Masters by Research Thesis

Presented to the General Office
Newton Building, University of Salford, England

This paper represents my own work. Any input or work done by other people is clearly noted and properly referenced.

An investigation into the impact of Unmanned Aerial Vehicles on soundscape perception in urban environments

Rory Kerr Nicholls

r.k.nicholls@edu.salford.ac.uk

Table of Contents

Contributing publications

The following papers have been published by myself and my peers during the extent of my Masters by Research course at the University of Salford, and contribute towards the research described in this thesis. Of particular note is the publication in the International Journal of Environmental Research and Public Health, with the submission being chosen as a cover paper for the journal.

1. Nicholls, R. & Torija, A. J., 2021. *An investigation into human response to unmanned aerial vehicle noise.* Proceedings of Internoise 2021. Washington, DC, Institute of Noise Control Engineering.

2. Torija, A. J. & Nicholls, R., 2022. Investigation of Metrics for Assessing Human Response to Drone Noise. *International Journal of Environmental Research and Public Health,* 19(6).

3. Nicholls, R., Torija, A. J., Green, N. & Romero, C. R., 2022. *Assessment of drone noise impact on urban soundscapes.* Proceedings of Internoise 2022. Glasgow, Institute of Noise Control Engineering.

1. INTRODUCTION

Unmanned aerial vehicles (UAV), or drones as they are more commonly termed, have scope to strengthen the infrastructure of urban and metropolitan areas through typical means such as delivery, maintenance, and surveillance. As supporting technology improves with the success of these vehicles, more impressive operations may become conventional, such as emergency response (first aid), firefighting, road traffic control and monitoring air pollution (Burchan, 2019). The implications of UAV integration into city operations are advantageous, however as these vehicles are introduced a need to investigate any potentially adverse health effects on the public in relation to their novel acoustic character becomes principal. When compared to ubiquitous forms of vehicle noise atypical of urban environments, such as road, rail and aircraft noise, UAV sound signatures are distinctive. This is due to the increased proportion of audible noise present in higher frequencies above 2 kHz (Torija, et al., 2019). Multi-rotors under real operations might have rotors operating at different rotational speeds to maintain vehicle stability, which leads to complex tonal content (i.e., multiple BPF harmonics interacting between each other and giving a feeling of 'roughness'). This is done by changing the thrust produced by each rotor to balance the vehicle, which results in audible fluctuations in the spectra of operating UAV introducing tonal and broadband characteristics (Cabell, et al., 2016). Due to these complex spectral and temporal characteristics, UAV can be dominant over a background soundscape environment, even at comparably low levels.

Previous research has investigated the capability of established sound quality metrics from similar noise sources to quantify the effects of UAV noise on communities. Conventional aircraft could be considered as the closest relating noise source to UAV, which is currently present in metropolitan soundscapes, and has an extremely developed catalogue of noise metrics which are used to assess the potential negative impact on public health from these vehicles. Research has shown that metrics such as loudness and sharpness are strong indicators of perceived annoyance, proving that a higher ratio of high to low frequency content indicate a higher perceived annoyance (Soeta & Kagawa, 2020). Similarly, conventional rotorcraft noise sources such as helicopters have been investigated by controlling for loudness, and using synthesised helicopter noise which vary in sound quality metric attributes. It was found that sharpness, tonality and fluctuation strength were key predictors for perceived annoyance when used in multilevel regression analysis (Boucher, et al., 2020).

Therefore, metrics that interrogated spectral content more robustly than a typical metric based on broadband SPL (sound pressure level) were developed, such as EPNL (effective perceived noise level), for conventional aircraft. The EPNL metric was used to better account for not only the perception of the loudness of aircraft, but also tonal influence and the duration of noise exposure. The effectiveness of EPNL to account for tonal content was investigated and compared to metrics from developed by the automotive industry, such as sharpness and tonality. It was found that EPNL was not refined enough to quantify the complex spectral characteristics of conventional aircraft, due to the metric only considering the most prominent tonal artefact of a signal. It was found however, that paring sharpness and tonality with a solely broadband level-based metric (PNL, perceived noise level), increased the accuracy of a linear model for predicting subjective values of preference towards several conventional aircraft and their associated engines (Torija, et al., 2019).

To decide upon the appropriate sound quality metrics to be used when assessing the potential impact on communities introduced by UAV noise, it is paramount to understand the key operational differences between this new source and its conventional counterparts. One key difference is the change in operational distance of UAV when compared to conventional aircraft. Due to the nature of potential UAV operations, i.e., delivery, UAV will be at much closer ranges than commercial aircraft which typically fly at 400 ft at airport boundaries. This means that any effects that air absorption has on conventional aircraft will be drastically reduced for UAV which, compounding with an already high-frequency heavy spectral characteristic, leads to significant increases in perceived annoyance when assessed in comparison with conventional aircraft (Torija & Clark, 2021).

Initial research has also begun to show how the presence of background noise may help to reduce the negative perception of UAV noise in urban soundscapes. UAV stimuli were introduced into a number of recorded urban audio-visual soundscapes, and participants were asked to rank the soundscapes in order of preference. It was found

that when a UAV stimulus was introduced into a roadside soundscape, the perceived annoyance increased by 1.3 times. However, when a UAV was introduced into a quieter city park soundscape, the perceived annoyance increased by up to 7 times (Torija, et al., 2020). This research, however, only quantified the soundscape and UAV stimuli using a broadband L_{Aeq} value, and did not investigate the control of spectral and temporal characteristics on subjective responses.

The research presented in this thesis aims to comprehensively investigate the spectral and temporal characteristics of UAV noise using discrete sound quality metrics which have the potential to surpass previously adopted broadband and limited frequency-discriminate metrics used in neighbouring industries such as conventional aircraft. This research uses statistical analysis techniques such as multilevel linear analysis and multidimensional scaling to not only evaluate the correlations between these sound quality metrics and subjective response values like perceived annoyance, but to also identify the key critical bands of these sound quality metrics which are contributing to perceived annoyance, and how they may be being produced by the mechanical processes in UAV flight and operation. The effects of noise masking upon UAV introduction into a soundscape is assessed in order to understand the controlling frequency ranges of UAV noise over perception, and which soundscape environments may be best suited to introduce UAV operations. The main objectives of this research are summarised as follows:

- To understand the key spectral and temporal characteristics of isolated UAV noise that influence human perception
- To quantify the effects of masking UAV noise with urban soundscape environments
- To gather fundamental statistical information on the spectral and temporal noise characteristics of UAV, for the development of assessment metrics relating human response to UAV noise

Following from this introduction is a literature review, which gives more detail on the previously discussed research. Then, metrics used for perception-based engineering in industries such as the commercial aircraft and the automotive industry are given and explained. After this, the thesis splits into two parts, including a methodology, analysis and results section for each. These two sections give detail on two separate experiments run to gather data for this thesis. The first half looks in-depth at UAV stimuli and the spectral characteristics they have, using sound quality metrics, a subjective listening experiment, and multilinear regression to give insight into the key sound quality metrics that control for perceived annoyance, loudness and pitch. The second half of the research involved introducing UAV stimuli into urban soundscapes, and gathering subjective responses of perceived annoyance, loudness, UAV dominance and soundscape pleasantness. The response values for this experiment were regressed against the results of a multidimensional scaling analysis of calculated specific sound quality metric values for each UAV stimuli. Each part of the research is subdivided into a methodology, analysis, and results section. Following both stages of the research will be a conclusion, which summaries the key results of the research, with potential future research to end.

2. LITERATURE REVIEW

2.1. Human response to conventional aircraft noise

2.1.1. Three-Dimensional Psychological Evaluation of Aircraft Noise and Prediction by Physical Parameters (Soeta & Kagawa, 2020)

Although both large aircraft, such as commercial passenger flights, and drones are aerial vehicles, the methods of propulsion and elevation used differ, and hence typical spectral, temporal, and operational characteristics of noise from either method of aviation differ. This does not, however, mean that strategies to evaluate the psychoacoustic impact of noise emission cannot be transferred, at least partially, from one aerial vehicle to another.

Aircraft noise measurements were taken of take-off and landing routines at roughly 100 m from the runway to the measurement position. Binaural microphones (B&K type 4101) were used to capture the aircraft noise, and was saved to a computer for acoustical analysis. Sound quality metrics were used to quantify the captured aircraft recordings, and included loudness, sharpness, roughness, and fluctuation strength.

From the recorded aircraft noise measurements, 16 were chosen with a mean L_{Aeq} of 80 dB were used as test stimuli in a subjective testing stage. Of the 16, 6 were measured under a take-off routine, and the remaining 10 were measured under a landing routine. Each stimulus was presented to the test participants binaurally, through a pair of Sennheiser HD800 headphones and a Sennheiser HDVD800 power amplifier. The stimuli were 2 seconds long and each stimulus was presented in a pair with each other stimulus for comparison testing.

Three different subjective aspects of the sound were used to build the scaled response system for the testing procedure, annoyance, loudness, and pitch. For each pair, the participant was asked to rate the annoyance, loudness and pitch on a 7-point scale. For example, for pair A, consisting of sounds Y and Z, the participant made one of the following statements: I perceived Y as strongly louder than Z (3 points); I perceived Y as moderately louder than Z (2 points); I perceived Y as slightly louder than Z (1 point); I perceived loudness equally (0 point); I perceived Z as slightly louder than Y (-1 point); I perceived Z as moderately louder than Y (-2 points); and, I perceived Z as strongly louder than Y (-3 point). This 7-point scale was also similarly used for responses of annoyance and pitch.

Two regression models were built, using calculated values from the stimuli. The first was constructed using time-domain based factors, such as the recording's L_{Aeq} and delay time and amplitude of the first maximum peak, and the second model was constructed using frequency-domain based factors, such as the loudness and sharpness. Stepwise regression was used to identify the most influential factors on the participant's responses of annoyance, loudness and pitch. Also, it was observed that the stimuli recorded during a landing routine, rather than a take-off routine, were perceived as statistically more annoying, higher in loudness and in pitch by the participants.

It was found that the most significant factors in the model constructed from the time-domain variables were the L_{Aeq}, the width of the first decay of the auto-correlation function, $W_{\phi(0)}$, and the standard deviation of the L_{Aeq}. For the model constructed from the frequency-domain variables, the most significant factors were the loudness, sharpness, roughness, the standard deviation of the loudness, the standard deviation of the sharpness, and the standard deviation of the fluctuation strength.

The results show that the loudness and L_{Aeq} of the aircraft routine and the variation in sound level (standard deviation of L_{Aeq}) were significant factors in annoyance prediction. Furthermore, the sharpness and the spectral content are also significant, where a higher sharpness, or larger proportion of high frequency to low frequency content, would indicate a higher perceived annoyance. These factors should be investigated in the context of drone noise emissions.

2.1.2. Sound Quality Metric Indicators of Rotorcraft Noise Annoyance using Multilevel Regression Analysis (Boucher, et al., 2020)

Other forms of aviation, aside from large aircraft such as planes, also can be used to hypothesise sound quality metrics which are likely to take precedence when modelling for perceived preferability and annoyance of small to mid-sized UAVs. Aviation techniques which are more akin to the drones include rotorcraft such as helicopters; and research around these modes of transport give insight into suitable metrics for quantifying the spectral and temporal characteristics of drone noise.

The research aimed to develop "The Rotorcraft Sound Quality Metric" (RoQM) psychoacoustic testing procedure, which incorporated multilevel regression analysis to correlate key acoustical factors of rotorcraft noise with perceived annoyance. First, before subjective testing could begin, helicopter noise stimuli were synthesized to remove the change in sound quality during a physical flyover, and to create stimuli that are more controlled in the variation of the sound quality metrics. A total of 105 synthesized, 5-second-long stimuli were created, with varying values of roughness, sharpness, fluctuation strength, tonality and impulsiveness, and were normalised for loudness.

40 test subjects listened to the stimuli pool and gave responses for perceived annoyance on a 10-point scale; not at all annoying, slightly annoying, moderately annoying, very annoying and extremely annoying, and were free to choose in-between these points as well. The responses were then used in multilevel regression analysis, which combines complete pooling of the response data and no pooling of the response data. Regression without pooling evaluates each subjects' responses individually, so subject-specific regression parameters that are not influenced by the whole subject response pool can be derived. Complete pooling evaluates the responses as an aggregated dataset, so the derived regression parameters apply to the whole subject pool.

Three models were built for predicting perceived annoyance, in order to assess the efficiency of each model comparatively. They used individual subject datasets to calculate intercept values that varied per subject. The first of these models used no sound quality metric predictors, so had a slope of 0. The second had a fixed slope metric which did not vary between subjects, and the final model had slope metrics that varied between the subjects.

Since the loudness of each test stimuli was normalised, the effects of the other sound quality metrics on the predicted annoyance from the regression models could be investigated more clearly. Sharpness, tonality, and fluctuation strength were found to be significant factors for predicting annoyance, having consistent slope values across all models. 2nd-order terms were also investigated, where the interactions between the sound quality metrics were analysed. It was found that interactions between some of the sound quality metrics were statistically significant, but due to the impracticality of interpreting what the effect of changing the product of two sound quality metrics would have on the annoyance, 1st-order sound quality metric terms are used for the regression slopes.

2.1.3. On the Assessment of Subjective Response to Tonal Content of Contemporary Aircraft Noise (Torija, et al., 2019)

For current modes of aerial transport, such as conventional passenger aircraft, metrics have been developed to objectify the impact of air traffic noise on communities close to aircraft operation. One metric includes the Effective Perceived Noise Level, or EPNL, which was developed by the Federal Aviation Administration (FAA) to include the impact of the Perceived Noise Level (PNL), duration effects, and inclusion of the strongest protruding tone of conventional aircraft noise. Torija et al investigated whether the EPNL Tone Correction was sufficient at quantifying the impact of the tonal content of twin-engine aircraft on perceived annoyance, and whether more modern metrics with more refined frequency resolution could be used to develop a more sufficient metric.

Two previously established metrics were considered to quantify the complex tonal and high frequency characteristics of aircraft noise. The first was Aures' tonality, which has been shown to appropriately quantify the presence of isolated tones, as well as harmonic series covering a large frequency band. The second metric was

Sharpness, which is a representation of the ratio between high frequency content and the whole frequency content of noise.

The research used a subjective testing methodology to gather response data to aircraft sound stimuli. The aircraft recorded for the stimuli used 6 different engine variants: CFM-56-5, GE90-76/85/92B, GE90-115B, LEAP/PW1127G, RB211-5, and TRENT1000. The aircraft stimuli were captured at a location roughly 900m away from the south runway at Heathrow Airport. All aircraft were recorded during take-off, using a class 1 Rion ML52 sound level meter with the microphone mounted on a 1 m diameter ground board, attached to the meter with a 5 m cable. The microphone was positioned in a flat, open area, and fitted with a hemi-spherical wind shield. A total of 48 stimuli were used, each being 4 s long, as well as 24 reference samples. Participants were presented sets of six stimuli, with a set comprising of 4 aircraft stimuli, and 2 reference samples. The participants would first rank the stimuli in order of preference, with the option to listen to the stimuli individually as many times as required. Then, once an order was decided, the participant would hear the stimuli in the order they had chosen consecutively. After this, the participant could then make final adjustments to the order of preference for the current set, then proceed to the next set of 6 stimuli.

The reference stimuli were included in each set of stimuli for each participant, to be used as a basis to build a relative preference magnitude scale between the stimuli. Using a double reference magnitude scale allowed for a more dynamic assignment of preference for each stimulus in relation to the reference stimuli. The reference samples 1 and 2 were given arbitrary magnitude values of 60 and 40, respectively. The other stimuli are given magnitude values based on the equal spacing of the stimuli in the participants ranking order, in relation to the reference stimuli magnitude values and their position in the ranking order. Figure 1 demonstrates this for a set of results from the listening experiment.

Sample Order	Assigned Magnitude
Reference 1	60.0
Sample A	56.0
Sample B	52.0
Sample C	48.0
Sample D	44.0
Reference 2	40.0
Magnitude Range	100%

Sample Order	Assigned Magnitude
Sample A	65.0
Reference 1	60.0
Sample B	55.0
Sample C	50.0
Sample D	45.0
Reference 2	40.0
Magnitude Range	125%

Sample Order	Assigned Magnitude
Sample A	66.7
Reference 1	60.0
Sample B	53.3
Sample C	46.7
Reference 2	40.0
Sample D	33.3
Magnitude Range	167%

Sample Order	Assigned Magnitude
Sample A	70
Reference 1	60
Sample B	50
Reference 2	40
Sample C	30
Sample D	20
Magnitude Range	250%

Sample Order	Assigned Magnitude
Sample A	100
Sample B	80
Reference 1	60
Reference 2	40
Sample C	20
Sample D	0
Magnitude Range	500%

Figure 1: Demonstration of preference magnitude scaling for the response values of a participant (Torija, et al., 2019) (reproduced with permission)

The magnitude-scaled preference ratings from the subjective experiment were used in regression models to assess the effectiveness of the predictors PNL, EPNL, sharpness and tonality. It was found that EPNL with a tonal correction was not suitable for explaining the effect of tonal content on the responses of preference rating given by the participants. EPNL improved the correlation of models with only PNL as independent and preference rating as the dependent variable, for engines that contain a physically dominant blade passing frequency (BPF) tone (LEAP/PW1127G), but failed to improve the correlation of models for engines with a more diverse tonal content. This is due to EPNL only discriminating against the most prominent tonal artefact in aircraft noise, and does not account for any subharmonics of that tone, or other complex tonal content. Furthermore, it was found that the one-third octave band resolution of EPNL led to some tonal content being unaccounted for in the correction calculation. Aures' tonality was found to improve the correlation of models including PNL for engines that contained more complex tonal content, as it could account for the effect of a physically dominant tone, such as a BPF tone, it's harmonics, and series of complex tones in the frequency spectra. Models including PNL and Aures' tonality improved correlation, when compared to models only containing PNL, for all engine variants, except the GE90-76/85/92B and TRENT1000. These engines were found to correlate strongly with sharpness, for the GE90-76/85/92B variant, and PNL for the TRENT1000 variant.

2.2. UAV noise characterisation

2.2.1. Ten Questions Concerning the use of Drones in Urban Environments (Watkins, et al., 2019)

Many questions are raised when envisioning the introduction of drone fleets into modern society. Areas of intrigue include what would they be used for? What would inhibit their adoption in urban environments and rural environments? How impactful would the perception of noise and other sensory interaction be on the health of the general population? How can they be regulated to ensure public approval and acceptance?

In the context of community noise impact, it is queried that the perception of the drone noise will depend on the mission, or reason for the drone's presence, and the person or people associated with the mission. Furthermore, factors relating to the context of the drone's presence may influence perceived annoyance, such as the distance of the drone from the observer, and the operational load, the time of day, and the individuals' sound sensitivity. Spectral and temporal factors of the noise have been shown to also explain differences in perceived annoyance in neighbouring industries, such as the automotive and aircraft industries, and should be investigated for this new sound source. It would therefore be beneficial if a combination of these factors could be correlated with responses of perceived annoyance from a group of test participants.

2.2.2. Psychoacoustic Characterisation of a Small Fixed-Pitch Quadcopter (Torija, et al., 2019)

Initial investigation into the psychoacoustic profile of typical UAV noise has been carried out in internal, anechoically controlled conditions, as well as in an external environment for comparison between fundamental UAV mechanical noise signatures and how they differ when put into more realistic operations. Mechanically, UAV employ propulsion techniques that are more intricate than other conventional aircraft, such as commercial planes or helicopters. A common configuration for sUAV is that of a quadcopter, where 4 rotor blades are mounted symmetrically upon electric motors, which are controlled by a central unit to balance the load to each electric motor, depending on the desired direction of flight. These vehicles are expected to be used for operations such as delivery (Yoo, et al., 2018), which could potentially affect a large percentage of urban communities, and hence the psychoacoustic characteristics of their noise profiles must be understood.

A database of audio files was collected of not only UAV, but also car pass-by events, motorbike pass-by events, and aircraft take-offs, for the sake of analysing differences in the spectral characteristics of each of these vehicles. The road vehicle samples were measured at the roadside, approximately 1.5 m away from the vehicles. The aircraft samples were measured at approximately 900 m from the end of the runway, with the planes passing overhead at heights of around 435 m. External UAV samples were recorded of straight-and-level flyovers at two heights above ground level (1 m and 2 m), two lateral distances between the microphone and the flight track (0 and 5m), and three

extra-payload conditions (0, 434 g and 656 g). Internal UAV samples were taken of the quadcopter fixed to a stand at 1.8 m above ground level. This meant that no distortion of the noise from the UAV was introduced through the UAV trying to counter-balance itself. The UAV was operated with each of its rotors operating and not operating, to assess how each component of the UAV's propulsion system effects the acoustic signature.

From the audio samples of the vehicles collected, 4 second clips were extracted which captured complete pass-by events of the road vehicles and the UAV, and aircraft clips were edited so the most prominent audio feature was presented. The audio samples were normalised to 65 dB L_{Aeq}, for comparison purposes. This meant that any differences between the acoustic signatures of these vehicles were attributable to the spectral and temporal characteristics of these noise sources. Sound quality metric values were calculated for each vehicle sample using HEAD ArtemiS Classic, including DIN 45631/A1 loudness, DIN 45692 sharpness, Aures' listening model for roughness and fluctuation strength, and tonality was calculated following publications by Terhardt and Aures (Terhardt, et al., 1982) (Aures, 1985).

The UAV measured internally in the anechoic chamber was put under 4 different operating conditions; only the electric motor working, the electric motor with 1 rotor operating, 2 rotors operating and 4 rotors operating. When just the electric motor was in use, high-frequency content was dominant in the spectrum, with tonal components being present between 2.4 kHz and 6 kHz. For the subsequent operating conditions, harmonics of the blade passing frequency (BPF) at low-to-mid frequency regions are close together. The 2-rotor operating condition showed more prominent BPF harmonics at higher frequencies, which increased for the 4-rotor operating condition. These results are shown in Figure 2. The spectrogram plots of the UAV measured externally illustrate the unsteadiness of the frequency spectra of UAV noise due to micro-adjustments to steady the UAV during flight, as well as to account for unbalances caused by wind (Cabell, et al., 2016). Figure 3 shows how, when compared to the steady spectra of the UAV which was measured internally and fixed in place, unsteady frequency content can be seen which is magnified at higher harmonics of the BPF.

