2023-01-1107 Published 08 May 2023

A Psychoacoustic Test for Urban Air Mobility Vehicle Sound Quality

Matthew Boucher NASA Langley Research Center

Menachem Rafaelof National Institute of Aerospace

Durand Begault NASA Ames Research Center

Andrew Christian, Siddhartha Krishnamurthy, and Stephen Rizzi NASA Langley Research Center

Citation: Boucher, M., Rafaelof, M., Begault, D., Christian, A. et al., "A Psychoacoustic Test for Urban Air Mobility Vehicle Sound Quality," SAE Technical Paper 2023-01-1107, 2023, doi:10.4271/2023-01-1107.

Received: 28 Feb 2023 Revised: 08 Mar 2023 Accepted: 08 Mar 2023

Abstract

This paper describes a psychoacoustic test in the Exterior Effects Room (EER) at the NASA Langley Research Center. The test investigated the degree to which sound quality metrics (sharpness, tonality, etc.) are predictive of annoyance to notional sounds of Urban Air Mobility (UAM) vehicles (e.g., air taxis). A suite of 136 unique (4.6 second duration) UAM rotor noise stimuli was generated. These stimuli were based on aeroacoustic predictions of a NASA reference UAM quadrotor aircraft under two flight conditions. The synthesizer changed rotor noise parameters such as the blade passage frequency, the relative level of broadband self-noise, and the relative level of tonal motor noise. With loudness constant, the synthesis parameters impacted sound quality in a way that created a spread of predictors both in synthesizer parameters and in sound quality metrics. Forty subjects listened to the suite of UAM noise stimuli in the EER and judged each sound individually on a standard scale of annoyance. Additionally, a subset of the UAM noise stimuli were compared to a reference sound that varied in loudness. From these responses, the relative effect of changes in loudness or changes in other sound quality metrics on annoyance was evaluated. This paper covers background and motivation for the test, details of how the sound stimuli were generated, and details of the test design and execution. Test results investigate how sound quality may affect perceived annoyance to UAM vehicle noise, indicating the importance of sharpness, tonality, impulsiveness, and roughness on annoyance to UAM noise.

Introduction

The Urban Air Mobility (UAM) concept involves a fleet of small aircraft, holding approximately 4-6 passengers, providing direct service between points within city centers (e.g., vertiports), or from a city center to an outlying transportation hub such as an airport [1]. These missions necessarily mean that UAM vehicles will be in close proximity to the public. Notionally, these vehicles will be novel rotorcraft configurations employing electric or hybrid-electric powertrains. Due to the wide vehicle design space offered by such new technologies and the unknown constraints that will be imposed by the integration of these vehicles into existing cities, there is a great deal of uncertainty about the noise that will result from such operations [2]. Will the noise be incessant and overwhelming, or will it not be heard at all over the existing noise of a city? Will the vehicles sound like any existing rotorcraft vehicles, or something completely new? What would an ideal sound or sound level be for this class of vehicles? What may people find annoying about this type of noise source? The goal of this study is to begin to answer this last question.

This paper describes the psychoacoustic Test of UAM Sound Quality (TUSQ) that was executed in 2022 in the Exterior Effects Room (EER) at the NASA Langley Research Center. This test leveraged a large amount of knowledge about the aeroacoustics of conventional rotorcraft (helicopters and UAVs), as well as psychoacoustic knowledge from past testing that has indicated the importance of sound quality (SQ). Specifically, perceptual aspects of noise such as tonality and roughness are important indicators in the perception of rotorcraft noise.

In terms of aeroacoustics, the mechanisms for noise production from UAM vehicles are largely the same as traditional rotorcraft. This includes periodic loading and thickness noise generated by the spinning rotor, various broadband self-noise mechanisms, and the possibility of electric motor noise [2]. Where UAM concepts differ is in the wide range of vehicle designs that are currently being pursued. These novel configurations may have rotors of widely varying size, tip speeds, and geometry, and may produce a similarly wide variation of noise signatures [3]. For instance, consider the balance between the periodic and broadband noise sources on a UAM vehicle. The balance between these two sources will be a function of rotor design and operation. For conventional helicopter designs, the periodic components tend to be the dominant source (though this is not to say that broadband sources are completely unimportant for helicopters [4]). A question for UAM is whether it may be beneficial for a UAM vehicle to have a broadband-dominated noise signature. Will residents of a city being served by this vehicle find that kind of noise to be less bothersome than the noise of a helicopter executing the same mission?

One way to answer this is to think of a noise as not just being described by a single number that quantifies its perceived magnitude – is something loud or not – but as having many qualitative dimensions that can vary. Thus, the broadband component of a helicopter is *sharper*, but less *impulsive* than the tonal component. These are examples of the sound qualities (SQs) over which various designs may range. Quantitative methods to estimate such SQs from recordings and predictions of noise have been maturing over the past few decades. The SQ metrics may be used as inputs to models that attempt to explain the annoyance or aversion people feel towards qualitatively different kinds of noise. Thus, while *loudness* may still be the most important SQ (as any annoying sound can always be "turned down"), these other SQs may be useful for vehicle design if low perceived annoyance at a human observer's location is a goal.

There are two main ways that researchers have related annoyance to SQ and vehicle design. The first approach, and one that has gained recent popularity, involves the creation of so called "Psychoacoustic Annoyance" (PA) models [5]. Here, researchers take annoyance ratings from psychoacoustic tests of sounds that cover a large range of SQ (and are nominally constant). They then formulate nonlinear combinations of the SQ parameters that correlate well with the test results. This is done in such a way that *absolute* predictions of annoyance in pressure-like units are generated – an attractive outcome as noise metrics used for certification and regulation are also necessarily on absolute scales. These PA models may then be used to guide the design of future noise-making machines. Over time this approach has been extended beyond the original slate of sounds to specific noise sources (e.g., fixed-wing aircraft) and beyond the original slate of SQ parameters to include aspects such as tonality [6, 7]. The most recent publication of this type comes from Torija et al. and is focused on the noise of small UAV rotors [8].

An alternative is to take a vehicle-centric approach. Here, a model of a particular vehicle is modified in line with realistic changes to the vehicle design. Psychoacoustic testing is then undertaken to determine the optimal design within this achievable space [9]. This approach aligns more with what is seen in industrial applications of designing for sound quality, where a machine is nominally defined already, and simply needs to be tuned to give the most pleasant user experience [10].

This study charts a course between these two concepts. On one hand, the information generated from the test may be useful for building models for predicting annoyance across a large swath of UAM vehicles. On the other hand, the slate of sounds that comprise the test come from a single model of a NASA UAM concept vehicle. The sound of this vehicle is generated by auralizing different flight conditions (i.e., the sound that the vehicle radiates is dependent on whether it is flying straight, ascending, turning, etc.). These different conditions are then subjected to a significant amount of post processing which allows the sound of the vehicle to change from that resembling the sound of a helicopter, to that of a propeller-driven plane, and beyond to more broadband-dominated sources reminiscent of jet aircraft. Ultimately, this dataset may be useful for extending existing PA-like models or developing new models based on other approaches (e.g., random forests [11]).

This study also builds upon the rotorcraft sound quality metric (RoQM) series of psychoacoustic tests previously undertaken at NASA Langley. These experiments investigated the role that SQ plays in the perception of conventional helicopter designs. The RoQM-I test, in particular, shares psychoacoustic methodology and sound design strategies with this work [12, 13]. That test determined that the SQs of sharpness, tonality, and fluctuation strength are all important in the perception of helicopter noise. The RoQM-II (A and B) tests looked at helicopter noise from the point of view of entire flyover events [14, 15]. They demonstrated that significant *qualitative* differences may exist between noise signatures of different helicopters, even within models of the same make and capability. These differences were also found for the *same* helicopter performing different maneuvers (turning, descending, etc.). The test demonstrated that the sound exposure level noise metric stood to be improved by incorporation of qualitative information.

In this work, certain technical challenges are addressed, including: (1) in the absence of a database of recordings, how can UAM noise stimuli be generated for use in a psychoacoustic test? (2) how can these stimuli be designed to span a significant range of sound quality?, and (3) how can a psychoacoustic test be executed to test sound quality factors that may contribute to annoyance?

After solutions to these challenges are described, research questions are addressed, which are concerned with how sound quality of UAM vehicle noise affects perceived annoyance, assuming that loudness is the dominant factor. Specific research questions are:

1. What is the relative difference between effects of loudness alone and effects due to a combination of other sound quality characteristics?
2. In terms of sound quality other than loudness, what are some of the important factors influencing annoyance responses to UAM noise?

Generation of UAM Noise Stimuli

The psychoacoustic test for UAM vehicle sound quality produced sound stimuli based on aeroacoustic noise predictions and auralizations of a quadrotor NASA reference vehicle with three blades on each rotor. This vehicle was a six passenger, 1200 lb. payload quadrotor turboelectric design [16]. The auralized noise at the observer location was then modified to span a range of sound quality, such as sharpness, tonality, impulsiveness, and fluctuation strength. The stimuli were presented to human test subjects who responded with annoyance ratings and comparisons. This section discusses different aspects of sound quality and addresses the first technical challenge described above: how UAM noise stimuli were generated in the absence of a database of UAM vehicle recordings.

