TITLE OF INFORMATION COLLECTION: Remote Psychoacoustic Test, Phase 1, for Urban Air Mobility Vehicle Noise Human Response

B. COLLECTIONS OF INFORMATION EMPLOYING STATISTICAL METHODS

1. **Describe (including a numerical estimate) the potential respondent universe and any sampling or other respondent selection method to be used.**

A minimum of 80 test subjects are needed for the Phase 1 remote psychoacoustic test of the UAM Vehicle Noise Human Response Study. This minimum requirement was initially formed by the following needs:

1. To assess the performance of the remote test method, the Phase 1 remote test responses will be compared a previous in-person test [4]. The in-person test used 40 test subjects, although only the responses from 38 test subjects were analyzed due to two subjects not completing the test. To eliminate the number of test subjects as a potential source of uncertainty in response differences, a minimum of 40 test subjects is sought for the Phase 1 remote test.

2. Phase 1 remote test subjects will be divided into two groups, where one group will be provided a contextual cue when asked to respond to sounds and the other group will not. See items 1 and 2, Part A, for additional details. A potential outcome is that the two contextual cue groups will have significantly different responses. If that is the outcome, only the group of test subjects who are not provided a contextual cue can be compared to the previous in-person test responses, since the in-person test did not provide a contextual cue. Therefore, a minimum of 80 test subjects are needed for the Phase 1 remote test if test subjects from the group provided contextual cue have significantly different responses than the group of subjects who are not provided a contextual cue.

While the above two reasons are why a minimum of 80 test subjects were initially sought for the Phase 1 remote test, if these reasons are neglected, another reason comes from the ability to analyze test subject responses from two distinct geographic regions. If test subjects from the group provided contextual cue have significantly different responses than test subjects not provided contextual cue, then comparing the overall mean annoyance responses between test subjects in two distinct geographic regions may be challenging if all 80 test subjects are used for the comparison. Note that test subjects from the same geographic region will also be in different contextual cue groups. Differences between the two contextual cue groups may potentially obscure differences between test subjects in different geographic regions. Therefore, analysis of geographic region annoyance response differences will need to occur within a contextual cue

group, either the group provided contextual cue or the group of subjects not provided contextual cue.

Analysis of Variance (ANOVA) was performed on simulated test subjects to understand how small the number of test subjects from a geographic region can be to still allow for overall mean annoyance response difference to be found, if it exists, between subjects from different regions. The simulated test subjects were divided into two groups, representing test subjects from two different geographic regions. The two geographic regions are not yet specified, since they can only be determined after collected test subject US ZIP code. With 40 test subjects in either the group provided contextual cue or not provided contextual cue, a maximum of 20 test subjects can be in each geographic region. ANOVA was performed using geographic region test subject groups with three different numbers of test subjects: 5, 10, and 20 test subjects. Having 5 or 10 test subjects each from the two geographic regions instead of 20 subjects would mean there only 10 or 20 test subjects, respectively, in each contextual cue group, despite the concerns mentioned at the beginning of item 1, Part B, with these smaller numbers of test subjects. For each number of test subjects per geographic region (5, 10, or 20), test subjects for one geographic region were simulated to have the same responses as the real previous in-person test subject responses [4]. Therefore, the ANOVA analysis presented here assumes that Phase 1 remote test responses for one of the geographic region groups will be the same as the in-person test responses.

The test subjects in the other geographic region were simulated to have in-person test responses that were all shifted by a certain annoyance rating. ANOVA has the task of uncovering the annoyance rating shifts. The in-person annoyance rating scale was numerically between 1 and 11 annoyance ratings. An annoyance rating of "2" corresponds to a response of "Not at all Annoying," and a rating of "10" corresponds to a response of "Extremely Annoying." If shifting an in-person annoyance rating caused the rating to be less than 1 or greater than 11, the shifted rating was kept at 1 or 11 annoyance ratings, respectively. A separate ANOVA, which is now already parameterized by the number of test subjects per geographic region, was performed for each annoyance rating shift, which were -0.30, -0.35, 0.30, and 0.35 annoyance ratings. One group of 5, 10, or 20 test subjects were given these four rating shifts to their simulated responses, and the other group, representing the other geographic region, were not given these shifts. These shifts represent the minimum shifts that ANOVA can detect based on initial analyses that are not presented in this document. For example, it was difficult for ANOVA to detect annoyance rating shifts of -0.20 ratings even with 20 test subjects.

