A Laboratory Method for Assessing Audibility and Localization of Rotorcraft Fly-In Noise

Stephen A. Rizzi[1]

Senior Researcher for Aeroacoustics

NASA Langley Research Center

Hampton, Virginia

Andrew W. Christian

Research Engineer

NASA Langley Research Center

Hampton, Virginia

Menachem Rafaelof

Senior Research Engineer

National Institute of Aerospace

Hampton, Virginia

[1] Corresponding author: NASA Langley Research Center, Mail Stop 463, Hampton, Virginia, 23681-2199, Tel: +1 757 864-3599, s.a.rizzi@nasa.gov

Abstract

The low frequency content of rotorcraft noise allows it to be heard over great distances. This factor contributes to the disruption of natural quiet in national parks and wilderness areas, and can lead to annoyance in populated areas. Further, it can result in the sound being heard at greater distances compared to higher altitude fixed wing aircraft operations. Human response studies conducted in the field are challenging since test conditions are difficult to control. This paper presents a means of quantitatively determining the audibility and localization of rotorcraft fly-in noise, under a specified ambient noise condition, within a controlled laboratory environment. It is demonstrated using synthetic fly-in noise of a light utility helicopter. The method is shown to resolve differences in audibility distances due to different ground impedances, propagation modeling methods, and directivity angles. Further, it demonstrates the efficacy of an accelerated test method.

Nomenclature

A	amplitude of main and tail rotor harmonics (Pa)
c	speed of sound (m/s)
ε	localization error (deg)
f	frequency (Hz)
k	wavenumber (1/m)
p	pressure (Pa)
r	slant range between the source and observer (m)
R_p	complex reflection coefficient (dimensionless)

t time (s)

Z acoustic impedance (Pa-s/m^3)

ϕ plane wave angle of incidence (deg)

φ phase of main and tail rotor harmonics (rad)

Subscript

a air

g ground

1,2 direct and ground reflected paths, respectively

Introduction

Noise from low flying rotorcraft is a source of annoyance in both natural and populated areas and is the impetus for recent rules affecting their operations (Refs. 1-3). For rotorcraft noise to be judged as annoying, it must first be audible. Audibility depends on both the aircraft sound at the receiver and the background or ambient sound. Several human audibility models, based on signal detection theory (Ref. 4), have been developed and incorporated in computer codes, e.g., ICHIN (Ref. 5), INM (Ref. 6), NMSIM (Ref. 7) and AAM (Ref. 8). Model validation using sound jury data based on recorded signals is often problematic (Refs. 9-12), because unsteadiness associated with both the source and its long propagation path, as well as the ambient, negatively affect repeatability and controllability, and introduce uncertainty. In addition, such testing can only be conducted on an already developed vehicle, limiting the ability to assess the effect of unique design or operational characteristics.

Previous work by the authors (Ref. 13) focused on simulation of rotorcraft fly-in noise to enable such investigations. The approach in that work first synthesized the source noise pressure time history using available source noise data, and then used physical models to propagate the sound to a distant observer. In doing so, it purposefully eliminated the source and path unsteadiness, resulting in a monotonically increasing level on approach, making it ideal for use in controlled laboratory studies.

This paper builds on that work through the development of a laboratory method for assessing audibility and localization of simulated rotorcraft fly-in noise in the presence of a specified ambient noise. Specifically, a test method capable of quantifying subjective differences between fly-in noise at different source emission angles and propagation effects was sought. A psychoacoustic test, using simulated fly-in and ambient noise, was designed to demonstrate the approach and associated data processing, while exploring intersubject and intrasubject variance in measured audibility distances. The efficacy of an accelerated test method, was also established. General guidance for design of future experiments using this test method is also offered.

Test Scenario Development

The test methodology was developed around the requirement for assessing audibility and localization in an outdoor listening environment. For the purpose of this effort, three main elements are required: the ability to simulate fly-in noise; the ability to simulate a prescribed ambient noise condition; and the ability to reproduce the sound in a calibrated, three-dimensional (3D) listening environment. Note, however, that other simulated outdoor listening scenarios exist. If, for example, recordings of actual fly-in noise are used, inclusive of the aforementioned unsteadiness in such data, then both the propagated source

signal and the ambient noise at the time of measurement are embedded in the recording. Under this condition, the only remaining element for a laboratory test is sound reproduction. Such an approach has advantages over traditional jury testing in the field because the same sound can be exactly reproduced for multiple groups of test subjects. Alternatively, if the test objective is to assess audibility but not localization, then 3D sound reproduction is not required. In the remainder of this section, a recap of the fly-in noise simulation method is first discussed, followed by description of the test facility and the method used to reproduce the fly-in and ambient noise.

Fly-In Noise Simulation

Fly-in noise simulation is comprised of two parts: source noise synthesis and propagation. An in-depth treatment of these was provided by the authors in Ref. 13; a brief recap is provided herein.

Source noise synthesis

The source noise synthesis itself is comprised of two parts. The first part entails the source noise description and the second part entails generation of the source noise pressure time history based on that description. Only the main and tail rotor sources are considered herein as these dominate the audibility of an approaching rotorcraft (Ref 9). Other noise sources such as turbine, gearbox, and airframe are not considered.

In this work, recently acquired flight test data from the Airbus/Eurocopter AS350B “AStar” light utility helicopter served as the basis for the source noise description (Ref. 14,15). Data used in this study were obtained at the Sweetwater test site (run 302202). The AStar helicopter has a 3-blade main rotor with a nominal blade passage frequency

(BPF) of 19 Hz and a 2-blade tail rotor with a nominal BPF of 70 Hz. The advancing side of the main rotor is on the port side of this vehicle. Separation of the main and tail rotor acoustic signals from the ground measurement, obtained over a range of emission angles in both the polar and azimuthal directions, follows the method developed by Greenwood and Schmitz (Ref. 16). Because the relative contribution of main and tail rotors varies with emission angle, multiple angles must be considered when assessing audibility and localization.

