NASA/TM–20205008997

Rotorcraft Sound Quality Metric Test 1: Stimuli Generation and Supplemental Analyses

Siddhartha Krishnamurthy
Langley Research Center, Hampton, Virginia

Matthew A. Boucher
Langley Research Center, Hampton, Virginia

Andrew W. Christian
Langley Research Center, Hampton, Virginia

Stephen A. Rizzi
Langley Research Center, Hampton, Virginia

September 2021

NASA STI Program Report Series

Since its founding, NASA has been dedicated to the advancement of aeronautics and space science. The NASA scientific and technical information (STI) program plays a key part in helping NASA maintain this important role.

The NASA STI program operates under the auspices of the Agency Chief Information Officer. It collects, organizes, provides for archiving, and disseminates NASA's STI. The NASA STI program provides access to the NTRS Registered and its public interface, the NASA Technical Reports Server, thus providing one of the largest collections of aeronautical and space science STI in the world. Results are published in both non-NASA channels and by NASA in the NASA STI Report Series, which includes the following report types:

- TECHNICAL PUBLICATION. Reports of completed research or a major significant phase of research that present the results of NASA Programs and include extensive data or theoretical analysis. Includes compilations of significant scientific and technical data and information deemed to be of continuing reference value. NASA counterpart of peer-reviewed formal professional papers but has less stringent limitations on manuscript length and extent of graphic presentations.

- TECHNICAL MEMORANDUM. Scientific and technical findings that are preliminary or of specialized interest, e.g., quick release reports, working papers, and bibliographies that contain minimal annotation. Does not contain extensive analysis.

- CONTRACTOR REPORT. Scientific and technical findings by NASA-sponsored contractors and grantees.

- CONFERENCE PUBLICATION. Collected papers from scientific and technical conferences, symposia, seminars, or other meetings sponsored or co-sponsored by NASA.

- SPECIAL PUBLICATION. Scientific, technical, or historical information from NASA programs, projects, and missions, often concerned with subjects having substantial public interest.

- TECHNICAL TRANSLATION. English-language translations of foreign scientific and technical material pertinent to NASA's mission.

Specialized services also include organizing and publishing research results, distributing specialized research announcements and feeds, providing information desk and personal search support, and enabling data exchange services.

For more information about the NASA STI program, see the following:

- Access the NASA STI program home page at http://www.sti.nasa.gov

- Help desk contact information:

https://www.sti.nasa.gov/sti-contact-form/
and select the "General" help request type.

NASA/TM–20205008997

Rotorcraft Sound Quality Metric Test 1: Stimuli Generation and Supplemental Analyses

Siddhartha Krishnamurthy
Langley Research Center, Hampton, Virginia

Matthew A. Boucher
Langley Research Center, Hampton, Virginia

Andrew W. Christian
Langley Research Center, Hampton, Virginia

Stephen A. Rizzi
Langley Research Center, Hampton, Virginia

National Aeronautics and
Space Administration

Langley Research Center
Hampton, Virginia 23681-2199

September 2021

Available from:

NASA STI Program / Mail Stop 148
NASA Langley Research Center
Hampton, VA 23681-2199
Fax: 757-864-6500

Table of Contents

1 Overview

A psychoacoustic test was conducted at the NASA Langley Research Center Exterior Effects Room (EER) to assess annoyance to simulated helicopter sounds over a range of sound quality (SQ) metric values. Initial findings identified important SQ metrics as sharpness, tonality, and fluctuation strength. This document is a supplement to the initial findings in which the following are discussed: (i) a detailed treatment of the sound generation process, (ii) the impact of analyzing results with stimuli measured in the EER instead of the intended synthesized stimuli, (iii) an evaluation of annoyance responses with certification metrics, and (iv) adjunct analyses related to the test methodology.

2 Introduction

The Rotorcraft Sound Quality Metric (RoQM-I-2017) psychoacoustic test took place in December 2017 at the NASA Langley Research Center EER to capture individual annoyance to simulated helicopter sounds of varying SQ. Each sound SQ was quantified by a set of SQ metrics. Krishnamurthy et al. [1] briefly discussed generation of test stimuli, described how the test was conducted, and provided initial analysis results on which SQ metrics were important to describing annoyance toward helicopter sounds. Boucher et al. [2] used multilevel analysis on the test data to account for variablity among test subjects when discerning which SQ metrics were important to the annoyance response. The SQ metrics considered in those previous works and in this document are loudness, sharpness, tonality, fluctuation strength, impulsiveness, and roughness. They will be defined and referenced in section 3. This paper documents how the RoQM-I-2017 test stimuli sounds were generated and documents supplemental analyses of test results not captured in references [1] or [2].

Section 3 of this paper details the generation of all RoQM-I-2017 test stimuli. Different test sounds were generated by perturbing blade passage frequency (BPF) harmonic parameters, such as magnitude and phase, of a baseline simulated helicopter sound. The perturbation methods were intended to keep the simulated sounds subjectively similar to actual helicopter sounds and vary a single SQ metric while leaving other metrics roughly constant. Krishnamurthy et al. described the sound generation process for the subset of test sounds that primarily changed in tonality. To provide documentation on how all test sounds were generated, this paper describes every method used to affect each SQ metric.

A post-test recording of test sounds at subject locations in the EER revealed SQ differences between the recordings and stimuli generated for playback in the EER. Section 4 of this paper discusses the impact of these SQ differences on the analysis results.

Section 5 explores the relationship between the annoyance response and aircraft noise certification metric values of test sounds such as their A-weighted sound pressure levels. Krishnamurthy et al. and Boucher et al. did not analyze these metrics with annoyance, as they instead focused on analyses with SQ metrics. This paper provides results using linear regression between certification metric values and the mean test subject annoyance responses to sounds.

Section 6 briefly documents adjunct analyses of the test methodology. This includes analyses on the probability distribution of test sound responses, the annoyance response to repeated test sounds, the mean annoyance response over test duration, and the annoyance response to practice session sounds, which were played to subjects before the main test. Section 6 also documents the annoyance response to all test sounds.

3 Test Signal Generation

This section will discuss generating simulated helicopter sounds for the RoQM-I-2017 psychoacoustic test. Section 12 explains how to obtain the audible sounds generated for this test online [3]. The ArtemiS Suite from HEAD Acoustics [4] was used to calculate the loudness, sharpness, tonality, roughness, impulsiveness, and fluctuation strength SQ metrics of the sounds. Boucher et al. [2] gives a more detailed introduction to these metrics, but their definitions and their calculation methods are also provided here:

- Loudness: the sensation that corresponds most closely with sound intensity [5] and indicates the perception of sound level. It is measured in Sones. ArtemiS calculated loudness using the DIN standard 45631/A1 time-varying loudness definition [6].
- Sharpness: a measure of the high-frequency content of a sound [7], measured in Acums. This test used the DIN standard 45692 calculation [8] in ArtemiS.
- Tonality: a measure of the prominence of narrowband tones compared to other tones in the sound spectrum, in units of TU. Tonality was calculated in ArtemiS using a Tonality vs. time metric that is based on Aures/Terhardt calculation of tonality [9].
- Roughness: a measure of sound modulations between 20 Hz and 300 Hz in Aspers. It is calculated in ArtemiS using a hearing model [10].
- Fluctuation strength: a measure of sound modulations below 20 Hz (units in Vacils) and peaks around 4 Hz [5]. It is also calculated with a hearing model [10].
- Impulsiveness: a measure of short bursts of sound with rapidly changing loudness [7], in units of IU. Its calculation also uses a hearing model [10].

This section will cover how sounds generated for the test targeted each of these SQ metrics.

3.1 Overview

A direct approach for human subject testing could be to present recordings of different rotorcraft flyovers made during NASA acoustic flight tests [11, 12, 13] to test subjects who would then rate their annoyance to the sounds. There are two main problems with this direct approach. First, the SQ values of the sound recordings can change by large amounts during a flyover. Determining the portion of the flyover that results in a subject's annoyance response is challenging. Second, the sound stimuli presented to subjects should vary in only one or two metrics while other metrics remain roughly constant to provide control over the specific SQ metric values presented to subjects. Finding such sequences of sounds from flyover recordings is difficult, especially if recording methods do not provide a means of determining source noise [14]. Determining source noise enables capturing and extending periodic portions of flyover sounds with constant SQ.

To address these issues, a baseline signal was extracted from one of the recordings and then manipulated to create a wide range of stimuli with controlled variations in specific SQ metrics. The sound generation method in this test began with a flyover recording of a Eurocopter AStar (AS350) helicopter [15]. This vehicle has a main rotor with three blades and a tail rotor with two blades. The blade passage frequency (BPF) of the main rotor is 19.2 Hz. The BPF of the tail rotor is 69.8 Hz. The time history of this recording is shown in figure 1 along with emission angle from the helicopter to the recording microphone. The SQ metrics of this recording, as calculated by ArtemiS, are shown in figure 2. Except for fluctuation strength, metric values change considerably over the course of the flyover.

Using methods from Greenwood and Schmitz [14], average pressure-time histories for single complete blade passages of the main and the tail rotors were extracted from a six second segment of the recording near the 21-second portion of the flyover, corresponding to emission angles of around 15 degrees elevation and approximately –1 (negative one) degrees azimuth. The blade passage signals are given in figure 3. Magnitudes and phases of the main and tail rotor harmonics can be found from these blade passage signals. These magnitudes and phases are used in an additive synthesis technique [16] to generate a longer duration sound from BPF harmonics. This technique effectively repeats the blade passage signals from figure 3 for any desired duration. The resulting simulated sound is referred to as the "AS350 baseline."

Figure 4 shows one second of the AS350 baseline. Note that the seemingly random variations in figure 4 are due to the combination of two deterministic periodic signals at different BPFs. The SQ metric values of this AS350 baseline signal are now nominally constant, as shown in figure 5.

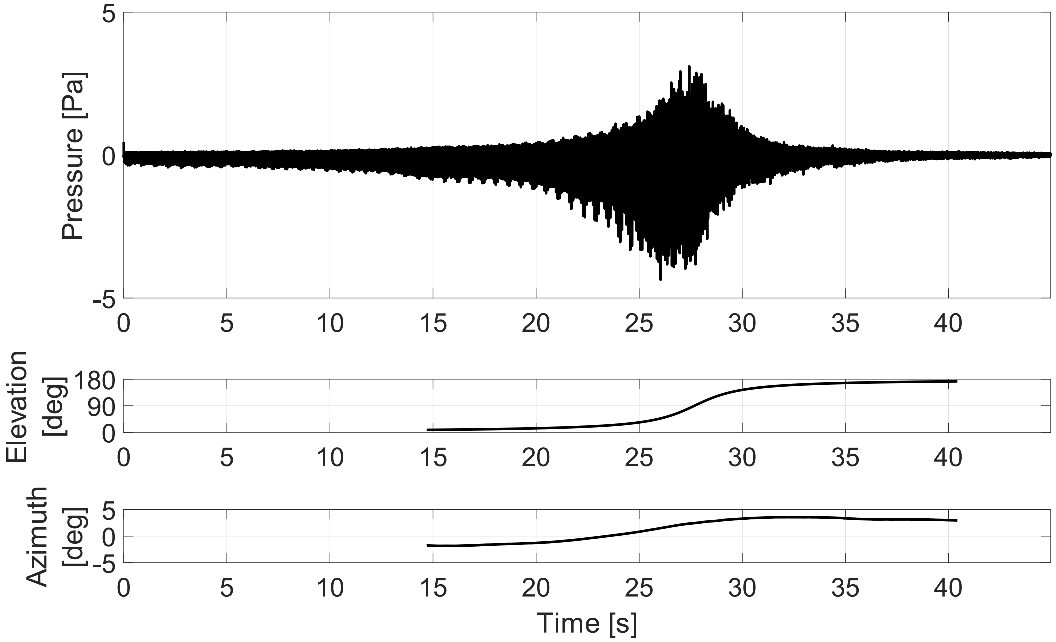

Figure 1. Time history of AS350 helicopter flyover recording.

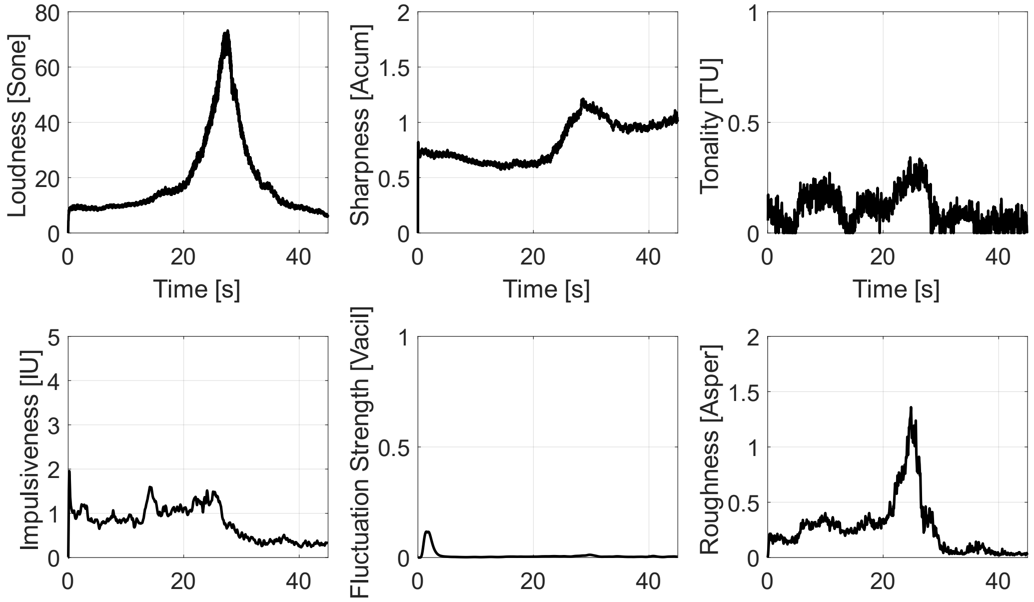

Figure 2. Sound quality metrics of AS350 helicopter flyover recording.

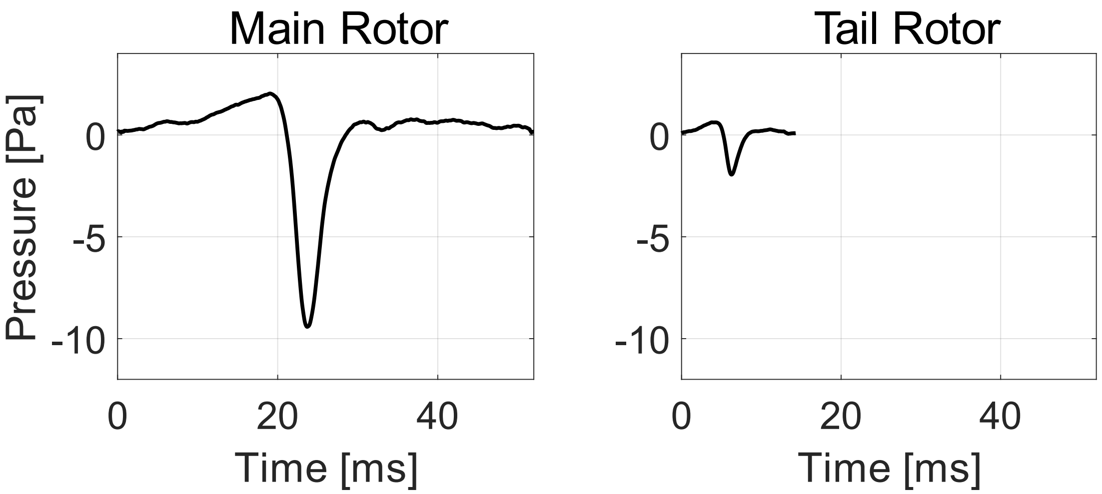

Figure 3. AS350 helicopter main and tail rotor blade passage signals.

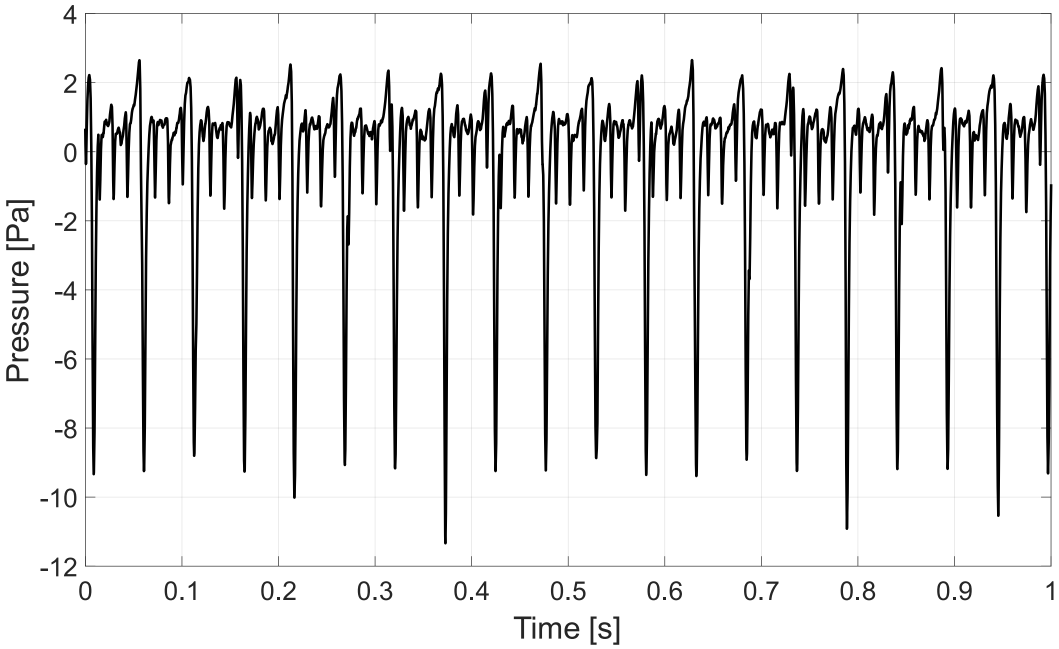

Figure 4. Simulated periodic AS350 baseline helicopter sound pressure time history.

3.1.1 Exceedance Levels

Since there are still relatively small variations in some metrics, such as tonality and roughness in the figure 5 example, the 5% exceedance levels of the metrics were used to describe the metric levels for the duration of the sound. The exceedance level is the value of the metric that is exceeded x% of the time within a sound segment. If an SQ value is nearly constant over the sound duration, as it is for sharpness, impulsiveness, and fluctuation strength in figure 5, the x% exceedance level will be approximately equal to the constant level.

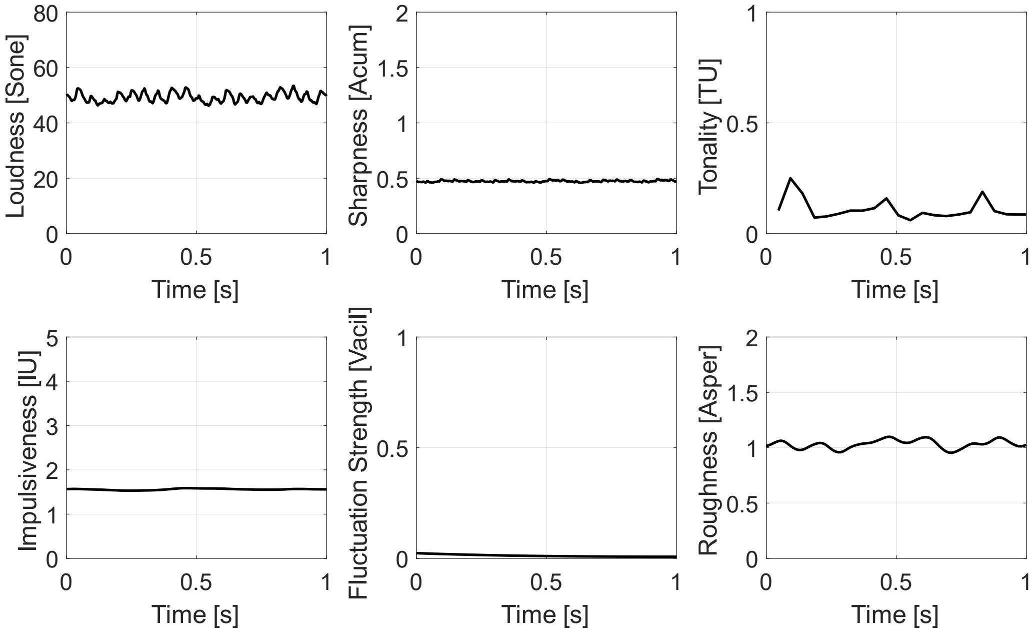

Figure 5. SQ metrics of AS350 baseline.

3.1.2 Parameter Perturbation Methods

To create sequences of constant simulated helicopter sound snippets that vary in one or two metric values, one could manipulate the baseline signal to achieve a range of SQ metric values. However, it was important that the simulated sounds for the psychoacoustic test be subjectively similar to a helicopter. This could be achieved if the manipulations maintained multiple blade passage harmonics of a main and tail rotor. A number of parametric perturbations to the baseline signal were devised, subject to this BPF harmonic constraint, where each perturbation produced a change in either a single SQ metric or a small number of SQ metrics. For each perturbation method, several sounds were generated to produce a range of 5% exceedance values for the affected SQ metric or metrics. All generated sounds were six seconds in duration. Sections 3.2-3.7 below describe the specific parametric perturbation methods used to target individual SQ metrics or combinations of metrics.

3.1.3 Metric Boundaries

Flyover recordings were examined to determine realistic minimum and maximum SQ metric values expected from helicopter sounds. These recordings consisted of 172 flyovers of the Bell 206, Messerschmitt-Bolkow-Blohm (Bo) 105, McDonnell Douglas (MD) 520N, MD 902, Mil (Mi) 8, and Bell XV-15 rotorcraft [11] [12] [13]. Table 1 shows the SQ metric limits that were used to bound the generated test sounds. These limiting values were not the absolute minimum and maximum values obtained from the recordings. They were obtained by examining the distribution of recorded SQ values to determine cutoff points that captured the majority of SQ values in the flyover recordings.

The SQ values for the AS350 baseline sound are provided in the right column for comparison with the minimum and maximum SQ values. The baseline SQ values trend closer to the lower range of SQ values, so there was some concern that an alternative baseline with SQ values closer to the midrange of table 1 would be preferred. As a check, simulated sounds with SQ values close to the maximum in table 1 were judged by the authors to be subjectively similar to helicopter noise, so the AS350 baseline was deemed acceptable for this test.

Table 1. Minimum and maximum SQ values used, based on recordings.

Sound Quality Metric (Unit)	Minimum Value	Maximum Value	AS350 Baseline Value
Fluctuation Strength (Vacil)	0.0064	0.8516	0.0803
Impulsiveness (IU)	0.0107	5.0470	1.4863
Tonality (TU)	0.0000	0.8895	0.0991
Sharpness (Acum)	0.3577	1.6400	0.4652
Roughness (Asper)	0.0529	2.0000	0.9760

3.1.4 Identical Loudness for all Sounds

Based on results of previous research, for example McMullen and Davies [17] and More [18], the loudness metric is assumed to dominate subject response to test sounds. The predictive abilities of the other metrics could be difficult to separate if sounds varied considerably in loudness. Therefore, each generated sound was adjusted during post processing so its loudness level was approximately 10 Sone, corresponding to the level of an automobile from several meters away. This level was selected through pilot testing the sound reproduction in the EER. Because each metric has some dependency on loudness, the metrics had to be examined after the loudness adjustment to see if they were still within their desired SQ range. The SQ metrics for the AS350 baseline after adjusting loudness to 10 Sone are shown by the blue traces in figure 6. The black traces are the SQ metric values from figure 5 before the loudness adjustment. Except for roughness, the SQ metric value changes were relatively small and all were still within the minimum and maximum SQ limits in table 1.

