Perception of Noise from Unmanned Aircraft Systems: Efficacy of Metrics for Indoor and Outdoor Listener Positions

Nathan Green,[1] Antonio J. Torija[1], Carlos Ramos-Romero[1]

[1] *Acoustics Research Centre, University of Salford, Manchester, M5 4WT, UK*

Abstract:

This paper presents the results of a listening experiment designed to assess annoyance and perceived loudness (*PL*) for several Unmanned Aircraft System (UAS) operations, with the listener simulated in indoor and outdoor positions. This research investigated (i) how participant responses change depending on UAS operation, (ii) which broadband metrics are most suitable for representing annoyance and *PL,* (iii) differences in noise level required to result in equal participant responses to different operations and (iv) which Sound Quality Metrics (SQMs) are significant for UAS noise perception. Results indicate annoyance and *PL* responses were greatest for landing operations with flyovers being least annoying or loud. L_{Aeq}, L_{ASmax} and Loudness (N_5) were the strongest predictors in representing Annoyance. Offset analysis predicted small differences in Annoyance responses between flyovers and other operations, but also indicated that flyovers would require an increase to L_{ASmax} of 3.3 to 6.3 dB compared to other operations to achieve equal *PL*. Loudness was the most significant SQM, with minor contributions from impulsivity for annoyance and *PL* when outside, and tonality for *PL* when indoors. These findings contribute to the understanding of UAS noise perception for the development of metrics and assessment methods accounting for the characteristics of UAS operations.

Keywords: *UAS Noise*, *Noise Annoyance, Perceived Loudness, Listening Experiment*, *Sound Quality Metrics*

I. INTRODUCTION

Over recent years, the interest in using Unmanned Aircraft Systems (UASs) for commercial operations has increased dramatically as numerous public and private sectors seek to leverage the potential benefits of autonomous vehicles. As a result, the requirement to better understand auditory perception and exposure-response to UAS noise and the potential associated impacts on exposed communities (e.g., noise annoyance) has also become of greater interest.

Significant strides have been made developing an understanding of the complex sound signature generated by UAS attributed to phenomena such as 'rotor / rotor interaction' (Torija *et al.*, 2021) and the unsteady tonal components of the sound present within outdoor flight (Cabell *et al.*, 2016; Alexander and Whelchel, 2019; Torija *et al.*, 2019). Torija & Clarke (2021), compared the sound spectra of two conventional aircraft (737 Max 8 and Airbus A320) and two multi-copter drones (DJI Matric 200 and Yuneec Typhoon)at a normalized broadband level of 65 dB(A). What was evident from the comparison was at equal overall sound level the two UAS exhibited significantly more noise above 2kHz than the conventional aircraft. Gwak et al. (2020b) made a similar observation, noting that one of the significant differences between the sound generated by conventional aircraft and UAS is the amount of high frequency energy within the sound signature. The same is also true when compared against noise from road traffic which peaks around the 1 kHz third octave band and quickly reduces above 1.6 kHz (Gjestland, 2008).

Research into the perception of UAS noise has also been advancing over recent years. For example, the Civil Aviation Authority (CAA) in the UK have recently produced an overview report (CAA, 2023) summarising the current knowledge on the effects of electric Vertical Take-Off and Landing (eVTOL) noise on humans. Schäffer et al. (2021) produced a systematic review of existing research on UAS noise including the effects on humans. Most of the literature discussed within these reviews highlight the significance of sound pressure level and its impact on Annoyance (Christian and

Cabell, 2017; Callanan *et al.*, 2020; Gwak *et al.*, 2020b). Others have investigated the use of Sound Quality Metrics (SQMs, refer to Zwicker and Fastl (2013) for a detailed introduction) to better understand what elements of the UAS sound the listener is responding to. Torija & Nicholls (2022) identified Sharpness (measuring perceptual effects of high frequency noise) and Fluctuation Strength (measuring perceptual effect of slow amplitude modulation) as being significant factors influencing annoyance and Perceived Loudness (*PL*). Hui et al. (2021) found that the Sound Quality Metric (SQM) Loudness (N_5), A-weighted Equivalent Continuous Sound Level (L_{Aeq}), and A-weighted Maximum Sound Level (L_{Amax}) all had a strong correlation with the Annoyance associated with a hovering drone, with L_{Amax}, $N_{5,}$ Sound Exposure Level (L_{AE}) and Roughness (R_5) all demonstrating a strong correlation with Annoyance for drone flyover events.

Torija & Clarke (2021) discussed how the noise signature of a given UAS will be different depending on whether the vehicle is taking-off, hovering, landing or flying over. This is particularly observed in relation to the directivity and prominence of tonal noise over broadband noise (Alexander and Whelchel, 2019). Gallo et al. (2022) found the psychoacoustic annoyance, based on the value of the Sound Quality Metrics Loudness, Sharpness, Fluctuation Strength and Roughness (Zwicker and Fastl, 2013) of flyover and transition manoeuvres to be higher than for hovering flight. However, to date, there is not a detailed study investigating the noise perception of UAS under different operational manoeuvres.

Hitherto, little research has been conducted to better understand the changes in perception of UAS noise when the listener is outdoors vs. indoors or, to understand how different noise metrics perform for UAS noise signatures transmitted through building partitions. Work previously undertaken by Ramos-Romero et al. (2022) began to investigate how noise from UAS flyovers could be predicted within indoor environments. That work provided a framework for how UAS noise propagation and transmission to an indoor environment might be predicted with the aim of determining a minimum Drone-Façade distance required to avoid excessive internal noise levels. The framework was based upon existing assessment metrics (e.g., 42 dB $L_{\mathrm{Amax,indoors}}$) but did not investigate the performance of such metrics for UAS noise perception. This paper presents the results of a listening experiment designed to investigate the perception of UAS noise from different operational procedures, and for three simulated listener positions (i.e., outdoors, indoors with a partially open-window, and indoors with a closed window). The goal of the paper is to advance the state of the art in UAS noise perception by answering the following research questions:

1. How do annoyance and *PL* of UAS noise change as the listener is positioned outdoors or indoors?
2. Which broadband noise metrics correlate best with annoyance and *PL* for UAS noise?
3. Are there operating procedures that are perceived as particularly loud or annoying?
4. What is the contribution of metrics accounting for loudness, tonality, frequency content and temporal characteristics on UAS noise perception?

The structure of this paper is as follows: Section II presents a literature review and technical justification for the research, Section III presents details of the UAS included within the experiment, the design and methodology of the listening experiment and introduces the method of statistical analysis, Section IV the results of the experiment and Section V the conclusions.

II. Review of Key Aspects for UAS noise assessment

A. Indoor vs Outdoor Noise

Exposure to excessive indoor noise, particularly at home or at people's place of rest, is a well-documented problem as it can lead to behavioural changes, increased annoyance, reduced speech intelligibility and sleep disturbance (Berglund *et al.*, 1999; Hurtley, 2009). The potential adverse impacts of excessive noise can also extend to physiological issues, such as increased risk of hypertension and heart disease (Babisch, 2011; Basner *et al.*, 2014; Foraster *et al.*, 2017; Kempen *et al.*, 2018), and the deterioration of mental health (Stansfeld *et al.*, 2000; Hardoy *et al.*, 2005; Clark and Paunovic, 2018; Clark *et al.*, 2020).

When considering the acoustic transmission through a typical residential façade, the window or glazed element is usually the weakest point. The sound reduction of windows can vary significantly depending on the design: single glazed windows may offer a weighted sound reduction performance (R_w) of around 24 dB, whereas much more robust designs, such as triple-glazing windows may offer up to 45 dB R_w (Waters-Fuller *et al.*, 2007). Most residential properties in urban and suburban areas in the UK have double-glazed windows and typically offer an R_w of between 30 and 35 dB (Hurtley, 2009) but assume the windows are closed. However, many properties will rely upon openable windows to provide adequate ventilation and prevent overheating during the summer months. By extension, when considering the external to internal transmission of noise, it is important to consider the transmission through both partially open and closed windows.

A comprehensive laboratory study (Waters-Fuller *et al.*, 2007), considering window configurations and receiving room representative of typical sensitive residential rooms (e.g., living room) in the UK, reported the sound reduction performance of a typical double glazing configuration, either fully closed or with the window partially open.

It is also worth noting the significance of the distance between the noise source and the façade. The greater the distance between the source and receiver, the less noise will be observed because of spherical spreading and atmospheric absorption which will reduce the ratio of high frequency energy before it arrives at the receiver or façade. Whilst it was outside the scope of this experiment to investigate how changes of distance between the UAS noise source and receiver/façade effect perception, research into this area could be of significant value.

B. Metrics – Impact Assessment, Noise Certification and ΔL

1. Noise Impact Assessment – Single Events

When determining the magnitude of impact of environmental noise, it is still most common to have assessment criteria defined as broadband noise metrics. Whilst there are significant limitations to what can be expressed about sound within a broadband noise metric, they remain indispensable, as they are simple to measure, require relatively basic equipment to capture, are long established and simple to understand and compare. For single events, the sound of a vehicle is typically represented using the A-weighted Equivalent Continuous Sound Pressure Level (L_{Aeq}), A-weighted Sound Exposure Level (L_{AE}), the Maximum A-weighted Sound Level (L_{Amax}). The Effective Perceived Noise Level (EPNL) is the metric generally used for the noise certification of larger commercial aircraft, both propeller and jet driven (Filippone, 2014). EPNL is a combination of the Perceived Noise Level (*PNL*), which accounts for the combined 'noisiness' of a noise event across the frequency spectra plus corrections for tones and duration of the event.

2. UAS Noise Certification

Noise certification refers to the process undertaken to determine the noise level associated with a vehicle when operating under specific conditions. New guidance from the European Union Aviation Safety Agency (EASA) for UAS below 600kg (2022) has recommended that L_{AE} is the metric to be reported for flyover or cruise operations whereas L_{Aeq} is recommended for hover operations. No metrics were recommended within this document for take-off or landing operations. In September 2022, the Federal Aviation Authority (FAA) certified the Matternet model M2, a quadcopter with a Maximum Take-Off Weight (MTOW) of 11.5 kg designed for parcel delivery. The certification process for flyover noise was broadly based on the noise certification method for small helicopters defined within Part 36, Subpart H, Appendix J of the Code of Federal Regulations (Archives, 2023) which stipulates noise levels are to be presented as an L_{AE}. Recently published consultation paper from EASA (2023) which admittedly is for the certification of larger Urban Air Mobility (UAM) vehicles has recommended EPNL, typically used for conventional aircraft, as the metric to be used for take-off, flyovers and landing operations and L_{Aeq} for hover operations.

3. Sound Quality Metrics

SQMs, such as Loudness, are unlike conventional metrics that are used to describe the physical properties of the sound or noise event (maximum or average sound pressure for example). Instead, SQMs are tailored to describe the human response to sound or hearing sensation. Psychoacoustic annoyance can be described through a combination of SQMs describing the Loudness, tone colour and temporal structure of a sound (Zwicker and Fastl, 2013). Early Psychoacoustic Annoyance models we derived using a combination of the following SQMs: Loudness, Sharpness (describing tone colour), Fluctuation Strength and Roughness to describe the temporal structure. Other SQMs have subsequently been developed to describe other spectral or temporal characteristics such as Impulsiveness and Tonality.