With the source noise now characterized by separate blade passage signatures of the main and tail rotor as a function of emission angle, it is possible to synthesize long, combined pressure time histories for subsequent propagation. In the most general case of an arbitrary trajectory, the synthesis would entail morphing the signatures over time with changing emission angle. In the present application, however, the process is made simpler by the fact that the helicopter is at long range and is flying at a low altitude. For a straight-in approach, this effectively eliminates the need to morph the signatures because only one azimuthal emission angle is required (from the nose of the aircraft) and the low elevation angle changes very little over the range of interest. Had the aircraft been at higher altitude and first audible nearer the overhead position, this simplification would not be possible. To assess audibility at nonzero azimuthal angles, as would occur for flight paths that are to the side of the observer, the authors developed a method to "crab" the vehicle, i.e., to fly it with its nose pointed in the direction of the intended flight path, but with its direction of travel along the desired azimuthal angle at its intended speed toward the observer, see Ref. 13. This nonphysical approach allows one to test audibility using a single fly-in event at

the desired azimuthal angle, greatly improving the efficiency of the test and eliminating the need for morphing between different azimuthal angles.

Synthesis of combined main and tail rotor pressure time histories is performed in the time domain using an additive synthesis approach, that is,

$$p(t) = \sum_{i=1}^{m} A_i^{MR} \cos\left(2\pi \hat{f}_i^{MR} t + \varphi_i^{MR}\right) + \sum_{j=1}^{n} A_i^{TR} \cos\left(2\pi \hat{f}_i^{TR} t + \varphi_i^{TR}\right) \quad (1)$$

in which A and φ are the amplitudes and phases of each harmonic; $\hat{f}$ are the BPFs of each harmonic (scaled, if necessary, for "crab" angle, see Ref. 13); m and n are the number of main and tail rotor harmonics, respectively; and the superscripts MR and TR denote main and tail rotor, respectively. The amplitudes and phases are obtained by performing a single Discrete Fourier Transform (DFT) of the individual main and tail rotor signatures over a time window equal to their respective periods. The pressure time histories are generated with sufficient length (typically hundreds of seconds) to allow simulated fly-ins over tens of kilometers. A snippet of pressure time history at the nose at a shallow polar angle of 15° is shown in Figure 1. Since the main and tail rotor BPFs are not harmonically related, their superposition gives a time-varying quality to the generated waveform that would be missing if only the main rotor was included.

Simulated Propagation

The process for propagating the sound to an observer near the ground entails application of a time-varying time delay, gain, and filter to the synthesized source noise, in order to simulate Doppler shift, spreading loss, atmospheric absorption, and ground attenuation. These are discussed in detail in Ref. 13. Because one of the test method objectives is the ability to discriminate audibility and localization results obtained using

different ground impedance and propagation modeling approaches, a brief summary of that propagation element is next given.

Incorporation of an appropriate ground plane attenuation model for the reflected path is critical for this application as it has a significant effect on the sound received by the observer. The pressure at the observer is given by Ref. 17 as

$$\frac{p}{p_0} = \frac{1}{r_1} e^{-ikr_1} + \frac{R_p}{r_2} e^{-ikr_2} + (1 - R_p) \frac{\kappa}{r_2} e^{-ikr_2} \tag{2}$$

in which p_0 is the pressure at the reference distance; r_1 and r_2 are the slant ranges between the source and observer along the direct and ground reflected paths, respectively; k is the wavenumber in air; and the third term on the right hand side is the spherical wave correction to the plane wave assumption, with the boundary loss factor κ given by Refs. 18,19. The complex reflection coefficient R_p is given by Ref. 17 as

$$R_p = \frac{\sin\phi - Z_a / Z_g}{\sin\phi + Z_a / Z_g} \tag{3}$$

in which ϕ is the angle of incidence under the plane wave assumption, and Z_a / Z_g is the ratio of air to ground impedance. In this work, the Attenborough four-parameter model (Ref. 20) is used to specify the ground impedance in terms of the grain shape factor g_f, the pore shape factor ratio s_f, the volume porosity Ω and the flow resistivity σ. The spherical wave correction has the effect of increasing the attenuation at low frequencies and decreasing the attenuation at higher frequencies, in a range-dependent manner, relative to the plane wave model; see Figure 11 in Ref. 13.

Test Stimuli

The simulated fly-in conditions, listed in Table 1, determined the sounds used in the human subject test. Note that sample IDs 151-155 were identical. The remainder differ either by azimuth angle, ground surface, propagation model, or simulation speed.

In each case, a uniform atmosphere at a temperature of 15 °C, pressure of 1 atmosphere, 50% relative humidity, and density of 1.225 kg/m^3 was specified. A Prandtl number of 0.724, needed in the calculation of ground impedance, was also specified. With the exception of sample ID 160, the simulation was performed using the same straight and level trajectory as the original recording, namely, an altitude of 74.1 m above ground level and an indicated airspeed of 60.4 m/s (117.3 KIAS), with the 45° advancing side sample (ID 159) employing the aforementioned "crab" angle technique.

Sample ID 160 was an attempt at an accelerated test method. It was simulated at the same altitude, but at twice the original speed and double the speed of sound, thus keeping the Doppler shift factor the same as the other samples. This halved the time to render the fly-in compared to sample IDs 151-155. If successful, this approach could be used in the future to lessen the test time and thereby reduce any effects of test subject fatigue.

