An annoyance model for urban air mobility vehicle noise in the presence of a masker

Tyler Tracy[1], Matthew Boucher, Andrew Christian, Siddhartha Krishnamurthy, Stephen Rizzi
NASA Langley Research Center,
Mail Stop 463, Hampton, VA, 23681-2199

Durand Begault
NASA Ames Research Center,
Moffett Field, CA,94035-1000

Kevin Shepherd
Distinguished Research Associate,
NASA Langley Research Center,
Mail Stop 463, Hampton, VA, 23681-2199

ABSTRACT

Proposed Urban Air Mobility (UAM) operations offer an alternative to road and rail traffic for local and regional movement of people and goods. To allow for large-scale adoption of UAM vertical takeoff and landing (VTOL) aircraft, it is critical to predict human annoyance response to the noise generated by these vehicles. This paper proposes a model that predicts an individual's perceived level of annoyance when presented with UAM VTOL aircraft noise in addition to a masking noise. The model predicts changes in annoyance based on subjective evaluation of detection, noticeability, and annoyance of noise in the presence of a masker. The application of this annoyance model will influence future work for a psychoacoustic annoyance model that incorporates UAM sound quality.

1. INTRODUCTION

Modern developments in vertical takeoff and landing (VTOL) vehicles have enabled the concepts of Advanced Air Mobility (AAM) and Urban Air Mobility (UAM) to address issues surrounding the movement of people and goods. These concepts would enable VTOL vehicles to fly within cities and between urban and suburban areas. As VTOL vehicle designs are refined for UAM operations, noise generated by their operations will be a pertinent factor impacting humans along

[1] tyler.d.tracy@nasa.gov

and in the vicinity of flight paths. Previous work has studied the noise generated by VTOL operations and the propagation of that noise to the ground (e.g., [1, 2]). These references provide insight into the nature of UAM noise, but the question of how humans respond to UAM VTOL noise remains an active topic of research.

Fastl and Zwicker [3] developed a widely adopted psychoacoustic annoyance model to predict human annoyance response to steady state noise associated with stationary machines. This model principally relates annoyance to loudness, with additional psychoacoustic parameters accounting for fluctuation strength, roughness, and sharpness. This model estimates the annoyance for a given sound and does not consider the soundscape. Further attempts to refine this model for specific target sounds have included spectral and temporal sound quality metrics such as impulsiveness [4] and tonality [5, 6]. These refinements were found to improve annoyance predictions for target sounds and were consequently applied toward UAM VTOL vehicle annoyance predictions by Boucher et al. [7, 8].

While these results provide insight into the estimated annoyance response for target UAM VTOL sounds in isolation, in practice UAM operations will occur within existing ambient noise in communities. We hypothesize that a correction factor could be formulated for annoyance prediction when UAM noise is masked by an ambient, evaluated at the instant the flyover occurs. This correction is referred to in this paper as a “discount” to account for the masking effect caused by the ambient.

A psychoacoustic test titled DNA-2023 was recently completed in the Exterior Effects Room at NASA Langley Research Center to investigate whether such a discount effect could be quantified in a laboratory condition [9]. The test generated data to determine the effect of this discount on annoyance by looking at detection, noticeability, and annoyance. Analyses of the data show that the addition of a masking noise can shift the onset of annoyance. This work seeks to fit those data to a predictive model of discounting that may then be applied as an augmentation to the kinds of psychoacoustic annoyance models described above. Such a model might then improve the prediction of an individual’s annoyance response to UAM VTOL noise in the presence of an ambient background noise.

2. MASKING DISCOUNT HYPOTHESIS

The basis of a discounting correction to predict annoyance in the presence of masking has been developed in recent years [10, 11]. A chapter in a NATO report was produced that covers the literature and past work that supports these formulations and discusses the application of masking to UAM noise [12]. Many texts exist that cover auditory masking as a general topic (e.g., Moore [13]). Briefly, the discounting concept can be derived from two ideas gleaned from past literature:

1. When a sound is prominent over an ambient so that it is clearly audible, the ambient does not strongly affect the annoyance response to the sounds.
2. When a sound source is at a low enough level to be at least partially masked by ambient sounds, a reduction in annoyance to the sound source is observed.

These two statements constrain the possible effects that masking can have on a prediction of annoyance. Consider the y = x dashed black lines in Figures 1 and 2. These lines represent a linear regression between a frequency-weighted decibel measure of a noise and a prediction of annoyance that has a decibel-like scale (that is, a base-10 logarithmic relationship to a pressure-like unit). This prediction would be expected to correlate linearly with the kinds of scales used in psychoacoustic annoyance tests (e.g., a 0-10 numerical rating scale) [14].

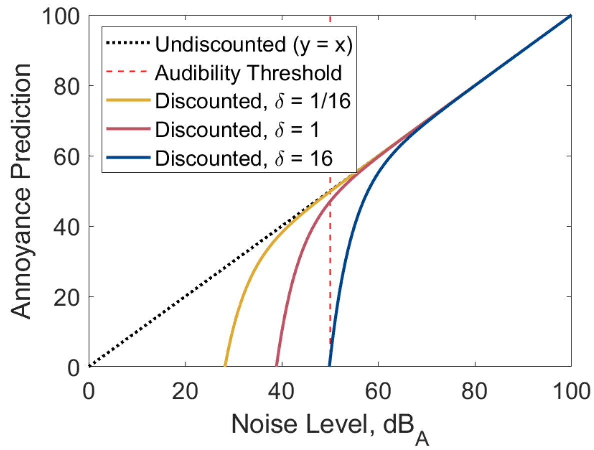

Figure 1: Noise Level vs Annoyance Prediction, varying δ.