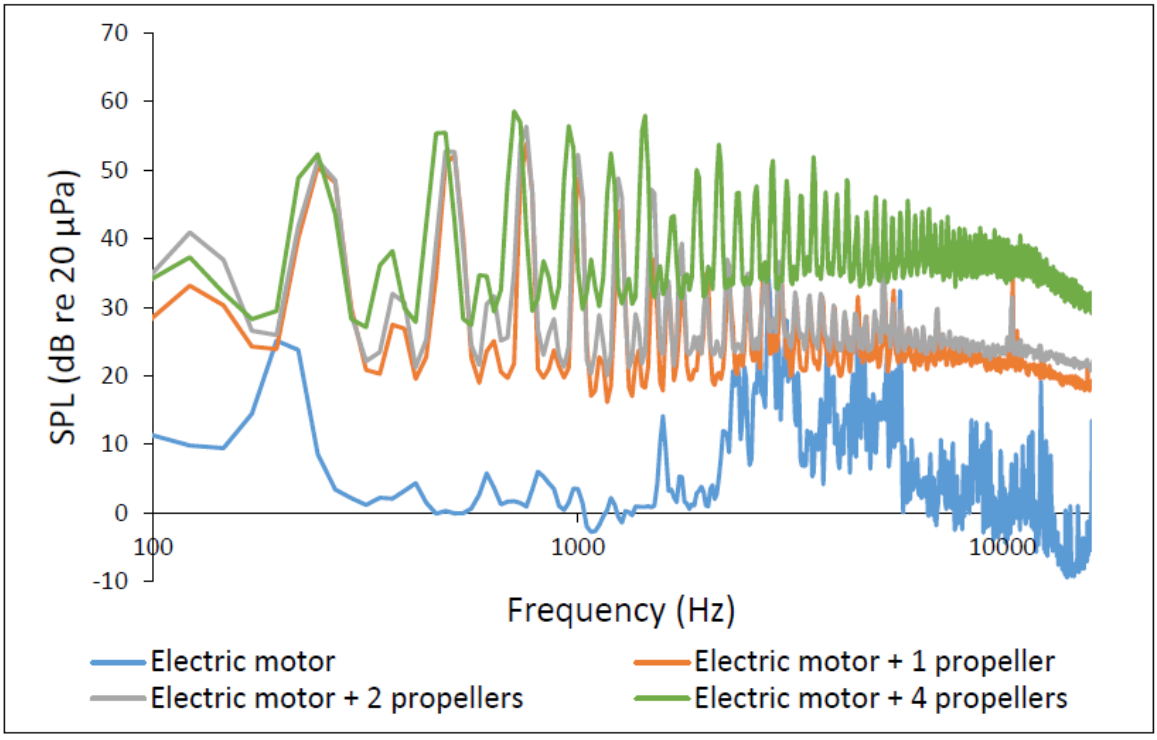

Figure 2: Frequency spectrum of the quadcopter tested; electric motor only, motor plus 1, 2, and 4 rotors (Torija, et al., 2019) (reproduced with permission)

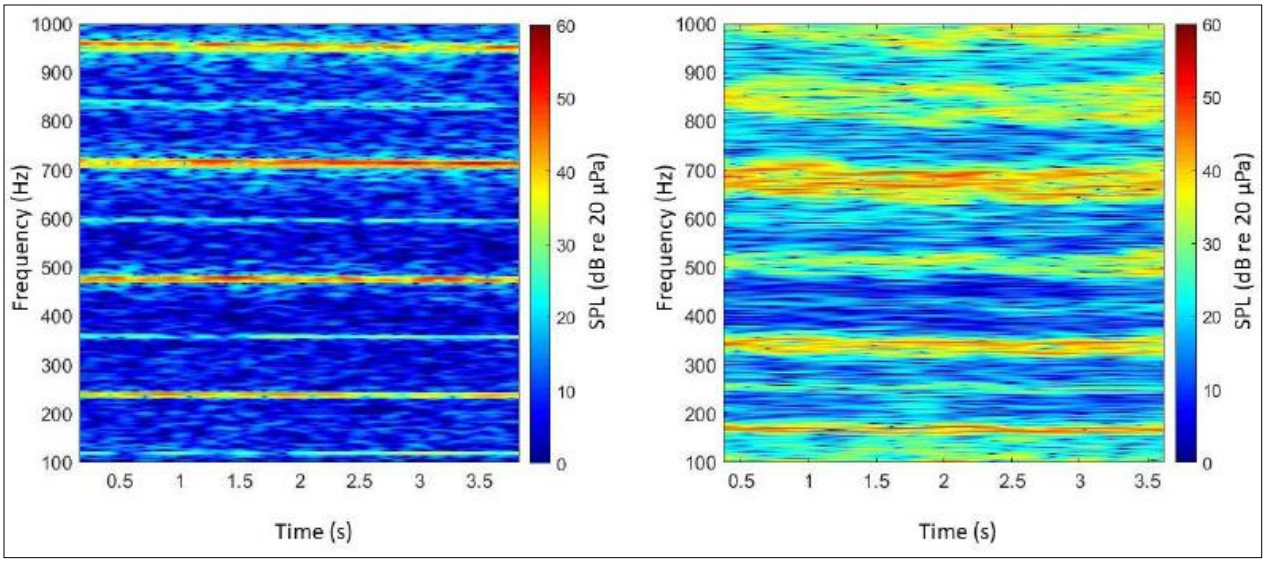

Figure 3: Spectrogram of the quadcopter measured in the anechoic chamber (left), and externally while hovering (Torija, et al., 2019) (reproduced with permission)

Furthermore, when analysing the frequency spectra of the external UAV flyover operations with varying payload complex tones up around 1.6 kHz are present, as well as significant high-frequency content introduced between 3.8 kHz and 6 kHz. As the payload increased from 0 g to 434 g, the required power to keep the UAV elevated increased, which in turn led to a higher pitch noise being emitted from the vehicle. This was caused by the electric motors being driven at higher power to maintain stability.

The analysis of the other vehicles measured, aircraft and road vehicles, revealed interesting results. With all the vehicle samples normalised to 65 dB L_{Aeq}, it was found that the quadcopter and the road vehicles achieved a similar value of average loudness, with the value of the aircraft being significantly lower. This was caused by the increased ratio of high-frequency content in their spectra, as the aircraft spectra are shaped by the effects of air absorption. Furthermore, the average values of tonality for the quadcopter and the aircraft were found to be higher than the value for road vehicles. For sharpness, it was found that the average value for the quadcopters was significantly higher than that of the road vehicles, and roughly double that of the aircraft, as shown in Table 1. Figure 4 shows how the frequency spectra of a typical aircraft significantly decrease at higher frequencies, compared to the quadcopter and road vehicles.

Table 1: Average sound quality metric values for aircraft, road vehicles and quadcopters tested (Torija, et al., 2019) (modified from source with permission)

	Aircraft	Quadcopter	Road Vehicles
Loudness (sone)	20.02 ± 2.47	27.21 ± 3.43	26.19 ± 3.12
Sharpness (acum)	1.69 ± 0.34	3.57 ± 0.24	2.64 ± 0.39
Roughness (asper)	2.31 ± 0.39	2.19 ± 0.19	2.79 ± 0.78
Fluctuation Strength (vacil)	0.11 ± 0.02	0.08 ± 0.01	0.09 ± 0.01
Tonality (tu)	0.32 ± 0.07	0.36 ± 0.07	0.24 ± 0.15

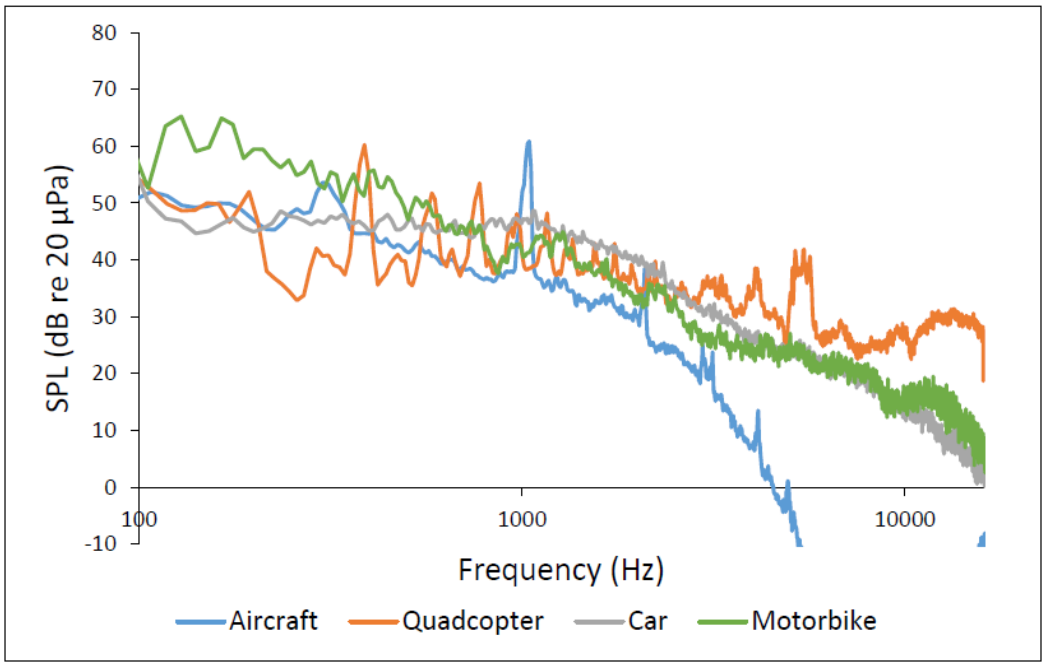

Figure 4: Frequency spectrum of examples of an aircraft, quadcopter, and road vehicle (Torija, et al., 2019) (reproduced with permission)

The modified Psychoacoustic Annoyance model (PA'), developed by Di et al (Di, et al., 2016) from Zwicker's model (Zwicker & Fastl, 2007), was used to calculate PA' values from the calculated sound quality metric values of the vehicle samples. It was found that the calculated PA' value for the quadcopters was 58.53±8.69, for road vehicles it was calculated to be 56.09±11.10, and the aircraft had a value of 41.11±6.53. Zwicker's model was also adjusted by More to create a value of PA_{mod} to account for the specific characteristics of aircraft noise, specifically the tonality effects (More, 2011). This model found the quadcopter to have a PA_{mod} value of 99.06±20.85, the road vehicles had a value of 68.69±14.17, and the aircraft had a value of 42.51±7.80. In both PA model versions, it was found that the quadcopter samples had significantly higher PA values when compared to aircraft noise, and higher values of PA when compared to road vehicles.

This research found that typical UAV, specifically those in quadcopter configuration, had statistically higher values of PA when compared with typical road vehicles and aircraft. This was due to the significant increase in high-frequency content present in the spectra of UAV when compared to aircraft, as well as complex tonal artefacts introduced by rotor interactions. The values of PA, however, were calculated using models not specifically tailored to UAV noise, rather to conventional aircraft. A progression from this would be to perform a psychoacoustic experiment using UAV sound stimuli to develop a PA value, like the regression methods employed in this thesis.

2.2.3. A Psychoacoustic Approach to Building Knowledge about Human Response to Noise of Unmanned Aerial Vehicles (Torija & Clark, 2021)

To predict public response to UAV operational noise, the typical noise characteristics of these vehicles must be explored and understood. Previous research has suggested that UAV may be quieter than conventional aircraft, but due to the significant decrease in the distance of operation between the noise source and communities, UAV could still imply a serious adverse effect on the public. Typically, conventional aircraft such as commercial flight vehicles

operate at heights of round 400 ft from airport boundaries, whereas UAV operations will typically be much closer, including delivery, building maintenance and blue light services.

A comparison was made between 2 UAV models, and 2 conventional aircraft models. The UAV assessed were the DJI M200 and the Yuneec Typhoon, and the 2 conventional aircraft were the Boeing 737 MAX 8 and the Airbus A320. Measurements of the conventional aircraft were made approximately 900 m away from the end of the south runway of Heathrow Airport during take-off operations, with the height of the aircraft during measurement being estimated as 435.3 ± 57.4 m from the measurement point. The UAVs were recorded during flyovers at a level distance of 45.7 m from the measurement position. Figure 5 presents the recorded frequency spectra of these operations. It was found that the UAV models' noise spectra included much more prominent high-frequency content when compared to the conventional aircraft models. This difference is predominantly due to the effects of air absorption over the large distances between the operating aircraft and the measurement positions, or ground level, as well as the fact that UAV generate higher frequency noise from smaller, high rpm rotor blades and motors (Gwak, et al., 2020). The rotor blades create a self-noise phenomenon, which includes turbulent boundary layer trailing edge noise (Alexander & Whelchel, 2019), the interactions between adjacent rotors (Torija, et al., 2019), and the interactions between armature and magnets of the electric motors (Cabell, et al., 2016).

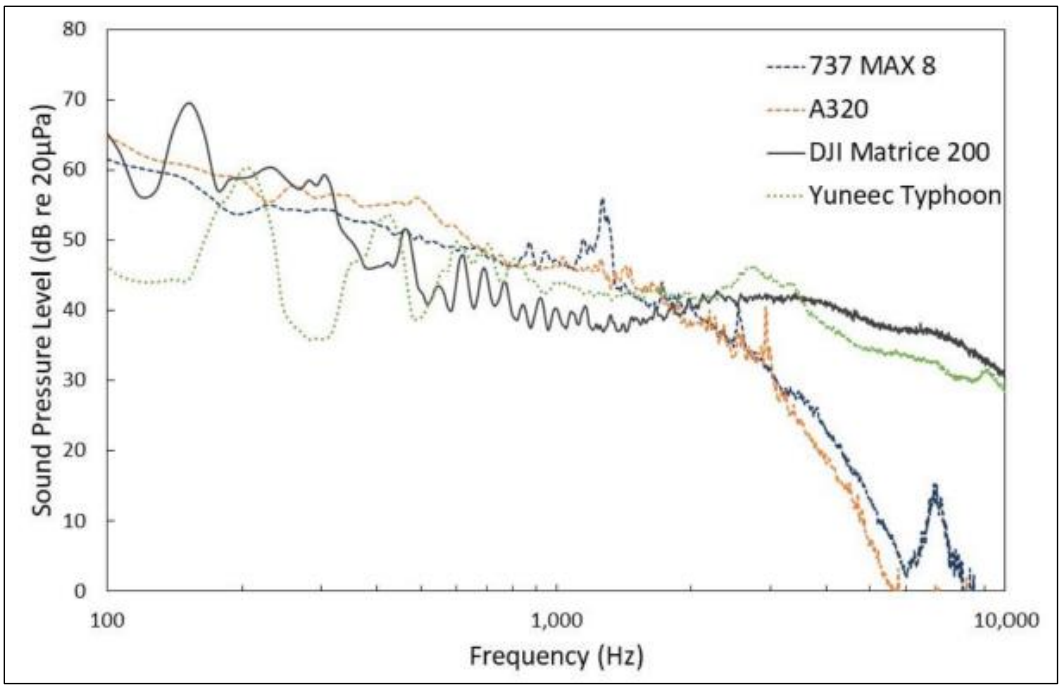

Figure 5: Frequency spectra from commercial aircraft and UAV operations (Torija & Clark, 2021) (reproduced with permission)

Aside from the high-frequency characteristics of the UAV, Figures 6 and 7 show how tonal the frequency content of UAV can be, with the frequency of harmonics being dependent on the blade passing frequency (BPF) of the UAVs rotors. Figure 6 shows that, at further distances, the tonal prominence of the UAV noise is less obvious, whereas at closer distances, as illustrated in Figure 7, the BPFs and harmonics are clearly discernible. In Figure 7, the two bright lines in the spectrogram at about 120 and 140 Hz are the BPFs for the two sets of rotors of the DJI M200 (left), whereas the Yuneec Typhoon (right) has its BPFs clustered around 200 Hz. What can also be seen in Figure 7, particularly in the spectrogram of the DJI M200 model, is that the tonal harmonics fluctuate slightly in the frequency domain over time. This is caused by the motors micro-adjusting to account for ambient weather conditions, such as wind gusts, while maintaining UAV stability. These fluctuations can influence the level, frequency, and temporal characteristics of UAV noise (Alexander, et al., 2019). Therefore, it can be said that the noise characteristics of

UAV noise does not resemble the noise characteristics of conventional aircraft (Christian & Cabell, 2017). Furthermore, it has been found that even if the L_{Aeq} is controlled (to 65 dBA) for both a UAV noise source and a conventional aircraft noise source, the preference rating given during a subjective assessment to a UAV stimuli was 33% lower than the rating given to a conventional aircraft (Torija & Li, 2020).

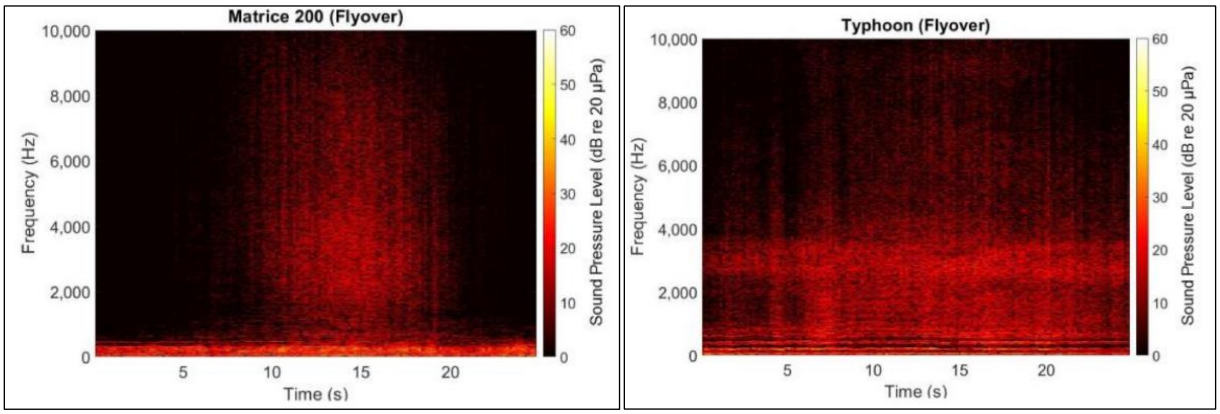

Figure 6: Spectrogram of UAV flyovers, measured at approx. 50 m (Torija & Clark, 2021)

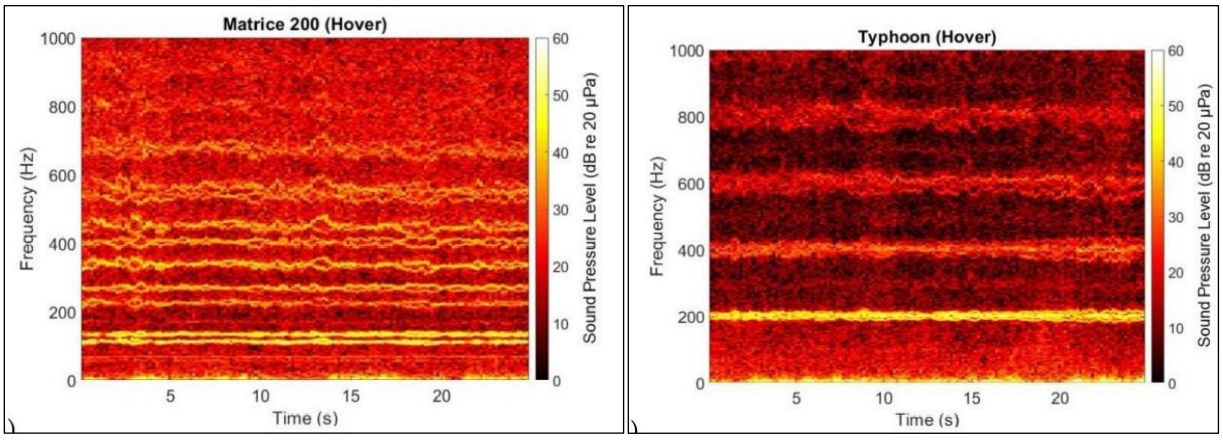

Figure 7: Spectrogram of UAVs hovering, measured at approx. 1.2 m (Torija & Clark, 2021) (reproduced with permission)

2.3. UAV noise introduced into urban soundscapes

2.3.1. Effects of a Hovering Unmanned Aerial Vehicle on Urban Soundscapes Perception (Torija, et al., 2020)

Previous research has investigated the perceived annoyance of unmanned aerial vehicles (UAVs), or drones, when introduced to different populated environments. This research was undergone to understand the impact on the general public's health as the use of UAVs for roles such as delivery increases.

Firstly, audio-visual recordings of two areas were captured. The first area was the Southampton Common Park and contained 4 individual measuring positions at varying distances from a busy road. The second area was a park located in the city centre of Southampton, containing 3 measurement positions at varying distances from a busy road junction. 1st order A-format ambisonic audio was recorded using 4 Micro Electrical-Mechanical System (MEMS) microphones, integrated into a panoramic camera which captured the visual stimuli.

Secondly, a small, quadcopter drone (DJI Phantom 3 standard) was used to gather audio-visual signals. The drone was recorded inside an anechoic chamber, with the drone being fixed in position at an unvarying elevation, so that all the quadcopters' rotors could move freely. A greenscreen with a high acoustic permeability was set up behind the drone so that the previously recorded environmental stimuli could be used as a background to virtually place the recorded drone on. The audio-visual recordings of the drone were captured using the same equipment as the environmental recordings.

Subjective listening tests were carried out by 30 healthy participants. The test stimuli included the audio-visual recordings of both the environmental recordings, and the environmental recordings with the drone audio-visual signals virtually introduced. Furthermore, these scenarios were presented without the visual recordings, leaving only audio playback of each environment without and with the introduction of a drone. The participants were asked to give responses for their perceptions of loudness, annoyance, and pleasantness for each test stimuli, using an 11-point scale (0 being not at all, and 10 being extremely).

The responses were analysed using multilevel linear modelling to evaluate the responses given for perceived loudness, annoyance, pleasantness, and to also observe any effect when playback included visual stimuli. The results showed that the introduction of drone noise in environments with a higher background level, mainly generated by traffic noise, the annoyance increased by around 1.3 times that of when the drone is not present. Where the environment had a lower level of background noise, the introduction of a drone increased the perceived annoyance by about 7 times that of when the drone is not present. It was also shown that using panoramic video stimuli yielded little effect on the responses for perceived loudness, but introduced large differences in the responses of annoyance and pleasantness than with solely audio stimuli, strengthening the previously suggested idea that using audio-visual stimuli aid in creating more accurate representations of the environments.

The responses for perceived annoyance, loudness and pleasantness were only evaluated against the L_{Aeq} of each of the test stimuli, which was found to not account for the noise features of drones, suggesting that the use of more sound quality metrics should be used to characterise drone test stimuli in future research.

2.3.2. Auditory Detection Probability of Propeller Noise in Hover Flight in Presence of Ambient Soundscape (Stalnov, et al., 2022)

Aside from investigating the spectral characteristics of UAV and how they affect perceived annoyance, other means of mitigation should be researched. In this paper, the probability that a UAV may be detected in a particular environment is explored, by considering the noise level of the background environment, the number of blades the UAV has, and the operational distance of the UAV from the observer. Since UAV operations are predicted to be typically carried out in metropolitan areas, where infrastructure is developed and populations are high, it is of value to try and understand how high ambient noise levels in urban soundscapes could potentially mitigate the adverse effects of UAV noise; although this research does also investigate the effects of rural soundscapes, for comparison.

This research was segmented into two parts. The first involved a two-step approach, to understand the probability of a UAV noise source being detected by an observer based on the noise level of the UAV compared to background noise in a forced choice experiment, and then to understand whether that UAV sample would be detected in an external environment where the listener is unaware of the experiment, in an unforced choice experiment methodology. The second part of the research models the auditory nerve system, to predict whether a UAV signal would be detected based on nerve fibre stimulation from the amplitude of the UAV signal.

The first step of the first part was carried out by presenting the experiment participants with two audio stimuli; the first included only the background environmental noise, and the second included the environmental and the UAV noise. The listener was then prompted to answer whether they heard the UAV stimuli present in the audio containing both UAV and background noise. Four potential outcomes are possible using this method; a UAV stimulus is present and the participant detects it (hit), a UAV stimuli is present and the participant does not detect it (miss), a UAV stimuli is not present and the participant believes there is (false alarm), and a UAV stimuli is not present and

the participant does not believe there is (correct rejection). From these outcomes, as well as other accepted probability assumptions, a hit probability function was derived, to predict whether a UAV signal would be detected if introduced into background environmental noise. The second part looked at quantifying the function of the auditory nerve systems within the ear, and used an arithmetic function derived to for the threshold of an acoustic stimulus to be detected by the auditory nerve system.

The UAV stimuli were recorded in an anechoic chamber, where a motor and propeller of a UAV were mounted in place in the middle of the chamber, at a height of 1 m above ground level. A total of 30 stimuli were recorded: 15 for a 2-blade propeller, at polar angles increasing in equal increments from 0^{o} to 105^{o}, as well as a 5-blade propeller. The background noise stimuli were measured in both rural and urban areas. The microphone for these recordings was positioned at 1.2 m above ground level. It was found that the urban environment had a typically increased level at regular intervals in the frequency spectra when compared to the rural environment. The rural environment exhibited a hump in the spectra at around 1 kHz, and the urban environment has a sharp increase in level around 6 kHz.

It was found that as operational distance increased, the threshold level for UAV detection in both rural and urban environments increased, however the threshold level in rural environments was consistently lower. Furthermore, for operational distances up to around 700 m, it was found that the difference between the threshold level for urban and rural environments was not statistically significant. However, at distances greater than around 700 m, it was observed that the detection threshold for the 5-blade propeller noise was higher than that of the 2-blade propeller noise, meaning that the probability to detect a 5-blade propeller-based vehicle was consistently lower than a 2-blade counterpart. It was found that UAV stimuli in rural areas have a higher probability of being detected than UAV stimuli in urban areas.

However, at smaller operational distances, below around 100 m, the calculated probability of detection for both UAV propeller configurations in each environment is very high, above 0.75. From the results of this research, and considering the expected proximity of UAV operations such as delivery to the public, it is clear that operations within these distances must be investigated to assess the feasibility of UAVs. Furthermore, a more detailed analysis into the spectral contributions of UAV to a soundscape environment should be undertaken, to ascertain whether key characteristics of UAV stimuli are the main contributors to a lower estimated threshold of detection value, and consequently an increased estimated detection probability value. Finally, this research draws attention to the fact that the complexity of urban soundscapes lends itself to being a more appropriate noise masking tool for UAV noise when compared with rural soundscape environments, supporting the case for research into urban soundscape masking.

