INTER-NOISE 2018
Impact of Noise Control Engineering
26-29 AUGUST
CHICAGO, ILLINOIS

Psychoacoustic Test to Determine Sound Quality Metric Indicators of Rotorcraft Noise Annoyance

Siddhartha Krishnamurthy[a]
Andrew Christian[b]
Stephen A. Rizzi[c]
NASA Langley Research Center
2 North Dryden Street
Hampton VA 23681

ABSTRACT

Noise certification metrics such as Effective Perceived Noise Level and Sound Exposure Level are used to ensure that helicopters meet regulations, but these metrics may not be good indicators of annoyance since noise complaints against helicopters persist. Sound quality (SQ) metrics, specifically fluctuation strength, tonality, impulsiveness, roughness, and sharpness, are explored to determine their relationship with annoyance. A psychoacoustic test was conducted at the NASA Langley Research Center Exterior Effects Room to assess annoyance to helicopter-like sounds over a range of SQ metric values. The amplitude, phase, and frequency of the AS350 helicopter main and tail rotor blade passage signal harmonics were manipulated to produce 105 unique helicopter-like sounds with prescribed values of SQ metrics. All sounds were set to roughly the same loudness level. These sounds were played to 40 subjects who rated each sound for annoyance. Analyses given in this paper point to which SQ metrics are important to the helicopter noise annoyance response.

1 INTRODUCTION

Community exposure to aircraft noise is detrimental to health on both psychological and physiological levels[1]. Noise related annoyance caused by helicopter operations presents a particularly complicated problem for both researchers and regulators. Its acoustic and socio-economic underpinnings have more, and deeper, facets than those for fixed-wing aircraft. The FAA released a study in 2004 that stated, among other things, that "additional development of models for characterizing the human response to helicopter noise should be pursued[2]." To date, a sparse amount of additional work has been done (for instance McMullen and Davies[3]), yet the

---

[a] email: siddhartha.krishnamurthy@nasa.gov
[b] email: andrew.christian@nasa.gov
[c] email: s.a.rizzi@nasa.gov

general understanding of the problem has not improved greatly since the original formulation of current helicopter noise regulations took place in the mid-1980s.

Regarding the state of the science of human response to helicopter noise, the recent report from the Airport Cooperative Research Program (ACRP) provides an excellent review of literature spanning more than 40 years[4]. A central question in much of this past research is whether or not helicopter noise is, in and of itself, more annoying to people than an equal amount of noise from fixed-wing aircraft (when measured on equal scales such as dB_A). There seems to be no wide scientific consensus on this point. However, what is generally agreed upon is that noise from helicopters is significantly more complex than noise from fixed-wing aircraft. This is not only true in the purely acoustical sense: helicopters have a much wider operational envelope which includes hovering (creating exposures of unlimited length) and operate in closer proximity to dense populations than fixed-wing aircraft.

Acoustically, there are multiple aspects of helicopter noise that are not shared by fixed-wing aircraft. Firstly, there are much fewer operations compared with fixed-wing aircraft noise. Helicopter operations are not distributed as regularly in space or time as other noise sources. This makes measurement and correlation with annoyance in real-world situations extremely difficult. Another acoustic issue is whether the impulsive quality of the noise increases annoyance. Many researchers have hypothesized that two sounds, played at similar intensity, will cause a differential annoyance response based only on the difference in the impulsive quality of the sound.a) The literature is split with regard to this hypothesis, with some studies finding that impulsiveness does play a significant role, and others finding that it does not, provided that some measure of overall intensity is held sufficiently constant[4].

A third acoustic consideration is the role of tonal components in helicopter noise. Noise certification regulations for helicopters are encoded in 14 CFR Part 36[5]. For light helicopters, A-weighted Sound Exposure Level (SEL-A) is the metric used for this purpose, but for large helicopters the Effective Perceived Noise Level (EPNL) metric is used, as it is for large fixed-wing aircraft. Since EPNL takes tonal content into consideration, its utilization suggests that tonality is important to helicopter noise annoyance.

Regulations such as 14 CFR Part 36 were formulated and have remained largely unchanged since the 1980s. Since that time, there has been a great deal of development on what are called "sound quality" (SQ) metrics. Computational tools have been developed to produce a quantitative measure of, for instance, the 'roughness' SQ metric of a sound, independent of its magnitude (which, as a quality, is commonly known as the 'loudness'). These metrics have found wide application in the design and manufacture of consumer products[6]. Recently, measures of SQ have been shown to be able to enhance the predictive capability of noise metrics for fixed-wing aircraft[7,8]. McMullen and Davies[3] found that annoyance responses to helicopter flyover recordings were highly correlated with loudness in addition to EPNL and SEL-A but did not sufficiently explore correlations between annoyance and other SQ metrics. Metrics that may be pertinent to rotorcraft noise beyond loudness are measures of: impulsiveness and tonality (clearly, given past work and regulation), as well as fluctuation strength, roughness, and sharpness.