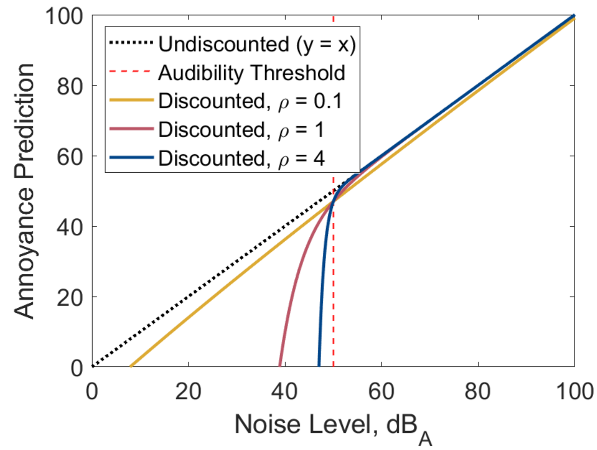

Figure 2: Noise Level vs Annoyance Prediction, varying ρ.

If an ambient is introduced but the target sound is at a high level relative to it, then, by point 1 above, there should be no effect on predicted annoyance to the target sound. As the level of the target sound is reduced, we expect the predicted annoyance to reduce linearly, based on the y = x relationship. However, as the target sound level is reduced such that it becomes masked by an ambient, the annoyance due to that sound will begin to decrease more than would be predicted by the reduction of the target sound level alone.

What is unknown is the functional relationship between annoyance and sound level below that masking point. A form for this change in annoyance prediction that is flexible to a wide range of possible behaviors was proposed by Christian [11]:

$$w_D = -\frac{\alpha}{\left(\frac{d'}{\delta}\right)^{\rho}} \tag{1}$$

The value w_D can be added to a decibel-like metric (e.g. dB(A), Zwicker phons, etc.) of the target sound to form a discounted version of that metric. Application of w_D within a psychoacoustic annoyance model is discussed in Section 6 below.

In Equation 1, d' comes from signal detection theory and is the unitless detectability index for a target sound in the presence of a masker. This is a value that ranges from $d' = 0$ for a sound that is undetectable, to $d' = +\infty$ when there is no masking noise, and the target sound can always be heard. For a given signal level and masker level, the value of d' can be different for each test subject. When $d' = 1$, a subject has approximately a 50% chance of hearing the target sound in the presence of the masker. In Figures 1 and 2, the red dashed trace is the signal level and masker level combination that give $d' = 1$ for a test subject, which is called the audibility threshold. Generally, a 3 dB change in the level of the target sound, or the masker, will change d' by a factor of two. For more information on signal detection theory in psychoacoustics, see the basic introduction in Moore's textbook [13], or the more thorough treatments given in either Macmillan and Creelman [15] or Green and Swets [16].

The parameters α, δ and ρ in Equation 1 govern how the discount impacts the prediction of annoyance as d' gets smaller and approaches 0. Without loss of generality, α is set to 3 dB for the current work, which is the value where the discount defines a reduction in the decibel-like measure of the target sound by a factor of two. The value δ, which is unitless, then defines the d' value where this 3 dB "knee" will be encountered. The parameter ρ, which is also unitless, defines the discount rate, or the rate of decay as d' gets smaller. Values for δ and ρ can therefore be manipulated to produce nearly any possible decay of the discount function in line with the two concepts above.

Figure 1 shows how the function changes with the parameter, δ. When δ increases, there is a drop in annoyance at higher target sound levels. This would indicate that the onset of annoyance to the target sound was at a level above the point where the target sound could be heard, indicating the sound needs to intrude upon the subject's attention before affecting annoyance. On the other hand, when δ is small (near 1 and below), the knee in the discount curve is at or below the threshold of audibility and there really is no effective discount.

Figure 2 shows that the effect of changing ρ. As ρ increases, the discount function steepens. For values of ρ greater than 1, the annoyance to the sound vanishes as soon as the target sound becomes even partially masked. On the other end of the scale, when ρ is relatively small (such as 0.1 as shown in the figure), the discounted line approaches that of the y = x line, and there is effectively no discount for the subject.

3. PSYCHOACOUSTIC TEST

The DNA-2023 psychoacoustic test took place in October 2023 in the Exterior Effects Room (EER) at the NASA Langley Research Center [17]. The test was designed to measure human annoyance to sound in the presence of background noise. A test hypothesis was that when a background noise masks a sound, a measure of the annoyance response changes according to Equation 1. To test this hypothesis, annoyance responses were collected in the presence of varying levels of a masking sound. Five test subjects participated in the study. Hearing tests were performed before and after each subject's participation, and the test protocol was approved by the NASA Institutional Review Board.

3.1. Acoustic Stimuli

Two target sounds (A and B) and one masker (Figure 3) were used in the psychoacoustic test. Sound A (Figure 3, blue trace) is a harmonic tone complex with the lowest tone at 80Hz and peaks separated by 40Hz. The speakers in the EER crossover around 80 Hz, and the fundamental was omitted in order to avoid significant energy coming from the subwoofers. The magnitudes of the peaks decay from approximately 120 Hz to approximately 360Hz. The masker (Figure 3, yellow trace) is broadband noise designed so that the probability of detection of sound A in the presence of the masker is 50% ($d' = 1$) for all one-third octaves within sound A based on the Sneddon et al. [18] model for a test subject's d' value (a thorough description of this algorithm is given in Rizzi et al. [19]). Sound B (Figure 3, orange trace) is also shaped broadband noise, but shifted up in frequency so that it is not masked by the masker. Sound B peaks at around 1kHz. A high-frequency tone complex was originally tried for sound B but was not used in the actual test since the sound-quality offsets (Q_B, discussed below) of some subjects to this sound were unreasonably large.