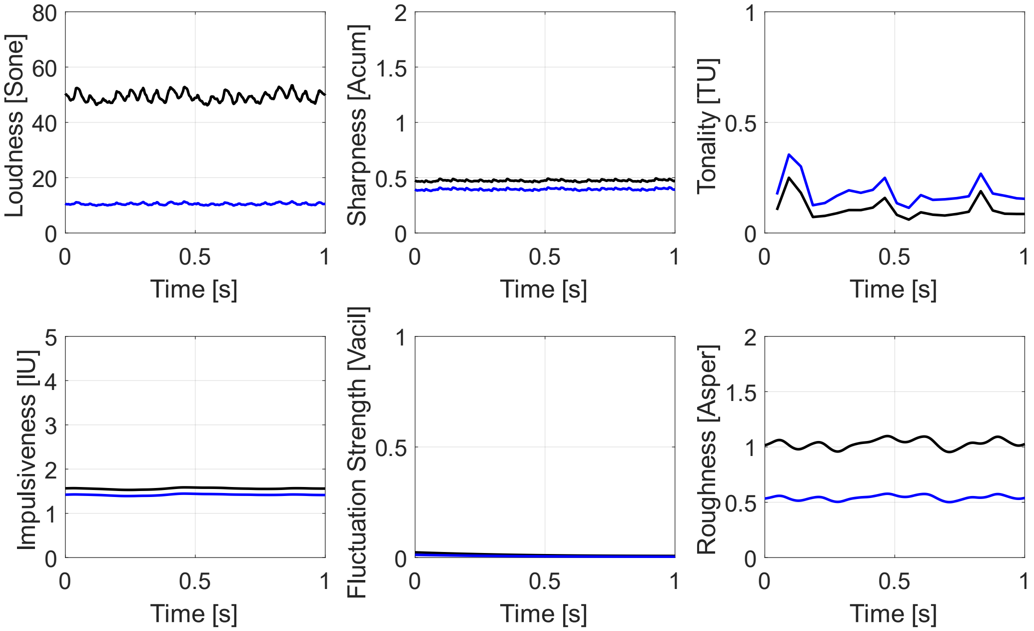

Figure 6. SQ metrics for AS350 baseline before (black traces) and after (blue traces) adjusting loudness to 10 Sone.

The method to adjust the loudness used the ISO 532-2 stationary loudness definition [19]. However, recall that the loudness of test sounds was calculated using the DIN standard 45631/A1 time-varying loudness definition [6]. As a consequence, the 5% exceedance levels of loudness for the test sounds varied slightly from 10 Sone. Before conducting the test, the slight deviations were considered acceptable. With fluctuation strength having a dependence on loudness, post-test analyses in Krishnamurthy et al. [1] showed the loudness deviations introduced an artificial uncertainty in the degree to which fluctuation strength described the annoyance response. Using the ISO 532-2 stationary loudness definition for quantifying test sound loudness may reduce the artificial uncertainty, but this approach has not been tested and was not used in the analyses covered by this current document.

3.2 Perturbation Methods to Affect Fluctuation Strength

The fluctuation strength SQ metric of the AS350 baseline can be affected by modulating the signal. Though modulations in both amplitude and frequency contribute to this auditory sensation [5], amplitude modulation was selected to affect fluctuation strength. Frequency or phase modulation was not explored. Two amplitude modulation methods were used: a deterministic method and a stochastic method. For both methods, the amplitude modulated signal, $a(t)$, is given by

$$a(t) = (1 + m(t)) \times b(t) \tag{1}$$

where $m(t)$ is the modulating signal, and $b(t)$ is the baseline signal.

The deterministic method affected the fluctuation strength by varying a modulation frequency and a modulation index. The modulating signal is given by

$$m(t) = M \cos(2\pi f_m t) \tag{2}$$

in which M is the modulation index and f_m is the modulation frequency. The modulation index is related to the root mean square (RMS) of the modulating signal, m_{RMS}, by

$$m_{RMS} = \frac{M}{\sqrt{2}} \ . \tag{3}$$

For the amplitude modulation perturbation method, four sets of five sounds each were formed, where each set had a different modulation frequency of either 2, 4, 8, or 12 Hz. Within each modulation frequency set, the modulation index was varied over the five sounds.

In addition, the amplitude modulation could be achieved by a stochastic pertubation method using bandlimited noise. This type of modulation could potentially come from unsteady vehicle operation. The resulting modulating signal is

$$m(t) = \frac{T_{RMS}}{\left\| W(t, f_m) \right\|} W(t, f_m) \tag{4}$$

in which $W(t, f_m)$ is a bandlimited random noise signal with a bandwidth of $f_m/\sqrt{2}$ centered around the modulation frequency, f_m. The quantity $\|W(t, f_m)\|$ is the norm of $W(t, f_m)$ taken over time, which is the RMS of the noise signal. The scalar quanitity T_{RMS} sets the RMS of the modulating signal, or

$$m_{RMS} = T_{RMS} \ . \tag{5}$$

Using Equation (4), the RMS of the modulating signal can be changed by changing the value of T_{RMS} without changing $W(t, f_m)$. This technique allows the time history of $W(t, f_m)$ to be identical in multiple stimuli with different modulating signal RMS values so that $W(t, f_m)$ is not considered as a perturbation parameter. Two sets of five sounds each were generated using this stochastic amplitude modulation perturbation method.

Table 2 lists the seven sets of perturbations using the two amplitude modulation perturbation methods (deterministic and stochastic) targeted to create test sounds with changing fluctuation strength. Sets 1–4 were formed using equations (1) and (2), the determinisitic method, and sets 5 and 6 were formed using equations (1) and (4), the stochastic method. Within each of the sets 5 and 6, $W(t, f_m)$ was identical for all sounds in the set, and only T_{RMS} was varied. Sets 1–6 used the AS350 baseline for $b(t)$. Set 7 was formed using equations (1) and (2), but the tail rotor was removed from the baseline signal and the main rotor blade passage harmonics had random phases.

Table 2. Sets of sounds to change fluctuation strength.

Set	**Base Signal**	**Modulator**	f_m **(Hz)**	**1st Sound** m_{RMS}	**2nd Sound** m_{RMS}	**3rd Sound** m_{RMS}	**4th Sound** m_{RMS}	**5th Sound** m_{RMS}
1	AS350 Baseline	Sine Wave	2	0.0000	0.3989	0.5439	0.6346	0.7071
2	AS350 Baseline	Sine Wave	4	0.0000	0.3989	0.5439	0.6346	0.7071
3	AS350 Baseline	Sine Wave	8	0.0000	0.3989	0.5439	0.6346	0.7071
4	AS350 Baseline	Sine Wave	12	0.0000	0.3989	0.5439	0.6346	0.7071
5	AS350 Baseline	Bandlimited Noise	4	0.0000	0.4113	0.5656	0.7199	0.9255
6	AS350 Baseline	Bandlimited Noise	8	0.0000	0.4113	0.5656	0.7199	0.9255
7	AS350 Baseline main rotor only, harmonics have random phase	Sine Wave	2	0.0000	0.3989	0.5439	0.6346	0.7071

Table 2 shows variations of the RMS of the modulating signal for the sounds in each set. For sounds generated using the sine wave modulator in equation (2), the RMS values in table 2 are calculated from equation (3). For sounds generated using the bandlimited noise modulator in equation (4), the RMS values in table 2 are calculated from equation (5). For the first sound in each set, the base signal, $b(t)$, is reproduced. Note that sound 1 is the same for sets 1–6. The modulating signal RMS value changes are identical for sets using equation (2) and are identical for the two sets using equation (4). RMS values are unitless, and range from 0 to 0.9255 for the sounds generated.

Modulating signal RMS values provide a means to compare stimuli against the baseline sound through a bound on the amplitude modulated signal mean value. Let $\overline{|a|}$ be the mean of the absolute value of the amplitude modulated signal, $a(t)$, and let b_{RMS} be the RMS of the baseline signal. Using equation (1), the Cauchy-Schwarz inequality, and the triangle inequality, one can compute an upper bound to $\overline{|a|}$ as

$$\overline{|a|} \leq \left(1 + m_{RMS}\right) \times b_{RMS}\,. \tag{6}$$

The RMS of the AS350 baseline is $b_{RMS} = 2.27$ Pa, and the RMS of the main rotor only with random harmonic phases is $b_{RMS} = 2.18$ Pa. Using these b_{RMS} values, equation (6), and the m_{RMS} values in table 2, one can determine how the mean of the absolute value of the amplitude modulated signal changes over the sound sets. As an example, $\overline{|a|}$ can be up to almost twice that of the baseline signal RMS for the fifth sounds in sets 5 and 6.

Figure 7 shows the fluctuation strength values of the five sounds in each of the seven sets from table 2. The 5% exceedance values of fluctuation strength over the sound duration are plotted versus modulator RMS. The magenta colored lines at the top and bottom of the figure are the fluctuation strength metric bounds obtained from rotorcraft flyover recordings as given in table 1.

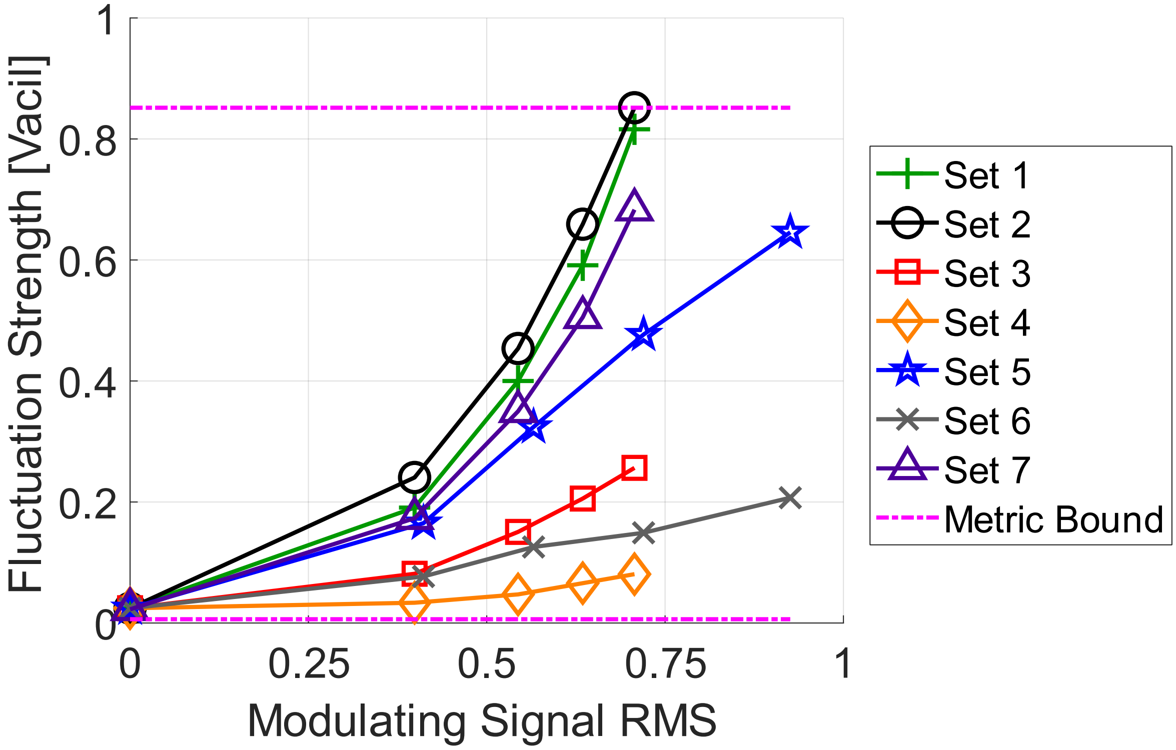

Figure 7. Fluctuation strength, 5% exceedance, of different modulating signal RMS perturbation sets.

For a particular modulation frequency and RMS, the stochastic method gives lower fluctuation strength than the deterministic method. Set 5 has lower fluctuation strength than set 2, although both sets use the same 4 Hz modulation frequency. Set 6 has lower fluctuation strength than set 3, although both sets use the same 8 Hz modulation frequency.

For a particular RMS value, the fluctuation strength increases with modulation frequency up to 4 Hz before decreasing since the maximum fluctuation strength for any arbitrary signal occurs at 4 Hz modulation [5]. Set 1, which uses 2 Hz modulation, has lower fluctuation strength than set 2, which uses 4 Hz modulation, but set 2 has higher fluctuation strength than sets 3 and 4, which use 8 Hz and 12 Hz modulation, respectively.

The perturbation methods were chosen to minimize changes to the other SQ metrics. However, the requirement to maintain multiple blade passage harmonics to maintain similarity to helicopter sounds made it challenging to always keep other metrics constant. Figure 8 shows the effects of the fluctuation strength perturbation methods on other SQ metrics. Here, the 5% exceedance values of the other SQ metrics including loudness are used. To change fluctuation strength over an acceptable range, some modest change in loudness was accepted. Effects of the loudness variation on analyses with fluctuation strength were discussed in Krishnamurthy et al. [1]. Other metrics were roughly constant, but there is some slight variation in tonality, and there is significant variation in impulsiveness for sets 3 and 4.

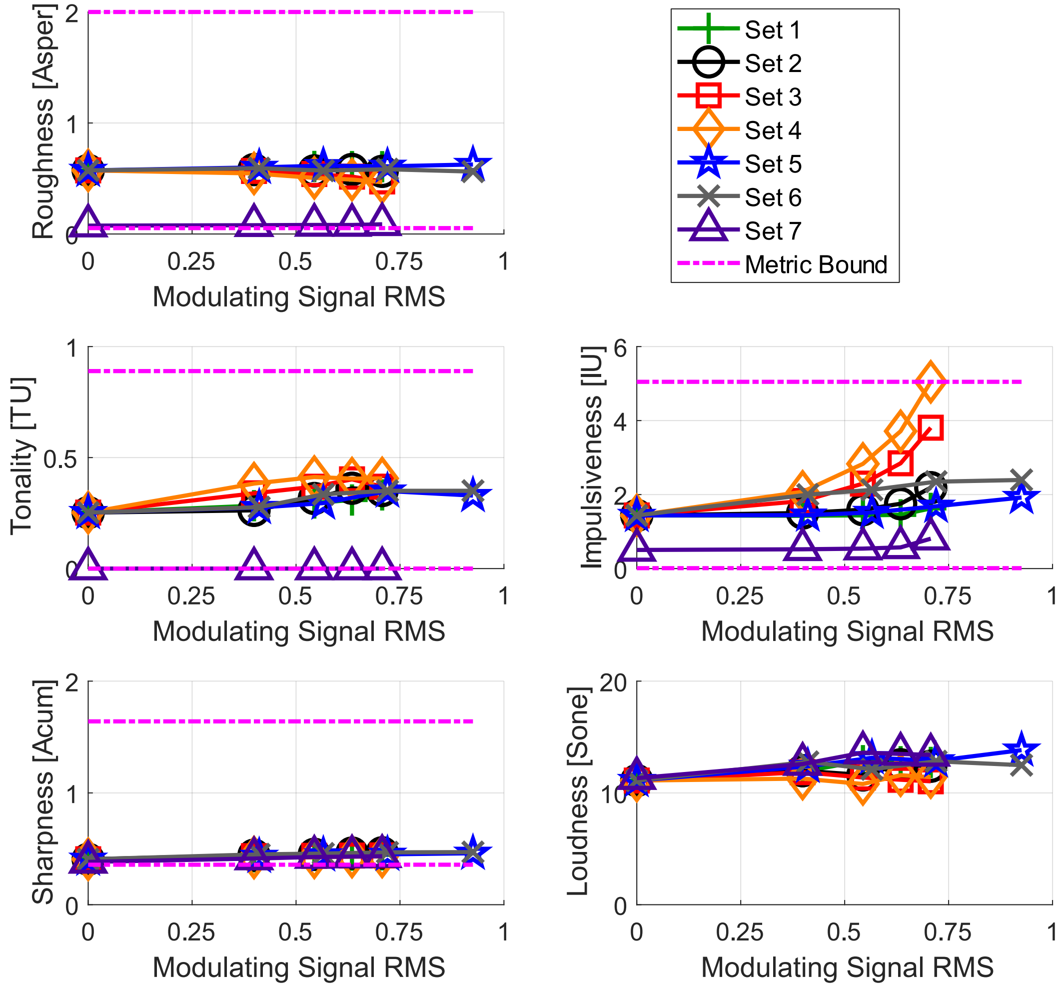

Figure 8. SQ metric 5% exceedance values of RMS perturbation sets targted at changing fluctuation strength.

3.3 Perturbation Method to Affect Impulsiveness

A perturbation method based on adjusting the phases of main rotor blade passage signal harmonic tones was used to primarily affect impulsiveness. No tail rotor was included in sounds that targeted impulsiveness. Due to the relatively high BPF of the tail rotor relative to the main rotor, including a tail rotor sound reduced the prominence of the main rotor pulses and the range over which impulsiveness varied.

Most of the sounds were based on perturbations of the AS350 baseline signal. However, an additional baseline, acquired from a recording of an AS350 helicopter experiencing blade-vortex interaction (BVI), was used to generate a subset of the sounds. This additional sound is referred to as the “BVI baseline.”

Three sets of sounds were generated for this method. For each sound in a set, the main rotor blade passage harmonic tone phases were created by linearly interpolating the phase at each tone between its corresponding value in the first and last sounds. The quantity $\phi_{k,n}$ is the phase of the nth main rotor blade passage harmonic tone out of N harmonics from sound number k in a set. The elements of vector $\vec{\phi}_k$ are the N harmonic phases of sound k so that

$$\vec{\phi}_k = \begin{bmatrix} \phi_{k,1} \\ \phi_{k,2} \\ \vdots \\ \phi_{k,N-1} \\ \phi_{k,N} \end{bmatrix} . \tag{7}$$

With up to K sounds in a set, $\vec{\phi}_1$ contains the blade passage harmonic tone phases of sound 1, and $\vec{\phi}_K$ contains the harmonic tone phases of sound K. The phases in equation (7) can be found by

$$\vec{\phi}_k = \left(\frac{K-k}{K-1} \right) \vec{\phi}_1 + \frac{k-1}{K-1} \vec{\phi}_K , k \in \{1, 2, \ldots, K-1, K\} . \tag{8}$$

The phase quantities in equations (7) and (8) are between –180 and 180 degrees.[1]
The three sets of sounds that targeted impulsiveness are:

1. Change AS350 baseline main rotor blade passage harmonic tone phases from being random to being that of original baseline phases over $K = 5$ sounds.
2. Change AS350 baseline main rotor blade passage harmonic tone phases from being random to being identical over $K = 6$ sounds. Identical phase is when all N harmonic tones of the blade passage signal have a phase of zero degrees. A signal with all equal phases (and equal magnitudes) would be an impulse train and have maximum impulsiveness for a given BPF.
3. Change AS350 baseline main rotor blade passage harmonic tone phases from being random to being the same as those of the BVI baseline over $K = 5$ sounds. For this set of sounds, harmonic magnitudes are also linearly interpolated over the five sounds from being that of the AS350 baseline main rotor blade passage to being that of the BVI baseline main rotor blade passage.

To create sounds for set 3, let the nth blade passage harmonic magnitude of the kth sound in the set be $M_{k,n}$. The vector $\vec{M}_k$ contains the elements of the harmonic magnitudes in

$$\vec{M}_k = \begin{bmatrix} M_{k,1} \\ M_{k,2} \\ \vdots \\ M_{k,N-1} \\ M_{k,N} \end{bmatrix} . \tag{9}$$

The elements of vector $\vec{M}_1$ are the blade passage harmonic magnitudes of the AS350 baseline main rotor, and the elements of vector $\vec{M}_5$ are the elements of the BVI baseline main rotor blade passage harmonic magnitudes. The elements of the vector in equation (9) can be found by

$$\vec{M}_k = \left(\frac{5-k}{5-1} \right) \vec{M}_1 + \frac{k-1}{5-1} \vec{M}_5 , k \in \{1, 2, 3, 4, 5\} . \tag{10}$$

---

[1] Spherical linear interpolation was not done, so phases in equation (8) are not always interpolated along the shortest circular arc between the phase values of the first and last sounds of a set [28].

Table 3 lists the first and last sounds for each set. Sound 1, the AS350 baseline main rotor with random phase, is the same for all sets. The same random number generation seed was used for all sets, and so sound 1 for the three sets is identical. Sound 1 for all of the sets is also identical to sound 1 for set 7 from the fluctuation strength set (see table 2). The sounds in sets 1 and 2 to affect impulsiveness use only equation (8) to affect harmonic tone phases. Set 3 uses both equations (8) and (10) to affect harmonic tone phases and magnitudes, respectively.

Table 3. Sets of sounds to change impulsiveness.

Sound Set	**Number of Sounds**	**First Sound**	**Last Sound**
1	5	AS350 Baseline Main Rotor, Random Phase for all Blade Passage Harmonic Tones	AS350 Baseline Main Rotor
2	6	AS350 Baseline Main Rotor, Random Phase for all Blade Passage Harmonic Tones	AS350 Baseline Main Rotor, Identical Phases (0 radians) for all Blade Passage Harmonic Tones
3	5	AS350 Baseline Main Rotor, Random Phase for all Blade Passage Harmonic Tones	BVI Baseline Main Rotor

Figure 9 shows that impulsiveness changes linearly with sound number for these perturbation methods. There is little effect on the other metrics except for roughness.

3.4 Perturbation Methods to Affect Tonality

The properties of the AS350 baseline tail rotor were adjusted to affect tonality. The Aures/Terhardt tonality calculation [9] increases weighting on tones that are more prominent than nearby tones and increases weighting on tones up to 700 Hz, above which the weighting is reduced. In the AS350 baseline, as a result of magnitude rolloff rates with increasing rotor harmonic number and the tail rotor having a higher BPF (approximately 70 Hz) than the main rotor (approximately 20 Hz), the magnitudes of the first few tail rotor harmonics are prominent relative to nearby main rotor harmonics. Therefore, increasing tail rotor harmonic magnitudes or increasing tail rotor BPF, as long as the BPF does not exceed 700 Hz, will increase the tonality of the resulting sound. Note that increasing the magnitude of main rotor harmonics will not have the same effect on tonality. With a BPF of approximately 20 Hz, the tonality weighting in the Aures/Terhardt calculation is small for the first few harmonics compared to the first few harmonics of the tail rotor. Since the harmonics are also closer together, main rotor harmonics do not become more prominent relative to each other with uniform magnitude increase. The perturbation methods targeting tonality to generate test sounds are:

1. Change AS350 baseline tail rotor blade passage harmonic magnitudes from no tail rotor to 5 dB above baseline. Main rotor is unchanged.
2. Change AS350 baseline tail rotor blade passage harmonic magnitues from no tail rotor to 5 dB above baseline. Tail rotor BPF set to 200 Hz. Main rotor blade passage harmonics set to have random phase.
3. Change AS350 baseline tail rotor BPF from 30.8 Hz (less than half the baseline BPF) to 200 Hz. Main rotor is unchanged.
4. Change AS350 baseline tail rotor BPF from less than half the baseline BPF to 200 Hz. Main and tail rotor blade passage harmonics have random phase.