This paper describes the first in what will be a series of psychoacoustic tests aimed at bringing the contemporary computational tools of synthesizing sounds and measures of SQ to bear on the problem of understanding annoyance to helicopter noise. It describes the formulation of a test that makes use of NASA psychoacoustic facilities, employs various SQ metrics, and uses synthesized helicopter-like sounds in order to attempt to answer the following questions:

---

a) This difference is independent of secondary sources that may be induced by an impulsive sound, such as those caused by rattle and structural vibration in the interior of a home.

1. Which SQ metrics significantly correlate with the annoyance response of human subjects?
2. Is it possible to order, or rank the relative importance of these metrics?

2 TEST DESCRIPTION

This section will discuss generating helicopter-like sounds for "The Rotorcraft Sound Quality Metric" (RoQM-I-2017) psychoacoustic test, followed by a description of how the test was conducted. All SQ metrics for test sounds were calculated using the ArtemiS Suite from HEAD Acoustics. The same methods used in Rizzi, et al[8] were also used to calculate loudness, sharpness, tonality, roughness, and fluctuation strength. Additionally, impulsiveness was computed in the ArtemiS Suite using a hearing model.

2.1 Signal Generation

Recordings of different rotorcraft flyovers are available from recent NASA acoustic flight tests[9-11]. These recordings could be directly played to human test subjects in the Exterior Effects Room (EER) at NASA Langley Research Center[12], who would then rate their annoyance to the sounds. There are two main problems with this direct approach. First, the SQ values can change by large amounts over a flyover duration, making the determination of when in the flyover a subject's annoyance response was registered more challenging. Another issue is that sequences of sounds need to be presented to subjects that vary in one or two metrics while other metrics remain roughly constant. Finding such sequences of sounds from flyover recordings is difficult.

To solve the first problem, the sound generation method in this test began with a flyover recording of an AS350 helicopter[13]. Using methods from Greenwood and Schmitz[14], the blade passage signals of the main and tail rotors of the helicopter were extracted from the recording. Repeating the blade passage signals in time produced a constant periodic helicopter source noise with repeating SQ values as shown in Figure 1, with the sharpness SQ metric as an example. This sound with repeated blade passage signals is referred to as the "AS350 baseline." The blade passage frequency (BPF) of the main rotor is 19.2 Hz. The BPF of the tail rotor is 69.8 Hz. Snippets of this sound were played to subjects.

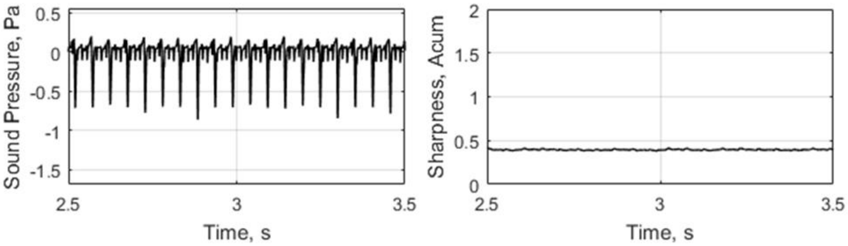

Figure 1. AS350 baseline time histories of sound pressure (left) and sharpness SQ metric (right)

To create sequences of constant helicopter-like sound snippets that vary in one or two metric values, one could manipulate the baseline signal to achieve a different SQ. Since it is important that the sounds be helicopter-like, the manipulations cannot be arbitrary. The approach taken was to focus on changing parameters that might correspond to helicopter design and operational characteristics to affect the SQ. An example was changing the BPFs of the main and tail rotors while maintaining the same amplitude and phase in the resulting harmonics of the respective signals. Twenty different parameter perturbations methods were developed, and subsets of these methods were each devoted to changing one of the SQ metrics while leaving others constant.

To leave room for discussing test results, this paper will not exhaustively cover parameter perturbation methods for test sound generation. Instead, we provide a brief overview using the subset of methods devoted to primarily affecting tonality as an example, and briefly touch upon how sounds were generated to affect other SQ metrics.