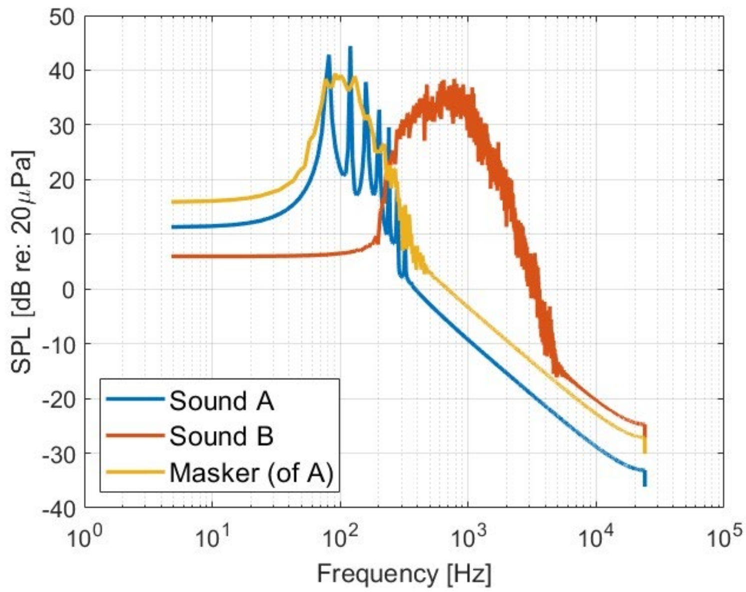

Figure 3: Acoustic stimuli used in the psychoacoustic test

Sounds A and B were synthesized based on representative tonal and broadband noise components created by a spinning rotor. As a result of the air volume displacement and air loading of a spinning rotor, thickness and loading noise occurs at the blade passage frequency and its harmonics [20]. Sound A thus represents this deterministic low frequency noise source of a UAM vehicle. On the other hand, blade self-noise is random and broadband in nature and dominates at higher frequencies [21]. Sound B can be thought of as representative of the self-noise component of rotorcraft noise. Actual UAM flight vehicles will have additional source noise components and added complexity due to maneuvers and unstable atmospheric conditions.

The target sounds were combined with the masker at numerous relative levels to create a robust dataset of test stimuli – the levels of all three sounds were manipulated throughout the test. Note that high levels of the masker would mask the tones in sound A. For lower levels of the masker, sound A was partially masked. Sound A was also played unmasked. When the masker was combined with sound B, the combined sound was still shaped broadband noise. The masker did not have a large effect on the audibility of sound B because masking a high frequency sound with a low frequency sound is quite difficult [3].

3.2. Measuring audibility

To estimate d' in Equation 1 for each test subject, the audibility of the target sounds in background noise was measured via a three-alternative forced choice (3AFC) method. For each trial, all three intervals contained the background masker noise and only one of the three intervals contained the target sound (either A or B). After hearing all three intervals, the following prompt appeared on computer tablets held by each test subject: “Which interval had the extra sound?”

Adaptive staircases were used with the 3AFC method in which the difference in sound pressure level of the target sound relative to the masker was varied [22]. In this type of testing, the level of the next 3AFC presentation is based upon whether the subject has gotten the last few responses correct. In this way, a performance ratio (of ‘hits’ and ‘misses’) that corresponds to a known d' can be targeted. An example of responses from one subject for sound A is shown in Figure 4. In this example, the apparent audibility threshold of sound A occurs when sound A is approximately 5 – 6 dB below the masker for this subject. The data from these two staircases are processed together to produce a single estimate of d' for each subject via the 'no pooling' method described in Boucher at al. [23].

3.3. Measuring annoyance

After measuring the audibility threshold for sounds A and B for each subject, test subjects compared sound A to sound B, with and without the masker present, and indicated which total sound was more annoying. This method is called "paired comparisons," as opposed to 3AFC. The paired comparisons were administered through adaptive staircases in which the masker level remained constant, and the relative level of the target sounds was adjusted to find an Equal Annoyance Point (EAP), which is used to determine subject's discount parameters. First, a baseline annoyance comparison was made between sounds A and B without the masker to measure the unmasked EAP. This process was then repeated with three levels of the masker, separated by 10 dB, added to both sound A and sound B to vary detectability, d'. Finally, the absolute levels of the target sounds were reduced by 10 dB and the paired comparison staircases were repeated for two levels of the background noise.

An example of the adaptive staircase method for annoyance paired comparisons is shown in Figure 5. Initially, one sound is much louder than the other such that both staircases move toward a relative level somewhere in between the extreme cases. However, once the relative level approaches the vicinity of subjective equality (i.e., the EAP), several reversals in the staircases appear, a reflection of the difficulty of the task. Unlike the 3AFC trials for audibility, there is no right or wrong answer for the annoyance comparisons.

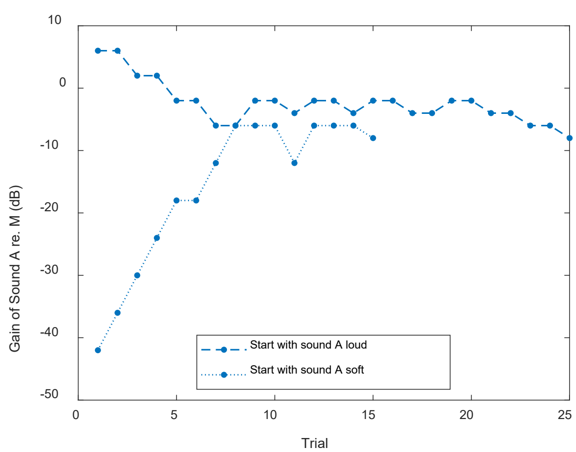

Figure 4: Example of 3AFC trials for measuring the audibility of sound A in the presence of the masker.