In all cases, the observer was set at a height of 1.2 m. Three of the four ground impedance parameters are dimensionless and were fixed at $g_f = 0.75$, $s_f = 0.875$, $\Omega = 0.715$ for all cases. Only the flow resistivity σ was changed between surfaces, and these were set to 10 MPa-s/m^2 for "new asphalt," 1.82 MPa-s/m^2 for "packed sandy silt," and 200 kPa-s/m^2 for grass (Ref. 21).

Figure 2 shows the unweighted overall sound pressure level (OASPL) as a function of distance from the observer for sample IDs 151, 156, 157, and 158. The distance has been

normalized with respect to the median audibility distance of sample IDs 151-155 (see Audibility Test Results and Analysis section), all of which are based on the same stimulus. This stimulus is subsequently referred to as the baseline. The OASPL of sample 160, when plotted against distance, is the same as sample IDs 151-155. In each case, a monotonic increase in OASPL with decreasing emission distance is indicated; first at a very slow rate at long distances and ending with a rapid rise at short distances. Because the human hearing threshold of pure tones has a standard deviation of about 4-6 dB in the 20–100 Hz one-third octave band, respectively, (Ref. 22), it is anticipated that subjective audibility distances of this low frequency rotor noise will vary greatly at distances far from the observer. The differences in OASPL between samples decrease with decreasing emission distance. At the farthest range, the difference in OASPL between ground surfaces is as much as 16 dB, between new asphalt (ID 156) and grass (ID 157). This large effect due to changes in ground impedance illustrates its importance in determining the distance at which audibility will occur, particularly for low frequencies where atmospheric absorption has little effect.

The comparison of propagation models indicates that the plane wave model (ID 158) underpredicts the attenuation relative to the spherical wave model at normalized distances greater than 0.4, giving an OASPL that is greater than 5 dB higher than the spherical wave model (ID 151) at the farthest distance. At normalized distances smaller than 0.4, the plane wave model overpredicts the attenuation. It is not immediately evident from eqn. (2) that this behavior should result from the spherical correction. This is a result of the interaction between the third term of eqn. (2) and spectral content of the source noise.

Test Facility Description

All testing was performed in the Exterior Effects Room (EER) at the NASA Langley Research Center in Hampton, Virginia, see Figure 3. The EER is a human-rated test facility, with a 31-channel, 3D, sound reproduction capability over a compensated frequency range of 16 Hz to 20 kHz; see Ref. 23 for additional details.

Reproduction of Fly-In Noise

The fly-ins were reproduced in the EER by positioning the sound source image in 3D space (the "virtual" position) at its emission position, that is, the location of the source when it generated the sound currently being heard. The emission position lags the current, or visual, position by the propagation time to the observer. The emission position is changed in time according to the flight trajectory, giving the illusion that the source is moving. All trajectories were rendered along a radial line toward the center of the room, but at potentially different azimuthal presentation angles. As previously noted, this simplified the source noise synthesis, and gave a time-invariant incoming angle for source noise localization.

Each fly-in event was typically 2-3 minutes in duration, and was initiated at a distant range. In this manner, they became audible in the presence of the persistent ambient noise, to be discussed below, at some time after the start of each presentation. The starting positions were determined through a pilot study performed using a group of researchers from the NASA Langley Research Center as test subjects, and were adjusted such that they all would not, on average, become audible around the same time after the start. A 10 s fade-in and 2 s fade-out was applied at the start and end of each presentation to avoid clicks or pops in the reproduction. All fly-ins were reproduced at the correct simulated level

using a single system gain, determined through a system calibration process employing a pink noise test signal, between 16 Hz and 5 kHz 1/3-octave band at a spectrum level of 64 dB, at multiple elevation and azimuth angles.

Simulation and Reproduction of Ambient Noise

A broadband, random ambient noise was synthesized from the prescribed 1/3-octave band SPL shown in Figure 4. The prescribed spectrum has an OASPL of 57.1 dB, and was selected on the basis of prior work. It is close to the rural area ambient spectrum of MIL-STD-1474E Level 1 (Ref. 24) above 100 Hz, and higher than that spectrum level at lower frequencies. The synthesis process converted the 1/3-octave band definition to a narrowband definition by assuming equal levels within each band, and assigned a random phase to each frequency component. Long-duration pressure time histories were generated via an overlap-add technique and were looped with a crossfade built into the ends.

Reproduction of the ambient noise was performed such that it would appear to be omnipresent. In general, this can be accomplished by simultaneously reproducing the sound at many locations about the test subjects. In this work, six virtual positions (5 overhead and 1 below the subject plane) were used. The signal path gain was adjusted to match the OASPL of the prescribed spectrum at a centrally located microphone position. It is seen in Figure 4 to agree well with the prescribed spectrum in the significant frequency range of the rotor noise source, that is, in the 1/3-octave band frequency range from 20 Hz to 1 kHz.

Psychoacoustic Test Methodology

Testing was performed in accordance with a test protocol approved by the NASA Langley Institutional Review Board. Each sample ID was evaluated by a group consisting of nominally 32 subjects each, in order to achieve a confidence interval (CI) of ±2 dB OASPL for resolving differences between samples of interest. This number of subjects was based upon past experience conducting similar tests, and is later confirmed in the Audibility Test Results and Analysis section of this paper. There were three subject groups and each group rated a collection of 32 fly-in sounds. Some of the 32 sounds were from the AStar vehicle and some were from other vehicles not pertinent to this study. The 32 sounds were divided into 4 sessions of 8 fly-ins each, with each session lasting about 20 minutes in duration. The test stimuli order was assigned according to a Latin Square scheme to minimize any bias due to presentation order. The AStar signatures were divided amongst subject groups a-c, as indicated in Table 1. Subjects were recruited from the local community and were required to have a hearing loss in either ear not greater than 40 dB, relative to reference hearing thresholds in Ref. 25, over the frequency range of 250 Hz to 6 kHz. Subjects were tested four at a time, as depicted in Figure 3. Groups a, b, and c contained 32, 32, and 33 subtest subjects, respectively, for a total of 97 subjects. Before the start of the test, each subject group was provided with written instructions, exposed to the range of sounds they would hear through a familiarization session, and became proficient with the use of the tablet PC for entering their responses through a practice session. Subjects were given breaks between sessions.