2.1.1 Signal Generation Example

Figure 2 shows how perturbing certain helicopter parameters can primarily affect tonality. Each trace color represents a different helicopter parameter that is being changed. Each point, or plot marker, along the change index axis is the metric 5% exceedance level value obtained from a sound generated by changing the helicopter parameter. The exceedance level is the value of the metric which is exceeded x% within a sound duration. The set of green traces, for example, are formed by increasing the magnitudes of the tail rotor harmonics from 0 Pa (meaning no tail rotor exists) at change index 1 to 3.2 times (5 dB above) those of the AS350 baseline at change index 5. As another example, the set of black traces increase the tail rotor BPF from 31 Hz at change index 1 to 200 Hz at change index 5. These perturbations cause an increase in tonality. The harmonic magnitude and tail rotor BPF values were chosen to produce an approximately equivalent interval in tonality for each change index.

The magenta colored dash-dot lines in Figure 2 are metric limits that were determined by examination of actual rotorcraft flyovers. They were obtained by computing the SQ metrics of 172 flyover recordings of the Bell 206, Bo 105, MD 520N, MD 902, Mi 8, and XV-15[9-11] and determining the minimum and maximum SQ metric values over all flyover sounds. The XV-15 aircraft is a tilt-rotor. The MD 520N and MD 902 use a NOTAR system having a fan inside their tail booms to counter the main rotor torque instead of a conventional tail rotor. Test sounds were generated so that the metric values stayed within these limits. Limits for loudness are absent in Figure 2, and an explanation follows.

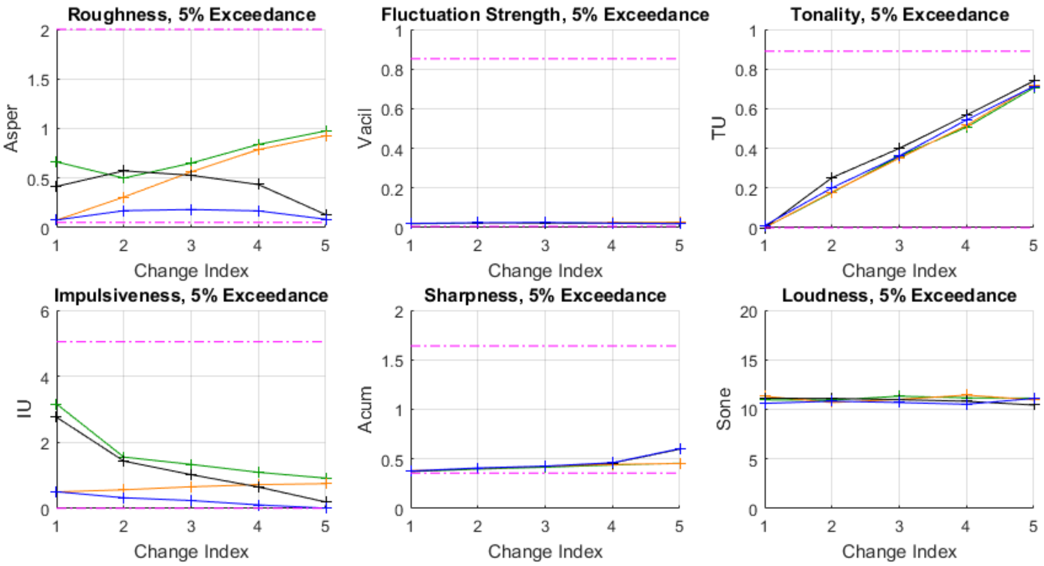

Figure 2. SQ metric values of sounds derived from AS350 baseline; Green trace: changing magnitudes of tail rotor harmonics from 0 Pa to 3.2 times baseline magnitudes; Yellow trace: changing magnitudes of tail rotor harmonics with 200 Hz BPF from 0 Pa to 3.2 times baseline magnitudes, main rotor harmonics have random phase; Black trace: changing tail rotor BPF from 31 Hz to 200 Hz; Blue trace: changing tail rotor BPF from 31 Hz to 200 Hz, main and tail rotor harmonics have random phase; Magenta dash-dot lines: SQ limits for helicopter-like sounds obtained from flyover recordings.

The perturbation methods in Figure 2 were chosen because they were considered to mainly affect tonality while having a small or negligible effect on other metrics. While this consideration proved mostly true, some strong dependencies between metrics could not be completely eliminated. For example, there is a non-negligible effect on impulsiveness for two of the perturbation methods represented by the green and black traces in Figure 2. The requirement that sounds remain helicopter-like is one reason why dependencies between metrics persist for parameter perturbation methods.

2.1.2 Identical Loudness for all Sounds

Based on results of previous research, for example McMullen and Davies[3] and More[7], the loudness metric is assumed to dominate subject response to test sounds. The predictive abilities of the other metrics could potentially be more difficult to separate if sounds varied considerably in loudness. Therefore, the gain of all sounds generated for the psychoacoustic test were adjusted so that their loudness level was approximately 10 Sone. This loudness level corresponds to the sound of an automobile one might hear from several meters away. This level was selected through pilot testing the sound reproduction in the EER room. The loudness was adjusted as a post-processing operation after perturbing helicopter parameters to change the desired SQ metric.