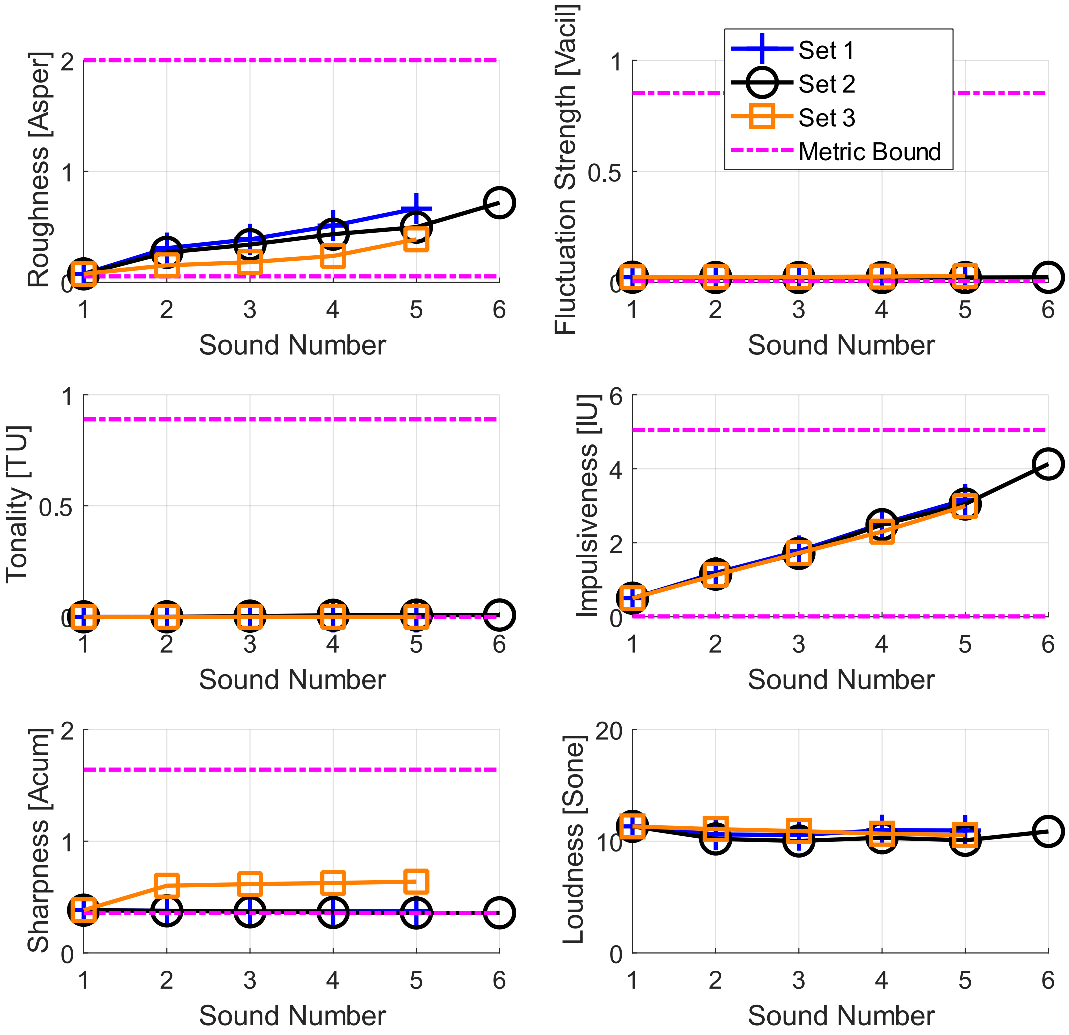

Figure 9. SQ metric 5% exceedance values of perturbation sets targted at changing impulsiveness

Set of five sounds each were generated using each of the above four methods. Descriptions of the sounds are given in table 4. The value of the parameter affected is shown under each sound in a set. Sound 2 of set 3 is the AS350 baseline. Sound 2 of set 4 has the tail rotor BPF of the AS350 baseline, but the main and tail rotor blade passage harmonics have random phase. The random phases of the main rotor blade passage harmonics for sets 2 and 4 are identical to the phase values of the first sound in set 7 that targeted fluctuation strength (see table 2).

Figure 10 shows that these four sets produced a linear change in tonality with sound number. Fluctuation strength, sharpness, and loudness do not have large variations. As a result of the phase relationship among blade passage harmonic tones, impulsiveness and roughness were more difficult to keep constant, and impulsiveness changed more with sets 1 and 3. Those changes are mitigated with the use of random phase in the harmonic tones of the rotor blade passages in sets 2 and 4.

Table 4. Sound perturbation methods to change tonality.

Set	Parameter Affected	Sound 1	Sound 2	Sound 3	Sound 4	Sound 5
1	Tail Rotor Harmonic Magnitudes Rel. to Baseline	No Tail Rotor	–0.9 dB	0.9 dB	3.1 dB	5.0 dB
2*	Tail Rotor Harmonic Magnitudes Rel. to Baseline	No Tail Rotor	–0.5 dB	1.6 dB	3.4 dB	5.0 dB
3	Tail Rotor BPF	30.8 Hz	69.8 Hz (Baseline)	87.2 Hz	108.9 Hz	200.0 Hz
4**	Tail Rotor BPF	30.8 Hz	69.8 Hz	87.2 Hz	108.9 Hz	200.0 Hz
*Main rotor blade passage harmonic tones have random phase. Tail rotor BPF is 200 Hz. ** Main and tail rotor blade passage harmonic tones have random phase.						

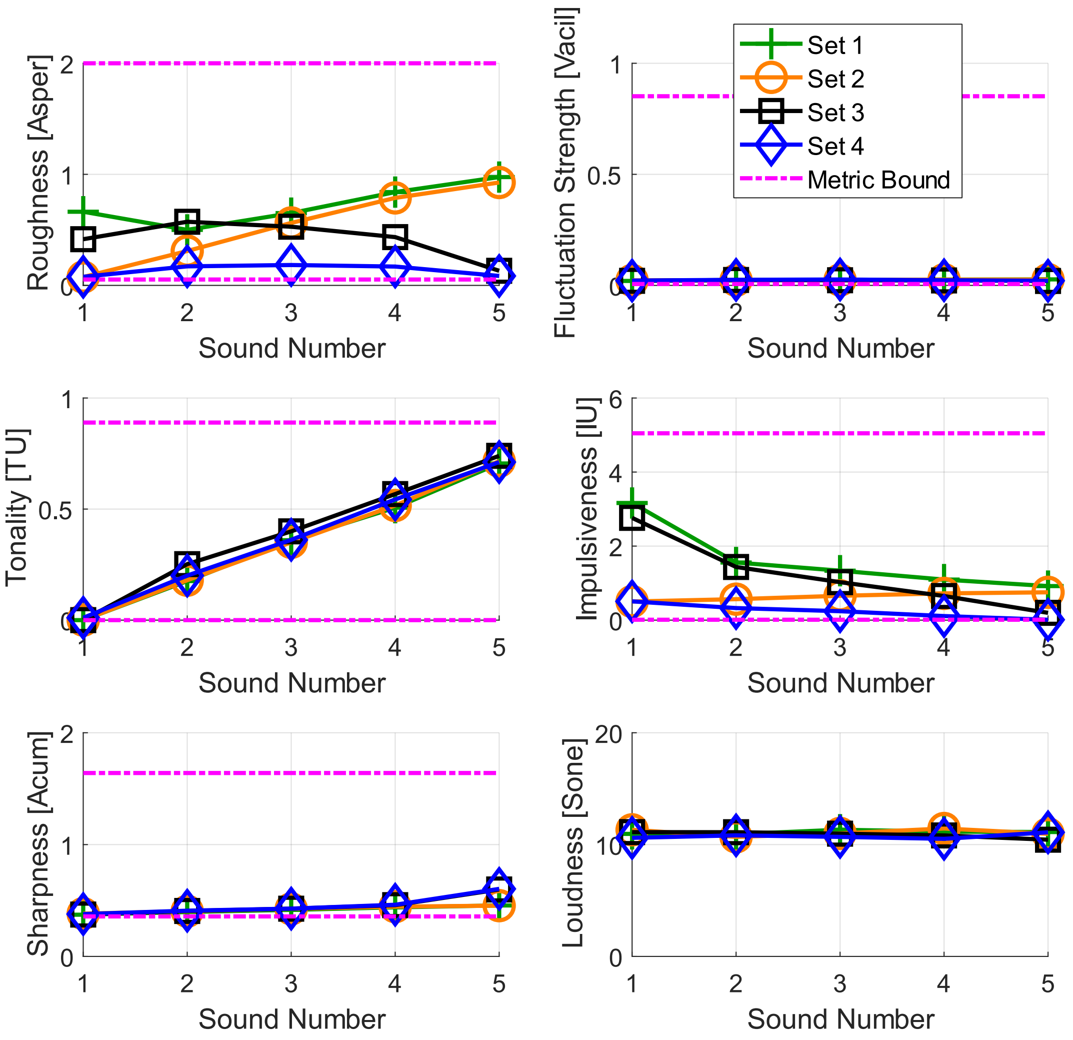

Figure 10. SQ metric 5% exceedance values of perturbation methods targeted at changing tonality.

3.5 Perturbation Methods to Affect Sharpness

Sharpness measures spectral balance and is weighted toward higher frequencies. It can be affected in the test sounds by adjusting the magnitude of high-frequency blade passage harmonics. Three perturbation methods targeted sharpness. The first two are:

1. Set AS350 baseline tail rotor BPF to 1 kHz. Change tail rotor blade passage harmonic magnitudes from no tail rotor to baseline tail rotor blade passage magnitudes. Main rotor is unchanged. The magnitude change is uniform across all tail rotor blade passage harmonics.
2. Change AS350 baseline tail rotor BPF from 69.83 Hz to 1 kHz. Main rotor is unchanged.

Five sounds were generated using each of these methods, as listed in table 5. Sound 1 of set 2 is the AS350 baseline.

Table 5. First two perturbation methods to change sharpness.

Set	Parameter Affected	Sound 1	Sound 2	Sound 3	Sound 4	Sound 5
1*	Tail Rotor Harmonic Magnitudes Rel. to Baseline	No Tail Rotor	-12.9 dB	-8.9 dB	-4.5 dB	0.0 dB
2	Tail Rotor BPF	69.8 Hz (Baseline)	260.6 Hz	475.3 Hz	713.8 Hz	1.0 kHz
*Tail rotor has BPF of 1 kHz.						

The third perturbation method used to affect sharpness was to increase rotor blade passage harmonic magnitude linearly with frequency. The magnitude increase over blade passage harmonic frequencies for the main or tail rotor is given by

$$A'_h = A_h + \frac{\Delta A'_N}{N-1}(h-1) \tag{11}$$

In equation (11), A'_h is the magnitude of harmonic tone h of either the main or tail rotor in dB, A_h is the original baseline magnitude of harmonic h in dB, N is the number of harmonics in either the main or tail rotor, and $\Delta A'_N$ is the dB change in magnitude of the Nth, or last, harmonic of the main or tail rotor. From equation (11), one can see that the magnitude of the BPF at $h = 1$ remains unchanged. The magnitudes of the other harmonics increase linearly in dB.

For the third perturbation method, four sets of five sounds each were produced. They are:

1. Set 3: Increase magnitude of AS350 baseline main and tail rotor blade passage harmonics with magnitude ramp of increasing slope in frequency.
2. Set 4: Only increase magnitude of AS350 baseline main rotor blade passage harmonics with magnitude ramp of increasing slope in frequency. Tail rotor left unchanged.
3. Set 5: Only increase magnitude of AS350 baseline tail rotor blade passage harmonics with magnitude ramp of increasing slope in frequency. Main rotor left unchanged.
4. Set 6: Only increase magnitude of AS350 baseline main rotor harmonics with magnitude ramp of increasing slope in frequency. Main rotor harmonics set to have random phase. No tail rotor.

Table 6 lists the test sounds for each of these four sets. The values in table 6 are the values of $\Delta A'_N$, the decibel magnitude change of the highest harmonic of the main or tail rotor blade passages relative to their baseline values. The first sounds for sets 3–5 are the AS350 baseline. For set 6, the first sound has random main rotor blade passage harmonic phases that are identical to the phase values for the first sound in set 7 that targeted fluctuation strength (see table 2). The values listed in table 6 were selected to produce approximately the same sharpness value for a particular sound number over sets 3-6 given that the adjustment to 10

Sone loudness would be applied. As a result, the nonzero values of $\Delta A'_N$ in table 6 are not the same for each sound number over the four sets.

Table 6. Values of $\Delta A'_N$ for third perturbation method sets of sounds to change sharpness.

Set	Sound 1 (dB)	Sound 2 (dB)	Sound 3 (dB)	Sound 4 (dB)	Sound 5 (dB)
3	0.0 (Baseline)	11.9	16.9	21.2	24.4
4	0.0 (Baseline)	16.1	21.2	25.5	28.5
5	0.0 (Baseline)	12.8	18.0	21.8	25.0
6	0.0	10.3	15.8	20.1	23.7

Figure 11 shows that sounds produced using the three methods effectively change sharpness with little variation in fluctuation strength and loudness. Figure 11 also shows that sharpness values for each sound number in the six sets are approximately equal. Impulsiveness varies among the sets but is relatively constant within a set. Tonality changes considerably over sounds for sets 1 and 5. It jumps significantly at the second sound in set 2. These sets of sounds focus on changing parameters of the tail rotor which contributes more heavily to tonality.

3.6 Dependence Between Roughness and Impulsiveness

For the simulated helicopter sounds in this test, impulsiveness and roughness were strongly dependent on each other. One reason for the strong dependence is that harmonic frequency separations due to the main and tail rotor produce amplitude modulations in the sound. The modulation frequencies are between 20 Hz and 300 Hz, which strongly affects roughness (roughness peaks at 70 Hz). Since the main and tail rotor periodic blade passages are similar to pulse trains at approximately 20 Hz and 70 Hz, respectively, roughness might not be a useful discriminator for sounds in this test. Therefore, for all sounds in this test, the roughness metric was allowed to freely vary with other metric changes.

3.7 Higher Order Effects

Sets of sounds were also created for the test where two or more SQ metric values were changed for each sound number in a set. To generate these sets of sounds, four sets of sounds from the methods previously mentioned were collected. These four sets were:

1. For fluctuation strength: Set 7 from table 2.
2. For impulsiveness: Set 1 from table 3.
3. For tonality: Set 2 from table 4.
4. For sharpness: Set 6 from table 6.

Figure 12 shows the metrics produced by these four sets. Aside from roughness, each set effectively changed only its targeted metric while leaving the other metrics constant. These sets are referred to as first-order effects. Sound 1 for the four sets was the same.

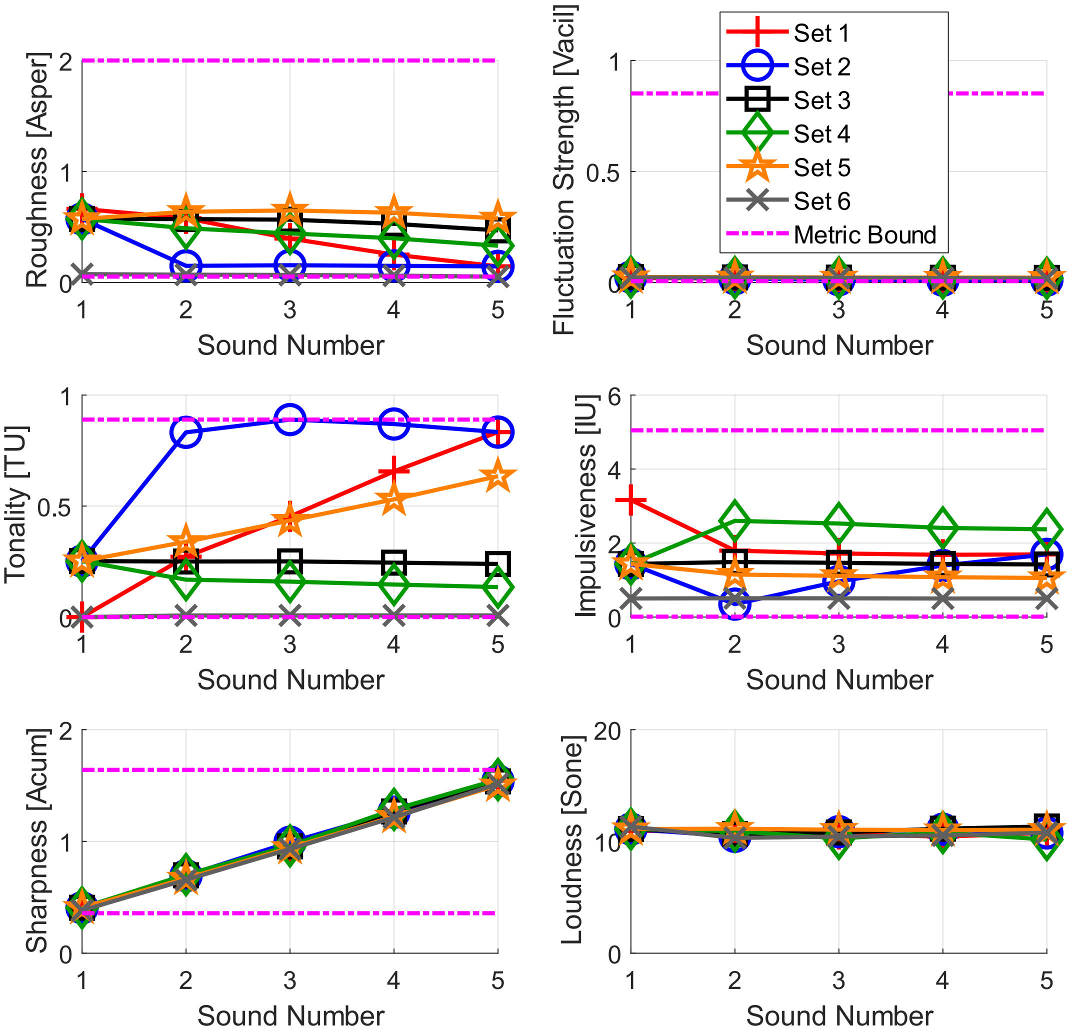

Figure 11. SQ metric 5% exceedance values of perturbation methods targeted at changing sharpness.

The parameter changes associated with these four sets were then applied concurrently to create sounds that changed two or more of the metrics. Combinations of parameter changes are referred to as higher order effects. For example, changing harmonic phase with set 1 for impulsiveness and high-frequency harmonic magnitude with set 6 for sharpness concurrently changed impulsiveness and sharpness together. Generation of these sounds was inspired by the central composite design of experiments. Central composite design of experiments is related to response surface methodolgy in experimental design [20]. Figure 13 shows the effects of all combinations of two of the above four sets or second-order effects.

Figure 14 shows the effects of all combinations of three of the four first-order effects. Each third-order method in figure 14 changes three of the metrics, not including roughness, and leaves one metric roughly constant. Each third-order method affected the metrics to varying degrees. For example, changing fluctuation strength, impulsiveness, and tonality (black trace) had a greater effect on fluctuation strength and tonality than on impulsiveness.

Finally, figure 15 shows the effect of applying all first-order methods at the same time to change all four metrics.

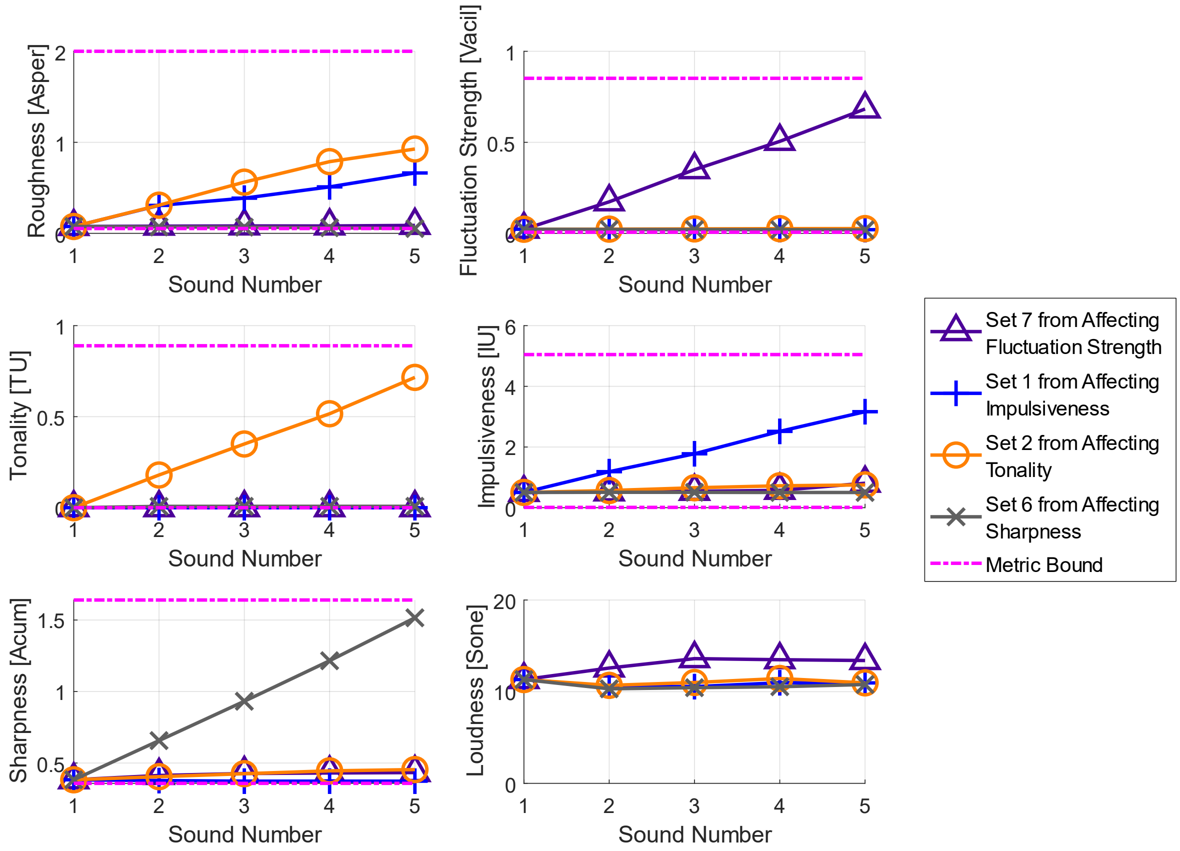

Figure 12. SQ metric 5% exceedance values of first-order effects.

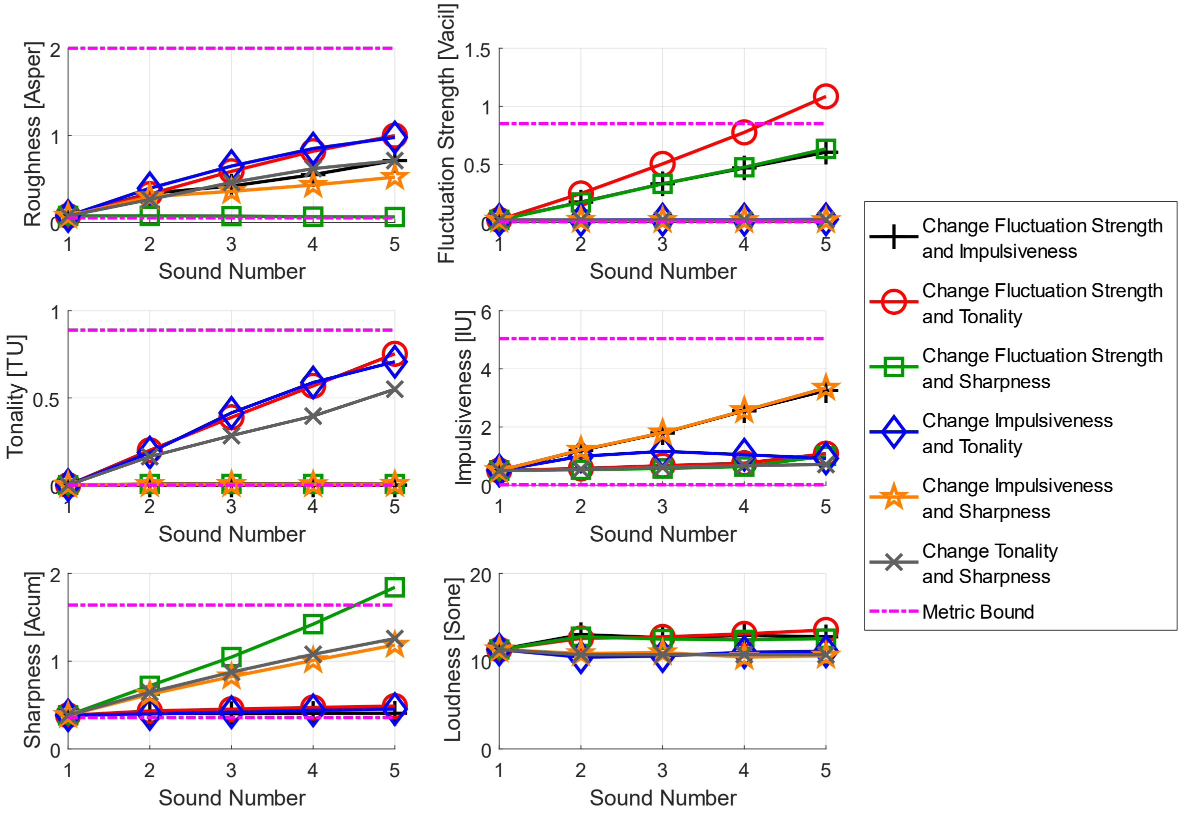

Figure 13. SQ metric 5% exceedance values of second-order effects.