Audibility and localization were measured by soliciting two different types of subject response. For the audibility task, subjects were asked to indicate on tablet PCs when they

thought they heard the incoming aircraft, and when they were sure they heard it, see Figure 5a. Subjects were allowed to answer "THINK" as many times as they wished, but were able to answer "SURE" only once. Further, subjects were allowed to answer "SURE" without first answering "THINK." Allowing subjects to answer "THINK" multiple times before answering "SURE" was an effort to minimize false "SURE" responses. A visual indicator was provided to confirm the subject's selection. The raw audibility data for each sample ID is the simulation time at which subjects indicated "SURE." Post-processing of the data using the simulated trajectories allowed the emission and visual positions to be determined.

Only after the subjects indicated "SURE" in response to the audibility question, were they asked to indicate where they thought the sound was coming from, see Figure 5b. Physical numbers were mounted on the walls of the EER at every 30 degrees (some are visible in Figure 3), with position 5 located at the front of the room. Subjects were able to change their answer as many times as they wanted until the end of each fly-in (before the overhead position). The tablet PC recorded the simulation time and room position (with a resolution of 0.3°) of each tablet entry. These were post-processed using the simulation trajectory and knowledge of the EER azimuthal presentation angle to obtain the emission and visual positions in the same manner as above, and the localization error, that is, the deviation of the azimuthal subject angle from the presentation angle (expressed in degrees).

Relation to Other Psychoacoustic Test Methods

The proposed test methodology was formulated to reproduce the scenario of sound jury testing in the field, but within the controlled environment afforded by the laboratory. It is worthwhile to compare this test proceedure to other existing psychoacoustic methods

typically used in psychoacoustic laboratories. With few counterexamples, nearly all other methods do not use stimuli that continuously change as do the fly-in sounds. Rather, other methods rely on repeatedly testing small segments of sound that are constant during each trial, and that are changed systematically between trials, e.g., their level is adjusted up or down depending on the past performance of the subject. The most common of these discrete methods is known as the method of limits, and is often employed in a descending fashion where the target sound is first presented at a clearly audible level and then reduced incrementally, e.g., as in basic audiometric tests.

A further delineation between psychoacoustic testing methods can be drawn based on the type of question posed to the subject. In "Yes/No" tests, subjects are presented with an interval of time and asked whether the target sound was present in the interval or not. In n-Alternative Forced-Choice (nAFC) tests, subjects are presented with n intervals (one of which contains the target sound) and asked to provide their best guess as to which interval contained the target sound. An important distinction can be drawn between these methods: the former relies on the subject to be a diligent and consistent reporter of when they, in fact, heard the target sound, whereas in the nAFC paradigm the experimenter can tell when the subject guessed by observing at what signal level the response accuracy tends toward 1/n. That is, for nAFC, the subject's decision process (as to when to respond that they can hear a target sound) is removed as a variable from their performance. Due to the advantage of control over the decision process provided by nAFC testing, it is most commonly employed in contemporary psychoacoustic tests.

With the above understanding, the fly-in method proposed is perhaps most similar to a test using the ascending method of limits with a Yes/No paradigm, implying that a subject's

decision process is still a factor in the data gathered in this study. Further information on both classical and contemporary psychophysical test methods and theories can be found in Ref. 26.

Audibility Test Results And Analysis

In the following, each set of audibility results has been normalized with respect to the median audibility range determined from the aggregation of 163 subject "SURE" responses associated with sample IDs 151-155.

Analysis of Individual Samples

A test of normality was first performed on the audibility data on the basis of the emission distance. The null hypothesis is that the data is normally distributed. Results from two different normality tests are provided in Table 2. A p-value ≤ 0.05 (shown highlighted) is an indicator, but not proof, that the null hypothesis should be rejected. In other words, when $p \leq 0.05$, it indicates that the data is not likely originated from a population with a normal distribution. The Kolmogorov-Smirnov (K-S) test rejects the null hypothesis for the fewest sample IDs (2 of 10), while the Lilliefors test rejects the null hypothesis for half of the sample IDs.

Plots of the most normal and least normal samples are shown in Figure 6 and Figure 7, for sample IDs 157 and 159, respectively. A visual indicator of the normality, or lack thereof, is provided through comparison of the response data with the probit function fit, the inverse of which is the cumulative distribution function of the normal distribution. The plots also show two figures of merit associated with the audibility data; the median and the mean audibility distances. The median audibility distance, x_{50}, represents the distance at

which half the subjects find the sample audible (Pr=0.5), while the mean audibility distance, $\bar{x}$, is that at which the average subject would find the sample audible. The probability associated with the mean audibility distance is determined here from interpolation of the subject data, not from the probit function fit. Whereas the mean and median are nearly identical in Figure 6, consistent with a normal distribution, they markedly differ in Figure 7, due to the skewness arising from several early responses combined with the balance accumulated over a relatively short range interval.

Confidence intervals of the median and mean were computed using different methods. For the median audibility distance, CIs were computed according to the method described in Ref. 27, where the upper and lower CI probabilities are given as

$$\Pr = \frac{1}{n}\left[\frac{(n+1)}{2} \pm \frac{z\sqrt{n}}{2}\right] \tag{4}$$

in which n is the sample size, and z is taken as 1.96 for the 95% CI. This expression is valid for any continuous distribution. Note that for most sample IDs, the total number of responses equaled the total number of subjects indicated in Table 1. The 95% CIs of the median distance x_{50} were determined directly through interpolation of the subject data at the CI probabilities. Consequently, the upper and lower CIs are not necessarily symmetric about x_{50}.