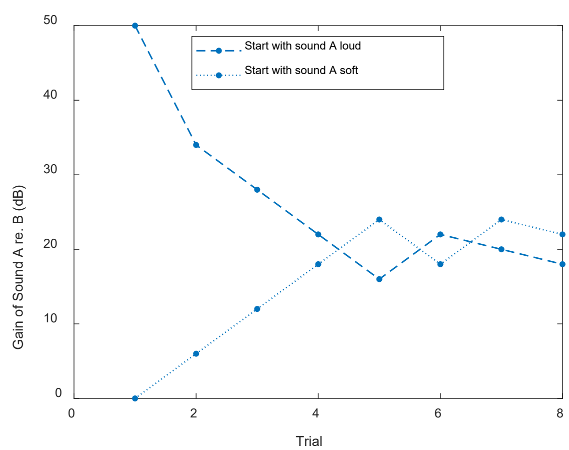

Figure 5: Example of paired comparison trials for measuring the annoyance of sounds A and B ($G_M = 15$ dB).

3.4. Measuring shift in Equal Annoyance Point due to masker

To calculate the shifts in the EAP from the unmasked to masked cases, the paired comparison annoyance responses were analyzed using the "fitglme" function in MATLAB R2024a, a random effects multilevel logistic regression with measured differences in gain and the staircase as the grouping variable. There were six EAP shifts using the staircases from combinations of the target sound levels (low or high) and the masker levels (not present, low, medium and high). An example is shown in Figure 6 using the raw responses from Figure 5. The black circles indicate the raw responses to the paired comparison trials with the masker present with a gain of $G_M = 35$ dB. Circles shown at probability = 0 indicate when sound B was judged to be more annoying than sound A. Circles at the top of the figure at probability = 1 indicate when the subject judged sound A to be more annoying than sound B. The dashed red line is the curve fit to raw responses without the masker present, shown as red X's. The solid black line is the curve fit to the black circles. The horizontal bar is the shift in EAP between unmasked and masked condition, corresponding to 10 dB for this example.

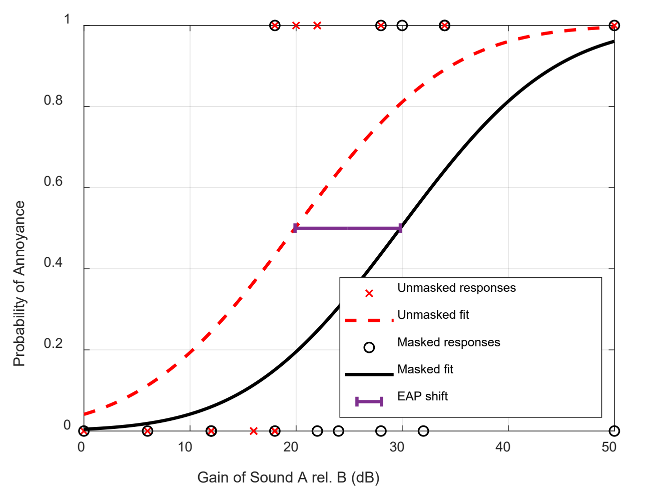

Figure 6: Measured shift in the Equal Annoyance Point between target sounds A and B when the masker ($G_M = 35$ dB) *is added to both target sounds (subject 2).*

4. FINDING MASKING DISCOUNT PARAMETERS

Based on the responses from the paired comparison test, we can formulate an equation describing the probability that a subject will be more annoyed by interval A than by interval B. This formulation starts with Equation 2, below, with an annoyance measure for target sound A (C_A) which takes the A-weighted decibel value of target sound A (dB_A), scaled by the interval gain for target sound A (G_A). A correction is added based on the subject's EAP between the unmasked sound A and the masker itself (Q_A), as measured in the paired comparison portion of the test. Then, we subtract the discount term of Equation 1 with $\alpha = 3$ dB and d' being a function of ΔG_{AM}, the difference in gains of the target sound A and the masker. The discount term, which is the rightmost term in Equation 2, is found from the EAP shift between an unmasked and masked sound A. The data used in these predictions is addressed in [9].

$$C_A = dB_A + G_A + Q_A - 3\left(\frac{d'(\Delta G_{AM})}{\delta}\right)^{-\rho} \tag{2}$$

Starting from a value of d' measured at $G_A = 0$ dB from the 3AFC portion of the test, further changes in ΔG_{AM} cause $d'(\Delta G_{AM})$ to adjust by $2^{\Delta G_{AM}/3}$. We similarly generate a C_B and C_M, a component annoyance measure for target sound B and the masker, respectively. For C_M, the value of Q_M is 0 dB and d' is a function of ΔG_{MA}. The audibility of sound B is mostly unaffected by the masker, and ΔG_{MA} is the reciprocal of ΔG_{AM}. The corresponding discount terms for C_B and C_M become negligible except for low values of sound B and the masker.