3. CONVENTIONAL AIRCRAFT NOISE METRICS

The increase in use of aircraft for transportation of passengers and goods, and the subsequent exposure of these noise sources to the public, has led to the development of noise metrics which give an objective way to assess the potential impact of aircraft operational noise on public health. The perceived noise level (PNL) and effective perceived noise level metric (EPNL) were developed to assess the impact of tonal aircraft noise on the public. The EPNL calculations was defined by the FAA (Federal Aviation Agency, 2017) and expands on the PNL method by including a tonal correction implemented in the calculation, via a 10-step algorithm. This algorithm was designed to detect and account for tonal content in one-third octave bands from 80 Hz to 10 kHz. The step involves calculating the SPL in the 80 Hz one-third octave band (critical-band rate of 3 Bark), and then calculating the change in SPL, s, with increasing one-third octave band centre frequency (Equations 1 through 4).

$$s(3,k) = no\ value \tag{1}$$

$$s(4,k) = SPL(4,k) - SPL(3,k) \tag{2}$$

...

$$s(i,k) = SPL(i,k) - SPL[(i-1),k] \tag{3}$$

…

$$s(24,k) = SPL(24,k) - SPL(23,k) \tag{4}$$

Where k is a time index which are measured in 0.5 s intervals during a flyover operation of the aircraft, and i is the critical-band rate. Step 2 is to encircle the value of the slope, $s(i,k)$, where the absolute value of the change in the slope is greater than 5, described in equation 5. Step 3 involves categorising the slope $s(i,k)$ for SPL evaluation. If the encircled value is positive and algebraically greater than the slope $s[(i-1),k]$, encircle $SPL(i,k)$. If the encircled value of the slope $s(i,k)$ is zero or negative and the slope $s[(i-1),k]$ is positive, encircle $SPL[(i-1),k]$. For all other cases of slope $s(i,k)$, no SPL value is to be encircled. Following from this, step 4 is to omit all SPL values encircled, and calculate new SPL values, $SPL'(i,k)$.

$$|\Delta s(i,k)| = |s(i,k) - s[(i-1),k]| > 5 \tag{5}$$

For non-encircled SPL values from step 3, $SPL'(i,k)$ is equal to $SPL(i,k)$. For encircled SPL values in critical-band rates 1 through 23, the new SPL value is equal to the arithmetic average of the preceding and following critical-band rate SPL values, as shown in equation 6. If the SPL value of the highest critical-band rate, 24 Bark, is encircled, then the new SPL value for that band is shown in Equation 7.

$$SPL'(i,k) = \frac{1}{2}\left[SPL[(i-1),k] + SPL[(i+1),k]\right] \tag{6}$$

$$SPL'(24,k) = SPL(23,k) + s(23,k) \tag{7}$$

Step 5 of the algorithm involves recomputing new slopes, $s'(i,k)$, for $1 \leq i \leq 25$, including an imaginary critical-band rate, 25 Bark. This process is like Equations 1 through 4, and is reiterated in Equations 8 through 12.

$$s'(3,k) = s'(4,k) \tag{8}$$

$$s'(4,k) = SPL'(4,k) - SPL'(3,k) \tag{9}$$

…

$$s'(i,k)SPL'(i,k) - SPL'[(i-1),k] \tag{10}$$

…

$$s'(24,k) = SPL'(24,k) - SPL'(23,k) \tag{11}$$

$$s'(25,k) = s'(24,k) \tag{12}$$

$$\bar{s}(i,k) = \frac{1}{3}\left[s'(i,k) + s'[(i+1),k] + s'[(i+2),k]\right] \tag{13}$$

Then, step 6 consists of computing the arithmetic average of the three adjacent slopes for critical-band rates 3 to 23, as in Equation 13. Step 7 involves computing the final adjustments to the one-third octave band SPL values, $SPL''(i,k)$, starting with critical-band rate 3 Bark to 24 Bark, as shown in Equations 14 through 17.

$$SPL''(3,k) = SPL(3,k) \tag{14}$$

$$SPL''(4,k) = SPL''(3,k) + \bar{s}(3,k) \tag{15}$$

…

$$SPL''(i,k) = SPL''[(i-1),k] + \bar{s}[(i-1),k] \tag{16}$$

…

$$SPL''(24,k) = SPL''(23,k) + \bar{s}(23,k) \tag{17}$$

$$F(i,k) = SPL(i,k) - SPL''(i,k) \tag{18}$$

Equation 18 represents step 8 of the algorithm, which is to calculate the differences, $F(i,k)$, between the original SPL value and the adjusted SPL values. Only values of $F(i,k)$ greater than zero are noted. Step 9 is to determine tonal correction factors from $F(i,k)$ for each of the 24 one-third octave bands, using Figure 8. Finally, step 10 is to designate the largest of the tonal correction factors, as determined in step 9, as $C(k)$. From this, and the perceived noise level (PNL) as calculated following Appendix B, paragraph B36.3 of FAR Part 36, the tone corrected perceived noise level (PNLT) can be calculated for each k^{th} interval, and subsequently the EPNL, as shown in Equations 19 and 20, where d is the time interval to the nearest 1 s during which $PNLT(k)$ is 10 dB of the maximum tone corrected perceived noise level (PNLTM).

$$PNLT(k) = PNL(k) + C(k) \tag{19}$$

$$EPNL = 10\log\left(\sum_{k=0}^{2d} 10^{\frac{PNLT(k)}{10}}\right) - 13 \tag{20}$$

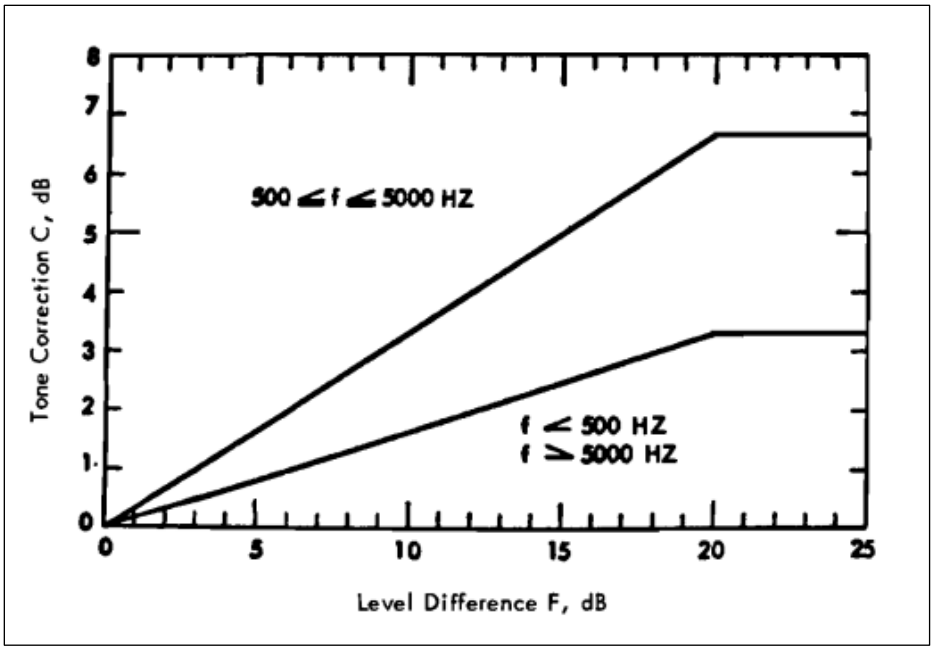

Figure 8: Tone correction as a function of level difference (Federal Aviation Agency, 2017)

For conventional rotorcraft such as helicopters, the sound exposure level (SEL) is often used as a metric for assessing the potential impact of noise from these vehicles on communities. The SEL of noise is the constant sound level that has the same amount of energy in one second as the original noise event. It is calculated as described by the FAA to assess helicopter noise (Federal Aviation Agency, 2017), to consider both helicopter flyover event duration and sound level. Equation 21 below shows the calculation of SEL, where L_{Aeq} is the A-weighted sound level measured for the total time the event elapsed, t.

$$SEL = L_{Aeq} + 10\,.Log_{10}(t) \tag{21}$$

Although these metrics are deemed suitable for the assessment of conventional aircraft, their application to noise from UAV has been deemed inappropriate (Torija, et al., 2019). This may be due to the operational characteristics of UAV, which infer an observable increase in mid- to high-frequency content with much more rapid fluctuations in tonal artifacts when compared to conventional aircraft noise signatures. Therefore, the development of metrics that quantify UAV noise specifically is crucial to the adoption of UAV operations in public areas, with this development being guided by the influence of UAV noise on human perception.

4. PERCEPTION-INFLUENCED ENGINEERING

The effects of transportation noise on public health have thus far been assessed predominantly through methods that scrutinise sound pressure and power levels from noise sources. Perception-influenced engineering aims to improve product design through existing knowledge of psychoacoustics, human physiology, and psychological response, which benefits from the use of more intricate sound quality metrics that quantify the functions of the human auditory system. It has been suggested that sound levels do not necessarily account for these intricacies (Greenwood, et al., 2022). An example of this can be seen in helicopter noise nuisance assessment, where current standing values used to quantify annoyance, such as the sound exposure level (SEL), have been seen to have some deficiency in predicting community response, and a level penalty of 4 dB must be applied (Boucher, et al., 2019). For helicopter noise, it has been suggested by recent research that a method incorporating several sound quality metrics should be used, to better evaluate the impact of helicopter noise on communities, and a similar approach could be beneficial

for the assessment of UAV community noise response. Similarly for aircraft noise, previous research has explored the impact of synthesised aircraft noise tonality on the annoyance response from participants in a subjective experiment, suggesting that sound quality metrics including a value of tonality should be used in the prediction of annoyance from actual aircraft noise (More & Davies, 2010). Sound quality metrics have been used to assess consumer product noise response in various industries, and have had significant development within the automotive industry, to quantify the effects of various frequency-based and temporal acoustic phenomena.

4.1. Sound quality metrics

Sound quality metrics have been developed to quantify acoustic phenomena perceived by the human ear. They are typically used in product design to mitigate perceived annoyance from noise sources ranging from washing machines to automotive vehicles (Davies, 2007). These metrics are particularly useful in regression analysis, to investigate the effects of acoustic characteristics of noise on human perception of the noise source. Standardised and industry accepted methods for calculating these metrics were implemented in this research to quantify the spectral and temporal characteristics of UAV noise.

4.1.1. Critical-band rates

The critical-band rate scale was derived to represent how the human ear discriminates between frequency components (Zwicker & Fastl, 2007). The critical-band rate scale, or Bark scale, converts the logarithmic frequency sensitivity of the ear into a linear range, representing frequencies from 0 to 15.5 kHz as 0 to 24 Bark on the critical-band rate scale. The relationship between frequency range and critical-band rate scale is illustrated in Table 2.

From the division of the ear frequency range into critical-band rates, the intensity of sound in these critical-bands for a given narrow-band noise can be considered. When a tone of a given frequency is incident on the basilar membrane, critical-bands adjacent to the critical-band related to that tone's frequency may also be excited. Figure 9 illustrates how the excitation level of the critical-band rates changes as a narrow-band noise passes up through the frequency range. This excitation function is used in sound quality metrics to identify the significance of frequency content, and is equal to the critical-band level in the range of main excitation. Critical-band level, L_G, is calculated using the critical-band intensity, I_G, as a function of the critical-band rate, z, as shown in Equations 22 and 23.

$$I_G(z) = \int_{z-0.5\ Bark}^{z+0.5\ Bark} \frac{dI}{dz} dz \tag{22}$$

$$L_G = 10 \cdot \log10\left(\frac{\mathrm{I_G}}{I_0}\right), \qquad I_0 = 10^{-12}\,\frac{W}{m^2} \tag{23}$$

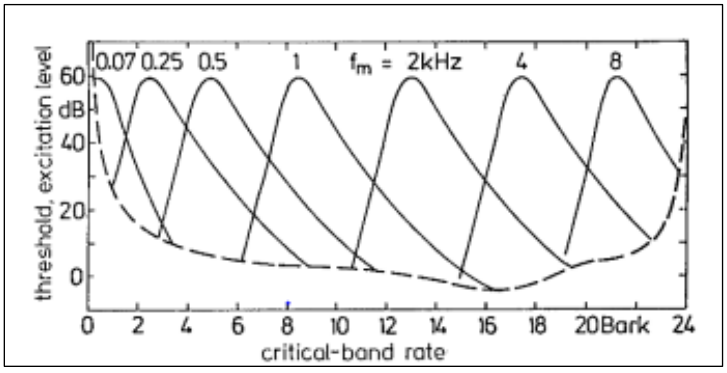

Figure 9: Plot of excitation level as a function of critical-band rate (Zwicker & Fastl, 2007)

Table 2: Relationship between critical-band and frequency (Zwicker & Fastl, 2007) (Reproduced from source)

Bark	Lower Band Frequency (Hz)	Centre Frequency (Hz)	Upper Band Frequency (Hz)
0	0	50	100
1	100	150	200
2	200	250	300
3	300	350	400
4	400	455	510
5	510	570	630
6	630	700	770
7	770	845	920
8	920	1000	1080
9	1080	1175	1270
10	1270	1375	1480
11	1480	1600	1720
12	1720	1860	2000
13	2000	2160	2320
14	2320	2510	2700
15	2700	2925	3150
16	3150	3425	3700
17	3700	4050	4400
18	4400	4850	5300
19	5300	5850	6400
20	6400	7050	7700
21	7700	8600	9500
22	9500	10750	12000
23	12000	13750	15500
24	15500		

4.1.2. Loudness

Loudness is a measure of the intensity of sound perceived by the ear (Zwicker & Fastl, 2007). Loudness values are in phon or sone, where phon is a logarithmic metric and sone is a linear metric. For example, the phon assumes that if a 1 kHz sine wave increases in level by 10 dB (10 phon), the perceived loudness would be double at any other frequency. However, through jury testing this was not found to be the case, as a 30 Hz sine wave was found to only need to increase by 5 dB to instigate a doubling in perceived loudness. Equal-loudness contours are lines that relate the loudness of tones at different frequencies to one another, and are displayed in Figure 10. This shows that tones of different frequencies are perceived at different loudness level, despite being of equal sound pressure level. Loudness is calculated by first calculating the specific loudness, N', in each critical-band rate, as shown in Equation 24.

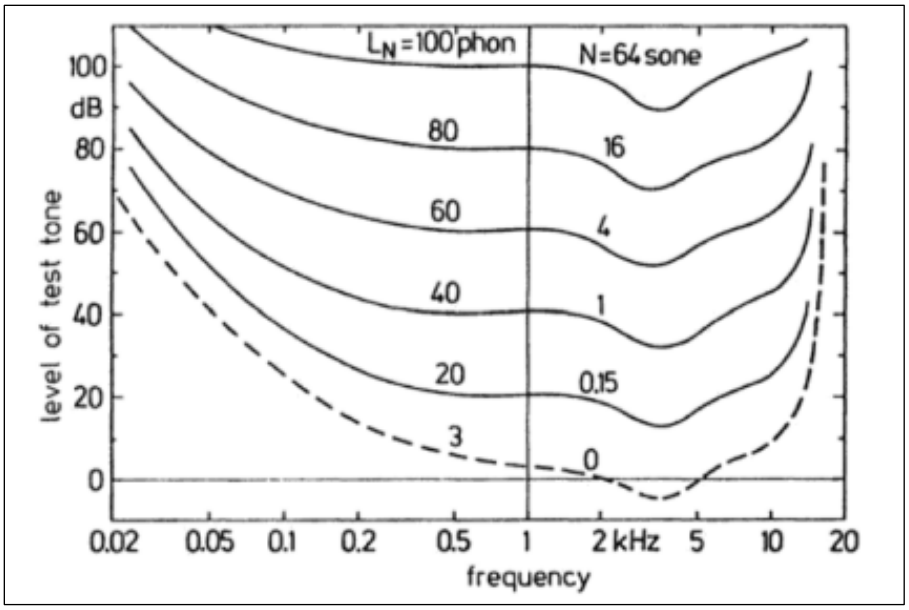

Figure 10: Equal-loudness contours for tones of varying frequency (Zwicker & Fastl, 2007)

$$N' = 0.08\left(\frac{E_{TQ}}{E_0}\right)^{0.23}\left[\left(0.5 + 0.5\frac{E}{E_{TQ}}\right)^{0.23} - 1\right]\frac{sone_G}{Bark} \tag{24}$$

E is the excitation level, E_0 is the excitation level corresponding to the reference intensity I_0, and E_{TQ} is the excitation level at threshold in quiet. The value of loudness is then the integral of all values of specific loudness, as shown in Equation25.

$$N = \int_0^{24\,Bark} N' \tag{25}$$

Loudness has been revisited by various researchers following this initial model to better account for various effects, such as post-masking and temporal non-linearities (Genuit, et al., 2009). The calculation standard DIN 45631/A1, which includes the A1 amendment "Calculation of the loudness of time variant sounds" accounts for these discrepancies in unsteady signals, and is recommended to be used when analysing more technical sound stimuli (Faslt, et al., 2009)

4.1.3. Sharpness

Sharpness can be described as the ratio of energy in higher frequency critical bands, to the energy in all critical bands. As the narrow-band centre frequency of a noise increases, the calculated sharpness of that noise also increases. Sharpness is measured in Acum, and was derived by Zwicker and Fastl (Zwicker & Fastl, 2007). The reference sound producing 1 Acum is a narrow-band noise, 1 critical band-width wide, centred at 1 kHz, with a level of 60 dB.

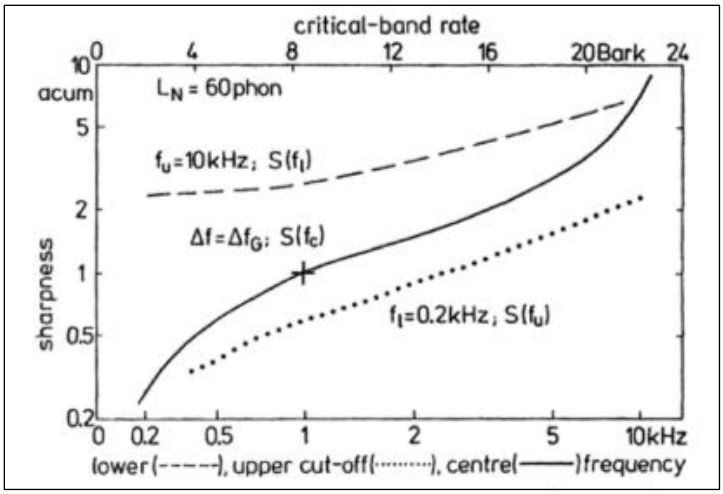

Figure 11: Plot of sharpness against frequency (Zwicker & Fastl, 2007)

Figure 11 shows how as the critical-band or frequency content increases, the calculated sharpness increases. The cross indicated in the figure at about 8.5 Bark is where the reference noise lies, giving a value of 1 Acum from a 1 kHz narrow-band noise at 60 dB. The dotted line in the figure illustrates how, from frequencies of about 1 kHz to about 3 kHz, the sharpness curve is linearly proportional to the centre-frequency. As the centre-frequency increases above 3 kHz, sharpness can be seen to increase at a more rapid rate than centre frequency. Sharpness can be calculated following Equation 26, where N' is the total loudness of all critical-band rates, 24 Bark is the highest critical-band rate, and $g(z)$ is a weighting function dependent on critical-band rate. Fastl and Zwicker defined $g(z)$ as shown in Figure 12.

$$S = 0.11 \frac{\int_0^{24\ Bark} N' g(z) z\ dz}{\int_0^{24\ Bark} N'\ dz} \tag{26}$$

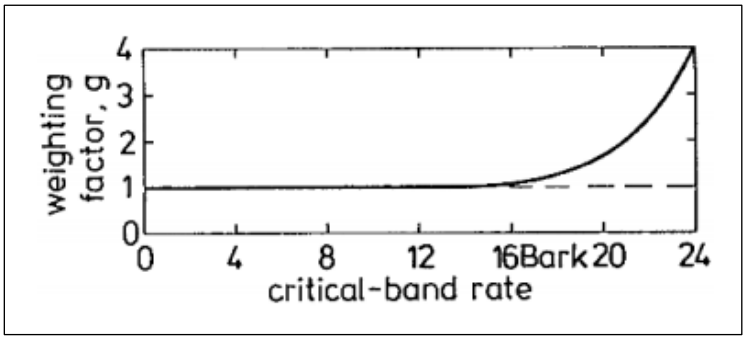

Figure 12: Weighting factor applied in sharpness calculation, dependent on critical-band rate, as defined by Zwicker and Fastl (Zwicker & Fastl, 2007)

Figure 12 shows how the weighting applied to the sharpness calculation in Equation 26 stays at 1 for critical-band rates 0 to 15.8 Bark, then increases exponentially from 15.8 to 24 Bark. Von Bismarck (von Bismarck, 1974) developed this metric by adjusting the weighting function applied to the critical-band rates, with the new weighting function being shown in Equation 27.

$$g_B(z) = \begin{cases} 1 & z \leq 15\ Bark \\ 0.2e^{0.308\left(\frac{z}{Bark} - 15\right)} & z > 15\ Bark \end{cases} \tag{27}$$

However, it has been seen that the overall loudness of a signal holds strong control over perception, but the sharpness model above is loudness normalised and independent of the overall loudness. Therefore, Aures developed a sharpness model which was loudness dependent, and appear as an appendix in DIN 45692. The equation for Aures' model of sharpness is similar Equation 26, except it includes a weighting function which incorporates signal loudness. This is shown in Equation 28, below, where $g_A(z)$ is the new frequency-dependent weighting function. Example curves are given that represent the weighting function for 3 loudness levels (1 sone, 10 sone and 100 sone) in Figure 13.

$$S = 0.11 \frac{\int_0^{24\ Bark} N' g_A(z) z\ dz}{\int_0^{24\ Bark} N'\ dz} \tag{28}$$

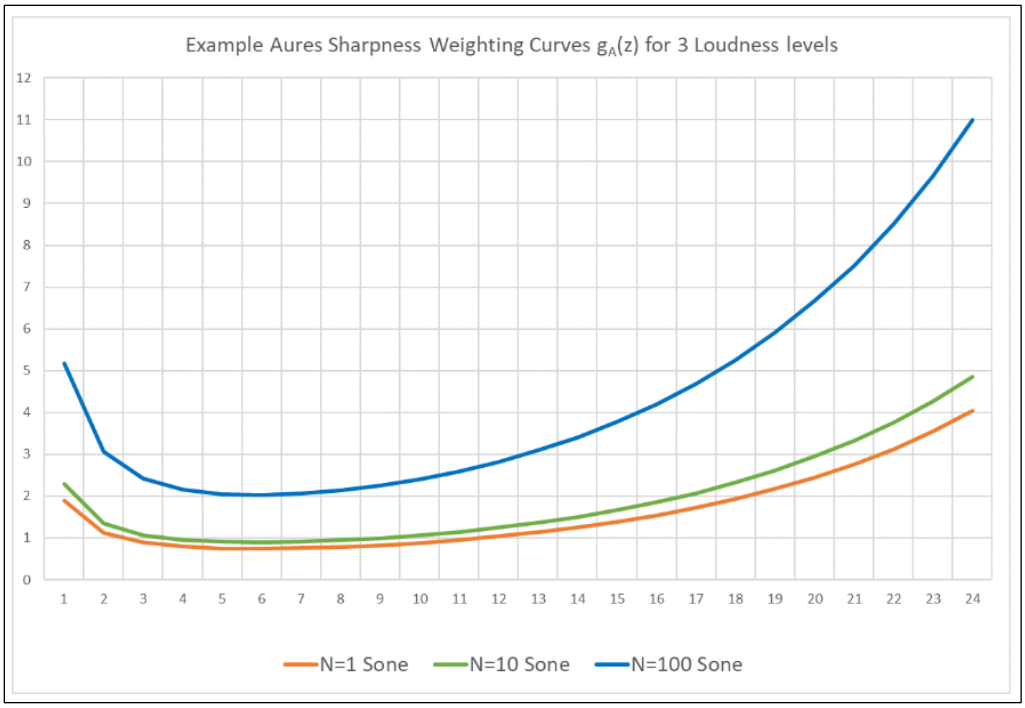

Figure 13: Aures' sharpness weighting curves $g_A(z)$ (Siemens Digital Industries Software, 2021)

4.1.4. Roughness, fluctuation strength and impulsiveness

Roughness is a sensation that is influenced by variation of a signal's amplitude or frequency. In the cases of pure tones which are modulated by amplitude or frequency variations, if the variations occur at frequencies below 10 Hz, the human ear can actively track the change, which is the impression of a pulse or a beating effect. As the frequency of variation increases from around 20 to 300 Hz, the sensation of roughness can be perceived, as the ear no longer becomes capable of actively tracking any changes in the temporal structure of the signal. Above 300 Hz, the centre frequency of the tone and the sidebands created due to amplitude and frequency modulated start to become individually perceptible. Roughness depends on the centre frequency, modulation frequency and modulation depth of the signal. As modulation depth increases, the sensation of roughness increases. For modulation frequency, as it increases, the roughness sensation is seen to have a band-pass characteristic; as modulation frequency tends towards very low or high values, roughness decreases. For an amplitude-modulated tone of carrier frequency 1 kHz, and maximum modulation depth, the maximum roughness value was perceived at a modulation frequency of 70 Hz. The unit of roughness, asper, is defined by a sinusoidal tone of 1 kHz with a level of 60 dB, with an amplitude-modulation frequency of 70 Hz with a modulation depth of 1 applied to it, which gives a value for roughness of 1

asper. Fluctuation strength is similar to roughness, in that it is a sensation caused by signal modulation, but at very low frequencies. The unit for fluctuation strength is vacil, and is defined by the same sinusoidal tone used to define asper, except amplitude-modulated at a frequency of 4 Hz instead of 70 Hz.