Confidence intervals for the mean audibility distance were computed using two different methods; one assumes normality of the data and one does not. For normally distributed data, the CI is given as (Ref. 27)

$$CI = \mu \pm \frac{t\sigma}{\sqrt{n}} \tag{5}$$

in which the central tendency, μ, is estimated by the sample mean, the standard deviation, σ, is estimated by the Bessel-corrected sample standard deviation, and t is derived from the Student's t-distribution with n-1 degrees of freedom, taken here for the 95% CI. The bias-corrected accelerated (BCa) bootstrapping method allows the CI to be calculated without the assumption of normality, see Ref. 28 for details. In this work, 10^6 sample means were calculated from sets of n random responses that were selected randomly, with replacement, from the subject data, allowing the CIs to be repeatable to 4 significant figures.

Median and mean audibility distances and their 95% CIs are shown in Figure 8. Confidence intervals calculated using eqn. (5) are always symmetric about the mean, whereas, in general, those calculated using the BCa method are not. These results show that the Student's t-distribution method is sufficient and that the more computationally intensive BCa method is unnecessary in this case. Figure 8 also indicates that

- Samples IDs using the baseline stimulus and tested in the same manner, i.e., 151-155, have comparable audibility distances.
- Sample ID 156 for the harder "new asphalt" surface indicates audibility at roughly twice the baseline distances, while sample ID 157 for the softer "grass" surface indicates a closer distance relative to the baseline distances. Even though audibility is not determined on the basis of OASPL, these results are consistent with the OASPLs of the simulated fly-ins (see Figure 2).
- The plane wave assumption used in the propagation model (ID 158) indicates audibility at ranges greater than the baseline distances, also consistent with its OASPL.

- Audibility distances of the 45° advancing side case (ID 159) are roughly half that of the baseline distances.
- The accelerated test method (ID 160) indicates comparable audibility distances relative to the baseline distances.
- There appears to be a weak dependence of the size of the CIs on the location of the absolute audibility range (in both median and mean). This may be due to the fact that, at closer mean detection distances, the rate of change of the sample sound (e.g., as measured in OASPL) is higher than at farther distances. Therefore, the subjects' responses are more likely to form a tighter cluster for samples having closer range detection, resulting in a smaller CI.

The data also shows that CIs of the mean audibility distance, computed using both methods, are comparable (within 6% of each other), even for those cases that do not pass the normality test (IDs 152, 155, 156, 159 and 160). In other words, violation of the normality assumption used in eqn. (5) nevertheless yielded comparable CIs to the alternative method, which did not assume normality.

At this point, it is worthwhile to return to the decision made during the test design phase with respect to the number of subjects tested. The stated objective was to achieve CIs within ± 2 dB OASPL. In order to calculate that, the test designer must have knowledge of the variance of the response data, which in the current test design was based on prior knowledge. Figure 9 shows the CIs based upon data from this experiment. Here, it is seen that the maximum CI is 1.9 dB OASPL for 32-33 subjects; the minimum CI is 1.1 dB. Using this data, and eqn. (5), Figure 9 shows how the expected CIs would vary as the

number of subjects included in the test changes. This would allow future tests to be designed with a specific CI in mind.

Analysis of Multiple Samples

Recall that the objective of this effort was to develop a test method for evaluating audibility that could discriminate between results obtained for different ground impedances, propagation modeling methods, and directivity angles. Beyond what can be inferred from the CIs, one-way analysis of variance (ANOVA) is next performed to more rigorously determine whether audibility distances from different samples have the same mean. ANOVA is conceptually similar to comparing CIs between pairs of samples, although the process is performed across multiple samples simultaneously and provides a measure of significance; see, for instance, Ref. 29.

In each of the following, one-way ANOVA is performed on selected pairs of data, with the null hypothesis that their means are equal. The observed value of the F test statistic is the ratio of the response variance between sample IDs to the response variance within sample IDs. The critical value of F depends on the degrees of freedom associated with F and the desired significance level α. The probability of an F value greater than the observed F is denoted as the p-value. There are two methods of evaluating the test hypothesis: if $F \geq F_{critical}$ or p-value $\leq \alpha$, then the null hypothesis is rejected. Both evaluations were performed at the 5% significance level. Note that the ANOVA analyses assume normally distributed data. The consequence of that assumption being violated in some cases is discussed below. Before evaluating the statistical difference between sample IDs that are expected to be different, it is worthwhile to evaluate the data for those sample IDs that are expected to be the same. For this purpose, consider the independent

evaluations of the baseline, that is, sample IDs 151, 152, and one of 153-155. Note the latter are not independent as they were evaluated by the same subject group. Table 3 shows the results of the one-way ANOVA between sample IDs 151 and 152, 151 and 153, and 152 and 153. In all cases, $F < F_{critical}$ and p-value > 0.05. Therefore, the null hypothesis is accepted, demonstrating that the mean audibility distance of the same stimuli between different subject groups is not significantly different. This is an important result because it demonstrates that not all stimuli need to be evaluated by all subjects when making comparisons of mean audibility distances.

Comparison of different ground impedances

The statistical significance of differences found in the mean audibility distances for different ground impedances was determined by performing separate one-way ANOVA between sample IDs 151 and 156, 151 and 157, and 156 and 157. These sample IDs were evaluated by the same subject group (a). The highlighted results in Table 3 show that the null hypothesis is rejected, indicating that the test method and data analysis are capable of resolving those differences. Note that the F values greatly exceeded $F_{critical}$, and the p-values << 0.05 in two of the three analyses.