The component annoyance measures for target sound A and the masker are added logarithmically to create an annoyance prediction value interval A, $\mathrm{AP}(A)$, shown in Equation 3. A similar annoyance prediction value, $\mathrm{AP}(B)$, is generated for interval B.

$$\mathrm{AP}(A) = 10 \cdot \log_{10}\left[10^{(C_A/10)} + 10^{(C_M/10)}\right] \tag{3}$$

The probability that a given subject will select interval A as more annoying than interval B is given by $\Pr(A)$ in Equation 4. It is modeled using a normal cumulative distribution function with zero mean and a standard deviation of 2 dB. This function is evaluated at the difference in annoyance predictor values between interval A and interval B, i.e., $\mathrm{AP}(A) - \mathrm{AP}(B)$.

$$\Pr(A) = \frac{1}{2} \cdot \left[1 + \mathrm{erf}\left(\frac{\mathrm{AP}(A) - \mathrm{AP}(B)}{2 \cdot \sqrt{2}}\right)\right] \tag{4}$$

4.1. Psychometric Surfaces

At a fixed masking level, the value of $\Pr(A)$ will vary as the gains of target sounds A and B change. Through many calculations of $\Pr(A)$ at different gains of sound A and sound B, a psychometric surface plot can be generated showing G_A vs G_B vs $\Pr(A)$ to better understand how the subject's perceived annoyance changes as a function of gains of the target sounds. We can generate a subject's psychometric surface for each of the four masking levels used in the DNA test, two of which are shown in Figures 7 and 8.

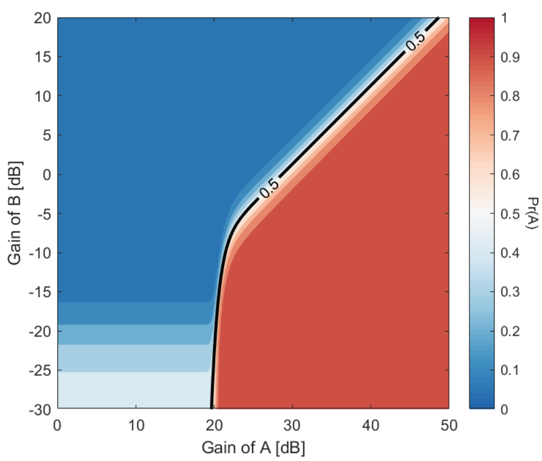

Figure 7: Trial psychometric surface Gain of A vs Gain of B vs Pr(A) (δ = 10 and ρ = 5) at a masker gain of 15 dB. At low values of masker gain, the subject is more likely to be annoyed by target sound A over target sound B.

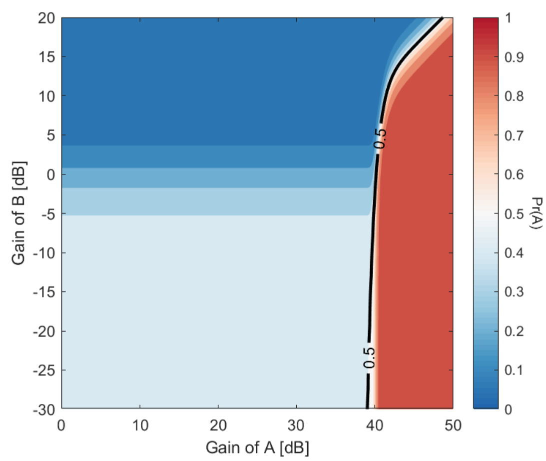

Figure 8: Trial psychometric surface Gain of A vs Gain of B vs Pr(A) (δ = 10 and ρ = 5) at a masker gain of 35 dB. At high values of masker gain, the subject is less likely to be annoyed by target sound A over target sound B.

These surfaces were generated with the gain of sound A ranging from 0 dB to 50 dB and the gain of sound B from -30 dB to 20 dB. The equal annoyance contour, where $\Pr(A) = 0.5$, i.e., the subject has a 50% chance of rating interval A as more annoying, is highlighted in black. Each psychometric surface is based on a specific value for δ ranging from 0.2 to 20 and a value for ρ ranging from 0.1 to 10. Psychometric surfaces with the extreme values of δ and ρ are shown in Figures 9 and Figure 10.

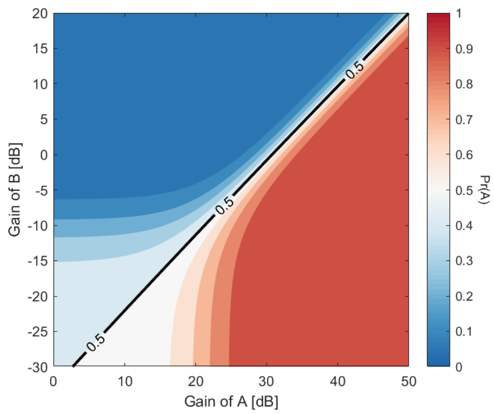

Figure 9: Trial psychometric surface (δ = 0.2 and ρ = 0.1) at a masker gain of 25 dB. These parameter values indicate the subject does not strongly discount annoyance of target sound A in the presence of a masker.

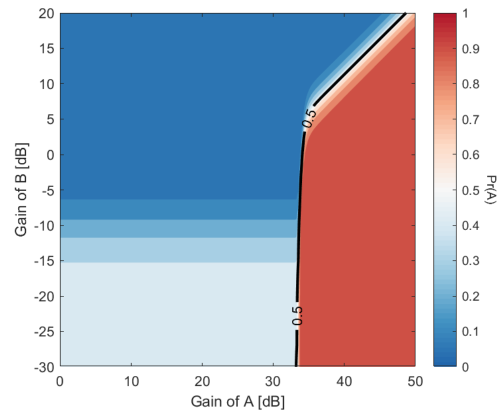

Figure 10: Trial psychometric surface (δ = 20 and ρ = 10) at a masker gain of 25 dB. These parameter values indicate the subject strongly discounts annoyance of target sound A in the presence of a masker.