With simulated responses derived from in-person test responses for 38 test subjects, there are 501942, 472733756, 33578000610 different combinations of test subject responses for simulated test subjects in groups of 5, 10, and 20, respectively. Performing an ANOVA for each different combination would be computationally impractical. On the other hand, performing ANOVA on only one combination of responses to simulate test subjects leaves open the possibility that

another combination of in-person responses will give a different ANOVA result for the overall mean response difference between test subjects from different geographic regions. A compromise is to perform a separate ANOVA each for 1000 different combinations of in-person responses that simulated each number of test subjects (5, 10, or 20) per geographic region. From the 1000 combinations, a probability value can be found that 5, 10, or 20 test subjects per geographic region with certain simulated annoyance rating shift (-0.30, -0.35, 0.30, and 0.35) between the two regions will allow the underlying overall mean annoyance response rating shift to be detected by ANOVA between subjects from the two regions.

In summary, parametrized ANOVA was performed for 1. The number of test subjects per geographic region (5, 10, or 20), 2. The annoyance rating shift in simulated test subject responses for one of the geographic regions (-0.30, -0.35, 0.30, or 0.35 annoyance rating), and 3. Each of 1000 response combinations derived from the original in-person responses for 38 test subjects. Hence, 12,000 ANOVA were performed. One response combination was assigned to test subjects in one geographic region, and a shifted version of the responses was assigned for simulated test subjects in the other geographic region. The 1000 response combinations were chosen randomly, and no effort was made to choose combinations with the least dependency among each other. The percentage of the 1000 response combinations for each combination of number of test subjects (5, 10, or 20) and response shifts (-0.30, -0.35, 0.30, or 0.35) for which ANOVA could detect the mean annoyance response shift was determined to be the probability of ANOVA finding a significant overall mean annoyance response difference between test subjects from two distinct geographic regions.

Figure 1 gives the results of the parametrized ANOVA. The ANOVA used a significance level of 0.05 to test whether the overall mean responses of the two simulated test subject groups, representing different geographic regions, are the same. The simulation results in Figure 1 show that having 20 test subjects guarantees ANOVA to detect a mean annoyance rating difference of at least 0.35 annoyance ratings between groups of test subjects in two different geographic regions and a strong possibility that a shift of 0.30 ratings can be detected. While using 10 test subjects still gives a strong possibility of detecting a 0.35 annoyance rating difference, it is not a guarantee. Using 5 test subjects leaves at least a 30% chance that a difference of 0.35 ratings will not be detected. As a reminder, the results in Figure 1 assume Phase 1 remote test responses will be similar to in-person test responses. If higher variance exists in the Phase 1 remote test responses relative to the in-person test responses, the probabilities of detecting an annoyance rating difference become lower than what Figure 1 projects.
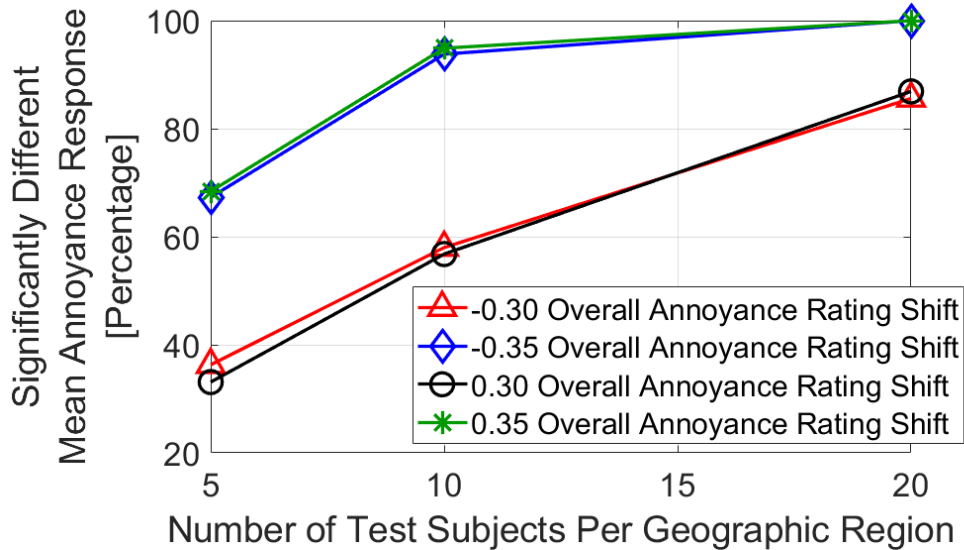
Figure 1. *ANOVA results comparing simulated responses from test subject in two different geographic regions.*

With the desire to detect as small of an annoyance rating difference between subjects in different geographic regions as possible, the results in Figure 1 indicating using 20 test subjects is preferrable, and the other concerns mentioned at the beginning of item 1, Part B, on why 80 test subjects should be used, a minimum of 80 test subjects are requested for the Phase 1 remote test. Since it is expected that some test subjects will not complete the Phase 1 remote test, at least 90 test subjects are sought.