Comparison of different propagation modeling

Next, consider the comparison between different propagation methods by the ANOVA between sample IDs 152 and 158. These sample IDs were evaluated by the same subject group, (b). The highlighted results in Table 3 again show that the null hypothesis is rejected, indicating that the test method and data analysis are capable of resolving the difference in propagation methods. The importance of this finding is that the appropriate

propagation method must be chosen for the scenario of interest. Like the ground impedance analyses, the F value greatly exceeds $F_{critical}$, and the p-values << 0.05.

Comparison of different source emission angles

The highlighted results in Table 3 indicate that the null hypothesis is rejected, see the comparison of sample IDs 152 and 159. These sample IDs were also evaluated by the same subject group, (b). Like the analysis of ground impedance and propagation method data, the F value greatly exceeds $F_{critical}$, and the p-values << 0.05.

Comparison of normal and accelerated test methods

The ANOVA results between sample IDs 151 and 160, shown in Table 3, indicate that the null hypothesis should be accepted, that is, there is no statistically significant difference in the mean audibility distance when testing with the accelerated test method. These sample IDs were also evaluated by the same subject group, (a). While this finding is promising, caution should be exercised in its application. In particular, the method assumes that the audibility range is independent of the rate of change of the sound level. In this application, the rate of change of the sound level near the mean audibility distances is slow, even at double speed. However, when the rate of change of the sound level is fast, as seen in Figure 2 close to the observer, then the accelerated method might yield statistically different ranges.

It is worthwhile to next discuss the results of the above ANOVA analyses in the context of the earlier results of non-normality of the sample responses. Since ANOVA is a method based on the assumption of normality of the distribution of the sample response, and some of these responses displayed non-normality, it could be argued that the results of

significance/insignificance between samples were due to the failure of the ANOVA method on the data. There are several reasons to believe that this is not the case. First, there does not appear to be a relationship between instances of normality/non-normality and significance/insignificance. For instance, responses from sample 152 (one of the repeats) indicated non-normality (see Table 2). However, the mean audibility distance was not significantly different between sample 152 and samples 151 and 153 (see Table 3). Further, the results of the ANOVA analysis make sense in the context of the samples themselves, i.e., that repeat sounds show insignificantly different sample responses as expected. Secondly, the *p*-values provided in Table 3 indicate that the differences between samples are typically emphatic, i.e., when samples are different, their *p*-value is typically far away from the critical value of $p = 0.05$. The only ANOVA result that is close to this threshold is in the comparison of IDs 151 and 157, but neither of those samples showed significant non-normality. Finally, the CIs from the Student's t-distribution method (based on an assumption of normality) and the non-parametric BCa method (not based on an assumption of normality) showed only a 6% difference, see Figure 8. Differences on this order are not likely to change the result of the ANOVA analyses from an insignificant difference to a significant difference, or vice versa.

Localization Test Results And Analysis

In the following, distances associated with subjective ratings of localization angles have been normalized with respect to the particular sample ID x_{50} audibility range. In this manner, the normalized emission distance at which 50% of the subjects correctly localize the sound, to within some specified error, ε, is always less than unity.

Analysis of Individual Samples

Typical localization results are shown in Figure 10, for sample ID 155. Here, all subject responses within ± 30º, ± 22.5º, and ± 15º of the EER presentation angle are plotted as a function of their corresponding normalized emission distance. There are more subject responses than there are subjects because each subject can make multiple localization responses, and all those within the specified localization error range are included. It is seen that subjects' ability to correctly localize the source improves at closer ranges, that is, the aircraft must be closer for subjects to localize with ± 15° than for subjects to localize within ± 30°. The high frequency content of the signal becomes greater with decreasing range due to reduced atmospheric absorption. The data therefore suggests that subjects' improved ability to localize the sounds is associated with increasing higher frequency content of the signal. This observation holds for all sample IDs tested, as shown in Figure 11 at the 50% probability level.

Analysis of Multiple Samples

Figure 10 and Figure 11 represent snapshots of localization acuity at three selected angles. Of course, there are a continuum of angles at which 50% of respondents are correct (think about slicing along the Pr=0.5 line in Figure 10). Such data is plotted in Figure 12 for all ten sample IDs. An estimation of the emission distance at which subjects correctly localize the source to within some specified error of the actual (presentation) angle can be obtained via linear regression giving

$$\text{Normalized Emission Distance, } x = 0.428 \log_{10} |\varepsilon| + 0.034 . \tag{6}$$

This fit accounts for 83% of the variance in the data. No attempt is made here to generalize this measure beyond the current dataset. The sample mean normalized emission distance and sample standard deviation at three selected error ranges (± 30º, 22.5º, 15º) are given in Table 4. It is seen that the localization error decreases with mean distance, consistent with Figure 10 and Figure 11, and the standard deviations are small relative to the means.

Conclusions

A laboratory method has been developed to assess the human audibility and localization of rotorcraft fly-in noise. The method utilizes high-fidelity simulated source and ambient noise to eliminate the unsteadiness found in actual recordings. Psychoacoustic test results demonstrate that this approach is capable of resolving the effects of ground impedance, propagation method, and source directivity on the audibility distance in a statistically significant manner. Mean audibility distances were found to be statistically invariant with respect to subject group, and differences between a normal speed and an accelerated test method were found to be statistically insignificant. Localization results indicated that subjects were better able to locate the source with decreasing distance, and a relationship between the localization error and distance was observed.

With this capability, it is possible to further explore the factors leading to human audibility of rotorcraft. Some of these have been explored in terms of the listening environment (ground impedance) and source directivity, while others remain to be explored, e.g., operations and weather. With this approach, it is also possible to assess the effects of unsteadiness in the received signal (whether due to unsteadiness of the source itself or that introduced by atmospheric turbulence) and unsteadiness in the ambient, through systematic introduction of these effects in a future experiment designed for this

purpose. Such an endeavor would help to better interpret human response data acquired using actual recordings or in the sound jury tests conducted in the field.