Measures of Sound Quality

This section describes the measures of sound quality that were studied in this test – where they come from and what they represent in the sample sounds. It also discusses the algorithms used to calculate the metrics, as some SQs have multiple standards and others have none.

Loudness is "the sensation that corresponds most closely to the perception of sound intensity of the stimulus" [5]. Measured in phon or sone, equal loudness contours based on pure tones show a strong frequency dependence similar to the hearing threshold in quiet (loudness of 0 sone) and are also level dependent. In this work, two similar implementations of the Zwicker loudness method for time varying sounds in free field were used. One used the DIN 45631/A1 standard [17] implemented in ArtemiS Suite 13.6 [18], and the other used ISO 532-1 [19] implemented in the NASA Auralization Framework (NAF) [20] Psychoacoustic Library (NAF-PAL). For reference, a pure tone at 1kHz with an SPL of 40dB is defined as 1 sone, and the loudness level doubles for each 10 dB increase in SPL.

Sharpness is a measure of spectral balance; it increases if acoustic energy is concentrated at higher frequencies. This work used the standard DIN 45692 (free field) [21], which was developed to compare sounds of similar loudness.

The hearing model of Sottek [22, 23], also implemented in ArtemiS Suite, was used to calculate the sound quality metrics of tonality, roughness, impulsiveness, and fluctuation strength which are described below. The hearing model mimics the signal processing that occurs in the human auditory system. Sound passes through filters that model the outer and middle ear, and overlapping bandpass filters represent the frequency selectivity of the inner ear. Subsequent steps consider the threshold of hearing and the limits of the auditory system to track time varying noise within a critical band.

A noise that is perceived to be tonal is one that has a concentration of acoustic energy in a narrow frequency band and that has low acoustic energy in adjacent bands. The extreme example of such a noise is a simple sinusoid. Tonality is measured in tonality units (TU). The hearing model for calculating tonality was designed to consider all aspects of tonality perception, not only those resulting from pure tones [24].

Fluctuation strength (measured in vacil) and roughness (measured in asper) are calculated in similar ways but measure different perceptual characteristics. Fluctuation strength is the result of slow temporal variations of the noise, typically below 20Hz, while roughness measures faster temporal variations, up to 300Hz, where the human auditory system is not capable of tracking real-time fluctuations. Maximum perceived responses occur at 4Hz and 70Hz for fluctuation strength and roughness, respectively.

Also calculated using the hearing model, impulsiveness is the perceptual effect resulting from sudden changes in sound intensity. It is measured in impulsiveness units (IU).

Perception of sound quality may differ for complex UAM sounds used in this experiment compared to simple sound stimuli used to derive some of the above SQ metrics. As an example, loudness is 3.5 times higher for uniform exciting (i.e., broadband) noise than it is for a pure tone at 1kHz if both are played at 60dB [5]. The noise stimuli in this test have both tonal and broadband components, so the perceived difference in loudness is most likely not as extreme. Nevertheless, the sound quality metrics listed here are used as potential predictors of annoyance; other calculation methods for these metrics are out of scope of this work.

Sound Stimuli

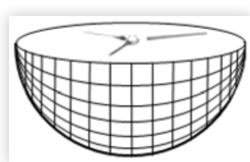
The Test for UAM Sound Quality consisted of 137 sound stimuli that varied in loudness and other sound quality metrics described above. The stimuli generation consisted of the following steps:

1. Aeroacoustic noise predictions of individual rotors of the quadrotor NASA reference vehicle
2. Auralizations of the quadrotor vehicle noise under level cruise and 5-degree descent flight conditions
3. Post-processing of auralizations to generate stimuli with varying sound quality

Aeroacoustic Noise Predictions The noise predictions started by calculating blade loading, motion, inflow velocity and effective angle of attack using the Comprehensive Analytical Rotorcraft Model of the Rotorcraft Aerodynamics and Dynamics (CAMRAD II) program [25]. Blade geometry, blade passage frequency and flight condition were among the inputs given to CAMRAD II. The flight conditions were for level cruise and 5-degree descent, and the blade passage frequency was 20Hz. The trailing edge thickness of the blades was 1.8mm.

The outputs of CAMRAD II (blade loading, motion, inflow velocity and effective angle of attack) were used as inputs to the Aircraft NOise Prediction Program 2 (ANOPP2) [26] to generate acoustic predictions on a hemisphere grid of points surrounding each rotor of the NASA quadrotor. Figure 1

FIGURE 1 Source noise hemisphere around rotor.



illustrates a source noise hemisphere surrounding a rotor. It is an exaggerated illustration because the hemisphere radius will normally be at least several times the rotor radius. Each grid point on the hemisphere contains an acoustic prediction.

ANOPP2 calculated a separate source noise hemisphere surrounding each vehicle rotor for loading and thickness (i.e., periodic) noise predictions using the Formulation 1A Internal Function Module (AF1AIFM). This is an implementation of Farassat's Formulation 1A [27, 28]. Each loading and thickness noise hemisphere grid point contained a blade passage pressure time history.

A separate noise hemisphere was calculated for self-noise (i.e., broadband) predictions using the ANOPP2 Self Noise Internal Functional Module (ASNIFM) for each rotor. This noise is predicted in 1/3-octave bands and is an implementation of the Brooks-Pope-Marcolini self-noise model in a rotating frame [29]. Each self-noise hemisphere grid point contained the 1/3-octave band sound pressure level time history over a single rotor revolution. A more in-depth description of the aeroacoustic modeling is given in [30].

Auralizations Source noise hemisphere acoustic predictions are not in forms that are directly audible. The loading and thickness noise blade passage signals are too short, the self-noise predictions lack phase information, neither prediction is at a time resolution required for audio playback, and they are only available at discrete points instead of being continuous over the hemisphere. A full auralization from source noise hemispheres requires separate synthesis of audible sound for each noise source (loading and thickness, and self-noise) followed by propagation to a simulated listener [31].

Although the noise predictions corresponded to level flight and 5-degree descent operating conditions, the auralizations were made by assuming the vehicle was stationary with respect to the observer. A stationary vehicle was used because the auralized sounds for this paper needed to be roughly constant in SQ. For a vehicle moving with respect to an observer, the SQ metric values may change considerably with rotor-to-observer emission angle.

The NASA Auralization Framework (NAF) [20] was used to synthesize each rotor noise source into audible sound. The advanced plugin for periodic additive sound synthesis [20] was used to auralize the loading and thickness noise. The Modulated Broadband Synthesis Plugin [32] was used to auralize the self-noise by modulating a stochastic signal. The observer was located flush with the ground, and the NASA quadrotor was 1000 ft above ground level with a constant emission angle of 60 degrees elevation and 0 degrees azimuth, relative to the observer. Each noise source was extended to 10 seconds in duration at a sampling rate of 48 kHz. Spherical spreading and atmospheric absorption were applied, but no ground reflection was included.

After auralizing both the loading and thickness noise as well as the self-noise, the results were simply added to get the total noise at the ground observer. Then, the loudness of the total noise was adjusted to 6 sones through an iterative process using the NAF-Psychoacoustics Analysis Library (NAFPAL) [33]. For a target loudness of N* and stimuli loudness of N, the total noise was multiplied by N*/N until the percent between N* and N was within 1%.

Post-processing After generating auralizations of a stationary quadrotor from acoustic predictions of two different flight conditions, a number of post-processing techniques were applied. This was done to create a range of SQ metric values for the psychoacoustic test. The post-processing applied to the two auralizations included:

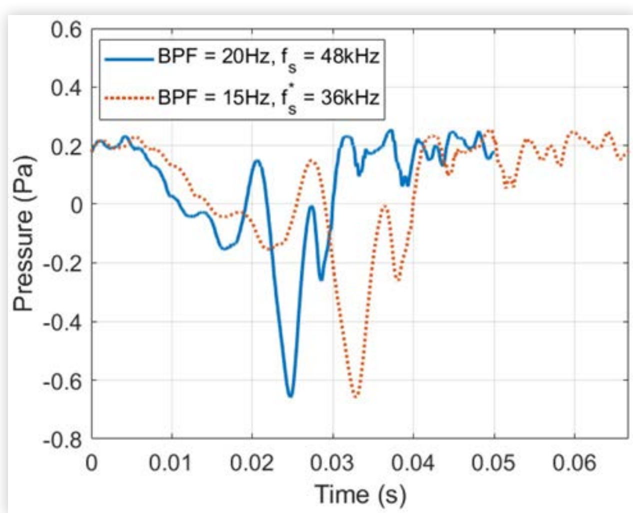
1. Resampling the loading and thickness noise (i.e., assuming a different time basis) to mimic a change in blade passage frequency and remodulating the self-noise to track the change in blade passage frequency
2. Applying a gain (positive or negative) to the broadband self-noise component of the source noise
3. Mimicking electric motor noise by adding a tone complex whose fundamental frequency was linked to the blade passage frequency
4. Applying a moving average to the loading and thickness noise component of the source noise
5. Applying amplitude modulation to the total noise

After the post-processing, the total noise was adjusted to a loudness of 6 sone.