Although these metrics are not used for the certification or assessment of environmental noise, they are highly valuable in predicting or understanding what characteristics of a sound listeners may be responding to when they rate the annoyance or *PL* of a sound.

4. dB Offset or ΔL

Examining the difference in noise levels (ΔL) required for two noise sources to result in equal annoyance is an important tool to help understand variations in response to road traffic, rail and conventional aircraft noise (Fields and Walker, 1982; Schreckenberg *et al.*, 1999). Recent studies have taken this principal and applied it to UAS noise and found it to be more annoying than other transportation vehicles, at the same sound level. A pioneering study by Christian and Cabell (2017) compared the annoyance of drone flyovers with road vehicle pass-bys. In their study Christian and

Cabell found drones to be equally annoying as road vehicles at 5.6 dB higher sound level; or in other words, road vehicles had to be 5.6 dB louder to be perceived as equally annoying as UAS. Similar findings have been found by other researchers. For instance, Torija and Li (2020) investigated the preference (i.e., an 'inverse indicator' of annoyance) of different transportation noise sources, finding a small quadcopter was 33% less preferred than a conventional civil aircraft taking-off (at the same sound level, 65 dBA); Gwak et al.(2020a; b) also found the annoyance of a hovering drone to be significantly higher than a take-off jet aircraft. Specifically, they found hovering drones equally annoying as a jet aircraft taking-off with a 4-10 dB higher sound level, depending on the size of the drone.

In these cases, this sound level difference, or offset, (ΔL) helps to understand the differences in exposure-response between vehicles. However, this research intended to use the same process for investigating the ΔL for different noise metrics (e.g., L_{Aeq} or L_{Amax}) between different UAS operations for equal exposure response. This process of deriving a dB offset value or delta (Δ) is illustrated in FIG. 1.

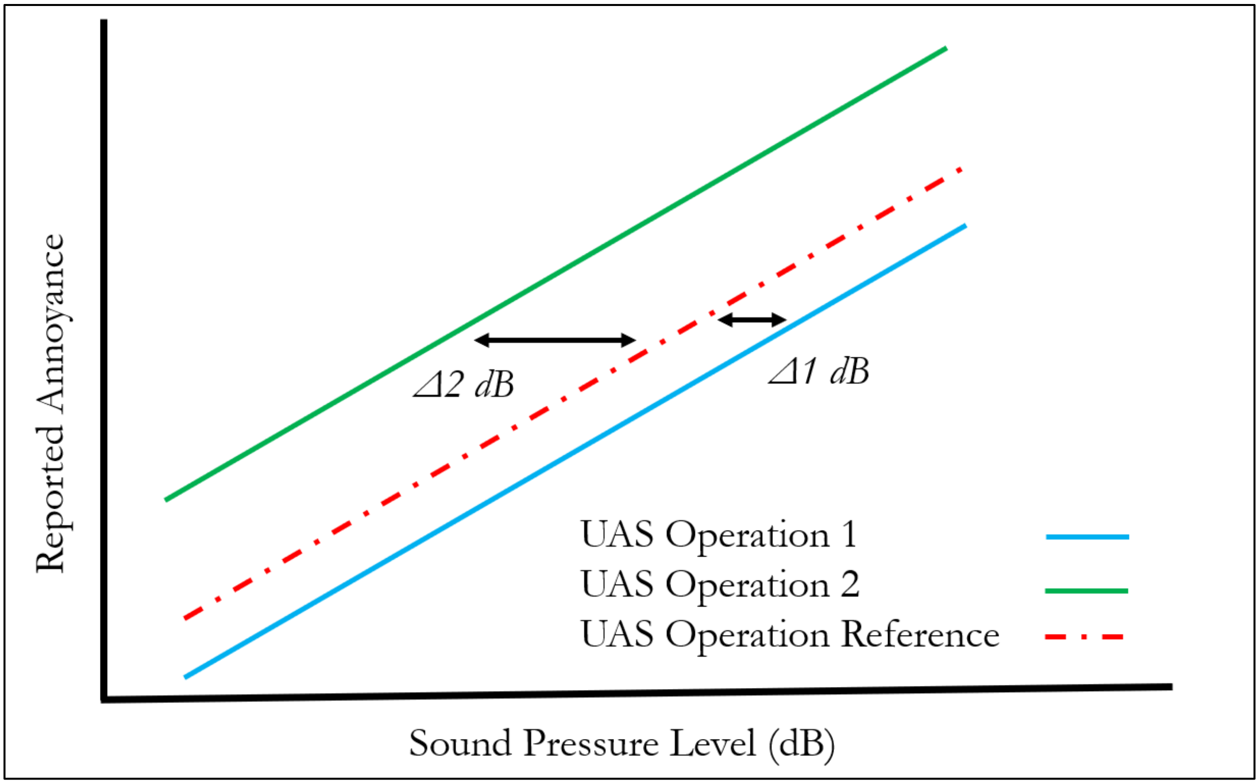


FIG. 1. (Colour Online) Illustration showing how the Δ dB value is derived from two trendlines summarising participant responses to two different sound sources

III. Methodology

A. Drone Noise Audio Database

The listening experiment used a database of audio files provided by the U.S Volpe National Transportation Systems Centre (Read *et al.*, 2020). The database includes recordings of three types of small multi-rotor UAS performing different flying operations, four of which have been included within this listening experiment (hovering, take-off, landing and flyover at 15 m/s,). Table I presents the design specifications of the multi-rotor UAS that were recorded and have subsequently been used with this listening experiment.

Table I. Specifications of the UAS used within the listening experiment

Multirotor Aircraft	Number	Drone Weight	MTOW* (Kg)	Largest

Models	of rotors	(Kg)		Dimension** (m)
Gryphon Dynamics GD28X	Four pairs – Contra-Rotating	11.8	31.7	2.1
DJI M200	4	4.0	6.1	0.9
Yuneec Typhoon	6	1.9	2.4	0.5

Full details of the measurement methods can be found in the following document (Read *et al.*, 2020). To summarise, the microphone was mounted on a tripod at 1.2 metres above ground. Flyover measurements were obtained with the microphone directly underneath the flightpath and the drone at an altitude of 150 feet above ground (~47.5 m). For the take-off measurements the drone flew to an altitude of 150 feet with a vertical ascent, then proceeded to move away from the measurement position, the landing measurements followed the same process, but in reverse. For the hover measurements, the drone hovered at an altitude of four feet (1.2 m) above the ground, held the position for 30 seconds and then rotated 90 degrees. For the take-off, landing and hover measurements the distance between the microphone and take-off/landing point was 30 feet (9.1m) from the microphone position.

Both audio and sound level data were recorded. The audio was recorded with a fidelity of 48kHz sample rate and 24-bit analogue to digital conversion. Sound level data was also recorded by feeding using a Larson Davis sound level meter, measurements were recorded at 1 second intervals with slow response A-weighted noise levels. These measured levels were used to calibrate the audio files and equipment used for the listening experiment. In total, 12 audio files were selected from the Volpe database, i.e., each UAS described in Table I performing a flyover, hover, landing and take-off operation.

FIG. 2 details the stages of work that were undertaken to prepare the audio files for the listening experiment and SQM analysis.

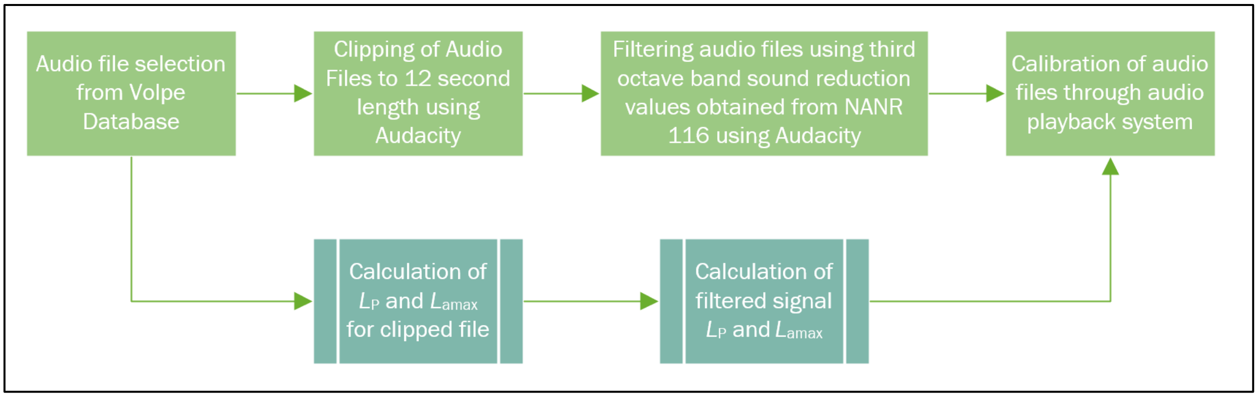


FIG. 2. (Colour Online) Preparation of Audio Files for Listening Experiment

B. Sound Transmission Through a Building Façade

To estimate the sound reduction through a building façade, with either a partially open or closed window, test data was obtained from the document titled 'NANR116: Sound Insulation Through Ventilated Domestic Windows' (Waters-Fuller *et al.*, 2007). The measurements presented within NANR116 are laboratory measurements but designed to emulate a typical residential receive room in terms of room dimensions and reverberation time. Therefore, the sound reduction values presented within the document are the *Apparent Sound reduction (R')* per third octave band or the *weighted Apparent Sound Reduction* (R'_w) denoting broadband performance. For this reason, no additional reverberation was applied to the audio files during the processing phase.

The sound reduction data used for this experiment were collected from measurements of a typical residential double glazing window configuration. More specifically, this window was an inward right hand swinging configuration with an area of 0.945 m^2. The glass/ airspace / glass arrangement was a 4 – 16 – 4 mm configuration. The partially open scenario had a free area of 0.05 m2, this free area was selected as the weighted sound reduction value (R'_w) of 12 dB and for the closed window the R'_w was 30 dB. The selected R'_w were also consistent with results presented within other research (Locher *et al.*, 2018) that summarised the results of numerous studies concluding that 'open window' scenarios typically exhibited sound reductions of 10 – 13 dB and 'window closed' scenarios typically between 26 and 31 dB.

For these measurements, only the third octave band R' data between 50 Hz and 5 kHz were recorded. For third-octave bands below 50 Hz, the same values as the 50 Hz third octave band were applied. As analysis of the drones' frequency content indicated they were not producing any significant levels of sound within this frequency range, the uncertainty introduced by this assumption is considered negligible. For frequencies above 5 kHz, for the partially open window an average of the previous three third octave bands (3.15, 4 and 5 kHz) were calculated and applied to all third octave bands up to 20 kHz. For the closed window, it was assumed that the mass law would dictate the sound reduction performance over 5 kHz which assumes a 6dB increase in performance per octave or, a 2 dB increase per third-octave. Both assumptions were considered reasonable in estimating the actual performance of the window at frequencies above 5 kHz although it must be acknowledged it does introduce a degree of uncertainty at very high frequencies.