The method to adjust the loudness used the ISO 532-2 stationary loudness definition[15]. However, the resulting loudness of test sounds was calculated using the DIN standard 45631/A1 time-varying loudness definition in ArtemiS[16]. As a consequence, the 5% exceedance levels of loudness for test sounds are only approximately constant at 10 Sone as seen in Figure 2. Before conducting the test, the slight loudness variations were considered acceptable, but they affected the subsequent analyses, as will be discussed in the Results and Discussion section.

2.1.3 Generating Sounds to Affect Other SQ Metrics

Other sets of helicopter parameter perturbation methods, not shown, focused on changing one of the other metrics. Fluctuation strength was varied by amplitude modulating the AS350 baseline with both sine functions and band-limited noise and changing modulation frequency and modulation index, which is the intensity of the modulation. The amplitude modulation can correspond, for example, to main and tail rotor BPF separation. Increasing the magnitude of the high frequency harmonics of the AS350 baseline or using a high frequency BPF affected sharpness. These perturbations may correspond with use of NOTAR systems.

The phases of main and tail rotor harmonics were adjusted to primarily affect impulsiveness, that may correspond to helicopter operational changes. Included among the perturbations was a highly impulsive sound from a helicopter experiencing blade-vortex interaction (BVI). This sound was acquired from a recording of an AS350 helicopter main rotor with BVI and is referred to as the “BVI baseline.” The relative phases of the BVI signal harmonics were perturbed over the change indices from the BVI baseline to being closer to random in order to produce a range of impulsiveness.

For the helicopter-like sounds in this test, a set of perturbation methods that changed either impulsiveness or roughness while keeping the other one constant could not be found. These metrics were strongly dependent on each other for the test sounds considered. One reason for the strong dependence is that harmonic frequency separations produced modulations in the sound. The modulation frequencies are between 20 Hz and 300 Hz, where the roughness of sounds is strongly affected. Therefore, for all sounds in this test, the roughness metric was allowed to freely vary with other metric changes.

Sounds that combined different parameter perturbation methods were also created for the test. The methods that best kept other metrics roughly constant while changing either tonality, fluctuation strength, sharpness, or impulsiveness were determined. The corresponding parameter changes were then applied concurrently to create sounds that changed two or more of the metrics along with the change index. For example, changing harmonic phase and high frequency harmonic magnitude concurrently changed impulsiveness and sharpness together. Only a subset of these sounds was played to subjects.

The sound generation just discussed produced 105 unique helicopter-like sounds that served as stimuli to human test subjects in the EER. In addition to the AS350 and BVI baselines, 29 sounds were generated to primarily affect fluctuation strength, 12 sounds were generated to primarily affect impulsiveness, 17 sounds were generated to primarily affect tonality, and 23 sounds were generated to primarily affect sharpness. An additional 22 sounds were generated that combined parameter perturbation methods as described above.

2.2 Psychoacoustic Test Description

The RoQM-I-2017 psychoacoustic test took place in December 2017 in the EER. The test was executed in accordance with a protocol approved by the NASA Langley Institutional Review Board.

Forty subjects participated in the RoQM test: 10 groups of 4 subjects each. Subjects were recruited from the surrounding community. They ranged in age from 18 to 69, with mean and median ages of 35.3 and 31 years respectively. An even split of 20 male and 20 female subjects participated in the test. All subjects were screened for signs of pathological hearing loss (hearing loss ≤ 30 dB re: audiometric 0, 250 Hz – 4 kHz, was acceptable).

All 40 subjects listened to all 105 sounds that were prepared via the methods described in the previous section. The sounds samples were each 6 seconds long: 5 seconds at the intended loudness with 0.5 second cosine tapers on either end. The order of the sounds was shuffled between groups of subjects using both Latin Square and random layers in order to try to prevent sequential contraction biases in the responses.

The EER uses a vector-based amplitude panning scheme between its 31 speakers in order to produce sound that can appear to come from anywhere around the room, and that virtual sound position can move with time. For this test, the sounds were reproduced so that they were stationary. They were played back from a single location, relatively overhead to the subjects, that was determined to minimize differences in the 1/3-octave band sound pressure level between the four subject locations. The absolute level of the signal playback was calibrated via a sound level meter placed between the subject locations.

After each sound was played, subjects were presented with a scale on a tablet PC on which to rate their annoyance of the sound. The scale was based on the recommendations by Fields[17], and contained five ticks delineating amounts of annoyance with the labels “Not At All Annoying,” “Slightly Annoying,” “Moderately Annoying,” “Very Annoying,” and “Extremely Annoying.” The scale was continuous (i.e., subjects could respond between labels), and allowed subjects to respond beyond the extreme labels in order to try and avoid sequential contraction biases. The response was coded to numerical values spanning the range of 1 to 11, with the five even numbers corresponding to the labels.