Sottek (Sottek, 1993) created a hearing model, that resembled the audio processing functions of the human ear, to calculate roughness and fluctuation strength. First the outer and middle ear are represented by filtering of the input audio signal, with a transfer function as shown in Figure 14. Then, the signal is passed through a filter bank, comprising of parallel overlapping band-pass filters. The centre frequencies of the band-pass filters are distributed evenly across critical-band rates. Once the signal is filtered, using the Hilbert transformation (Kak, 1973), the envelopes of partial band signals are determined. Next, 3rd order low-pass filters are applied, with cut-off frequencies dependent on the filtered signal frequencies. For the signal filtered at 1 kHz, the low-pass filter has a cut-off frequency of 120 Hz. This filtering accounts for the modulation frequency above which the human ear cannot actively track the modulation change. A nonlinear filtering function is applied to the critical-band rate filtered signals, which is an exponential function with an exponent of 0.125. 3rd order high-pass filtering is then applied to the critical-band rate filtered signals, as well as a frequency dependent amplitude factor, $g_R(z_i)$.

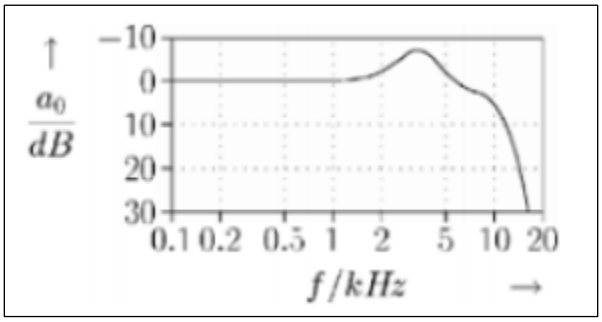

Figure 14: Transfer function from the outer to inner ear (Sottek, 1993)

The high-pass and low-pass filtering create the band-pass variation of perceived roughness as a function of modulation frequency. This gives the specific roughness or fluctuation strength per critical-band rate, depending on the modulation frequency applied to the signal. Roughness and fluctuation strength can be calculated by integrating the specific roughness or fluctuation strength, respectively. Figure 15 shows a functional block diagram of the hearing model.

The impulsiveness of a signal relates to the sensations perceived by the ear due to significant dynamic fluctuations of the level of the sound. Sottek developed a method for calculating impulsiveness from his already influential hearing model (Sottek, et al., 1995). First, the excitation function, $e(z_j, t)$, or e_j for short, is calculated for each input channel j, versus frequency and time using the previously described hearing model. Nonlinear processing is then applied the distribution of excitation. If background noise is present in the signal, the impulsiveness sensation is reduced considerably, therefor a function is applied that is approximately linear for small signal amplitudes, but resembles a power function with an exponent, a, of 0.15 for greater signal amplitudes. This is illustrated in Figure 16, which shows that as the amplitude of the input pulse increases, the effect of the compressive, nonlinear processing changes.

Once all specific impulse (I') values are calculated from the individual critical-band filters, as described in Sottek's hearing model, the total impulsiveness can be calculated from summation. A 4th order high-pass filter is applied with $f_{3\,dB} = 10\,Hz$ to the difference in the numerator of the summation and I', so to reproduce the dependency on the impulse rate frequency. In Equation 29, k_j is a channel dependent weighting factor. Then, total impulsiveness, I, is calculated from Equation 30.

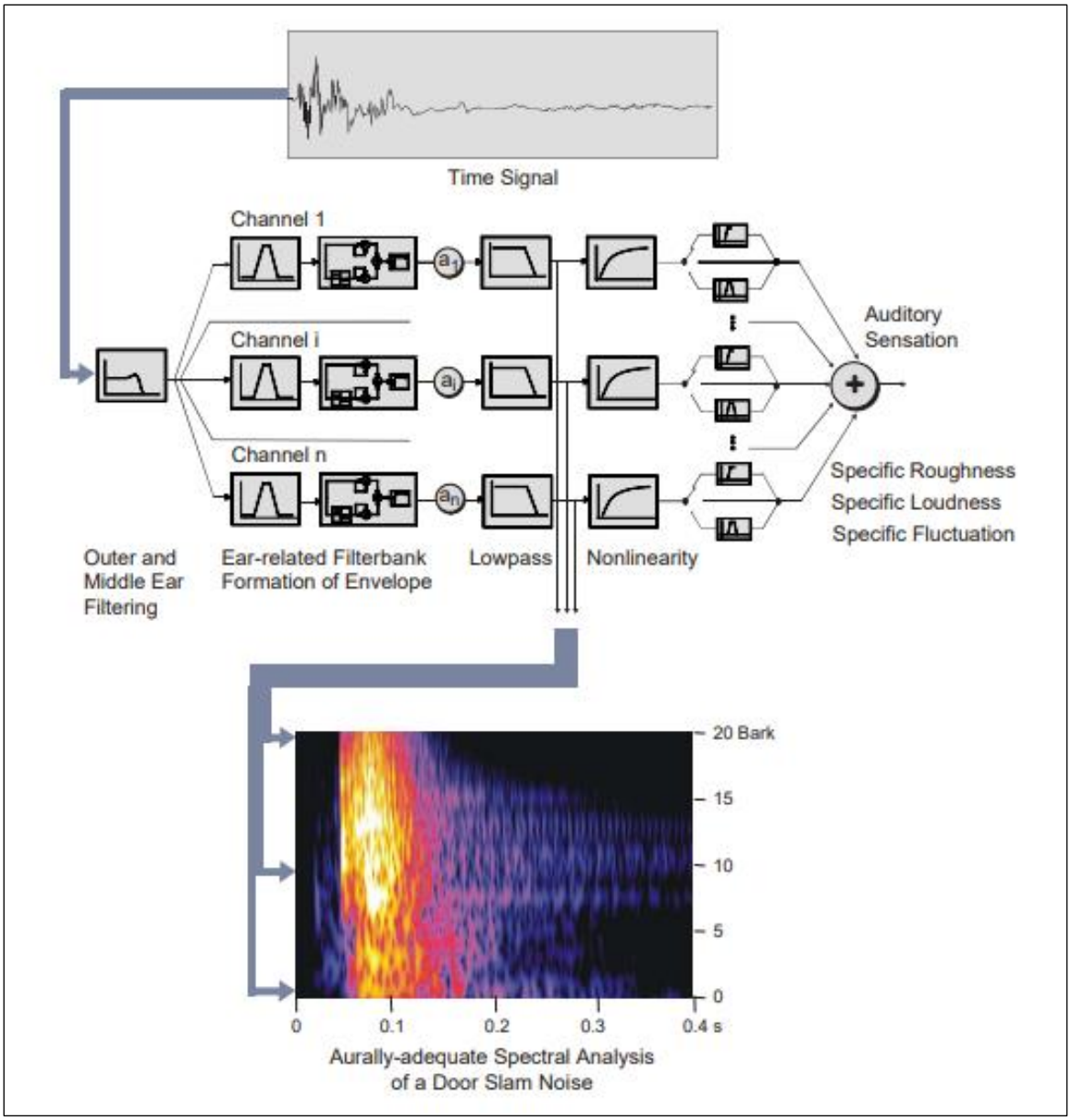

Figure 15: Functional block diagram of hearing model used to calculate specific roughness and fluctuation strength (Sottek, 1993)

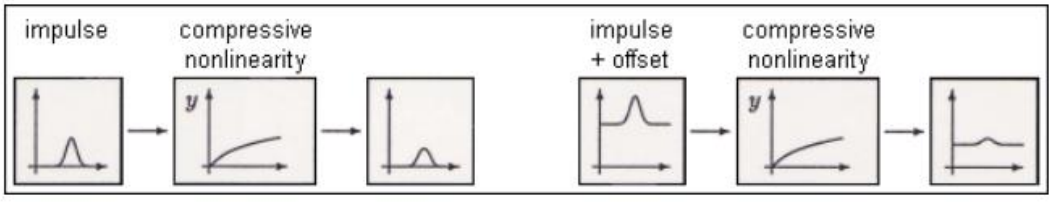

Figure 16: Effect of the nonlinear processing on an input impulse, with the weighting dependent on the input level (Sottek, 1993)

$$I' = \sum_{j=1}^{N} k_j \cdot \frac{\overline{\left(y(e_j) - \overline{y(e_j)}\right)^n}}{\overline{y(e_j)}^m} \tag{29}$$

$$I = 0.055556 \cdot I' + heavi(I' - 1.8) \cdot (1.271367 \cdot I' - 2.288461) + heavi(I' - 7) \cdot (2.288461 - 0.326923 \cdot I')$$

$$heavi(x) = \begin{Bmatrix} 0, & x < 0 \\ 1, & x \geq 0 \end{Bmatrix} \tag{30}$$

4.1.5. Tonality

Tonality quantifies the dominance of tonal components in signals. An early version of a tonality method was presented by Terhardt et al (Terhardt, et al., 1982), which took the form of a hearing model (Sottek, 1993) designed to detect pitch and pitch salience. First, a signal is converted from the time domain to frequency domain, via a fast Fourier transform (FFT), with a Hanning window. From this, tonal components are interrogated following Equation 31, where L_i is the relative SPL of the i[th] spectral sample, L_{i-1} is the SPL of the previous sample, and L_{i+1} is the SPL of the next spectral sample. If a spectral sample meets this criterion, then it is investigated further following Equation 32. If both conditions are met by a spectral sample, then sample $i - 3$ to $i + 3$ are considered to be a tonal component.

$$L_{i-1} < L_i \geq L_{i+1} \tag{31}$$

$$L_i - L_{i+j} \geq 7\ dB\ ; j = -3, -2, +2, +3 \tag{32}$$

$$f_c = f_i + 0.46(Hz/dB)(L_{i+1} - L_{i-1}) \tag{33}$$

The frequency of the tonal component, f_c, is then found following Equation 33. This is repeated until all eligible tonal components have been found in the signal's frequency spectrum. This gives the values N, the total number of eligible tonal components found in the spectrum, f_c for each tonal component, and L_c, the SPL of each tonal component. Then, the effect of masking on each tonal component is approached. That is to say, how dominant that tonal component is compared to the rest of the frequency spectrum content. SPL excess, LX_μ $(1 \leq \mu \leq N)$, is calculated for each tonal component, μ, following Equation 34.

$$LX_\mu = L_\mu - 10\log_{10}\left[\left(\sum_{\substack{v=1 \\ v \neq \mu}}^{N} 10^{\frac{L_{Ev}(f_\mu)}{20\ dB}}\right)^2 + I_{N\mu} + 10^{\frac{L_{TH}(f_\mu)}{10\ dB}}\right] \tag{34}$$

$$L_{Ev}(f_\mu) = L_v - s(z_v - z_\mu) \tag{35}$$

$$s = \begin{bmatrix} 27\ dB/Bark & if\ f_\mu \leq fv \\ -24 - (0.23\ kHz/f_v) + (0.2L_v/dB)\ dB/Bark & if\ f_\mu > f_v \end{bmatrix} \tag{36}$$

$$z = \{13\arctan(0.76(f/kHz)) + 3.5\arctan^2(f/7.5\ kHz)\} \tag{37}$$

$$L_{TH}(f_\mu) = 3.64 f_\mu^{-0.8} - 6.5e^{(-0.6f_\mu - 3.3)^2} + (f_\mu^4 \times 10^{-3}) \tag{38}$$

$L_{ev}(f_\mu)$ is the excitation level which is produced at the frequency f_μ by the v^{th} tonal component. The v^{th} tonal component is not the same as the μ^{th}, however, as the μ^{th} tonal component's SPL level is skipped in the summation

within Equation 34. $L_{TH}(f_\mu)$ is the hearing threshold of tonal frequency f_μ. $I_{N\mu}$ is the intensity of noise within the critical-band of the μ^{th} tonal component, calculated from adding the sound intensities of the spectrum samples corresponding to the critical band-rates of the μ^{th} (z_μ) and v^{th} (z_v) tonal components, calculated in Equation 36. Using Equations 34 through 38, the SPL excess, LX_μ, can be calculated. If LX_μ has a positive value, then the tonal component μ is considered significant. If the value of LX_μ is 0 or negative, then the tonal component μ is considered insignificant. Aside from SPL excess, the effects of simultaneous spectral components interacting with the auditory system are accounted for. H_μ, the spectral pitch of the μ^{th} tonal component, is measured in pitch units (pu). For tonal components with a value of LX_μ greater than 0, v_μ, the induced pitch shift caused by interacting tonal components, is calculated following Equations 39 through 42.

$$H_\mu = (f_\mu)(1 + v_\mu) \tag{39}$$

$$v_\mu = 2(L_\mu - 60)(f_\mu - 2) \times 10^{-4} + 1.5e^{-\frac{LX'_\mu}{20}}(3 - \ln(f_\mu)) \times 10^{-2} + 3e^{-\frac{LX''_\mu}{20}}(0.36 + \ln(f_\mu)) \times 10^{-2} \tag{40}$$

$$LX'_\mu = L_\mu - 20log10\left(\sum_{v=1}^{\mu-1} 10^{\frac{L_{Ev}(f_\mu)}{20}}\right) \tag{41}$$

$$LX''_\mu = L_\mu - 20log10\left(\sum_{v=\mu+1}^{N} 10^{\frac{L_{Ev}(f_\mu)}{20}}\right) \tag{42}$$

LX'_μ is the SPL excess of the lower frequency tone within the interaction, and LX''_μ is the SPL excess of the higher frequency tone within the interaction. The process of the masking algorithm collectively gives R, a number of relevant tonal components, where $R \leq N$, the SPL excess (LX_μ) of the relevant tonal components, and the spectral pitch (H_μ) of the relevant tonal components. These are used to appropriately apply a weighting factor to the tonal components, with the most significant factors that dictate the components weighting being SPL excess and frequency. The spectral-pitch weight, WS_μ, is described in Equations 43 and 44.

$$WS_\mu = \left[1 - e^{-\frac{LX_\mu}{15}}\right]\left[1 + 0.07\left(\frac{f_\mu}{0.7} - \frac{0.7}{f_\mu}\right)^2\right]^{-\frac{1}{2}}, if\ LX_\mu \geq 0 \tag{43}$$

$$WS_\mu = 0, if\ LX_\mu < 0 \tag{44}$$

Virtual pitch is the perceived pitch of a complex tone where the fundamental frequency is perceived by the brain from the series of harmonics in a signal, and may correspond to a frequency which is not included in the harmonic structure. This is accounted for in Terhardt's model by including a virtual pitch evaluation. Virtual pitch candidates are specified as subharmonics of one of the relevant tonal components. Virtual pitch in this method, H_{im}, of the m^{th} subharmonic of the i^{th} relevant tonal component is calculated following Equation 45.

$$H_{im} = m^{-1}(f_i/Hz)(1 + v_i - \text{sign}(m-1)\,10^{-3}\{18 + 2.5m - (50 - 7m)(f_i/kHz)m^{-1} + 0.1[m^{-1}(f_i/kHz)]^{-2}\}) \tag{45}$$

Aures (Aures, 1984) suggested an updated method to this previous hearing model of tonality, by accounting for frequency, bandwidth, and level of all tonal components, as well as noise included in a signal. Figure 17 shows a working block diagram of Aures' method for tonality.

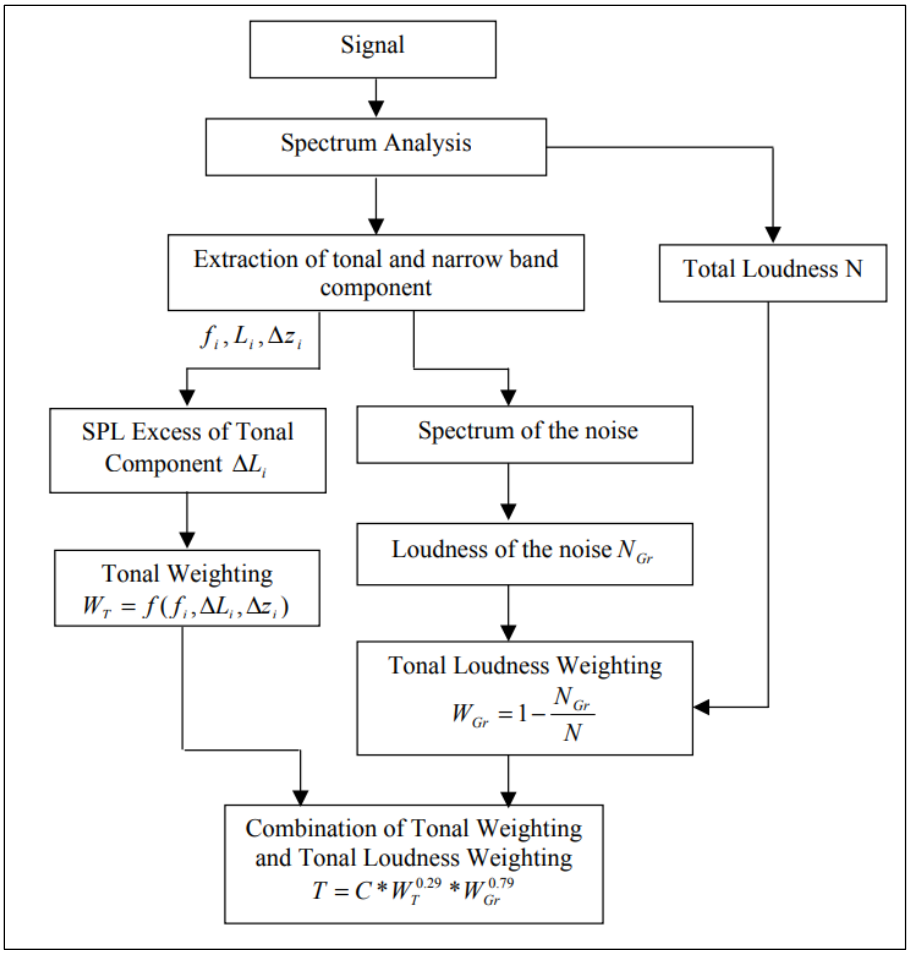

Figure 17: Working block diagram of Aures' tonality method

In this method, the calculation of SPL excess is the same as the Terhardt method, but introduces identifying tones and applying weighting functions based on the bandwidth (Equation 46), centre frequencies (Equation 47), and prominence of each tonal component present (Equation 48).

$$w_1(\Delta z_i) = \frac{0.13}{\Delta z_i + 0.13} \tag{46}$$

$$w_2(f_i) = \left(\frac{1}{\sqrt{1 + 0.2\left(\left(\frac{f_i}{700}\right) + \left(\frac{700}{f_i}\right)\right)^2}} \right)^{0.29} \tag{47}$$

$$w_3(\Delta L_i) = \left(1 - e^{-\frac{\Delta L_i}{15}}\right)^{0.29} \tag{48}$$

Where Δz_i is the bandwidth of the tonal component i, expressed in Bark, f_i is the frequency of the tonal component i, in Hz, and ΔL_i is the level of the tonal component i above the broadband masking noise, which is explained in equation 4 to 8 of Terhardt et al. (Terhardt, et al., 1982). An overall weighting function is applied by combining Equations 46 through 48, and is shown in Equation 49.

$$w_T = \sqrt{\sum_{i=1}^{n} \left[\left(w_1 (\Delta z_i)^{\frac{1}{0.29}} \right) \left(w_2 (f_i)^{\frac{1}{0.29}} \right) \left(w_3 (\Delta L_i)^{\frac{1}{0.29}} \right) \right]^2} \tag{49}$$

$$K = c \cdot w_T^{0.29} \cdot w_{Gr}^{0.79} \tag{50}$$

$$W_{Gr} = 1 - \frac{N_{Gr}}{N} \tag{51}$$

Aures' tonality (K), in tonality units (tu), is calculated as shown in Equation 50. The weighting function, W_{Gr}, that accounts for the overall loudness of the tone to noise ratio is calculated following Equation 51, where N_{Gr} is the loudness of the broadband noise, and N is the total loudness of the sound. The calibration constant, c, is equal to 1.09, and is given so that a 1 kHz pure tone with a sound level of 60 dB has a tonality value of 1 tu.

Simpler forms of quantifying the tonality of a signal have also been used, where a signal does not contain as much complex tonal content. For example, the tone-to-noise ratio (TTNR) approach considers a tonal artifact that is 8 dB or more over the adjacent masking noise in the signal. Masking noise is the background noise that, if high enough, makes the tonal content of a signal perceptually inaudible. The process for calculating the tone-to-noise ratio of a tonal artefact starts by finding the sound pressure level of the tonal component. Then the average sound pressure level of the critical band that the tonal component is in (including the tone) is calculated. The sound level of the tone is then subtracted from the sound level of the critical band to leave the masking noise level. The tone-to-noise ration of that tonal component is then calculated following Equation 52, where L_T is the sound level of the tonal component, and L_M is the sound level of the masking noise. A tonal component as assessed by this method is said to be significant if the TTNR is greater than 8 dB.

$$TTNR = 10 * \log_{10} \left(\frac{L_T}{L_M} \right) \tag{52}$$

Alternatively, the prominence ratio (PR) method takes the whole critical band containing a tonal component, and assesses the level of that critical band against the adjacent bands. First, the sound level of the critical band containing the tonal component is calculated. Then the sound levels of the two adjacent critical bands are calculated. The PR is then found using the following Equation 53, where L_A is the sound level of the tonal critical band, and L_B and L_C are the sound levels of the adjacent critical bands. The PR value of a tonal critical band must be greater than 9 dB for it to be considered significant.

$$PR = 10 * \log_{10} \left(\frac{L_A}{(L_B + L_C) * \ 0.5} \right) \tag{53}$$

5. STATISTICAL ANALYSIS TECHNIQUES

5.1. Regression analysis and multilevel linear modelling

Linear regression is a statistical analysis technique that is used to investigate correlations between dependent and independent variable datasets. Multilevel linear modelling can be used to assess regression factors across different groups, or datasets which have a hierarchal or clustered structure. In multilevel modelling, there are several levels of model that can be built. Level 1 models resemble a simple linear regression model, but with repeated calculation for multiple observations per group, which follows Equation 54.