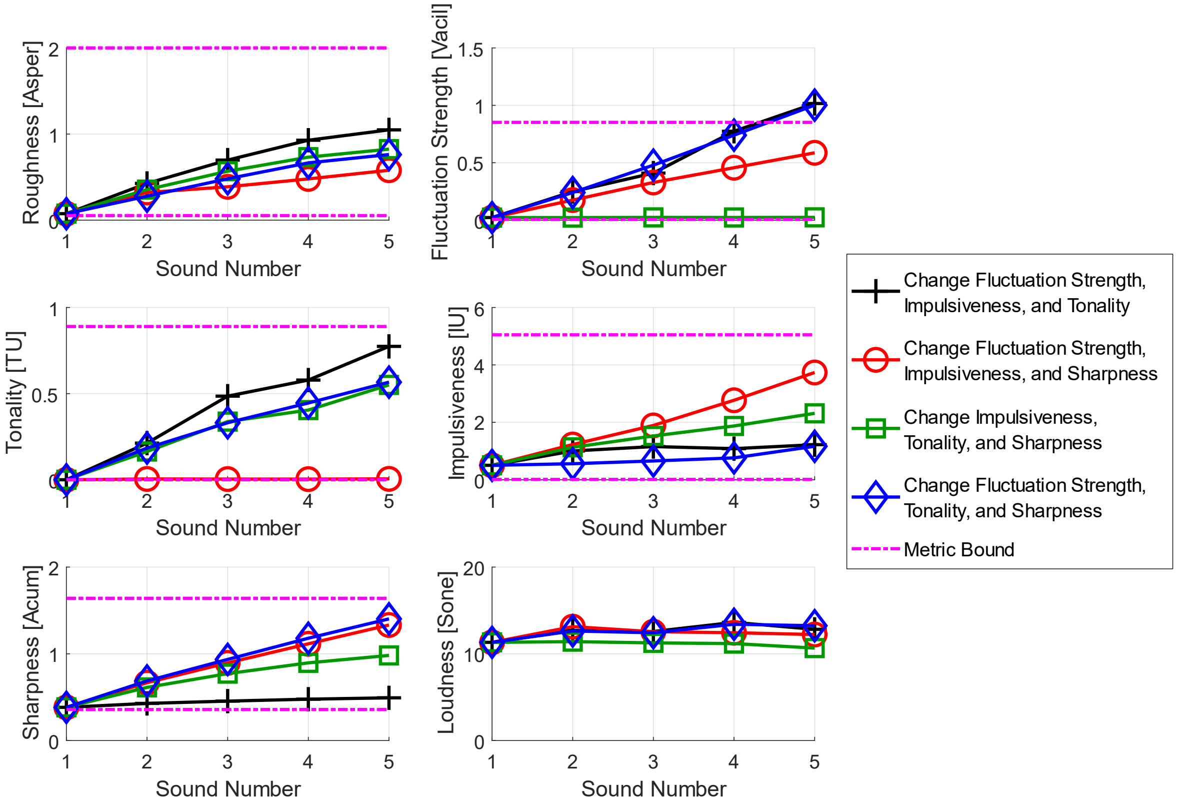

Figure 14. SQ metric 5% exceedance values of third-order effects.

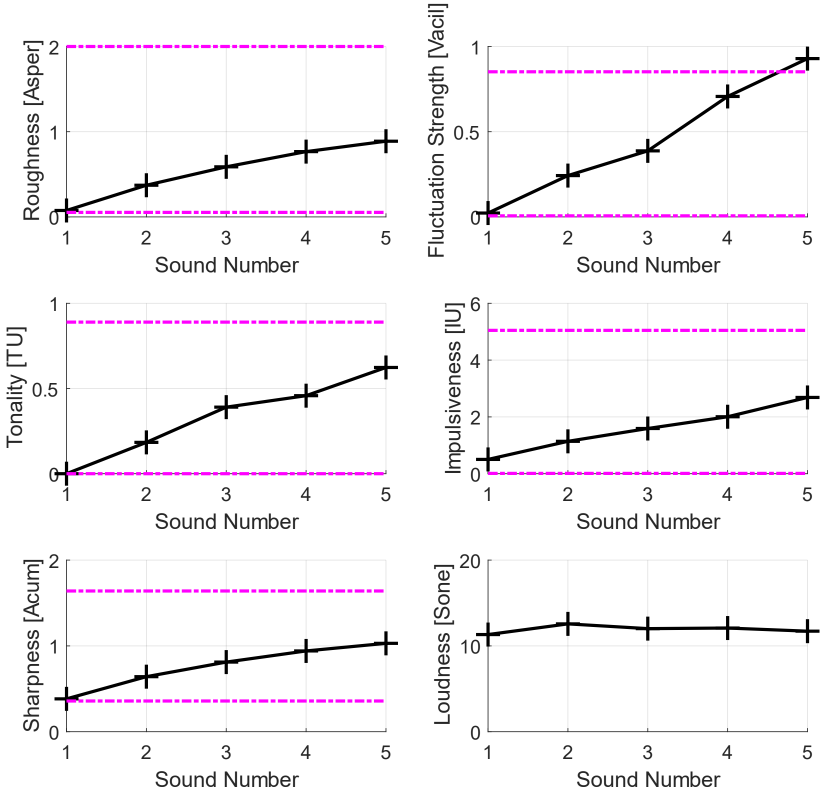

Figure 15. SQ metric 5% exceedance values of fourth-order effects.

Though many sounds for second-, third-, and fourth-order effects were produced, only sound numbers 2 and 3 from each effect and each order were played to subjects in the psychoacoustic test. This decision was made to limit the time needed to conduct the test to limit test subject fatigue. Though all five sounds from each first-order method were played to subjects, only 22 sounds from the second, third, and fourth-order methods were tested for annoyance. The number 22 comes from there being six second-order effects, four third-order effects, one fourth-order effect, and only two sounds being selected to be played to test subjects from each higher order effect (6+4+1 higher order effects x 2 sounds per effect = 22 sounds).

3.8 Final Test Sounds

A total of 105 unique sounds, each six seconds in duration, were generated for the psychoacoustic test using the methods described above. Some of these sounds were repeated between different perturbation methods bringing the total number of sounds generated for the test to 128. The number 128 was chosen as a design goal for the number of test sounds so that they could be divided among four sessions and be played in a different random order for each group of test subjects [1]. Table 7 lists the number of sounds generated to target each metric. The “Description of Repeated Sounds” column lists which sounds were used for each sound generation category more than once or were already created in the sound generation categories given in previous rows. For example, targeting fluctuation strength produced 30 unique sounds including the AS350 baseline. The AS350 baseline was also repeated five more times to target fluctuation strength. The AS350 baseline main rotor with random harmonic phases was also used to target fluctuation strength. It is repeated three more times to generate sounds for impulsivenss. It is not included in the 13 unique sounds targeted for impulsiveness since it was already used to target fluctuation strength.

Table 7. Summary of the sounds generated for the test.

Sound Generation Category	Number of Sounds Generated	Number of Unique Sounds	Description of Repeated Sounds
Targeting Fluctuation Strength	35	30	5 repetitions of AS350 baseline.
Targeting Impulsiveness	16	13	3 repetitions of AS350 baseline main rotor only with random harmonic phase.
Targeting Tonality	20	17	1 repetition of AS350 baseline main rotor only. 1 repetition of AS350 baseline main rotor only with random harmonic phase. 1 repetition of AS350 baseline.
Targeting Sharpness	30	23	1 repetition of AS350 baseline main rotor only. 4 repetitions of AS350 baseline. 1 repetition of AS350 baseline with 1 kHz BPF tail rotor. 1 repetition of AS350 baseline main rotor only with random harmonic phase.
Second, Third, and Fourth-Order Effects	22	22	None.
Other Sounds that were Repeated	5	0	1 repetition of AS350 baseline modulated by 4 Hz sine wave with 0.7071 RMS modulation (Sound 5 of set 2 affecting fluctuation strength). 1 repetition of AS350 baseline main rotor only with all harmonics at identical phase (Sound 6 of set 2 affecting impulsiveness). 2 repetitions of BVI baseline. 1 repetition of AS350 baseline with 200 Hz tail rotor BPF (Sound 5 of set 3 affecting tonality).
All Sounds	128	105	(See rows above).

The five sounds listed in the row “Other Sounds that were Repeated” were generated to test subject consistency. The sounds that produced the largest values of fluctuation strength, impulsiveness, and tonality were repeated. Sounds producing the largest values of sharpness were already repeated during sound generation to target sharpness. The BVI baseline was also repeated an additional two times. Results of this check of subject consistency are discussed in Section 6.2.

Each of the 128 test sounds was given a unique numerical identifier. The list of identifiers, a reference to the table where they are described, and which SQ metrics they target are given in Section 9. Section 12.1 explains how to find summary descriptions of each sound.

4 Measured Sound Quality Metric Values

After the psychoacoustic test was conducted, the sounds generated for the test, referred to as the synthesized sounds, were played in an empty EER and recorded at the four seat locations where test subjects sat [1]. The recorded sounds at the four seats will be referred to as seat recordings. Although test equipment was carefully calibrated, including loudspeaker equalization, delay compensation and level calibration, slight differences in SQ metrics occurred between the synthesized sounds and the seat recordings. Although some seat to seat variation was expected based on data collected during loudspeaker calibration in the EER [21], it was assumed that the SQ metrics of the synthesized sounds would be representative of what was heard at each seat location and any differences in SQ metrics at the seats would not change analysis conclusions.

The extent to which the SQ metrics of the synthesized sounds matched those of the seat recordings is shown in figure 16. In each subfigure, the abscissa is the synthesized metric value and the ordinate is the metric value from the seat recordings. The measurement unit for each metric is given in brackets. The solid lines are *y* = *x* for each metric. If the data points lie on this line, it means that the recorded metric is a perfect match to the synthesized metric.

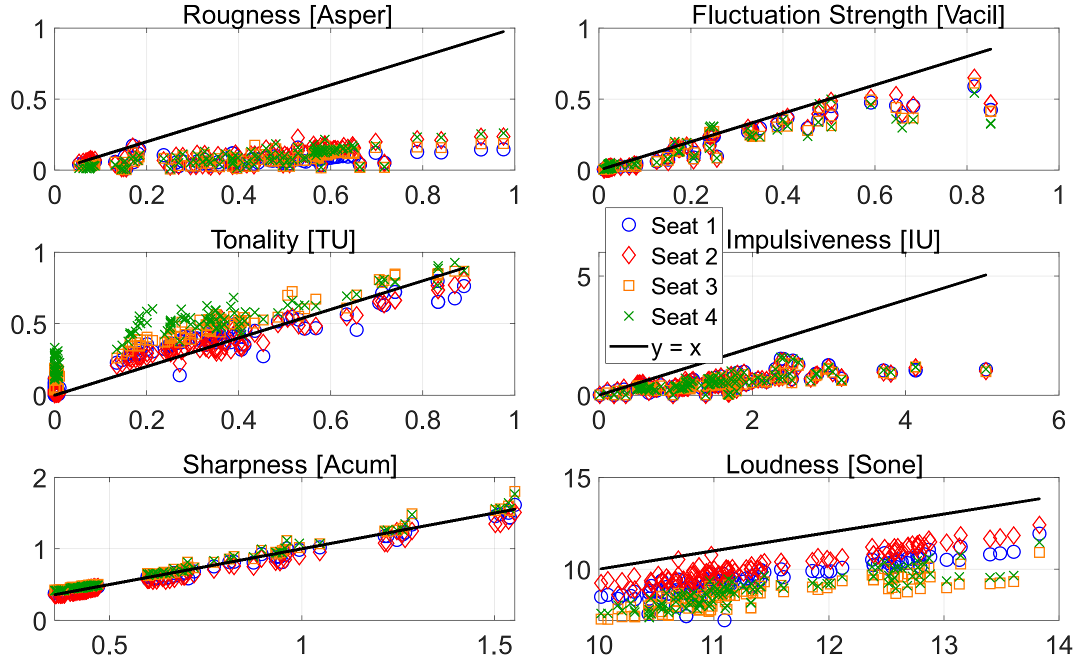

Figure 16. Sound quality metric values, 5% exceedance, seat recordings (ordinate) vs. synthesized sounds (abscissa). Solid line indicates perfect reproduction of intended metric.

The loudness values calculated from the seat recordings are between 1 and 3 Sones lower than the synthesized values. The loudness at seat 3 was generally the lowest, while the highest loudness occurred at seat 2. Although loudness was reproduced at slightly lower levels than intended, the slope of the recorded loudness matches very well with the *y* = *x* line. This indicates that changes in loudness were generally well-reproduced at all four seats, and the recorded loudness level bias is not expected to have a large effect on analyses primarily dealing with linear relationships.

As a check on the recorded loudness levels and to serve as a reference for ambient noise recordings, sound pressure levels for the sounds were calculated using the ANOPP2 Acoustic Analysis utility [22]. In figure 17, the synthesized sound pressure levels, in dBA, are plotted against the recorded sound pressure levels at each seat. The levels roughly correspond with the trend in loudness levels from figure 16 and verify that seat recordings were of lower sound level than the synthesized sounds. As verification on the seat recording sound pressure levels, note that their variation is similar to the 4–5 dB variation found during EER calibration [21].

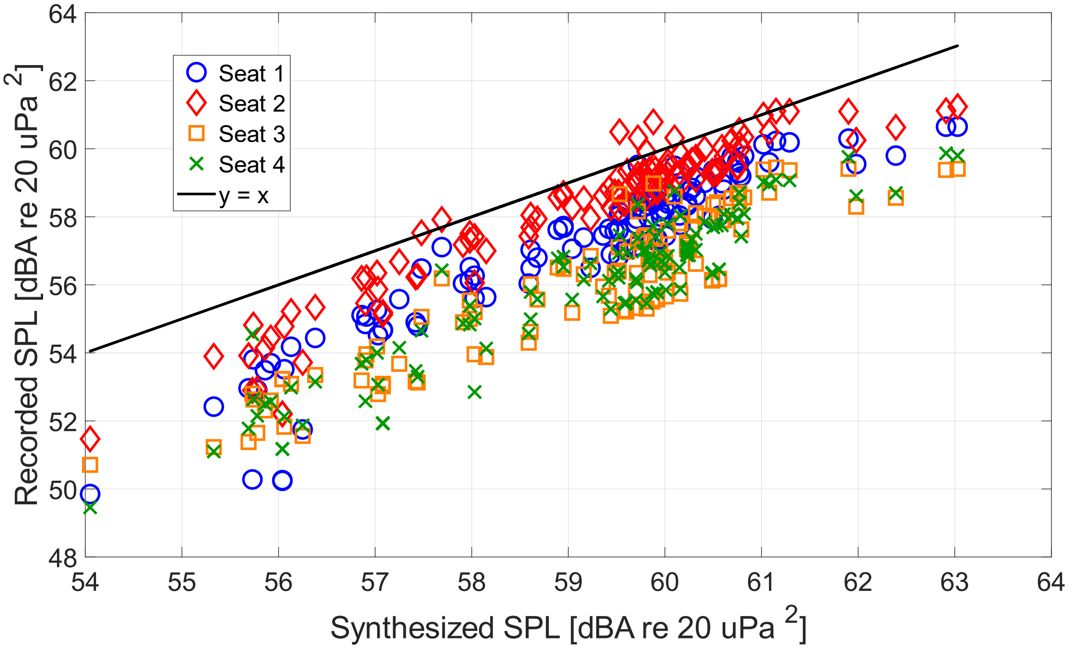

Figure 17. Measured sound pressure level, in dBA, at the 4 seat locations for 105 synthesized sounds, 5% exceedance.

Due to room acoustic effects, the relative phases of tones in the synthesized sounds are generally not conserved in the seat recordings. This modification of tone phase relationships may be the reason for the poor agreement between impulsiveness in the seat recordings and synthesized sounds shown in figure 16. The generally positive slope of the impulsiveness values in figure 16 is evidence that transfer functions between speakers and seats maintained some amount of the phase relationships among tones in the synthesized sounds. The dependence with impulsiveness also caused roughness values to be less for seat recordings than for synthesized sounds.

Fluctuation strength values shown in figure 16 were approximately reproduced at low values but not as well at high values. One hypothesis for this apparent saturation effect is that the seat recordings are not only modified by the transmission path and room effects, but also include an amount of irreducible ambient noise present in the EER (e.g., from HVAC sources). As the intended fluctuation strength increases, the modulations being applied to the baseline signals have troughs that reach farther down in signal level. If the sound becomes modulated down beyond the ambient noise floor, then the recording will not capture the full extent of the modulations, and the observed saturation effect may occur. Different methods of generating fluctuation strength samples may be more or less prone to this kind of corruption—the randomly modulated samples required a larger modulation depth to achieve the same fluctuation strength value as the ones that were sinusoidally modulated.

Tonality and sharpness are the two metrics that matched most closely when comparing the synthesized sounds and seat recordings. This is most likely because the perturbation methods used to adjust these metrics are related to adjusting the spectral balance, which is less sensitive to room acoustic effects. Additionally, the calculation of sharpness and tonality does not include absolute differences in loudness, so the loudness level difference between synthesized sounds and seat recordings did not affect these two metrics.

Although recorded tonality values generally follow the y = x line, there is a large spread below 0.5 TU. Even when no tonality is present in a synthesized sound, tonality may still be recorded at the seats due to the presence of ambient noise. Figure 18 shows the ambient SQ values recorded at each of the seats. The ambient recorded sound pressure levels, shown in the bottom plot, are much lower than the sound pressure levels of the seat recordings (see figure 17). Since SQ values don't necessarily scale with sound pressure level, the ambient SQ values can be as large as the SQ values of the seat recordings, particularly for tonality and sharpness. For example, the addition of ambient noise to some synthesized sounds without tonality caused the corresponding seat recordings to have tonality values between 0.2 and 0.35 TU. The effect of ambient noise on the SQ value at the seats is nonlinear with sound pressure level, hence seat recordings of some synthesized sounds without tonality had low tonality values (below 0.2 TU). The nonlinear effect is also why many seat recordings have sharpness values below the minimum ambient sharpness value of 1.35 Acum.

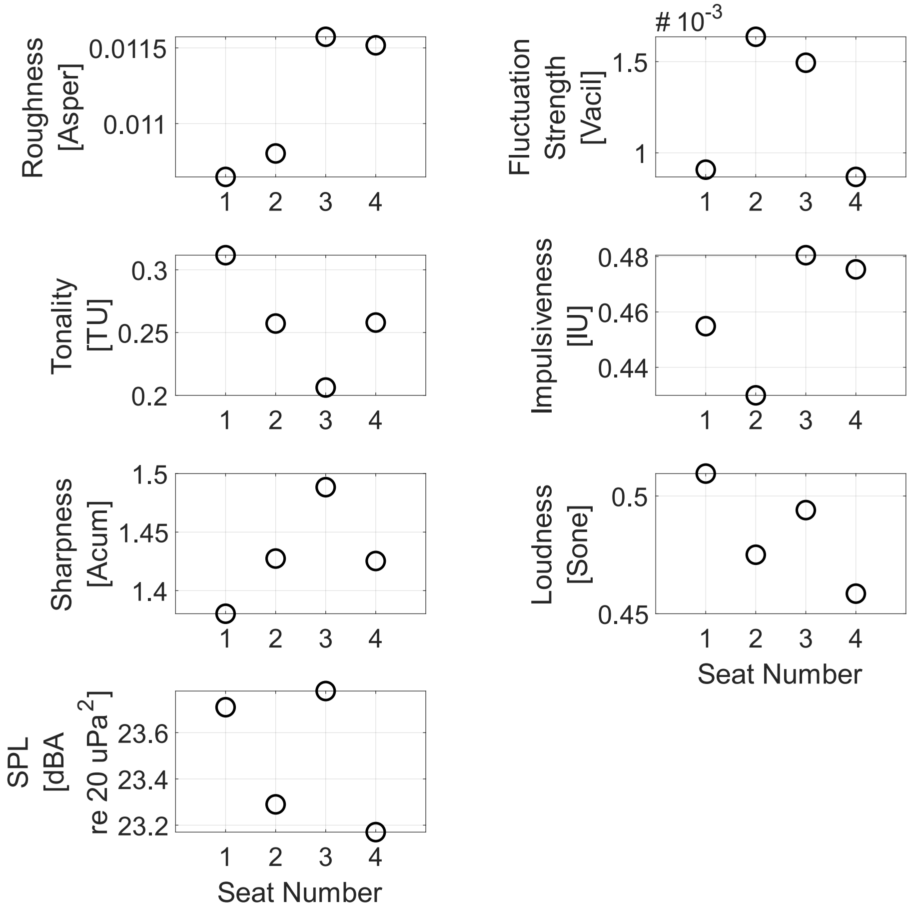

Figure 18. Recorded ambient sound quality metrics and sound pressure level, 5% exceedance.

The stimuli were intended to cover the range of SQ metrics in table 1, however, the data from figure 16 show that the stimuli heard by the test subjects did not fully cover the desired test space with respect to roughness and impulsiveness. If the just noticeable difference [23], or JND, of the impulsiveness values were known, the test stimuli might still provide useful insight into the relationship between impulsiveness and annoyance. Because the JND was unknown, data from this test were not useful for establishing such a relationship. These considerations cast doubt on an earlier claim made by the authors [1] that impulsiveness did not have a noticeable effect on the annoyance response. Increasing the range of impulsiveness values and understanding the JND of impulsiveness are needed to strengthen conclusions between this metric and human response to helicopter sounds.

Conversely, test stimuli did cover the desired test space for fluctuation strength, tonality and sharpness in the seat recordings. Hence, the ranking of SQ metrics for describing annoyance that was determined in earlier analyses [1, 2] is still valid. That ranking from most to least important in describing the annoyance response was sharpness, tonality, and fluctuation strength.

Section 12.3 explains how to obtain the seat recordings online [3].

4.1 Annoyance Variation in Seat Recordings Explained by Regression

The ultimate purpose of the psychoacoustic test described here was to explore the relationship between SQ metrics and annoyance. In this section, we repeat the analysis from the original reference relating annoyance and metric values [1], except here, the SQ metrics of the seat recordings are used instead of the metrics of the synthesized sounds. The specifics of the test were described in the earlier paper [1] and are repeated here briefly. Multiple linear regressions were executed relating subject annoyance response with combinations of SQ metrics of seat recordings. The annoyance response was quantified using rating labels based on recommendations by Fields [24]. These labels were the choices presented to subjects to rate their annoyance to a sound. Confidence intervals for mean annoyance responses to all test sounds were determined using bias-corrected and accelerated percentile bootstrap estimations [25] with simulations of the mean annoyance involving 100,000 samples. Comparing the regressions from the synthesized sounds and seat recordings, values of r^2 were lower with the seat recordings than with the synthesized sounds. Centering the SQ metrics about their respective mean values before doing the linear regression did not change the r^2 values. This indicates that using what was actually heard by the subjects had less predictive accuracy for annoyance than using what the subjects were intended to hear. This section comments on possible reasons for the lower r^2 values.