The test was broken up into 4 sessions, each lasting between 5 and 10 minutes. Before the sessions were administered, the subjects were allowed to listen to a suite of 10 sounds from the test in order to familiarize themselves with the range of sounds that would be heard. They were also given a practice session of 26 sounds and were given instructions on the response task. In total, each subject’s participation (including pre- and post-test audiograms) required approximately an hour.

3 RESULTS AND DISCUSSION

It is expected that multiple metrics could serve as indicators of annoyance. As a first step, a multiple linear regression was performed on the responses to all 105 unique helicopter-like sounds acquired in the test. Before the regression was performed, the 5% exceedance values of SQ metrics, including loudness, were normalized so that the minimum 5% exceedance value of a metric over all sounds was zero, and its maximum 5% exceedance value over all sounds was one. The annoyance response data were not normalized and remained between values of 1 and 11. Since roughness is highly dependent with impulsiveness, it was not included in the regression.

Multiple linear regression between mean annoyance of the test subjects and combinations of the remaining five metrics produced r^2 values that are given in Table 1. The r^2 values measure the percentage of variation in annoyance explained by the regression. The coefficients of the linear equation for each combination relating the metrics to mean annoyance are not given in the table. They range between 0.5 and 1.9 and are all positive. The 5% exceedance values of loudness, sharpness, tonality, fluctuation strength, and impulsiveness are denoted N5, S5, T5, F5, and I5, respectively.

For combination 1 at the top of Table 1, a linear regression between mean annoyance and all five metrics gave an r^2 value of 0.57. Reasons why the r^2 values are not closer to unity will be discussed shortly. 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 1. 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 0.35, which is the difference between 0.53 and 0.18 when removing S5 from combination 2, or between 0.38 and 0.03 when removing N5 from combination 3. The highlighted entries in Table 1 indicate higher Δr^2 values (≥ 0.17).

Table 1. The r^2 values of a multiple linear regression with mean annoyance and different metric combinations and the reduction in r^2, Δr^2, with a regression after removing a metric from a combination.

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	0.57	0.04	**0.19**	**0.18**	0.01	0.04
2	S5, T5, F5, I5	0.53	-	**0.18**	**0.17**	**0.18**	0.03
3	N5, T5, F5, I5	0.38	0.03	-	**0.26**	0.01	0.02
4	N5, S5, F5, I5	0.39	0.03	**0.27**	-	0.02	0.02
5	N5, S5, T5, I5	0.56	**0.21**	**0.19**	**0.19**	-	0.04
6	N5, S5, T5, F5	0.53	0.03	**0.18**	**0.17**	0.02	-

3.1 Sharpness and Tonality as Indicators of Annoyance

The first result from Table 1 is that the multiple linear regression points to sharpness and tonality both being important to describing annoyance towards the helicopter-like sounds. Removing sharpness or tonality from all the metric combinations produces a relatively large reduction in the r^2 values.

3.2 Impulsiveness as an Indicator of Annoyance

The second result from Table 1 is that impulsiveness does not describe annoyance to helicopter-like sounds well. The reduction in the r^2 values is relatively small after impulsiveness is removed from all the metric combinations compared to the reduction seen when removing sharpness or tonality.

A cursory look might indicate this finding to be contrary to expectations since BVI noise is considered to be annoying and impulsive. As mentioned previously, the test included a BVI baseline accompanied by a series of sounds that diminished the impulsiveness of the baseline by adjusting the phase of harmonics. Table 2 lists these sounds in which Sound 1 (the BVI baseline) was transformed into Sound 4 by changing harmonic phase. The second column of Table 2 gives decreasing impulsiveness values from Sound 1 to Sound 4 before their gain was adjusted to 10 Sone loudness. The third column shows the loudness of these sounds are also decreasing. After adjusting the gain of the sounds to 10 Sone loudness, the fourth column shows the impulsiveness value for each sound does not change significantly from its value before the gain adjustment. Subject responses to these sounds after adjusting loudness to 10 Sone are given in the last column of Table 2. For reference, an annoyance rating value of 6 is "Moderately Annoyed" and an annoyance rating value of 8 is "Very Annoyed." Sound 1 with BVI noise is not rated more annoying than the less impulsive sounds, and the confidence intervals around the mean values overlap. We speculate that if the sounds were played to subjects before adjusting the gain to 10 Sone loudness, subjects would have rated the BVI baseline to be more annoying that the other sounds. Therefore, in this study, it seems that annoyance to BVI noise was not driven by impulsiveness. Also, as mentioned in the Introduction, impulsiveness can lead to secondary noise sources like rattle and vibration. In this way, impulsiveness may still be important to predicting annoyance in real world situations, but producing this relationship was beyond the scope of this test.