2. Describe the procedures for the collection of information.

Annoyance responses to stimuli will be collected electronically as described in the document, "TestApplicationProcedure.pdf." Responses will be returned to a NASA cloud service (the NASA Amazon Web Services) as they are entered by test subjects through the remote test application, which is accessed by test subjects on web browsers from their computers. Test subject response data from the NASA cloud service will be transferred to other NASA-managed computers for analyses at the conclusion of the Phase 1 remote test. Data will not be retained on the NASA cloud service.

3. Describe methods to maximize response rates and to deal with issues of non-response.

At least 90 test subjects will be given the Phase 1 remote test to account for some test subjects not completing the test so that responses over the full test from at least 80 test subjects are available for analyses. If a test subject does not complete the Phase 1 remote test, the responses that were provided may be used for analyses. Analyses approaches for Phase 1 remote test

results are amenable to unequal numbers of responses to test stimuli. Test subjects will be given an email to take the test. Test subjects who do not attempt the Phase 1 remote test will not be contacted again to complete the test.

4. Describe any tests of procedures or methods to be undertaken.

As described in item 16, Part A, augmented linear regression combined with percentile bootstrapping of regression model parameters will help compare responses from the Phase 1 remote test to the previous in-person test. Specifically, the Phase 1 remote test analysis will determine if the results in Figure 2 can be replicated for the in-person test. Here, augmented linear regression produced two regression lines relating annoyance with the A-Weighted Sound Exposure Level of the test stimuli. As a reminder, an annoyance rating of "Not at All" corresponds to a numerical annoyance rating of "2," and an annoyance rating of "Very" corresponds to a numerical annoyance rating of "8." The black circle markers represent the mean annoyance response to each of the recorded small-unmanned aircraft sounds (sUAS). The blue triangle markers represent the mean annoyance response to each of the recorded ground vehicle sounds. Figure 2 shows an offset of -5.84 dB in Sound Exposure Level between responses to small-unmanned aircraft and ground vehicles. A percentile bootstrap of the data gave 95% confidence intervals to the offset, not shown in Figure 2, of [-7.4, -4.4] dB. Since the confidence intervals do not overlap with 0 dB, and the coefficient of determination (R^2) of the augmented regression is relatively high, the offset was found to indicate a significant difference between annoyance responses to small-unmanned aircraft and ground vehicles [4]. The same stimuli as shown in Figure 2 are among the 84 sound stimuli (see item 3, Part A) that will be used in the Phase 1 remote test. Augmented linear regression of the Phase 1 remote test results to reproduce Figure 2 will determine the performance of the remote test method in terms of the ability to obtain a significant difference between responses to sounds of small-unmanned aircraft and ground vehicles [5].
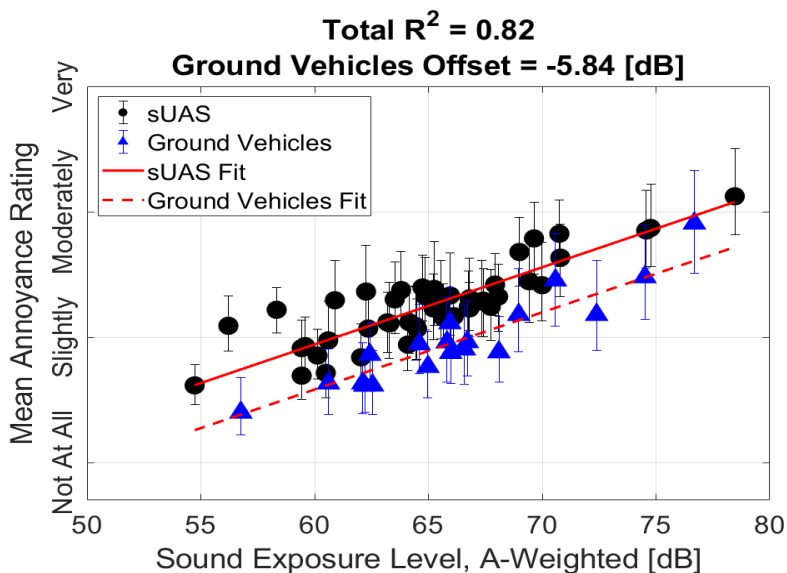
Figure 2. *Augmented linear regression results from previous in-person test.*