Opportunities for future research include an assessment of the effectiveness of audibility models incorporated in various computer codes using the data acquired, and a study to relate a subject's performance in this fly-in setting to alternative modes of testing (e.g., an nAFC task) for comparable sounds.

Acknowledgments

This research was supported by the National Aeronautics and Space Administration, Aeronautics Research Mission Directorate, Advanced Air Vehicles Program, Revolutionary Vertical Lift Technology (RVLT) Project. The authors wish to thank Dr. Kevin P. Shepherd, Distinguished Research Associate, NASA Langley Research Center (retired), for sharing his experience and guidance during the development of this capability, and Susan Gorton, RVLT Project Manager, NASA Langley Research Center, for her support.

References

[1]"Special flight rules area in the vicinity of Grand Canyon National Park: Actions to substantially restore natural quiet," National Park Service, US Department of the Interior, Draft Environmental Impact Statement DES 10-60, 2011.

[2]"Report on the Los Angeles Helicopter Noise Initiative," Federal Aviation Administration 2013.

[3]"The New York North Shore Helicopter Route," Federal Aviation Administration, Department of Transportation, 14 CFR Part 93, Docket No. FAA-2010-0302, 2010.

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

[5]Mueller, A.W., Smith, C.D., Shepherd, K.P., and Sullivan, B.M., "A new version of the helicopter aural detection program, ICHIN," NASA TM-87745, 1986.

[6]Boeker, E.R., et al., "Integrated Noise Model (INM) Version 7.0 Technical Manual," FAA AEE-08-01, 2008.

[7]Plotkin, K.J., "The role of aircraft noise simulation models," *InterNoise 2001*, The Hague, The Netherlands, 2001.

[8]Page, J.A., Wilmer, C., Schultz, T., Plotkin, K.J., and Czech, J., "Advanced acoustic model technical reference and user manual," SERDP Project WP-1304, May 2009.

[9]Hartman, L. and Sternfeld, H., "An experiment in aural detection of helicopters," US Army Air Mobility R&D Laboratory USAAMRDL Technical Report 13.50 under Contract DAAJ02-71-C-0065, AD917355, Ft. Eustis, VA, 1973.

[10]Abrahamson, L.A., "Correlation of actual and analytical helicopter aural detection criteria, Volume 1," US Army Air Mobility R&D Laboratory USAAMRDL-TR-74-102A, Ft. Eustis, VA, January 1975.

[11]Miller, N.P., Anderson, G.S., Horonjeff, R.D., Menge, C.W., Ross, J.C., and Newmark, M., "Aircraft noise model validation study," HMMH Report No. 295860.29, January 2003.

[12]Fleming, G.G., Plotkin, K.J., Roof, C.J., Ikelheimer, B.J., and Senzig, D.A., "Assessment of tools for modeling aircraft noise in the national parks," Federal Interagency Committee on Aviation Noise (FICAN), March 2005.

[13]Rizzi, S.A. and Christian, A., "A method for simulation of rotorcraft fly-in noise for human response studies," *InterNoise 2015*, San Francisco, CA, 2015.

[14]Watts, M.E., Greenwood, E., and Stephenson, J.H., "Measurement and characterization of helicopter noise at different altitudes," *AHS 72nd Annual Forum*, West Palm Beach, FL, 2016.

[15]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.

[16]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.

[17]Piercy, J.E., Embleton, T.F.W., and Sutherland, L.C., "Review of noise propagation in the atmosphere," *Journal of the Acoustical Society of America*, Vol. 61, No. 6, 1977, pp. 1403-1418.

[18]Attenborough, K., Hayek, S.I., and Lawther, J.M., "Propagation of sound above a porous half-space," *Journal of the Acoustical Society of America*, Vol. 68, No. 5, 1980, pp. 1493-1501.

[19]Rasmussen, K.B., "Sound propagation over grass covered ground," *Journal of Sound and Vibration*, Vol. 78, No. 2, 1981, pp. 247-255.

[20]Attenborough, K., "Acoustic impedance models for outdoor ground surfaces," *Journal of Sound and Vibration*, Vol. 99, No. 4, 1985, pp. 521-544.

[21]Embleton, T.F.W. and Daigle, G.A., *Aeroacoustics of Flight Vehicles: Theory and Practice. Volume 2: Noise Control*, Atmospheric Propagation, H.H. Hubbard, Editor. 1991, NASA RP 1258: Hampton, VA.

[22]"Acoustics - Statistical distribution of hearing thresholds of otologically normal persons in the age range from 18 years to 25 years under free-field listening conditions," International Organization for Standardization ISO 28961, 2012.

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

[24]"Design criteria standard noise limits," Department of Defense MIL-STD-1474E, 2015.

[25]"Acoustics - Reference zero for the calibration of audiometric equipment - Part 1: Reference equivalent threshold sound pressure levels for pure tones and supra-aural earphones," International Organization for Standardization ISO 389-1:1998(E), 1998.

[26]Gescheider, G.A., Psychophysics: The fundamentals, 3rd Ed. Lawrence Erlbaum Associates Inc., Mahwah, NJ, 1997.

[27]Snedecor, G.W. and Cochran, W.G., Statistical Methods, 6th Edition. The Iowa State University Press, Ames, IA, 1967.

[28]Efron, B., "Better bootstrap confidence intervals," *Journal of the American Statistical Association*, Vol. 82, No. 397, 1987, pp. 171-185.

[29]Montgomery, D.C., Design and Analysis of Experiments, 3rd Ed. John Wiley & Sons, New York, NY, 1991.