Table 2. Metric values for series of test sounds based on original BVI sound.

Sound Number	I5 (IU), before adjusting Loudness to 10 Sone	N5 (Sone), before adjusting Loudness to 10 Sone	I5 (IU), after adjusting Loudness to 10 Sone	Mean Annoyance Rating + (95% Confidence Interval) to 10 Sone Loudness
Sound 1, BVI baseline	3.4	72.7	3.0	6.8 + (-0.7, 0.6)
Sound 2	2.5	70.7	2.3	6.9 + (-0.7, 0.6)
Sound 3	1.9	69.2	1.7	6.9 + (-0.8, 0.7)
Sound 4, Least BVI-like	1.2	66.0	1.1	6.6 + (-0.6, 0.6)

3.3 Fluctuation Strength as an Indicator of Annoyance

The third result from Table 1 involves loudness and fluctuation strength. Removing these metrics individually from most of the combinations produces relatively small reductions in r^2 compared with removing sharpness or tonality individually. Since the loudness values of all sounds in the test were approximately 10 Sone, this result for loudness is expected. However, when fluctuation strength is removed from combination 2 or loudness is removed from combination 5, the reductions in the r^2 values are just as large as those when sharpness or tonality are individually removed. Notice that removing fluctuation strength or loudness from combinations 2 or 5, respectively, gives the same metric combination of sharpness, tonality, and impulsiveness. These combinations are the only ones that do not contain either fluctuation strength or loudness after removing a metric from one of the combinations 1 to 6 in Table 1. This result suggests that at least one of them (fluctuation strength or loudness) is needed to better explain the annoyance response.

Figure 3 helps explain why there is a dependency between these two metrics. Each different colored trace represents a different perturbation method of the AS350 baseline to primarily affect fluctuation strength. In these cases, the modulation index for different modulation frequencies was changed. Although the fluctuation strength of the generated sounds changes as desired, the loudness has more variation than desired. The loudness did not vary as much away from 10 Sone (as seen in Figure 3) when perturbations were applied to primarily affect tonality, sharpness, or impulsiveness (not shown). Also, fluctuation strength was held almost perfectly constant for perturbation methods targeted at the other metrics. The sounds in Figure 3 were the main ones available in the test to ascertain subject response to fluctuation strength.

Johnson, et al[18] showed that the just-noticeable difference (JND) of loudness around the 10 Sone level is approximately 1.5 dB, or approximately four Sone. Since test sounds produced loudness changes that were within the JND, it is reasonable to assume that the subjects were responding to changes in fluctuation strength and not the loudness variation. Without a way to incorporate JNDs into the analysis, the linear regression assumes subjects were responding to the loudness variation. With that in mind, the analysis also points to fluctuation strength being important to annoyance prediction.

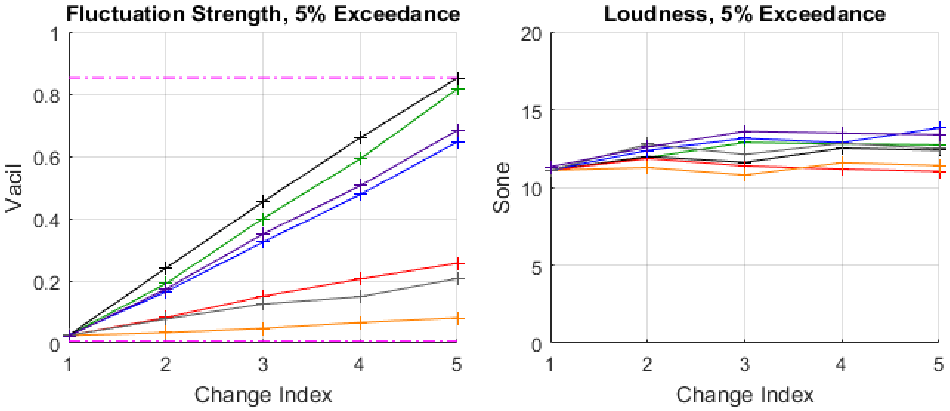

Figure 3. Fluctuation strength and loudness for subset of test sounds generated by perturbing parameters to primarily affect fluctuation strength.

3.4 Ranking of SQ metrics

The reductions in r^2 in Table 1 suggest a ranking of the SQ metrics. A ranking of most important to least important to describing annoyance is: sharpness, tonality, fluctuation strength, and impulsiveness. Loudness is not included in the ranking since the test was not designed to test response to this metric, but it is assumed to be more important than sharpness. For the 22 test sounds that combined helicopter parameter perturbation methods, firm conclusions have not yet been made from their separate analyses.

