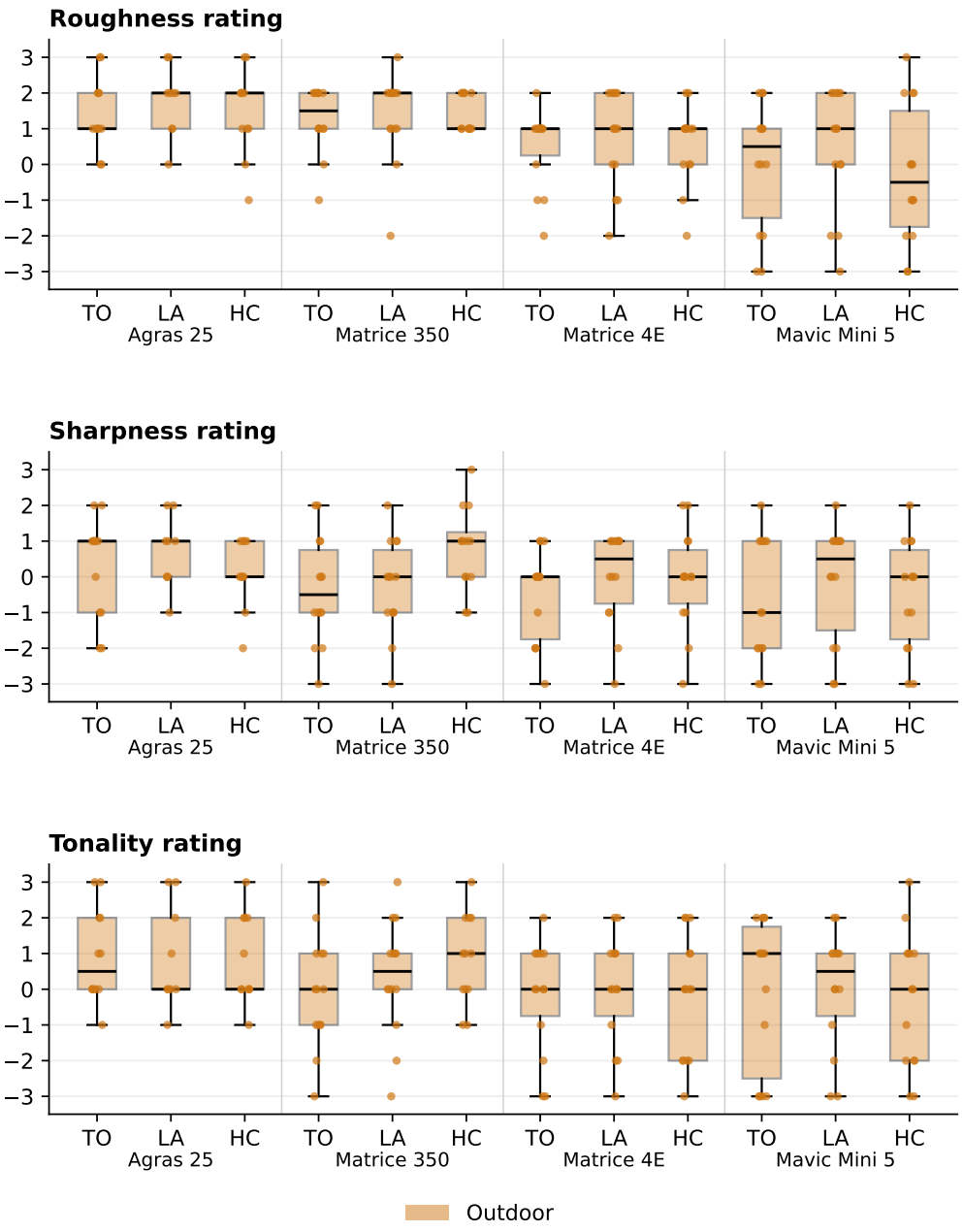


Figura 5.5: Outdoor subjective distributions, roughness, sharpness and tonality. All available respondents are retained. Boxes show median and quartiles; whiskers extend to the most extreme values within 1.5 IQR, and all individual responses are plotted. Three manoeuvres are adjacent within each UAV. Axis limits include the full response scale and graphical margins.

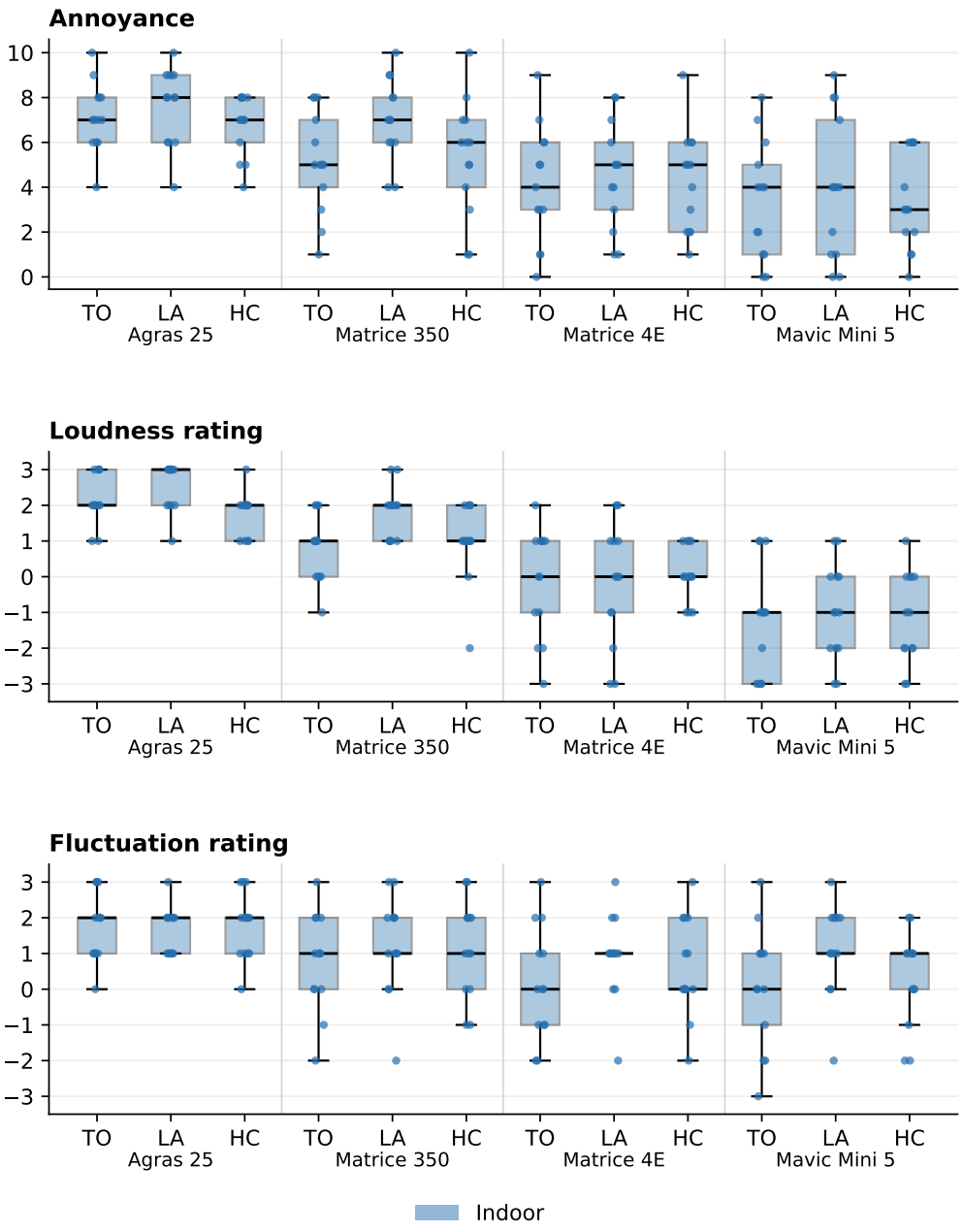


Figura 5.6: Indoor subjective distributions, numerical annoyance, loudness and fluctuation. All available respondents are retained. Boxes show median and quartiles; whiskers extend to the most extreme values within 1.5 IQR, and all individual responses are plotted. Three manoeuvres are adjacent within each UAV. Axis limits include the full response scale and graphical margins.

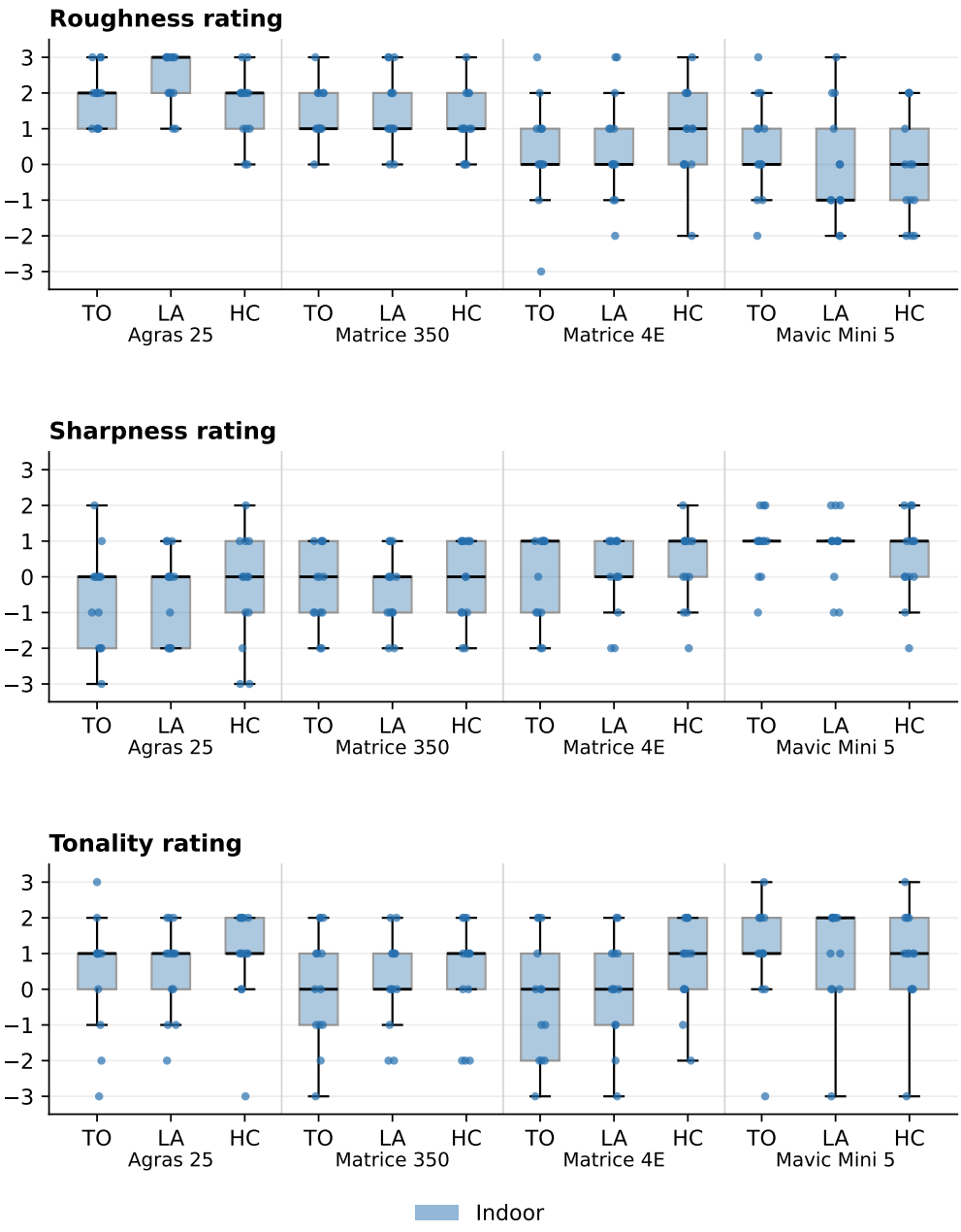


Figura 5.7: Indoor subjective distributions, roughness, sharpness and tonality. All available respondents are retained. Boxes show median and quartiles; whiskers extend to the most extreme values within 1.5 IQR, and all individual responses are plotted. Three manoeuvres are adjacent within each UAV. Axis limits include the full response scale and graphical margins.

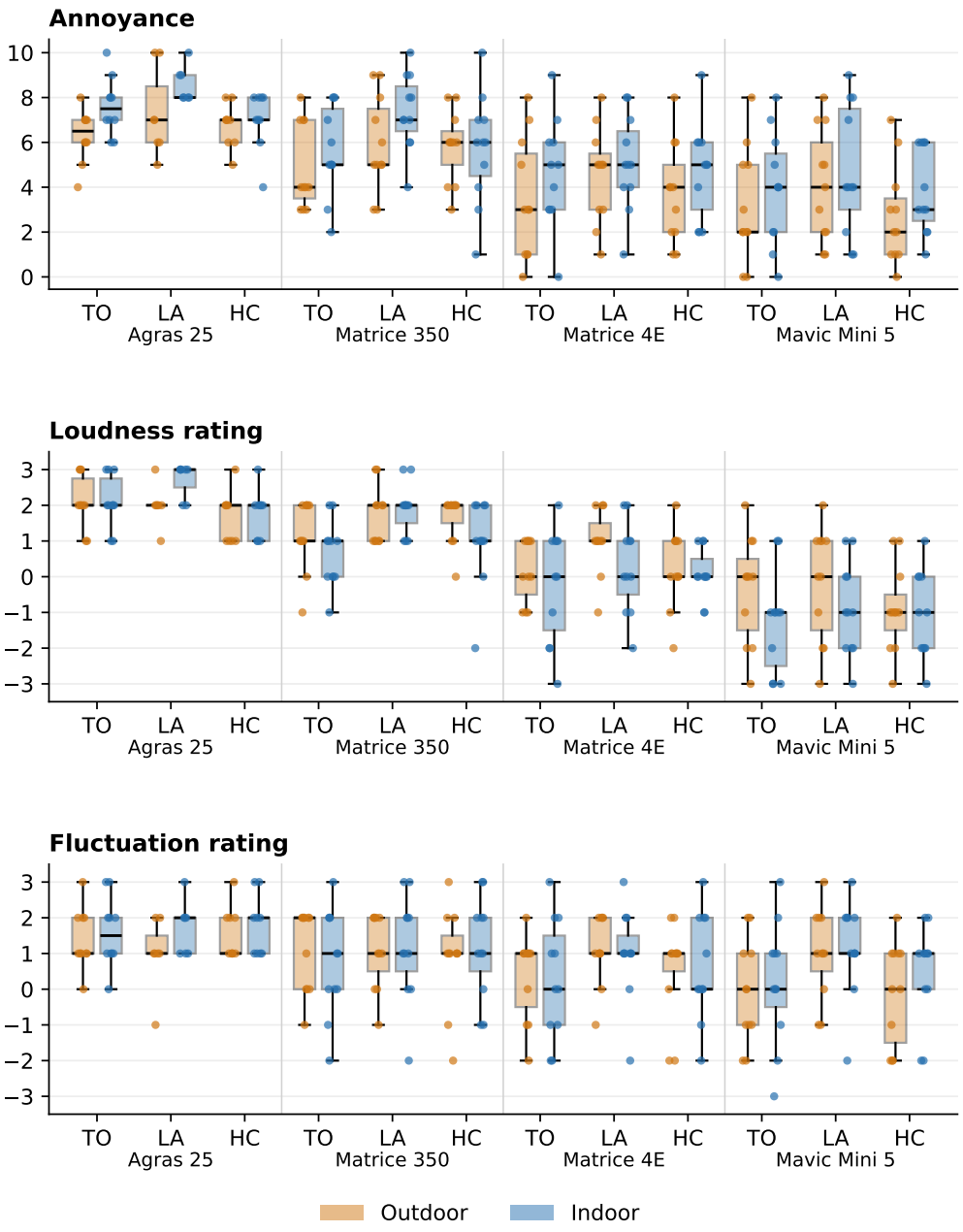


Figura 5.8: Outdoor–indoor comparison restricted to the same participants for each UAV–manoeuvre condition. Orange denotes outdoor and blue indoor. There are 125 pairs from eleven returning listeners; pair counts are 7, 10 and 9 for Agras landing, take-off and cruise, and 11 elsewhere. Boxplot and point conventions match the separate-sample figures.