Change of Blade Passage Frequency. The source noise definitions for the quadrotor reference vehicle had a blade passage frequency (BPF) of 20Hz. However, UAM vehicles are expected to operate over a range of BPFs, which will affect the time and frequency content of the noise and may be important to perception. Therefore, two methods were developed to modify the BPF of the auralizations. The BPF of the loading and thickness noise was changed by assuming a different time basis, and the broadband self-noise was remodulated at the new BPF.

The process for changing the BPF of the loading and thickness noise is depicted in Figure 2. First, the effective period of one revolution is changed by calculating the ratio of sampling frequencies. For example, to change to a new BPF,

FIGURE 2 Assuming a lower sampling frequency for the loading and thickness noise extends the period of one revolution and mimics a change in BPF.

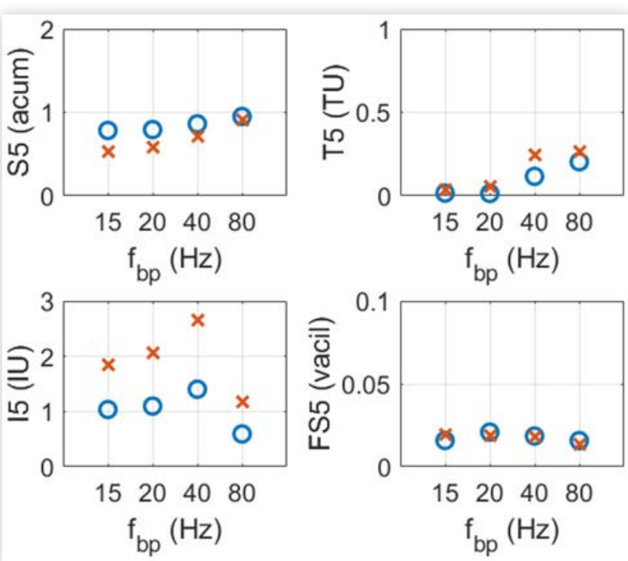


the following relationship holds: $f^*_{bp}/f_{bp} = f^*_s/f_s$, where f_{bp} is the BPF of the auralization, f_s is the sampling frequency of the auralization and '*' indicates the new value. Lowering the BPF will extend the waveform, as shown in Figure 2. All frequency content of the loading and thickness noise is lowered in this case. If the BPF increases, the loading and thickness noise frequency content is also raised. Since the other noise components and total noise must be at a sampling rate of 48kHz, the pressure time history is resampled to get the pressures back to the desired time basis.

The second part of changing the BPF of the auralization is to re-modulate the broadband self-noise. Each 1/3-octave band of the broadband self-noise auralization is characterized by a modulation rate equal to the BPF and a unique modulation depth. This modulation profile over one revolution is applied to regenerate self-noise, assuming the new duration of one revolution by the method outlined by Christian et al. [34]. This method shapes white noise (resulting in the same spectrum as the original self-noise auralization) and then applies a fluctuating time envelope, at the new BPF with the appropriate modulation depth from the original modulation profile, to create the new modulated broadband noise.

After combining the individual sources, the authors conducted an informal audition of various BPFs between 15 and 200Hz, considered to be the plausible range of operating conditions for UAM vehicles. It was judged that BPFs of 15, 20, 40 and 80 Hz sufficiently spanned the range of aural impressions perceived during the audition. Figure 3 shows the 5% exceedance values of sharpness (S5), tonality (T5), impulsiveness (I5), and fluctuation strength (F5) for four BPFs and two flight conditions. The exceedance level is the value of the metric that is exceeded x% of the time within a sound segment. The blue circles represent the level cruise condition, and the red crosses represent the five-degree descent condition. Sharpness and tonality slightly increase for higher BPFs. Impulsiveness also increases up to 40Hz but then decreases substantially for 80Hz. Fluctuation strength is mostly constant and low for these BPF values.

FIGURE 3 Effect of a change in blade passage frequency on the 5% exceedance values of various SQ metrics. Points at f_{bp}=20Hz are the original auralizations. Circle is level cruise and cross is 5-degree descent.



Adjust Spectral Weighting to Change Sharpness. To change the sharpness SQ metric, a spectral weighting parameter, g_{hf}, modified the spectral balance of an auralization to lower or higher frequencies. This parameter was applied to the self-noise component of the auralization only. It controls the spectral balance in two ways: (1) applying a coloring filter to adjust the weighting of acoustic energy at lower or higher frequencies and (2) applying a frequency independent gain factor given by $10^{g_{hf}/20}$. Figure 4 shows the first effect of the spectral weighting parameter on the self-noise. For g_{hf}>0 dB, the spectral centroid is higher. For g_{hf}<0 dB, the spectral centroid is lower, as is the overall level.

After the spectral centroid is modified, the frequency independent gain factor is applied, and the resulting modified broadband noise is added to the loading and thickness noise. Then, the total noise is adjusted to a loudness of 6 sones, as described previously.

The effect of the spectral weighting parameter on spectrum of the level cruise auralization is shown in Figure 5. The g_{hf} = 0 dB trace is for the original auralization. For g_{hf} < 0 dB, the SPL is reduced for frequencies above approximately 900Hz and increased for lower frequencies, which reduces the sharpness. The reverse is true for g_{hf} > 0 dB; higher frequencies are much more prominent, which increases the sharpness.

Add Tonal Component to Increase Tonality. The use of electric propulsion in UAM vehicles may lead to the presence of highly annoying pure tones produced by electric motors. To assess the annoyance of such tones, tone complexes were constructed mimicking the commutation of electrical current in Brushless DC motors which may be related to motor spin speed (RPM) and BPF. Additive synthesis is used here to construct an electric motor tone complex consisting of a

FIGURE 4 Use of the spectral weighting factor, g_{hf}, to apply a coloring filter to the self-noise (f_{bp}=20Hz, level cruise flight condition) and to raise or lower the spectral centroid.

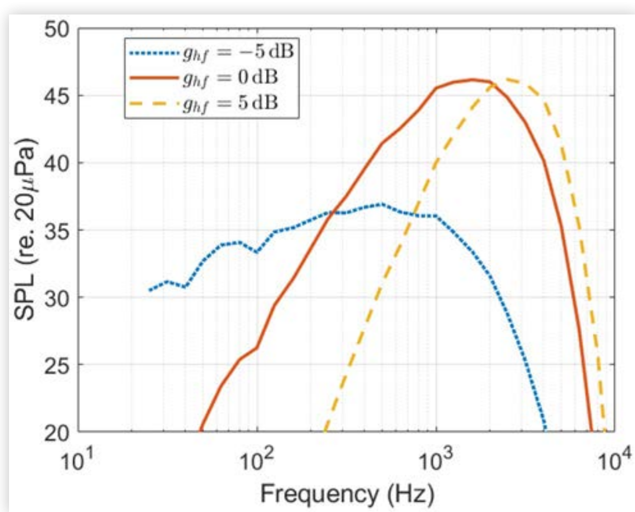
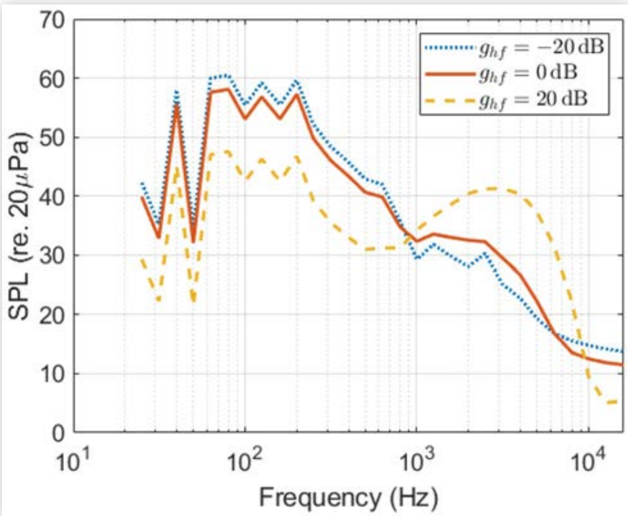


FIGURE 5 Effect of the spectral weighting parameter, g_{hf}. The 1/3-octave band spectra shown are the sum of the loading and thickness noise and the self-noise for 20Hz blade passage frequency for the level flight condition.



harmonic series of sine waves. The inclusion of the tone complex affects the tonality SQ. The time series of the tone complex is given by

$$y_{tone}(t) = A\sum_{n=1}^{18} a_n \sin(2\pi f_{com} nt + \phi_n) \quad (1)$$

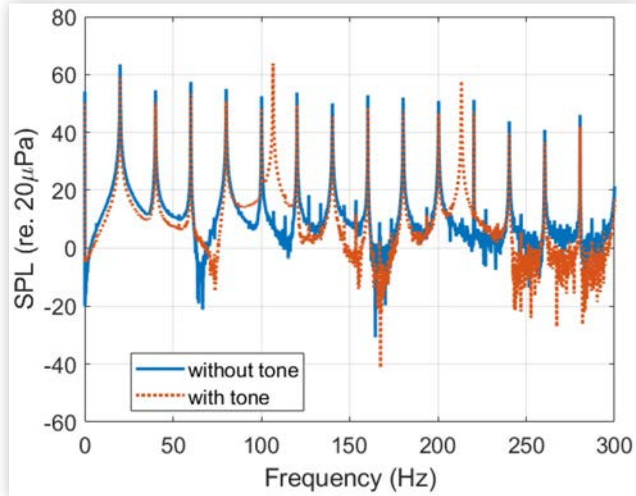
where n is the harmonic number, a_n is the amplitude of each harmonic, given by 0.5^{n-1}, f_{com} is the commutation frequency, t is time, and ϕ_n is a random phase for each harmonic. The tonal amplitude factor is A, which is used to modify the contribution of the tone to the total noise. The RPM is given by $\omega = f_{bp} * 60/n_b$, where n_b is the number of blades and f_{bp} is the blade passage frequency. The commutation frequency of the electric motor is $f_{com} = \omega N_{pp}/60$, and N_{pp} is the number of motor pole pairs.