C. Creation of Audio Files

The audio editing software 'Audacity' was used as it contains all the audio editing functions that were required to prepare the files for the experiment. The first step was to clip the audio files to the desired length (12 seconds each). Using the 12 audio files obtained from the Volpe outdoor measurement database, the files were then filtered using third-octave band sound reduction values to simulate the transmission through the double-glazed window, which was either partially open or closed, creating 36 audio files in total. This filtering was done using the 'Filter EQ Curve' tool within

Audacity. FIG. 3 below presents the *R'* values used for each third octave band for the two window conditions.

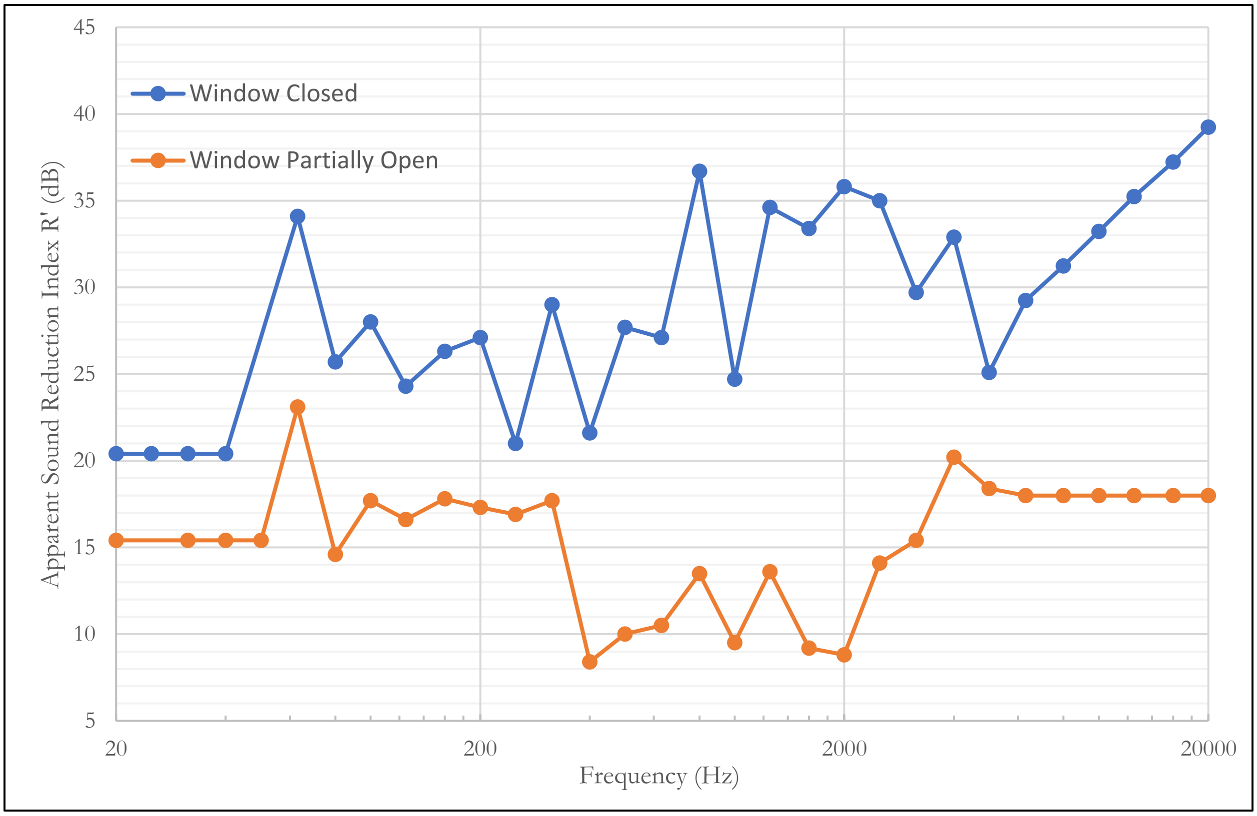


FIG. 3 (Colour Online) Third Octave Sound Reduction values – Partially-open and closed windows

A short 'fade-in' and 'fade-out,' approximately 100ms in length were applied to the audio files to avoid startling the participant. When 'clipping' the audio file zero crossing points were selected to avoid clicks or pops appearing at the beginning or end of the track.

D. Audio Reproduction System, Listening Room, and Calibration

The listening experiment was conducted within the 'Listening Room' at the University of Salford. This room is acoustically treated to reduce both reverberation and ambient sound levels. The L_{Aeq} value was measured as being between 20 – 22 dB. Although the listening experiment was administered through headphones, the headphones that were used were an open-back design and provide minimal isolation from external sounds. Any noise generating equipment such as laptops were positioned away from the participant and covered with acoustic foam. During pilot sessions, the noise from the listening experiment equipment was monitored, and it was concluded that any contributions to the ambient noise levels within the listening room were negligible.

The audio reproduction system used for the experiment was a laptop with Matlab software, external sound card (Motu 4Pre – Audio Interface, Cambridge Massachusetts, US), Headphone Amplifier (Little Labs 'Monotor', Los Angeles, California, US) and headphones (Beyer Dynamic DT 1990 Pro, Heilbronn, Germany). The calibration equipment included a Brüel & Kjær 2250 Class 1 Sound Level Meter (SLM) and Brüel & Kjær Artificial Ear Type 4153. The calibration process consisted of playing the audio files through the audio playback system to measure the L_{ASmax} and L_{Aeq} values. These values were compared with those measured by Volpe during the outdoor measurement campaign, gain corrections were then applied to the audio files within 'Audacity' to correct the broadband sound levels of the audio files. The 'corrected' audio files were then remeasured using the SLM to check the level. Priority was given to L_{ASmax} value during the calibration process as this would be unaffected by the 'fade-in' and 'fade-out' applied to the audio file. A calibration level within 0.5 dB of the L_{ASmax} value presented within the measured Volpe data (or value derived once third octave band sound reductions had been applied to the measured data) was considered suitably calibrated. FIG. 4 (A) presents the audio reproduction system and FIG. 4 (B) the calibration process. Table II presents the calibrated L_{Aeq} and L_{ASmax} noise levels of the 36 audio files used within the listening experiment are shown in Table II.

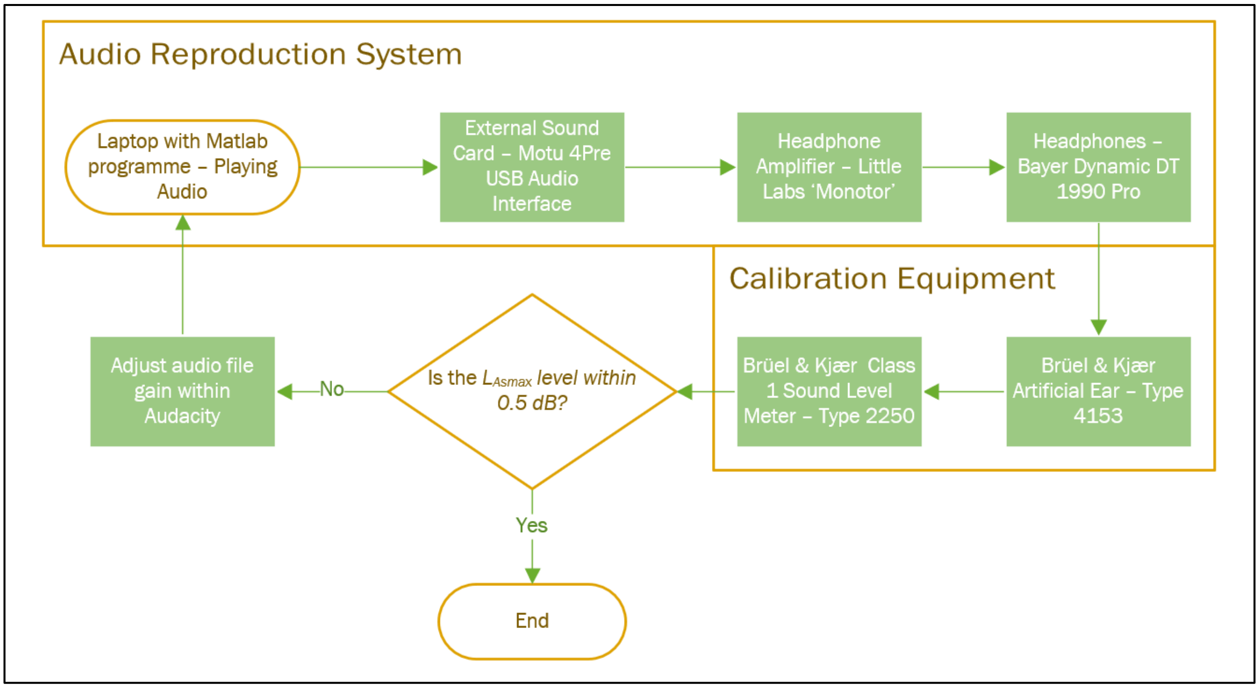


FIG. 4 (A) (Colour Online) Calibration process and audio reproduction equipment & FIG. 4 (B) Sound Level Meter and Artificial Ear Being Used for Calibration

Table II Calibrated L_{Aeq} and L_{ASmax} values of the audio files

Operation	Drone	L_{Aeq}	L_{ASmax}

		Outdoor	Part-Open	Closed	Outdoor	Part-Open	Closed
Flyover	GD28X	66.6	54.1	39	70.3	57.9	42.7
	M200	51.4	35.9	24.7	54.6	39.3	27.4
	TYPHOON	52.3	38.3	24.3	54.8	41.6	26.0
Landing	GD28X	69.8	56.5	40.6	72.1	58.8	43.4
	M200	64.2	51.3	35.6	67.0	55.0	38.6
	TYPHOON	61.9	48.8	31.6	64.2	51.3	34.4
Take-Off	GD28X	72.7	59.9	44.3	74.4	61.5	46.0
	M200	58.0	44.6	29.6	59.8	47.2	31.1
	TYPHOON	56.3	43.4	27.4	60.3	47.2	32.2
Hover	GD28X	75.4	59.7	46.5	76.3	60.7	47.4
	M200	58.8	46.1	30	59.6	46.9	30.5
	TYPHOON	55.8	41.8	26.4	56.9	42.9	27.2

The calibrated L_{Aeq} and L_{ASmax} noise levels of the 36 audio files presented within Table II show a wide range of sound levels used within the listening experiment ranged between 26.0 – 76.3 dB and 24.3 – 75.4 dB for the L_{ASmax} and L_{Aeq} respectively. The rationale for presenting the stimuli across this wide range of sound levels was to simulate the actual level of noise that would be experienced by the listener in each of the indoor / outdoor listening positions. Although, it should be noted that without the presence of masking noise to simulate real-world listener scenarios, participants were asked to rate stimuli, some with very low noise levels which may be rendered negligible or inaudible once even a small amount of ambient sound is introduced. However, the selected method was decided upon for the study to focus solely on how changes to the stimuli affect participant responses. The research team recognise that further research considering the acoustic context of various listener positions and other non-acoustic factors will be important in understanding more about human response to UAS noise.