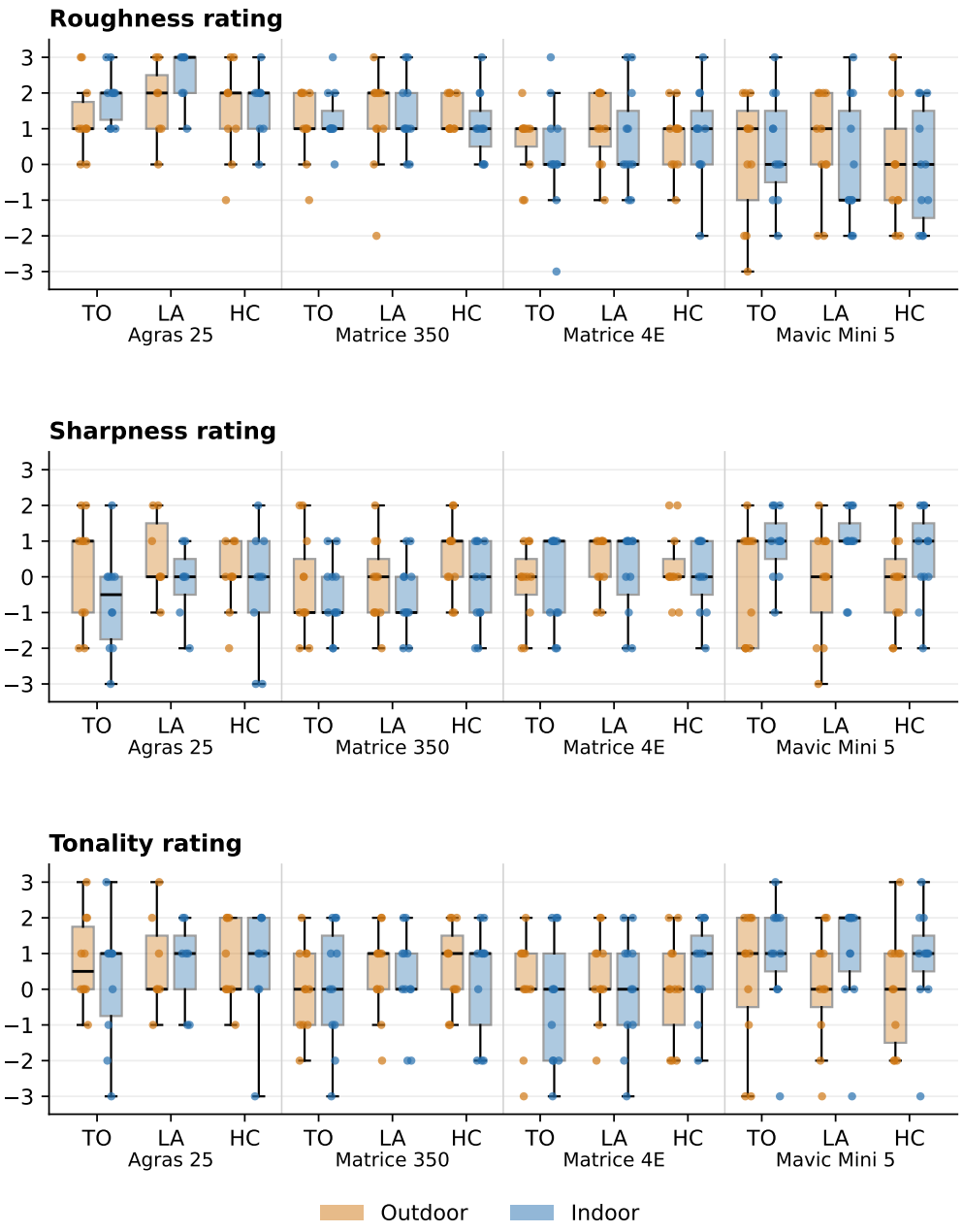


Figura 5.9: Outdoor–indoor comparison restricted to the same participants for each UAV–manoeuvre condition. Orange denotes outdoor and blue indoor. There are 125 pairs from eleven returning listeners; pair counts are 7, 10 and 9 for Agras landing, take-off and cruise, and 11 elsewhere. Boxplot and point conventions match the separate-sample figures.

only complete pairs and the same denominator for its adjacent outdoor and indoor bars. Differences between these two displays can therefore reflect the restriction to returning participants as well as changes in the category distributions.

5.4 Paired Wilcoxon Signed-Rank Tests

Table 5.4 contains all 72 two-sided tests, with the separate matched-sample medians, W and the original Jamovi probability. Three original probabilities are below 0.05:

- Agras landing, subjective loudness: outdoor median 2 and indoor median 3; $W = 15$, $p = 0.037$. There are seven complete pairs and five non-zero differences.
- Matrice 350 landing, numerical annoyance: outdoor median 5 and indoor median 7; $W = 39.5$, $p = 0.048$. There are eleven pairs and nine non-zero differences.
- Mini 5 horizontal cruise, subjective sharpness: outdoor median 0 and indoor median 1; $W = 21$, $p = 0.031$. There are eleven pairs and six non-zero differences.

Each of these comparisons points towards higher indoor ratings in its matched sample. The other 69 tests have original $p \geq 0.05$. This does not establish equality or perceptual equivalence between environments, particularly where there are few non-zero differences.

Tabella 5.4: All condition-specific paired Wilcoxon signed-rank tests. Medians refer to the matched sample; W is the positive-rank sum for Indoor minus Outdoor. Original two-sided Jamovi p values below 0.05 are bold. No multiplicity adjustment is applied in this table.

UAV	Phase	Outcome	Pairs	Med. Out	Med. In	W	p
Agras 25	TO	Annoyance	10	6.5	7.5	44.5	0.085
Agras 25	TO	Loudness	10	2	2	5	1.000
Agras 25	TO	Fluctuation	10	1	1.5	13.5	0.589
Agras 25	TO	Roughness	10	1	2	18.5	0.105
Agras 25	TO	Sharpness	10	1	-0.5	15	0.212
Agras 25	TO	Tonality	10	0.5	1	10.5	0.320
Agras 25	LA	Annoyance	7	7	8	19.5	0.071
Agras 25	LA	Loudness	7	2	3	15	**0.037**
Agras 25	LA	Fluctuation	7	1	2	3	0.371
Agras 25	LA	Roughness	7	2	3	6	0.174
Agras 25	LA	Sharpness	7	0	0	1.5	0.269

Continued on next page

UAV	Phase	Outcome	Pairs	Med. Out	Med. In	W	p
Agras 25	LA	Tonality	7	0	1	11	1.000
Agras 25	HC	Annoyance	9	7	7	18.5	0.490
Agras 25	HC	Loudness	9	2	2	9	0.766
Agras 25	HC	Fluctuation	9	1	2	4.5	0.586
Agras 25	HC	Roughness	9	2	2	16	0.792
Agras 25	HC	Sharpness	9	0	0	9	0.824
Agras 25	HC	Tonality	9	0	1	8	1.000
Matrice 350	TO	Annoyance	11	4	5	27.5	0.200
Matrice 350	TO	Loudness	11	1	1	3.5	0.071
Matrice 350	TO	Fluctuation	11	2	1	14	0.326
Matrice 350	TO	Roughness	11	1	1	9	0.766
Matrice 350	TO	Sharpness	11	-1	-1	9	0.824
Matrice 350	TO	Tonality	11	0	0	12.5	0.746
Matrice 350	LA	Annoyance	11	5	7	39.5	**0.048**
Matrice 350	LA	Loudness	11	2	2	10	0.572
Matrice 350	LA	Fluctuation	11	1	1	13	0.930
Matrice 350	LA	Roughness	11	2	1	10	1.000
Matrice 350	LA	Sharpness	11	0	-1	1.5	0.586
Matrice 350	LA	Tonality	11	1	0	7.5	1.000
Matrice 350	HC	Annoyance	11	6	6	16	0.831
Matrice 350	HC	Loudness	11	2	1	9.5	0.250
Matrice 350	HC	Fluctuation	11	1	1	15	0.374
Matrice 350	HC	Roughness	11	1	1	3	0.233
Matrice 350	HC	Sharpness	11	1	0	2	0.089
Matrice 350	HC	Tonality	11	1	1	5	0.572
Matrice 4E	TO	Annoyance	11	3	5	23	0.149
Matrice 4E	TO	Loudness	11	0	0	8.5	0.386
Matrice 4E	TO	Fluctuation	11	1	0	12	0.796
Matrice 4E	TO	Roughness	11	1	0	8.5	0.386
Matrice 4E	TO	Sharpness	11	0	1	23	1.000
Matrice 4E	TO	Tonality	11	0	0	3	0.276
Matrice 4E	LA	Annoyance	11	5	5	26	0.286
Matrice 4E	LA	Loudness	11	1	0	2	0.089
Matrice 4E	LA	Fluctuation	11	1	1	6	0.766
Matrice 4E	LA	Roughness	11	1	0	9	0.430
Matrice 4E	LA	Sharpness	11	1	1	18	0.624
Matrice 4E	LA	Tonality	11	0	0	10	0.548
Matrice 4E	HC	Annoyance	11	4	5	39	0.256
Matrice 4E	HC	Loudness	11	0	0	5	0.583
Matrice 4E	HC	Fluctuation	11	1	0	20	0.821
Matrice 4E	HC	Roughness	11	1	1	12	0.830
Matrice 4E	HC	Sharpness	11	0	0	10	1.000
Matrice 4E	HC	Tonality	11	0	1	15.5	0.339
Mavic Mini 5	TO	Annoyance	11	2	4	27.5	0.582
Mavic Mini 5	TO	Loudness	11	0	-1	16	0.134
Mavic Mini 5	TO	Fluctuation	11	0	0	16	0.792

Continued on next page

UAV	Phase	Outcome	Pairs	Med. Out	Med. In	W	p
Mavic Mini 5	TO	Roughness	11	1	0	18.5	0.482
Mavic Mini 5	TO	Sharpness	11	1	1	39	0.053
Mavic Mini 5	TO	Tonality	11	1	1	13	0.168
Mavic Mini 5	LA	Annoyance	11	4	4	19	0.430
Mavic Mini 5	LA	Loudness	11	0	-1	8.5	0.098
Mavic Mini 5	LA	Fluctuation	11	1	1	6	0.850
Mavic Mini 5	LA	Roughness	11	1	-1	2.5	0.105
Mavic Mini 5	LA	Sharpness	11	0	1	23.5	0.124
Mavic Mini 5	LA	Tonality	11	0	2	25.5	0.058
Mavic Mini 5	HC	Annoyance	11	2	3	17	0.203
Mavic Mini 5	HC	Loudness	11	-1	-1	19	0.713
Mavic Mini 5	HC	Fluctuation	11	0	1	26	0.286
Mavic Mini 5	HC	Roughness	11	0	0	9	0.824
Mavic Mini 5	HC	Sharpness	11	0	1	21	**0.031**
Mavic Mini 5	HC	Tonality	11	0	1	24	0.105

5.5 Systematic Correlation Results

The following matrices use Spearman coefficients, with nominal two-sided p values displayed in each off-diagonal cell. Coefficient magnitude and direction are the principal descriptive information. The probabilities are unadjusted and subject to the small-sample and grouping limitations stated in Section 4.7. They should not be read as a ranking of the physical importance of each descriptor.

5.5.1 Subjective–Subjective Relationships

Outdoor mean annoyance is strongly associated with mean subjective roughness ($\rho = 0.975$, $p < 0.001$), loudness ($\rho = 0.967$, $p < 0.001$) and fluctuation ($\rho = 0.956$, $p < 0.001$). Indoors, the corresponding annoyance–loudness and annoyance–roughness coefficients are 0.970 and 0.958, both with nominal $p < 0.001$. Subjective loudness and roughness themselves correlate strongly: 0.960 outdoors and 0.935 indoors.

Indoor sharpness ratings follow a different ordering: their correlations with subjective loudness and annoyance are negative ($\rho = -0.958$ and -0.942, respectively; both nominal $p < 0.001$). This is a result of the recorded questionnaire values and is not converted into a positive relationship by reversing the scale after analysis. The matrices therefore show both shared response patterns and attributes that do not behave as synonyms for annoyance.

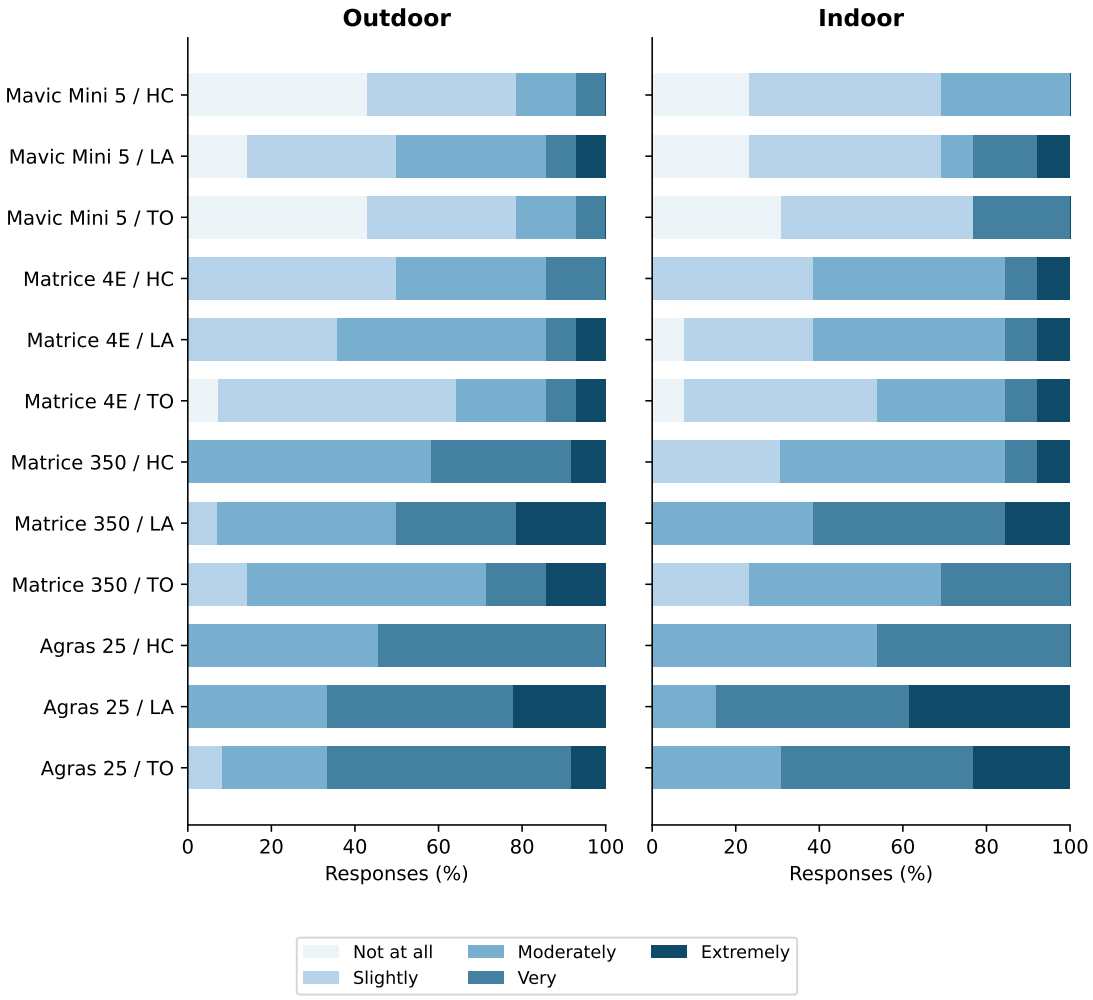


Figura 5.10: Verbal annoyance by UAV and manoeuvre, using the full sample in each environment. Each bar totals 100% of the available responses for that condition; categories are ordered from not at all to extremely annoyed. The original Italian wording is reproduced in Appendix A.