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + e_{ij} \tag{54}$$

Where Y_{ij} is the dependent variable for an individual observation, i, for group j, X_{ij} is a level 1 predictor for that dependent variable, β_{0j} is the intercept of the dependent variable for group j, β_{1j} is the predictor coefficient, or slope for the relationship between dependent variable and the level 1 predictor for group j, and e_{ij} is the residual error. For level 1 models, intercepts and slopes of the regression lines between dependent and independent variables can either be fixed per group, vary randomly between each group, or be non-randomly varying by including a level 2 predictor variable in the level 1 regression coefficients, illustrated in Equations 55 and 56.

$$\beta_{0j} = \gamma_{00} + \gamma_{01}W_j + u_{0j} \tag{55}$$

$$\beta_{1j} = \gamma_{10} + u_{1j} \tag{56}$$

Where γ_{00} is the overall intercept of the model, or mean of all dependent variables when all predictor variables equal zero. γ_{01} is the regression coefficient relating the dependent variable and the level 2 predictor, W_j. u_{0j} is the random error component for the deviation of the group j and the overall intercept. γ_{10} is the overall regression coefficient for all groups between the dependent variable and the level 1 predictor variable and u_{1j} is the error component for the slope. Random intercept models, or models which allow the intercepts of regression slopes to vary between group but maintains fixed regression slopes across groups can be used to evaluate the intraclass correlation coefficient (ICC). The ICC is a descriptive statistic that can be used to assess which dependent variable responses correlate within a group. The ICC is calculated using the residual variance, σ^2_{eij}, and the variance of the subject-dependent intercepts, σ^2_{u0j}, and is shown in Equation 57.

$$ICC = \frac{\sigma^2_{u0j}}{\sigma^2_{u0j} + \sigma^2_{eij}} \tag{57}$$

5.2. Principle component analysis

Principle component analysis can be used to understand key characteristics that can explain correlations in a dataset. The main goal of PCA is to dimensionally reduce large datasets, through transformation resulting in a smaller dataset that still describes most of the information of the previous large dataset. The main trade-off between a large and small dataset is between accuracy of the data and the simplicity of interpreting results from the data. The first stage of PCA is typically to standardise the values of continuous variables in the dataset, so that contributions of the variables towards the analysis are equal. This is done by subtracting the mean of a variable from any individual variable value, and then dividing by the standard deviation of that variable, as shown in Equation 58 below.

$$z_n = \frac{x_n - \bar{x}}{\sigma_x} \tag{58}$$

Where z_n is the normalised variable value, x_n is a value from the initial variable dataset, $\bar{x}$ is the mean of the variable dataset, and σ_x is the standard deviation of the variable dataset. This stage ensures that all variables used in the PCA will be transformed to the same scale. The next stage is to compute the covariance matrix of the dataset. This step highlights any correlation between the variables being included in the PCA. The covariance matrix is a $p \times p$ sixed symmetric matrix, where p is the number of dimensions describing the dataset, or the number of variables. For example, a dataset including 3 variables (x, y and z) would have a covariance matrix as illustrated below.

$$\begin{bmatrix} Cov(x,x) & Cov(x,y) & Cov(x,z) \\ Cov(y,x) & Cov(y,y) & Cov(y,z) \\ Cov(z,x) & Cov(z,y) & Cov(z,z) \end{bmatrix} \tag{59}$$

The values down the main diagonal, where covariances are calculated between a variable and itself (i.e., $Cov(x,x)$), would all equal 1. Furthermore, the $Cov()$ functions which are interrogating the same variables (i.e., $Cov(x,y)$ and $Cov(y,x)$) would be equal. Eigenvectors and their corresponding eigenvalues are calculated from the covariance matrix, with the total number of eigenvectors found being equal to the number of variables in the initial dataset (p). Eigenvectors from the variance matrix can be seen as the direction in which the most variance occurs in a dataset, and give the directions of the principal components. The corresponding eigenvalues are used to rank the significance of each eigenvector, by quantifying the amount of variance carried in each principal component. To find the percentage of the variance explained by one principal component, the eigen value for principal component is divided by the sum of all eigenvalues for that dataset. The percentage of the variance explained is an intuitive metric for assessing the important of principal components relating to a dataset of variables.

The next step is to form a feature vector, which includes eigenvectors for each variable in the covariance dataset up to a maximum value, or until the point where the sum of the covariance for each of the variables included in the feature vector equal a minimum criterion. Typically, the latter method is taken, with the minimum variance explained by the eigenvectors included being set as 95%. For example, if a dataset had 4 variables, with covariance explained values of 76%, 15%, 6% and 3% respectively, then only the eigenvectors of the first two variables would be included in the feature vector. Now the principal components have been chosen based on their eigenvalues, scores can be calculated for each input variable that translate them onto each principal component. These scores give indication of the significance each variable has on a principal component.

6. PART 1 – HUMAN RESPONSE TO SMALL UAV NOISE

6.1. Methodology

The first part of the subjective testing aimed to correlate psychoacoustic characteristics of isolated drone noise stimuli with human response values given by participants during subjective testing. Psychoacoustic characteristics of the UAV stimuli were described using sound quality metrics, and the participants gave responses of annoyance, loudness and pitch to the UAV stimuli. These values were then correlated using simple regression and multilevel modelling with calculated SQMs to assess the effectiveness of the metrics to predict values of annoyance, loudness and pitch of UAV stimuli. Details of the methodology are given in the following sub-sections.

6.1.1. UAV sound stimuli

A total of 44 sound stimuli were gathered from industry partners, which represented 8 different drone models. These models varied in weight and size, and were captured performing various operations, such as take-offs, hovering, and landing at different distances. Each stimulus was edited to be 4 seconds long, with any pass-by event taking place halfway through the 4 second period (Torija, et al., 2019). The UAV stimuli are presented in detail in Table 3. It has been previously shown in research that UAV constantly adjust to counter adverse weather conditions when in flight. These micro-adjustments are of the rotational speeds of the drone rotors, and are to maintain the stability of the vehicle. Due to these micro-adjustments, rapid fluctuations of frequency are introduced into the acoustic signature, and often have a negative impact on the overall perception of UAV (Torija, et al., 2019). It was therefore deemed appropriate to include a range of operations, to evaluate which operating conditions are deemed to be the most annoying. A range of operating distances were also included, to observe if audible distance can be correlated to perceived annoyance, loudness and pitch. To maintain relative L_{Aeq} between the UAV stimuli, a calibration system was implemented, using measurement data provided with the stimuli. Since the effect of operational distance is being assessed, keeping the relative levels between the sound stimuli is important for correlating with annoyance. The calibration setup is presented in Figure 18.

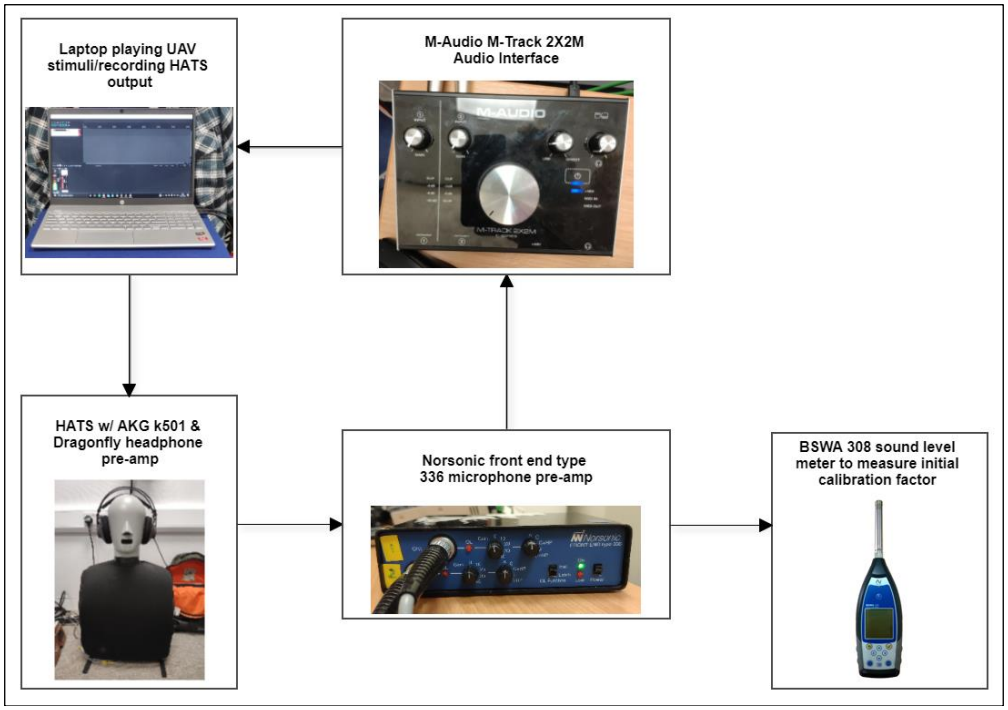

Figure 18: Calibration set up used for subjective test 1 UAV stimuli

Table 3: UAV stimuli included in subjective test 1

Sound	UAV	UAV Weight (kg)	Operation	Distance (m)	Calibrated LAeq
1	DJI Inspire	2.85	Flyover	15	52
2	DJI Inspire	2.85	Flyover	7.5	58
3	DJI Inspire	2.85	Landing	7.5	64
4	DJI Inspire	2.85	Takeoff	2	70
5	Intel Falcon	1.2	Flyover	30	54
6	Intel Falcon	1.2	Flyover	60	47
7	DJI Matrice 600	9.1	Takeoff	3	71
8	DJI Matrice 600	9.1	Hover	40	65
9	DJI Matrice 600	9.1	Flyover	40	57
10	DJI Mavic	0.743	Flyover	15	51
11	DJI Mavic	0.743	Flyover	30	46
12	DJI Mavic	0.743	Flyover	60	37
13	DJI Mavic	0.743	Maneuvering	7.5	51
14	DJI Mavic	0.743	Maneuvering	7.5	53
15	DJI Mavic	0.743	Takeoff	7.5	59
16	DJI Phantom 3	1.216	Maneuvering	2	68
17	DJI Phantom 3	1.216	Takeoff	2	64
18	DJI Phantom 3	1.216	Landing	2	62
19	DJI Phantom 3	1.216	Hover	2	69
20	DJI Phantom 3	1.216	Ascending	2	64
21	DJI Phantom 3	1.216	Flyover	2	61
22	DJI Phantom 3	1.216	Flyover	2	63
23	DJI Phantom 3	1.216	Flyover	2	66
24	DJI Phantom 3	1.216	Flyover	5.4	56
25	DJI Phantom 3	1.216	Flyover	5.4	59
26	DJI Phantom 3	1.216	Flyover	5.4	57
27	DJI Phantom 3	1.216	Hover	2.2	62
28	DJI Phantom 3	1.216	Hover	5.1	56
29	DJI Phantom 3	1.216	Hover	2.2	67
30	DJI Phantom 3	1.216	Hover	3.6	67
31	DJI Matrice 200	4	Flyover	46	56
32	DJI Matrice 200	4	Flyover	46	45
33	DJI Matrice 200	4	Takeoff	30	50
34	DJI Matrice 200	4	Landing	30	52
35	DJI Matrice 200	4	Hover	1.2	56
36	Yuneec Typhoon	2	Flyover	46	48
37	Yuneec Typhoon	2	Flyover	46	44
38	Yuneec Typhoon	2	Takeoff	30	46
39	Yuneec Typhoon	2	Landing	30	52
40	Yuneec Typhoon	2	Hover	1.2	57
41	Gryphon GD28X	11.8	Takeoff	30	53
42	Gryphon GD28X	11.8	Landing	30	54
43	Gryphon GD28X	11.8	Maneuvering	30	57
44	Gryphon GD28X	11.8	Hover	1.2	60

The calibration system included a BSWA 308 sound level meter, a Head and Torso System (HATS), a Norsonic front end type 336 microphone, an M-Audio M-Track 2x2M audio interface, and a laptop to play and record the sound stimuli through the system, a pair of AKG k501 headphones, and a Dragonfly headphone pre-amplifier. Firstly, the BSWA 308 meter was used to measure the calibration factor of the system. The meter was calibrated to a 94 dB, 1 kHz sine wave, and when no significant drift from the previous calibration measurement was observed, was plugged into the output of the Norsonic microphone pre-amp. Then, a 94 dB, 1 kHz sine wave was played from the laptop, through the headphones, into the microphones situated in the HAT systems ears, and into the microphone pre-amp. The level of the sine wave was measured by the sound level meter, and the difference between this level and the initially measured 94 dB was the calibration factor to be applied to the UAV stimuli. The sounds were calibrated to this system, and re-recorded through this system, so that the analysis would be of the UAV stimuli presented to the participants using the exact system that they were to use during the subjective testing. This was so that any frequency colouration added by the headphones or other parts of the system would be accounted for. Due to the COVID-19 pandemic, and the restrictions of the lockdown, it was impossible to invite participants into the university to attend any subjective testing, so an online experiment was designed to allow participants to give response data from home.

6.1.2. Online subjective testing

Participants were recruited for both parts of this research via email to members of a group that have given the University of Salford's Acoustics Research department consent to be contacted regarding upcoming research being facilitated by department. The group were sent the researcher's contact information, as well as an information form detailing the reasoning and methodology behind the subjective experiments. Furthermore, colleagues shared the information via social media platforms, such as LinkedIn, to their peers. Interested peers then contacted the researcher via email, were given further detail of the subjective experiment, gave written consent to undertake the experiments if they wished to do so, and were given an identifying ID. Participants' data was stored securely, and the data gathered was anonymised. A total of 49 participants successfully took part in both experiments. Of the 49 participants, 34 were male (69.4%), and 15 were female (30.6%).

An online testing interface was built, using the Web Audio Evaluation Toolkit (WAET) (Jillings, et al., 2015), to allow for participants to listen to the 44 UAV stimuli from home, and respond with values of perceived annoyance, loudness and pitch using sliders on the interface. The inclusion of perceived loudness and pitch as response values, as well as perceived annoyance, to investigate the correlation between the calculated sound quality metrics and perceived loudness and pitch, as well as the relationships between perceived loudness and pitch, and perceived annoyance. The part 1 online subjective test took about 20 minutes to complete, and all UAV stimuli were presented to the participants. When a UAV stimulus was presented, the participant could listen to the stimuli as many times as they wanted, by pressing on a slider. Once the participant had listened to the current UAV stimulus, they would proceed to rate the sound using the sliders to give values of perceived annoyance, loudness and pitch. If the participant was satisfied with their responses, they could progress to the next stimulus, or if they were not satisfied, could relisten to the sound and change their responses. For each participant, the order of the UAV stimuli presented was randomised by an internal function of the WAET.

Before the participant response stage commenced, however, there were several stages to ensure that the UAV stimuli was being played back correctly and at a safe and appropriate level. Since the test was accessed by the participants online, each participant would be using a different system for the UAV stimuli playback. To mitigate potential risks to the participants regarding excessive noise exposure, the first pre-experiment stage was a safety precaution stage. The participants were presented with a practice page, which was similar to the design of the main participant response page. The practice page included the sliders for perceived annoyance, loudness and pitch, but had 5 different UAV stimuli that the participant could listen to, ranging in L_{Aeq} from the stimuli which was the quietest, to the stimuli which was the loudest. The participants were then asked to adjust the level of their playback system so that the UAV sound with the highest level was not at an uncomfortable level, but the UAV sound with the lowest level was still audible. The participants were then asked to not change the level of their playback system for the duration of the subjective experiment. Of course, this would introduce some stimuli level differences between

participants, which would have been controlled for by calibrating the stimuli and playback system prior to a laboratory based subjective test. Participants were also asked to adjust the levels of a series of tones so that they were perceived to be of equivalent level. This data was stored and could be assessed if a participant responds with particularly anomalous response values, as these could be due to hearing losses. A stage was included to make sure that the playback of the sound stimuli was in stereo. Participant IDs were given to each participant with an invitational email to the online experiment, which the participant would enter into the experiment interface prior to commencing the experiment. 49 participants agreed to participate and complete the online experiment, and consent was given to take part in the experiment by the participants using tick boxes, which were required to be checked before the test commencing.

6.1.3. Analysis

For the first subjective test, the analysis was split into 3 key stages. The first stage was an initial, simple analysis of the response data, which was gathered from the online subjective testing, and the calculated sound quality metrics of the UAV sound stimuli. The subjective response data and objective sound quality metrics were assessed for correlation using simple linear regression. Values of loudness, sharpness, roughness, tonality, fluctuation strength and impulsiveness were calculated using the HEAD Acoustics ArtemiS SUITE 12.5 software. Loudness was calculated following the DIN 45631/A1 standard. Sharpness was calculated following the Aures model, due to the observably large variance in the loudness of the stimuli (Aures, 1984). Tonality was also calculated following Aures' model (Aures, 1984). Roughness, fluctuation strength and impulsiveness were calculated following the methods derived by Sottek (Sottek, 1993) (Sottek, et al., 1995). Furthermore, to compare the suitability between metrics quantifying loudness, the $L_{Aeq,4s}$ and PNL values for each UAV stimuli were calculated using MATLAB. These metrics were chosen following the literature review and perception-influenced engineering review sections of this thesis, sections 2 & 3. The value that is exceeded for 5% of the stimuli time interval, or the 5th percentile, is commonly used in psychoacoustic calculations of objective metrics, to mitigate the effect of noise in stimuli. Furthermore, the first 0.5 s of each stimulus was omitted, as to remove any transient effects introduced by editing and cropping the UAV stimuli files (Torija, et al., 2021). The calculated 5th percentile sound quality metrics values were used in the linear regression analysis, which was implemented using IBM SPSS statistics software, which efficiently carries out statistical analysis calculations and generates statistical plots. The simple linear regression was carried out to evaluate the statistical significance of the sound quality metrics. This regression creates an equation that describes the correlation between the sound quality metrics and the gathered responses of perceived annoyance, loudness and pitch from the subjective testing. The general equation for the regression is described in Equation 60.

$$Y_i = \gamma_0 + \gamma_1 X_{1i} + \cdots + \gamma_n X_{ni} + e_i \tag{60}$$

Y_i is the average perceived annoyance of sound i, γ_0 is the y-axis intercept of the model, γ_n is the correlation coefficient that pairs with X_{ni}, the n-th sound quality metric 5th percentile value of sound i, and e_i, the residual error. To choose the most statistically significant sound quality metrics to be included in the regression model, a backwards stepwise method was implemented. Backwards stepwise regression determines which variables to include in a model by first including all variables in the model, and then removing the metric that has the smallest reduction in R^2 value, or the most statistically insignificant variable. This process was repeated until no variables could be removed without a significant reduction in R^2. The 2nd of the analysis was an initial investigation into the effects of distance on the perceived annoyance, loudness and pitch for the UAV stimuli performing flyover operations. A total of 18 of the UAV stimuli were of flyover operations at distances varying from 2 metres to 60 metres. The intrusiveness of UAVs will no doubt be a key factor in the acceptance of UAV as a viable form of delivery, as well as other services. If the relationships between operational distance and perceived annoyance, loudness and pitch can be understood, then these relationships can be used to determine acceptable situations where UAV can operate effectively while also mitigating any potentially negative effects of their presence as a sound source.

Stage 3 of the analysis consisted of a multilevel linear regression analysis, to identify the significance of subject-dependent responses of perceived annoyance. Multilevel linear regression has been used previously to investigate the factors contributing to annoyance for rotorcraft and small UAV, and has found to be a useful tool in discovering

key variables (Gwak, et al., 2020) (Boucher, et al., 2020). Multilevel linear regression is a method that integrates no pooling and complete pooling of data between subjects. No pooling would mean that a regression analysis for each subject's response data would be built, meaning that a regression relationship would be described for each subject. Complete pooling suggests an aggregation of all response data, so a regression analysis would build a correlation between the independent data and the response data for the whole subject group. This multilevel regression groups by subject, therefore assuming a partial pooling of the subject data and a normal distribution across subjects of regression. It is first useful to determine whether a multilevel regression model is appropriate for understanding any grouping effects that may be causing variance in the response data. This was done by using a very simple, fixed intercept model with no predictor variables to understand the effect of clustered data on the dependent variable by assessing the intraclass correlation coefficient (ICC), which is a ratio of the variance of the subject-dependent intercept estimates from the simple model, and the sum of this variance with the variance of the fixed intercept value estimated by the model. The model is described by Equations 61 and 62:

$$Y_{ij} = \beta_{0j} + e_{ij} \tag{61}$$

$$\beta_{0j} = \gamma_{00} + u_{0j} \tag{62}$$

Where Y_{ij} is the perceived annoyance of sound i from participant j, β_{0j} is the sum of γ_{00}, the overall mean intercept for all subjects, and u_{0j}, the subject-dependent intercept offset. e_{ij} is the residual error per subject. From this, estimates of the variance of subject-dependent intercept values and the estimate of variance of the fixed intercept value can be used to calculate the ICC. If the ICC calculation yields a statistically significant result, then it can be assumed that clustering effects in the model contribute to the value of the dependent variable, and a more detailed multilevel model should be introduced. The next stage is to introduce predictor variables into the mixed model. A multilevel regression model, with a variable intercept per participant but fixed slopes of sound quality metrics, has a general equation which is described by Equation 63:

$$Y_{ij} = \beta_{0j} + \gamma_{10}X_{1i} + \cdots + \gamma_{n0}X_{ni} + e_{ij} \tag{63}$$

γ_{n0}, which does not vary per subject, is the regression coefficient of the n-th sound quality metric 5th percentile value of sound i. Furthermore, introducing subject-dependent regression slope coefficients for each sound quality metric can reveal more information about the variance between how participants perceive these metrics, but previous literature has found that introducing subject-dependent slopes for this style of sound quality metric analysis yielded little improvement to model accuracy when compared to the increase in accuracy introduced by including subject-dependent intercepts (Boucher, et al., 2020). Therefore, a subject-dependent slope and intercept model was omitted from this analysis. Metrics were removed from this model to investigate the reduction in R^2 between the model's predicted values of annoyance and the measured response values, to determine the significance of each metric as predictor variables.

6.2. Results

6.2.1. Descriptive statistics of response data

It has previously been shown that in research where participants are asked to subjectively assess sound stimuli during an experiment, a large amount of variability can be observed in response data (Torija & Flindell, 2015). Aside from this, the research methodology involved participants using the testing interface from home with their own playback equipment due to the COVID-19 pandemic, which may have introduced a greater degree of variability than if the experiment was carried out in laboratory conditions as initially planned. Therefore, it is deemed appropriate to carry out a statistical analysis of the response data, to ascertain the degree of variability which may have been introduced by the adopted experimental methodology, and to deduce whether the data collected fulfils the requirements for the subsequent analyses. To assess this, the non-parametric statistical value Kendall's coefficient of concordance (Kendall's W) was employed, giving an indication of the agreement between participants of how they

rated each sound stimuli for perceived annoyance, loudness and pitch. Kendall's W is calculated using the following equations.

$$R = \sum_{i=1}^{k} (R_i - \bar{R})^2 \tag{64}$$

$$W = \frac{12R}{m^2(k^3 - k)} \tag{65}$$

Where R_i is the sum of the ratings given for subject i, k is the total number of subjects, j is the current rater, and m is the total number of raters. $\bar{R}$ is the mean of all values of R_i. A value of W of 1 indicates the participants have been unanimous in their rating of each subject, and a value of 0 indicates no agreement between participants. Table 4 shows the calculated values of Kendall's W for the response values gathered from the participants. Perceived annoyance and loudness showed strong agreement between participants, with Kendall's W values of greater than 0.6. Perceived pitch showed statistically significant agreement, although not to the same magnitude as perceived annoyance and loudness. These values indicate that the response values gathered from the participants have appropriate variance necessary to continue with the subsequent analyses.

Table 4: Kendall's W values calculated for the response data

Stastical Value	Perceived Annoyance	Perceived Loudness	Perceived Pitch
Kendall's W	0.60	0.64	0.41
Significance	<0.001	<0.001	<0.001

6.2.2. Correlation analysis

A simple correlation analysis looked at regression relationships between the calculated sound quality metrics and the response data gathered during online subjective test 1. These results were presented in the proceedings of Inter-Noise 2021 (Nicholls & Torija, 2021). Firstly, the sound quality metrics and dependent variables (responses of perceived annoyance, loudness and pitch) were regressed against each other to illustrate any obvious correlations between variables, and the correlation coefficients are presented in Table 5. Some of the sound quality metrics yield a strong correlation with the response values given by the test participants. For perceived annoyance, the most significant correlations were those with $L_{Aeq,4s}$, PNL, loudness, sharpness and fluctuation strength, indicating that these sound quality metrics may be good candidates for predicting perceived annoyance in a more developed regression model. Furthermore, perceived annoyance also correlated strongly with perceived loudness and perceived pitch. This result would suggest that participants thought that the loudness and frequency content relating to pitch were significant factors influencing annoyance. Perceived loudness correlated strongly with $L_{Aeq,4s}$, PNL, loudness, sharpness and fluctuation strength, meaning that these metrics could aid in controlling the perceived loudness of UAV noise. Perceived pitch correlated strongly with PNL, loudness, sharpness, tonality, roughness and impulsiveness, meaning these metrics may be useful in quantifying how the frequency content of UAV noise could be manipulated to reduce perceived annoyance.