The result of adding a tone complex to the auralization of the level cruise flight condition ($f_{bp} = 20Hz$) is shown in Figure 6. The curve 'without tone' is the unchanged auralization. For the low frequencies shown, the auralization is dominated by a harmonic series based on the blade passage frequency. However, some effect of the broadband self-noise is visible in between the peaks at lower SPL.

When the tone complex is added, a second harmonic series is introduced. With N_{pp}=16, the first and second harmonics occur at 107 and 213Hz, which are clearly visible in the spectra and have comparable amplitudes to the loading and thickness noise. The peaks at harmonics of the BPF are lowered in the spectrum with the added tone, because the two spectra are of equal loudness. Changing the amplitude factor, A, affects the tonality of a sound.

Apply Moving Average to Reduce Impulsiveness. The impulsive character of helicopter noise is associated with periodic "chopping" sounds corresponding to each rotor rotation. The largest contribution to this character comes from

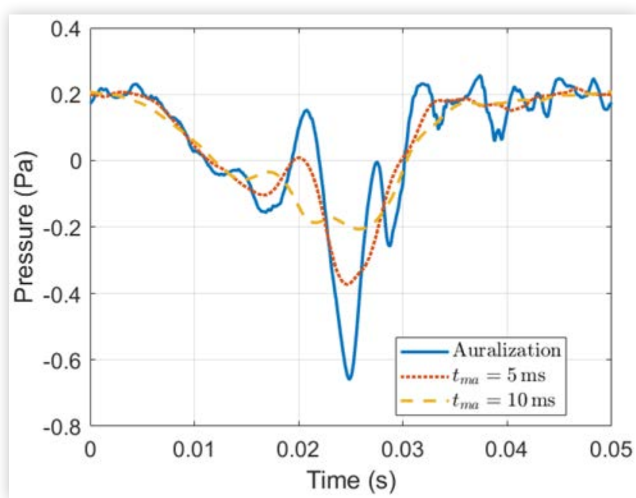
FIGURE 6 Effect of adding a motor tone complex to the auralized sound. With N_{pp} = 16, the first and second harmonics are at 107 and 213Hz. The case without tone is the level cruise at 20Hz blade passage frequency (loading and thickness noise and self-noise included).



the loading and thickness noise. By time averaging the waveform, the impulsive character can be reduced. Figure 7 demonstrates this process on the loading and thickness noise shown previously in Figure 2. For the unchanged auralization, t_{ma} is effectively $1/f_s$. For higher values of the moving average parameter, t_{ma}, sudden changes in amplitude are diminished, and the impulsiveness metric is reduced.

Amplitude Modulation to Increase Fluctuation Strength. Modulations in rotorcraft noise have been found to be an important characteristic of the audible rotor noise. Here, amplitude modulation is applied to the total noise to affect the fluctuation strength SQ.

FIGURE 7 Acoustic pressure time history for the level cruise condition at 20Hz blade passage frequency after applying a moving average to the loading and thickness noise.



The amplitude modulator is given by

$$a(t) = b(t)\left(1 + m(t)\right)$$
$$m(t) = M \sin\left(2\pi f_m t\right) \quad (2)$$

In Eq. (2), M is the modulation index, f_m is the modulation frequency, m(t) is the modulating signal, a(t) is the amplitude modulated signal, and b(t) is the signal to be modulated. This modulation approach is equivalent to the deterministic amplitude modulation method used in previous work on helicopter sound quality [12].

The effect of amplitude modulation is applied to the level cruise flight condition auralization (f_{bp}=20Hz) in Figure 8. Five blade passages are shown, which corresponds to one modulation period. With no modulation (M=0), which is the original auralization, the peak amplitudes for the blade passages remain roughly constant. For M=0.2, higher and lower peak amplitudes are evident throughout the period.

Structure of the Algorithm. The steps of the UAM noise stimuli generator are as follows:

1. Start with separate loading and thickness noise and self-noise for the quadrotor vehicle for a given flight condition (level cruise or 5-degree descent) with a BPF of 20Hz
2. If desired BPF is not 20Hz
   a. Assume a different time basis for the loading and thickness noise
   b. Remodulate the self-noise
3. If g_{hf} is not 0 dB (to adjust the sharpness SQ)
   a. Apply spectral coloration to the self-noise
   b. Apply gain to self-noise
4. If moving average is not equal to $1/f_s$, apply moving average to the loading and thickness noise to lower the impulsiveness SQ

FIGURE 8 Effect of applying amplitude modulation to the auralization. With a modulation frequency of 4Hz, the time window shown is for one period.

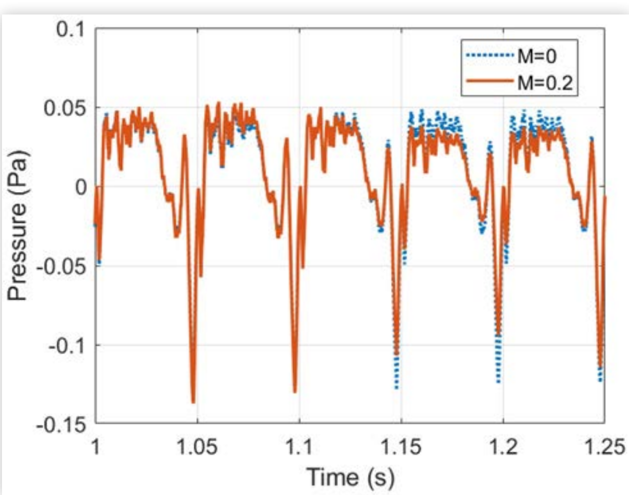
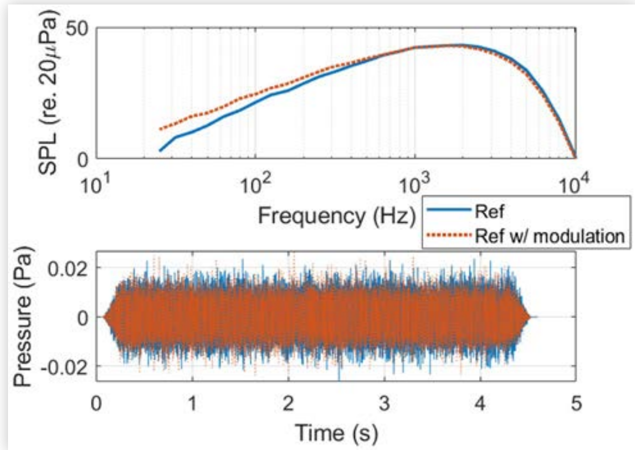


FIGURE 9 Spectrum and waveform of the reference (Ref) sound used in the psychoacoustic experiment. A 20Hz modulation was removed and the reference was set to 6 sones, resulting in slightly lower spectral level at low frequencies.



5. If tone amplitude, A, is not equal to 0, add tone complex to increase tonality
6. To adjust the fluctuation strength SQ, apply amplitude modulation in Eq. (2)
7. Add the different components
   a. Loading and thickness noise
   b. Self-noise
   c. Tone complex
8. Equalize to desired loudness level (6 sones for this test)

Reference Sound to Test Loudness Effects To test annoyance responses based on loudness, a reference sound was generated. From a pilot test involving NASA colleagues, the 10 least annoying sounds were considered as candidates to be a reference sound. The intent was to have a sound that did not have any dominant sound quality characteristics, so that differences in annoyance responses would be mostly due to changes in loudness. The level cruise auralization with a BPF of 20Hz was one of the sounds found to be least annoying. To make this sound more bland and potentially less annoying, temporal variations were removed. This was done by not including the loading and thickness noise and removing the modulation of the broadband self-noise. Figure 9 shows the reference sound (blue trace) has roughly the same spectrum as the self-noise of the level cruise auralization (red trace). Small differences in SPL resulted from setting the reference sound to 6 sones after removing the 20Hz BPF harmonics.

Test Design and Execution

The previous section described how UAM noise stimuli were generated from predictions, auralizations and post-processing techniques. This section addresses how the psychoacoustic test was designed and executed using the above noise generation tools. First, a full factorial design was applied to 8 baseline sounds to span a significant range of sound quality. Then, details of how the psychoacoustic test was executed are given.

Design of Noise Stimuli Based on Sound Quality

Test stimuli were generated starting from source noise hemispheres of individual NASA quadrotor reference vehicle rotors from level cruise and 5-degree descent flight conditions. Rotor loading and thickness noise and self-noise auralizations for a 20 Hz BPF were generated assuming a stationary vehicle with respect to the ground observer. Since the auralized sounds from all four rotors were roughly the same SQ, the auralization of only one rotor for each flight condition was selected to be modified into test stimuli. A single rotor sound will be added three more times to itself to approximate the full quadrotor sound. This approximation was considered acceptable for far-field sounds.