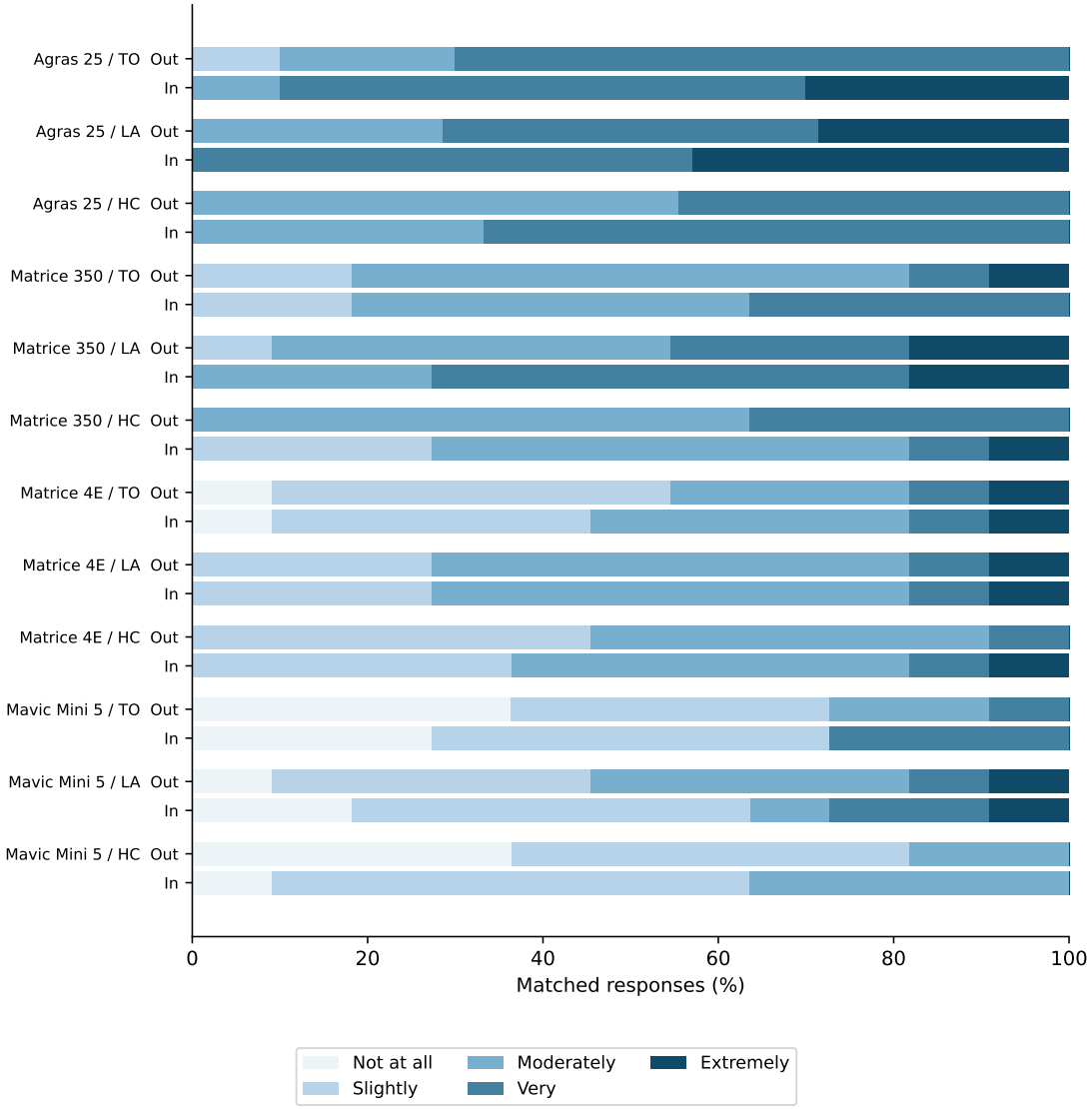


Figura 5.11: Verbal annoyance in the matched sample. Each adjacent Out/In pair uses the same respondents for one UAV–manoeuvre condition. Denominators are seven for Agras landing, ten for Agras take-off, nine for Agras cruise and eleven for all other conditions.

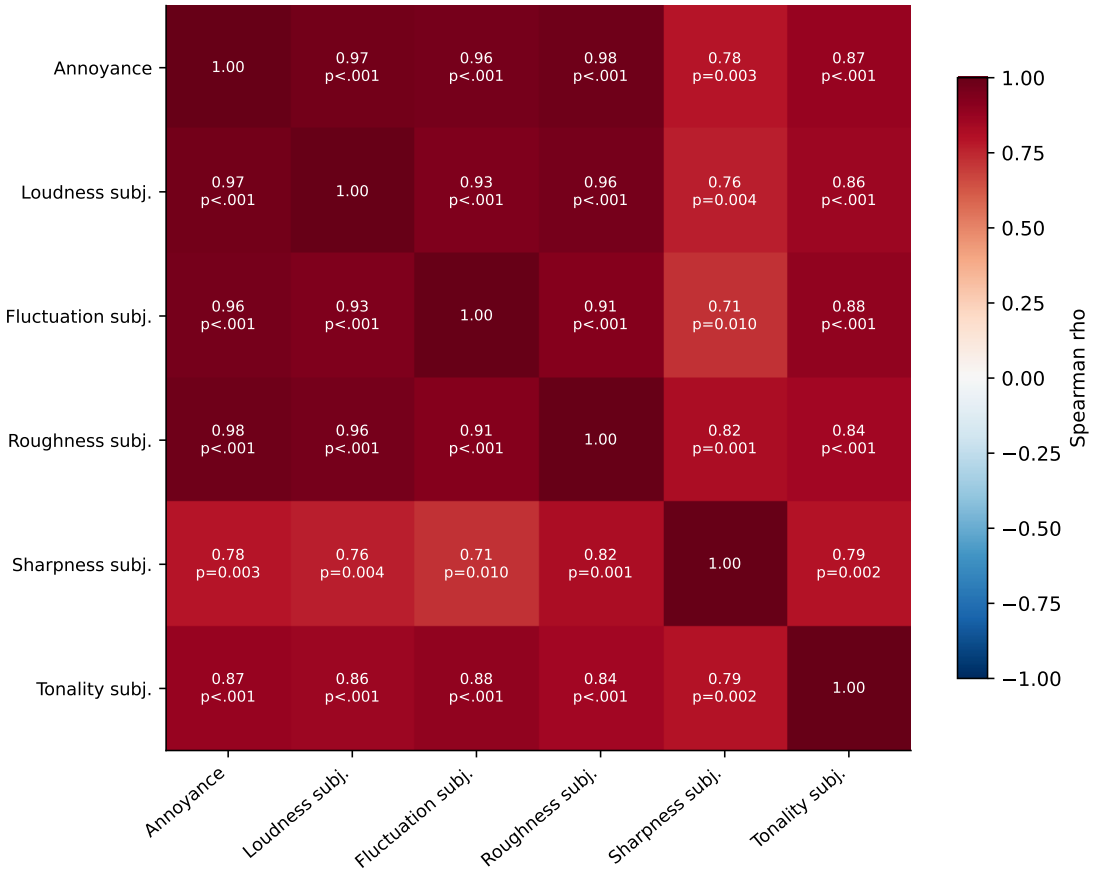


Figura 5.12: Outdoor subjective–subjective matrix of twelve condition means. Cells contain Spearman rho and the nominal two-sided p-value. Means use all available respondents per condition; they are not 156 independent acoustic observations.

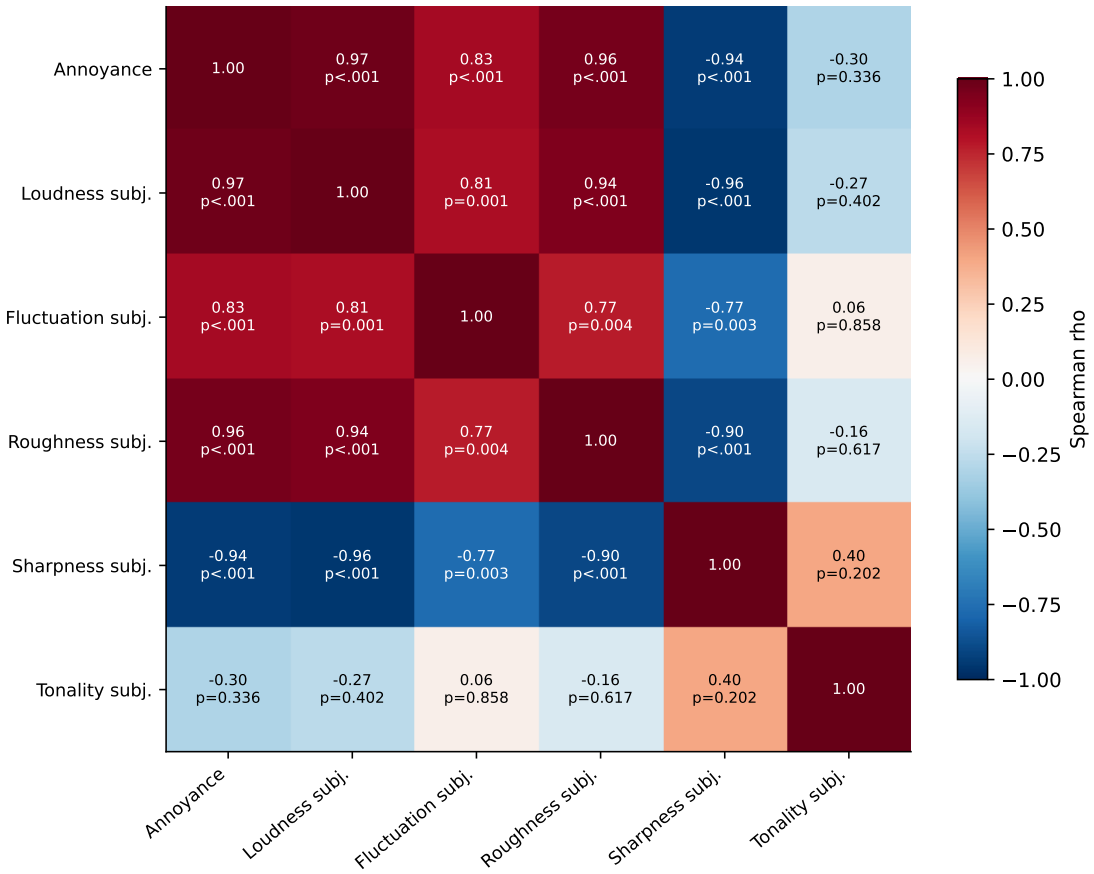


Figura 5.13: Indoor subjective–subjective matrix of twelve condition means. Cells contain Spearman rho and the nominal two-sided p-value. Means use all available respondents per condition; they are not 156 independent acoustic observations.

5.5.2 Objective–Objective Relationships

Across all 28 events, N5 and N10 are almost redundant in rank ordering ($\rho = 0.998$, nominal $p < 0.001$). HM loudness is also strongly associated with N5 and N10 ($\rho = 0.975$ and 0.978). Fluctuation strength is associated with HM loudness ($\rho = 0.861$), indicating that part of its pooled variation follows the same broad aircraft–condition pattern.

The extended HDF–XL2 matrix is reported in Appendix D. XL2 equivalent level and SEL correlate at $\rho = 0.980$, while N10 and SEL correlate at $\rho = 0.876$; the nominal probabilities are below 0.001. The event-level scatterplots in Figure 5.15 show the underlying grouping. These results support a common magnitude/exposure component, while retaining the different definitions, durations and acquisition perspectives of the metrics. They do not imply that one instrument's value can be substituted numerically for another's.

5.5.3 Objective Descriptors and Mean Responses

Objective loudness has the clearest cross-domain association with annoyance. For HM loudness, $\rho = 0.970$ outdoors and 0.986 indoors; for N5 and N10 the coefficients are 0.981 outdoors and 0.993 indoors. All corresponding nominal probabilities are below 0.001. The same HM descriptor is strongly related to its subjective counterpart ($\rho = 0.986$ outdoors and 0.963 indoors; both $p < 0.001$).

Other descriptors have more differentiated relationships. Objective roughness correlates with subjective roughness at $\rho = 0.655$ outdoors ($p = 0.021$) and 0.762 indoors ($p = 0.004$). Objective fluctuation strength correlates with its subjective counterpart at 0.839 outdoors ($p < 0.001$) and 0.676 indoors ($p = 0.016$). Objective sharpness and subjective sharpness have a positive outdoor association (0.592, $p = 0.043$), but a negative indoor association (-0.851, $p < 0.001$).

Objective tonality does not display a stable positive relationship with rated tonality. The outdoor association is negative ($\rho = -0.712$, $p = 0.009$), whereas the indoor coefficient is close to zero ($\rho = -0.032$, $p = 0.923$). Its correlations with annoyance are also negative, at -0.764 outdoors and -0.650 indoors. These pooled associations require interpretation in relation to aircraft differences and the particular metric and rating definitions; they do not demonstrate a beneficial causal effect of tonality.

The XL2 quantities also track mean annoyance. SEL has $\rho = 0.914$ outdoors and 0.916 indoors, both with nominal $p < 0.001$. Full XL2–subjective matrices, paired-sample sensitivity results and individual-response summaries are supplied in the supplement. The four HM-loudness relationships illustrated in Figure 5.18 show the underlying condition points alongside the simple regression lines discussed below.

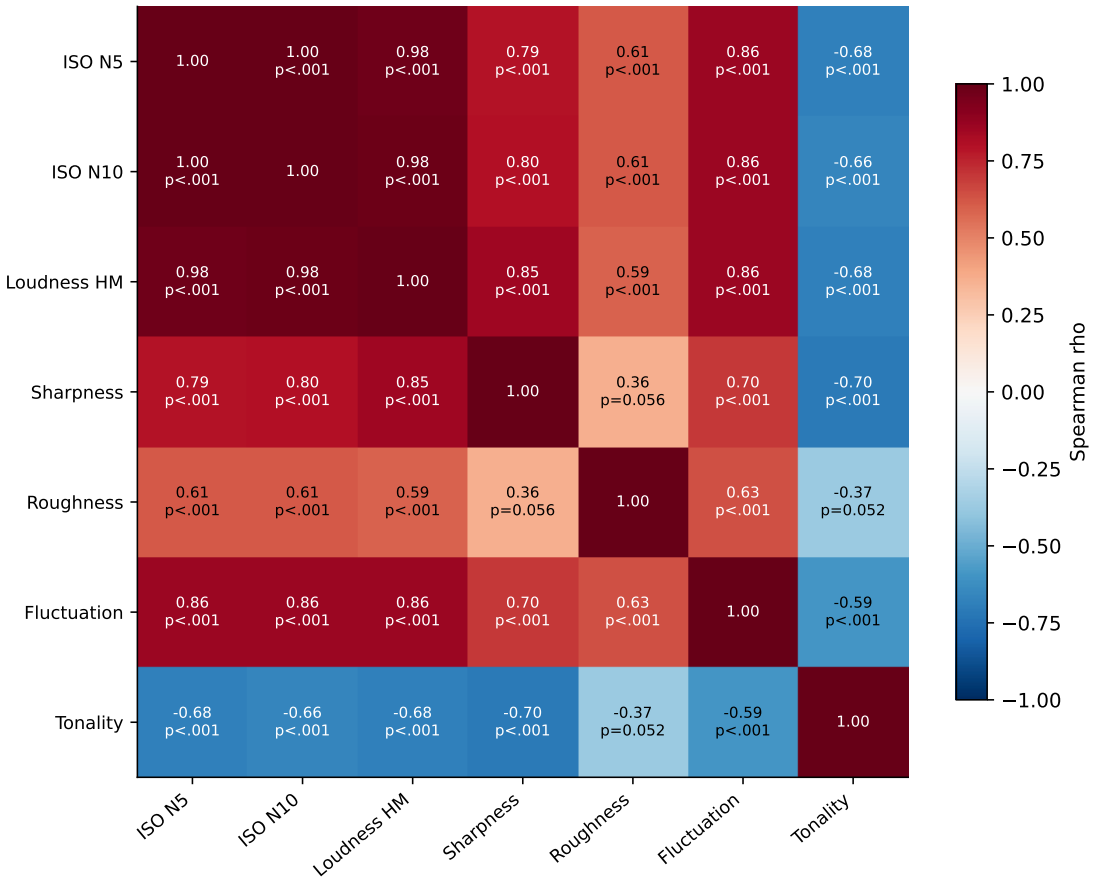


Figura 5.14: Objective–objective Spearman matrix using all 28 independent HDF events. Unequal repetition counts are retained. Cells show rho and nominal p; the observations are grouped within twelve conditions and four UAVs.

Tabella 5.5: Spearman correlations between objective descriptors and mean annoyance for the twelve evaluated conditions per environment. Probabilities are nominal, two-sided and unadjusted.