When comparing the metrics quantifying the sensation of loudness, those being $L_{Aeq,4s}$, PNL and loudness, it can be seen in Table 5 that the strongest predictor for perceived annoyance and perceived loudness is PNL, followed by loudness, and then $L_{Aeq,4s}$. Furthermore, $L_{Aeq,4s}$ did not yield a significant correlation with perceived pitch, whereas PNL and loudness did. These results suggest that L_{Aeq} may not be the most suitable metric for quantifying the sensation of loudness, and therefore PNL was carried through to the next stages of analysis for part 1. Using these metrics, simple linear regression models were built to predict response values of perceived annoyance, loudness and pitch using a backward stepwise method to remove underperforming metrics. The first iteration of the model included all metrics, and was correlated to perceived annoyance with an adjusted R^2 of 0.935. The final iteration of this model included PNL, sharpness and impulsiveness. The annoyance model statistics of each iteration during the backwards stepwise method can be seen in Table 6. The predictor variables for this model, PNL, sharpness and impulsiveness had standardised coefficients of 0.696, 0.289 and -0.088, respectively.

Table 5: Correlation coefficients between 5th percentile sound quality metrics and response values

		LAeq,4s	PNL	Loudness	Sharpness	Fluctuation strength	Tonality	Roughness	Impulsiveness	Perceived Annoyance	Perceived Loudness	Perceived Pitch
Perceived Annoyance	Adjusted R Squared	0.810	0.956	0.899	0.899	0.401	0.254	0.193	-0.160		0.960	0.614
	Significance	<0.001	<0.001	<0.001	<0.001	0.007	0.097	0.210	0.298		<0.001	<0.001
Perceived Loudness	Adjusted R Squared	0.820	0.983	0.921	0.871	0.467	0.150	0.290	-0.083	0.960		0.438
	Significance	<0.001	<0.001	<0.001	<0.001	0.001	0.331	0.056	0.594	<0.001		0.003
Perceived Pitch	Adjusted R Squared	0.190	0.467	0.501	0.549	-0.003	0.477	-0.337	-0.368	0.614	0.438	
	Significance	0.250	0.001	0.001	<0.001	0.983	0.001	0.025	0.014	<0.001	0.003	

Then, a simple regression model was built to correlate loudness with the sound quality metrics. It was found that PNL and fluctuation strength were strong predictor metrics for perceived loudness, which in turn can be seen to correlate strongly with perceived annoyance. The iteration statistics for the simple loudness model can be seen in Table 8. The final iteration of the loudness model, including PNL and fluctuation strength as predictor variables, had an adjusted R^2 of 0.973. The chosen predictor variables for this model are both measures of sound magnitude phenomena, and had standardised regression coefficients of 0.946 and 0.094, respectively. This result highlights the fact that PNL plays a significant role in quantifying the perceived loudness of UAV noise.

Finally, a simple regression model was built to correlate perceived pitch with the sound quality metrics. The final iteration of this model included sharpness, roughness and tonality, which are all metrics that represent frequency-based phenomena. This model had a final adjusted R^2 of 0.585. The iteration statistics for the simple pitch model can be seen in Table 10. Sharpness, tonality and roughness had standardised regression coefficients of 0.563, 0.320 and -0.452, respectively. This shows that these metrics all play a similarly significant role in quantifying the perceived pitch experienced by the subjective test participants.

Table 6: Simple annoyance model iteration statistics

Model Iteration	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	Metric Removed	Final Metrics Included
1	0.944	0.935	0.0429	0.944		
2	0.943	0.935	0.0429	-0.001	Tonality	
3	0.941	0.935	0.0429	-0.002	Fluctuation Strength	
4	0.939	0.935	0.0432	-0.002	Roughness	PNL, Sharpness, Impulsiveness

Table 7: Final simple annoyance model iteration predictor coefficients

Model Iteration	Predictor	Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
4	(Constant)	-0.818	0.100		-8.194	<0.001
	PNL	0.014	0.002	0.696	8.719	<0.001
	Sharpness	0.065	0.018	0.289	3.631	0.001
	Impulsiveness	-0.181	0.081	-0.088	-2.240	0.031

Table 8: Simple loudness model iteration statistics

Model Iteration	R Squared	Adjusted R Squared	Std. Error of the Estimate	R Square Change	Metric Removed	Final Metrics Included
1	0.976	0.973	0.0429	0.976		
2	0.976	0.973	0.0429	0.000	Impulsiveness	
3	0.976	0.974	0.0429	0.000	Tonality	
4	0.976	0.974	0.0432	0.000	Roughness	
5	0.974	0.973	0.0299	-0.001	Sharpness	PNL, Fluctuation Strength

Table 9: Final simple loudness model iteration predictor coefficients

Model Iteration	Predictor	Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
5	(Constant)	-1.354	0.049		-27.688	<0.001
	PNL	0.021	0.001	0.946	34.678	<0.001
	Fluctuation Strength	0.612	0.177	0.094	3.452	0.001

Table 10: Simple pitch model iteration statistics

Model	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	Metric Removed	Final Metrics Included
1	0.668	0.615	0.0864	0.668		
2	0.655	0.610	0.0870	-0.013	PNL	
3	0.636	0.598	0.0883	-0.020	Fluctuation Strength	
4	0.614	0.585	0.0897	-0.022	Impulsiveness	Sharpness, Tonality, Roughness

Table 11: Final simple pitch model iteration predictor statistics

Model	Predictor	Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
4	(Constant)	0.129	0.074		1.741	0.089
	Sharpness	0.105	0.019	0.563	5.382	<0.001
	Tonality	0.470	0.150	0.320	3.135	0.003
	Roughness	-1.414	0.316	-0.452	-4.474	<0.001

6.2.3. Multilevel linear model analysis

A simple, fixed intercept model was built to test whether any variance in the response values of perceived annoyance is controlled by grouping or clustering effects within the data. This simple model is created following Equations 61 and 62. The estimated fixed intercept for the simple model is shown in Table 12, and is the mean average of all responses of perceived annoyance for each individual UAV stimuli. Using Equation 57, and the values given in Table 12, the ICC was calculated to be 0.185, which means that 18.5% of the variance in perceived annoyance is explained by the individual participant, leaving the rest of the variance to be explained by other factors, such as the calculated sound quality metrics related to each stimulus. Since the UAV stimuli vary largely in each sound quality metrics, this is to be expected, and gives a good case for using a more complex multilevel model to investigate any clustering or grouping effects within the data. The next multilevel model built included the sound quality metrics, with varying y-axis intercepts but fixed regression slopes across participants.

Table 12: Estimates of fixed and covariance parameters for fixed intercept model with no predictor variables

		Estimate	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Fixed	Intercept γ_{00}	0.6088	0.0162	<0.001	0.0576	0.0641
Covariance Parameters	Residual $\sigma^2_{\varepsilon ij}$	0.0514	0.0016	<0.001	0.0484	0.0546
	Intercept σ^2_{u0j}	0.0117	0.0026	<0.001	0.0075	0.0182

Table 13 shows the fixed effect statistics, or the statistics of the sound quality metrics used in the mixed model, with fixed regression slopes but variable intercepts per participant. PNL, sharpness, fluctuation strength, roughness and impulsiveness all were deemed statistically significant predictors in the fixed slope, variable intercept model. Previously, in the simple linear model of perceived annoyance, only PNL, sharpness and impulsiveness were deemed statistically significant predictors for perceived annoyance. Fluctuation strength and roughness have now become more significant after using variable y-axis intercepts per participant, which could mean that relationships between these metrics and perceived annoyance vary between participants. The metrics roughness and impulsiveness have negative estimates, meaning that for roughness, as the metric increases by a value of 1 asper, the predicted perceived annoyance decreases by 0.199. Similarly, as impulsiveness increases by a value of 1, the predicted perceived annoyance from the model decreases by 0.121. Figure 19 shows the reduction in R^2 value per each metric removed from the mixed model with subject-dependent y-axis intercepts, but fixed regression slopes. PNL shows to have the largest reduction in R^2 value when removed from the model, followed by sharpness at a much smaller magnitude. Fluctuation strength, tonality, roughness and impulsiveness reduce the R^2 value of the model by a negligible amount when these metrics are removed while PNL is included.

Table 13: Estimates of fixed effects for multilevel model with subject-dependent intercepts and fixed regression slopes

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	-0.877	0.060	1922	-14.641	<0.001	-0.995	-0.760
PNL	0.014	0.001	2101	15.780	<0.001	0.013	0.016
Sharpness	0.060	0.010	2101	6.183	<0.001	0.041	0.080
Fluctuation strength	0.275	0.134	2101	2.055	0.040	0.013	0.537
Tonality	0.077	0.040	2101	1.908	0.056	-0.002	0.156
Roughness	-0.199	0.083	2101	-2.400	0.016	-0.362	-0.036
Impulsiveness	-0.121	0.047	2101	-2.588	0.010	-0.212	-0.029

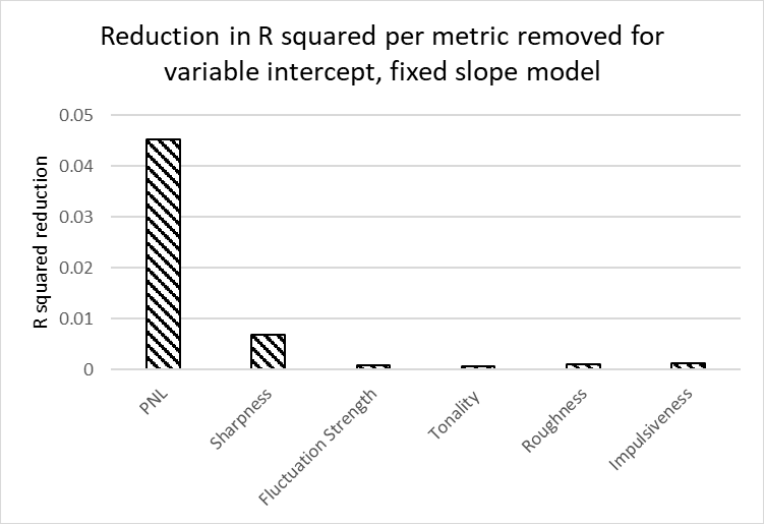

Figure 19: Reduction in R^2 value per metric removed for mixed model using variable intercepts and fixed regression slopes per participant

6.2.4. Perceived responses as a function of distance

The UAV stimuli which represented the UAV performing flyover operations were used to assess how operational distance could be deemed to be a significant factor in introducing UAV into logistical services. The average values of perceived annoyance, loudness and pitch were taken from the subjective experiment and plotted against operational distance. Perceived annoyance yielded a very strong logarithmic correlation with operational distance, with an R^2 value of 0.7408, as did perceived loudness, with an R^2 value of 0.8158. Although perceived pitch does yield a negative, logarithmic correlation with operational distance, larger residual distances from the trendline can be observed when compared to the plots of perceived annoyance and perceived loudness against distance. As operational distance increases, you would expect a decrease in sharpness to come into play as air absorption effects occur, which mainly attenuates higher frequency content. Considering the typical frequency content of UAV noise being high frequency, and the number of UAV that could potentially be operating at a given time in an urban environment, it would be crucial to use operational distance as a controlling factor to mitigate any adverse effects on public health.

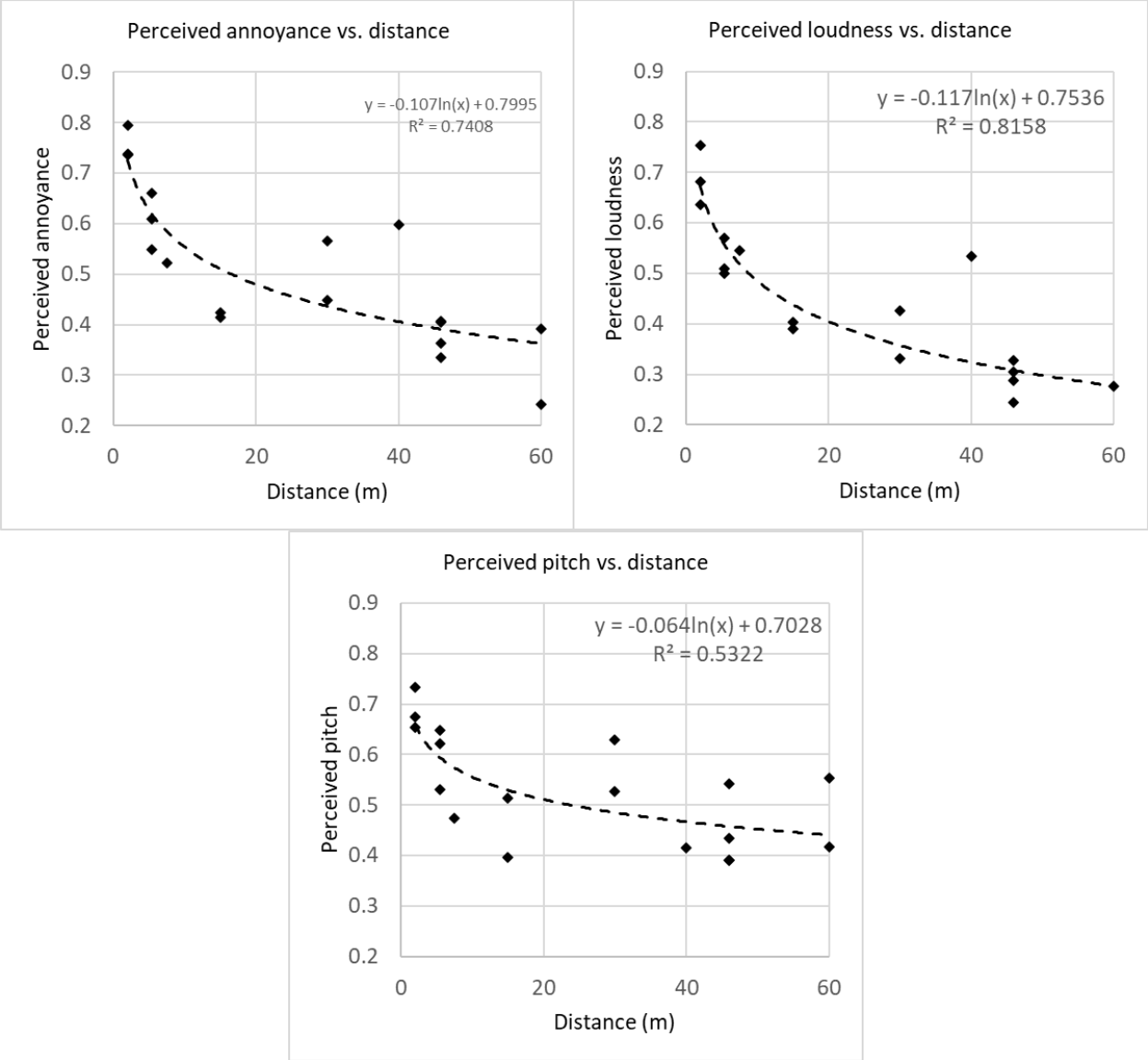

Figure 20: Plots of average perceived annoyance, loudness and pitch against distance for UAV stimuli performing flyover operations

7. PART 2 – IMPACT OF UAV NOISE ON URBAN SOUNDSCAPE PERCEPTION

7.1. Methodology

The second part of this research aims to quantify the change in perception of urban soundscapes when UAV noise stimuli are introduced, and relate the key characteristics of UAV noise which contributes to this change in perception with the mechanical functions typically employed by these vehicles for flight. The methodology for the second part of this research adopts a similar approach as the first. An online subjective experiment was carried out, to gather response data from participants assessing UAV stimuli in the context of various urban soundscapes. A smaller group of the UAV stimuli from the first part of the research were selected that varied in operation, UAV size and operational distance. These stimuli were then introduced into urban soundscapes which were recorded in urbanised areas. The effect each soundscape had on the perception of the UAV stimuli was quantified by calculation critical-band rate specific sound quality metrics for each UAV and soundscape stimulus, and then calculating the difference in each specific sound quality metric imposed by the introduction of the UAV into the soundscape. The specific sound quality metric differences were then used in principal component analysis (PCA) to find key bandwidths for each sound quality metric. The principal components found using PCA were then used in a regression analysis to give insight into how these bandwidths could potentially control human response.

7.1.1. UAV and soundscape stimuli

20 UAV stimuli were chosen from the previously gathered sound used in the first part of the research. The stimuli chosen to cover the various operations such as flyovers, take-offs, hovering, landing, and mid-flight manoeuvres at distances from 1.2 m to 60 m. 6 soundscape environments were recorded in urban areas including city squares, canal paths, parks and pedestrianised streets, each being 10 s long. The 20 UAV stimuli were then combined with these 6 soundscape recordings to create 120 soundscape stimuli, each being a combination of a UAV and soundscape. The 20 UAV and 6 soundscape stimuli used are presented in Tables 14 and 15. The soundscape stimuli were also calibrated as described in Section 6.1.1, to maintain appropriate relative levels between the soundscapes during playback in the online subjective experiment.

The differing soundscape environments will be assessed for their ability to potentially mask unwanted spectral characteristics of UAV which could contribute to negative human response. UAV typically have complicated spectral content, created by fundamental frequencies of their sets of rotors. This is shown in Figure 21. In busier, more populated areas, where noise from various sources is introduced, the spectral content of background noise is more complicated. Soundscape environment 3, for example, includes a busy pedestrianised area, which is dominated by pedestrians talking, and includes distant traffic and occasional noise from advertisements over PA systems. In comparison to soundscape environment 4, which captures a quiet city park with distant road traffic noise, soundscape environment 3 has the potential to mask more spectral content of UAV. This is illustrated in Figure 22, which shows how a UAV hovering at 40 m introduces an obvious spectral pattern into soundscape environment 4, whereas the same UAV does not create as significant a pattern in soundscape environment 3.

Table 14: Environments used in soundscapes

Environment Sound No.	Environment Description
1	Canal scene with train noise
2	Harehills Ln, road traffic noise
3	Millenium Square, pedestrian noise and adverts
4	Park with distant road traffic noise
5	Peel Park, early morning
6	Market Str, very busy with pedestrians

Table 15: UAV stimuli used in soundscapes

UAV Sound No.	UAV Model	UAV Weight (kg)	UAV Operation	UAV Distance (m)
1	DJI Inspire	2.85	Flyover	15
2	DJI Inspire	2.85	Flyover	7.5
3	DJI Inspire	2.85	Landing	7.5
4	DJI Inspire	2.85	Take-off	2
5	DJI Matrice 600	9.1	Hover	40
6	DJI Matrice 600	9.1	Flyover	40
7	DJI Mavic	0.743	Flyover	15
8	DJI Mavic	0.743	Flyover	30
9	DJI Mavic	0.743	Flyover	60
10	DJI Mavic	0.743	Manoeuvring	7.5
11	DJI Mavic	0.743	Take-off	7.5
12	DJI Matrice 200	4	Flyover	46
13	DJI Matrice 200	4	Flyover	46
14	DJI Matrice 200	4	Take-off	30
15	DJI Matrice 200	4	Landing	30
16	DJI Matrice 200	4	Hover	1.2
17	Gryphon GD28X	11.8	Take-off	30
18	Gryphon GD28X	11.8	Landing	30
19	Gryphon GD28X	11.8	Manoeuvring	30
20	Gryphon GD28X	11.8	Hover	1.2

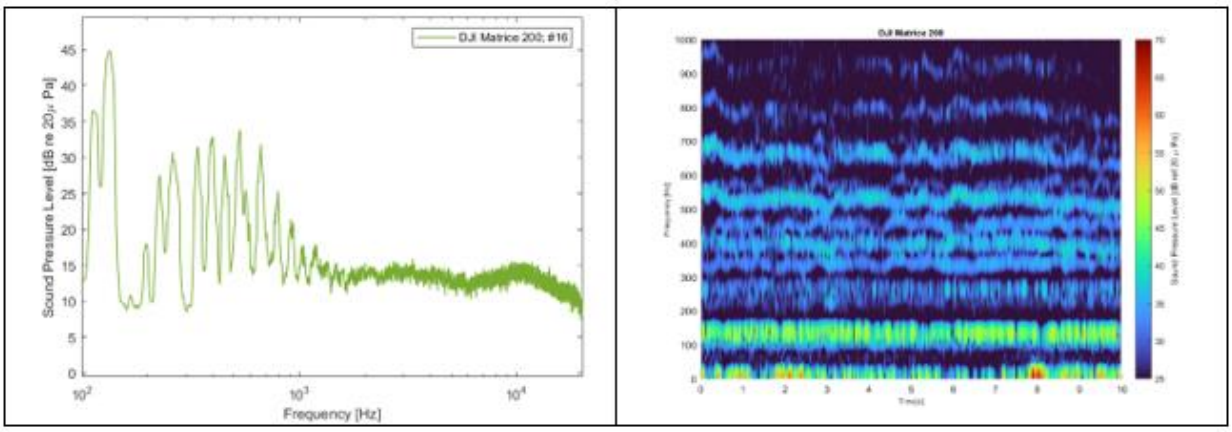

Figure 21: Frequency spectra and spectrogram for the DJI Matrice 200 hovering at 1.2 m from microphone

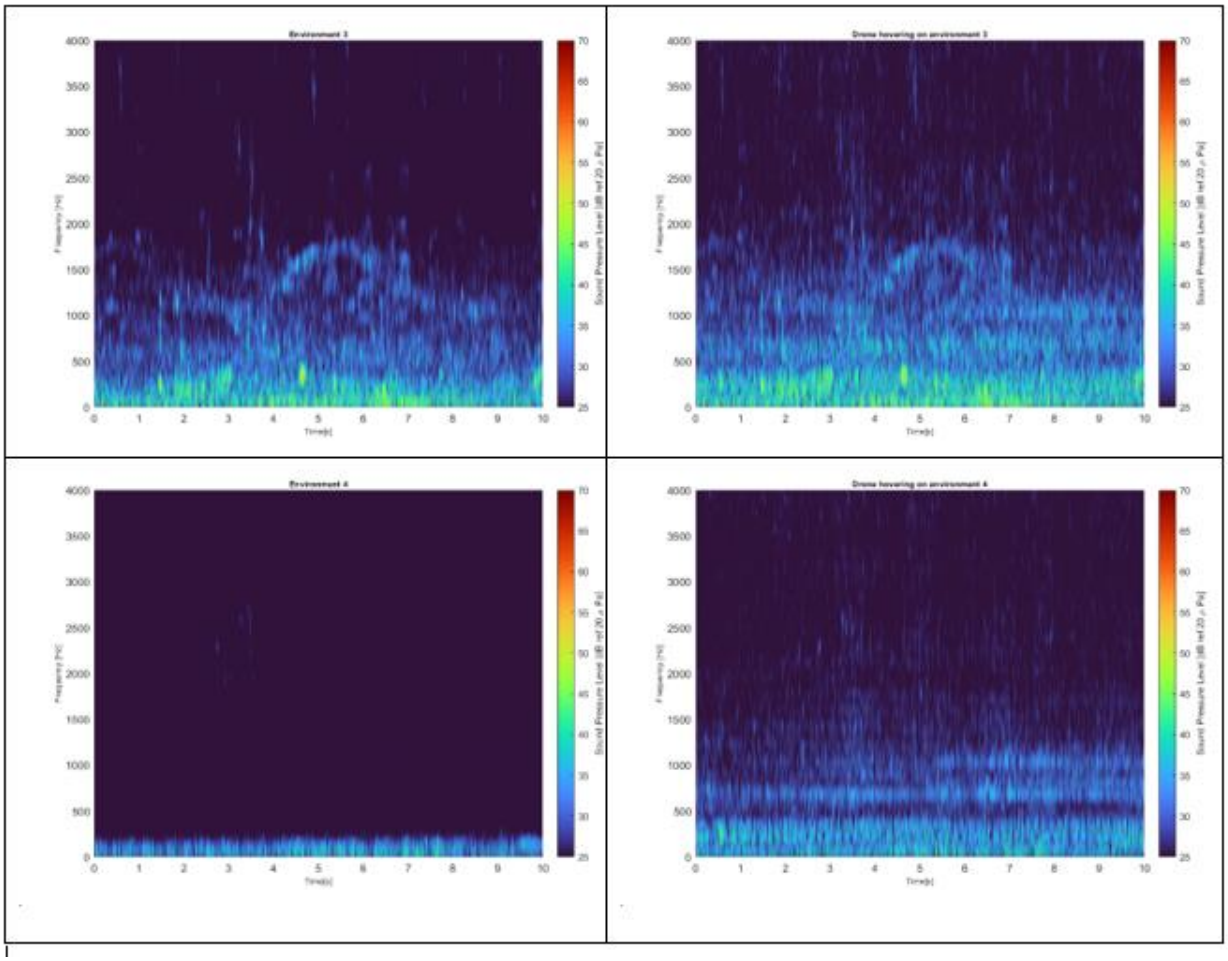

Figure 22: (Top) Spectrogram of Environment 3 without (left) and with (right) DJI Matrice 600 hovering at 40 m from the microphone; (Bottom) Spectrogram of Environment 4 without (left) and with (right) DJI Matrice 600 hovering at 40 m from the microphone

7.1.2. Online subjective experiment

For the second online subjective experiment, a similar methodology was used as the first part of the research. Participants were asked to give response values of perceived annoyance, loudness, UAV dominance and soundscape pleasantness. These response metrics were rated using sliders. The loudness slider had extremes of "very quiet" and "very loud", the annoyance slider had extremes of "not very annoying" and "very annoying", and the drone dominance slider had extremes of "not very dominant" to "very dominant". The pleasantness slider had extremes of "not very pleasant" to "very pleasant". The participant could listen to each soundscape as many times as they required to make their decision on each response metric, and then continue to the next soundscape. The order of the soundscapes was randomised for each participant. The response values were saved anonymously on a numerical scale from 0 to 1, corresponding to the position of the slider placement for each response metric. The experiment took around 40 minutes in total, and participants had the option to take a 5-minute break halfway through the stimuli, to mitigate the effects of fatigue. The order of the soundscape stimuli was randomised for each test participant, and the participants could listen to the sounds as many times as required for them to make their choices.