E. Uncertainty within the Audio Stimuli

It should be acknowledged that the processing and calibration of the data, outlined within Subsections C and D, presents a risk in terms of introducing distortions or unwanted artifacts within the sound. Measures were taken at every stage of the stimuli preparation to reduce the risk of introducing unwanted audio distortions. Professional grade audio reproduction hardware was used within the experiment with care taken at all stages of the process to reduce the risk of introducing noise to the signal. Whilst the exact magnitude of distortion within the audio files has not been quantified, it is thought the risk of audible distortions arising from the processing and calibration is very low.

F. Questionnaire and Interface

Before the experiment, participants were provided with an overview of the task and format of the experiment, the task overview stated that they will be listening to sounds from UAS but not that filtering had been applied to simulate external to internal transmission. Once the participants had been given time to read the instructions and ask questions they put on the headphones and were presented with four 'familiarisation sounds', these sounds were not used within the main experiment but were selected as they highlighted the range of sounds the participant would be presented. Participants were able to replay the sounds if they wanted, once the participant had listened to each of these sounds, they were given one more opportunity to ask questions before the experiment began.

The listening experiment interface was created within Matlab (Version R2022a). The interface of the experiment presented the participant with a single audio file randomly selected from the 36 files. The interface had a 'Play Sound' button, two 11-point sliders (0 to 10) one to rate the 'Annoyance' and the other to rate the 'Perceived Loudness'. At the bottom, a next button which could only be pressed once the 'Play Sound' button has been pressed, this was to avoid participants accidentally missing the stimuli. Finally, there was a counter in the top right corner showing remaining stimuli to help maintain concentration.

G. Participant Information

The participants that took part within the experiment were sourced from a mailing list which is maintained by the university to notify students, staff and alumni from across the entire university about upcoming listening experiments.

Initially, 31 participants took part within the experiment. However, due to an error with the data collection for one of the participants, their response was excluded from the analysis. Therefore, the final number of participants was 30. A sample size of 30 has been previously adopted for prior UAS listening experiments and is typically considered to be the smallest size required to be able to test for both Type I errors (rejection of a true null hypothesis) with 95% confidence and having a sufficiently large statistical power to reduce the risk of Type II errors (acceptance of a false null hypothesis) (Torija *et al.*, 2020b; Lakens, 2022). However, it is noted that if additional resources were available at the time of the experiment, increasing the sample size would have resulted in a greater statistical power. Table III presents the demographical information collected for each of the participants.

Table III Demographical Information of Participants

Gender			
Category	Male	Female	Other
Number of Participants	22 (73%)	7 (23%)	1 (3%)

Age Range					
Category	18 – 24	25 - 34	35 – 44	45 – 54	55 - 64
Number of Participants	11 (37%)	9 (30%)	6 (20%)	3 (10%)	1 (3%)
English as Native Language					
Category	Yes		No		
Number of Participants	19 (63%)		11 (37%)		
Self-Identify as having a Hearing Impairment					
Category	Yes		No		
Number of Participants	3 (10%)		27 (90%)		

As the information in Table III demonstrates, most participants were male and tended to fall into the younger age ranges, either '18 – 24' or '25 – 34'. Of those who identified as having a hearing impairment the additional information provided highlighted some loss of high frequency response to their hearing and one participant mentioned they suffer from mild tinnitus but also noted that it did not interfere with their daily life. Based on the participant information provided, all response data was included within the analysis.

H. Statistical Analysis

The association between acoustic metrics and SQMs with subjective responses (i.e., Annoyance and *PL*) for the series of UAS sound samples presented to the participants was investigated using linear regression and multilevel models. As shown in Eq. (1), the subjective response (either Annoyance or *PL)* Y_i for the ith sound is predicted as:

$$Y_i = \gamma_0 + \gamma_1 \chi_{1i} + e_i \qquad (1)$$

Where γ_0 is an intercept, γ_1 is the slope with respect to acoustic metric or SQM χ_1, χ_{1i} is the value of the acoustic metric or SQM of the ith sound, and e_i is the residual error. In a linear regression approach, a 'complete pooling' takes place, as all participants' responses are aggregated in the analysis. In contrast, in a multilevel analysis, a 'partial pooling' is possible allowing regression parameters (i.e., intercept and slope) to vary randomly across participants (Hox *et al.*, 2017).

A multilevel analysis augments a linear regression analysis by providing both participant-specific and aggregate regression parameters in one analysis. Boucher et al. (2023) provide a detailed description of the multilevel analysis, and its use for the analysis of transportation noise. A multilevel analysis approach was implemented by Torija et al. (2020a) to investigate the contribution of a series of acoustic and non-acoustic factors to the perception of different urban soundscapes with a UAS hovering.

The formulation of a multilevel regression analysis is shown in Eqs. (2-4).

$$Y_{ij} = \beta_{0j} + \beta_{1j}\chi_{1i} + e_{ij} \tag{2}$$

Where Y_{ij} is the subjective response, either Annoyance or *PL,* for the ith sound and the jth participant, and e_{ij} is the residual error,

$$\beta_{0j} = \gamma_{00} + \mu_{0j} \tag{3}$$

Where γ_{00} is an overall mean intercept for all participants, and μ_{0j} is a participant-specific intercept offset, and

$$\beta_{1j} = \gamma_{10} + \mu_{1j} \tag{4}$$

Where γ_{10} is an overall mean slope for the metric χ_1, and μ_{1j} is a participant-specific slope offset. Both μ_{0j} and μ_{1j} are assumed to follow a normal distribution.

These statistical analyses were conducted with the IBM SPSS Statistics package (version 29).

IV. Results

A. Annoyance and Perceived Loudness as a function of UAS operational procedure and listener position

FIG. 5 presents the average participant response data for annoyance and *PL* separated by the listener response position.

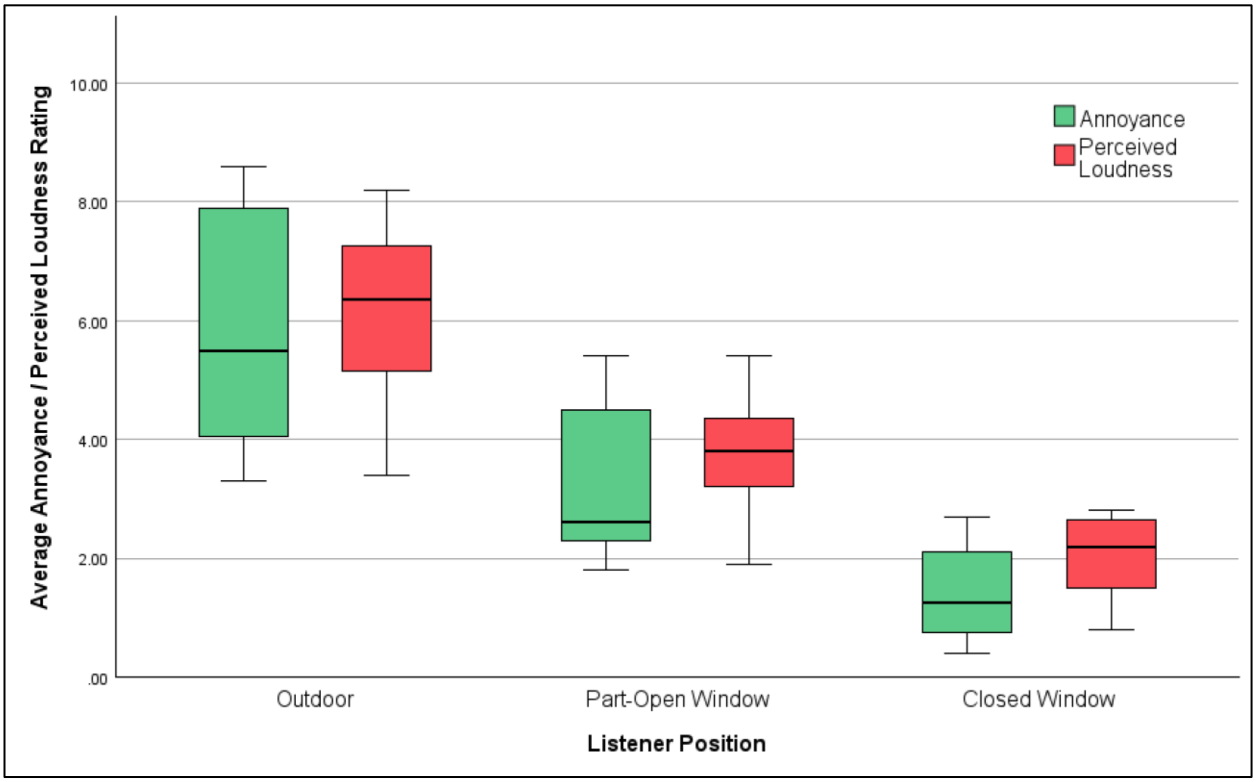


FIG. 5 (Colour Online) Boxplots presenting Average Annoyance and *PL* Ratings Separated by Listener Positions.

The participant responses show a clear trend for both annoyance and *PL* with responses being the highest (i.e., most annoying, and loudest) when the listener was in the outdoor position followed by indoors 'partially-open window' then the 'closed window' scenario. The significance of the listener position on annoyance and *PL* was evaluated using a one-way Analysis of Variance (ANOVA) which found the differences in average responses to be statistically significant (Annoyance: *F* [2,33]

= 28.124 P = <0.001) (PL: F [2,33] = 49.844 P = <0.001). The significance of the listener position on effect size can be better visualised using a Cumming estimation plot, making use of bootstrap resampling to determine the confidence interval (CI) of the effect size, see (Ho *et al.*, 2019) for more details. FIG. 6 presents the individual participant annoyance ratings separated by the listener position (left), and the mean difference and 95% CI between the listener positions with 'Outdoors' being the control group and the other 'test' groups (right). For annoyance, results demonstrated a mean difference between Outdoors and Open Window scenarios is -2.59 (upper and lower bounds of 95% CI [-2.93 and -2.27]). The mean difference between Outdoors and Closed Window is -4.5 (upper and lower bounds of 95% CI [-4.8 and -4.21]). For PL, the mean difference between Outdoors and Open Window scenarios is -2.34 (upper and lower bounds of 95% CI [-2.68 and -1.98]). The mean difference between Outdoors and Closed Window is -4.05 (upper and lower bounds of 95% CI [-4.37 and -3.72]).