Section 10 contains tables of r^2 values and regression equations for mutliple linear regression with annoyance and combinations of SQ metrics of seat recordings for each of the four seats. The r^2 tables also give the change in r^2 when removing an SQ metric from a combination, which is an indication of the metric's importance in explaining the annoyance response. A cursory analysis of the results in section 10 suggests the same ranking of importance of SQ metrics as was found in Krishnamurthy et al. and Boucher et al.: sharpness, tonality, and fluctuation strength. Section 12.4 explains how to obtain subject responses to test sounds online [3].

Reasons for the r^2 value differences are potentially:

1. Effects of ambient noise in the seat recordings.
2. A nonlinear relationship between mean annoyance responses and SQ metrics of sound stimuli. Figure 19 shows a nonlinear relationship between subject annoyance rating and tonality values of synthesized sounds and seat recordings (see section 3.4).
3. As seen in figure 19, the confidence intervals for the seat recordings are larger than for the synthesized sounds due to the smaller sample size at each test seat. Figure 20, which shows the annoyance response for sounds targeting sharpness (see section 3.5), shows similar confidence interval sizes for responses to the synthesized sounds and seat recordings as seen in figure 19. The reduced sample size may also explain some of the irregular variation seen in figures 19 and 20.

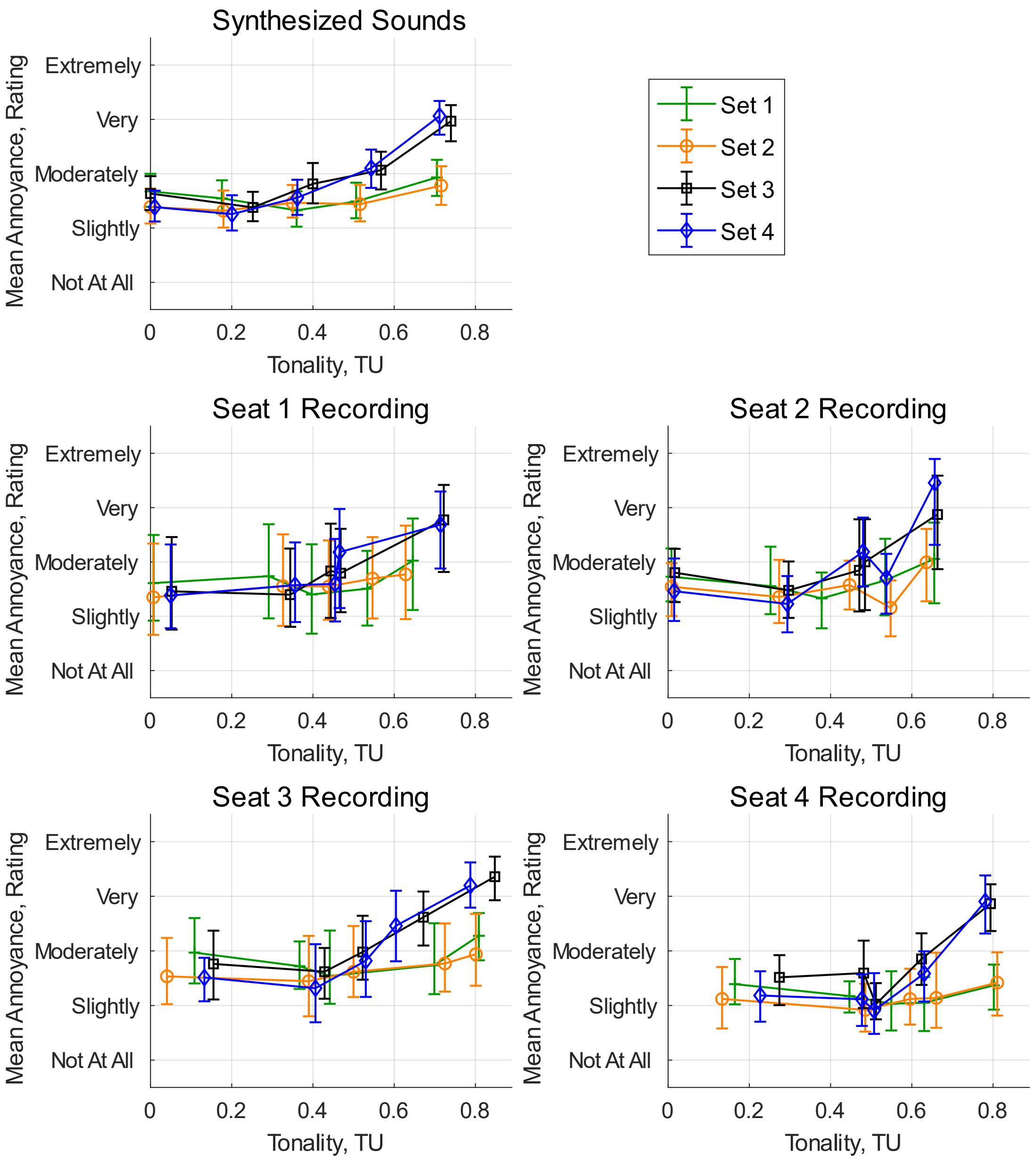

Figure 19. Annoyance response to sounds targeting tonality.

With respect to the noise floor, subjects have an ability to focus on the sounds and to resolve them as auditory objects that are distinct from the ambient noise. This ability is referred to as "auditory streaming." Many auditory cues can be used in this streaming process, such as timbre (or SQ), location, and temporal patterns. This process is an automatic and universal cognitive effect that subjects would quickly learn during the practice (see section 6.4) and familiarization sessions (i.e., what is the target sound and what is the room ambient), similar to the way that people quickly learn to focus on a single conversation in a crowded room (the cocktail party effect) [26]. A microphone is unable to perform this kind of processing. So, paradoxically, even though the seat recordings are a more accurate representation of the acoustic field that was presented to the subjects, the synthesized sounds might be a more accurate representation of what the subjects focused on, and hence, what drove their responses. This could explain the marginally higher r^2 for model derived from the synthesized sounds.

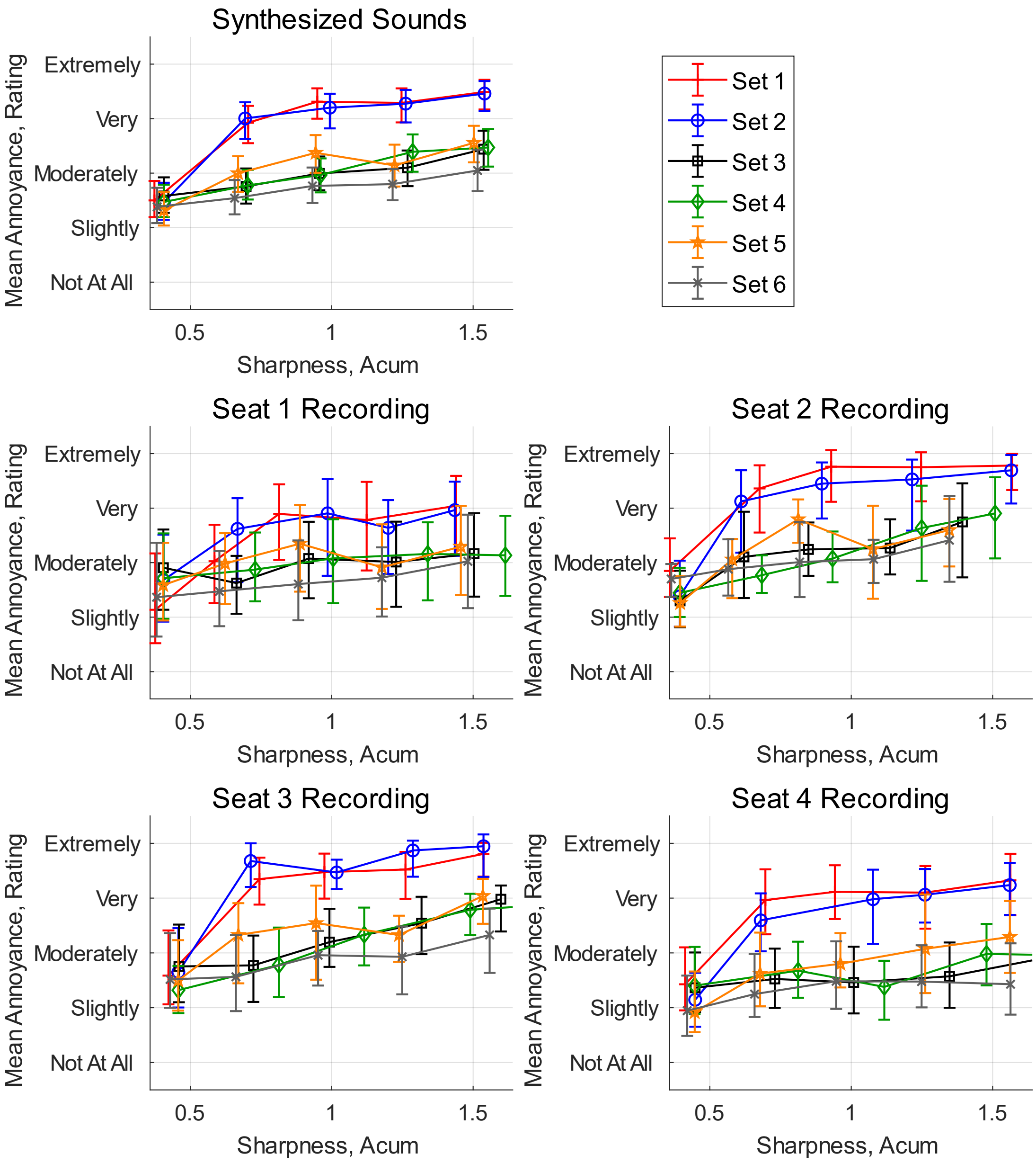

Figure 20. Annoyance response to sounds targeting sharpness.

Although auditory streaming may explain the higher r^2 values for the synthesized sounds, the impulsiveness data demonstrates the value in using recorded values when the sound reproduction and not room ambient noise leads to a corruption of the intended SQ. Reproduction errors cannot be filtered by auditory streaming but the effects of ambient noise can be filtered.

5 Certification Metrics

In addition to SQ metrics, the relationship between annoyance to the sounds in this test and aircraft noise certification metrics was explored. Since the sounds were of short duration and constant, those certification metrics involving summation and integration over time, such as sound exposure level or effective perceived noise level, were not analyzed. Instead, the constituent metrics of these certification metrics were analyzed. These constituent metrics are A-weighted sound pressure level (ASPL), Z-weighted (or unweighted) sound pressure level (ZSPL), perceived noise level (PNL), and perceived noise level, tone-corrected (PNLT). These metrics were computed using the ANOPP2 Acoustic Analysis utility [22]. The results of a linear regression between the 5% exceedance levels of each metric and annoyance responses, with 95% confidence intervals, to all synthesized test sounds are given in figure 21. The x-axes metric values are normalized to the maximum (normalized value of 1) and minimum (normalized value of 0) of each metric computed over all synthesized sounds. Mean annoyance responses are marked by ‘x’s. Confidence interval bars around the mean annoyance responses were found through bias-corrected and accelerated percentile bootstrap estimation [25] with simulations of the mean annoyance to each sound involving 100,000 samples. Figure 21 gives the linear regression trends as solid black traces across every plot. The r^2 value of each trend is given in the plot title next to the metric abbreviations.

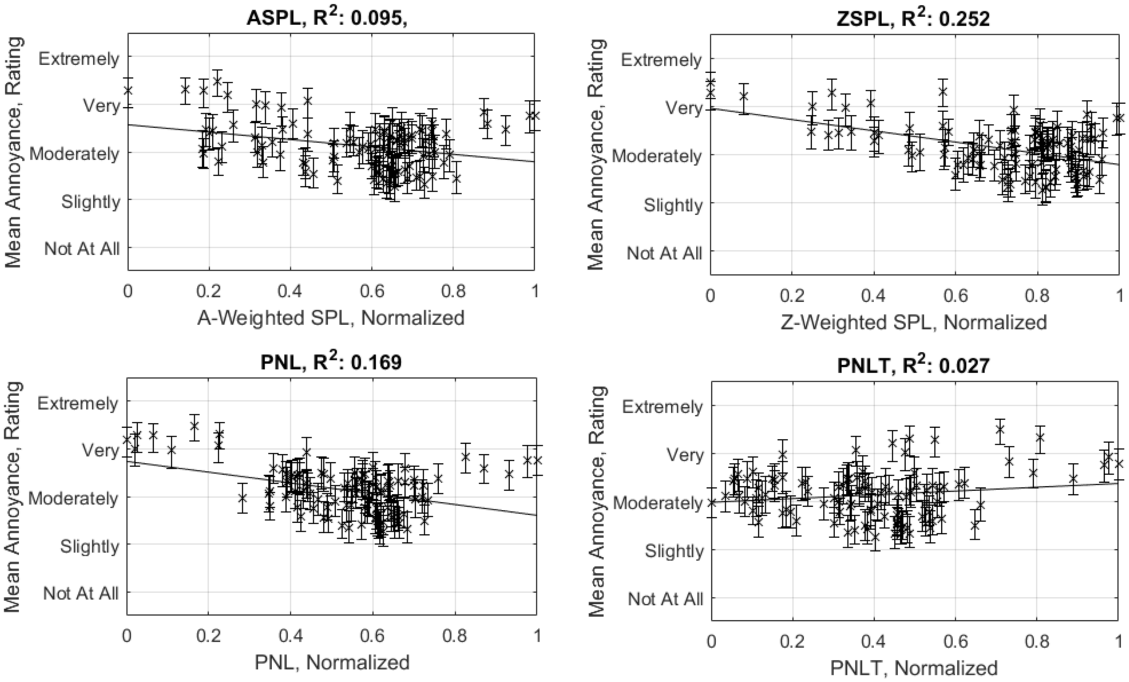

Figure 21. Linear regression between constituent metrics and annoyance.

The following are observations from figure 21:

1. There is relatively large scatter in the data, especially where metric values are clustered. An example is the scatter in mean annoyance to ASPL values between 0.6 and 0.8.
2. There is negative slope for the linear regression for ASPL, ZSPL, and PNL with the annoyance response.
3. PNLT has a small positive slope computed for the linear regression with annoyance response, but as the relatively small r^2 value of 0.027 suggests, the predictive utility of PNLT is small.

These observations would seem to contradict the expectation that annoyance increases as constituent metric level increases. This is likely a result of keeping sounds at constant loudness in this test and manipulating elements other than loudness. This result points to the possibility that factors contributing to annoyance of helicopter sounds not captured by the constituent metrics are being captured by the SQ metrics.

Comparing changes in the constituent metrics with changes in SQ metrics over different perturbation methods can provide insight into negative slopes for the linear regressions with ASPL, ZSPL, and PNL in figure 21. The top row of plots in figures 22 and 23 repeat the data from figures 11 and 12, respectively, that describe the sound perturbation methods. The top left plots in figures 22 and 23 show the tonality and sharpness SQ metric values, respectively, for stimuli sets that target these metrics. Loudness values of the stimuli sets are given in the top right of these figures. For every set of stimuli, tonality and sharpness increase with sound number. As tonality and sharpness change over the desired range of metric values, loudness is nominally constant.

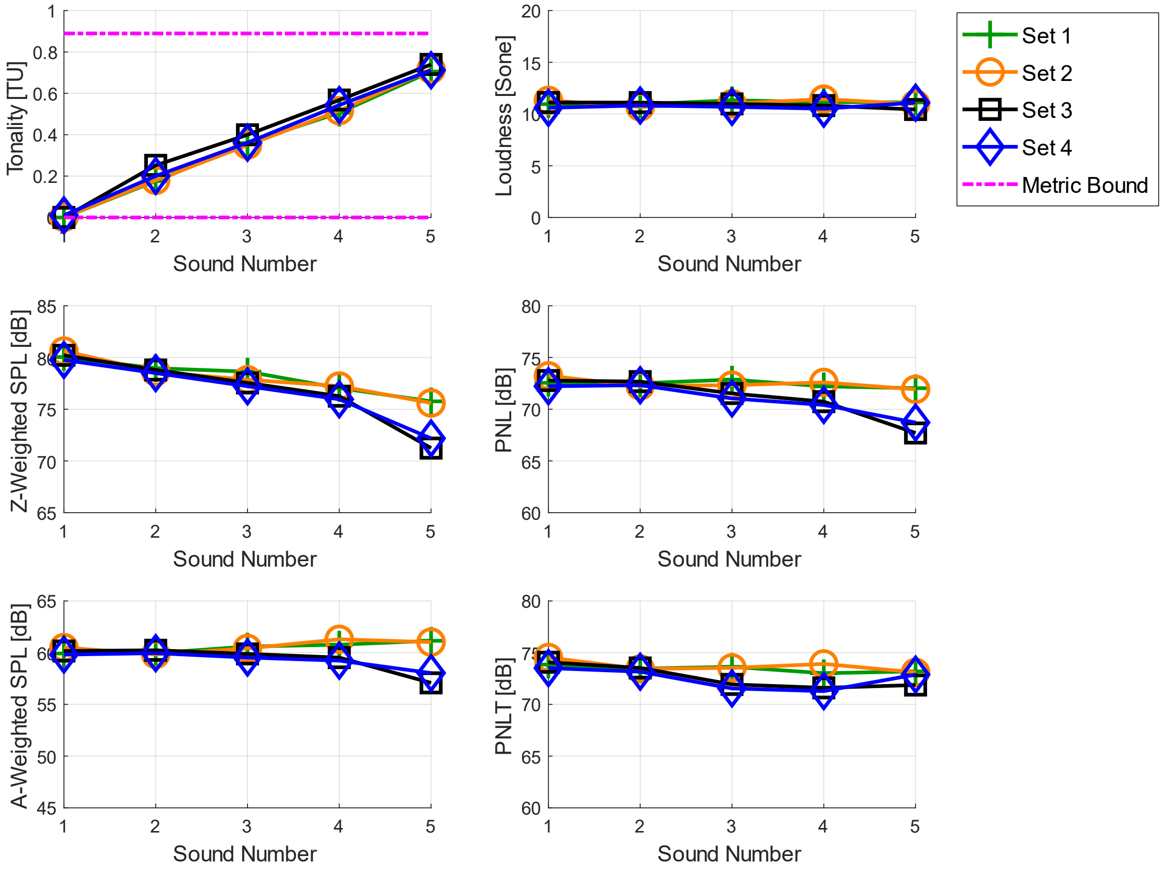

Figure 22. Metric 5% exceedance values of perturbation methods targeting tonality.

The ZSPL, PNL, and ASPL values for the same stimuli are given in the middle left, middle right, and bottom left plots of figures 22 and 23, respectively. ZSPL, PNL, and ASPL either reduce overall or remain roughly constant over five sounds. While there are increases in these metrics between sounds for some of the stimuli sets, such as PNL increasing after sound 3 for set 2 that targets sharpness, the increases are not large enough to cause an overall noticeable increase in the constituent metric values over five sounds.

The negative linear regression slopes for ZSPL, PNL, and ASPL with annoyance are a reflection of the inverse dependence these metrics have with tonality and sharpness. As evidenced by the large scatter in annoyance response in figure 21 to the constituent metrics and with loudness nominally constant over the stimuli in figures 22 and 23, test subjects were likely not responding to the intensity aspects of ZSPL, PNL, and ASPL. Instead, since increased annoyance was found to be associated with larger tonality and sharpness values [1], they were responding to aspects that produced a net reduction in ZSPL, PNL, and ASPL over the same set of stimuli for which tonality and sharpness increased. This response produced the linear regression trends in figure 21 of annoyance reducing with increasing values of ZSPL, PNL, and ASPL.

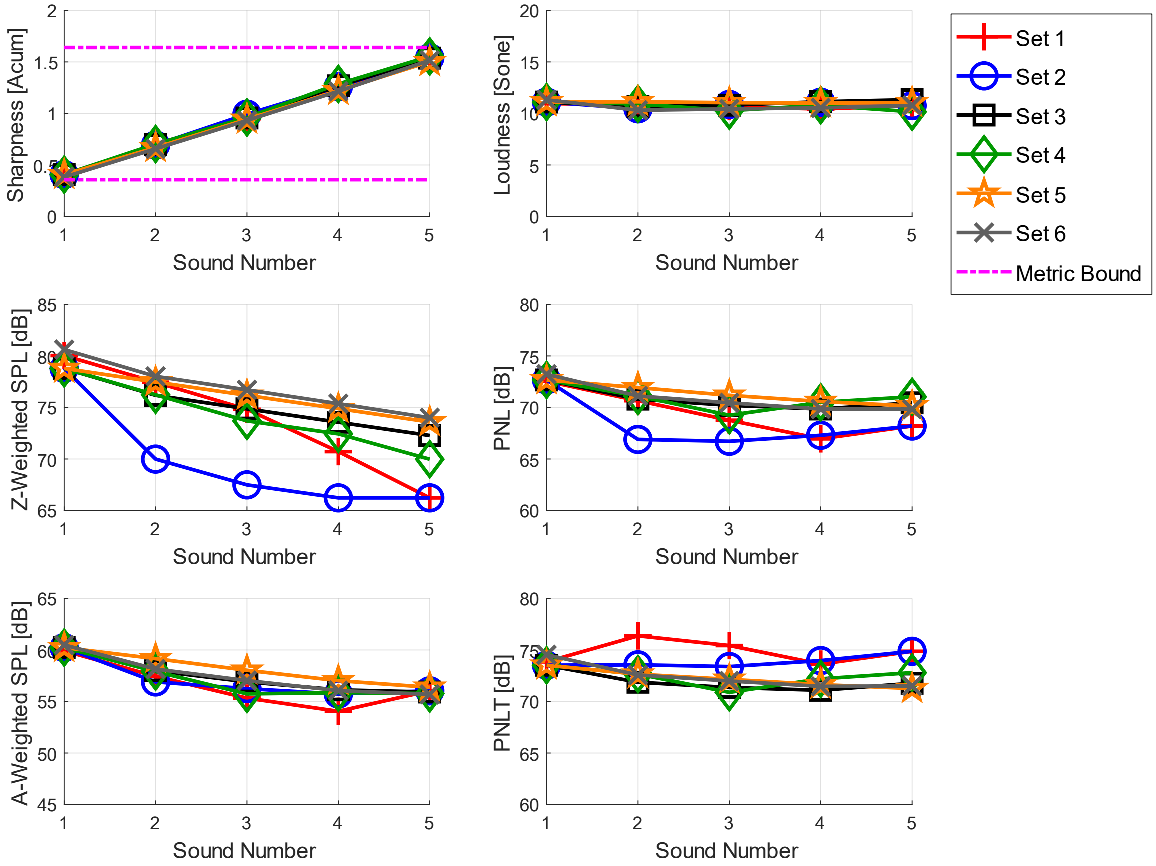

Figure 23. Metric 5% exceedance values of perturbation methods targeting sharpness.

The bottom right plots of figures 22 and 23 show that PNLT remains roughly constant over five sounds for all sets targeting tonality and sharpness. There are perturbations in the PNLT values for some sets, such as for sets 3 and 4 targeting tonality, which reduce in PNLT before increasing. Overall, figures 22 and 23 show that the PNLT metric appears to be marginally less affected by changes in the SQ metrics. Although PNLT incorporates a tone penalty, keeping the loudness constant appears to keep this metric also nominally constant.

Results of linear regression between annoyance and constituent metric values of seat recordings are similar to the results in figure 21 for the synthesized sounds. For each of the four test subject seats, table 8 shows the r^2 between the four constituent metrics calculated with seat recordings and mean annoyance. Metric r^2 values are similar to the respective r^2 results in figure 21.

Table 8. Linear regression r^2 values between constituent metric values of seat recordings and mean annoyance response.