7.1.3. Analysis

Once the response values of perceived annoyance, loudness, UAV dominance and soundscape pleasantness were gathered through the online subjective test methodology, the analysis was undertaken using MATLAB. Similar to the first experiment, sound quality metrics were calculated using the HEAD Acoustics ArtemiS Suite for each soundscape stimuli including UAV noise, as well as the soundscapes by themselves. The sound quality metrics calculated were critical-band rate specific values of loudness, tonality, roughness, fluctuation strength, and impulsiveness. Having the metrics calculated for both the stimuli including the UAV noise and the soundscape environments by themselves meant that the difference in each specific sound quality metric values when a UAV stimulus is introduced could be calculated. This approach is summarised in Equation 66, below.

$$\Delta SQM_{s,e,b} = SQM_{s,b} - SQM_{e,b} \tag{66}$$

Where $SQM_{s,b}$ is one of the specific sound quality metrics for a soundscape (i.e., environment + drone), s, at critical-band rate, b, and $SQM_{e,b}$ is the specific sound quality metric for the corresponding environment, e, used in that soundscape. From this, an individual dataset was built for differences of each individual specific sound quality metrics. To assess the significance of each sound quality metric dataset in relation to the response metrics gathered during the online experiment, a principal component analysis was undertaken. This meant that the large datasets of differences in each sound quality metric for each soundscape stimuli across the critical-bands could be interrogated for key critical-bands that explain most of the variance in each dataset. The first 3 principal components were found for each dataset, and the scores which relate the specific sound quality metric values to the principal component. Then, the principal components of each individual sound quality metric dataset were used in linear regression against the response values gathered during the online subjective testing. From this, it can be seen which critical-band rates, as illustrated by the principal component analysis, are potential predictors of perception of UAV stimuli in urban soundscapes, and which critical-band rates would be the most suitable to mask using a soundscape environment.

7.2. Results

7.2.1. Linear regression of SQM datasets against response values

The sound quality metric difference datasets were used in linear regression against the response values of perceived annoyance, loudness, UAV dominance and soundscape pleasantness. Furthermore, the broadband $L_{Aeq,4s}$ difference values were included in the linear regression analysis, to assess the performance of $L_{Aeq,4s}$ as a predictor variable for the above response values. A summary of the results can be seen in Table 16. It can be seen that the $L_{Aeq,4s}$ dataset yielded the largest adjusted R^2 values when correlated with the response metrics perceived annoyance, loudness, and soundscape pleasantness, and the second highest value when regressed against UAV dominance (only surpassed marginally by loudness). However, we know from the first part of this research that the perceived effects of frequency content correlate significantly with the perception of UAV noise. Therefore, it is deemed appropriate to assess the effects of the loudness of discrete frequency ranges (such as critical band-rates, or in relation to the Bark scale), which is impossible by simply using the broadband $L_{Aeq,4s}$ metric. Consequently, the specific loudness metric was used in the subsequent PCA analysis to investigate how specific frequency ranges impact the response metrics.

Aside from the $L_{Aeq,4s}$ dataset, the difference in loudness introduced by the presence of a UAV on a soundscape caused the strongest reaction in the subjective test participants. Perceived annoyance, loudness and UAV dominance correlated to the loudness dataset with adjusted R^2 values of 0.77, 0.74 and 0.75, respectively, with a smaller value of 0.32 for the correlation with soundscape pleasantness. Roughness yielded a significant adjusted R^2 value when regressed against UAV dominance, but less significant values when regressed against 0.38, 0.32 and 0.096. When regressed against perceived annoyance, fluctuation strength yielded an adjusted R^2 of 0.44, but showed less significant results with the other response values, however. Impulsiveness had its highest adjusted R^2 value when regressed against UAV dominance, but did not correlate as strongly with the other response values. Tonality did not perform well when correlated against any of the response values given by the participants.

Table 16: Adjusted R^2 values for regression models between SQM datasets and response values

SQM Dataset	Perceived Annoyance	Perceived Loudness	UAV Dominance	Soundscape Pleasantness
LAeq,4s	0.84	0.88	0.71	0.42
Specific Loudness	0.77	0.74	0.75	0.32
Specific Roughness	0.38	0.32	0.44	0.096
Specific Fluctuation Strength	0.44	0.34	0.42	0.033
Specific Impulsiveness	0.28	0.27	0.37	0.23
Specific Tonality	0.13	0.094	0.19	0.21

7.2.2. Principal component analysis

The principal component analysis of each specific sound quality metric dataset revealed critical-band, or frequency ranges for each sound quality metric that had strong influence on the principal components of each sound quality metric dataset. The results in this section have previously been presented by the author at Inter-Noise 2022 (Nicholls, et al., 2022). The values of the scores have been normalised ($\tilde{t}_{PC_{i,n}}$) for the purpose of data interpretation, using the following equation, where $\left|t_{PC_{i,n}}\right|$ is the absolute magnitude of the score for critical-band rate i, within principal component n, and η_n is the variance explained by principal component n.

$$\tilde{t}_{PC_{i,n}} = \left|t_{PC_{i,n}}\right| \times \eta_n \tag{67}$$

The first dataset interrogated was the loudness dataset. The first principal component found for the loudness dataset explained 84.1% of the variance in the data. This principal component included scores at all critical bands which were of increased magnitude when compared to principal components from later datasets. However, an increase in the principal component scores relating to the critical bands between roughly 2 kHz and 6.4 kHz can be observed, as illustrated in Figure 23. The second principal components yielded much different scores when compared to the first. The second principal component has slightly stronger scores associated with the critical bands at 200 to 300 Hz, and 6.4 to 7.7 kHz. The second principal component explained only 6.5% of the variance in the loudness dataset. The third principal component for loudness explained 3.4% of the variance of the dataset, with scores that show less of a clear pattern when compared to the first two principal components. The scores for the principal components found for the roughness dataset can been in Figure 24. The first principal component explained 70.3% of the variance in

the dataset, and can be seen to have a prominent score increase relating to the critical bands between 1.3 kHz and 5.3 kHz. The second principal component of the sharpness dataset has a similar region of increased scores, but the critical-band rate range is lower, showing an increase in the scores in the frequency range between 400 Hz and 910 Hz, as well as at 3.7 kHz. The second principal explained 13.1% of the variance in the roughness dataset. Finally, the third principal component of the roughness dataset explained 6.1% of the variance in the dataset. This principal component can be seen to have high scores associated with the 100 and 200 Hz critical bands.

The scores for the principal components of the fluctuation strength dataset are presented in Figure 25. The first principal component of the fluctuation strength dataset can be seen to be dominated by the scores associated with the critical-band rates between 200 and 400 Hz, and slight increases in scores between 2.7 to 4.4 kHz. The first principal component explains 58.2 % of the variance in the fluctuation strength dataset, an observably lower value compared to the first principal components of the previous two sound quality metric datasets. The second principal component explains 28% of the variance in the fluctuation strength dataset, and has two controlling critical-band rate regions. The first region is related to the frequency range between 300 Hz to 510 Hz. The second region relates to the frequency range between 3.2 kHz and 5.3 kHz. The third principal component for the dataset explained only 4.9% of the variance in the dataset and did not show any significant scores. Impulsiveness yielded principal component scores as illustrated in Figure 26, with larger scores being achieved in two separate frequency ranges. The first range is between critical-band rates with centre frequencies of 2.3 kHz and 3.2 kHz, and the second range being between critical-band rates with centre frequencies of 6.4 kHz and 9.4 kHz. This principal component explains 70.8% of the variance in the impulsiveness dataset. The second principal component for the impulsiveness dataset explained 8.7% of the variance and had significant scores relating to critical-band rates with centre frequencies from 2.3 kHz to 2.7 kHz, and 3.7 kHz to 6.4 kHz. The third principal component calculated explains 6.5% of the variance in the impulsiveness dataset. The critical-band rate with centre frequency 7.7 kHz had a high score for this principal component. The first principal component for the tonality dataset yielded scores with slightly increased values for critical-band rates with centre frequencies from 200 Hz to 1 kHz, and 6.4 kHz to 7.7 kHzj. The second principal component for this dataset had a very high score for the critical-band rate with centre frequency 7.7 kHz. The third principal component had very large scores for the critical-band rates with centre frequencies between 2 kHz and 3.15 kHz. The principal components of the tonality dataset explained 54.5%, 19.7%, and 11.79% of the variance, respectively. Table 17 below summarises the results of the principal component analysis.

Table 17: Summary of results from PCA

	PC1		PC2		PC3		Total Explained Variance
Specific SQM	**Explained variance (%)**	**Frequency Range (Hz)**	**Explained variance (%)**	**Frequency Range (Hz)**	**Explained variance (%)**	**Frequency Range (Hz)**	
Specific Loudness	84.1	Broadband	6.5	200 - 300, 6.4k - 7.7k	3.4		94
Specific Roughness	70.3	1.3k - 5.3k	13.1	400 - 910, 3.7k	6.1	100 - 200	89.5
Specific Fluctuation Strength	58.2	200 - 400, 2.7k - 4.4k	28	300 - 510, 3.2k - 5.3k	4.9		91.1
Specific Impulsiveness	70.8	2.3k - 3.2k, 6.4k - 9.4k	8.7	2.3k - 2.7k, 3.7k - 6.4k	6.5	7.7k	86
Specific Tonality	54.5	200 - 1k, 6.4k - 7.7k	19.6	100 - 300, 3.7k - 7.7k	11.7	2k - 3.2k	85.8

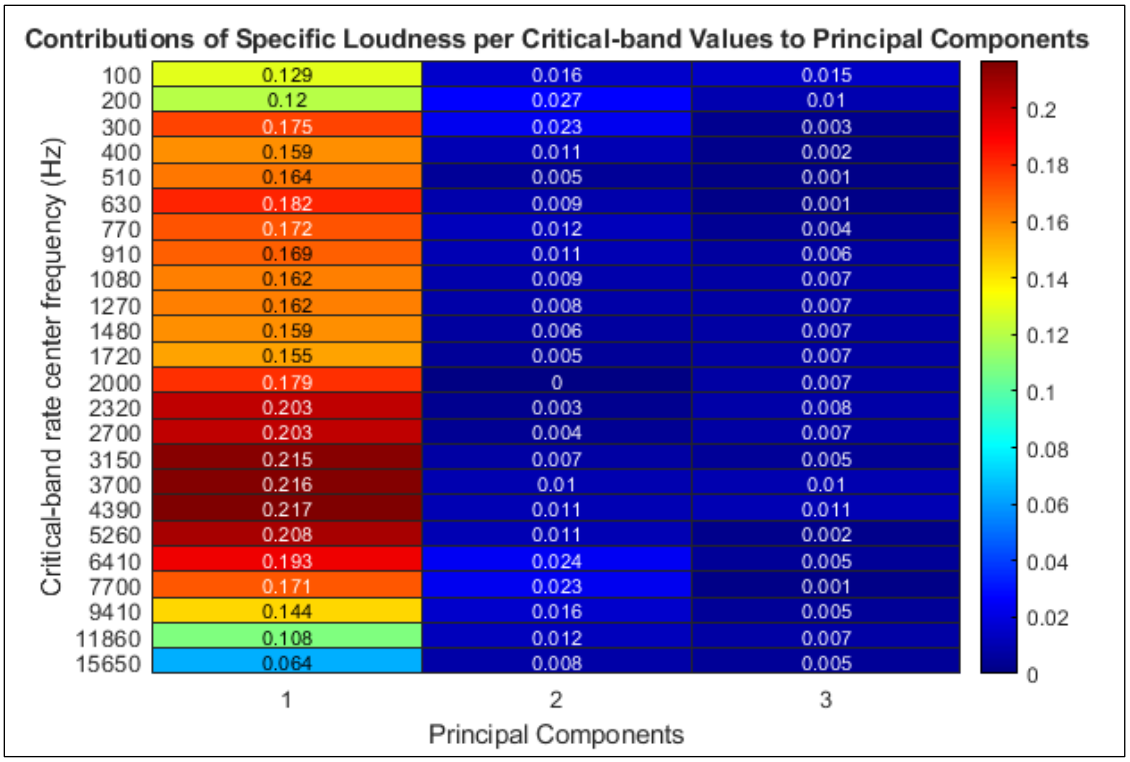

Figure 23: Scores of ΔSQM (Loudness) per critical-band rate for each principal component

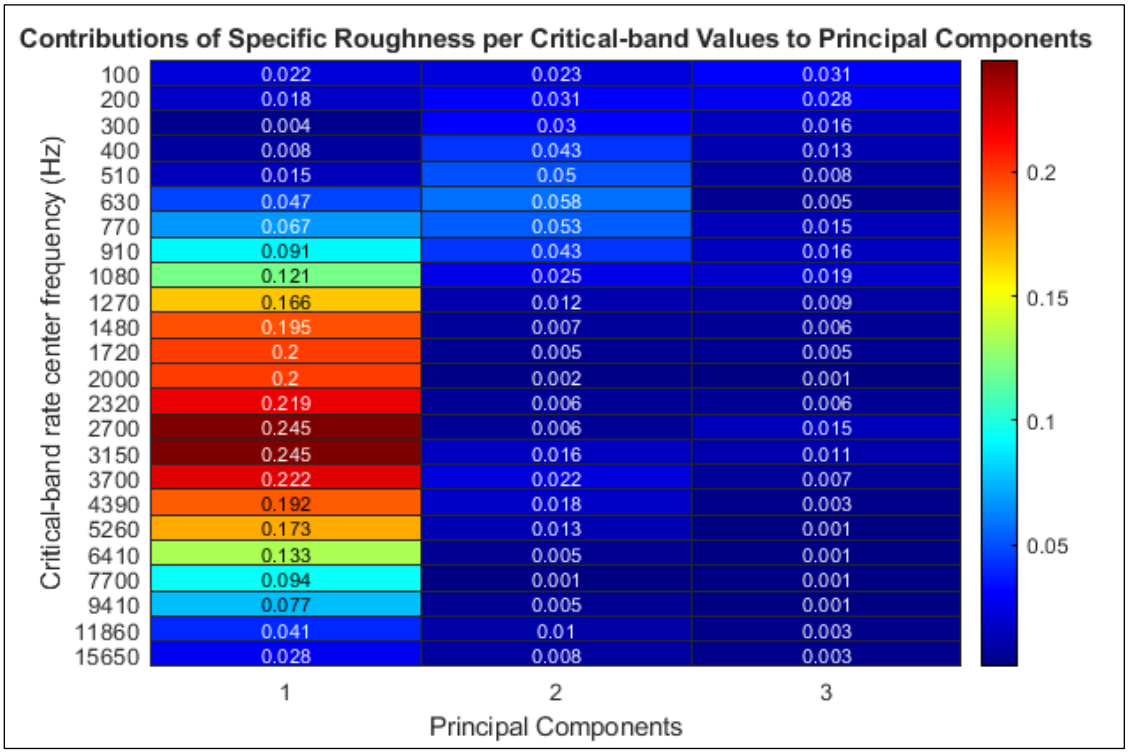

Figure 24: Scores of ΔSQM (Roughness) per critical-band rate for each principal component

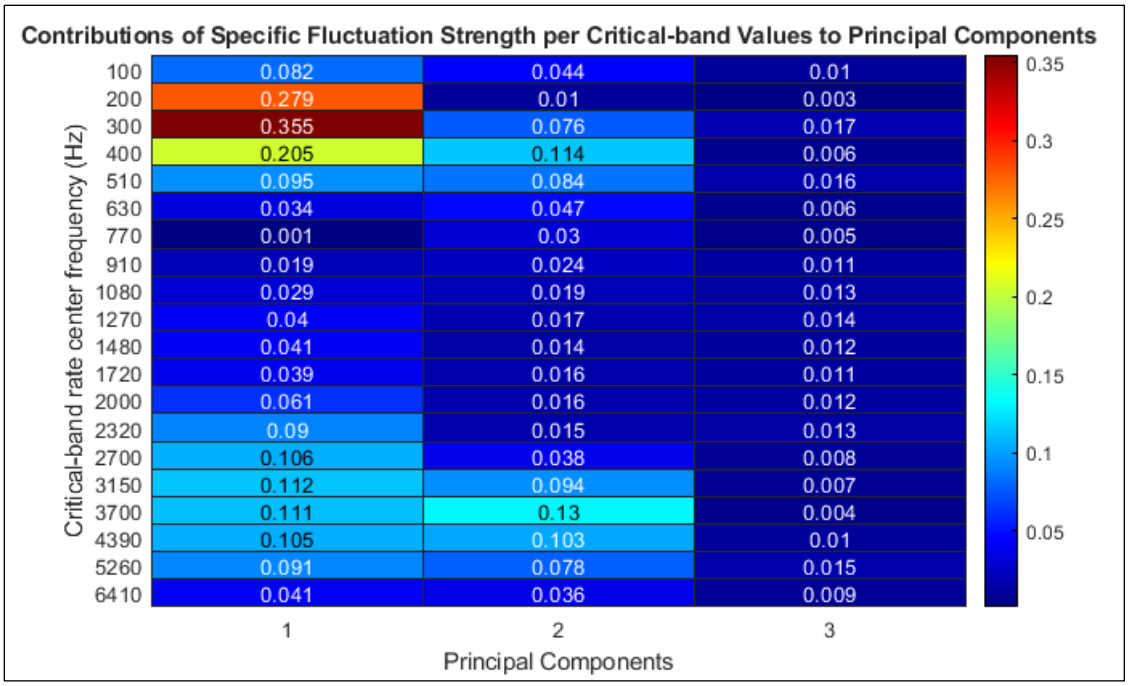

Figure 25: Scores of ΔSQM (Fluctuation Strength) per critical-band rate for each principal component

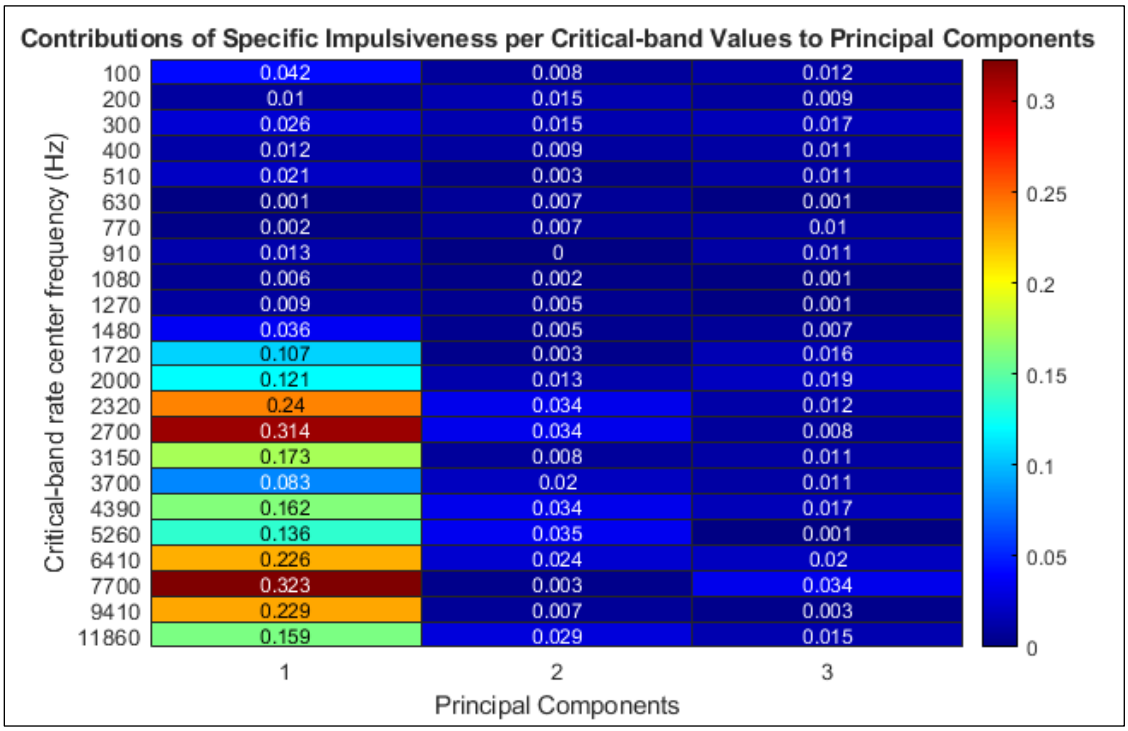

Figure 26: Scores of ΔSQM (Impulsiveness) at each critical-band rate for each principal component

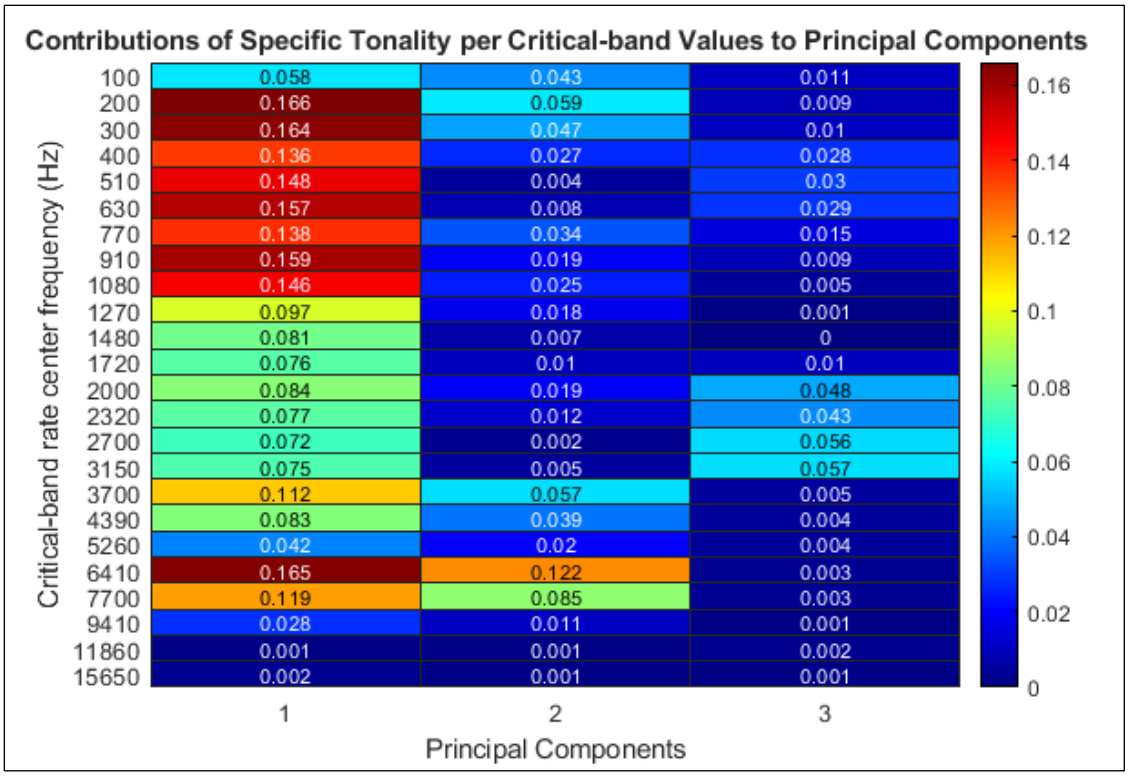

Figure 27: Scores of ΔSQM (Tonality) at each critical-band rate for each principal component

8. DISCUSSION

The results of part 1 of the research are presented in Section 6.2. Initial correlation analysis showed that participants' response values of perceived annoyance correlated strongly with their responses for perceived loudness and pitch. This would suggest that perceived loudness and pitch are key factors in the perception of annoyance for UAV noise, and means of assessing and mitigating the amplitude and frequency phenomena relating to these perceived effects should be developed. For this to be possible, however, it must be understood which phenomena are causing increases in these response metrics. For perceived loudness, the metrics PNL, loudness, sharpness and fluctuation strength were found to have significant correlation statistics. PNL had a larger R^2 value when correlated with perceived loudness compared to the sound quality metric loudness, and hence was chosen instead of loudness in the subsequent regression models to quantify the amplitude phenomena related to perceived loudness. The PNL metric, as developed by Kryter (Kryter, 1960), was designed to assess jet engine noise, and seems to be quantifying the spectral and amplitude characteristics associated with the UAV stimuli effectively. The strong correlation between perceived loudness and sharpness strengthens the case that the perception of both amplitude and frequency phenomena are related, and should both be treated appropriately when trying to mitigate perceived annoyance.