Both flight conditions were modified to produce stimuli at BPFs of 15, 40 and 80 Hz. Including the original auralizations at 20Hz BPF, this created 8 baseline stimuli (two flight conditions at four BPFs each). From each baseline, a full factorial, two-level design was created based on four factors (sharpness, tonality, impulsiveness, and fluctuation strength). An example of a 3-factor, two-level design is shown in Figure 10, which is a cube with sharpness, tonality, and impulsiveness axes. Each of the eight corners of the cube is a different combination of low or high values of the three factors considered. What values are "low" or "high" will be explained shortly. The corners of the cube in Figure 10 may represent all the stimuli at low fluctuation strength (as shown in Table 3), and another cube would show the combinations at high fluctuation strength.

The full factorial design generated 16 categories of UAM noise stimuli, based on different combinations of low/high values of sharpness, tonality, impulsiveness, and fluctuation strength. These categories are detailed in Table 1. These noise qualities were generated using specific values of the post-processing parameters described earlier.

FIGURE 10 3-factor, 2-level (2^3) full factorial design with sharpness, tonality, and impulsiveness as factors.

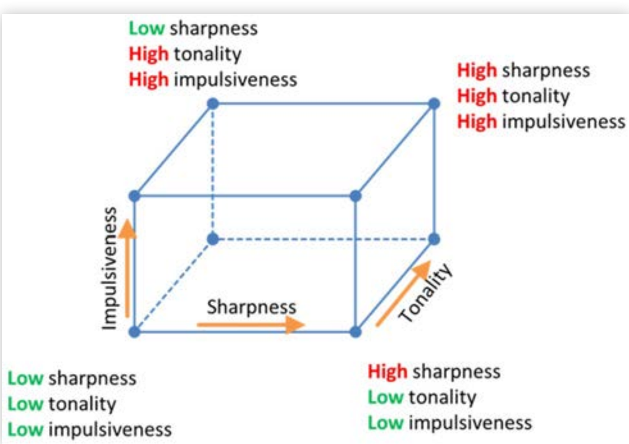


TABLE 1 Categories of UAM noise stimuli generated from a 4-factor, 2-level full factorial design. Categories 9-16 are identical to categories 1-8, except that they have high fluctuation strength.

Category	Sharpness	Tonality	Impulsiveness	Fluctuation Strength
1	Low	Low	Low	Low
2	High	Low	Low	Low
3	Low	High	Low	Low
4	High	High	Low	Low
5	Low	Low	High	Low
6	High	Low	High	Low
7	Low	High	High	Low
8	High	High	High	Low

For the level cruise flight condition with BPF=20Hz, the synthesis parameters used to achieve the desired qualitative aspects are shown in Table 2. The spectral weighting parameter was -10dB for low sharpness and +10dB for high sharpness. The moving average time was 0.010s for low impulsiveness and 1/fs for high impulsiveness. The low fluctuation strength was achieved with a modulation index of 0.075; high fluctuation strength had a modulation index of 0.15. For low tonality, the tonal amplitude factor was set to 1. For high tonality, the tonal amplitude was varied until the tone complex was clearly audible above the loading and thickness noise and the self-noise, resulting in values that ranged between 10 and 24 for this case.

Figures 11 and 12 show how the synthesis parameters affected sharpness, tonality, impulsiveness, and fluctuation strength using the 20 Hz level cruise baseline. Alternating low and high values for each SQ metric can be seen for the 16 categories of the full factorial design. In Figure 11 (top), all eight 'high' sharpness stimuli have higher sharpness than all 'low' sharpness stimuli. However, this is not the case for impulsiveness in Figure 12 (top). For example, the impulsiveness of category 1 is higher than the impulsiveness of category 8. For this case, it is more meaningful to compare categories 4 and 8, where both stimuli have high sharpness and tonality and low fluctuation strength, the only difference being the

TABLE 2 Synthesis parameters for level cruise flight condition at BPF=20Hz. Categories correspond to combinations of low/high sound quality values shown in Table 1. Categories 9-16 (M=0.15) had similar parameters.

Category	Spectral Weighting Parameter, g_{hf} (dB)	Tonal amplitude factor, A	Moving average, t_{ma} (s)	Modulation index, M
1	-10	1	0.010	0.075
2	10	1	0.010	0.075
3	-10	10	0.010	0.075
4	10	10	0.010	0.075
5	-10	1	$1/f_s$	0.075
6	10	1	$1/f_s$	0.075
7	-10	24	$1/f_s$	0.075
8	10	15	$1/f_s$	0.075

FIGURE 11 Sharpness and tonality of UAM noise stimuli generated from the level cruise, 20Hz auralization. Blue are low values; red are high values. See Table 3 for definition of categories. Dotted lines are SQ values of baseline.

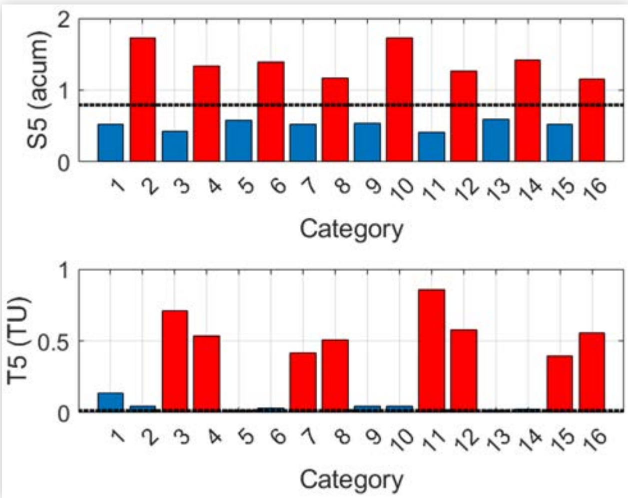
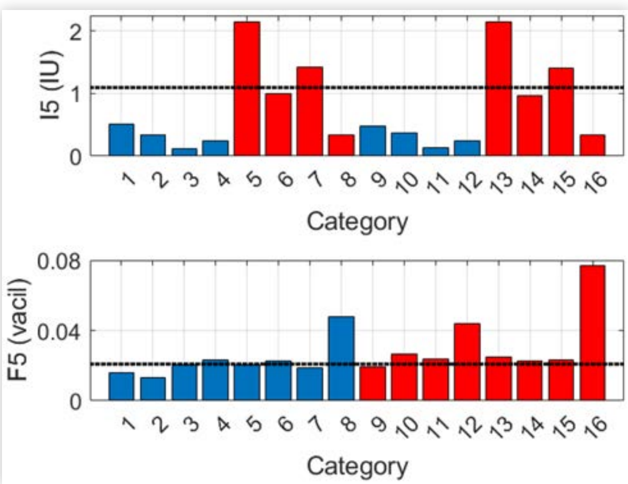


FIGURE 12 Impulsiveness and fluctuation strength of UAM noise stimuli generated from the level cruise, 20Hz auralization. Blue are low values; red are high values. See Table 1 for definition of categories. Dotted lines are SQ values of baseline.



level of impulsiveness. This is because the synthesis parameters, although targeted at one SQ, may affect more than one SQ at a time, which is a major challenge in designing a psychoacoustic test with UAM noise stimuli.

Despite this challenge, similar patterns of low/high sound quality metrics were found for the other seven baselines using the available synthesis parameters. Including the baselines, this resulted in 136 unique UAM noise stimuli at the same loudness level. Various statistics of the SQ values are shown in Table 3. It may be that future acoustic measurements show that UAM vehicle sound quality could extend outside the ranges shown in Table 3. Nevertheless, the ranges shown are assumed to span what is plausible for UAM vehicles.

TABLE 3 Minimum, mean, maximum, and standard deviation, σ, of the sound quality metrics for 136 UAM noise stimuli generated for the psychoacoustic test.

	N (sones)	S (acum)	T (TU)	R (asper)	I (IU)	F (vacil)
Min	5.8	0.38	0.01	0.01	0.02	0.01
Mean	6.1	0.9	0.3	0.3	0.8	0.02
Max	6.4	1.8	0.9	2.0	3.0	0.08
σ	0.07	0.4	0.3	0.4	0.7	0.01

TABLE 4 Mean sound quality metrics for the reference sound. The reference was presented at five different levels, spanning 20dB in 5dB increments, resulting in loudness values between 3.8 and 15.7 sones.

S (acum)	T (TU)	R (asper)	I (IU)	F (vacil)
1.6	0.04	0.03	0.32	0.01

As described earlier, a reference sound was also generated by removing the modulation in the broadband self-noise for the level cruise flight condition at a BPF of 20Hz. This reduced tonality, roughness, impulsiveness, and fluctuation strength. The SQ metrics for the reference sound are shown in Table 4.