Certification Metric	Seat 1	Seat 2	Seat 3	Seat 4
ASPL	0.15	0.19	0.08	0.10
ZSPL	0.28	0.28	0.24	0.23
PNL	0.20	0.27	0.08	0.12
PNLT	0.00	0.03	0.02	0.00

6 Adjunct Analyses Results

6.1 Sound Response Normal Distribution Test

To aid further analyses, it is of interest to determine if the subject responses for each individual sound in the test are normally distributed. In the multilevel analysis in Boucher et al. [2], subject responses to indvidual sounds were assumed to be normally distributed. A check on this assumption is presented in this paper.

This paper uses a technique based on the 1954 method by H. Chernoff and E.L. Lehmann [27] to determine if responses over all subjects to individual test sounds are normally distributed. In this method, a histogram of annoyance responses is compared to a histogram generated under an assumption of normality. To employ this method on test data, the annoyance response scale was divided into 11 intervals between 1 and 11 with “Not at All” annoying having a value of 2 and “Extremely” annoying having a value of 10. Each interval represented a bin of the histogram. Responses to each sound were then assumed to be normally distributed with mean and variance equal to the sample mean and sample variance of the subject responses for the sound. The response histogram and normal histogram were then compared and a p-value determined to reflect the strength of the comparison. The null hypothesis that annoyance responses for an individual sound are normally distributed is rejected if the p-value is less than a signficance level of 0.05, or 5%. See DeGroot and Schervish [27] for details on calculating the p-value.

Figure 24 shows p-values for each of the 128 stimuli used in the test. There are 24 sounds for which the null hypothesis can be rejected at a significance level of 0.05. Specific sounds for which the null hypothesis were rejected are indicated in table 10 in section 9 of this document.

In some cases, the lack of normality is due to a large number of subjects rating the sound at one end of the annoyance scale. Specifically, sounds 65, 70, and 73-80 were all ranked closer to being extremely annoying. This skewed the responses away from being normally distributed. Reasons the null hypotheses was rejected for the remaining 14 sounds are not as easily determined. The rejection of the normality assumption in 20% of the test stimuli suggests a reexamination of the impact of the normality assumption on the results described in Boucher et al. [2] may be warranted.

Section 12.4 explains how to obtain subject responses to test stimuli online [3].

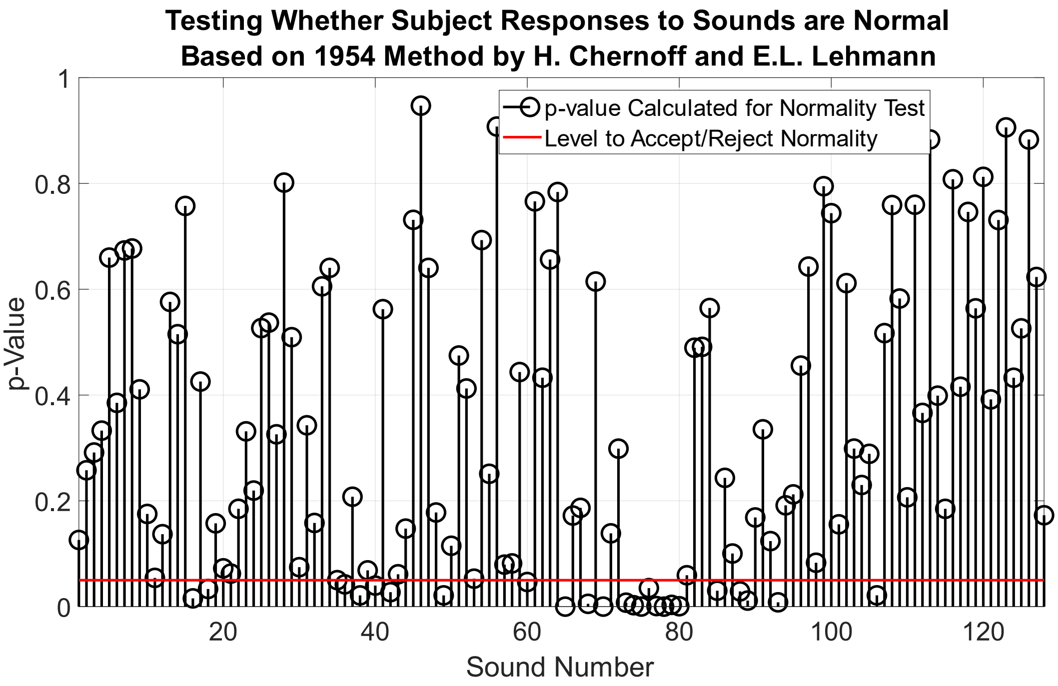

Figure 24. Test to determine normally distributed subject responses.

6.2 Response to Repeated Sounds

This section examines the consistency of responses averaged across all subjects to repeated test sounds. If the average annoyance rating is consistent for sound repetitions, it will provide more confidence in the responses to sounds that were only played once. As described in section 3.8, of the 105 simulated helicopter sounds generated for this test, eight of the sounds were repeated at least once to produce a total of 128 test sounds. Table 10 of section 9 gives the reference for each sound number. Sound 1, the AS350 baseline, was repeated the most during the test. A different order of sounds was played to each group of subjects and test subjects did not hear most sound repetitions consecutively.[2] For a list of the test sound sequence for each subject group, see section 11.

Figures 25–28 show the subject responses to each of the eight repeated sounds The blue dots are the individual subject responses to each sound repetition. The mean annoyance responses with 95% confidence intervals are given by the solid red traces. Confidence intervals for mean annoyance responses to all test sounds were determined using bias-corrected and accelerated percentile bootstrap estimations [25] with simulations of the mean annoyance involving 100,000 samples. Although a bias correction is involved, the bootstrap estimations assume a normally distributed response or that the responses can be transformed into normally distributed random variables. A green square marker is placed on the mean annoyance rating if the responses to the sounds were found to be normally distributed using the criteria discussed in section 6.1 and indicated in figure 24. The legend in figure 25 is applicable to figures 26–28.

---

[2] Table 10 in section 9 and table 19 in section 11 can be used to find the sound number sequences in each subject group where the same sound was repeated consecutively. These sound sequences are not listed in this paper.

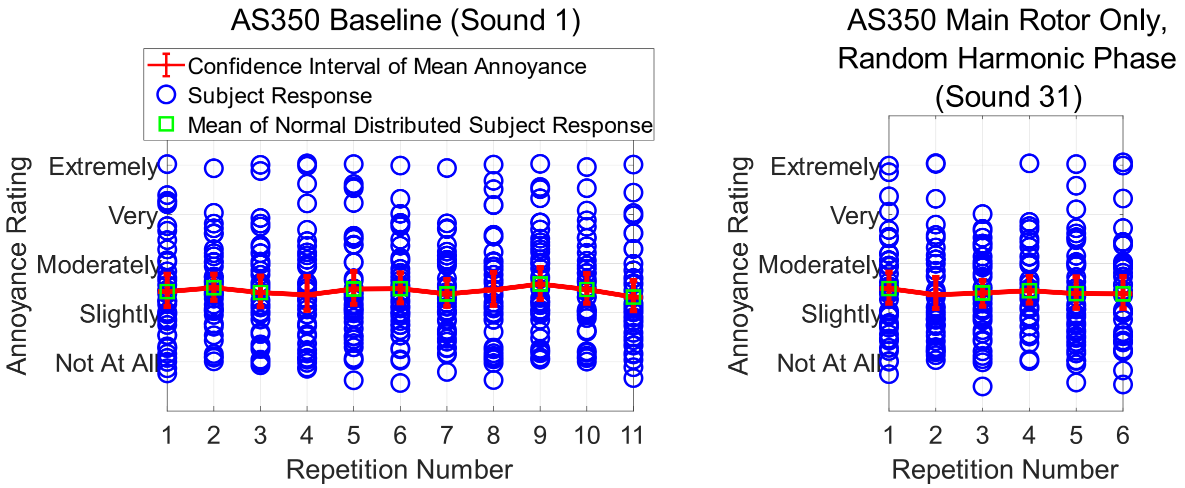

Figure 25. Annoyance ratings to repetitions of sound 1, AS350 baseline, and sound 31, AS350 main rotor only, random harmonic phase.

The bootstrap procedure used to compute the annoyance confidence intervals assumes a normally distributed response or that the responses can be transformed into normally distributed random variables. It is not known if responses not found to be normally distributed in section 6.1 can still be transformed into normally distributed random variables. As a result, the confidence intervals for sound responses without a green square at the mean response value should be interpreted with caution.

Determining how significantly different the mean response of one sound is from another can be accomplished by looking at whether confidence intervals overlap. On the left side of figure 25, mean responses to repetitions of sound 1 are not significantly different from each other. Overlapping confidence intervals indicate that, on average, subjects were consistent in their response to these repeated sounds, with the overlap possibly being more valid for confidence intervals of responses determined to be normally distributed.

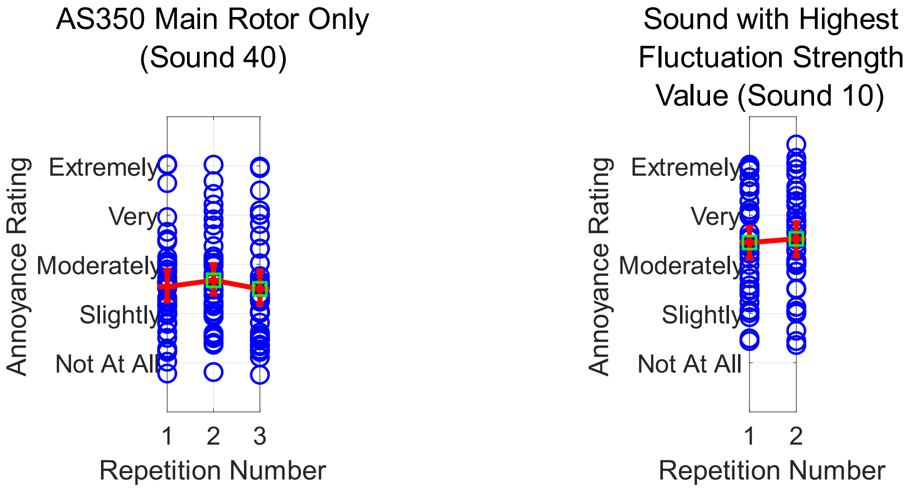

Figure 26. Annoyance ratings to repetitions of sound 40, AS350 main rotor only and sound 10, sound with highest fluctuation strength value.

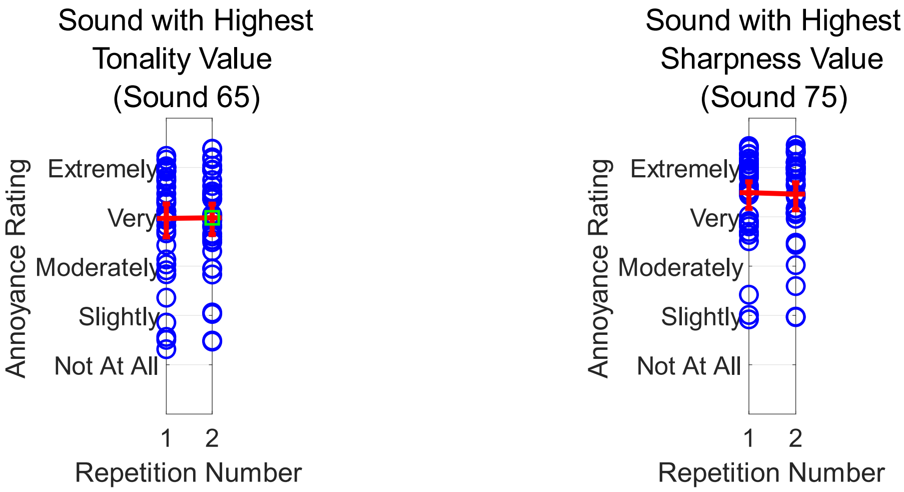

Figure 27. Annoyance ratings to repetitions of sound 65, sound with highest tonality value, and sound 75, sound with highest sharpness value.

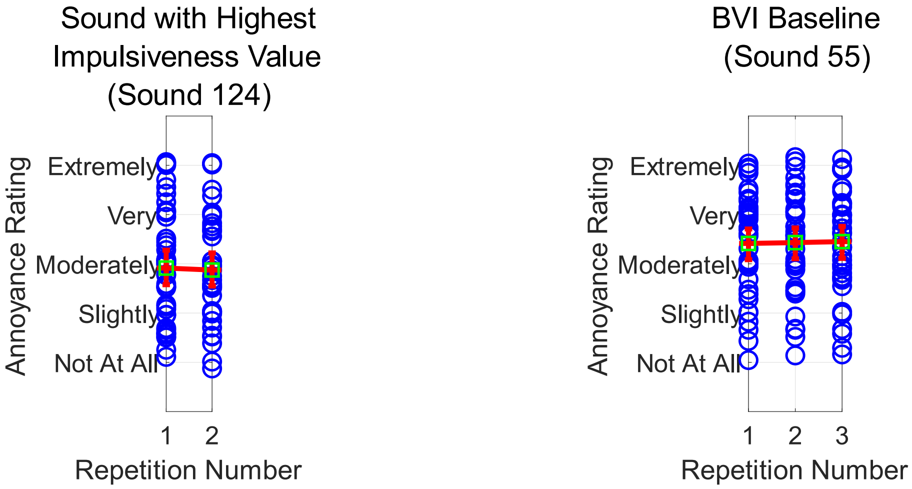

Figure 28. Annoyance ratings to repetitions of sound 124, sound with highest impulsiveness value, and sound 55, BVI baseline.

Confidence intervals also overlap around mean responses to repetitions of sounds 31 and 40, on the right and left in figures 25 and 26, respectively. They are sounds of the AS350 main rotor only with sound 31 modified to have a random phase in the blade passage harmonics.

Sounds in figures 26–28 produced extreme values of the SQ metrics for synthesized sounds. Sound 10, on the right side of figure 26, produced the highest fluctuation strength value in the test. Sound 65, on the left side of figure 27, produced the highest tonality. Sound 75, on the right side of figure 27, produced the highest sharpness. Sound 124, on the left side of figure 28, produced the highest impulsiveness. Overlapping confidence intervals within repetitions of these sounds indicate that subjects were consistent in their response to them. As a note, green square markers were deliberately omitted from responses to sound 75 in figure 27. Both repetitions of sound 75 were not found to be normally distributed in section 6.1.

Figure 28 indicates that subjects were consistent in their response to repetitions of sound 55, which was the BVI baseline.

The plots in figures 25–28 reveal that although subjects responded over almost the full annoyance rating range for each sound, they were consistent in their overall mean annoyance response for individual sounds. Focusing on sounds for which responses passed the normality test, the mean annoyance confidence intervals for repetitions of each sound overlap considerably. Overlaps among repetitions for individual sounds occur despite sounds being played in a different order to each group of subjects during the main test.

An alternative quantitative analysis approach is finding p-values to the null hypothesis that a response distribution to one sound is the same as the response distribution to another sound. It requires an involved method to calculate p-values with response distributions that have variable mean and variable variance. A relatively quick analysis on the mean response to repeated sounds was desired for this paper, and the quantitative approach was not attempted.

The large spread of responses by subjects in figures 25–28 suggests considering variation among subjects in analyses. While the variation in response between subjects is not shown in this document, it was addressed and analyzed in Boucher et. al. [2].

6.3 Annoyance Response over Test Duration

The analysis in this section was used to explore if there was fatigue or test duration bias to subject responses. Figure 29 shows mean responses over the ten subject groups as a function of sound sequence number. For the test duration, the mean annoyance response was close to moderate over all groups. Since a different order of sounds was played to each group, each sound sequence number represents a different sound for each group. In this manner, the sound sequence numbers correspond to the elapsed test duration. The generally flat behavior of the mean annoyance rating as a function of test duration indicates subjects did not appear to be biased by test duration as a result of fatigue or other factors. For a list of the test sound sequences for each subject group, see section 11, and to obtain the subject group test sound sequences online [3], see section 12.4.

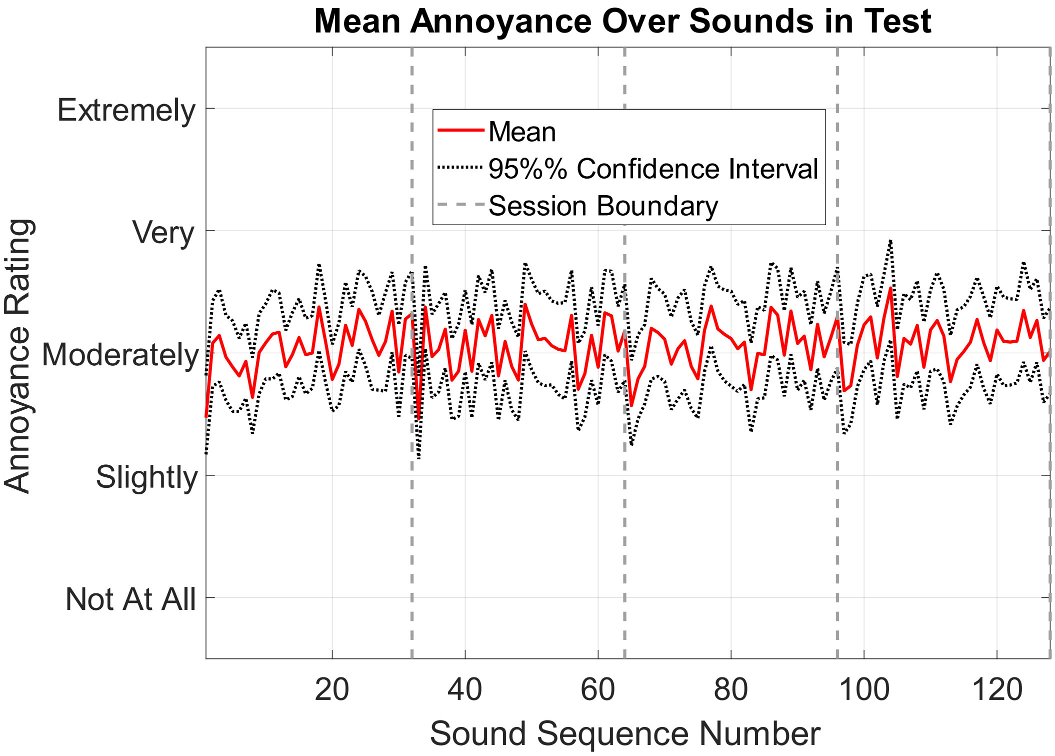

Figure 29. Mean annoyance response during test duration.

6.4 Practice Session Results

This section analyzes subject mean responses to the practice session sounds to determine the effectiveness of test practice in helping subjects prepare for the main test. Before the main test sessions, practice sounds were played to each of the ten groups of four subjects. The practice session was the first opportunity for subjects to input their annoyance response to sounds. These practice sounds consisted of 27 sounds chosen to reasonably cover the ranges of SQ metrics found in the main test; the practice sounds were presented to all subjects in the same order. These 27 practice sounds were also played to subjects again during the main test, randomly intermixed with the other test sounds.

The red trace in figure 30 shows the mean annoyance response, along with 95% confidence intervals, to each of the 27 practice sounds. The blue trace in figure 30 shows how subjects responded to these same sounds during the main test. In the comparison, the responses are shown to be different at the beginning of the practice session. After the first 10 or 11 practice session sounds were played, the annoyance responses begin to match more closely to the main test session responses. This result indicates that a practice session was not only helpful in teaching subjects how to respond to the test, but it was also beneficial in helping the subjects become consistent in their overall responses to the main test sounds.

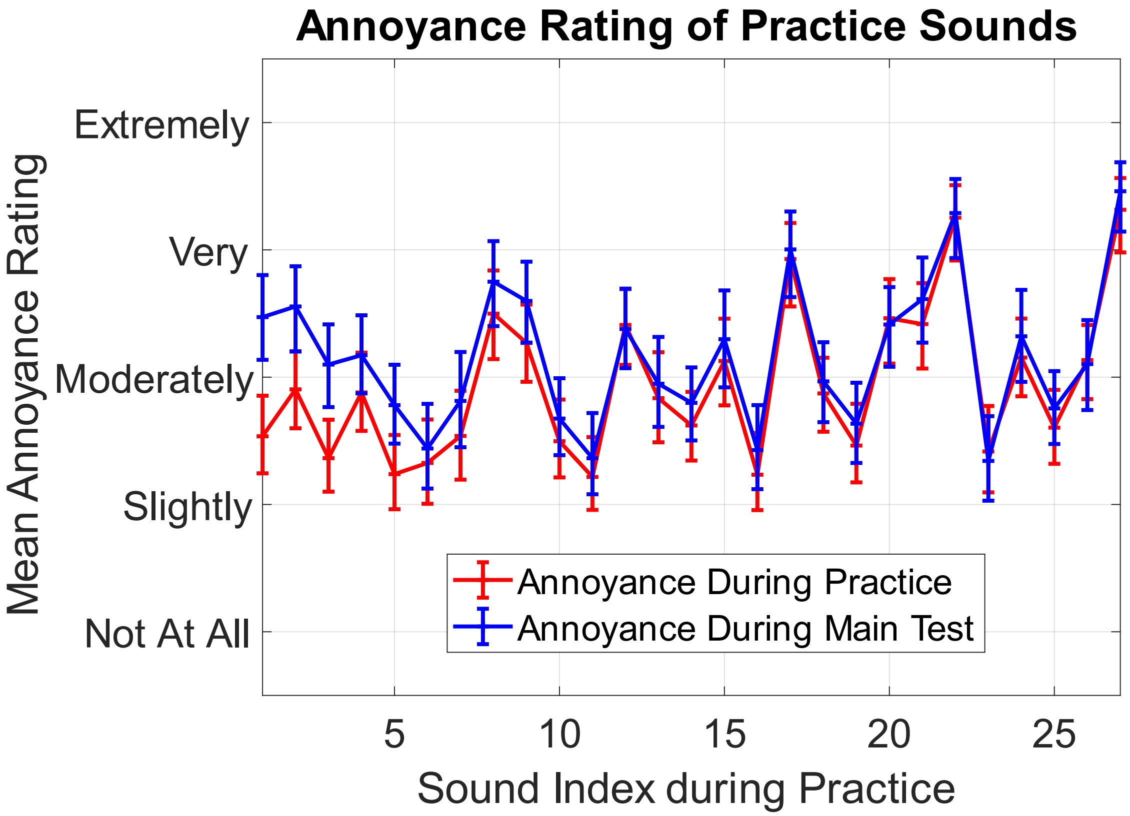

Figure 30. Annoyance Response to Practice Session Sounds.

The order of sounds used in the practice session is given in table 9 by sound number. Descriptions of the sounds by sound number are given in table 10 in section 9. Section 12.4 explains how to obtain subject sound responses to practic sounds online [3].

Table 9. Test sound number corresponding to sound index, or sound order, during practice session.

Sound Index/Order During Practice	Test Sound Number	Sound Index/Order During Practice	Test Sound Number	Sound Index/Order During Practice	Test Sound Number
1	5	10	41	19	61
2	95	11	36	20	55
3	84	12	23	21	9
4	103	13	18	22	74
5	109	14	99	23	115
6	59	15	13	24	34
7	63	16	1	25	50
8	30	17	77	26	69
9	121	18	88	27	80

6.5 Responses by Sound Number

This section presents an alternative view of the annoyance rankings to test sounds by showing the annoyance response to all sounds on one graph. Figure 31 shows the mean annoyance response with 95% confidence intervals for each sound by sound number given in table 10. Responses are colored by the SQ metric targeted by sets of sounds and the order effects from section 3.7. Sounds that are repeats of other sounds are not shown in figure 31. Hence, the annoyance response to only the 105 unique test sounds are given.