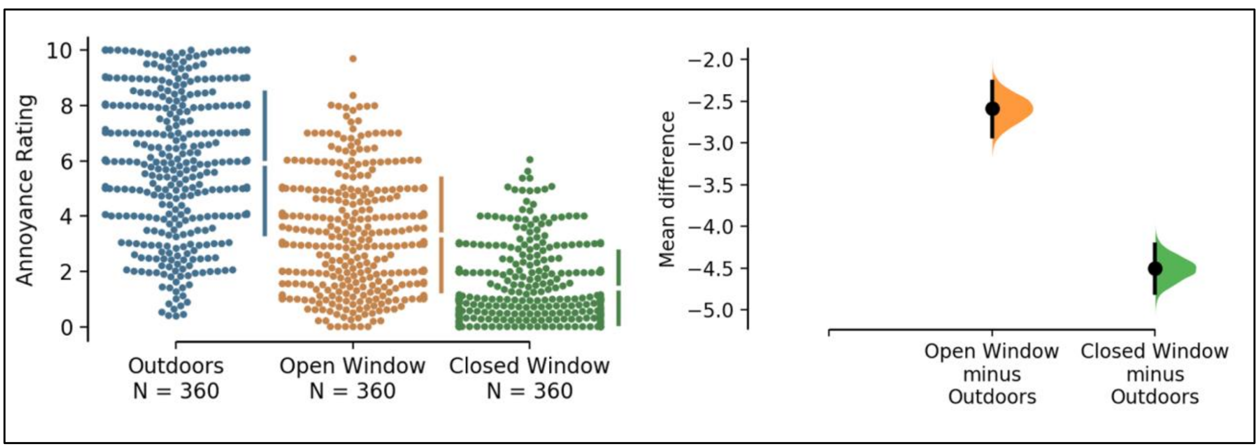


FIG. 6 (Colour Online) Cumming estimation plot presenting annoyance ratings separated by listener position and mean difference and 95% CI upper and lower bounds between the control and test groups. N = number of participant resposnes

The data can also be separated by operation. FIG. 7 presents an estimation plot presenting participant annoyance ratings separated by the UAS operation and using the flyover operation as the control group. The mean difference along with the lower and upper bound of the 95% CI between the control and test operations are presented in Table IV.

Table IV Differences in Participant Annoyance and PL for the four UAS Operations

Operation	Annoyance			Perceived Loudness		
	Mean Difference	Lower Bound	Upper Bound	Mean Difference	Lower Bound	Upper Bound
Take-off	0.85	0.42	1.3	1.38	0.93	1.83
Landing	1.33	0.90	1.79	1.48	1.04	1.91
Hover	0.72	0.28	1.17	1.03	0.58	1.46

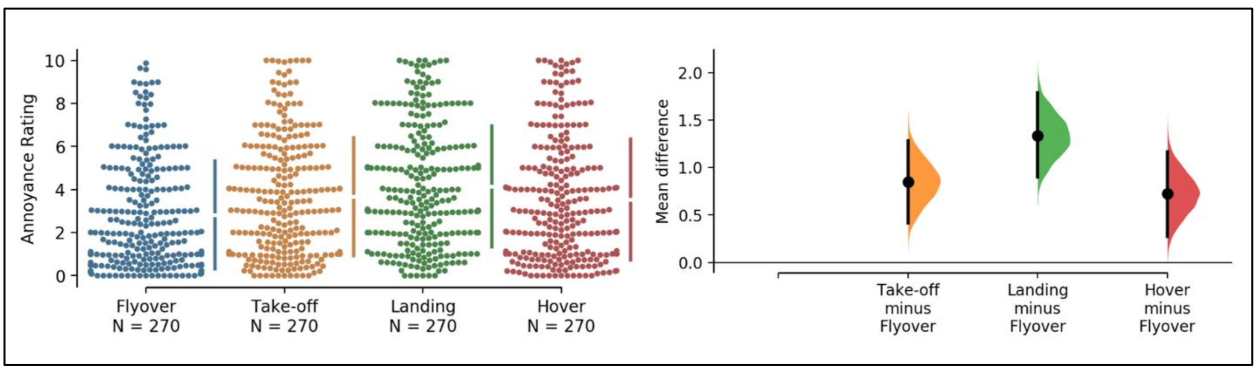


FIG. 7 (Colour Online) Cumming estimation plot presenting annoyance ratings separated by UAS operation and mean difference and 95% CI upper and lower bounds between the control and test groups

For both annoyance and *PL*, responses to each of the operations appear to follow the same trend with landing exhibiting the greatest mean difference, followed by takeoff, hover and flyover. This suggests that the possibility of annoyance is increased when exposured to noise from UAS operations other than flyovers. However, as the sound levels of the audio files were not standardised, some of these variations within the participant responses are likely a result of changes to noise levels. For example, the difference between the L_{ASmax} for the 'Outdoor' flyover and hover operations of the GD28X drone were +6.0 dB for the hover. Therefore, it is not entirely clear at this stage whether the increased annoyance and *PL* resulted from the characteristics of the operation or, whether the participants were responding to differences in noise level. The significance of the differences in noise level have been investigated further in Section C.

B. Comparison of Loudness Metrics

Previous research into the perception of UAS and other environmental noise sources has demonstrated that 'loudness' is the most significant characteristic of the sound when assessing it for both Annoyance and *PL* (Gwak *et al.*, 2020b; Nicholls, 2021). To better understand the efficacy of different loudness metrics, six metrics have been used to model the participant response data. The conventional metrics of L_{Aeq}, L_{Asmax} and L_{AE} along with other metrics such as Perceived Noise Level (*PNL*), Effective Perceived Noise Level (*EPNL*) and the Sound Quality Metric (SQM) Loudness – DIN 45631/A1 model (N_5). Note that the standard 5th percentile value of Loudness (i.e., loudness exceeded 5% of the time) was used for the analysis; and that the first 0.5 s of the sound sample was excluded from the calculation to avoid transient effects of the digital filters used for the calculation of the metric (Torija *et al.*, 2021).

Each of the metrics mentioned above have been used to model annoyance and *PL* response data using simple linear regression analysis and bootstrapping to calculate the Confidence Intervals (CIs). The R^2 values and 95% CIs for each of the loudness metrics (independent variable) are presented within Table V.

Table V Results of Regression Analysis for each for the Broadband Metrics with 95% Confidence Interval

Metric	Annoyance R^2	Annoyance CI	PL R^2	PL CI
L_{ASMax}	0.93	0.91, 0.96	0.90	0.84, 0.95
L_{Aeq}	0.93	0.91, 0.96	0.90	0.82, 0.95
L_{AE}	0.91	0.88, 0.95	0.88	0.80, 0.93
PNL	0.91	0.87, 0.95	0.88	0.80, 0.94
EPNL	0.91	0.86, 0.94	0.89	0.82, 0.94
Loudness (N_5)	0.94	0.92, 0.97	0.87	0.82, 0.93

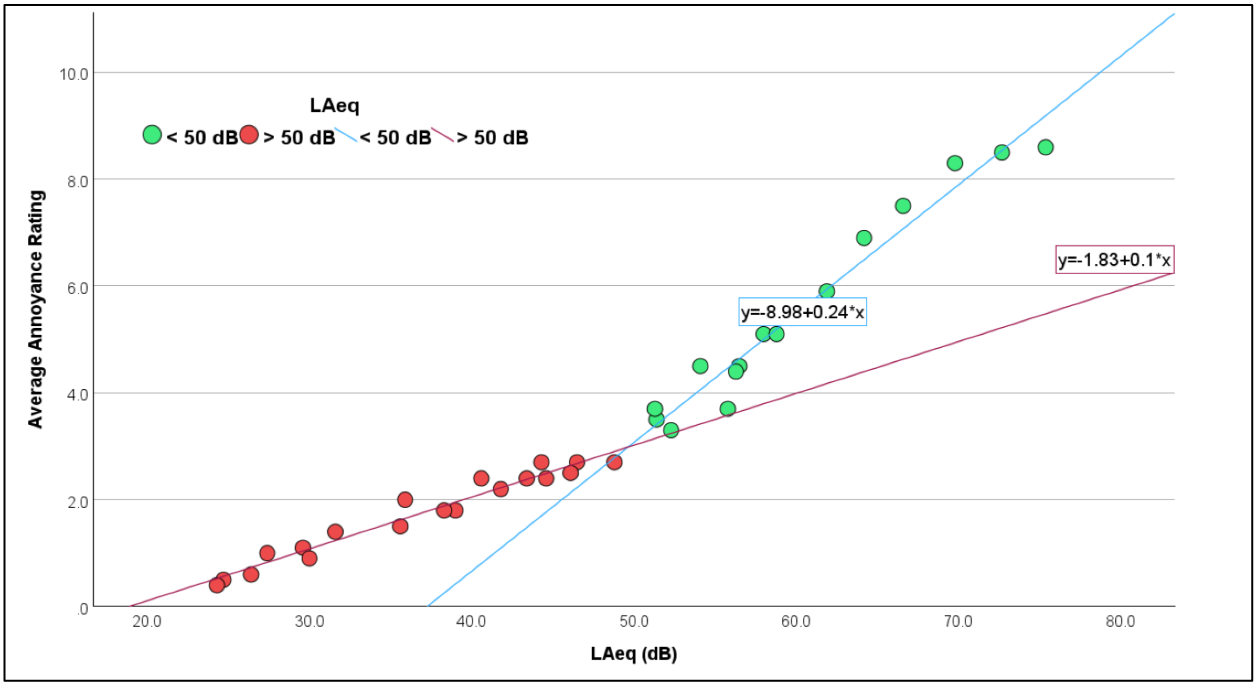
Results of the regression analysis for Annoyance show that the SQM Loudness performed best with an R^2 value of 0.94. L_{Aeq} and L_{Asmax} scored marginally lower with R^2 values of 0.93. PNL, EPNL and L_{AE} scored slightly lower with R^2 values of 0.91. For modelling *PL*, L_{ASMax}, L_{Aeq} both scored an R^2 value of 0.90. *EPNL* scored slightly lower with 0.89, *PNL* and L_{AE} with 0.88 and Loudness the lowest with 0.87.

The A-weighted metrics generally performed better than PNL and EPNL. This is most likely a result of the A-weighted metrics better representing how humans are sensitive to the spectral content of the stimuli. A small difference was observed between the results for *PL* and *PNL* or *EPNL*, this may be a result of the tonal corrections applied within *EPNL* which could help predict the effect of tones on the *PL* of a stimuli.

The positive relationship between sound level or Loudness and average annoyance can also be seen by plotting average annoyance against L_{Aeq} and N_5, as can be seen in FIG. 8 (A) and FIG. 8 (B). For

L_{Aeq} noise levels between 20 – 50 dB, the average annoyance response increases steadily as the sound level increases, with scatter plot slope of 0.1. From approximately 50 dB L_{Aeq} and above, increases to loudness appear to cause a sharper increase to annoyance as the slope increases to 0.24 for L_{Aeq}. When plotting the average annoyance against Loudness (Sones), the scale has a more linear representation of average annoyance which suggests Loudness is the more consistent performer across different sound levels.

For L_{Aeq}, the change in gradient of the slope is consistent with the response data of the 'outdoor' stimuli, the significance of Sharpness was investigated (as high frequencies were more attenuated than the other frequency regions during the outdoors-to-indoor propagation). Multiple linear regression analysis was used to plot L_{Aeq} and Sharpness (Aures model) against Average Annoyance. The inclusion of Sharpness resulted in a statistically significant improvement (Adj. R^2 change = 0.007, Sig. F Change [1,33] = 0.034) of the Adjusted R^2 value suggesting Sharpness could be a factor influencing the greater increase to recorded annoyance. Alternatively, the reason could be associated with the overall noise level or loudness of the noise events of around 50 dB being the onset level of a more adverse response. In the UK, a Survey of Noise Attitudes (SoNA) for aircraft (CAA, 2017) recommended that the 'Lowest Observable Adverse Effect Level' and the 'onset of significant annoyance' be set at 51 dB and 54 dB $L_{Aeq,16hr}$ respectively. The analysis of L_{Aeq} data corresponds closely with the thresholds recommended within the SoNA report.