Psychoacoustic Test

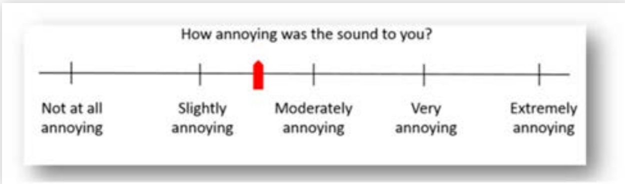
The psychoacoustic Test of UAM Sound Quality (TUSQ) was executed in 2022 in the Exterior Effects Room (EER) at NASA Langley Research Center. Forty test subjects were recruited from the local area and were tested in groups of four. TUSQ was executed in accordance with the NASA Institutional Review Board and respected the prevailing COVID health and safety guidance for the summer of 2022. The test was divided into two parts to measure subjective annoyance to UAM noise: (1) annoyance ratings on an 11-point scale to 136 unique UAM noise stimuli of equal loudness and (2) annoyance comparisons between a subset of UAM noise stimuli of equal loudness and a reference sound that varied in loudness.

Annoyance Ratings With 8 baselines and 16 stimuli generated for each baseline, 136 unique UAM noise stimuli were generated. With 4 replicates of the baselines, a total of 160 (8*12+8*4) stimuli were presented to the test subjects. Each stimulus had a loudness of 6 sones and a duration of 4.6 s. The stimuli were presented over 4 sessions of 40 questions each.

A familiarization session was played before the test began to give the test subjects an impression of the range of sounds that would be contained in the test. These sounds included both of the auralizations at 20Hz BPF, the other baselines of the level cruise with BPFs of 15, 40 and 80Hz, and also the two sounds that had the highest values of sharpness, tonality, impulsiveness and fluctuation strength (a total of 2+3+8=13 sounds).

A practice session of 20 sounds was played to give the test subjects the opportunity to become familiar with the computer tablets and how to record their annoyance response. After each stimulus, the question, “How annoying was the sound

FIGURE 13 Question to test subjects for sounds of equal loudness. The test subject replies by sliding the cursor on a computer tablet screen.



to you" appeared on computer tablets (see Figure 13). Each subject responded by moving a slider on an annoyance rating scale with responses ranging from "Not at all annoying," at a level of 2 to "Extremely annoying" at a level of 10. Therefore, a numerical rating of "1" corresponded with an annoyance rating below "Not at All Annoying," and a numerical rating of "11" corresponded with an annoyance rating above "Extremely Annoying." After the practice session, four regular test sessions followed resulting in 6400 annoyance ratings (160 questions times 40 subjects).

Annoyance Comparisons After 4 sessions of giving annoyance ratings to the sound stimuli with the same loudness, another familiarization session was presented to the subjects to introduce the reference sound. The reference sound was played 5 times over a range of 20 dB, with the sound increasing in level every presentation. This was done, not only to familiarize the subjects with the reference sound, but also to make it clear that this sound would be presented at significantly different levels – that level was an auditory cue to be used in their judgements of the pairs.

Another practice session gave the test subjects the opportunity to become familiar with the A-B comparison task and to practice recording their response on the computer tablets. This practice session consisted of 13 A-B comparisons.

A fifth test session (Session 5) consisted of A-B comparisons between the reference sound and 26 stimuli selected from the 136 unique UAM noise stimuli. The original auralizations (level cruise and 5-degree descent, both at a BPF of 20Hz) were two of the 26 sounds, and 24 other stimuli were selected at random. The reference sound was played at 5 different loudness levels, spanning 20dB, corresponding to 3.8, 5.6, 8.0, 11.3 and 15.7 sones.

The test was completed with the test subjects responding to the Weinstein noise sensitivity questionnaire [35].

Results

The main results from the annoyance responses collected from the Test for UAM Sound Quality are organized as follows:

1. Analysis of A-B comparisons between the reference sound (which varied in loudness) and other UAM noise stimuli (fixed loudness)
2. ANOVA analysis of changes in sharpness, tonality, impulsiveness, and fluctuation strength for noise stimuli at fixed loudness
3. Regression analysis when roughness is also included
4. Analysis using a psychoacoustic annoyance model

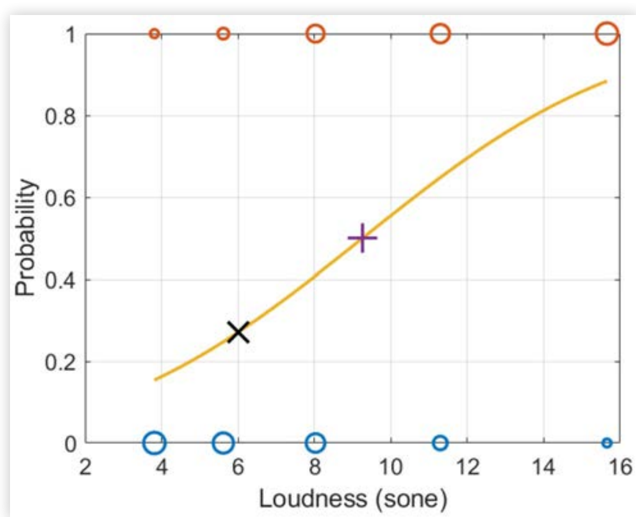
Effect of Loudness

Loudness is known to be the most important factor (relative to other sound quality metrics) in the annoyance response to noise. The annoyance responses from Session 5 compared UAM noise stimuli with a reference stimulus. The UAM noise stimuli were presented at 6 sones while the loudness of the reference stimulus was presented at loudness levels between approximately 4 and 15 sones.

The annoyance comparisons between the UAM noise stimuli and the reference stimulus are presented in Figure 14. The responses are given by the circles and plotted as a function of loudness of the reference stimulus. A 'zero' response indicates that the reference stimulus was less annoying than the UAM noise stimuli, and a 'one' response indicates that the reference stimulus was more annoying than the noise stimuli. Larger circles indicate more responses, and smaller circles indicate less responses. For higher reference stimulus loudness, the reference is found to be more annoying than the UAM noise stimuli, indicated by larger circles with a 'one' response. Likewise, for lower reference stimulus loudness, the reference is found to be less annoying than the UAM noise stimuli, indicated by larger circles with a 'zero' response.

The circles plotted in Figure 14 represent the total number of individual responses at each level of the reference stimulus. All individual responses are considered part of a binomial distribution, and a logistic curve is fit to these data. The resulting S-curve quantifies the general observations made in the previous paragraph. What is most useful about this analysis is that the loudness level of the reference stimulus needed to make it equally annoying to the UAM noise stimuli can be found. This occurs when the probability that the reference stimulus is more annoying is equal to 0.5. The '+' symbol

FIGURE 14 Probability that the reference sound is more annoying than the UAM noise stimuli as a function of loudness of the reference sound. Larger circles indicate more responses.



in Figure 14 shows that this equal annoyance point occurs for a reference stimulus at a loudness of 9.3 sones. This is a large difference from the loudness of the UAM stimuli (6 sones) and is similar to a 6.3dB difference in SPL.

The other marker on the curve ('x' at 6 sones) gives a probability of around 0.27. This means that when the reference stimulus is presented at the same loudness as the UAM noise stimuli, the UAM noise stimuli are found to be more annoying 73% of the time.

These results indicate that loudness is not the only factor people are responding to when listening to UAM noise. There are other sound quality characteristics (e.g., spectral or temporal) that affect annoyance. In other words, for a reference sound at 6 sones, changes in sound quality may result in perceptual differences that are similar to a loudness change of 3.3 sones (or a change in SPL of 6.3dB).

Effects of Sharpness, Tonality, Impulsiveness, and Fluctuation Strength

The analysis with respect to loudness suggests that there were sound quality characteristics in the UAM noise stimuli that were more annoying than what loudness alone would indicate. The analysis in this section seeks to find the sound quality metrics that caused this difference. Here, an Analysis of Variance (ANOVA) is applied to the 2-level (low or high) factorial test design (see Tables 1 and 2) involving sharpness (S), tonality (T), impulsiveness (I) and fluctuation strength (F). This analysis is for the 136 UAM noise stimuli at the same loudness of 6 sones, so any significant differences in annoyance are not due to differences in loudness defined in ISO 532-1.

The Type III sum of squares ANOVA analysis is summarized in Table 5. This includes main effects based on only one SQ metric as well as interaction effects that depend on the product of two SQ metrics. The shaded rows indicate the factors that are statistically significant (i.e., when the value in the p-Value is less than the significance level of 0.05).

The factors that are found to be significant in this test are the main effects of sharpness ($F_{1,39} = 71.2$, $p < .0001$), tonality ($F_{1,39} = 38.9$, $p < .0001$) and impulsiveness ($F_{1,39} = 130.7$, $p < .0001$). Although fluctuation strength was not found to be significant in this test, it may be because of the limited range present among the final noise stimuli (maximum value was 0.08 vacil). Higher values can be generated with the amplitude modulation described in this paper but were judged during pilot testing to be less than plausible for a real UAM vehicle.

The annoyance effects due to low and high values of the significant main effects (sharpness, tonality, and impulsiveness) are shown in Figure 15. High tonality and impulsiveness both lead to higher annoyance responses. On the other hand, high sharpness reduced annoyance. This effect of sharpness was not expected. It indicates that UAM noise with higher broadband self-noise levels are less annoying than noise with lower broadband levels. Since the noise stimuli were equalized in terms of loudness, low broadband levels resulted in higher levels of loading and thickness noise and higher levels of the added motor tone complex, which may have contributed to the increase in annoyance.