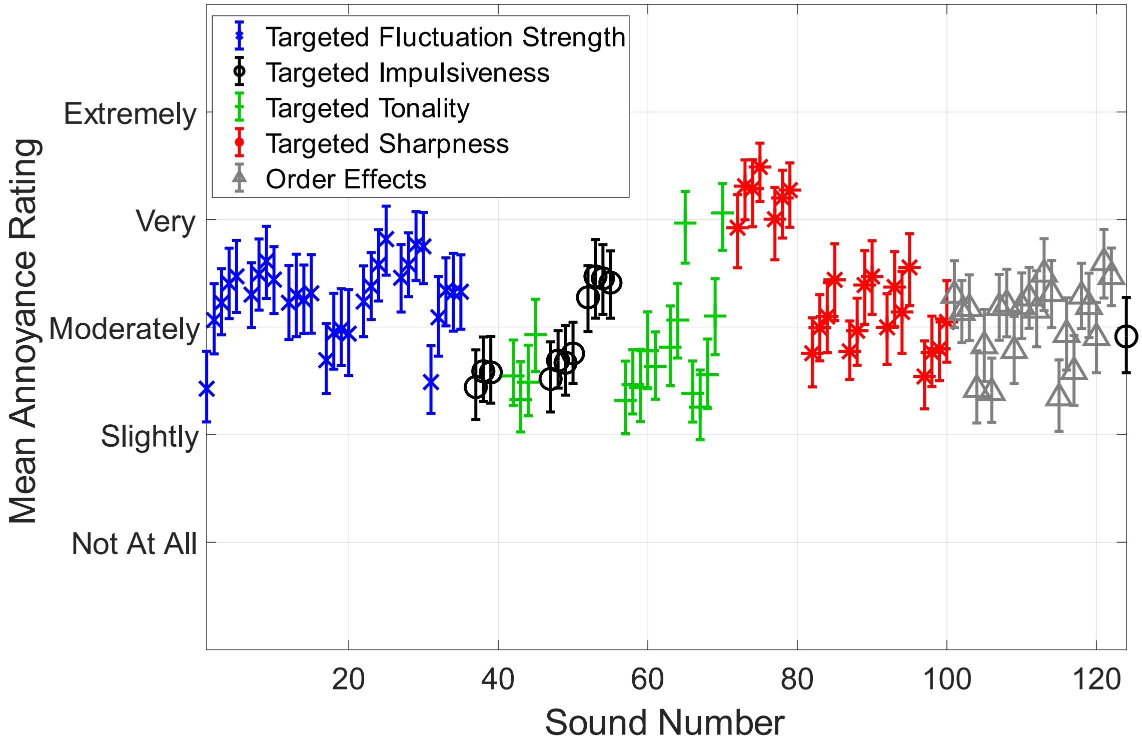

Figure 31. Mean annoyance response by sound number.

Figure 31 shows the range of responses elicited for sounds targeting each SQ metric. Responses to all sounds were above slightly annoyed. Responses to sounds targeting tonality and sharpness ranged to just above very annoyed. Responses to sounds targeting fluctuation strength ranged to just below very annoyed. Sounds targeting impulsiveness elicited the smallest range in the annoyance response. The range of annoyance responses for sounds created using higher order effects was also small. Conclusions have not been drawn about the parameter combinations used to create those sounds, but the small annoyance response range makes analyses challenging.

Section 12.4 explains how to obtain subject sound responses online [3].

7 Summary

This paper described how the RoQM-I-2017 psychoacoustic test was prepared, conducted, and analyzed. The methods used to generate, or synthesize, sets of test sound signals that targeted the SQ metrics of fluctuation strength, impulsiveness, tonality, and sharpness were provided in section 3. Synthesis methods mostly attempted to change one SQ metric while leaving other metrics roughly constant. The range of targeted SQ metric values was informed by analyzing rotorcraft flyover recordings. The loudness SQ metric was anticipated to be the dominant determinant of the subject response to test sounds, and so all test sounds were adjusted to have roughly the same loudness of ten Sone. During test sound generation, it was found that the roughness and impulsiveness metrics are highly dependent, so roughness was allowed to freely vary in the test sound generation process.

Measurements of test sounds at the subject seat locations in the test room were acquired after test completion in order to compare the sound metric values heard by the test subjects with the intended metric values. The measured loudness of the test sounds was slightly lower than the synthesized loudness but still roughly constant. Fluctuation strength, tonality, and sharpness did not change signficantly between synthesized and measured sounds. However, there was a significant change in the impulsiveness between synthesized and recorded sounds, with the result that the the impulsiveness of the measured sounds did not cover the range of values originally sought for the test. Because the just noticeable difference [23] of impulsiveness is unkown, conclusions about the relationship between impulsiveness and subject response cannot be made with the methods used in this analysis.

Other conclusions of this paper are:

- The poor predictive utility of metrics ASPL, ZSPL, PNL, and PNLT, which constitute certification metrics, for annoyance response compared with the predictive utility of the SQ metrics suggests that helicopter sounds have annoyance characteristics that are not captured by the constituent metrics but are captured by the SQ metrics.
- Responses to 104 test sounds may be considered normally distributed, but 24 sounds have responses that may not be considered normally distributed. This lack of normality complicates analyses based on confidence intervals.
- Subjects responded consistently to repeated test sounds.
- Test duration did not noticeably affect subject response.
- The practice session was important for helping subjects become familiar with answering test questions and achieving consistency in their responses.

8 References

[1] S. Krishnamurthy, A. Christian and S. A. Rizzi, "Psychoacoustic Test to Determine Sound Quality Metric Indicators of Rotorcraft Noise Annoyance," in *InterNoise 2018*, Chicago, IL, 2018.

[2] M. Boucher, S. Krishnamurthy, A. Christian and S. A. Rizzi, "Sound quality metric indicators of rotorcraft noise annoyance using multilevel regression analysis," *Proceedings of Meetings on Acoustics,* vol. 36, p. 040004, 2019.

[3] "Aircraft flyover simulation," NASA, 2021. [Online]. Available: https://stabserv.larc.nasa.gov/flyover/.

[4] HEAD acoustics GmbH, NVH Division, "ArtemiS Suite 12.2," 2016.

[5] Zwicker E. and H. Fastl., Psychoacoustics, Facts and Models, Second Updated Edition, Springer, 1999.

[6] "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," Deutsches Institut für Normung (DIN), 2009.

[7] U. Salford, "An introduction to Sound Quality testing," The Crescent, Salford, M5 4WT, UK - 0161 295 5000, 2020.

[8] "Measurement technique for the simulation of the auditory sensation of sharpness," Deutsches Institut für Normung (DIN) 45692, 2009.

[9] W. Aures, "Berechnungsverfahren fur den sensorischen Wohlklang beliebiger Schallsignale," *Acustica,* vol. 59, pp. 130-141, 1985.

[10] R. Sottek, "Modelle zur Signalverabeitung im menschlichen Gehör (Models for signal processing in the human ear)," in *Electrical Engineering and Information Technology*, Aachen, 1993.

[11] C. Burley, C. Smith and D. Conner, "Flight Acoustic Testing and Data Acquisition for the Rotorcraft Noise Model (RNM)," in *American Helicopter Society 62nd Annual Forum*, Phoenix, 2006.

[12] M. E. Watts, D. A. Conner and C. D. Smith, "Joint Eglin Acoustic Week III Data Report," NASA/TM-2010-216206, Hampton, VA, United States, 2010.

[13] D. Conner, M. Marcolini, B. Edwards and J. Brieger, "XV-15 Tiltrotor Low Noise Terminal Area Operations," in *American Helicopter Society 53rd Annual Forum*, Virginia Beach, VA, 1997.

[14] E. Greenwood and F. H. Schmitz, "Separation of Main and Tail Rotor Noise from Ground-Based Acoustic Measurements," *Journal of Aircraft,* vol. 51, pp. 464-472, 3 2014.

[15] M. E. Watts, E. Greenwood, B. Sim, J. Stephenson and C. D. Smith, "Helicopter Acoustic Flight Test with Altitude Variation and Maneuvers," NASA TM-2016-219354, Hampton, VA, 2016.

[16] A. Christian, D. D. Boyd Jr., N. S. Zawodny and S. A. Rizzi, "Auralization of tonal rotor noise components of a quadrotor flyover," in *InterNoise 2015*, San Francisco, CA, 2015.

[17] A. McMullen and P. Davies, "Results of a test using simulated rotorcraft flyovers designed to compare ratings of rotorcraft noise," in *Noise-Con 2014*, Fort Lauderdale, FL, 2014.

[18] S. R. More, "Aircraft Noise Characteristics and Metrics," Purdue University, PhD Thesis, West Lafayette, IN, 2011.

[19] "ISO 532-2:2017, Acoustics -- Methods for calculating loudness -- Part 2: Moore-Glasberg method," Acoustics, ISO/T. C. 43, 6 2017. [Online]. Available: https://www.iso.org/standard/63078.html.

[20] D. C. Montgomery, Design and Analysis of Experiments, Hoboken, NJ: John Wiley & Sons, Inc., 2006.

[21] K. J. Faller II, S. A. Rizzi and A. R. Aumann, "Acoustic performance of a real-time three-dimensional sound reproduction system," NASA/TM-2013-218004, 2013.

[22] L. V. Lopes and C. L. Burley, "ANOPP2 User's Manual," 2016.

[23] G. Fechner, Elemente der Psychophysik, Leipzig: Breitkopf and Hartel, 1862.

[24] J. M. Fields, R. G. D. Jong, T. Gjestland, I. H. Flindell, R. F. S. Job, S. Kurra, P. Lercher, M. Vallet, T. Yano, R. Guski, U. Felscher-suhr and R. Schumer, "Standardized General-purpose Noise Reaction Questions for Community Noise Surveys: Research and a Recommendation," *Journal of Sound and Vibration,* vol. 242, pp. 641-679, 2001.

[25] B. Efron, "Better Bootstrap Confidence Intervals," *Journal of the American Statistical Association,* vol. 82, no. 397, pp. 171-185, 1987.

[26] B. C. Moore, An Introduction to the Psychology of Hearing, 6th ed., Brill, 2013.

[27] M. H. DeGroot and M. J. Schervish, "Testing Whether a Distribution Is Normal," in *Probability and Statistics*, D. Lynch, Ed., Pearson Education, Inc., 2012, pp. 636 - 637.

[28] S. Krishnamurthy, S. A. Rizzi, D. D. Boyd and A. R. Aumann, "Auralization of Rotorcraft Periodic Flyover Noise from Design Predictions," in *Proceedings of the 74th Annual Forum, AHS International*, Phoenix, 2018.

[29] N. A. o. Sciences, Assessing Community Annoyance of Helicopter Noise, V. Mestre, S. Fidell, R. D. Horonjeff, P. Schomer, A. Hastings, B. G. Tabachnick and F. A. Schmitz, Eds., Washington, DC: The National Academies Press, 2017.

[30] P. D. Schomer, "High-energy impulsive noise assessment," *The Journal of the Acoustical Society of America,* vol. 79, pp. 182-186, 1986.

9 Appendix A, List of Test Sound Numbers

Table 10 lists all the sounds in the test by their unique sound number. Of the 128 sounds played to test subjects, 23 were repetitions of a previous sound. For those repetitions only, column 2 lists the sound number of the original sound. For example, Sound 6 is a repetition of Sound 1, the AS350 baseline. The third column gives a reference to where more details on the sound are found and which SQ metrics the sound is targeting. Section 12.1 explains how to obtain summary descriptions of each sound that includes the information in table 10. The fourth column indicates the result of the normality test discussed in section 6.1.

Table 10. List of sound numbers.

Sound Number	Sound is Same As Sound #	Reference	Outcome of Normality Test Indicates Normally Distributed Subject Responses?
1	-	AS350 baseline; sound 1, set 1, affecting fluctuation strength, see table 2	Yes
2	-	Sound 2, set 1, affecting fluctuation strength, see table 2	Yes
3	-	Sound 3, set 1, affecting fluctuation strength, see table 2	Yes
4	-	Sound 4, set 1, affecting fluctuation strength, see table 2	Yes
5	-	Sound 5, set 1, affecting fluctuation strength, see table 2	Yes
6	1	Sound 1, set 2, affecting fluctuation strength, see table 2	Yes
7	-	Sound 2, set 2, affecting fluctuation strength, see table 2	Yes
8	-	Sound 3, set 2, affecting fluctuation strength, see table 2	Yes
9	-	Sound 4, set 2, affecting fluctuation strength, see Table 2	Yes
10	-	Sound 5, set 2, affecting fluctuation strength, see table 2	Yes
11	1	Sound 1, set 3, affecting fluctuation strength, see table 2	Yes
12	-	Sound 2, set 3, affecting fluctuation strength, see table 2	Yes
13	-	Sound 3, set 3, affecting fluctuation strength, see table 2	Yes
14	-	Sound 4, set 3, affecting fluctuation strength, see table 2	Yes
15	-	Sound 5, set 3, affecting fluctuation strength, see table 2	Yes
16	1	Sound 1, set 4, affecting fluctuation strength, see table 2	No
17	-	Sound 2, set 4, affecting fluctuation strength, see table 2	Yes
18	-	Sound 3, set 4, affecting fluctuation strength, see table 2	No
19	-	Sound 4, set 4, affecting fluctuation strength, see table 2	Yes
20	-	Sound 5, set 4, affecting fluctuation strength, see table 2	Yes
21	1	Sound 1, set 5, affecting fluctuation strength, see table 2	Yes
22	-	Sound 2, set 5, affecting fluctuation strength, see table 2	Yes
23	-	Sound 3, set 5, affecting fluctuation strength, see table 2	Yes
24	-	Sound 4, set 5, affecting fluctuation strength, see table 2	Yes
25	-	Sound 5, set 5, affecting fluctuation strength, see table 2	Yes
26	1	Sound 1, set 6, affecting fluctuation strength, see table 2	Yes
27	-	Sound 2, set 6, affecting fluctuation strength, see table 2	Yes
28	-	Sound 3, set 6, affecting fluctuation strength, see table 2	Yes
29	-	Sound 4, set 6, affecting fluctuation strength, see table 2	Yes

Table 10. Continued.

Sound Number	Sound is Same As Sound #	Reference	Outcome of Normality Test Indicates Normally Distributed Subject Responses?
30	-	Sound 5, set 6, affecting fluctuation strength, see table 2	Yes
31	-	AS350 baseline main rotor only, harmonics have random phase; sound 1, set 7, affecting fluctuation strength, see table 2	Yes
32	-	Sound 2, set 7, affecting fluctuation strength, see table 2	Yes
33	-	Sound 3, set 7, affecting fluctuation strength, see table 2	Yes
34	-	Sound 4, set 7, affecting fluctuation strength, see table 2	Yes
35	-	Sound 5, set 7, affecting fluctuation strength, see table 2	Yes
36	31	Sound 1, set 1, affecting impulsiveness, see table 3	No
37	-	Sound 2, set 1, affecting impulsiveness, see table 3	Yes
38	-	Sound 3, set 1, affecting impulsiveness, see table 3	No
39	-	Sound 4, set 1, affecting impulsiveness, see table 3	Yes
40	-	AS350 baseline main rotor; sound 5, set 1, affecting impulsiveness, see table 3	No
41	40	Sound 1, set 1, affecting tonality, see table 4	Yes
42	-	Sound 2, set 1, affecting tonality, see table 4	No
43	-	Sound 3, set 1, affecting tonality, see table 4	Yes
44	-	Sound 4, set 1, affecting tonality, see table 4	Yes
45	-	Sound 5, set 1, affecting tonality, see table 4	Yes
46	31	Sound 1, set 2, affecting impulsiveness, see table 3	Yes
47	-	Sound 2, set 2, affecting impulsiveness, see table 3	Yes
48	-	Sound 3, set 2, affecting impulsiveness, see table 3	Yes
49	-	Sound 4, set 2, affecting impulsiveness, see table 3	No
50	-	Sound 5, set 2, affecting impulsiveness, see table 3	Yes
51	31	Sound 1, set 3, affecting impulsiveness, see table 3	Yes
52	-	Sound 2, set 3, affecting impulsiveness, see table 3	Yes
53	-	Sound 3, set 3, affecting impulsiveness, see table 3	Yes
54	-	Sound 4, set 3, affecting impulsiveness, see table 3	Yes
55	-	BVI baseline; sound 5, set 3, affecting impulsiveness, see table 3	Yes
56	31	Sound 1, set 2, affecting tonality, see table 4	Yes
57	-	Sound 2, set 2, affecting tonality, see table 4	Yes
58	-	Sound 3, set 2, affecting tonality, see table 4	Yes
59	-	Sound 4, set 2, affecting tonality, see table 4	Yes
60	-	Sound 5, set 2, affecting tonality, see table 4	No
61	-	Sound 1, set 3, affecting tonality, see table 4	Yes
62	1	Sound 2, set 3, affecting tonality, see table 4	Yes
63	-	Sound 3, set 3, affecting tonality, see table 4	Yes
64	-	Sound 4, set 3, affecting tonality, see table 4	Yes

Table 10. Continued.

Sound Number	Sound is Same As Sound #	Reference	Outcome of Normality Test Indicates Normally Distributed Subject Responses?
65	-	Sound 5, set 3, affecting tonality, see table 4	No
66	-	Sound 1, set 4, affecting tonality, see table 4	Yes
67	-	Sound 2, set 4, affecting tonality, see table 4	Yes
68	-	Sound 3, set 4, affecting tonality, see table 4	No
69	-	Sound 4, set 4, affecting tonality, see table 4	Yes
70	-	Sound 5, set 4, affecting tonality, see table 4	No
71	40	Sound 1, set 1, affecting sharpness, see table 5	Yes
72	-	Sound 2, set 1, affecting sharpness, see table 5	Yes
73	-	Sound 3, set 1, affecting sharpness, see table 5	No
74	-	Sound 4, set 1, affecting sharpness, see table 5	No
75	-	Sound 5, set 1, affecting sharpness, see table 5	No
76	1	Sound 1, set 2, affecting sharpness, see table 5	No
77	-	Sound 2, set 2, affecting sharpness, see table 5	No
78	-	Sound 3, set 2, affecting sharpness, see table 5	No
79	-	Sound 4, set 2, affecting sharpness, see table 5	No
80	75	Sound 5, set 2, affecting sharpness, see table 5	No
81	1	Sound 1, set 3, affecting sharpness, see table 6	Yes
82	-	Sound 2, set 3, affecting sharpness, see table 6	Yes
83	-	Sound 3, set 3, affecting sharpness, see table 6	Yes
84	-	Sound 4, set 3, affecting sharpness, see table 6	Yes
85	-	Sound 5, set 3, affecting sharpness, see table 6	No
86	1	Sound 1, set 4, affecting sharpness, see table 6	Yes
87	-	Sound 2, set 4, affecting sharpness, see table 6	Yes
88	-	Sound 3, set 4, affecting sharpness, see table 6	No
89	-	Sound 4, set 4, affecting sharpness, see table 6	No
90	-	Sound 5, set 4, affecting sharpness, see table 6	Yes
91	1	Sound 1, set 5, affecting sharpness, see table 6	Yes
92	-	Sound 2, set 5, affecting sharpness, see table 6	Yes
93	-	Sound 3, set 5, affecting sharpness, see table 6	No
94	-	Sound 4, set 5, affecting sharpness, see table 6	Yes
95	-	Sound 5, set 5, affecting sharpness, see table 6	Yes
96	31	Sound 1, set 6, affecting sharpness, see table 6	Yes
97	-	Sound 2, set 6, affecting sharpness, see table 6	Yes
98	-	Sound 3, set 6, affecting sharpness, see table 6	Yes
99	-	Sound 4, set 6, affecting sharpness, see table 6	Yes
100	-	Sound 5, set 6, affecting sharpness, see table 6	Yes

Table 10. Continued.

Sound Number	Sound is Same As Sound #	Reference	Outcome of Normality Test Indicates Normally Distributed Subject Responses?
101	-	Sound 2 of 5, second-order effect to change fluctuation strength and impulsiveness, see figure 13	Yes
102	-	Sound 2 of 5, second-order effect to change fluctuation strength and tonality, see figure 13	Yes
103	-	Sound 2 of 5, second-order effect to change fluctuation strength and sharpness, see figure 13	Yes
104	-	Sound 2 of 5, second-order effect to change impulsiveness and tonality, see figure 13	Yes
105	-	Sound 2 of 5, second-order effect to change impulsiveness and sharpness, see figure 13	Yes
106	-	Sound 2 of 5, second-order effect to change tonality and sharpness, see figure 13	No
107	-	Sound 2 of 5, third-order effect to change fluctuation strength, impulsiveness, and tonality, see figure 14	Yes
108	-	Sound 2 of 5, third-order effect to change fluctuation strength, impulsiveness, and sharpness, see figure 14	Yes
109	-	Sound 2 of 5, third-order effect to change impulsiveness, tonality, and sharpness, see figure 14	Yes
110	-	Sound 2 of 5, third-order effect to change fluctuation strength, tonality, and sharpness, see figure 14	Yes
111	-	Sound 2 of 5, fourth-order effect to change fluctuation strength, impulsiveness, tonality, and sharpness, see figure 15	Yes
112	-	Sound 3 of 5, second-order effect to change fluctuation strength and impulsiveness, see figure 13	Yes
113	-	Sound 3 of 5, second-order effect to change fluctuation strength and tonality, see figure 13	Yes
114	-	Sound 3 of 5, second-order effect to change fluctuation strength and sharpness, see figure 13	Yes
115	-	Sound 3 of 5, second-order effect to change impulsiveness and tonality, see figure 13	Yes
116	-	Sound 3 of 5, second-order effect to change impulsiveness and sharpness, see figure 13	Yes
117	-	Sound 3 of 5, second-order effect to change tonality and sharpness, see figure 13	Yes
118	-	Sound 3 of 5, third-order effect to change fluctuation strength, impulsiveness, and tonality, see figure 14	Yes
119	-	Sound 3 of 5, third-order effect to change fluctuation strength, impulsiveness, and sharpness, see figure 14	Yes
120	-	Sound 3 of 5, third-order effect to change impulsiveness, tonality, and sharpness, see figure 14	Yes
121	-	Sound 3 of 5, third-order effect to change fluctuation strength, tonality, and sharpness, see figure 14	Yes
122	-	Sound 3 of 5, fourth-order effect to change fluctuation strength, impulsiveness, tonality, and sharpness, see figure 15	Yes
123	10	Sound 5, set 2, affecting fluctuation strength, see table 2	Yes

Table 10. Concluded.

Sound Number	Sound is Same As Sound #	Reference	Outcome of Normality Test Indicates Normally Distributed Subject Responses?
124	-	Sound 6, set 2, affecting impulsiveness, see table 3	Yes
125	124	Sound 6, set 2, affecting impulsiveness, see table 3	Yes
126	55	BVI baseline, see table 3	Yes
127	55	BVI baseline, see table 3	Yes
128	65	Sound 5, set 3, affecting tonality, see table 4	Yes

10 Appendix B, Multiple Linear Regression Results for Measured Sounds

This appendix gives results of multiple linear regressions between annoyance and SQ metrics of the seat recordings. It is the same process used as in Krishnamurthy et al. [1], but here, seat recordings are used instead of the synthesized sounds. To apply the multiple linear regression, a first step is that the 5% exceedance values of the SQ metrics from seat recordings, including loudness, are normalized so that the minimum 5% exceedance value of a metric over all sounds is zero, and its maximum 5% exceedance value over all sounds is one. The annoyance response data are not normalized and remain between values of 1 and 11. As in Krishnamurthy et al., since roughness is highly dependent with impulsiveness, it is not included in the regressions.

The seat recordings differed at each of the 4 subject seat locations. Tables 11, 13, 15, and 17 list r^2 values of each multiple linear regression between mean annoyance and SQ metric combinations for each test seat (numbered 1-4). The 5% exceedance values of loudness, sharpness, tonality, fluctuation strength, and impulsiveness are denoted N5, S5, T5, F5, and I5, respectively. The fifth column of each table lists the reduction in r^2 when a single metric is removed from the regression.