4.2. Parameter Optimization

For each subject we seek values of the discount parameters δ and ρ that best match the psychometric surfaces generated by the annoyance predictor to the subject's EAPs from the DNA test. To do that, we create trial psychometric surfaces with varying values of δ and ρ to find best-fit parameter values matching the measured EAP data for each subject. As δ ranges from 0.2 to 20 and ρ ranges from 0.1 to 10, both parameters have three logarithmically sized steps for trial values less than 1 and seven linear steps for trial values greater than 1. After calculating a subject's psychometric surfaces for trial values of δ and ρ, we calculate the surfaces' error relative to the six measured EAPs using the cost function ε:

$$\varepsilon = \sum_{n=1}^{6} |\Delta \Pr(A)_n| \cdot \sqrt{\frac{1}{\frac{1}{(\Delta G_{An})^2} + \frac{1}{(\Delta G_{Bn})^2}}} \tag{5}$$

For a measured EAP n, $\Delta \Pr(A)_n$ represents the difference in probability between $\Pr(A) = 0.5$ and the value of $\Pr(A)$ computed from Equation 4 at the EAP n's gain coordinates on the trial psychometric surface. The term ΔG_{An} represents the difference in gain of A at the measured EAP n to the trial surface's equal annoyance contour (the line where $\Pr(A) = 0.5$). Similarly, ΔG_{Bn} represents the difference in gain of B at the measured EAP n to the trial surface's equal annoyance contour. This cost function was selected to account for cases where the values of $\Delta \Pr(A)_n$ and

ΔG_{An} are small compared to the value of ΔG_{Bn}, as may be the case with high trial values of ρ as shown in Figure 10.

For an individual subject, the error is summed across six measured EAPs, with each individual EAP n measured at various gains of the target sound and the masker. This sum produces a composite error ε representing how well the psychometric surfaces fit to the measured EAP shifts in the DNA test. A subject's composite error for a trial pair of δ and ρ values is compared to the composite errors of every other pair of trial parameter values, and the minimum composite error is selected corresponding to the best-fit parameter values of δ and ρ. Figure 11 shows an optimized psychometric surface with a minimized composite error at a masker level of 25 dB.

The two EAPs measured at this masker level are shown as yellow X's, and their projection onto the psychometric surface are shown as black circles. One yellow X is for high gain values of sounds A and B, and the other yellow X is for low gain value of the sounds. The differences between measured and predicted EAPs, ΔG_{An}, ΔG_{Bn}, and $\Delta \mathrm{Pr}(A)_n$ are shown as dashed grey lines connecting the yellow X's to the black circles. Note that the composite error sums measured vs predicted errors for each EAP across all levels of the masker. Therefore, some measured EAPs match the prediction very well while others are slightly offset.

To better understand how variations in δ and ρ affect the composite error for an individual subject, a heat map of the parameter space showing trial values for pairs of δ and ρ versus composite error ε is shown in Figure 12. The chosen parameter search space explained in Section 4.2 is based on representative parameter values as described in Section 2. Note the error interpolation between trial values of δ and ρ was done using the "scatteredInterpolant" function with a natural interpolation method in MATLAB R2024a, resulting in the surface bumps seen between trial parameter values. The minimum composite error found through this method is local, and not guaranteed to be global. More sophisticated minimalization algorithms could be employed to find values of δ and ρ that correspond to a global minimum.

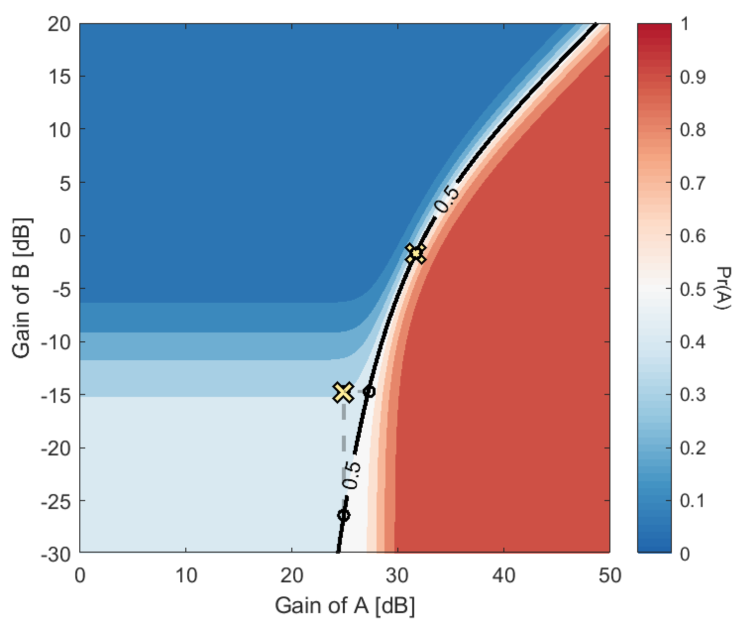

Figure 11: Measured Equal Annoyance Points (yellow X's) and their projections onto the optimized psychometric surface (black circles). The underlying psychometric surface Gain of A vs Gain of B vs Pr(A) has parameter values δ = 17 and ρ = *1, with a masker gain of 25 dB.*