This suggests that certain interaction effects emerge for UAM noise stimuli of equal loudness. Indeed, Table 5 shows interaction effects are significant for sharpness with impulsiveness ($F_{1,39} = 26.5$, $p < .0001$) and for tonality with impulsiveness ($F_{1,39} = 69.7$, $p < .0001$). These effects are shown in Figure 16. When impulsiveness and tonality both increase, annoyance responses also increase. For low impulsiveness, the annoyance responses increase significantly (i.e., nonoverlapping confidence intervals) when tonality is high. Again, higher sharpness was perceived to be less annoying. The interaction between sharpness and impulsiveness suggests that when sharpness is low, lower broadband self-noise levels make the impulsive character of loading and thickness noise more

TABLE 5 4-way ANOVA including 128 UAM noise stimuli (BPFs of 15, 20, 40 and 80 Hz, two flight conditions) based on a 4-factor, 2-level full factorial design. Shaded rows indicate significant factors.

Factor	Sum of Squares	Degrees of Freedom	Mean Squares	F	p-Value
S	317.66	1	317.66	71.21	0
T	173.54	1	173.54	38.9	0
I	583.11	1	583.11	130.72	0
F	3.79	1	3.79	0.85	0.3568
S*T	13.69	1	13.69	3.07	0.0799
S*I	118.35	1	118.35	26.53	0
S*F	0.01	1	0.01	0	0.9576
T*I	69.65	1	69.65	15.61	0.001
T*F	0.14	1	0.14	0.03	0.8620
I*F	0.9	1	0.86	0.19	0.6602
Error	22789.3	5109			
Total	24070.1	5119			

FIGURE 15 Significant main effects contributing to annoyance when considering sharpness (S), tonality (T), impulsiveness (I) and fluctuation strength (F).

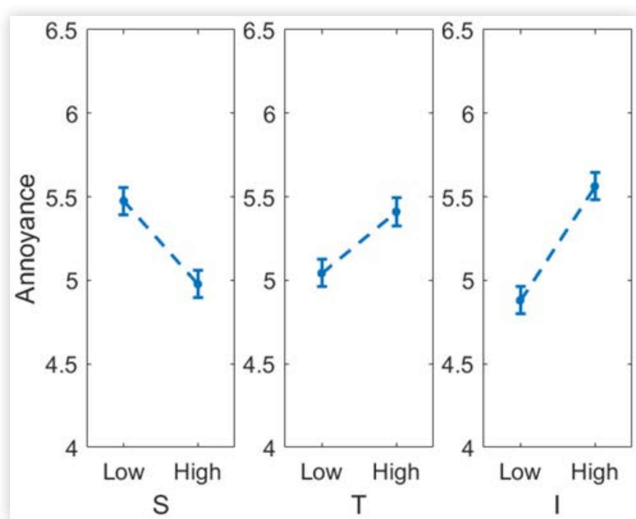


FIGURE 16 Significant interaction effects contributing to annoyance when considering sharpness (S), tonality (T), impulsiveness (I) and fluctuation strength (F).

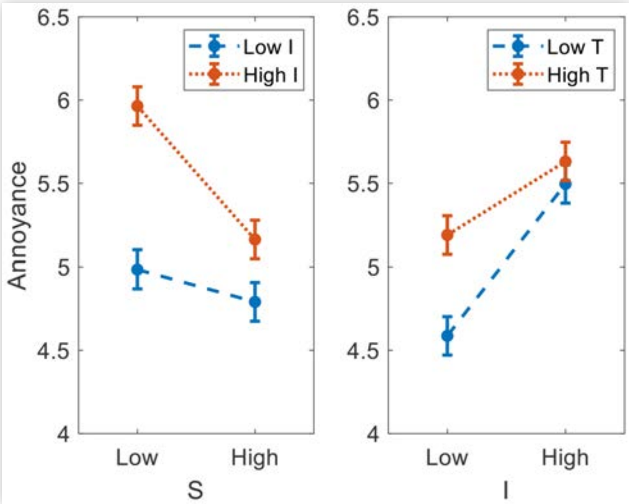
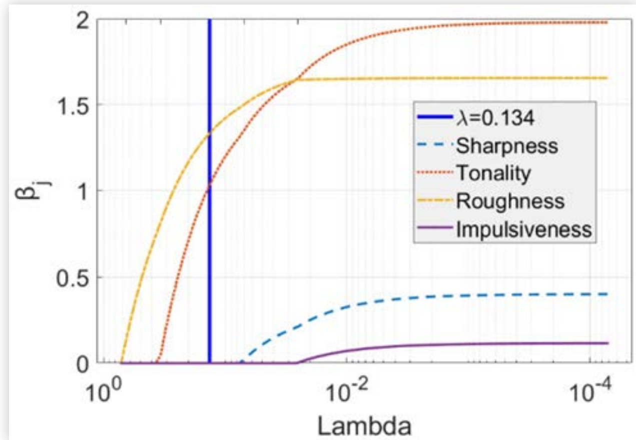


FIGURE 17 Coefficients of sound quality predictors in using the Lasso method for various values of the regularization parameter λ. The solid vertical line represents the λ value with minimum Mean Squared Error (MSE) when the model is validated using tenfold cross-validation.



evident, which increases the annoyance rating. Likewise, high sharpness (i.e., high self-noise) may significantly reduce the annoyance response to impulsiveness (dotted red line on the left-hand plot in Figure 16).

Effect of Roughness

The factorial test design did not include roughness, because a post-processing technique was not found that could manipulate roughness independently of the other sound quality metrics [12]. This excluded roughness from the above ANOVA analysis. However, roughness is expected to influence the annoyance to UAM noise, so the annoyance responses were also analyzed using a regression technique including roughness as a predictor.

The Least Absolute Shrinkage and Selection Operator (LASSO) method is like linear regression but applies a regularization technique to search for a model that includes only the most important predictors. LASSO minimizes the following cost function to yield the predictor coefficients, β_j.

$$\text{cost} = \frac{1}{2}\sum_{i=1}^{N}\left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j\right)^2 + \lambda\sum_{j=1}^{p}\left|\beta_j\right|. \quad (3)$$

Here, λ is the regularization parameter to be varied, and y_i represents the i_{th} response among N observations. x_{ij} is the data at observation i for predictor j among p predictors. β_o and β are scaler and a vector of length p, respectively, which are optimized during the minimization process. Further details can be found in [36].

The LASSO method was applied to the annoyance responses to the 136 stimuli presented at the same loudness. Since loudness did not vary, it was excluded from the analysis. Fluctuation strength was also excluded, because it was found not to be an important factor in this dataset (see Table 5).

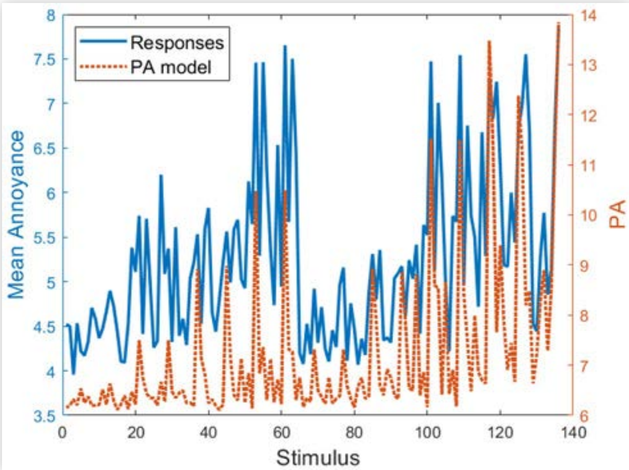
The result from LASSO using sharpness, tonality, roughness, and impulsiveness is shown in Figure 17. The minimum error occurs for λ=0.134 (vertical blue line). In the resulting sparse model, only tonality and roughness remain as predictors; sharpness and impulsiveness are discarded. The resulting β_j coefficients associated with tonality and roughness are 1.03/TU and 1.3/asper, respectively. This indicates that higher tonality and roughness increases annoyance and that changes in sharpness and impulsiveness have smaller effects than tonality and roughness.

Although sharpness and impulsiveness were found to be important predictors in the ANOVA analysis, LASSO showed that roughness and tonality emerge as more important predictors. For a smaller value of λ, however, sharpness would be the next metric to be included in the model and would have a positive coefficient value for the corresponding element in the vector β, indicating a positive correlation with annoyance, a result that differs from the ANOVA analysis shown in Table 5.

To further investigate the importance of roughness, Psychoacoustic Annoyance (PA) using Zwicker's model [5] was calculated and compared with the mean annoyance values from this TUSQ test. It was found that PA and mean annoyance are strongly correlated (Pearson correlation coefficient=0.67). This further strengthens the importance of roughness, because it was the only important contributor in the PA model. Loudness was constant, fluctuation strength was low and found to be insignificant by ANOVA, and tonality and impulsiveness are absent from the Zwicker PA model. Furthermore, only 3 of 136 stimuli had a sharpness value above the threshold to contribute to PA (meaning the sharpness discrepancy between ANOVA and LASSO may be inconsequential), leaving roughness as the only important contributor to the PA calculation. The result of this comparison is shown in Figure 18 where a clear correlation between mean annoyance and the PA prediction can be seen.