The correlations between perceived annoyance and the calculated sound quality metrics are similar to those between perceived loudness and the sound quality metrics, whereas perceived pitch does introduce some deviations. Perceived pitch yields larger adjusted R^2 values with metrics that quantify frequency phenomena, those being tonality and roughness, as well as impulsiveness. Interestingly, as roughness increases, perceived pitch decreases, which suggests that when more pronounced and less complex tonal content is present in UAV noise, the perceived pitch is higher than when UAV stimuli has a more complex tonal structure, as described in previous research (Torija & Clark, 2021). Similarly, as impulsiveness increases, perceived pitch is predicted to decrease. Literature has previously been inconclusive on whether impulsiveness is a strong predictor variable for perceived annoyance. In cases in literature that have deemed impulsiveness a significant indicator for perceived annoyance, the main source of impulsiveness is accredited to the main rotor blade-vortex interaction (BVI). This is also known as "blade slap", and the effects of this phenomena are more prevalent in larger rotorcraft (Mestre, et al., 2017). For UAV, although related to large rotorcraft by using rotor blades for elevation and propulsion, the scale of the rotor blades may not create the same "blade slap" effect as larger rotorcraft, therefore impulsiveness may need further investigation into how applicable it is as a predictor for UAV noise annoyance.

The results of the backwards stepwise regression model analysis show that for perceived annoyance PNL, sharpness and impulsiveness were the highest performing predictor metrics. Table 6 shows that the reduction in the R^2 per metric removed during the stepwise elimination routine was no larger than 0.002, with the final adjusted R^{2-} value being 0.935, showing a highly significant relationship between this model and perceived annoyance. When analysing the model, it can be seen in Table 7 that PNL (a metric quantifying the perception of loudness) has the highest standardised beta coefficient of 0.696, which is much greater in magnitude than the standardised coefficient of impulsiveness, being -0.088. So, although impulsiveness was chosen to be included in this regression model, its significance is still unclear, and further investigation should be carried out, as stated above. For the regression model predicting perceived loudness, it was found that PNL and fluctuation strength held the most significance following the backward stepwise routine. These metrics quantify the perception of loudness, and the inclusion of fluctuation strength may be quantifying the perceived effects of the interactions between rotor blade fundamental frequencies (even during stationary hover, rotor blades operate at varying RPM, which introduces beating effects), but further research should be carried out to investigate this result. PNL is the main contributor to the prediction of perceived loudness within this model, with a standardised beta coefficient of 0.946, compared to that of fluctuation strength, 0.094. Although impulsiveness is a metric which is used to quantify an amplitude phenomenon, it was removed during the first instance of the backward stepwise regression routine.

The perceived pitch model concluded the backwards stepwise regression routine with the metrics sharpness, tonality and roughness included, with the standardised beta coefficients all within a similar magnitude, as shown in Table 11. Roughness appears in this application to be accounting for complex spectral effects occurring due to the high-frequency interactions of rotor-blades and their corresponding motors, which is supported by previous research

(Torija, et al., 2019). The adjusted R^2 for this model gives a less profound correlation when compared to the perceived annoyance and loudness models however, with a value of 0.585 compared to those of 0.935 and 0.973, respectively. Although the adjusted R^2 value for the perceived pitch model illustrates significance, the decrease in magnitude may have arisen from the ambiguity of the response value, and participant's interpretation of what perceived pitch is meant to be describing. Pitch in a musical theory sense is referring to a relative quantity between musical notes, and therefore requires a reference to make pitch comparisons. Therefore, participants may have been constantly comparing the current UAV stimuli to the previous, introducing the potential for distortion in the response data. However, rephrasing the response value to something more objective, such as "high frequency content", could be less intuitive to a participant group of scientific laypeople.

The multilevel model with varying regression intercepts but fixed slopes revealed interesting relationships between the sound quality metrics and perceived annoyance. It was found that when the regression slopes between participants were allowed to vary, the sound quality metrics fluctuation strength and roughness were of a higher significance to the performance of the model when compared to the backward stepwise regression model, where they were removed before the final iteration. This suggests that the perception of the temporal and frequency phenomena these metrics are quantifying varies between test participants. Figure 19 shows that when PNL is removed from the fixed intercept, variable slope model, the largest reduction in R^2 occurs, proving that the loudness of UAV noise holds the greatest significance for the subjective perception across participants.

From the data presented in Figure 20, it is obvious that the operational distance of UAV will be paramount in mitigating the potential adverse effects of UAV noise on communities. The exponential decrease in perceived annoyance, loudness and pitch when increasing operational distance will be an extremely useful tool for the inception of UAV operations in urban areas, however environmental and logistical restraints could give rise to issues when solely relying upon operational distance as a mitigating factor. For delivery, for example, the very nature of the UAVs operation will be to manoeuvre to locations which are practical spaces for item delivery, and convenient for the end recipient. Previous research, however, has stated that the perceived annoyance of UAV noise does not vary significantly with operational distance (Christian & Cabell, 2017). In this research only one UAV type was assessed, as opposed to the various units analysed throughout this thesis, which could be the cause of the variation in the conclusions. The operational distances of UAV should be investigated using a highly controlled assessment methodology, as the implications could be significant for the adoption of these vehicles in urban infrastructure.

The results of the part 2 of this research illustrated significant findings on the effectiveness of the calculated sound quality metrics to quantify the perceived change in urban soundscapes due to the introduction of UAV. Although the change in $L_{Aeq,4s}$ was seen to correlate well with the response metrics as a broadband metric, it does not give any indication into whether a specific frequency range is controlling the correlation. The first part of the experiment showed that sharpness yielded greater correlation with perceived annoyance, loudness and pitch than $L_{Aeq,4s}$, which suggests that these response metrics are controlled by the level of higher frequency content more so than broadband frequency content. Therefore, due to the discrete nature of specific loudness, being discriminate between critical bands, a more accurate presentation of how specific frequency ranges are contribution to the broadband loudness, and how these frequency ranges impact the response metrics.

Tonality was found to perform poorly when correlating with all perceived response metrics, as can be seen in Table 16. It was initially expected that tonality would be of significance when predicting perceived annoyance for UAV noise, due to the dominance of harmonic tones in a typical UAV signature. This result is consistent with previous research, however (Gwak, et al., 2020), and may imply that either tonality is already being accounted for by loudness due to the tonal dominance of the UAV signatures, or that the tonality calculation used is not appropriately quantifying the perception of the tonal content of UAV noise.

It can be seen that the effects of amplitude phenomena, particularly the overall level of the soundscape as quantified by specific loudness, influences the response values of perceived annoyance, loudness, UAV dominance highly (R^2 values of 0.77, 0.74 and 0.75, respectively), and to a lesser extent soundscape pleasantness. This is supported by previous research, where the calculated loudness of UAV stimuli was found to be the main factor contributing to

perceived annoyance (Gwak, et al., 2020). From the values of percentage variance explained for each of the principal components of the loudness dataset (84.1%, 6.5% and 3.4%), the first principal is explaining the most variance by a significant margin. The scores for this principal component suggest that the effect of loudness on perceived annoyance, loudness and drone dominance is largely broadband, and the whole frequency spectrum holds accountability for mitigating negative perception.

The first principal component calculated for roughness (70.3% variance explained) showed significant scores for critical bands with centre frequencies between 1.3 kHz and 5.3 kHz. This mid to high frequency range has been previously related to the complex spectral fluctuations introduced into UAV signatures through micro-adjustments of the rotor blades controlled by the onboard flight computer system (Torija, et al., 2019). These micro adjustments are primarily to counteract adverse weather conditions during flight, and are essential to the successful operation of UAV (Alexander, et al., 2019). This result therefore suggests that roughness could be used to account for these interactions, typically being generated at harmonics of the blade passing frequencies (BPFs) of the rotors. The strongest correlation between specific roughness and the response values was with perceived UAV dominance, yielding an adjusted R^2 values of 0.41. Although this value is smaller in comparison to the R^2 values of specific loudness, it is still statistically significant, and shows that specific roughness could be quantifying the perception of how dominant a UAV signature is over an environment's soundscape.

The first principal component for the fluctuation strength dataset explained 58.2% of the variance, and had high scores at critical bands with centre frequencies of 200 Hz to 400 Hz. Previously, fluctuation strength has been attributed to quantify the low frequency amplitude modulation caused by rotor blade beating effects at rotor BPFs (typically between 200 Hz and 400 Hz) and could even be contributed to up to frequencies relative first harmonics of the BPF (typically 400 Hz to 630 Hz) (Torija, et al., 2022). These upper harmonics may be illuded to by the scores of the second principal component for fluctuation strength, with higher scores for critical band with centre frequencies between 300 Hz and 630 Hz. This principal component also explained 28% of the variance in the dataset, which is the highest value attained by a second principal component in this analysis. The principal component analysis for impulsiveness showed a notable correlation with perceived UAV dominance, yielding an R^2 value of 0.35, however the correlation statistics with perceived annoyance and loudness were less conclusive. The first principal component explained 70.8% of the variance in the impulsiveness dataset, with scores in two separate critical band ranges being significant. These ranges were critical bands with centre frequencies between 2.3 kHz and 3.15 kHz, and between 6.41 kHz and 9.41 kHz. As previously stated, previous research investigating the effectiveness of impulsiveness to quantify aviation noise characteristics is yet to be resolving, however when assessing the metric in the context of UAV noise, it has been shown that impulsiveness may be able to account for the perceptual effects caused by blade-vortex interaction (BVI) (Krishnamurthy, et al., 2018) (Torija, et al., 2021).

9. CONCLUSIONS

This thesis describes research which aimed to evaluate UAV noise spectra, and quantify the key characteristics of UAV noise that contribute towards perceived annoyance. Two subjective online experiments were carried out, the first of which was to gather response values of perceived annoyance, loudness and pitch from participants evaluating isolated UAV stimuli captured during various operational parameters. The second was to gather response values of perceived annoyance, loudness, UAV dominance and soundscape pleasantness from participants evaluating UAV noise in the context of the urban environment soundscapes. A regression analysis was carried out on the data from the first subjective experiment, including simple correlation analysis of calculated sound quality metrics and the response values, a linear regression analysis, and a multilevel linear model analysis. Using the UAV and the soundscape stimuli from the second subjective experiment, the change introduced in specific sound quality metric values by UAV noise was calculated and used in principal component analysis to identify key frequency ranges which are significant to the perception of temporal and spectral phenomena generated by UAV operational mechanisms.

From the analysis of the data from the first subjective experiment, it was found that PNL correlated with perceived annoyance, loudness and pitch stronger than the loudness metric, so was used to quantify the perception of broadband level for the subsequent analysis. The linear regression analysis for the perceived annoyance response data showed that PNL was the most significant metric for predicting perceived annoyance, followed by sharpness. Impulsiveness was included by the backwards stepwise regression routine for this model, but the significance of this result is unconclusive. PNL and fluctuation strength were significant predictors for perceived loudness. The metrics sharpness, tonality and roughness were found to correlate well with the perceived pitch response values, and are quantifying complex spectral phenomena associated with UAV noise. Furthermore, perceived pitch and loudness correlate directly with perceived annoyance, and should be considered when mitigating negative community response to UAV noise. The multilevel linear regression analysis concluded that PNL, sharpness, roughness and fluctuation strength were statistically significant metrics for the prediction of perceived annoyance of UAV noise. It was observed that perceived annoyance, loudness and pitch decreased exponentially with operational distance. This has implications for the operational convenience of UAV, and research should investigate further the practicalities of increasing UAV operational distance from communities.

The principal component analysis of the data from the second subjective experiment found that the loudness metric was quantifying the effects of the broadband amplitude level of UAV noise, and correlated strongly with perceived annoyance, loudness and UAV dominance. This illustrates the significance of assessing the amplitude of UAV noise when predicting community impact, which has been a recurring result throughout the analyses of this research. Roughness was found to have significance in the frequency ranges between 1.3 kHz and 5.3 kHz, which appears to be quantifying the complex spectral content introduced by rotor blade micro-adjustments, typical of UAV flight in adverse weather conditions. Roughness correlated with statistical significance to perceived UAV dominance. The first two fluctuation strength dataset principal components highlighted the frequency ranges of 200 Hz to 400 Hz, and 300 Hz to 630 Hz (which are typically the regions of fundamental and first harmonics of UAV rotor BPFs) to be of statistical significance. The principal component analysis for impulsiveness showed the frequency ranges between 2.3 kHz and 3.15 kHz and between 6.41 kHz and 9.41 kHz to be of statistical significance to the first principal component, and the metric correlated significantly with perceived UAV dominance. The effectiveness of impulsiveness for quantifying temporal features of UAV noise should be investigated further, due to the inconclusive results between previous research, and within the analyses described in this thesis.

10. FURTHER WORKS

This research gives a basis on which the development of UAV noise metrics can be built. An immediate extension of this research would be to build numerical predictor equations which contain the most significant specific sound quality metrics at key Bark values as predictor variables. A new dataset of UAV noise samples would be gathered (ideally through measurement rather than through peer donation, which this research was limited to due to COVID-19) and a similar subjective experiment to the methodology detailed in this research would be carried out to gather response data for a validation dataset. This validation dataset would be used to investigate the suitability of the created predictor equations.

Beyond this, the introduction of more complex reproduction methods to a subjective experiment involving UAV noise could give insight into how UAV noise perception could be influenced by more realistic operational or environmental characteristics. For example, a visual representation of UAV distance and flightpath using a virtual reality (VR) headset may alter the participants' perception of the UAV. Furthermore, the use of more elaborate audio reproduction such as ambisonic speaker systems could create a sense of immersion which influences a participants' perception. Giving the participants a description of a UAV operation, including the reason for the UAV being present, before playing the VR/ambisonic could alter a participant's attitude toward the UAV's presence in the soundscape environment. Any means of decreasing negative perception of UAV should be investigated, as it may aid in implementing UAV operations into urban infrastructure.

11. REFERENCES

Alexander, W. N. & Whelchel, J., 2019. *Flyover Noise of Multi-Rotor sUAS.* Madrid, Internoise .
Alexander, W. N., Whelchel, J., Intaratep, N. & Trani, A., 2019. *Predicting Community Noise of sUAS.* Delft, AIAA/CEAS Aeroacoustics Conference .
Aures, W., 1984. *Berechnungsverfahren fur den Wohlklang beliebiger Schallsignale, ein Beitrag zur gehorbezogenen Schallanalyse,* Munchen: PhD Thesis.
Aures, W., 1985. Berechnungsverfahren für den sensorischen Wohlklang beliebiger Schallsignale (A procedure for calculating sensory pleasantness of various sounds). *Acta Acustica united with Acustica,* 59(2), pp. 130-141.
Boucher, M., Krishnamurthy, S., Christian, A. & Rizzi, S., 2020. Sound Quality Metric Indicators of Rotorcraft Noise Annoyance using Multilevel Regression Analysis. *Journal of the Acoustical Society of America ,* 36(1).
Boucher, M., Krishnamurthy, S., Christian, A. & Rizzi, S. A., 2019. Sound quality metric indicators of rotorcraft noise annoyance using multilevel. *Proceedings of Meetings on Acoustics,* 36(1).
Burchan, A., 2019. *Public acceptance of drones: Knowledge, attitudes, and practice.* 59 ed. s.l.:Technology in Society.
Cabell, R., McSwain, R. & Grosveld, F., 2016. *Measured Noise from Small Unmanned Aerial Vehicles.* Providence, NOISE-CON.
Christian, A. W. & Cabell, R., 2017. *Initial Investigation into the Psychoacoustic Properties of Small Unmanned Aerial System Noise.* Denver, AIAA Aviation Technology, Integration and Operations Conference.
Davies, P., 2007. Perception-Based Engineering: Integrating Human Responses into Product and System Design. *The Bridge,* 37(3), pp. 18-24.
Di, G.-Q.et al., 2016. Improvement of Zwicker's psychoacoustic model aiming at tonal noises. *Applied Acoustics,* Volume 105, pp. 164-170.
Faslt, H., Volk, F. & Straubinger, M., 2009. *Standards for calculating loudness of stationary or time-varying sounds.* Ottawa, Inter-Noise.
Federal Aviation Agency, 2017. *14 CFR Part 36 - Noise Standards: Aircraft type and airworthiness certification.* s.l.:U.S. Department of Transport.
Genuit, K., Sottek, R. & Fiebig, A., 2009. *Comparison of Loudness Calculation Procedures in the Context of Different Practical Applications.* Ottawa, Canada, Inter-Noise.
Greenwood, E., Brentner, K. S., Rau, R. F. & Ted Gan, Z. F., 2022. *Challenges and opportunities for low nmoise electric aircraft,* s.l.: s.n.
Gwak, D. Y., Han, D. & Lee, S., 2020. Sound quality factors influencing annoyance from hovering UAV. *Journal of Sound and Vibration,* Volume 489.
Jeon, J. Y. & Jo, H. I., 2019. Effects of audio-visual interactions on soundscape and landscape perception and their influence on satisfaction with the urban environment. *Building and Environment,* Volume 169.
Jillings, N., De Man, B., Moffat, D. & Reiss, J. D., 2015. *Web Audio Evaluation Tool: A Browser-Based Listening Test Environment.* s.l., s.n.
Kak, S. C., 1973. Hilbert transformation for discrete data. *International Journal of Electronics,* 34(2), pp. 177-183.
Krishnamurthy, S., Christian, A. & Rizzi, S., 2018. Psychoacoustic test to determine sound quality metric indicators of rotorcraft noise annoyance. *Inter-Noise,* 258(7), pp. 317-328.
Kryter, K. D., 1960. The Meaning and Measurement of Perceived Noise Level. *Noise Control,* Issue 6, pp. 12-27.
May, D. N. & Watson, E. E., 1980. *Correction Procedures For Aircraft Noise Data Volume IV: Tone Perception,* Springfield: The National Technical Information Service.
Mestre, V. et al., 2017. *Assessing Community Annoyance of Helicopter Noise.* Washington, DC, s.n.
More, S., 2011. *Aircraft noise metrics and characteristics,* s.l.: Purdue University.
More, S. & Davies, P., 2010. Human response to the tonalness of aircraft noise. *Noise Control Engineering Journal,* 58(4), pp. 420-440.
Nicholls, R. & Torija, A. J., 2021. *An investigation into human response to unmanned aerial vehicle noise.* Washington, DC, Institute of Noise Control Engineering.
Nicholls, R., Torija, A. J., Green, N. & Romero, C. R., 2022. *Assessment of drone noise impact on urban soundscapes.* Glasgow, Institute of Noise Control Engineering.

Siemens Digital Industries Software, 2021. *Sharpness in Simcenter Testlab.* [Online]
Available at: https://community.sw.siemens.com/s/article/Sharpness-in-Simcenter-Testlab
[Accessed 26 July 2021].
Soeta, Y. & Kagawa, H., 2020. Three Dimensional Psychological Evalutation of Aircraft Noise and Prediciton by Physical Parameters. *Building and Environment* .
Sottek, R., 1993. *Modelle zur Signalverarbeitung im menschlichen Gehör (Models for signal processing in the human ear).* Aachen: RWTH Aachen University.
Sottek, R., Vranken, P. & Busch, G., 1995. *Ein Modell zrur Berechnung der Impulshaltigkeit (A model for calculating impulsiveness).* Saarbrucken, Germany, s.n.
Stalnov, O., Faran, M., Koral, Y. & Furst, M., 2022. Auditory detection probability of propeller noise in hover flight in presence of ambient soundscape. *The Journal of the Acoustical Society of America* , 151(6), pp. 3719-3728.
Terhardt, E., Stoll, G. & Steewann, M., 1982. Algorithm for Extraction of Pitch and Pitch Salience from Complex Tonal Sounds. *Journal of the Acoustical Society of America* , Volume 71, pp. 679-688.
Torija, A. J., Chaitanya, P. & Li, Z., 2021. Psychoacoustic analysis of contra-rotating propeller noise for unmanned aerial. *The Journal of the Acoustical Society of America,* Issue 149, pp. 835-846.
Torija, A. J. & Clark, C., 2021. A Psychoacoustic Approach to Building Knowledge about Human Response to Noise of Unmanned Aerial Vehicles. *International Journal of Environmental Research and Public Health,* Volume 682, p. 18(2).
Torija, A. J. & Flindell, I. H., 2015. The subjective effect of low frequency content in road traffic noise. *The Journal of the Acoustical Society of America,* 1(137), pp. 189-198.
Torija, A. J. & Li, Z., 2020. *Metrics for Assessing the Perception of Droine Noise.* Lyon, e-Forum Acusticum.
Torija, A. J., Li, Z. & Chaitanya, P., 2022. Psychoacoustic modelling of rotor noise. *The Journal of the Acoustical Society of America,* 151(3), pp. 1804-1815.
Torija, A. J., Li, Z. & Self, R., 2020. Effects of a hovering unmanned aerial vehicle on urban soundscapes perception. *Transportation Research Part D,* Volume 78.
Torija, A. J. & Nicholls, R., 2022. Investigation of Metrics for Assessing Human Response to Drone Noise. *International Journal of Enfironmenta;l Research and Public Health,* 19(6).
Torija, A. J. et al., 2019. On the assessment of subjective response to tonal content of contemporary aircraft noise. *Applied Acoustics,* Volume 146, pp. 190-203.
Torija, A. J., Self, R. H. & Lawrence, L. T., 2019. *Psychoacoustic Characterisation of a Small Fixed-pitch Quadcopter.* Madrid, s.n.
von Bismarck, G., 1974. Sharpness as an Attribute of the Timbre of Steady State Sounds. *Acustica,* Volume 30, pp. 159-172.
Watkins, S. et al., 2019. Ten Questions Concerning the use of Drones in Urban Environments. *Building and Environment* , Volume 167.
Wickelmaier, F., 2003. *An Introduction to MDS.* Aalborg: Sound Quality Research Unit, Aalborg University.
Yoo, W., Yu, E. & Jung, J., 2018. Drone deliver: Factors affecting the public's attitude and intention to adopt. *Telematics and Informatics,* Volume 35, pp. 1687-1700.
Zhang, Z. & Shrestha, M., 2003. *Sound Quality User-defined Cursor Reading Control - Tonality Metric,* s.l.: s.n.
Zwicker, E. & Fastl, H., 2007. Critical Bands and Excitation. In: *Psychoacoustics: Facts and Models.* Berlin: Springer-Verlag, pp. 149-173.
Zwicker, E. & Fastl, H., 2007. Loudness. In: *Psychoacoustics: Facts and Models.* Berlin: Springer-Verlag, pp. 203-238.
Zwicker, E. & Fastl, H., 2007. *Psychoacoustics: Facts and Models.* Third ed. Berlin: Springer-Verlag.
Zwicker, E. & Fastl, H., 2007. Sharpness and Sensory Pleasantness. In: *Psychoasoustics: Facts and Models.* Berlin: Springer-Verlag, pp. 239-246.