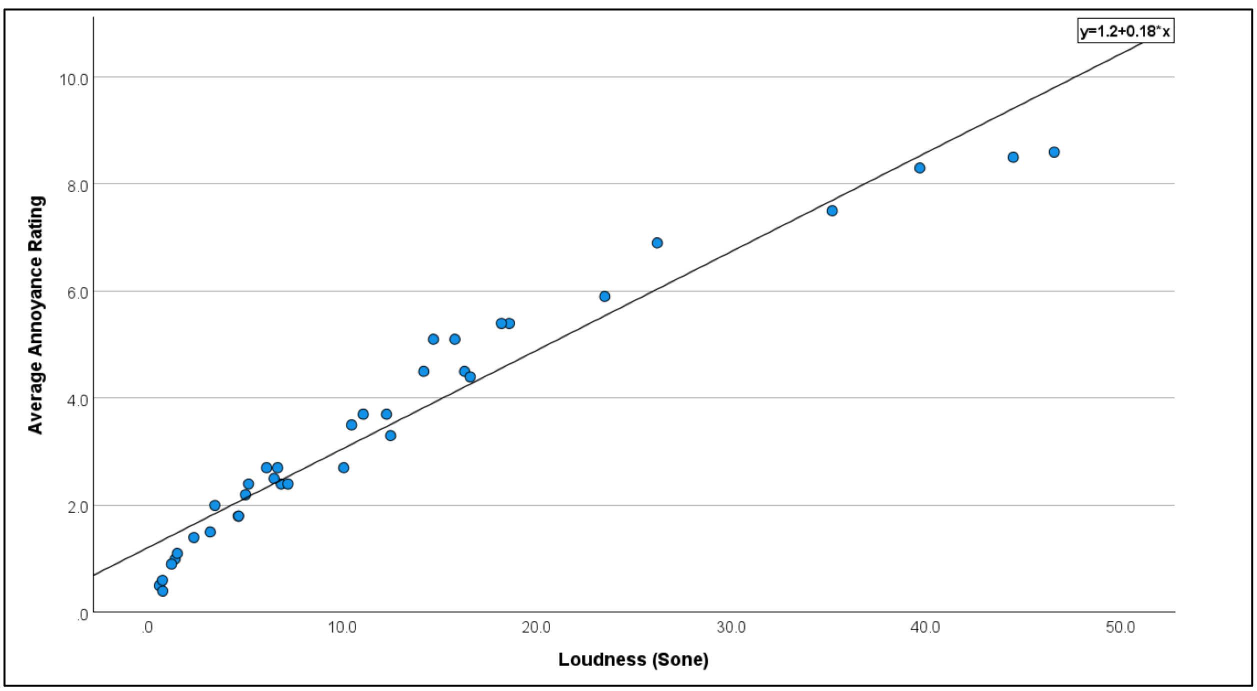


FIG. 8 (A) (Colour Online) Scatter graph plotting Average Annoyance against L_{Aeq} separated at 50 dBA level & FIG. 8 (B) Scatter graph plotting Average Annoyance against Loudness

Interestingly, the SQM Loudness performed the least effectively of all the analysed metrics for predicting *PL*. Although the exact reason is unknown, this could be a limitation of the metric itself as the DIN 45631 / A1 Loudness model is known to have some limitations when calculating the Loudness of time-varying sounds such as many of those presented within the experiment (Sottek, 2014; 2016; Völk, 2016)

C. Offset Analysis

Following a procedure suggested by Christian and Cabell (2017), the presence of systematic differences between UAS operational conditions in terms of annoyance and *PL* were investigated. A series of linear regression analyses adding a binary term C were conducted (see Eq. 5).

$$Y_i = \gamma_0 + \gamma_1 \chi_{1i} + \gamma_2 C_i + e_i \quad (5)$$

This binary term C is a dummy variable representing the type of operational condition; and γ_2 is the slope with respect to the binary term C. This augmented linear regression was conducted for each operational condition investigated, where C_i=1 corresponded to flyover and C_i= 0 corresponded to either hover, take-off, or landing. The offset, measured in the units of the specific acoustic metric or SQM, was calculated as γ_2/γ_1.

As shown in Table V, the inclusion of the binary predictor C in the linear regression analysis for annoyance was found to be non-significant in all cases (p-value > 0.05). Consequently, the explanatory value (R^2) of the model for all acoustics metrics and Loudness was not improved. The offset value for most metrics, in their respective units, was reduced. Looking at the offset values for Loudness (N_5) in sones, the metric with the highest correlation with annoyance (see Table IV), the results suggest that the Loudness of the hover, take-off and landing operations would all need to be reduced by either 1.0 or 1.1 sone to be rated as equally annoying as the flyover operation.

Table VI. Offset or Δ values between metrics for hover, take-off and landing operations vs flyovers for Annoyance. Offset values are shown in the respective metric's unit. R^2 values are shown in brackets.

	Hover	Take-off	Landing
L_{ASMax} (dB)	-0.9 (0.93)	-0.5 (0.93)	-0.5 (0.92)
L_{Aeq} (dB)	1.4 (0.94)	0.0 (0.94)	0.0 (0.93)
L_{AE} (dB)	0.7 (0.80)	-0.1 (0.92)	1.0 (0.90)
PNL (PNdB)	2.6 (0.91)	0.3 (0.90)	1.3 (0.90)
EPNL (EPNdB)	2.1 (0.90)	0.4 (0.91)	0.8 (0.90)
Loudness (N_5, sone)	-1.1 (0.93)	-1.0 (0.94)	-1.1 (0.97)

In Table VII, it is shown that the inclusion of the binary predictor C in the linear regression analysis for *PL* is significant at a p-value < 0.05 or 0.1 for the metrics L_{ASMax}, L_{Aeq}, L_{AE}, PNL and Loudness (N_5) when comparing flyover to take-off operations. This is also true for L_{ASmax} for the comparison of flyover and hover as it has been predicted that the L_{ASmax} of the hover operation would need to be 4.2 dB quieter than the flyover operation to have the same *PL*. The explanatory value (R^2) of the model for these metrics increases consequently. In this case, there is great consistency between the different metrics, clearly indicating that (1) flyover operations are perceived as less loud than hover, landing and take-off operations, and (2) take-off operations generally require the greatest offset from flyovers to achieve equal *PL* with L_{Aeq}, L_{ASmax}, L_{AE} for take-off operations all requiring corrections in the range of -5.9 to -6.6 dB to achieve equal *PL*.

Table VII. Offset or Δ values between metrics for hover, take-off and landing operations vs flyovers for Perceived Loudness. Offset values are shown in the respective metric's unit. R^2 values are shown in brackets. * p-value < 0.1, ** p-value < 0.05

	Hover	Take-off	Landing

L_{ASMax} (dB)	-4.2 (0.93)*	-6.3 (0.91)**	-3.3 (0.93)
L_{Aeq} (dB)	-1.9 (0.93)	-5.9 (0.89)**	-2.8 (0.93)
L_{AE} (dB)	-3.3 (0.77)	-6.6 (0.85)*	-2.1 (0.91)
PNL (PNdB)	-1.2 (0.91)	-6.6 (0.87)*	-1.9 (0.91)
EPNL (EPNdB)	-1.5 (0.92)	-6.3 (0.88)	-2.3 (0.92)
Loudness (N_5, sone)	-4.0 (0.89)	-6.1 (0.89)**	-3.6 (0.95)**

D. Multilevel Analysis

A multilevel analysis was carried out to investigate the contribution of psychoacoustic features other than Loudness to annoyance and *PL*. A series of multilevel regression analyses were performed, according to Eq. 2, with annoyance or *PL* as dependent variables, and Loudness (ANSI S3.4 2007), Sharpness DIN 45692 , Fluctuation Strength, Roughness, Tonality, and Impulsiveness as predictors. Fluctuation Strength, Roughness, Impulsiveness and Tonality metrics were calculated using the hearing model developed by Sottek (1993). Similar to the Loudness metric, the 5th percentile of the Sharpness, Fluctuation Strength, Roughness, Tonality and Impulsiveness metrics were used for the analysis; and the first 0.5 s of the sound sample were excluded from the calculation.

Four multilevel analysis models were built for each listening condition, outdoors, indoor with partially open window and indoor with closed window:

- Model M0, with fixed intercept (γ_0) and fixed slopes (γ_1). This is equivalent to a conventional multiple linear regression.

- Model M1, with variable intercept ($\gamma_{00} + \mu_{0j}$) and no predictors. This model accounts for the participants using different ranges of the annoyance and *PL* scales (Boucher *et al.*, 2023).
- Model M2, with variable intercept ($\gamma_{00} + \mu_{0j}$) and fixed slopes (γ_{10}).
- Model M3, with variable intercept ($\gamma_{00} + \mu_{0j}$) and variable slopes ($\gamma_{10} + \mu_{1j}$). This is to account for different changes in annoyance and *PL* as a function of changes in psychoacoustic features between participants.

The explanatory values (R^2) of each multilevel analysis model for annoyance and *PL* for the three listener positions are presented in Table VII. As shown in Table VII, the R^2 values of the models for annoyance are consistently higher than for *PL*. The only exceptions are M1 models, where the R^2 values are higher for *PL* than for annoyance. M1 models for *PL* have also R^2 values higher than M0 values. This suggests a different interpretation and use of the *PL* scale between participants. Moreover, comparing R^2 values of the M1 models between outdoors and indoor listener positions, it can be seen the influence of the scale use for both annoyance and PL in the quieter environments. The R^2 values consistently increase when variable intercept and slopes are used in the multilevel analysis.

Table VIII. R^2 values for each multilevel analysis model and listener position, for both annoyance and Perceived Loudness

	Outdoors		Indoor – Partially Open Window		Indoor – Closed Window	
	Annoyance	Perceived Loudness	Annoyance	Perceived Loudness	Annoyance	Perceived Loudness
M0	0.54	0.28	0.38	0.15	0.34	0.10

M1	0.23	0.34	0.34	0.51	0.36	0.45
M2	0.76	0.61	0.71	0.66	0.70	0.55
M3	0.83	0.76	0.80	0.75	0.85	0.69

The contribution of each SQM to annoyance and *PL* was assessed by the reduction in R^2 when such SQM is removed from the multilevel analysis model. A small reduction in R^2 implies that the specific SQM is of less importance; while a substantial reduction in R^2 implies that the SQM' importance is large.

As shown in FIG. 9, Loudness (N_5) is the main contributor to annoyance in the three listener positions. Outdoors, Impulsiveness (I_5), seems to play an important role for both annoyance and PL. After further exploration, I_5 values are significantly higher for landing than for the other operational conditions (see FIG. 9), which seems to suggest that the contribution of this SQM to annoyance and PL might be due to Blade Vortex Interaction (BVI) nose (Yung, 2000). For PL, the contribution of N_5 seems to be smaller, with important contributions of other SQMs, such as Fluctuation Strength (FS_5), I_5 and Tonality (T_5). In indoor environments, T_5 seems to be the most important predictor for *PL*.