Descriptor	ρ_{Out}	p_{Out}	ρ_{In}	p_{In}
ISO N5	0.981	< 0.001	0.993	< 0.001
ISO N10	0.981	< 0.001	0.993	< 0.001
Loudness HM	0.970	< 0.001	0.986	< 0.001
Sharpness	0.907	< 0.001	0.874	< 0.001
Roughness	0.616	0.033	0.671	0.017
Fluctuation	0.851	< 0.001	0.874	< 0.001
Tonality	-0.764	0.004	-0.650	0.022

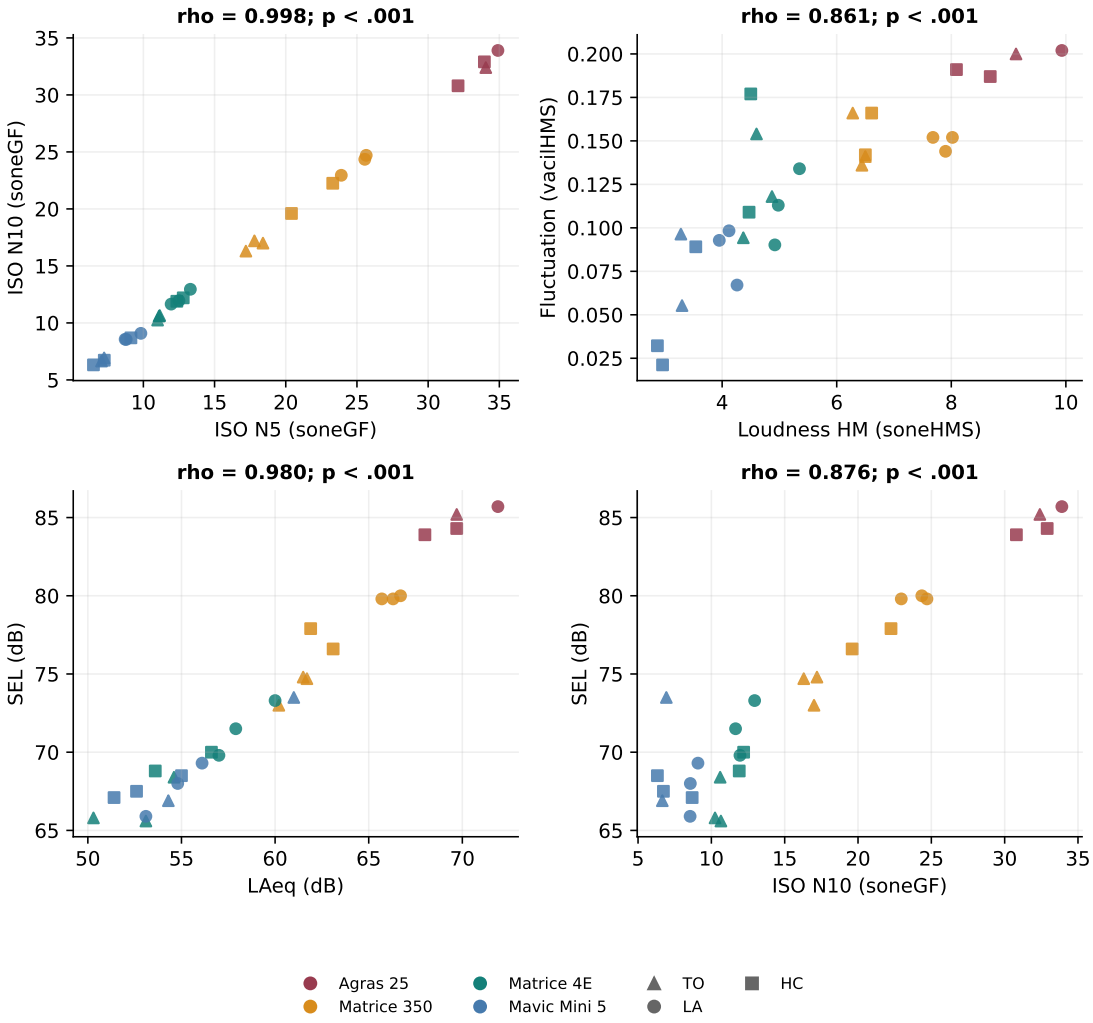


Figura 5.15: Selected objective–objective associations across all 28 events. Colours identify UAVs and marker shapes identify manoeuvres. Each point is one physical event. Coefficients and nominal probabilities are the corresponding cells of the extended objective matrix; the grouping shows why a pooled association need not describe a within-UAV causal relationship.

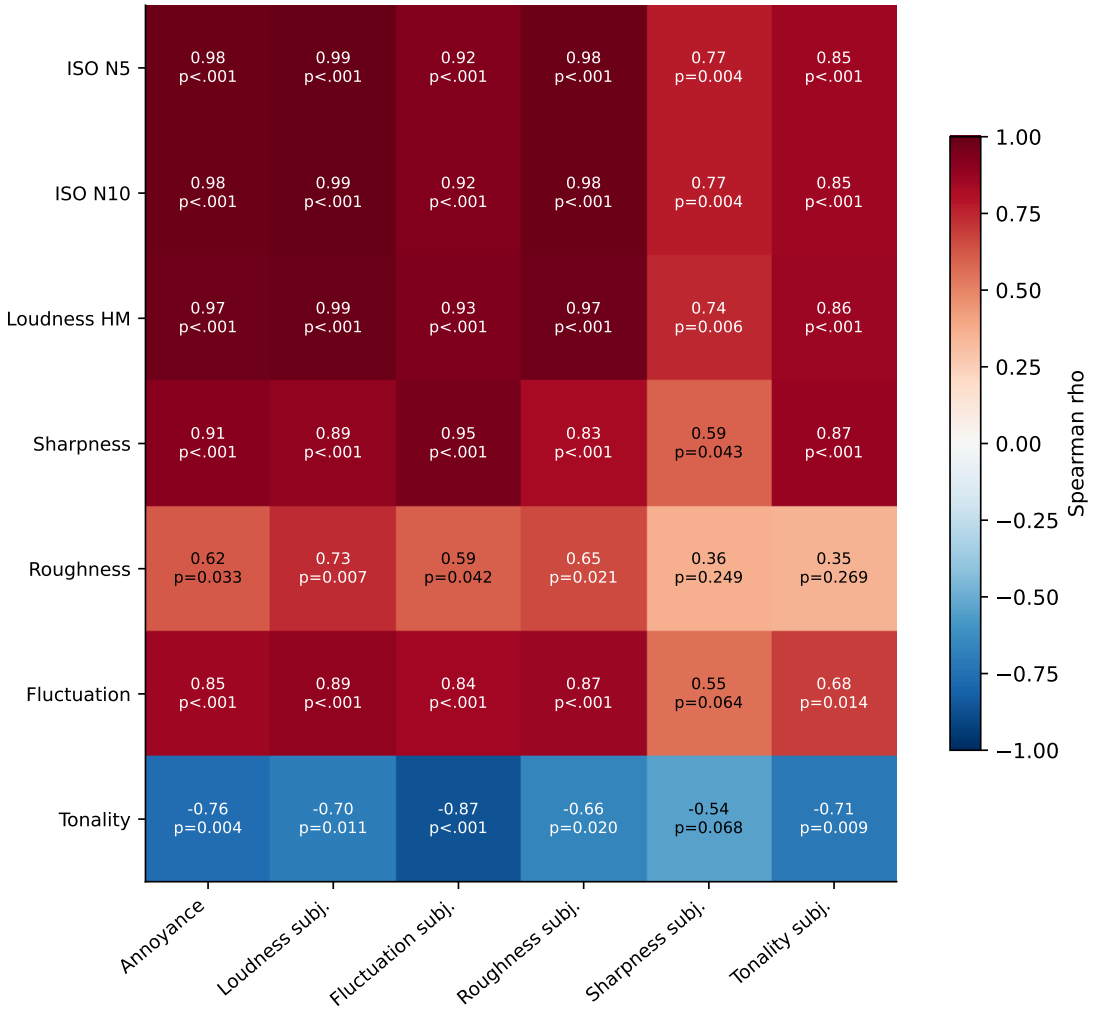


Figura 5.16: Outdoor objective–subjective Spearman matrix. One row of the analysis dataset corresponds to one of the twelve evaluated recordings and its condition-mean responses. Additional flight repetitions are excluded from this cross-domain analysis. Cells give rho and nominal p.

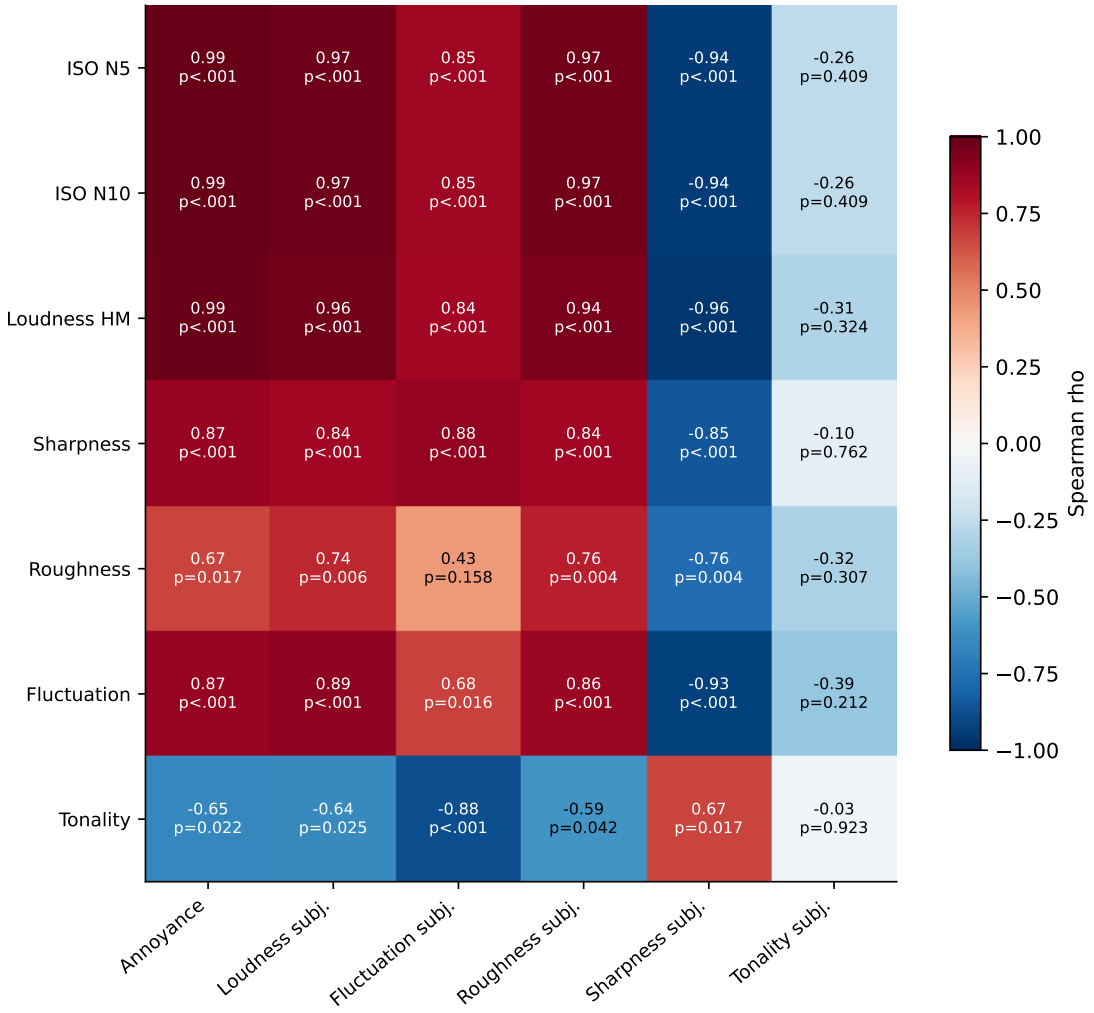


Figura 5.17: Indoor objective–subjective Spearman matrix. One row of the analysis dataset corresponds to one of the twelve evaluated recordings and its condition-mean responses. Additional flight repetitions are excluded from this cross-domain analysis. Cells give rho and nominal p.

5.5.4 Individual-Response and Sample-Composition Sensitivity

The participant-centred linear coefficient between HM loudness and annoyance is 0.665 outdoors and 0.744 indoors. Its exploratory crossed-bootstrap intervals are 0.429–0.842 and 0.493–0.884. The corresponding coefficients for subjective loudness are 0.746 and 0.834. Table D.1 reports these supplementary results. They support a positive individual-response association, while being weaker than several condition-mean associations. Averaging across listeners reduces part of the response variation, so a very high correlation of means should not be interpreted as equally accurate prediction of an individual listener.

Restricting the condition means to matched respondents retains the principal HM-loudness association with annoyance: $\rho = 0.965$ outdoors and 0.970 indoors, both nominal $p < 0.001$. This sensitivity result supports the broad loudness pattern, although it does not remove the limited number of UAVs or shared-listener structure.

5.6 Simple Regression Models

5.6.1 Predicting Mean Annoyance from One Objective Descriptor

Table 5.6 presents the eight objective-only single-predictor fits per environment. HM loudness explains 93.0% of the observed between-condition variance in outdoor mean annoyance and 97.1% indoors. The fitted equations, with HM loudness N_{HM} in soneHMS, are

$$\widehat{A}_{\mathrm{Outdoor}} = 0.632733 + 0.689095\, N_{\mathrm{HM}}, \tag{5.1}$$

$$\widehat{A}_{\mathrm{Indoor}} = 1.219109 + 0.665324\, N_{\mathrm{HM}}. \tag{5.2}$$

The coefficients come from the least-squares calculation in Equation 4.7. For example, 0.632733 is the outdoor intercept and 0.689095 is its slope, in annoyance-scale points per soneHMS. Neither number is a correlation coefficient. The intercept is part of the fitted equation and is not evidence of the response to a silent event outside the measured range.

N5 and N10 give slightly higher outdoor R^2 values, 0.953 and 0.954, than HM loudness. Indoors, HM loudness gives the highest R^2 among these candidates. SEL also provides strong in-sample fits ($R^2 = 0.914$ outdoors and 0.905 indoors). These values do not establish that one loudness implementation is universally superior: they concern the same small set of measured conditions, and the competing loudness predictors are highly correlated.

Roughness alone gives substantially lower linear fits ($R^2 = 0.245$ outdoors and 0.253 indoors) and slope probabilities of 0.102 and 0.096. This illustrates the difference between its positive Spearman association with annoyance and the evidence from an ordinary straight-line model. Sharpness, fluctuation and tonality have nominally non-zero slopes in both environments, but their fits and signs should be read together with the objective correlations rather than treated as independent effects.

Tabella 5.6: Single-predictor OLS models of mean numerical annoyance. Each fit uses 12 condition means. R^2 and adjusted R^2 describe in-sample fit; p is the nominal two-sided slope test.

Predictor	R^2_{Out}	$R^2_{\mathrm{adj,Out}}$	p_{Out}	R^2_{In}	$R^2_{\mathrm{adj,In}}$	p_{In}
Loudness HM	0.930	0.923	< 0.001	0.971	0.968	< 0.001
ISO N5	0.953	0.949	< 0.001	0.935	0.928	< 0.001
ISO N10	0.954	0.950	< 0.001	0.937	0.931	< 0.001
SEL	0.914	0.906	< 0.001	0.905	0.896	< 0.001
Roughness	0.245	0.169	0.102	0.253	0.178	0.096
Fluctuation	0.733	0.706	< 0.001	0.727	0.700	< 0.001
Sharpness	0.746	0.721	< 0.001	0.822	0.805	< 0.001
Tonality	0.508	0.458	0.009	0.453	0.398	0.017

5.6.2 Predicting the Corresponding Subjective Attribute

Table 5.7 distinguishes predicting annoyance from predicting the descriptor's perceptual counterpart. HM loudness gives strong fits to subjective loudness ($R^2 = 0.918$ outdoors and 0.963 indoors). Fluctuation strength gives $R^2 = 0.798$ outdoors and 0.535 indoors. Roughness gives weaker fits, with a nominally non-zero slope indoors but not outdoors.

Sharpness provides an important qualification. Its outdoor linear slope is not below the nominal 0.05 threshold, despite the positive rank association. Indoors the fitted slope is negative, with $R^2 = 0.568$ and $p = 0.005$. Tonality gives a negative outdoor slope and an essentially null indoor fit ($R^2 = 0.001$, $p = 0.932$). These outcomes prevent a general claim that each objective metric directly predicts the similarly named subjective attribute.