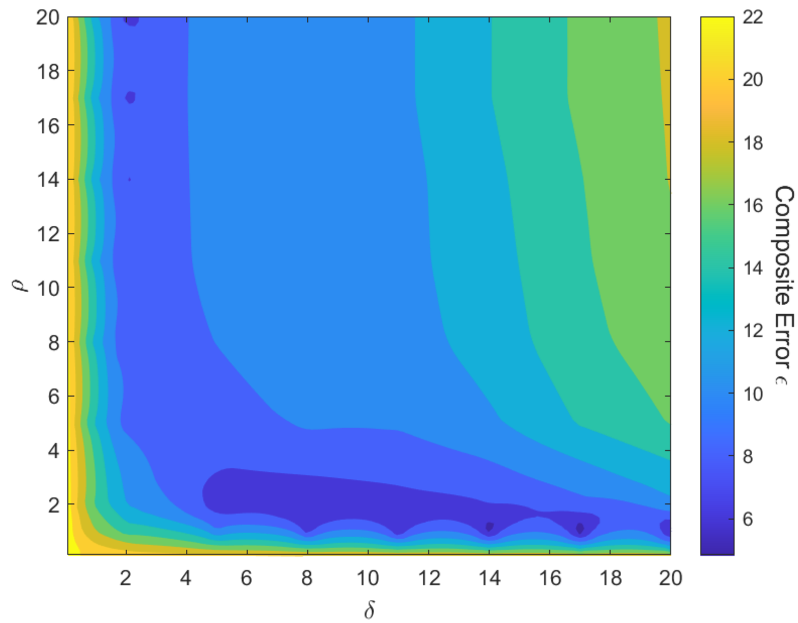

Figure 12: An individual subject's plot of δ versus ρ versus composite error. Minimizing the composite error cost function produces values of δ and *ρ that best fit a subject's measured equal annoyance points.*

5. RESULTS

Test subjects were categorized into one of three groups based on how they responded to a masked sound: 1. With a discounted annoyance when a masker is present, 2. Without a discounted annoyance but only if the sound is audible above the masker, and 3. Without a discounted annoyance until the sound level is well below the masker level. Table 1 shows the optimized parameter values for each subject in the DNA test. Figure 13 illustrates the corresponding annoyance versus noise level curves for the subjects, similar to curves shown in Figures 1 and 2. The Annoyance Prediction values on the y-axis define a one-to-one relationship with Noise Level in dB_A for the undiscounted case.

Table 1: Optimized parameter values

Subject	δ	ρ
1	14	2
2	2	6
3	20	0.1
4	1	0.1
5	17	1

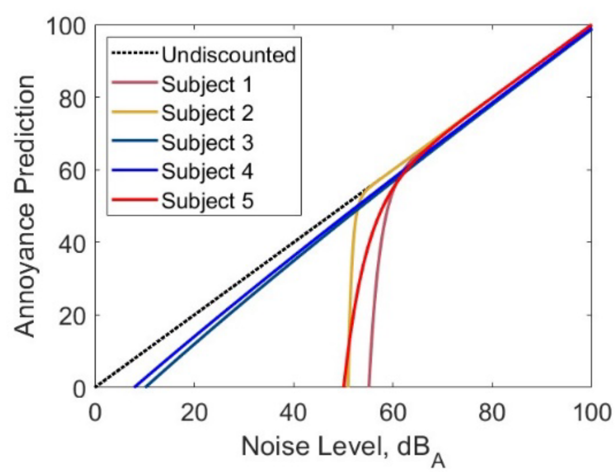

Figure 13: Noise Level vs Annoyance Prediction, for optimized parameter values per subject.

Generally, values of ρ much less than 1 indicate the subject does not have a strong masking discount, as is the case with Subjects 3 and 4. Expanding the range of δ and ρ in the optimization process is likely to produce the same result, as the low ρ value dominates the behavior of the discount curve. On the other hand, high values of δ combined with values of $\rho \geq 1$ indicate the subject's annoyance responses are sensitive to a masking discount, as is the case with Subjects 1 and 5. For Subject 2, a value of δ near 1 combined with a value of ρ much greater than 1 indicates that the discount term applies close to the threshold of audibility for sound A. This means that if sound A is at all audible, there is not much reduction in probability they will be more annoyed by a partially masked sound A compared to sound B. Conversely, subjects 1 and 5 with high values of δ may be less annoyed by a clearly audible, but partially masked, sound A compared to sound B.

6. CONCLUSIONS AND FUTURE WORK

The results of this analysis indicate that a masking discount does exist for some individuals and may be useful in describing the change in an individual's perceived annoyance of UAM vehicles due to the presence of ambient sound. The results also highlight the extreme variability of sound perception among individual subjects. The equal annoyance points, detection thresholds, and qualitative judgements have high variability across individuals (high variability with high certainty, in almost all cases). Based upon the variability seen in the values of these parameters, it is not surprising that the composite parameters of δ and ρ also have high variability. It also could be that the analyses used (thus far) are ineffective at discerning a discount for certain combinations of these intermediate values. The design of the test itself may complicate this analysis, as this is a first-ever attempt to determine values for these parameters, and the test is quite complex (involving multiple testing methods and multiple stages of analysis) relative to more fundamental psychoacoustic tests.

The main result of this first attempt to determine a discount function may simply be to indicate that the path ahead is quite complex, and it will be difficult to confidently quantify a discount effect that is applicable across subjects. The variability in behavior observed in this first study and analysis may be real or may be an artifact of the methods used. More work in the future, both with different testing modalities, and different analysis approaches, are needed to settle this impasse.

Beyond that, this research lays a foundation for more application-oriented work on a loudness-based annoyance metric for masked UAM sounds. By demonstrating that a masking discount exists in some subjects based on gain of target and masker signals, we can apply our findings to work by Boucher [8] in modeling Psychoacoustic Annoyance, which applies a sound quality adjustment to loudness, for UAM vehicles. Instead of finding an annoyance discount with the annoyance measure being the sound level or gain, an annoyance discount to a masker may be found with loudness as the annoyance measure. Combined with the work from Boucher [8], an adjustment to the loudness of UAM vehicles may be generated that accounts for both sound quality and the presence of a masker.

ACKNOWLEDGEMENTS

This work supports the NASA Revolutionary Vertical Lift Technology Project of the Advanced Air Vehicles Program. We also thank test participants and Erin Thomas of Analytical Mechanics Associates, Inc., for recruiting and screening participants.