3.5 Reasons for Low r^2

One reason for the low r^2 values in Table 1 is that the relationship between the metrics and annoyance appear to be nonlinear. In Figure 4, the response to a subset of test sounds appears to be nonlinear. There seems to be more sensitivity to higher values of tonality.

Responses towards different methods of perturbing helicopter parameters also contribute to low r^2 values. As an example, Figure 5, shows the responses for sounds generated by changing modulation index using two different perturbation methods to affect fluctuation. One perturbation method modulates the AS350 baseline with a 4 Hz sine wave, and the other modulates the baseline with 8 Hz band-limited noise. Despite producing lower fluctuation strength, the 8 Hz noise modulation elicits approximately the same annoyance response as the 4 Hz sine wave modulation. Either lower fluctuation strength values need to be generated with the 4 Hz modulation, or other factors besides fluctuation strength are influencing the responses.

Due to the nonlinearities and responses for different perturbation methods, other types of analyses will need to be performed on the data going forward to determine the significance level of each metric's predictive ability and possibly establish a predictive relationship. Examples are non-linear regression or transforming metric values so linear regression becomes more applicable. Another approach is to calculate information theoretic annoyance uncertainty reduction given sound quality metric values.

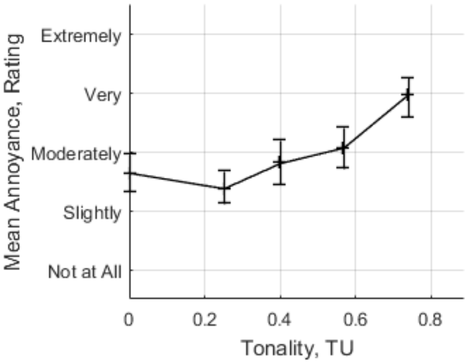

Figure 4. Response to subset of test sounds that changed in tonality by changing tail rotor BPF from 30 Hz to 200 Hz.

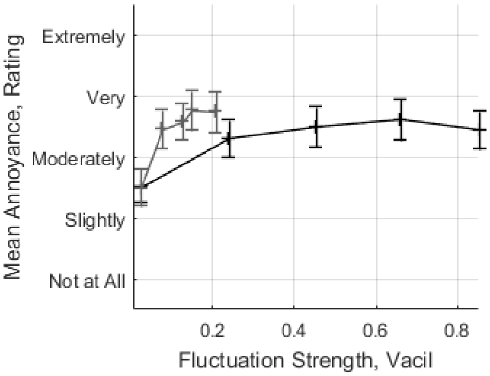

Figure 5. Response to subset of test sounds generated by modulating baseline by 4 Hz sine wave (black trace) and 8 Hz band-limited noise (gray trace).

4 CONCLUSIONS AND FUTURE WORK

The RoQM-I-2017 psychoacoustic test captured the annoyance response to 105 unique helicopter-like sounds of varying SQ. Test stimuli were generated from a periodic AS350 helicopter baseline sound. An AS350 experiencing BVI was also used to generate a set of sounds to test response to BVI. Different sounds were generated by perturbing helicopter parameters like the blade passage frequency and harmonic magnitudes. The perturbation methods were intended to keep the sounds helicopter-like and vary a single SQ metric while leaving others roughly constant. The gains of all sounds in the test were adjusted to have the same loudness (approximately 10 Sone) so that a variation in loudness would not dominate the subject response over the other SQ metrics.

Multiple regression analyses pointed to the sharpness, tonality, and fluctuation strength metrics being important to describing annoyance towards helicopter-like sounds in addition to loudness. Although test sounds were set approximately to 10 Sone, slight variations in loudness remained. This variation caused a dependence between fluctuation strength and loudness for the test sounds. After considering that the loudness of sounds was within a just-noticeable difference, it was hypothesized that subjects were responding to fluctuation strength variation and not to loudness variation. Impulsiveness was not found to directly reflect annoyance to helicopter-like sounds well. Since helicopter BVI is considered both annoying and impulsive, we speculate that annoyance to BVI noise is driven more by loudness than impulsiveness.

The linear regression analysis is not sufficient by itself. Reasons for the shortcoming are nonlinearities in the data and separate responses to different helicopter parameter perturbation methods. Other analyses and types of regression will need to be performed on the data to augment the linear regression. Further analysis will also determine if the SQ metrics are better indicators of annoyance for this data than other metrics used for certification such as Tone-corrected Perceived Noise Level used in EPNL and dB_A used in SEL-A.