Further investigation points to the influence of higher BPFs on the roughness of sound. Here, an increase in BPF (along with its harmonics) from 15 to 80Hz impacts the rate of amplitude modulation and hence the sensation of

FIGURE 18 Comparison between mean annoyance responses and the Zwicker psychoacoustic annoyance model for 136 UAM noise stimuli of the same loudness.



roughness perceived between 15 to 300Hz with peak at 70Hz. A moderate degree of correlation was found between roughness and BPF (Pearson correlation coefficient=0.44) and between BPF and annoyance (Pearson correlation coefficient=0.30).

Summary/Conclusions

This work concerns the annoyance response of human test subjects to simulated noise of an Urban Air Mobility vehicle. With data collected in the NASA Langley Exterior Effects Room, human test subjects gave annoyance ratings to 136 different noise stimuli that were presented at the same loudness level. This part of the test consisted of a full factorial, two-level design based on sharpness, tonality, impulsiveness, and fluctuation strength. In a separate task, the subjects compared their annoyance to a subset of the noise stimuli with a reference sound that was varied in level. The reference sound was a UAM noise stimulus in which the tonal and impulsive characteristics were removed, leaving a steady broadband noise.

The equal annoyance point between the reference sound and the UAM noise stimuli occurred when the reference was 3.3 sones (approximately 6.3dB) higher than the UAM noise stimuli. Since the reference had some temporal variations removed, this shows that some annoying aspects of UAM noise were missing in the bland reference sound and that loudness is not the only sound quality that is important for the subjective rating of UAM noise. In fact, results show that differences in sound quality may result in perceptual differences that are similar to a change in loudness of 3.3 sones.

Qualitative analysis showed the importance of sharpness, tonality, and impulsiveness in the perception of UAM noise. The importance of roughness and tonality were confirmed through a regression analysis method that searched for a sparse model. A psychoacoustic annoyance model confirmed the importance of roughness. Analysis using psychoacoustic models that also include tonality and impulsiveness terms is needed to assess their correlation with mean annoyance responses to UAM noise.

References

1. Thipphavong, D., Apaza, R., Barmore, B., Battiste, V. et al., “Urban Air Mobility Airspace Integration Concepts and Considerations,” in *2018 Aviation Technology, Integration, and Operations Conference*, 2018.
2. Rizzi, S., Huff, D., Boyd, D.D., Bent, P. et al., “Urban Air Mobility Noise: Current Practice, Gaps, and Recommendations,” 2020.
3. Pettingill, N. and Zawodny, N., “Identification and Prediction of Broadband Noise for a Small Quadcopter,” in *Vertical Flight Society 75th Forum*, 2019.
4. Snider, R., “Full-Scale Rotorcraft Broadband Noise Prediction and its Relevance to Civil Noise Certification Criteria,” in *American Helicopter Society 69th Forum*, 2013.
5. Fastl, H. and Zwicker, E., *Psychoacoustics: Facts and Models*, 2nd ed. (Springer, 2013)
6. More, S., *Aircraft Noise Characteristics and Metrics* (West, 2010)
7. Di, G.-Q., Chen, X.-W., Song, K., Zhou, B. et al., “Improvement of Zwicker’s Psychoacoustic Annoyance Model Aiming at Tonal Noises,” *Applied Acoustics* 105 (2016): 164-170.
8. Torija, A., Li, Z., and Chaitanya, P., “Psychoacoustic Modelling of Rotor Noise,” *The Journal of the Acoustical Society of America* 151 (2022): 1804-1815.
9. Bowen, D., “Methods for Providing Design Guidance to Improve Drone Sound Using Community Input,” in *Quiet Drones International e-Symposium on UAV/UAS Noise*, 2020.
10. Lyon, R., *Designing for Product Sound Quality* (Marcel Dekker, Inc., 2000)
11. Rafaelof, M. and Schroeder, A., “Investigation of Machine Learning Algorithms to Model Perception of Sound,” 2018.
12. Krishnamurthy, S., Boucher, M., Christian, A. and Rizzi, S., “Rotorcraft Sound Quality Metric Test 1: Stimuli Generation and Supplemental Analyses,” 2021.
13. Boucher, M., Krishnamurthy, S., Christian, A., and Rizzi, S., “Sound Quality Metric Indicators of Rotorcraft Noise Annoyance Using Multilevel Analysis,” *The Journal of the Acoustical Society of America* 153 (2023): 867-876.
14. Boucher, M., Christian, A., Krishnamurthy, S., and Rizzi, S., “Further Evaluation of the Efficacy of SEL Related to Annoyance to Helicopter Maneuvers,” in *NASA Acoustics Technical Working Group Meeting*, 2020a.
15. Boucher, M., Christian, A., Krishnamurthy, S., and Rizzi, S., “A Perceptual Evaluation of the Efficacy of Sound Exposure Level in the Rating of Annoyance to Helicopter Noise,” in *Vertical Flight Society 76th Forum*, 2020b.
16. Silva, C., Johnson, W., Solis, E., Patterson, M. et al., “VTOL Urban Air Mobility Concept Vehicles for Technology Development," in *2018 Aviation Technology, Integration, and Operations Conference*, 2018.

17. DIN 45631/A1, DIN 45631/A1, “Calculation of Loudness Level and Loudness from the Sound Spectrum - Zwicker Method - Amendment 1: Calculation of the Loudness Of Time-Variant Sounds,” 2009.
18. HEAD Acoustics, “ArtemiS Suite 13.6.”
19. ISO, “ISO 532-1:2017 Acoustics–Methods for Calculating Loudness–Part 1: Zwicker Method,” 2017.
20. Aumann, A., Tuttle, B., Chapin, W., and Rizzi, S., “The NASA Auralization Framework and Plugin Architecture,” in *InterNoise 2015*, San Francisco, CA, USA, 2015.
21. DIN 45692, “DIN 45692: Measurement Technique for the Simulation of the Auditory Sensation of Sharpness,” 2009.
22. Sottek, R., “Modelle zur Signalverarbeitung im menschlichen Gehör (in German),” 1993.
23. Sottek, R. and Genuit, K., “Models of Signal Processing in Human Hearing,” *AEU - International Journal of Electronics and Communications* 59 (2005): 157-165.
24. Bray, W., “A New Psychoacoustic Method for Reliable Measurement of Tonalities According to Perception,” in *Inter-Noise 2018*, Chicago, 2018.
25. Johnson, W., *Comprehensive Analytical Rotorcraft Model of Rotorcraft Aerodynamics and Dynamics, Version 4.10*, 2017.
26. Lopes, L. and Burley, C., “ANOPP2 User’s and Manual and Version 1.2,” in NASA/TM-2016-219342, 2016.
27. Farassat, F. and Succi, G., “A Review of Propeller Discrete Frequency Noise Prediction Technology with Emphasis on Two Current Methods for Time Domain Calculations,” *J. of Sound and Vib.* 71 (1980): 399-419.
28. Brentner, K., “Prediction of Helicopter Rotor Discrete Frequency Noise,” 1986.
29. Brooks, T., Pope, D., and Marcolini, M., “Airfoil Self-Noise and Prediction,” in NASA/RP-1218, 1989.
30. Rizzi, S., Letica, S., Boyd, D., and Lopes, L., “Prediction-Based Approaches for Generation of Noise-Power-Distance Data with Application to Urban Air Mobility Vehicles," in *28th AIAA/CEAS Aeroacoustics 2022 Conference*, 2022.
31. Krishnamurthy, S., Rizzi, S., Cheng, R., Boyd, D. et al., “Prediction-Based Auralization of a Multirotor Urban Air Mobility Vehicle,” 2021.
32. Krishnamurthy, S., Aumann, A., and Rizzi, S., “A Synthesis Plugin for Auralization of Rotor Self Noise,” in *AIAA Aviation 2021 Forum, Virtual Event*, 2021.
33. Tuttle, B., Rafaelof, M., Aumann, A., Rizzi, S. et al., “NASA Auralization Framework Psychoacoustic Analysis Library (NAFPAL),” NASA [Unpublished], Hampton, VA, 2021.
34. Christian, A., Caston, J., and Greenwood, E., “Regarding the Perceptual Significance and Characterization of Broadband Components of Helicopter Noise,” in *Vertical Fligth Society Forum 75*, 2019.
35. Luz, G., “Noise Sensitivity Rating of Individuals,” *Sound and Vibration* 39 (2005): 14-17.
36. Hastie, T., Ribshirani, R., and Friedman, J., *The Elements of Statistical Learning, Data Mining Inference and Perdition* (New York: Springer, 2017)

Contact Information

For further information on this work, please contact the first author at matthew.a.boucher@nasa.gov.

Acknowledgments

The authors would like to thank Brian Tuttle (Analytical Mechanics Associates, Inc.) for help preparing the auralizations and executing the test, Aric Aumann (Analytical Services and Materials, Inc.) for help executing the test, Erin Thomas (Analytical Mechanics Associates, Inc.) for recruiting test subjects and performing hearing screenings, and Kevin Shepherd for useful discussions. The authors also gratefully acknowledge the support from Ran Cabell, Susan Gorton, Benny Lunsford, Noah Schiller, and the NASA Revolutionary Vertical Lift Technology Project.

ISSN 0148-7191