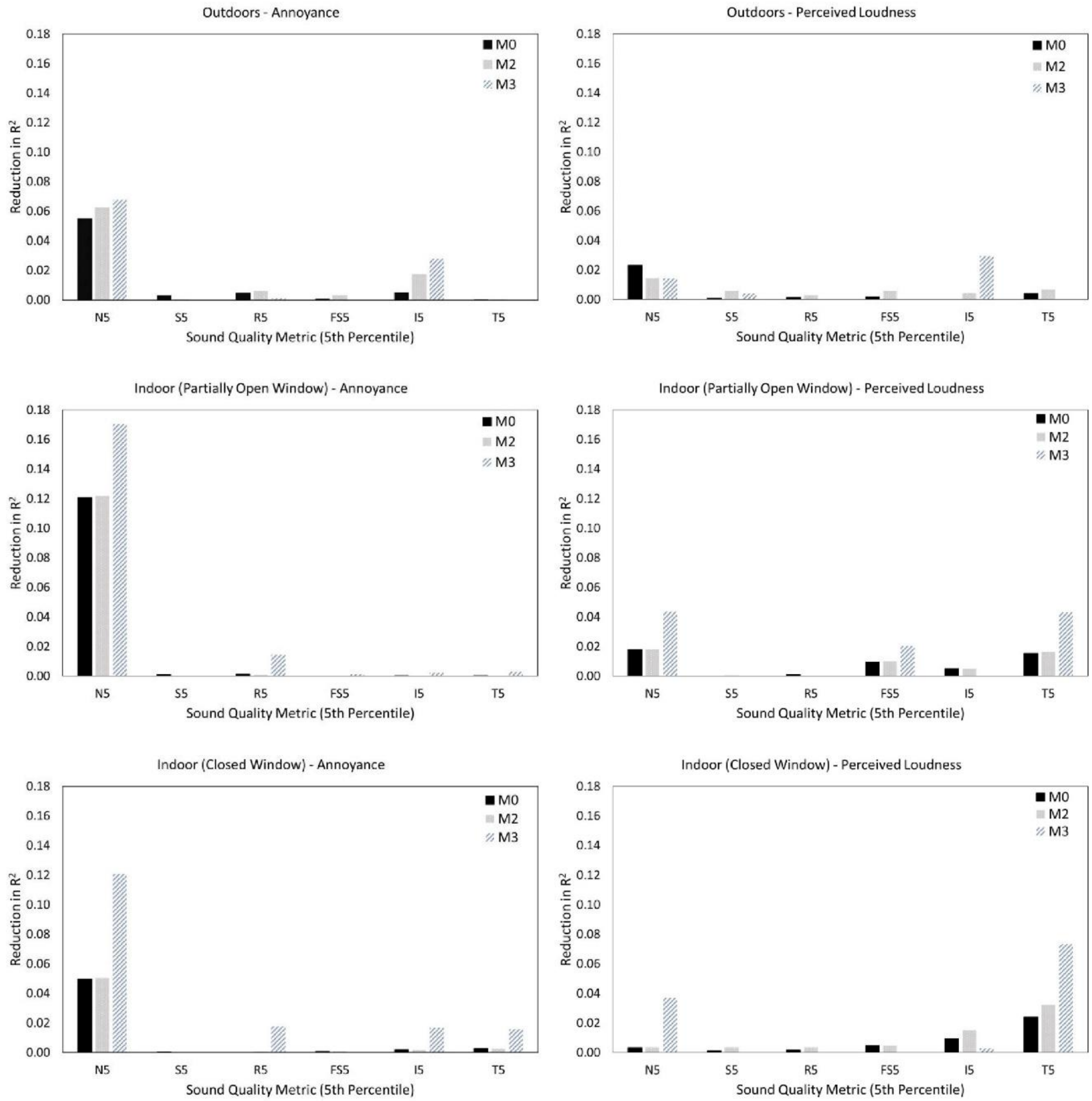


FIG. 9 Reduction in R^2 for the Sound Quality Metrics Loudness (N_5), Sharpness (S_5), Roughness (R_5), Fluctuation Strength (FS_5), Impulsiveness (I_5) and Tonality (T_5) when predicting annoyance (left) and Perceived Loudness (right), for outdoors (top), indoor with partially open window (middle) and indoor with closed window (bottom).

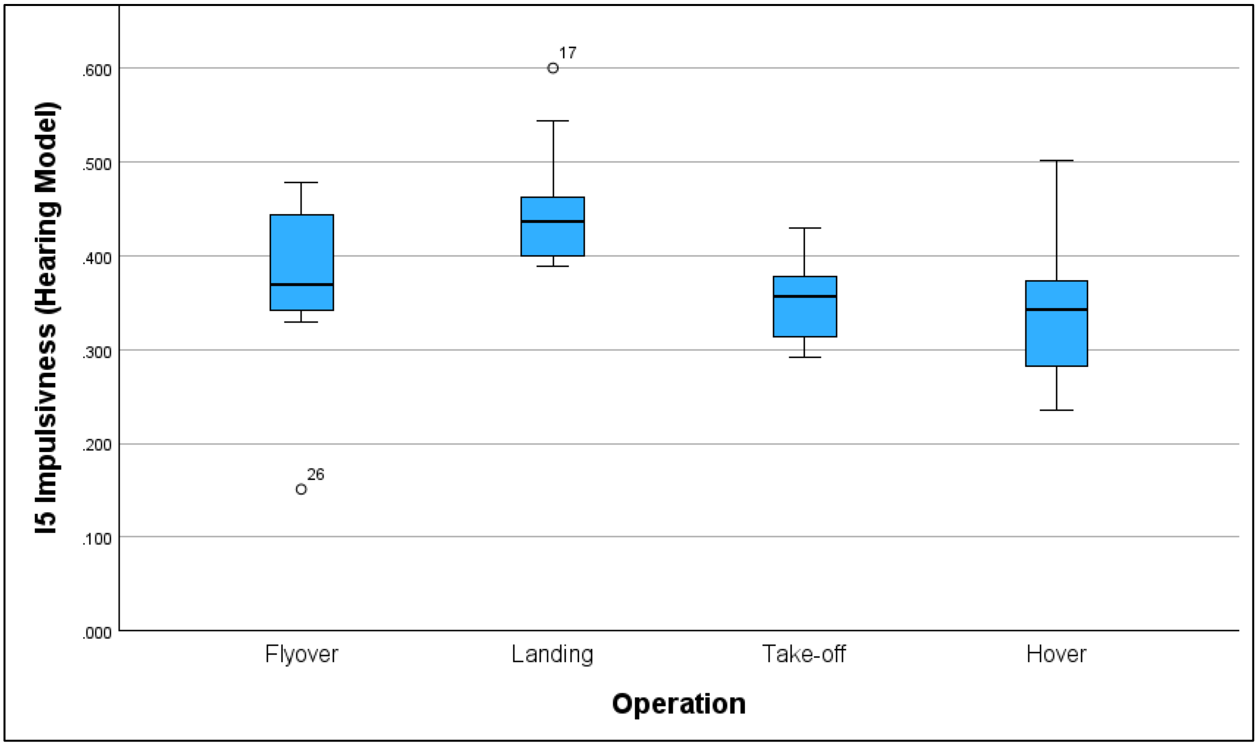


FIG. 10 (Colour Online) Box and Whisker plot presenting Impulsiveness for each UAS Operation

V. CONCLUSION

This paper investigates the changes in responses of annoyance and *PL* (i) with different UAS operations (i.e., flyover, hover, landing and take-off), and (ii) when a listener is in either an indoor or outdoor position. This paper also investigated the performance of a series of loudness based metrics, and complementary SQMs accounting for spectral and temporal characteristics, to explain such changes in annoyance and *PL*.

The participants' responses demonstrated that there was a statistically significant variation when comparing annoyance or *PL* simulated in different listener positions. Landing operations were considered the loudest and most annoying, followed by take-off and hover. Flyovers were perceived to be the least loud and annoying of the different operations.

Broadband noise metric analysis was undertaken to understand which metrics have the highest efficacy in predicting average annoyance and *PL*. The results of the analysis for annoyance indicate that across the different operations Loudness (N_5) performed the best, with an R^2 value of 0.94, followed by L_{Aeq} and L_{ASmax}, both with a value of 0.93. For *PL*, L_{ASmax} and L_{Aeq} performed the best, with an R^2 value of 0.90, followed by PNL, EPNL and L_{AE} 0.88 and Loudness (N_5) at 0.87. Further analysis of the L_{Aeq} metric suggests annoyance begins to increase at a faster rate when noise levels are above the 50 dB level.

The differences in response to the different UAS operations have been quantified through Offset analysis. Specifically, this analysis method was used to understand the noise level difference required to achieve an equal Annoyance or *PL* level between the different operations. Results of the offset analysis showed that only minor differences, less than 1 dB, are required for L_{Aeq}, L_{ASmax} and L_{AE} to achieve equal annoyance between the different operations. However, the differences become more pronounced when analysing *PL* with a flyover required to be 3.3 to 6.6 dB louder (when considering the L_{ASmax}) to be considered equally loud as other operations. For Loudness, flyovers would need to increase by between 3.6 and 6.1 sone to be considered equally loud as the other operations.

Results of the multilevel regression analysis demonstrate that Loudness was the principal factor for predicting annoyance and *PL*. However, there were contributions from other SQMs in specific scenarios. Impulsiveness influences responses of annoyance and *PL* when the listener was outdoors, which is thought to be potentially associated with Blade Vortex Interaction noise during the landing operation. For *PL* when the listener was in either of the two indoor scenarios tonality appears to play a role which became more significant in the window closed scenario. Further research will be conducted to investigate the perception of the tonal UAS noise in indoor environments, as this research demonstrates that it could be a key factor in their perception.

Several of the limitations of this study has been considered below along with the steps on how they could be addressed in future research. A single window opening type was considered within this experiment. Further research could look at multiple window configurations or, additional openings to better understand the effect of different configurations on UAS noise. Similarly, the number of UAS or UAS configurations could be increased from the three included within this study. Whilst the three UAS included all varied in their configurations, the nature of UAS means there are many more configurations dependant on size, weight, number of rotors, rotor diameter, number of blades etc. All of which contribute to the sound character generated by the UAS and may be perceived differently depending on whether the listener is indoors or outdoors. The limitations of the DIN 45631/A1 Loudness model were acknowledged for *PL* analysis as it has seemingly not dealt well with the time-varying nature of some stimuli particularly well. Further research could expand the analysis to evaluate the efficacy of other Loudness models, some of which may be better suited to predict the *PL* of time-varying sounds.

ACKNOWLEDGMENTS

The authors would like to acknowledge the funding provided by the UK Engineering and Physical Sciences Research Council for the Drone Noise project (EP/V031848/1).

The authors would like to thank VOLPE for making their library of drone audio measurements available to us.

The authors would like to acknowledge Lara Harris (formerly University of Salford Acoustics Research Centre [ARC]) and Connor Welham (ARC) for their assistance with organizing the listening experiment and designing the MATLAB program. Also, Michael Lotinga (ARC) for his advice and input into the statistical analysis and technical review.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts of interest to declare. All co-authors have seen and agree with the contents of the manuscript and there is no financial interest to report. We certify that the submission is original work and is not under review at any other publication.