5.6.3 Does Roughness Add Information to Loudness?

Adding objective roughness to HM loudness changes outdoor annoyance R^2 from 0.929667 to 0.931100 and indoor R^2 from 0.970572 to 0.972296. The increases are small. Adjusted R^2 decreases in both environments, from 0.922634 to 0.915788

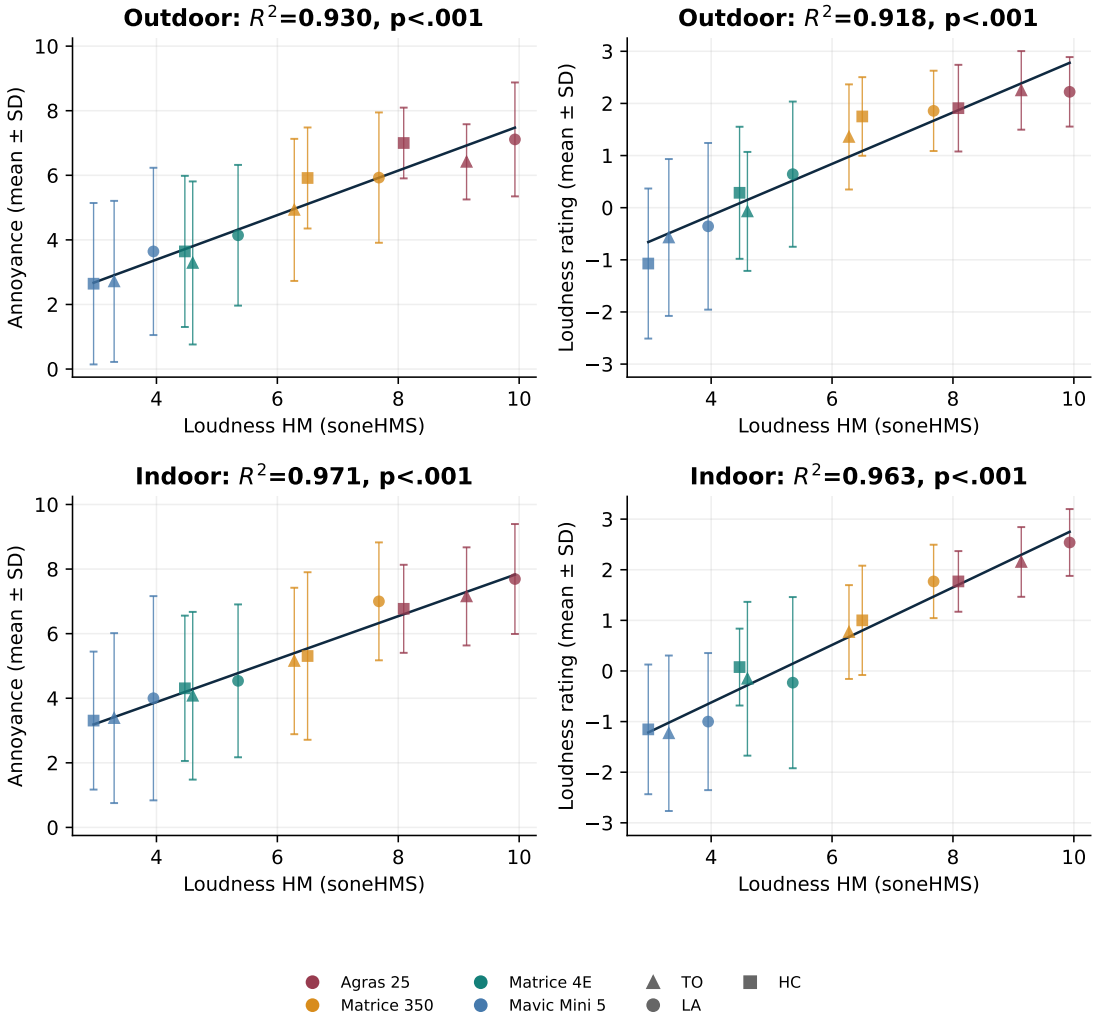


Figura 5.18: HM loudness versus mean numerical annoyance and mean subjective loudness, separately outdoors and indoors. Each point is one evaluated UAV–manoeuvre condition; colour identifies UAV and marker identifies manoeuvre. Vertical bars show the sample standard deviation of the ratings in each condition, not uncertainty of the fitted mean. Lines are OLS fits over the observed predictor range. Reported R-squared and p-values concern in-sample linear models, not external predictive validation.

Tabella 5.7: OLS models of the corresponding subjective attribute. b is the unstandardised slope; units are rating points per objective unit. Each model has 12 condition means.

Predictor / attribute	b_{Out}	R^2_{Out}	p_{Out}	b_{In}	R^2_{In}	p_{In}
Loudness HM	0.493	0.918	< 0.001	0.569	0.963	< 0.001
Sharpness	0.681	0.223	0.121	-1.259	0.568	0.005
Roughness	1.334	0.291	0.070	1.918	0.404	0.026
Fluctuation	10.229	0.798	< 0.001	7.570	0.535	0.007
Tonality	-5.926	0.348	0.044	0.250	0.001	0.932

outdoors and from 0.967630 to 0.966140 indoors. The conditional roughness probabilities are 0.675 and 0.473. Thus, the added coefficient does not provide persuasive incremental evidence in these fits.

The subjective models tell a similar story. Subjective loudness alone gives $R^2 = 0.931$ outdoors and 0.957 indoors. Adding subjective roughness gives $R^2 = 0.933$ and 0.957, with conditional roughness probabilities of 0.563 and 0.829. Strong correlation between the two ratings limits the independent information contributed by the second variable. These models remain relationships among subjective means; they do not provide an objective estimator that can operate without listening-test responses.

Tabella 5.8: Two-predictor comparisons for mean annoyance. The last column tests the roughness coefficient conditional on loudness.

Environment	Predictors	R^2	R^2_{adj}	$p_{\text{roughness}}$
Outdoor	Objective HM + roughness	0.931	0.916	0.675
Outdoor	Rated loudness + roughness	0.933	0.919	0.563
Indoor	Objective HM + roughness	0.972	0.966	0.473
Indoor	Rated loudness + roughness	0.957	0.948	0.829

5.6.4 Diagnostics and Evidential Boundaries

The complete set of 34 fits and the two-predictor coefficients are reported in Appendix E. Diagnostic summaries do not justify presenting the equations as validated general predictors. In particular, the indoor HM-annoyance fit has a Shapiro–Wilk probability of 0.021 and a maximum absolute studentised residual of approximately 4.21. The indoor N5 and N10 annoyance fits also show departures from residual normality in these summaries. These findings qualify the ordinary coefficient probabilities despite their very small nominal values.

The principal result is therefore the strong, interpretable in-sample loudness pattern, together with the limited additional information from roughness. No held-out prediction error or performance on a new UAV is claimed. The complete candidate results, including non-significant slopes, are retained so that the reported interpretation is not based only on successful fits.

Capitolo 6

Discussion

6.1 What the Study Establishes

The combined analyses identify a consistent relation between acoustic magnitude and subjective response in the tested UAV events. Loudness-related descriptors track condition-mean annoyance and subjective loudness in both environments, while the roles of roughness, fluctuation, sharpness and tonality are more dependent on the particular response and analysis. The paired results provide limited evidence for individual outdoor–indoor differences, but they do not establish that the two listening situations are equivalent.

These findings operate at different levels. The 28 acoustic recordings describe physical-event variability. The subjective distributions describe the listeners' ratings. The paired tests address within-person changes for particular conditions. The mean-based correlations and regressions describe relationships among twelve condition summaries. A strong result at one level cannot automatically be transferred to another: a close fit to twelve means is not a close fit to every individual rating, and a correlation across different UAVs is not necessarily a relationship within repeated flights of one UAV.

6.2 UAV, Manoeuvre and Event Variability

The higher levels and loudness values of the Agras conditions, together with generally higher annoyance ratings, are consistent with a magnitude-related interpretation of the broad platform differences. Matrice 350 also generally occupies higher-annoyance conditions than the smaller platforms. However, the Matrice 4E and Mini 5 field levels overlap in a manoeuvre-dependent way. A single ranking of four aircraft is therefore less informative than a comparison that retains the three manoeuvres.

The programmed trajectories change source height, distance and orientation during the event. These physical changes offer plausible reasons for differences in exposure, spectral content and temporal behaviour. The recordings do not provide independently verified rotor-speed or loading histories, so those mechanisms cannot be identified quantitatively from the response plots. In particular, it would be inappropriate to explain every perceptual difference solely through UAV size without separating operating condition and geometry.

Independent-event variability further qualifies the use of a representative recording. Repeated Matrice 4E take-offs show appreciable variation in roughness, while the Agras horizontal examples show that loudness and roughness need not change in the same direction between repetitions. The first evaluated event remains the correct source for objective–subjective correspondence, but it should not be regarded as an exact acoustic average of every possible execution of that manoeuvre.

The number of repetitions is too small and uneven for a general characterisation of repeatability. A box based on two or three events is primarily a compact display of those observations, rather than evidence of a well-estimated population distribution. The singleton Agras conditions cannot provide a sample standard deviation. These limitations favour retaining the event points and explicit counts, as done here.

6.3 Outdoor Exposure and Indoor Binaural Listening

Comparing returning listeners reduces confounding from stable differences between people, but the outdoor and indoor conditions still differ in several respects. Participants experienced a live visible event outdoors, whereas indoors they heard a recording from the headset operator's perspective. The experiments occurred in different months, and the indoor sessions were distributed across the Audio Space Lab and departmental offices. Familiarity from the earlier session, attention, expectation and the surrounding setting may all contribute to a change in judgement.

The fixed playback configuration provides a consistent procedure within the indoor experiment. It does not demonstrate that all aspects of the field experience were reproduced. The BHS II is a wearable binaural acquisition device, and its two channels reflect the recording position and the operator's ear-related geometry. The listeners did not necessarily occupy that same position outdoors. The approximate one-metre separation from the ZYLIA/XL2 position is also relevant when linking instrumental levels to the binaural descriptors.

The study did not measure localisation or externalisation as separate outcomes. Consequently, annoyance and attribute ratings cannot establish complete spatial

fidelity. This is consistent with the distinction in the binaural literature between reproduction of ear-related signals and the perceptual effects of the headphone transfer and head-motion cues [15, 14]. The outdoor–indoor analysis should be described as a comparison between these two experimental situations, rather than a controlled test of the recording device alone.

6.4 Interpreting the Paired Wilcoxon Evidence

The three nominally significant comparisons concern different outcomes and conditions: loudness for Agras landing, annoyance for Matrice 350 landing and sharpness for Mini 5 cruise. They do not define a uniform increase in every subjective attribute indoors. Each should be interpreted with its medians, the paired differences and the limited number of non-zero ranks.

The general Wilcoxon table retains the medians because they describe the response distribution without requiring an arithmetic-mean interpretation of the attribute scale. Nevertheless, the signed-rank statistic uses all ranked non-zero differences. It is not a test calculated from the displayed median pair alone, nor does a significant result imply that every listener changed in the same way.

The condition-specific results are interpreted with their small paired samples, tied and zero differences, and the normal approximation underlying the exported probabilities. Low power and absence of rejection must not be presented as evidence of equivalence; that would require a prespecified equivalence margin and an appropriate analysis.

6.5 Correlations in the Physical Context of Drone Sound

6.5.1 Loudness and the Common Magnitude Component

The agreement among N5, N10 and HM loudness, together with their associations with conventional level and exposure, suggests that a substantial component of the dataset is organised by acoustic magnitude. Their strong relationships with mean annoyance and subjective loudness are therefore physically interpretable. The result is consistent with literature identifying loudness as an important correlate of UAV annoyance [9, 1].

This interpretation also explains why a matrix can contain many large coefficients without demonstrating many independent perceptual mechanisms. If loudness, fluctuation and sharpness all increase across the same broad aircraft contrast, each may correlate with annoyance even if it contributes little additional information

after loudness is included. The full objective matrix is essential for recognising that redundancy before proposing combinations of predictors.

N5 and N10 describe closely related exceedance statistics of the same loudness signal. Their near-identical rank ordering is not a reason to combine both in a small regression as if they were independent sources of information. Similarly, the strong SEL and equivalent-level association does not remove the conceptual importance of duration; it only indicates that the observed events show similar overall ordering under these two descriptors.

6.5.2 Roughness and Fluctuation Strength

Roughness and fluctuation strength concern temporal properties of sound, but their algorithms and subjective labels do not represent interchangeable quantities. Their positive associations with the corresponding mean ratings indicate perceptually relevant structure in these recordings. Fluctuation gives a particularly clear outdoor relationship, while objective roughness produces a weaker straight-line annoyance fit than loudness.

The distinction between rank and linear analyses is important here. A positive Spearman coefficient can occur when conditions are mostly ordered together even if their numerical spacing does not align closely with a straight line. It is therefore consistent for objective roughness to have a nominally significant rank association with annoyance while its single-predictor OLS slope does not meet the same nominal threshold. The two analyses address different aspects of the relationship and use different working assumptions.

Subjective roughness is highly correlated with subjective loudness and annoyance. This may reflect shared perceptual responses, common use of the scales or correlated acoustic features. The present design does not separate these explanations. The limited benefit of adding roughness to loudness is more informative for model parsimony than the size of the roughness–annoyance correlation alone.

6.5.3 Sharpness and Tonality

The opposite signs of the outdoor and indoor sharpness associations prevent a simple interpretation of objective sharpness as a universal estimator of the questionnaire's sharpness item. The original item refers to how *acuto* the sound is; its interpretation can involve spectral pitch-related impressions that are not identical to the Aures sharpness algorithm. Context and response-scale use may also matter. The analysis retains the recorded coding rather than reversing it to produce an expected sign.

Tonality provides another example of the need for physical context. Its negative pooled relationship with annoyance does not show that increasing a tonal component reduces annoyance. Aircraft with different loudness and spectral signatures

can occupy different regions of both variables, creating a negative cross-platform association. Moreover, the objective Hearing-Model tonality and the subjective judgement of how *fischiante* a sound is need not measure precisely the same perceptual construct. The near-zero indoor objective–subjective tonality relationship is a substantive limitation on using this metric as a stand-alone predictor of that rating.

These findings support analysing each objective descriptor against both annoyance and its nominally corresponding subjective attribute. A descriptor can help order annoyance without reproducing the similarly named rating, and the reverse can also occur. The terminology alone does not establish measurement validity.

6.5.4 Means and Individual Responses

The condition means provide a readable summary of the relation between the evaluated events and the group's responses. However, averaging removes part of the between-listener variation. The lower participant-centred HM-loudness coefficients relative to the mean-based coefficients demonstrate why conclusions about individual prediction must remain separate.

The supplementary individual analysis preserves repeated-listener information through centring and exploratory resampling, but its intervals do not fully resolve grouping within four UAVs. It is therefore used as a check on the direction and persistence of associations, not as an additional set of ordinary significance tests. Agreement of the main loudness pattern in the paired-mean sensitivity analysis is reassuring about sample composition, while still leaving the limited range of aircraft as a major boundary.

6.6 What the Simple Models Add

The regressions translate selected correlations into explicit numerical relationships. The outdoor HM-annoyance equation, for example, expresses how the mean annoyance estimate changes per soneHMS within the observed range. This is more directly interpretable for the present dataset than a model with many transformations or a complex selection procedure. It also makes clear that the intercept and slope are estimated from the paired condition values, rather than read from a correlation coefficient.

HM loudness is a useful reference model because it gives strong fits in both environments and closely follows subjective loudness. The slightly higher outdoor fit of N5/N10 and the higher indoor fit of HM should be regarded as dataset-specific comparisons. These metrics are strongly related, and twelve conditions are insufficient to establish a general winner among formulations.

The combined models answer a narrower but useful question: does roughness add information after loudness has already been included? In both objective and subjective versions the increase in ordinary R^2 is very small, adjusted R^2 decreases, and the conditional roughness probability is not small. On this evidence, the simpler loudness model is the more defensible summary. This does not prove that roughness can never matter; it shows that the present sample does not support the additional coefficient as a useful independent contribution.

High R^2 must be interpreted at the level at which it was calculated. It describes between-condition variation in group means on the same data used for fitting. It is not the proportion of individual annoyance explained, and it does not measure accuracy on new drones. The indoor residual diagnostics also show that a very strong fit can coexist with departures from the model's distributional assumptions. For these reasons, the equations are reported as exploratory empirical models and not as validated tools for noise regulation or aircraft design optimisation.

6.7 Limitations and Priorities for Further Work

The most consequential limitation is the narrow set of acoustic conditions: four UAVs and three manoeuvres at one field site. Conditions share aircraft characteristics and respondents, while repetitions are uneven. Correlation probabilities and ordinary OLS tests therefore provide less secure inference than their numerical values alone might suggest.

The listener samples are also small, and outdoor attendance varies by condition. Returning listeners provide valuable paired data, but the fixed order and the interval between the two experiments prevent separation of environment, familiarity and order effects. The two indoor-only listeners contribute to indoor descriptions without a paired counterpart, and equivalent preliminary demographic information is unavailable for them. The exact wording of all indoor instructions and a participant-by-room allocation are not documented in the material available for the analysis.

Acquisition geometry, event duration and background require similar care. The HDF–XL2 mapping identifies physical-event correspondence, not identical measurement conditions. Background checks show insufficient margins in some frequency bands of quieter recordings. Since the waveforms and psychoacoustic metrics were not recomputed after energetic band correction, these contributions remain part of the analysed sound.

A follow-up study should expand the number of UAV platforms and independently recorded events, retain exact event identifiers across acquisition chains, and collect balanced repetitions. Subjective tests should balance or randomise the main sequence, document the room and listening conditions per session, and maintain

consistent questionnaire wording. New listeners and new recordings would permit assessment of whether the present simple loudness relationships generalise. Specific tests of the contribution of roughness, sharpness or tonality would benefit from designs that vary these characteristics while limiting changes in loudness and other correlated factors.

6.8 Synthesis

The study supports a clear magnitude-related organisation of the tested drone sounds and a strong association between loudness and group-mean annoyance. It also shows why the analysis cannot stop at large coefficients: event variability, respondent distributions, inconsistent attribute relationships and residual diagnostics all qualify the interpretation. Within these limits, the final simple regressions provide concise descriptions, while the paired comparisons and systematic matrices preserve the distinction between environmental changes, subjective constructs and objective acoustic descriptors.