REFERENCES

[1] S. Krishnamurthy, S. A. Rizzi, R. Cheng and D. D. Boyd, Jr., "Prediction-Based Auralization of a Multirotor Urban Air Mobility Vehicle," in *AIAA SciTech Forum, Paper 0587*, Virtual, 2021.

[2] J. Ko, Y. Kim, J. Jeong and S. Lee, "Prediction-based psychoacoustic analysis of multirotor noise under gusty wind conditions," *The Journal of the Acoustical Society of America,* vol. 154, no. 5, pp. 3004-3018, 2023.

[3] E. Zwicker and H. Fastl, Psychoacoustics: Facts and Models, 2nd ed., Springer, 2013.

[4] A. J. Torija, Z. Li and P. Chaitanya, "Psychoacoustic modelling of rotor noise," *The Journal of the Acoustical Society of America 151,* pp. 1804-1815, 2022.

[5] G.-Q. Di, X.-W. Chen, K. Song, B. Zhou and C.-M. Pei, "Improvement of Zwicker's psychoacoustic annoyance model aiming," *Applied Acoustics 105,* pp. 164-170, 2016.

[6] S. R. More, Aircraft Noise Characteristics and Metrics, West Lafayette, Indiana: Purdue University, 2010.

[7] M. Boucher, M. Rafaelof, D. Begault, A. Christian, S. Krishnamurthy and S. A. Rizzi, "A Psychoacoustic Test for Urban Air Mobility Vehicle Sound Quality," in *SAE Technical Paper Series*, 2023.

[8] M. Boucher, A. Christian, S. Krishnamurthy, T. Tracy, D. Begault, K. Shepherd and S. A. Rizzi, "Toward a psychoacoustic annoyance model for Urban Air Mobility vehicle noise," NASA/TM-20240003202, Hampton, VA, 2024.

[9] M. Boucher, A. Christian, S. Krishnamurthy, T. Tracy, K. Shepherd, D. Begault and S. A. Rizzi, "A psychoacoustic test on the effect of masking on annoyance to Urban Air Mobility vehicle noise," in *189th Meeting of the Acoustical Society of America*, Ottawa, 2024.

[10] A. Christian, "Towards the incorporation of auditory masking effects into assessments of comunity noise," in *Quiet Drones*, Doi: 10.13140/RG.2.2.16965.58088, Virtual, 2020.

[11] A. Christian, "A construction for the prediction of noise-induced annoyance in the presence of auditory masking," *The Journal of the Acoustical Society of America,* vol. 150, no. 4 Supplement, p. A261, 2021.

[12] A. Christian, "The effect of background noise on human response," in *Assessment and reduction of propeller and rotor nosie from unmanned aircraft systems*, NATO, STO-TR-AVT-314, 2023, p. 24.

[13] B. C. Moore, An introduction to the psychology of hearing, Brill, 2013.

[14] S. Krishnamurthy, S. Rizzi, R. Biziorek, J. Czech, J. Berg, D. Tannler, D. Bean, A. Ayrapetyan, A. Nguyen and J. Wivagg, "Remotely Administered Psychoacoustic Test for sUAS Noise to Gauge Feasibility of Remote UAM Noise Study," in *SAE Noise and Vibration Conference 2023, Paper 1106*, Grand Rapids, MI, 2023.

[15] N. A. Macmillan and C. D. Creelman, Detection Theory, Lawrence Erlbaum, p. 504.

[16] D. M. Green and J. A. Swets, Signal Detection Theory and Psychophysics, New York: John Wiley and Sons Inc., 1966.

[17] 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.

[18] M. Sneddon, K. Pearsons and S. Fidell, "Laboratory study of the noticeability and annoyance of low signal-to-noise ratio sounds," *Noise Control Eng. J.,* vol. 51, no. 5, pp. 300-305, 2003.

[19] S. A. Rizzi, A. W. Christian, S. Letica and S. Lympany, "Annoyance model assessments of Urban Air Mobility vehicle operations," in *30th AIAA/CEAS Aeroacoustics Conference*, Rome, 2024.

[20] H. H. Hubbard, "Aeroacoustics of Flight Vehicles: Theory and Practice. Volume 1: Noise Sources," 1991.

[21] T. Brooks, D. S. Pope and M. Marcolini, "Airfoil self-noise and prediction," in *NASA/RP-1218*, 1989.

[22] H. Levitt, "Transformed up-down methods in psychoacoustics," *J. Acoust. Soc. Am.,* vol. 49, p. 467–477, 1971.

[23] M. Boucher, A. W. Christian, M. Rafaelof, K. P. Shepherd, S. A. Rizzi and J. H. Stephenson, "Comparisons of analysis methods applied to alternative forced choice psychoacoustic data," NASA/TM-2023-0008547, Hampton, VA, 2023.

[24] E. Z. Hugo Fastl, Psyhoacoustics: Facts and Models, New York: Springer, 2006.

[25] J. Ko, J. Jeong, H. Cho and S. Lee, "Real-time prediction framework for frequency-modulated multirotor noise," *Physics of Fluids,* vol. 34, no. 2, p. 027103, 2022.

[26] M. Boucher, A. Christian, S. Krishnamurthy and S. A. Rizzi, "Perceptual evaluation of Sound Exposure Level in annoyance ratings to helicopter noise," *Journal of the American Helicopter Society,* https://doi.org/10.4050/JAHS.69.032006, 2024.

[27] M. Pérez-Ortiz, A. Mikhailiuk, E. Zerman and V. Hulusic, "From Pairwise Comparisons and Rating to a," *IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 29,* pp. 1139-1151, 2020.

[28] M. Boucher, M. Rafaelof, D. Begault, A. Christian, S. Krishnamurthy and S. Rizzi, "A Psychoacoustic Test for Urban Air Mobility Vehicle Sound Quality," *SAE International,* 2023.