Table 11 will be used as an example on reading the r^2 tables. For combination 1 at the top of table 11, a linear regression between mean annoyance and all five metrics gives an r^2 value of 0.397. Removing a single metric from combination 1 produces one of the next five combinations. This method is a simple way to provide rough guidance on the relative importance of a metric. Rerunning a regression without a metric reduces the r^2 value by an amount Δr^2 as shown in table 11. To demonstrate, removing S5 from combination 2 or N5 from combination 3 both produce the same metric combination of T5, F5, and I5. Regression with these three metrics gives an r^2 of approximately 0.234, which is the difference between approximately 0.396 and 0.162 when removing S5 from combination 2. An r^2 value of 0.234 is also the difference between 0.273 and 0.038 when removing N5 from combination 3. Tables 13, 15, and 17 can be read the same way but for different test subject seats.

Compared with the reduction in r^2 when removing metrics from combinations in Krishnamurthy et al., the Δr^2 values in tables 11, 13, 15, and 17 are slightly lower. A cursory view of the values appear to show mostly consistent values between the subject seats. There is some variation among the seats in the Δr^2 values when removing fluctuation strength, but this metric can still be discerned to be third in importance to describing the annoyance response after sharpness and tonality.

Tables 12, 14, 16, and 18 give the multiple linear regression relationship between mean annoyance and SQ metric combinations for test seats 1, 2, 3, and 4. In these tables, some regression coefficients for loudness and impulsiveness are negative. Also, the negative coefficients are not consistently negative over the seats except for loudness in combination 3 and impulsiveness in combination 4. Krishnamurthy et al. stated that all regression coefficients for multiple linear regression with synthesized sounds were positive although the coefficient values were not shown. Since loudness and impulsiveness both have relatively small Δr^2 values when removed from metric combinations, the sign variation in their regression coefficients may be because of the same reasons for low r^2 values discussed in section 4.1.

Table 11. The r^2 and Δr^2 values of a multiple linear regression with mean annoyance and different metric combinations from measured sounds for seat 1.

Combination Number	**Metric Combination**	r^2	Δr^2, **Removing N5**	Δr^2, **Removing S5**	Δr^2, **Removing T5**	Δr^2, **Removing F5**	Δr^2, **Removing I5**
1	N5, S5, T5, F5, I5	0.397	0	0.124	0.081	0.068	0.001
2	S5, T5, F5, I5	0.396	-	0.162	0.081	0.14	0.001
3	N5, T5, F5, I5	0.273	0.038	-	0.14	0.122	0.013
4	N5, S5, F5, I5	0.316	0.001	0.183	-	0.054	0.015
5	N5, S5, T5, I5	0.329	0.073	0.178	0.067	-	0
6	N5, S5, T5, F5	0.396	0	0.136	0.094	0.067	-

Table 12. Multiple linear regression equation from measured sounds for seat 1.

Combination Number	**Metric Combination**	**Regression Equation with Mean Annoyance (A)**
1	N5, S5, T5, F5, I5	A = (–0.159 x N5) + (1.685 x S5) + (1.253 x T5) + (1.628 x F5) + (0.162 x I5) + 5.216
2	S5, T5, F5, I5	A = (1.722 x S5) + (1.242 x T5) + (1.547 x F5) + (0.141 x I5) + 5.158
3	N5, T5, F5, I5	A = (–1.617 x N5) + (1.609 x T5) + (2.127 x F5) + (0.598 x I5) + 5.882
4	N5, S5, F5, I5	A = (0.249 x N5) + (1.996 x S5) + (1.447 x F5) – (0.585 x I5) + 5.703
5	N5, S5, T5, I5	A = (1.645 x N5) + (1.968 x S5) + (1.136 x T5) – (0.115 x I5) + 4.702
6	N5, S5, T5, F5	A = (–0.097 x N5) + (1.717 x S5) + (1.191 x T5) + (1.592 x F5) + 5.264

Table 13. The r^2 and Δr^2 values of a multiple linear regression with mean annoyance and different metric combinations from measured sounds for seat 2.

Combination Number	**Metric Combination**	r^2	Δr^2, **Removing N5**	Δr^2, **Removing S5**	Δr^2, **Removing T5**	Δr^2, **Removing F5**	Δr^2, **Removing I5**
1	N5, S5, T5, F5, I5	0.454	0.008	0.177	0.068	0.03	0
2	S5, T5, F5, I5	0.447	-	0.187	0.071	0.155	0
3	N5, T5, F5, I5	0.278	0.018	-	0.137	0.095	0.005
4	N5, S5, F5, I5	0.387	0.011	0.246	-	0.025	0.022
5	N5, S5, T5, I5	0.425	0.133	0.242	0.063	-	0
6	N5, S5, T5, F5	0.454	0.008	0.182	0.089	0.029	-

Table 14. Multiple linear regression equation from measured sounds for seat 2.

Combination Number	**Metric Combination**	**Regression Equation with Mean Annoyance (A)**
1	N5, S5, T5, F5, I5	A = (0.707 x N5) + (2.145 x S5) + (1.188 x T5) + (1.208 x F5) + (0.071 x I5) + 4.941
2	S5, T5, F5, I5	A = (1.927 x S5) + (1.214 x T5) + (1.693 x F5) + (0.080 x I5) + 5.171
3	N5, T5, F5, I5	A = (–0.955 x N5) + (1.644 x T5) + (2.069 x F5) + (0.395 x I5) + 5.597
4	N5, S5, F5, I5	A = (0.843 x N5) + (2.461 x S5) + (1.107 x F5) – (0.711 x I5) + 5.538
5	N5, S5, T5, I5	A = (1.808 x N5) + (2.402 x S5) + (1.145 x T5) + (0.033 x I5) + 4.680
6	N5, S5, T5, F5	A = (0.709 x N5) + (2.154 x S5) + (1.158 x T5) + (1.204 x F5) + 4.976

Table 15. The r^2 and Δr^2 values of a multiple linear regression with mean annoyance and different metric combinations from measured sounds for seat 3.

Combination Number	**Metric Combination**	r^2	Δr^2, **Removing N5**	Δr^2, **Removing S5**	Δr^2, **Removing T5**	Δr^2, **Removing F5**	Δr^2, **Removing I5**
1	N5, S5, T5, F5, I5	0.443	0.002	0.181	0.086	0.047	0
2	S5, T5, F5, I5	0.441	-	0.179	0.128	0.134	0.001
3	N5, T5, F5, I5	0.262	0	-	0.134	0.041	0.006
4	N5, S5, F5, I5	0.357	0.043	0.229	-	0.017	0.017
5	N5, S5, T5, I5	0.396	0.089	0.175	0.056	-	0
6	N5, S5, T5, F5	0.443	0.003	0.187	0.103	0.047	-

Table 16. Multiple linear regression equation from measured sounds for seat 3.

Combination Number	**Metric Combination**	**Regression Equation with Mean Annoyance (A)**
1	N5, S5, T5, F5, I5	A = (0.360 x N5) + (1.909 x S5) + (1.386 x T5) + (1.415 x F5) + (0.096 x I5) + 4.928
2	S5, T5, F5, I5	A = (1.880 x S5) + (1.489 x T5) + (1.634 x F5) + (0.143 x I5) + 4.975
3	N5, T5, F5, I5	A = (–0.100 x N5) + (1.707 x T5) + (1.322 x F5) + (0.395 x I5) + 5.216
4	N5, S5, F5, I5	A = (1.428 x N5) + (2.121 x S5) + (0.801 x F5) – (0.582 x I5) + 5.412
5	N5, S5, T5, I5	A = (1.587 x N5) + (1.876 x S5) + (1.056 x T5) – (0.081 x I5) + 4.897
6	N5, S5, T5, F5	A = (0.390 x N5) + (1.921 x S5) + (1.345 x T5) + (1.395 x F5) + 4.969

Table 17. The r^2 and Δr^2 values of a multiple linear regression with mean annoyance and different metric combinations from measured sounds for seat 4.

Combination Number	**Metric Combination**	r^2	Δr^2, **Removing N5**	Δr^2, **Removing S5**	Δr^2, **Removing T5**	Δr^2, **Removing F5**	Δr^2, **Removing I5**
1	N5, S5, T5, F5, I5	0.381	0.009	0.131	0.099	0.074	0.002
2	S5, T5, F5, I5	0.371	-	0.153	0.091	0.084	0
3	N5, T5, F5, I5	0.249	0.031	-	0.166	0.076	0.012
4	N5, S5, F5, I5	0.282	0.001	0.199	-	0.046	0.016
5	N5, S5, T5, I5	0.307	0.02	0.134	0.07	-	0
6	N5, S5, T5, F5	0.379	0.008	0.141	0.112	0.072	-

Table 18. Multiple linear regression equation from measured sounds for seat 4.

Combination Number	**Metric Combination**	**Regression Equation with Mean Annoyance (A)**
1	N5, S5, T5, F5, I5	A = (–0.767 x N5) + (1.586 x S5) + (1.637 x T5) + (1.643 x F5) + (0.236 x I5) + 5.175
2	S5, T5, F5, I5	A = (1.674 x S5) + (1.427 x T5) + (1.221 x F5) + (0.112 x I5) + 5.085
3	N5, T5, F5, I5	A = (-1.367 x N5) + (2.070 x T5) + (1.673 x F5) + (0.568 x I5) + 5.394
4	N5, S5, F5, I5	A = (0.266 x N5) + (1.900 x S5) + (1.266 x F5) – (0.582 x I5) + 5.761
5	N5, S5, T5, I5	A = (0.777 x N5) + (1.602 x S5) + (1.356 x T5) + (0.030 x I5) + 5.096
6	N5, S5, T5, F5	A = (–0.682 x N5) + (1.619 x S5) + (1.524 x T5) + (1.604 x F5) + 5.269

11 Appendix C, List of Test Sounds for Each Subject Group

Table 19 lists the sequence of sounds that were played to each of the ten test subject groups where a sound is identifed by its unique number from table 10. Section 12.4 explains how to obtain the data in table 19 online [3].

Table 19. List of Sound Numbers for Each Test Subject Group

Sound Sequence Number	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Group 7	Group 8	Group 9	Group 10
1	5	94	123	56	117	26	111	88	21	126
2	85	78	7	108	69	78	79	104	61	34
3	77	114	23	60	53	66	119	44	105	74
4	93	42	59	4	9	58	75	12	33	50
5	109	106	31	16	89	86	83	60	65	78
6	97	110	35	112	61	38	107	116	81	42
7	49	90	3	8	21	6	3	92	69	38
8	57	10	67	68	1	118	91	32	93	98
9	125	54	11	44	37	98	15	128	29	66
10	25	6	83	40	41	74	35	4	37	86
11	113	126	15	20	65	2	67	84	5	22
12	105	18	55	48	105	18	99	52	53	114
13	37	62	71	28	97	114	87	124	101	106
14	33	82	63	84	25	42	11	28	113	10
15	21	46	75	124	101	50	123	20	13	18
16	117	14	127	64	113	110	103	120	109	6
17	17	22	119	120	17	22	47	112	121	26
18	89	2	19	80	33	90	19	48	9	118
19	13	98	51	88	85	70	127	16	17	58
20	1	118	99	100	93	102	71	64	1	90
21	121	26	39	76	121	126	63	80	97	122
22	69	30	27	72	73	106	59	100	41	94
23	81	102	103	128	49	54	31	40	85	46
24	53	50	95	36	45	30	95	56	77	102
25	45	70	79	32	125	94	51	96	89	62
26	9	122	115	12	29	46	115	76	73	110
27	101	86	107	116	57	34	39	72	57	30
28	29	66	43	92	5	62	55	8	125	70
29	41	74	111	24	13	14	23	24	49	14
30	61	34	47	104	77	82	43	68	117	82
31	73	58	87	96	81	122	7	108	45	54
32	65	38	91	52	109	10	27	36	25	2
33	94	115	44	37	38	115	96	109	122	127
34	102	119	52	93	34	111	64	73	30	79
35	78	51	60	49	26	107	124	65	14	59

Table 19. Continued.

Sound Sequence Number	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Group 7	Group 8	Group 9	Group 10
36	26	127	80	117	82	39	40	33	82	47
37	126	55	8	65	46	127	48	25	106	7
38	98	31	64	105	66	87	84	125	58	3
39	46	15	84	1	14	99	68	9	78	87
40	82	83	32	9	90	119	120	49	54	27
41	14	99	96	125	126	31	44	21	38	71
42	90	35	116	25	78	59	24	121	114	99
43	54	7	4	89	114	11	12	69	86	83
44	58	107	72	77	74	83	128	85	110	23
45	6	39	56	97	62	67	80	53	22	43
46	38	27	112	121	118	35	8	41	18	111
47	10	59	36	85	98	63	36	29	2	11
48	66	87	40	57	18	95	92	57	94	51
49	74	75	24	113	94	55	60	5	42	63
50	50	43	28	33	30	15	88	17	66	35
51	22	47	108	17	86	123	32	93	98	15
52	118	19	20	73	106	71	108	113	34	107
53	2	11	128	109	22	7	20	97	102	95
54	30	63	104	41	50	103	104	1	50	119
55	122	111	124	61	102	47	4	89	74	115
56	62	91	76	101	54	3	52	13	6	75
57	18	71	16	13	110	51	100	81	46	67
58	114	67	88	45	10	91	56	101	90	103
59	86	103	48	5	6	75	28	61	26	123
60	42	95	12	21	2	19	112	117	62	39
61	70	3	92	69	42	79	116	45	118	31
62	34	123	68	29	70	23	76	77	126	91
63	110	23	120	81	58	27	16	37	10	19
64	106	79	100	53	122	43	72	105	70	55
65	55	45	22	67	19	13	94	99	71	49
66	71	81	2	11	123	65	54	87	119	17
67	111	9	46	83	27	101	62	71	103	53
68	15	117	58	79	15	1	70	3	43	25
69	103	89	50	35	103	93	114	103	95	9
70	123	41	94	119	95	109	38	91	3	45
71	35	5	34	87	31	57	10	83	31	61
72	63	85	42	3	75	37	22	107	35	41
73	3	101	66	43	99	17	30	63	75	29
74	107	13	86	15	111	45	66	11	11	113

Table 19. Continued.

Sound Sequence Number	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Group 7	Group 8	Group 9	Group 10
75	31	49	6	19	59	121	102	39	99	89
76	91	33	26	39	119	73	34	51	15	65
77	11	37	30	99	87	5	90	35	63	13
78	51	53	90	51	115	89	126	119	107	101
79	39	105	78	95	51	69	106	15	79	125
80	79	21	106	107	55	77	58	19	83	105
81	87	1	82	115	79	25	46	111	55	93
82	127	61	126	91	47	117	122	95	123	5
83	23	29	38	31	91	41	86	67	7	97
84	83	121	18	59	107	81	74	75	115	33
85	27	97	114	7	11	105	42	115	67	37
86	7	125	74	27	83	61	118	79	23	69
87	95	109	70	55	7	29	18	23	59	85
88	67	77	54	63	23	125	110	7	39	21
89	119	113	122	103	3	85	50	123	19	77
90	75	65	14	71	67	21	98	47	27	1
91	47	93	118	111	127	113	26	43	47	57
92	115	57	102	47	43	9	14	59	51	117
93	59	25	98	75	39	33	6	127	127	81
94	43	17	110	23	63	97	82	55	111	109
95	99	69	62	123	35	49	2	27	87	73
96	19	73	10	127	71	53	78	31	91	121
97	124	4	17	66	96	100	17	110	32	80
98	116	8	57	46	60	112	117	90	36	32
99	24	108	97	114	28	108	125	62	44	92
100	12	120	77	14	64	80	69	42	84	68
101	88	72	29	86	108	124	101	98	128	52
102	112	16	93	82	32	4	85	106	108	16
103	44	40	117	90	72	8	53	74	16	56
104	108	96	33	74	80	20	81	2	120	24
105	60	84	49	62	44	36	73	66	96	40
106	120	128	101	106	128	12	25	10	56	112
107	56	36	113	122	8	92	33	18	24	76
108	68	68	121	2	100	16	9	70	52	100
109	96	60	45	54	124	64	105	86	60	124
110	72	48	89	30	68	28	21	118	88	96
111	40	12	65	78	88	68	49	94	48	128
112	100	116	21	18	24	52	37	126	64	72
113	92	80	37	34	104	40	109	102	92	120

Table 19. *Concluded.*

Sound Sequence Number	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Group 7	Group 8	Group 9	Group 10
114	52	124	69	26	52	56	113	38	100	88
115	76	44	41	10	116	120	29	58	72	12
116	84	56	73	22	48	76	93	14	12	8
117	128	88	105	70	40	88	41	114	112	20
118	4	28	53	58	84	44	89	34	40	28
119	8	24	61	38	112	96	45	54	68	116
120	104	104	109	118	16	24	57	22	80	84
121	32	32	85	6	4	60	121	122	124	44
122	20	76	9	42	36	84	61	30	28	36
123	36	100	5	110	92	116	65	46	20	48
124	80	112	81	94	120	104	77	26	76	4
125	28	64	1	126	56	32	97	82	8	60
126	16	20	25	50	12	72	5	50	4	64
127	64	52	13	98	76	128	1	6	104	104
128	48	92	125	102	20	48	13	78	116	108

12 Appendix D, Test Sound Files and Response Data

12.1 Reading Sound File Descriptions

A description of the sound for each test stimulus with the details from sections 3 and 9 is found online [3] in the folder “stimuliDescriptions.” In this folder, stimuli descriptions can be found in the file “stimuliDescriptions.pdf.” Stimuli descriptions can also be read through the *.xlsx file “stimuliDescriptions.xlsx.” The sheet, “Stimuli Descriptions,” within the “stimuliDescriptions.xlsx” file, contains descriptions of all the 128 test sounds. Each row of the sheet contains one of the sound numbers from section 9 and the corresponding sound description.

12.2 Reading Synthesized Sound Files

The synthesized sound files for all the sound stimuli are available online [3]. They are in the folder labeled “synthesized_ROQM1_soundFiles.” Each sound file is named “ROQM_xxx.wav,” where “xxx” is the sound number. The particular sounds the sound numbers refer to are given in table 10, section 9. Sound 1 and sound 128, for example, are sound files “ROQM_001.wav” and “ROQM_128.wav,” respectively. All sound files are 32 bits per sample.

12.3 Reading Seat Recordings

The sound files for all the sound stimuli from seat recordings are available online [3] and located in the folder labeled “recorded_ROQM1_soundFiles.” Each sound file is named “ROQM_rec_seatA_cal_soundB.wav,” where “A” is the seat number and “B” is the sound number. Seat numbers are subject seats 1–4. A total of 128 sound numbers are associated with each seat number. The particular sounds the sound numbers refer to are given in table 10, section 9. The “rec” in the sound file name means the sound is recorded, and “cal” means the sound file was generated by adjusting the raw recorded sound by a calibration signal. All sound files are 32 bits per sample.

12.4 Reading Response Data

Test response data can be read through MATLAB with the binary file, “ROQMSubjectResponse.mat” given in the folder, “ROQM1_responseData,” online [3]. Variables contained in this binary file are explained in table 20 below.

Table 20. List of variables in response data binary file.

Variable	Description
numGroups	Number of test subject groups. It is equal to 10 groups.
numSessionsPerGroup	Number of test sessions for each subject group. It is equal to 4 sessions.
numSounds	Number of test stimuli. It is equal to 128 sounds.
numSoundsPerSession	Number of test stimuli in each test session. It is equal to 32 sounds.
numSoundsPractice	Number of sounds played in the practice session. It is equal to 27 sounds.
numSubjects	Number of test subjects. It is equal to 40 subjects.
numSubjectsPerGroup	This variable is a vector numGroups long where each element indexes the test group number and indicates the number of test subjects in that group. All elements are equal to 4 subjects.

Table 20. Concluded.

Variable	Description
annoyanceHistory	This matrix contains the annoyance response time history for each test subject. The matrix is 40x155 (40 rows by 155 columns) where each row indexes the subject number. Each column indexes the sound sequence, or order, number. The first 27 columns are responses to the practice sound questions. The remaining 128 columns are responses to each sound during the main test. In row i for subject i, column j contains the annoyance response of subject i to the jth sound played to the subject. Annoyance responses are given a double precision numerical value between 1–11 in units of annoyance rating.
annoyanceMatrix	This matrix contains the annoyance response to each sound number by each test subject. The sound number is not the same as the sound sequence, or order, number. The sound number refers to the number given for a sound in the first column of table 10, section 9. This matrix is 40x128 (40 rows by 128 columns). Each row indexes the subject number, and each column indexes the sound number. In row i for subject i, column j contains the annoyance response of subject i to sound number j. The annoyance responses are from the main test and not from the practice session. Annoyance responses are given a double precision numerical value between 1–11 in units of annoyance rating.
groupSoundSequenceVec	This matrix contains the order of sound numbers played to each subject group. The matrix is 10x155 (10 rows by 155 columns) where each row indexes the subject group number. Each column indexes the sound sequence, or order, number. The first 27 columns index the practice session sounds, and the remaining columns index the sounds in the main test. An element in row i and column j contains the sound number of the jth sound played to subject group i. This matrix contains the same data as table 19, section 11, but also includes the sound numbers for the practice session given in table 9, section 6.4.
groupTimeVec	This matrix contains the time stamp, in seconds, that each sound was played to all subjects in each group. The matrix is 10x155 (10 rows by 155 columns). Each row indexes the subject group number. Each column indexes the sound sequence, or order, number. The first 27 columns index the practice session sounds, and the remaining columns index the sounds in the main test.
sessionOnTimes	This 10x5 (10 rows by 5 columns) contains the time stamp, in seconds, when each session of the test began for each subject group. Each row indexes the subject group number. Each column indexes the session number. Session 1 is the practice session. Sessions 2–5 are the main test sessions.
sessionOffTimes	This 10x5 (10 rows by 5 columns) matrix contains the time stamp, in seconds, when each session of the test ended for each subject group. Each row indexes the subject group number. Each column indexes the session number. Session 1 is the practice session. Sessions 2–5 are the main test sessions.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYY)*
01-09-2021

2. REPORT TYPE
Technical Memorandum

3. DATES COVERED *(From - To)*
October 2018 - September 2021

4. TITLE AND SUBTITLE
Rotorcraft Sound Quality Metric Test 1: Stimuli Generation and Supplemental Analyses

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Siddhartha Krishnamurthy (NASA),
Matthew A. Boucher (NASA),
Andrew W. Christian (NASA),
Stephen A. Rizzi (NASA)

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER
WBS 664817.02.07.03.02.05

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
NASA Langley Research Center
Hampton, VA 23681-2199

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
National Aeronautics and Space Administration
Washington, DC 20546-0001

10. SPONSOR/MONITOR'S ACRONYM(S)
NASA

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
NASA-TM-20205008997

12. DISTRIBUTION/AVAILABILITY STATEMENT
Unclassified
Subject Category 71
Availability: NASA STI Program (757) 864-9658

13. SUPPLEMENTARY NOTES

14. ABSTRACT

A psychoacoustic test was conducted at the NASA Langley Research Center Exterior Effects Room (EER) to assess annoyance to simulated helicopter sounds over a range of sound quality (SQ) metric values. Initial findings identified important SQ metrics as sharpness, tonality, and fluctuation strength. This document is a supplement to the initial findings in which the following are discussed: (i) a detailed treatment of the sound generation process, (ii) the impact of analyzing results with stimuli measured in the EER instead of the intended synthesized stimuli, (iii) an evaluation of annoyance responses with certification metrics, and (iv) adjunct analyses related to the test methodology.

15. SUBJECT TERMS
Psychoacoustic test, psychoacoustics, sound quality metrics, rotorcraft noise, helicopter noise, annoyance, linear regression, loudness, tonality, impulsiveness, sharpness, roughness, fluctuation strength, human response

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			STI Help Desk (email: help@sti.nasa.gov)
					19b. TELEPHONE NUMBER *(Include area code)*
U	U	U	UU	61	(757) 864-9658