Capitolo 7

Conclusions

7.1 Main Findings

This thesis connected field acquisition, outdoor subjective assessment and subsequent binaural indoor listening for four UAVs during take-off, landing and horizontal cruise. A traceable correspondence was established between the twelve evaluated recordings, the subjective responses and the associated XL2 field events. The wider archive of 28 distinct recordings was retained for objective variability and objective–objective correlations, without assigning subjective ratings to unassessed repetitions.

The first research question concerned acoustic differences and event variability. Conventional levels and psychoacoustic descriptors varied across aircraft and manoeuvres, with the Agras events generally occupying the higher-magnitude part of the selected dataset. Independent executions of the same nominal manoeuvre also differed. Explicit event counts were essential: one to three recordings per condition support a description of the available variability, rather than a general estimate of measurement precision. Background checks further identified less favourable margins in some bands of the quieter recordings.

The second question concerned outdoor and indoor subjective responses. Distributional plots preserved numerical annoyance, verbal categories and the five attribute ratings as distinct outcomes. The matched comparison used eleven returning listeners and 125 complete condition pairs. Of 72 condition-specific Wilcoxon signed-rank tests, three had original two-sided $p < 0.05$: Agras landing loudness, Matrice 350 landing annoyance and Mini 5 cruise sharpness. The available results therefore identify a few nominal differences, but do not establish either a uniform environmental effect or equivalence of field exposure and binaural playback.

The third question concerned systematic associations. Loudness-related metrics gave the clearest cross-domain relationships. HM loudness correlated with mean

annoyance at Spearman $\rho = 0.970$ outdoors and 0.986 indoors, and also closely tracked mean subjective loudness. The full objective matrix showed that several descriptors share a common ordering across events, so separate strong correlations do not imply independent contributions. Fluctuation and roughness showed positive relationships with their perceptual counterparts, but sharpness and tonality were less consistent across environments. These results underline the difference between a physical descriptor, its similarly named questionnaire item and annoyance.

The fourth question concerned simple models. Unstandardised OLS regressions of the twelve condition means produced transparent equations, with HM-loudness annoyance fits of $R^2 = 0.930$ outdoors and 0.971 indoors. N5, N10 and SEL also gave strong in-sample relationships. Adding roughness to loudness produced only very small increases in ordinary R^2, reduced adjusted R^2 and yielded non-significant conditional roughness coefficients. The simpler loudness models therefore provide the more parsimonious summary of these data.

7.2 Contribution and Applicability

The main contribution is an integrated and traceable analysis of dynamic UAV sound under two listening situations. Separating independent flight events, evaluated stimuli, condition means and repeated listener responses allowed each analysis to address its intended question without artificially increasing the number of independent observations. The final presentation of distributions, medians, original two-sided paired-test probabilities, systematic matrices and simple equations makes the evidence inspectable at each stage.

The findings support loudness as a useful starting descriptor for explaining the relative annoyance of these recordings. They do not establish that loudness is sufficient for every drone sound or listening context. The inconsistent sharpness and tonality results and the limited added value of roughness indicate where further controlled investigation is needed, rather than showing that these sound qualities are inherently irrelevant.

The regression equations describe the measured conditions and their group means. They have not been validated on new UAVs, recordings or participants. High in-sample fit, small nominal probabilities and the apparent similarity of two listening conditions must therefore not be translated into general predictive or equivalence claims.

7.3 Future Development

The next step is to test the stability of the observed loudness relationships using more aircraft, balanced independent repetitions and additional listeners. Exact

cross-instrument event correspondence, consistent questionnaire instructions and documented listening-room conditions should be retained throughout. Designs that vary spectral or temporal attributes while controlling loudness would allow a more direct assessment of their contribution to annoyance. Such work would build on the present empirical models while addressing the limitations that prevent their immediate generalisation.

Appendice A

Outdoor Subjective Evaluation Questionnaire

The original Italian outdoor questionnaire is reproduced on the following five pages. Its numerical annoyance, verbal annoyance and five sound-quality items define the response variables used in the study. The available indoor dataset uses the same response variables and coding; this reproduction does not establish word-for-word equivalence of every instruction in the indoor form. The blank sixth page of the source document is omitted.

Seconda Parte

Test Psicoacustico

* Indica una domanda obbligatoria

1. ID partecipante *

Inserisci la seconda lettera del nome di tua madre, le ultime due cifre del tuo anno di nascita, l'ultima lettera del nome di tuo padre

2. Drone *

Contrassegna solo un ovale.

- Matrice 350
- Mavic Mini 5
- Mavic Pro/ Matrice 4
- Agras 25

3. Manovra *

Contrassegna solo un ovale.

- Decollo
- Atterraggio
- Volo orizzontale

4. 1.Pensando all'esperienza uditiva appena vissuta, **quanto** il rumore del drone che hai appena ascoltato ti ha **infastidito**, **disturbato** o **irritato**? *

Contrassegna solo un ovale.

- ◯ Estremamente
- ◯ Molto
- ◯ Moderatamente
- ◯ Poco
- ◯ Per niente

5. 2. Pensando all'esperienza uditiva appena vissuta, quale **numero da zero a dieci** rappresenta quanto il rumore del drone ti ha **infastidito**, **disturbato** o **irritato**? *

Contrassegna solo un ovale.

	0	1	2	3	4	5	6	7	8	9	10	
Per	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Estremamente

3. Valuta le **caratteristiche** del rumore che hai ascoltato

6. 3.1 Intensità *

-3 = Per niente **forte**
3 = Estremamente **forte**
Es. “La TV è **forte**, non riusciamo a sentirci”

Contrassegna solo un ovale.

- -3
- -2
- -1
- 0
- 1
- 2
- 3

7. 3.2 Fluttuazione *

-3 = Per niente **fluttuante**
3 = Estremamente **fluttuante**
Es. “Il rumore è **fluttuante** come le onde del mare”

Contrassegna solo un ovale.

- -3
- -2
- -1
- 0
- 1
- 2
- 3

8. 3.3 Ruvidezza *

-3 = Per niente **ruvido**
3 = Estremamente **ruvido**
Es. “La fontana scrosciava con un suono **ruvido**”

Contrassegna solo un ovale.

- -3
- -2
- -1
- 0
- 1
- 2
- 3

9. 3.4 Acutezza *

-3 = Per niente **acuto**
3 = Estremamente **acuto**
Es. “Lo strillo **acuto** del bambino svegliò il fratello”

Contrassegna solo un ovale.

- -3
- -2
- -1
- 0
- 1
- 2
- 3

10. 3.5 Carattere tonale *

-3 = Per niente **fischiante**

3 = Estremamente **fischiante**

Es. “Quel suono era così **fischiante** che mi ha sconvolto”

Contrassegna solo un ovale.

- -3
- -2
- -1
- 0
- 1
- 2
- 3

Il questionario è concluso, ti ringraziamo per il tuo contributo e per la partecipazione a questo test.

Questi contenuti non sono creati né avallati da Google.

Google Moduli

Appendice B

Recording Correspondence and Objective Variability

B.1 HDF–XL2 Correspondence

Table B.1 lists all available physical events. The mapping follows the documented chronological rule and the explicit identification of Mini 5 second take-off as XL2 3 and Matrice 4E second cruise as XL2 21. “Evaluated” identifies the first event supplying the objective descriptors for the subjective comparisons. Matching physical events does not imply matching signal duration or receiver position.

Tabella B.1: Complete mapping of all 28 physical events. The selected event in each condition supplies the objective predictors for subjective analyses.

Event	HDF basename	XL2	Evaluated
D1_LA_R1	`MatriceAgras_atterraggio`	41	Yes
D1_TO_R1	`MatriceAgras_decollo`	40	Yes
D1_HC_R1	`MatriceAgras_voloorizzontale`	42	Yes
D1_HC_R2	`MatriceAgras_voloorizzontale1`	43	No
D2_LA_R1	`Matrice350_atterraggio`	28	Yes
D2_LA_R2	`Matrice350_atterraggio1`	29	No
D2_LA_R3	`Matrice350_atterraggio2`	30	No
D2_TO_R1	`Matrice350_decollo`	24	Yes
D2_TO_R2	`Matrice350_decollo1`	25	No
D2_TO_R3	`Matrice350_decollo2`	26	No
D2_HC_R1	`Matrice350_voloorizzontale`	33	Yes
D2_HC_R2	`Matrice350_voloorizzontale1`	34	No
D4_LA_R1	`Matrice4E_atterraggio`	16	Yes
D4_LA_R2	`Matrice4E_atterraggio1`	17	No
D4_LA_R3	`Matrice4E_atterraggio2`	18	No

Continued on next page

Event	HDF basename	XL2	Evaluated
D4_TO_R1	`Matrice4E_decollo`	11	Yes
D4_TO_R2	`Matrice4E_decollo1`	12	No
D4_TO_R3	`Matrice4E_decollo2`	13	No
D4_HC_R1	`Matrice4E_voloorizzontale`	19	Yes
D4_HC_R2	`Matrice4E_voloorizzontale1`	21	No
D3_LA_R1	`MavicMini5_atterraggio`	5	Yes
D3_LA_R2	`MavicMini5_atterraggio1`	6	No
D3_LA_R3	`MavicMini5_atterraggio2`	7	No
D3_TO_R1	`MavicMini5_decollo1`	1	Yes
D3_TO_R2	`MavicMini5_decollo2`	3	No
D3_HC_R1	`MavicMini5_voloorizzontale`	8	Yes
D3_HC_R2	`MavicMini5_voloorizzontale1`	9	No
D3_HC_R3	`MavicMini5_voloorizzontale2`	10	No

B.2 Objective Event Summaries

TO denotes take-off, LA landing and HC horizontal cruise. The scalar definitions and units follow Table 4.2. A sample standard deviation is reported only with at least two independent events.

Tabella B.2: Independent-event descriptors by condition. SD is undefined for a single event and is shown as a dash.

UAV	Phase	Descriptor	n	Mean (SD)	Min–max
Agras 25	TO	ISO N5	1	34.050 (—)	34.050–34.050
		ISO N10	1	32.400 (—)	32.400–32.400
		Loudness HM	1	9.130 (—)	9.130–9.130
		Sharpness	1	2.760 (—)	2.760–2.760
		Roughness	1	0.800 (—)	0.800–0.800
		Fluctuation	1	0.200 (—)	0.200–0.200
		Tonality	1	0.087 (—)	0.087–0.087
Agras 25	LA	ISO N5	1	34.900 (—)	34.900–34.900
		ISO N10	1	33.900 (—)	33.900–33.900
		Loudness HM	1	9.930 (—)	9.930–9.930
		Sharpness	1	2.690 (—)	2.690–2.690
		Roughness	1	0.376 (—)	0.376–0.376
		Fluctuation	1	0.202 (—)	0.202–0.202
		Tonality	1	0.103 (—)	0.103–0.103
Agras 25	HC	ISO N5	2	33.025 (1.308)	32.100–33.950
		ISO N10	2	31.850 (1.485)	30.800–32.900
		Loudness HM	2	8.385 (0.417)	8.090–8.680
		Sharpness	2	2.643 (0.067)	2.595–2.690
		Roughness	2	0.577 (0.089)	0.514–0.640

Continued on next page

UAV	Phase	Descriptor	n	Mean (SD)	Min–max
		Fluctuation	2	0.189 (0.003)	0.187–0.191
		Tonality	2	0.054 (0.004)	0.051–0.056
Matrice 350	TO	ISO N5	3	17.800 (0.600)	17.200–18.400
		ISO N10	3	16.833 (0.473)	16.300–17.200
		Loudness HM	3	6.407 (0.114)	6.280–6.500
		Sharpness	3	2.428 (0.033)	2.400–2.465
		Roughness	3	0.583 (0.123)	0.474–0.716
		Fluctuation	3	0.148 (0.016)	0.136–0.166
		Tonality	3	0.116 (0.022)	0.095–0.139
Matrice 350	LA	ISO N5	3	25.033 (0.983)	23.900–25.650
		ISO N10	3	24.000 (0.926)	22.950–24.700
		Loudness HM	3	7.867 (0.172)	7.680–8.020
		Sharpness	3	2.528 (0.013)	2.515–2.540
		Roughness	3	0.230 (0.049)	0.184–0.282
		Fluctuation	3	0.149 (0.005)	0.144–0.152
		Tonality	3	0.132 (0.016)	0.118–0.149
Matrice 350	HC	ISO N5	2	21.850 (2.051)	20.400–23.300
		ISO N10	2	20.925 (1.874)	19.600–22.250
		Loudness HM	2	6.555 (0.078)	6.500–6.610
		Sharpness	2	2.215 (0.064)	2.170–2.260
		Roughness	2	0.173 (0.069)	0.125–0.222
		Fluctuation	2	0.154 (0.017)	0.142–0.166
		Tonality	2	0.148 (0.007)	0.143–0.152
Matrice 4E	TO	ISO N5	3	11.083 (0.076)	11.000–11.150
		ISO N10	3	10.500 (0.218)	10.250–10.650
		Loudness HM	3	4.613 (0.250)	4.370–4.870
		Sharpness	3	2.118 (0.028)	2.090–2.145
		Roughness	3	0.465 (0.139)	0.359–0.623
		Fluctuation	3	0.122 (0.030)	0.094–0.154
		Tonality	3	0.176 (0.038)	0.153–0.220
Matrice 4E	LA	ISO N5	3	12.583 (0.679)	11.950–13.300
		ISO N10	3	12.183 (0.681)	11.650–12.950
		Loudness HM	3	5.083 (0.233)	4.920–5.350
		Sharpness	3	2.150 (0.015)	2.135–2.165
		Roughness	3	0.211 (0.053)	0.150–0.244
		Fluctuation	3	0.112 (0.022)	0.090–0.134
		Tonality	3	0.146 (0.013)	0.137–0.161
Matrice 4E	HC	ISO N5	2	12.575 (0.318)	12.350–12.800
		ISO N10	2	12.050 (0.212)	11.900–12.200
		Loudness HM	2	4.485 (0.021)	4.470–4.500
		Sharpness	2	2.007 (0.081)	1.950–2.065
		Roughness	2	0.487 (0.177)	0.362–0.612
		Fluctuation	2	0.143 (0.048)	0.109–0.177
		Tonality	2	0.176 (0.016)	0.166–0.188
Mavic Mini 5	TO	ISO N5	2	7.152 (0.131)	7.060–7.245
		ISO N10	2	6.793 (0.194)	6.655–6.930

Continued on next page

UAV	Phase	Descriptor	n	Mean (SD)	Min–max
		Loudness HM	2	3.290 (0.014)	3.280–3.300
		Sharpness	2	2.093 (0.011)	2.085–2.100
		Roughness	2	0.130 (0.006)	0.126–0.134
		Fluctuation	2	0.076 (0.029)	0.055–0.096
		Tonality	2	0.225 (0.006)	0.221–0.229
Mavic Mini 5	LA	ISO N5	3	9.113 (0.617)	8.720–9.825
		ISO N10	3	8.733 (0.305)	8.550–9.085
		Loudness HM	3	4.110 (0.155)	3.950–4.260
		Sharpness	3	2.327 (0.046)	2.275–2.365
		Roughness	3	0.054 (0.016)	0.038–0.070
		Fluctuation	3	0.086 (0.017)	0.067–0.098
		Tonality	3	0.169 (0.068)	0.092–0.224
Mavic Mini 5	HC	ISO N5	3	7.625 (1.353)	6.495–9.125
		ISO N10	3	7.247 (1.267)	6.320–8.690
		Loudness HM	3	3.123 (0.364)	2.870–3.540
		Sharpness	3	1.772 (0.099)	1.685–1.880
		Roughness	3	0.099 (0.023)	0.078–0.123
		Fluctuation	3	0.047 (0.036)	0.021–0.089
		Tonality	3	0.165 (0.033)	0.132–0.199

B.3 Preliminary Ambisonics Branch

The nineteen-channel ZYLIA recordings were converted to sixteen-channel third-order Ambisonics with ZYLIA Ambisonics Converter. Wind contamination limited their suitability for the final listening tests. This processing branch was not used to generate the binaural stimuli and contributes no subjective results to the present analysis. The final perceptual experiment used the separately acquired BHS II/SQuadriga III recordings.

Appendice C

Subjective Descriptive Tables

The tables report all available responses separately by environment. The mean is followed by the sample standard deviation in parentheses. Median and quartiles are based on the same observations. These summaries are distinct from the medians of the restricted paired samples in Table 5.4.

Tabella C.1: Outdoor ratings: all available respondents. Mean (sample SD), median and interquartile range.