The RoQM-I-2017 psychoacoustic test highlights a need for further testing. Using test sounds that more precisely change certain SQ metrics while leaving others constant may provide a clearer view of the relationship between annoyance and the metrics.

5 ACKNOWLEDGEMENTS

The authors would like to thank Matthew T. Hayes, Menachem Rafaelof (National Institute of Aerospace), and Brian C. Tuttle (Analytical Mechanics Associates, Inc.) at NASA Langley Research Center for assisting in the RoQM-I-2017 psychoacoustic test. We would also like to thank Regina D. Johns and Erin E. Thomas (Analytical Mechanics Associates, Inc.) at NASA Langley for recruiting and evaluating test subjects.

This research was supported by the National Aeronautics and Space Administration, Aeronautics Research Mission Directorate, NASA Advanced Air Vehicles Program, Revolutionary Vertical Lift Technology Project.

6 REFERENCES

1. Basner, M., Clark, C., Hansell, A., Hileman, J., Janssen, S., Shepherd, K., and Sparrow, V., "Aviation Noise Impacts: State of the Science," *Noise and Health*, Vol. 19, (87), 2017, pp. 41–50

2. Federal Aviation Administration (2004) Report to Congress: Nonmilitary Helicopter Urban Noise Study. Report of the Federal Aviation Administration to the United States Congress Pursuant to Section 747 of the Wendell H. Ford Aviation Investment and Reform Act for the 21st Century (AIR-21), Washington, D.C.

3. A. McMullen and P. Davies, "Results of a test using simulated rotorcraft flyovers designed to compare ratings of rotorcraft noise", Proceedings of Noise-Con 2014, Fort Lauderdale, FL, September 8-10, 2014

4. Mestre, V., et al., "Assessing Community Annoyance of Helicopter Noise," Report of the Airport Cooperative Research Program #181, The National Academies Press, Washington, D.C., 2017.

5. Noise standards: aircraft type and airworthiness certification, CFR Title 14, Chapter 1 - Federal Aviation Administration, Department of Transportation, Part 36.

6. Lyon RH. Product sound quality – from perception to design. Sound and Vibration. 2003;37(3):18-23.

7. More, S.R., *Aircraft Noise Characteristics and Metrics*, Ph.D thesis, Purdue University, West Lafayette, IN, July 2011.

8. Rizzi, Stephen A., Casey L. Burley, and Russell H. Thomas "Auralization of NASA N+2 Aircraft Concepts from System Noise Predictions,"

9. Burley, C.; Smith, C.; and Conner, D.: Flight Acoustic Testing and Data Acquisition for the Rotorcraft Noise Model (RNM). American Helicopter Society 62nd Annual Forum, American Helicopter Society, Phoenix, AZ, May 2006.

10. Watts, M.; Conner, D.; and Smith, C.: Joint Eglin Acoustic Week III Data Report. Technical Report NASA/TM-2010-216206, National Aeronautics and Space Administration, March 2010.

11. Conner, D.; Marcolini, M.; Edwards, B.; and Brieger, J.: XV-15 Tiltrotor Low Noise Terminal Area Operations. American Helicopter Society 53rd Annual Forum, American Helicopter Society, Virginia Beach, VA, April 1997.

12. Faller II, K.J., Rizzi, S.A., and Aumann, A.R., "Acoustic performance of a real-time three-dimensional sound reproduction system," National Aerospace and Space Administration Technical Report, NASA TM-2013-218004, June 2013.

13. Watts, M.E., Greenwood, E., Sim, B., Stephenson, J.H., and Smith, C.D., "Helicopter acoustic flight test with altitude variation and maneuvers," NASA TM-2016-219354, 2016.

14. Greenwood, E. and Schmitz, F.H., "Separation of main and tail rotor noise from ground-based acoustic measurements," *AIAA Journal of Aircraft*, Vol. 51, No. 2, 2014, pp. 464-472.

15. "ISO 532-2:2017, Acoustics -- Methods for calculating loudness -- Part 2: Moore-Glasberg method." https://www.iso.org/standard/63077.html

16. "Calculation of loudness level and loudness from the sound spectrum – Zwicker method – Amendment 1: Calculation of the loudness of time variant sound," Deutsches Institut für Normung (DIN) 45631/A1, 2009.

17. Fields, J. M. et al., "Standardized General-Purpose Noise Reaction Questions for Community Noise Surveys: Research and a Recommendation," *Journal of Sound and Vibration*, Vol. 242, No. 4, 2001, pp. 641-679.

18. Johnson, Jennifer H., Christopher W. Turner, Jozef J. Zwislocki and Robert H. Margolis. "Just noticeable differences for intensity and their relation to loudness." *The Journal of the Acoustical Society of America*. Volume 93, Number 2, 1993, pp 983 – 991.