Determining if providing contextual cues affects responses to sounds will initially be analyzed using augmented linear regression. Instead of analyzing responses between small-unmanned aircraft and ground vehicle sounds, responses will be analyzed between the test subject groups with and without a contextual cue. Figure 3 shows augmented linear regression to artificially created responses to small-unmanned aircraft (sUAS) sounds from the two groups of test subjects. The artificial responses in Figure 3 were created by randomly and uniformly perturbing each original small-unmanned aircraft sound response from the previous in-person test [4] within 0.8 annoyance ratings from the original response. For the right plot, an extra step of shifting down one annoyance rating was taken for the responses hypothetically coming from the group provided a contextual cue. No annoyance rating was permitted to shift below 1 or above 11. The same analysis will be repeated for ground vehicle sound responses. If the results are similar to the left graph for both vehicle classes, it will support that providing a contextual cue does not provide a significant change in response [5].
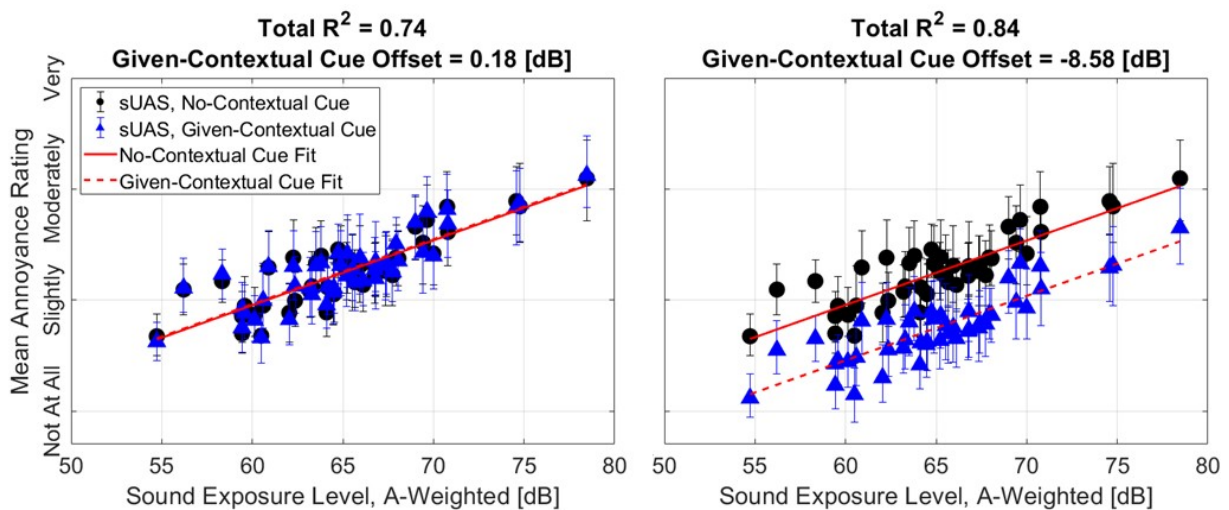

Figure 3. *Two potential Phase 1 remote test response outcomes.*

An ANOVA can also be performed on the overall mean annoyance responses of the two contextual cue groups. The ANOVA will be like as described in item 3, Part B, for responses to sounds from test subjects in two different geographic regions. The ANOVA can serve as additional support to conclusions drawn from augmented linear regression.

Analysis of test subject responses from different geographic regions is described in item 3, Part B, and will not be described further here.

5. Provide the name and telephone number of individuals consulted on statistical aspects of the design and the name of the agency unit, contractor(s), grantee(s), or other person(s) who will actually collect and/or analyze the information for the agency.

Andy Christian, a federal employee at NASA Langley Research Center, was minimally (less than 1 hour) consulted on the statistical aspects of analyzing test subject responses from different geographic regions.

The main contractor on NASA Technical Direction Notice (TDN) C2.01.023, Analytical Mechanics Associates, Inc., and its subcontractor, Harris Miller Miller & Hanson, will administer the Phase 1 remote test and collect test subject responses. Analyses of test subject responses will be performed by Siddhartha Krishnamurthy, a federal employee at NASA Langley Research Center.

References

[1] L. Gipson, "Advanced Air Mobility Project," National Aeronautics and Space Administration, 7 October 2021. [Online]. Available: https://www.nasa.gov/aeroresearch/programs/iasp/aam/description/. [Accessed 8 March 2022].

[2] S. A. Rizzi, D. L. Huff, D. D. Boyd, Jr., P. Bent, B. S. Henderson, K. A. Pascioni, D. C. Sargent, D. L. Josephson, M. Marsan, H. He and R. Snider, "Urban Air Mobility Noise: Current Practice, Gaps, and Recommendations," *NASA/TP–2020-5007433,* October 2020.

[3] 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, no. 4, pp. 641 - 679, 2001.

[4] A. Christian and R. Cabell, "Initial Investigation into the Psychoacoustic Properties of Small Unmanned Aerial System Noise," in *AIAA AVIATION Forum, Paper 4051*, Denver, Colorado, 2017.

[5] S. Krishnamurthy and S. A. Rizzi, "Feasibility study for remote psychoacoustic testing of human response to urban air mobility vehicle noise," in *INCE NOISE-CON 2022, Paper 676,* Lexington, KY, 2022.