UAV	Phase	Outcome	n	Mean (SD)	Median	Q1–Q3
Agras 25	TO	Annoyance	12	6.42 (1.16)	6.5	6–7
		Loudness	12	2.25 (0.75)	2	2–3
		Fluctuation	12	1.58 (0.90)	1.5	1–2
		Roughness	12	1.42 (1.00)	1	1–2
		Sharpness	12	0.25 (1.42)	1	-1–1
		Tonality	12	0.92 (1.31)	0.5	0–2
Agras 25	LA	Annoyance	9	7.11 (1.76)	7	6–7
		Loudness	9	2.22 (0.67)	2	2–3
		Fluctuation	9	1.33 (1.12)	1	1–2
		Roughness	9	1.78 (0.97)	2	1–2
		Sharpness	9	0.67 (1.00)	1	0–1
		Tonality	9	0.89 (1.45)	0	0–2
Agras 25	HC	Annoyance	11	7.00 (1.10)	7	6.5–7.5
		Loudness	11	1.91 (0.83)	2	1–2.5
		Fluctuation	11	1.64 (0.81)	1	1–2
		Roughness	11	1.45 (1.21)	2	1–2
		Sharpness	11	0.09 (0.94)	0	0–1
		Tonality	11	0.82 (1.25)	0	0–2

Continued on next page

UAV	Phase	Outcome	n	Mean (SD)	Median	Q1–Q3
Matrice 350	TO	Annoyance	14	4.93 (2.20)	4	3.25–7
		Loudness	14	1.36 (1.01)	1.5	1–2
		Fluctuation	14	0.93 (1.44)	1.5	0–2
		Roughness	14	1.29 (0.91)	1.5	1–2
		Sharpness	14	-0.36 (1.50)	-0.5	-1–0.75
		Tonality	14	-0.07 (1.59)	0	-1–1
Matrice 350	LA	Annoyance	14	5.93 (2.02)	5.5	5–7.75
		Loudness	14	1.86 (0.77)	2	1–2
		Fluctuation	14	1.14 (1.10)	1	0.25–2
		Roughness	14	1.36 (1.22)	2	1–2
		Sharpness	14	-0.29 (1.33)	0	-1–0.75
		Tonality	14	0.36 (1.60)	0.5	0–1
Matrice 350	HC	Annoyance	12	5.92 (1.56)	6	5.5–7
		Loudness	12	1.75 (0.75)	2	1.75–2
		Fluctuation	12	1.00 (1.35)	1	1–2
		Roughness	12	1.42 (0.51)	1	1–2
		Sharpness	12	0.75 (1.22)	1	0–1.25
		Tonality	12	0.83 (1.27)	1	0–2
Matrice 4E	TO	Annoyance	14	3.29 (2.52)	3	1–5
		Loudness	14	-0.07 (1.14)	0	-1–1
		Fluctuation	14	0.29 (1.14)	1	-0.75–1
		Roughness	14	0.50 (1.09)	1	0.25–1
		Sharpness	14	-0.50 (1.29)	0	-1.75–0
		Tonality	14	-0.14 (1.56)	0	-0.75–1
Matrice 4E	LA	Annoyance	14	4.14 (2.18)	5	2.25–5
		Loudness	14	0.64 (1.39)	1	0.25–1.75
		Fluctuation	14	0.57 (1.55)	1	0.25–1.75
		Roughness	14	0.71 (1.33)	1	0–2
		Sharpness	14	0.00 (1.30)	0.5	-0.75–1
		Tonality	14	0.00 (1.52)	0	-0.75–1
Matrice 4E	HC	Annoyance	14	3.64 (2.34)	3.5	2–5.5
		Loudness	14	0.29 (1.27)	0.5	0–1
		Fluctuation	14	0.21 (1.72)	1	-1.5–1
		Roughness	14	0.57 (1.09)	1	0–1
		Sharpness	14	-0.07 (1.38)	0	-0.75–0.75
		Tonality	14	-0.21 (1.72)	0	-2–1
Mavic Mini 5	TO	Annoyance	14	2.71 (2.49)	2	0.5–4.5
		Loudness	14	-0.57 (1.50)	0	-1.75–0
		Fluctuation	14	-0.14 (1.56)	0	-1–1
		Roughness	14	0.00 (1.80)	0.5	-1.5–1
		Sharpness	14	-0.64 (1.74)	-1	-2–1
		Tonality	14	-0.07 (2.09)	1	-2.5–1.75
Mavic Mini 5	LA	Annoyance	14	3.64 (2.59)	3.5	1.25–5
		Loudness	14	-0.36 (1.60)	0	-1.75–1
		Fluctuation	14	0.71 (1.54)	1	0.25–1.75
		Roughness	14	0.43 (1.70)	1	0–2

Continued on next page

UAV	Phase	Outcome	n	Mean (SD)	Median	Q1–Q3
Mavic Mini 5	HC	Sharpness	14	-0.14 (1.66)	0.5	-1.5–1
		Tonality	14	0.00 (1.66)	0.5	-0.75–1
		Annoyance	14	2.64 (2.50)	2	1–3.75
		Loudness	14	-1.07 (1.44)	-1	-2—0.25
		Fluctuation	14	-0.36 (1.78)	0	-2–1
		Roughness	14	-0.29 (1.94)	-0.5	-1.75–1.5
		Sharpness	14	-0.50 (1.56)	0	-1.75–0.75
		Tonality	14	-0.29 (1.90)	0	-2–1

Tabella C.2: Indoor ratings: all available respondents. Mean (sample SD), median and interquartile range.

UAV	Phase	Outcome	n	Mean (SD)	Median	Q1–Q3
Agras 25	TO	Annoyance	13	7.15 (1.52)	7	6–8
		Loudness	13	2.15 (0.69)	2	2–3
		Fluctuation	13	1.69 (0.95)	2	1–2
		Roughness	13	1.92 (0.76)	2	1–2
		Sharpness	13	-0.62 (1.39)	0	-2–0
		Tonality	13	0.46 (1.61)	1	0–1
Agras 25	LA	Annoyance	13	7.69 (1.70)	8	6–9
		Loudness	13	2.54 (0.66)	3	2–3
		Fluctuation	13	1.62 (0.65)	2	1–2
		Roughness	13	2.38 (0.77)	3	2–3
		Sharpness	13	-0.46 (1.20)	0	-2–0
		Tonality	13	0.54 (1.27)	1	0–1
Agras 25	HC	Annoyance	13	6.77 (1.36)	7	6–8
		Loudness	13	1.77 (0.60)	2	1–2
		Fluctuation	13	1.77 (0.93)	2	1–2
		Roughness	13	1.62 (0.96)	2	1–2
		Sharpness	13	-0.38 (1.56)	0	-1–1
		Tonality	13	0.92 (1.38)	1	1–2
Matrice 350	TO	Annoyance	13	5.15 (2.27)	5	4–7
		Loudness	13	0.77 (0.93)	1	0–1
		Fluctuation	13	0.77 (1.36)	1	0–2
		Roughness	13	1.38 (0.77)	1	1–2
		Sharpness	13	-0.31 (1.11)	0	-1–1
		Tonality	13	0.08 (1.61)	0	-1–1
Matrice 350	LA	Annoyance	13	7.00 (1.83)	7	6–8
		Loudness	13	1.77 (0.73)	2	1–2
		Fluctuation	13	1.23 (1.36)	1	1–2
		Roughness	13	1.54 (1.05)	1	1–2
		Sharpness	13	-0.38 (1.04)	0	-1–0
		Tonality	13	0.23 (1.30)	0	0–1
Matrice 350	HC	Annoyance	13	5.31 (2.59)	6	4–7

Continued on next page

UAV	Phase	Outcome	n	Mean (SD)	Median	Q1–Q3
		Loudness	13	1.00 (1.08)	1	1–2
		Fluctuation	13	1.15 (1.34)	1	0–2
		Roughness	13	1.15 (0.90)	1	1–2
		Sharpness	13	-0.08 (1.19)	0	-1–1
		Tonality	13	0.38 (1.50)	1	0–1
Matrice 4E	TO	Annoyance	13	4.08 (2.60)	4	3–6
		Loudness	13	-0.15 (1.52)	0	-1–1
		Fluctuation	13	0.15 (1.57)	0	-1–1
		Roughness	13	0.38 (1.45)	0	0–1
		Sharpness	13	0.00 (1.22)	1	-1–1
		Tonality	13	-0.31 (1.70)	0	-2–1
Matrice 4E	LA	Annoyance	13	4.54 (2.37)	5	3–6
		Loudness	13	-0.23 (1.69)	0	-1–1
		Fluctuation	13	0.92 (1.19)	1	1–1
		Roughness	13	0.54 (1.51)	0	0–1
		Sharpness	13	0.08 (1.12)	0	0–1
		Tonality	13	0.00 (1.47)	0	-1–1
Matrice 4E	HC	Annoyance	13	4.31 (2.25)	5	2–6
		Loudness	13	0.08 (0.76)	0	0–1
		Fluctuation	13	0.62 (1.39)	0	0–2
		Roughness	13	0.85 (1.28)	1	0–2
		Sharpness	13	0.31 (1.11)	1	0–1
		Tonality	13	0.69 (1.25)	1	0–2
Mavic Mini 5	TO	Annoyance	13	3.38 (2.63)	4	1–5
		Loudness	13	-1.23 (1.54)	-1	-3—1
		Fluctuation	13	0.08 (1.71)	0	-1–1
		Roughness	13	0.46 (1.39)	0	0–1
		Sharpness	13	0.92 (0.86)	1	1–1
		Tonality	13	1.00 (1.47)	1	1–2
Mavic Mini 5	LA	Annoyance	13	4.00 (3.16)	4	1–7
		Loudness	13	-1.00 (1.35)	-1	-2–0
		Fluctuation	13	1.15 (1.28)	1	1–2
		Roughness	13	-0.08 (1.61)	-1	-1–1
		Sharpness	13	0.85 (0.99)	1	1–1
		Tonality	13	1.00 (1.47)	2	0–2
Mavic Mini 5	HC	Annoyance	13	3.31 (2.14)	3	2–6
		Loudness	13	-1.15 (1.28)	-1	-2–0
		Fluctuation	13	0.31 (1.32)	1	0–1
		Roughness	13	-0.15 (1.52)	0	-1–1
		Sharpness	13	0.54 (1.20)	1	0–1
		Tonality	13	0.85 (1.46)	1	0–2

Appendice D

Supplementary Correlation Results

The extended objective matrix includes the 28 assigned HDF–XL2 event pairs. Two additional matrices relate the three XL2 descriptors to mean ratings for the twelve evaluated events, separately by environment. As in the main results, cell probabilities are nominal, asymptotic and unadjusted. Field-level descriptors are not headphone-ear measurements.

D.1 Individual-Response Associations

The participant-centred coefficients in Table D.1 are supplementary linear descriptions. Their bootstrap intervals retain participant and stimulus resampling, but no ordinary independent-observation probability is reported. They should not be read as Spearman results or as a second confirmatory testing family.

Tabella D.1: Supplementary participant-centred linear associations with objective HM loudness. Intervals are crossed-bootstrap percentile intervals (2,000 resamples); they are exploratory, not ordinary independent-observation confidence intervals. No p value was estimated for this analysis.

Environment	Response	Ratings	People	r_{within}	95% interval
Outdoor	Annoyance	156	14	0.665	0.429 to 0.842
Outdoor	Loudness rating	156	14	0.746	0.593 to 0.852
Indoor	Annoyance	156	13	0.744	0.493 to 0.884
Indoor	Loudness rating	156	13	0.834	0.689 to 0.915

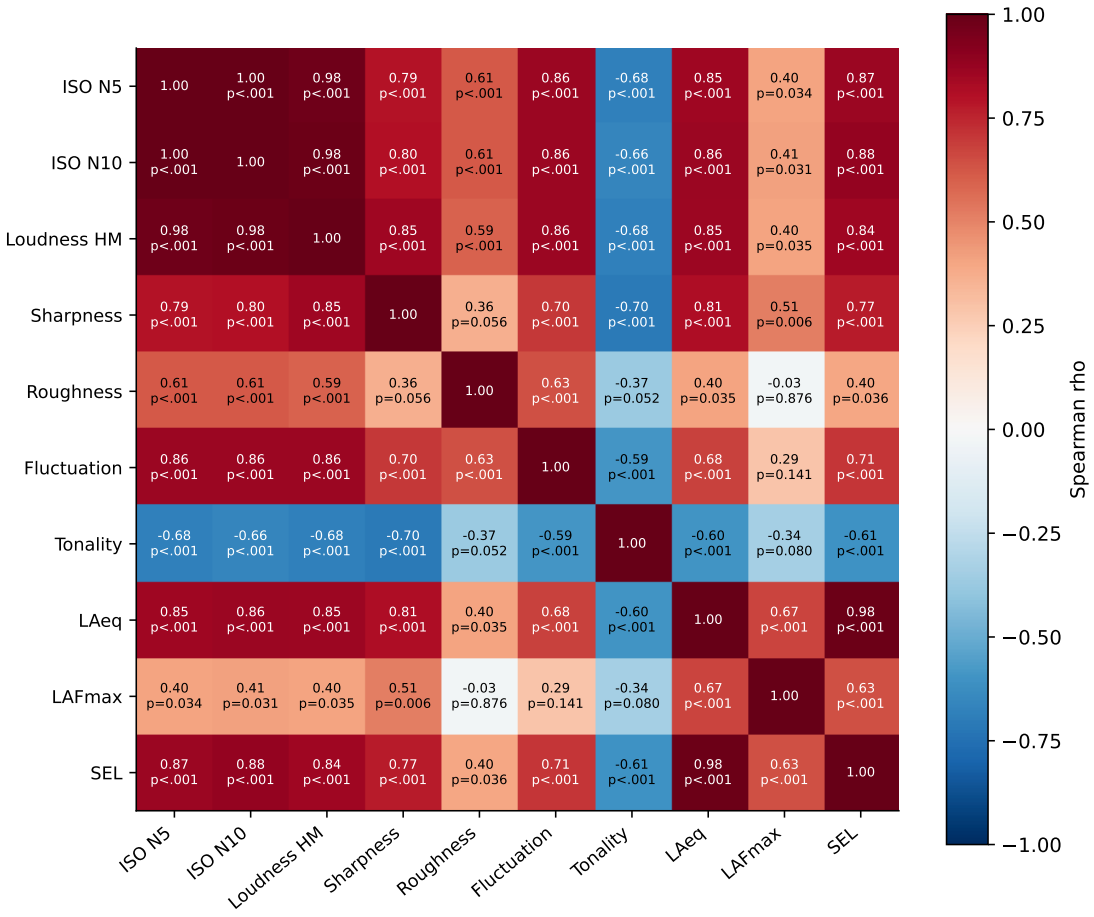


Figura D.1: Extended Spearman matrix of objective HDF and XL2 descriptors across 28 physical events. Each off-diagonal cell displays the coefficient and nominal p-value.

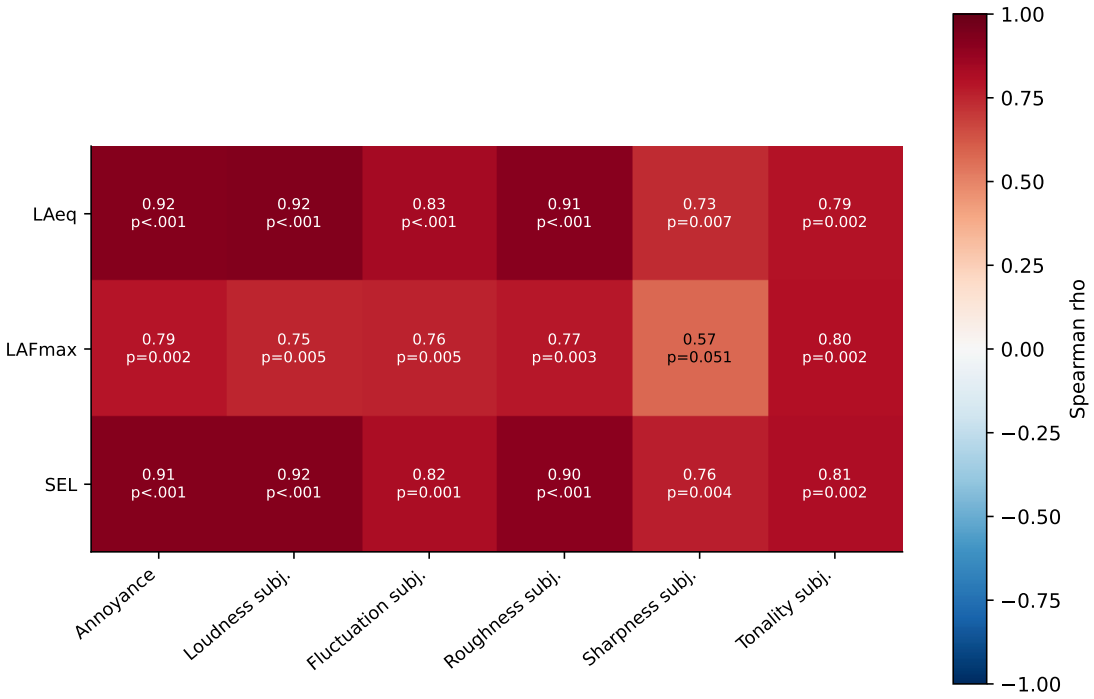


Figura D.2: Outdoor XL2–subjective matrix of twelve condition means. Cells give Spearman rho and the nominal two-sided p-value.