Ethics Approval

The listening experiment methodology and data collection procedures were presented to and granted approval by the University of Salford’s ethics committee (ID 6088) and are consistent with the ethical principles of the Acoustical Society of America. Informed consent was obtained for all participants involved in experiment described in this article prior to undertaking the experiment. Participant privacy was also of critical importance, with any sensitive information being collected and stored in line with General Data Protection Regulation (GDPR). For this experiment, participants were offered a modest inducement (£15) as an incentive to take part.

DATA AVAILABILITY

Audio stimuli, participant response and acoustic and sound quality metric data created for use within the listening experiment and described within this article is available here: http://tinyurl.com/mr42wesb

VI. REFERENCES

Alexander, W. N., and Whelchel, J. (**2019**). "Flyover noise of multi-rotor sUAS," in *Inter-Noise and Noise-Con Congress and Conference Proceedings* (Institute of Noise Control Engineering), pp. 2548-2558.

US Federal Government. US Code of Federal Regulations Title 14, Subchapter C, Part 36, Appendix J to Part 36 — Alternative noise certification procedure for helicopters under Subpart H having a maximum certificated takeoff weight of not more than 7,000 pounds. National Archives and Records Administration Office of the Federal Register. Available from: https:// www.ecfr. gov/ curre nt/title- 14/ chapt er-I/ subch apter-C/part-36/appendix-Appen dix% 20J%20to% 20Part% 2036.
Babisch, W. (**2011**). "Cardiovascular effects of noise," Noise & Health **13**, 201.
Basner, M., Babisch, W., Davis, A., Brink, M., Clark, C., Janssen, S., and Stansfeld, S. (**2014**). "Auditory and non-auditory effects of noise on health," The Lancet **383**, 1325-1332.
Berglund, B., Lindvall, T., Schwela, D. H., and Organization, W. H. (**1999**). "Guidelines for community noise."
Boucher, M. A., Krishnamurthy, S., Christian, A. W., and Rizzi, S. A. (**2023**). "Sound quality metric indicators of rotorcraft noise annoyance using multilevel analysis," The Journal of the Acoustical Society of America **153**, 867-876.
CAA (**2017**). "CAP 1506: Survey of Noise Attitudes 2014: Aircraft," edited by C. A. Authority.
CAA (**2023**). "The effects of eVTOL aircraft noise on humans (CAP2505)."
Cabell, R., Grosveld, F., and McSwain, R. (**2016**). "Measured noise from small unmanned aerial vehicles," in *Inter-Noise and Noise-Con Congress and Conference Proceedings* (Institute of Noise Control Engineering), pp. 345-354.
Callanan, J., Ghassemi, P., DiMartino, J., Dhameliya, M., Stocking, C., Nouh, M., and Chowdhury, S. (**2020**). "Ergonomic impact of multi-rotor unmanned aerial vehicle noise in warehouse environments," Journal of Intelligent & Robotic Systems **100**, 1309-1323.
Christian, A. W., and Cabell, R. (**2017**). "Initial investigation into the psychoacoustic properties of small unmanned aerial system noise," in *23rd AIAA/CEAS aeroacoustics conference*, p. 4051.
Clark, C., Crumpler, C., and Notley, A. H. (**2020**). "Evidence for Environmental Noise Effects on Health for the United Kingdom Policy Context: A Systematic Review of the Effects of Environmental Noise on Mental Health, Wellbeing, Quality of Life, Cancer, Dementia, Birth, Reproductive Outcomes, and Cognition," Int J Environ Res Public Health **17**.
Clark, C., and Paunovic, K. (**2018**). "WHO Environmental Noise Guidelines for the European Region: A Systematic Review on Environmental Noise and Quality of Life, Wellbeing and Mental Health," International Journal of Environmental Research and Public Health **15**, 2400.
EASA (**2022**). "Guidelines on Noise Measurement of Unmanned Aircraft Systems Lighter than 600 kg Operating in the Specific Category (Low and Medium Risk)."
EASA (**2023**). "Environmental Protection Technical Specifications applicable to eVTOL powered by multiple, vertical, non-tilting, evenly distributed rotors," edited by E. U. A. S. Agency.
Fields, J., and Walker, J. (**1982**). "Comparing the relationships between noise level and annoyance in different surveys: A railway noise vs. aircraft and road traffic comparison," Journal of Sound and Vibration **81**, 51-80.
Filippone, A. (**2014**). "Aircraft noise prediction," Progress in Aerospace Sciences **68**, 27-63.
Foraster, M., Eze, I. C., Schaffner, E., Vienneau, D., Heritier, H., Endes, S., Rudzik, F., Thiesse, L., Pieren, R., Schindler, C., Schmidt-Trucksass, A., Brink, M., Cajochen, C., Marc Wunderli, J., Roosli, M., and Probst-Hensch, N. (**2017**). "Exposure to Road, Railway, and Aircraft Noise and Arterial Stiffness in the SAPALDIA Study: Annual Average Noise Levels and Temporal Noise Characteristics," Environ Health Perspect **125**, 097004.
Gallo, E., Beaulieu, G., and Schram, C. F. (**2022**). "Annoyance factors of a maneuvering multicopter drone," in *28th AIAA/CEAS Aeroacoustics 2022 Conference*, p. 2837.
Gjestland, T. (**2008**). "Background noise levels in Europe."

Gwak, D. Y., Han, D., and Lee, S. (**2020a**). "Sound quality factors influencing annoyance from hovering UAV," Journal of Sound and Vibration, 115651.

Hardoy, M. C., Carta, M. G., Marci, A. R., Carbone, F., Cadeddu, M., Kovess, V., Dell'Osso, L., and Carpiniello, B. (**2005**). "Exposure to aircraft noise and risk of psychiatric disorders: the Elmas survey: Aircraft noise and psychiatric disorders," Social Psychiatry and Psychiatric Epidemiology **40**, 24-26.

Ho, J., Tumkaya, T., Aryal, S., Choi, H., and Claridge-Chang, A. (**2019**). "Moving beyond P values: data analysis with estimation graphics," Nature methods **16**, 565-566.

Hox, J. J., Moerbeek, M., and Van de Schoot, R. (**2017**). *Multilevel analysis: Techniques and applications* (Routledge).

Hui, C. T. J., Kingan, M. J., Hioka, Y., Schmid, G., Dodd, G., Dirks, K. N., Edlin, S., Mascarenhas, S., and Shim, Y. M. (**2021**). "Quantification of the Psychoacoustic Effect of Noise from Small Unmanned Aerial Vehicles," Int J Environ Res Public Health **18**.

Hurtley, C. (**2009**). *Night noise guidelines for Europe* (WHO Regional Office Europe).

Kempen, E. V., Casas, M., Pershagen, G., and Foraster, M. (**2018**). "WHO Environmental Noise Guidelines for the European Region: A Systematic Review on Environmental Noise and Cardiovascular and Metabolic Effects: A Summary," Int J Environ Res Public Health **15**.

Lakens, D. (**2022**). "Sample size justification," Collabra: Psychology **8**, 33267.

Locher, B., Piquerez, A., Habermacher, M., Ragettli, M., Roosli, M., Brink, M., Cajochen, C., Vienneau, D., Foraster, M., Muller, U., and Wunderli, J. M. (**2018**). "Differences between Outdoor and Indoor Sound Levels for Open, Tilted, and Closed Windows," Int J Environ Res Public Health **15**.

Nicholls, R. K. T., Antonio J. (**2021**). "An investigation into human response to unmanned aerial vehicle," in *Inter Noise* (Washington).

Ramos-Romero, C., Green, N., Roberts, S., Clark, C., and Torija, A. J. (**2022**). "Requirements for Drone Operations to Minimise Community Noise Impact," Int J Environ Res Public Health **19**.

Read, D. R., Senzig, D. A., Cutler, C. J., Elmore, E., and He, H. (**2020**). "Noise Measurement Report: Unconventional Aircraft-Choctaw Nation of Oklahoma: July 2019," (John A. Volpe National Transportation Systems Center (US)).

Schäffer, B., Pieren, R., Heutschi, K., Wunderli, J. M., and Becker, S. (**2021**). "Drone Noise Emission Characteristics and Noise Effects on Humans—A Systematic Review," International Journal of Environmental Research and Public Health **18**, 5940.

Schreckenberg, D., Schuemer-Kohrs, A., Schuemer, R., Griefahn, B., and Moehler, U. (**1999**). "An interdisciplinary study on railway and road traffic noise: Annoyance differences," in *Proceedings of Forum Acusticum*.

Sottek, R. (**1993**). "Modelle zur Signalverarbeitung im menschlichen Gehr (Models for signal processing in the human auditory system)," (Dissertation, TU Aachen).

Sottek, R. (**2014**). "Improvements in calculating the loudness of time varying sounds," in *INTER-NOISE and NOISE-CON Congress and Conference Proceedings* (Institute of Noise Control Engineering), pp. 3306-3312.

Sottek, R. (**2016**). "A hearing model approach to time-varying loudness," Acta Acustica united with Acustica **102**, 725-744.

Stansfeld, S., Haines, M., Burr, M., Berry, B., and Lercher, P. (**2000**). "A review of environmental noise and mental health," Noise & Health **2**, 1-8.

Torija, A. J., Chaitanya, P., and Li, Z. (**2021**). "Psychoacoustic analysis of contra-rotating propeller noise for unmanned aerial vehicles," J Acoust Soc Am **149**, 835.

Torija, A. J., and Clark, C. (**2021**). "A Psychoacoustic Approach to Building Knowledge about Human Response to Noise of Unmanned Aerial Vehicles," Int J Environ Res Public Health **18**.

Torija, A. J., and Li, Z. (**2020**). "Metrics for assessing the perception of drone noise," in *e-Forum Acusticum 2020* (European Acoustics Association (EAA), Lyon, France).

Torija, A. J., Li, Z., and Self, R. H. (**2020b**). "Effects of a hovering unmanned aerial vehicle on urban soundscapes perception," Transportation Research Part D: Transport and Environment **78**.

Torija, A. J., and Nicholls, R. K. (**2022**). "Investigation of Metrics for Assessing Human Response to Drone Noise," Int J Environ Res Public Health **19**.

Torija, A. J., Self, R. H., and Lawrence, J. L. (**2019**). "Psychoacoustic characterisation of a small fixed-pitch quadcopter," in *Inter-Noise and Noise-Con Congress and Conference Proceedings* (Institute of Noise Control Engineering), pp. 1884-1894.

Völk, F. (**2016**). "Predicting the Loudness of Non-Stationary Sounds: Zwicker's Original Envelope Extraction vs. DIN 45631/A1: 2010," in *INTER-NOISE and NOISE-CON Congress and Conference Proceedings* (Institute of Noise Control Engineering), pp. 6383-6389.

Waters-Fuller, T., Lurcock, D., Mackenzie, R., and MacKenzie, R. (**2007**). "NANR116 Open-Closed Window Research Report."

Yung, H. Y. (**2000**). "Rotor blade–vortex interaction noise," Progress in Aerospace Sciences **36**, 97-115.

Zwicker, E., and Fastl, H. (**2013**). *Psychoacoustics: Facts and models* (Springer Science & Business Media).