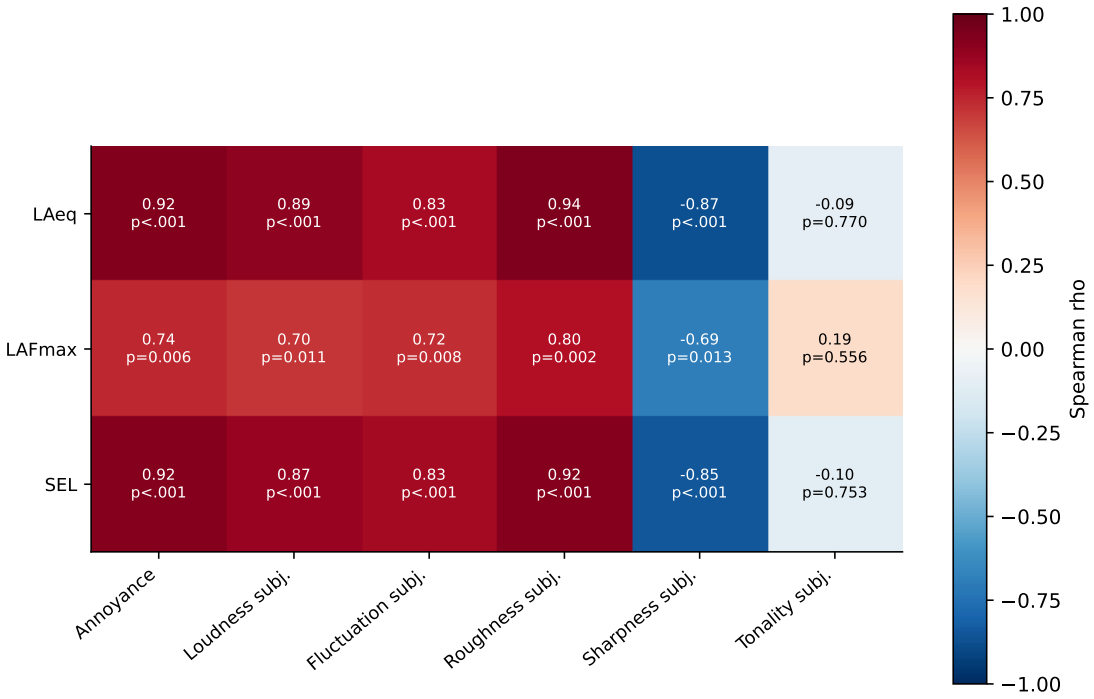


Figura D.3: Indoor XL2–subjective matrix of twelve condition means. Cells give Spearman rho and the nominal two-sided p-value.

D.2 Complete Numerical Archive

The accompanying numerical supplement contains all calculated Pearson and Spearman pairs, the paired-mean and twelve-selected-event sensitivity results, and the complete individual-response coefficient table. Every row identifies its dataset block, coefficient type, sample size and probability or interval convention. This archive preserves weak and non-significant results as well as the stronger associations illustrated in the thesis.

Appendice E

Complete Simple Regression Results

Each environment contributes seventeen OLS models fitted to twelve condition means. The a and b columns below are the unstandardised intercept and single-predictor slope. Two-predictor slopes and nominal intervals are listed separately. The full coefficient and diagnostic files, including standard errors and fitted values, accompany the thesis sources.

Tabella E.1: Outdoor: complete set of 17 exploratory OLS fits, each with $n = 12$. For two-predictor models, individual slopes are given in Table E.3.

Predictor(s)	Outcome	a	b	R^2	R^2_{adj}	p_{model}
Loudness HM	Annoyance	0.633	0.689	0.930	0.923	< 0.001
ISO N5	Annoyance	1.922	0.152	0.953	0.949	< 0.001
ISO N10	Annoyance	1.940	0.158	0.954	0.950	< 0.001
SEL	Annoyance	-10.668	0.207	0.914	0.906	< 0.001
Roughness	Annoyance	3.662	3.075	0.245	0.169	0.102
Fluctuation	Annoyance	1.427	24.854	0.733	0.706	< 0.001
Sharpness	Annoyance	-5.767	4.599	0.746	0.721	< 0.001
Tonality	Annoyance	8.074	-24.616	0.508	0.458	0.009
Loudness HM	Loudness subj.	-2.119	0.493	0.918	0.910	< 0.001
Sharpness	Sharpness subj.	-1.625	0.681	0.223	0.146	0.121
Roughness	Roughness subj.	0.401	1.334	0.291	0.220	0.070
Fluctuation	Fluctuation subj.	-0.638	10.229	0.798	0.777	< 0.001
Tonality	Tonality subj.	1.045	-5.926	0.348	0.282	0.044
Loudness rating	Annoyance	3.642	1.339	0.931	0.924	< 0.001
Roughness rating	Annoyance	2.651	2.403	0.914	0.905	< 0.001
HM + roughness	Annoyance	0.629	–	0.931	0.916	< 0.001
Rated loud. + rough.	Annoyance	3.362	–	0.933	0.919	< 0.001

Tabella E.2: Indoor: complete set of 17 exploratory OLS fits, each with $n = 12$. For two-predictor models, individual slopes are given in Table E.3.

Predictor(s)	Outcome	a	b	R^2	R^2_{adj}	p_{model}
Loudness HM	Annoyance	1.219	0.665	0.971	0.968	< 0.001
ISO N5	Annoyance	2.550	0.142	0.935	0.928	< 0.001
ISO N10	Annoyance	2.564	0.148	0.937	0.931	< 0.001
SEL	Annoyance	-9.302	0.195	0.905	0.896	< 0.001
Roughness	Annoyance	4.149	2.955	0.253	0.178	0.096
Fluctuation	Annoyance	2.068	23.391	0.727	0.700	< 0.001
Sharpness	Annoyance	-5.241	4.563	0.822	0.805	< 0.001
Tonality	Annoyance	8.162	-21.962	0.453	0.398	0.017
Loudness HM	Loudness subj.	-2.898	0.569	0.963	0.959	< 0.001
Sharpness	Sharpness subj.	2.926	-1.259	0.568	0.524	0.005
Roughness	Roughness subj.	0.302	1.918	0.404	0.344	0.026
Fluctuation	Fluctuation subj.	-0.066	7.570	0.535	0.488	0.007
Tonality	Tonality subj.	0.454	0.250	0.001	-0.099	0.932
Loudness rating	Annoyance	4.625	1.140	0.957	0.953	< 0.001
Roughness rating	Annoyance	3.401	1.823	0.877	0.865	< 0.001
HM + roughness	Annoyance	1.215	–	0.972	0.966	< 0.001
Rated loud. + rough.	Annoyance	4.701	–	0.957	0.948	< 0.001

Tabella E.3: Coefficients for the two-predictor models. Intervals are nominal 95% OLS intervals; const denotes the intercept.

Environment	Model	Term	B	SE	p	95% CI
Outdoor	Objective	const	0.629	0.401	0.151	-0.278 to 1.536
Outdoor	Objective	HM	0.707	0.075	< 0.001	0.538 to 0.876
Outdoor	Objective	Roughness	-0.281	0.649	0.675	-1.750 to 1.188
Outdoor	Subjective	const	3.362	0.497	< 0.001	2.237 to 4.486
Outdoor	Subjective	Rated loud.	0.983	0.606	0.140	-0.389 to 2.354
Outdoor	Subjective	Rated rough.	0.659	1.097	0.563	-1.824 to 3.142
Indoor	Objective	const	1.215	0.240	< 0.001	0.672 to 1.759
Indoor	Objective	HM	0.684	0.045	< 0.001	0.582 to 0.785
Indoor	Objective	Roughness	-0.291	0.389	0.473	-1.171 to 0.589
Indoor	Subjective	const	4.701	0.358	< 0.001	3.891 to 5.511
Indoor	Subjective	Rated loud.	1.203	0.292	0.003	0.541 to 1.864
Indoor	Subjective	Rated rough.	-0.109	0.488	0.829	-1.213 to 0.996

E.1 Diagnostic Qualification

The diagnostic archive includes residual Shapiro–Wilk and Breusch–Pagan probabilities, maximum Cook's distance and maximum absolute studentised residual for

each fit. In particular, several indoor annoyance fits show residual-normality departures, and influential observations are possible in a set of only twelve conditions. No observation was deleted to improve the reported fit. These diagnostics qualify ordinary OLS probabilities and do not establish performance on new data.

Bibliografia

[1] Beat Schäffer, Reto Pieren, Kurt Heutschi, Jean Marc Wunderli e Stefan Becker. «Drone Noise Emission Characteristics and Noise Effects on Humans—A Systematic Review». In: *International Journal of Environmental Research and Public Health* 18.11 (giu. 2021), p. 5940. ISSN: 1660-4601. DOI: 10.3390/ijerph18115940. (Visitato il 01/04/2026) (cit. alle pp. 1, 4–13, 17, 19, 23, 79).

[2] Antonio J. Torija e Charlotte Clark. «A Psychoacoustic Approach to Building Knowledge about Human Response to Noise of Unmanned Aerial Vehicles». In: *International Journal of Environmental Research and Public Health* 18.2 (gen. 2021), p. 682. ISSN: 1660-4601. DOI: 10.3390/ijerph18020682. (Visitato il 24/07/2026) (cit. alle pp. 1, 4, 17).

[3] Antonio J. Torija, Zhengguang Li e Rod H. Self. «Effects of a Hovering Unmanned Aerial Vehicle on Urban Soundscapes Perception». In: *Transportation Research Part D: Transport and Environment* 78 (gen. 2020), p. 102195. ISSN: 13619209. DOI: 10.1016/j.trd.2019.11.024. (Visitato il 27/03/2026) (cit. alle pp. 5, 17, 21).

[4] C. T. Justine Hui, Michael J. Kingan, Yusuke Hioka, Gian Schmid, George Dodd, Kim N. Dirks, Shaun Edlin, Sean Mascarenhas e Young-Min Shim. «Quantification of the Psychoacoustic Effect of Noise from Small Unmanned Aerial Vehicles». In: *International Journal of Environmental Research and Public Health* 18.17 (ago. 2021), p. 8893. ISSN: 1660-4601. DOI: 10.3390/ijerph18178893. (Visitato il 27/03/2026) (cit. alle pp. 6, 7, 14, 18, 19, 21, 24).

[5] Mansour Alkmim, João Cardenuto, Elisa Tengan, Thomas Dietzen, Toon Van Waterschoot, Jacques Cuenca, Laurent De Ryck e Wim Desmet. «Drone Noise Directivity and Psychoacoustic Evaluation Using a Hemispherical Microphone Array». In: *The Journal of the Acoustical Society of America* 152.5 (nov. 2022), pp. 2735–2745. ISSN: 0001-4966, 1520-8524. DOI: 10.1121/10.0014957. (Visitato il 27/03/2026) (cit. alle pp. 6–8, 10, 12–15).

[6] Antonio J. Torija e Rory K. Nicholls. «Investigation of Metrics for Assessing Human Response to Drone Noise». In: *International Journal of Environmental Research and Public Health* 19.6 (mar. 2022), p. 3152. ISSN: 1660-4601. DOI: 10.3390/ijerph19063152. (Visitato il 27/03/2026) (cit. alle pp. 7, 11–16, 18).

[7] Carlos Ramos-Romero, Nathan Green, Antonio J. Torija e César Asensio. «On-Field Noise Measurements and Acoustic Characterisation of Multi-Rotor Small Unmanned Aerial Systems». In: *Aerospace Science and Technology* 141 (ott. 2023), p. 108537. ISSN: 12709638. DOI: 10.1016/j.ast.2023.108537. (Visitato il 24/07/2026) (cit. alle pp. 8–10, 23).

[8] R. Zamponi, G. Gioli Torrione, V. Zůbek, E. Gallo, A. Zarri e C. Schram. «Maneuvering Drone Noise Investigation in a Reverberant Environment». In: *The Journal of the Acoustical Society of America* 157.6 (giu. 2025), pp. 4468–4481. ISSN: 1520-8524. DOI: 10.1121/10.0036900. (Visitato il 27/03/2026) (cit. alle pp. 8, 23).

[9] Ronja König, Laura Babetto, André Gerlach, Janina Fels e Eike Stumpf. «Prediction of Perceived Annoyance Caused by an Electric Drone Noise through Its Technical, Operational, and Psychoacoustic Parameters». In: *The Journal of the Acoustical Society of America* 156.3 (set. 2024), pp. 1929–1941. ISSN: 0001-4966. DOI: 10.1121/10.0028514. (Visitato il 27/03/2026) (cit. alle pp. 13–15, 17, 22, 79).

[10] Andrea Vaiuso, Marcello Righi, Oier Coretti e Moreno Apicella. *Drone Acoustic Analysis for Predicting Psychoacoustic Annoyance via Artificial Neural Networks*. Ott. 2024. DOI: 10.48550/arXiv.2410.22208. arXiv: 2410.22208 [cs]. (Visitato il 27/03/2026) (cit. alle pp. 14, 16, 22).

[11] Jun Zhang, Kean Chen, Hao Li, Xingshu Chen e Ningjuan Dong. «The Effects of Rating Scales and Individual Characteristics on Perceived Annoyance in Laboratory Listening Tests». In: *Applied Acoustics* 202 (gen. 2023), p. 109137. ISSN: 0003682X. DOI: 10.1016/j.apacoust.2022.109137. (Visitato il 18/06/2026) (cit. alle pp. 17, 18).

[12] Claudia Kawai, Jonas Jäggi, Fotis Georgiou, Jonas Meister, Reto Pieren e Beat Schäffer. «Short-Term Noise Annoyance towards Drones and Other Transportation Noise Sources: A Laboratory Study». In: *The Journal of the Acoustical Society of America* 156.4 (ott. 2024), pp. 2578–2595. ISSN: 0001-4966. DOI: 10.1121/10.0032386. (Visitato il 27/03/2026) (cit. a p. 18).

[13] J.M. Fields et al. «STANDARDIZED GENERAL-PURPOSE NOISE REACTION QUESTIONS FOR COMMUNITY NOISE SURVEYS: RESEARCH AND A RECOMMENDATION». In: *Journal of Sound and Vibration* 242.4

(mag. 2001), pp. 641–679. ISSN: 0022460X. DOI: 10.1006/jsvi.2000.3384. (Visitato il 14/05/2026) (cit. a p. 18).

[14] W. Owen Brimijoin, Alan W. Boyd e Michael A. Akeroyd. «The Contribution of Head Movement to the Externalization and Internalization of Sounds». In: *PLOS ONE* 8.12 (2013), e83068. DOI: 10.1371/journal.pone.0083068 (cit. alle pp. 19, 21, 79).

[15] Frederic L. Wightman e Doris J. Kistler. «Headphone simulation of free-field listening. I: Stimulus synthesis». In: *The Journal of the Acoustical Society of America* 85.2 (1989), pp. 858–867. DOI: 10.1121/1.397557 (cit. alle pp. 20, 79).

[16] Michael J. B. Lotinga, Marc C. Green e Antonio J. Torija. «Development of Psychoacoustic Prediction Models for Short-Term Noise Annoyance Responses to Unmanned Aircraft Systems». In: *The Journal of the Acoustical Society of America* 158.3 (set. 2025), pp. 2062–2082. ISSN: 1520-8524. DOI: 10.1121/10.0039056. (Visitato il 12/05/2026) (cit. alle pp. 22, 24).

[17] HEAD acoustics GmbH. *BHS II: Binaural Headset for Aurally Accurate Recording and Playback*. Code 3322. Data sheet, last update November 2023. HEAD acoustics GmbH. 2023 (cit. a p. 30).

[18] HEAD acoustics GmbH. *SQuadriga III: Mobile Recording and Playback System with Sound Level Meter Function*. Code 3324, revision 07. Data sheet, last update April 2026. HEAD acoustics GmbH. 2026 (cit. a p. 30).

[19] HEAD acoustics GmbH. *BPU Bundle: Binaural USB Playback Unit and Headphones*. Data sheet, Japanese edition, August 2023. HEAD acoustics GmbH. 2023. URL: https://cdn.head-acoustics.com/fileadmin/data/global/Datasheets/Archive/BPU-Playback-Unit-244X-Data-Sheet-jp.pdf (visitato il 18/09/2026) (cit. a p. 31).

[20] R Core Team. *Wilcoxon Rank Sum and Signed Rank Tests: wilcox.test*. 2026. URL: https://stat.ethz.ch/R-manual/R-devel/library/stats/html/wilcox.test.html (visitato il 20/09/2026) (cit. a p. 41).

[21] SciPy developers. *scipy.stats.spearmanr: API Reference*. 2026. URL: https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.spearmanr.html (visitato il 20/09/2026) (cit. a p. 43).