List of Figures

List of Tables

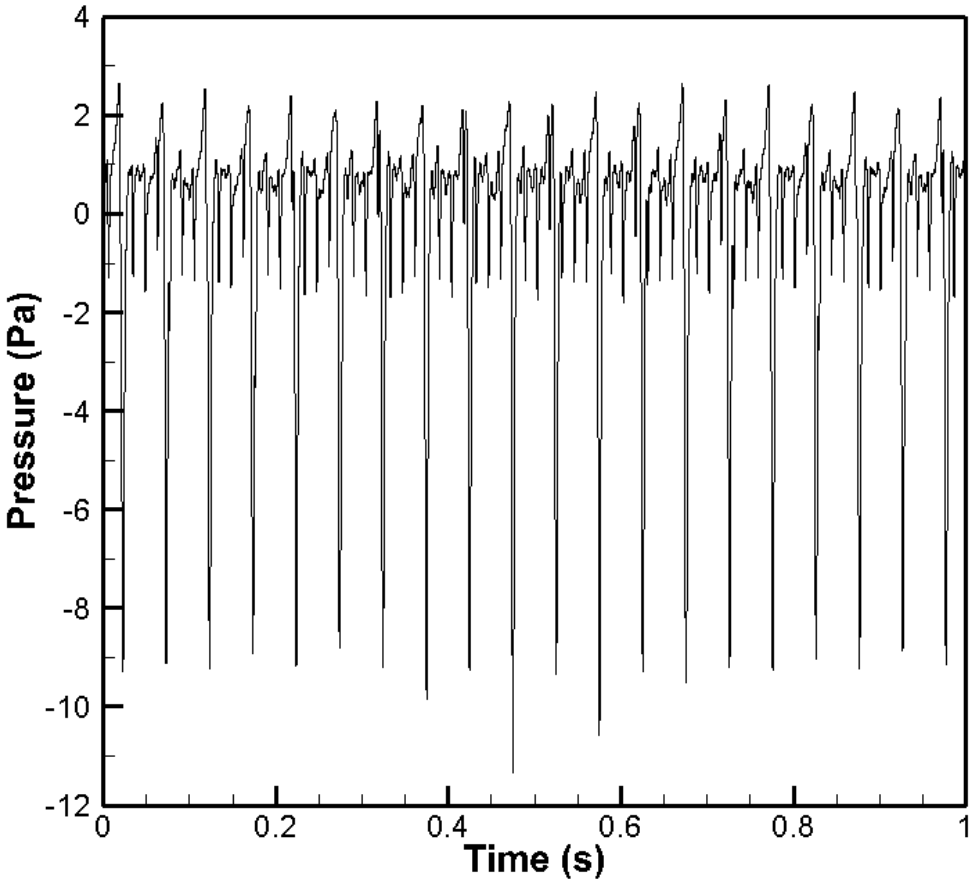


Figure 1. Snippet of synthesized pressure time history.

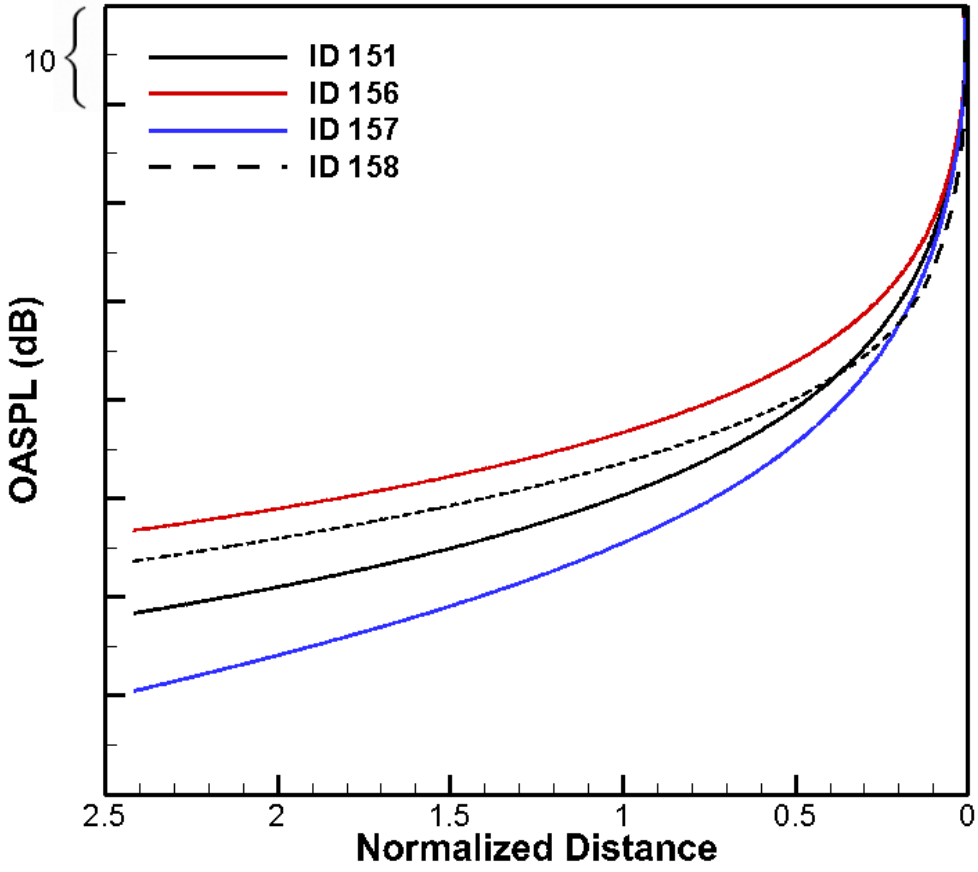


Figure 2. Comparison of overall SPL for simulated fly-ins with three ground impedances.

Figure 3. Photo depicting NASA personnel posing as test subjects in the NASA Langley Exterior Effects Room.

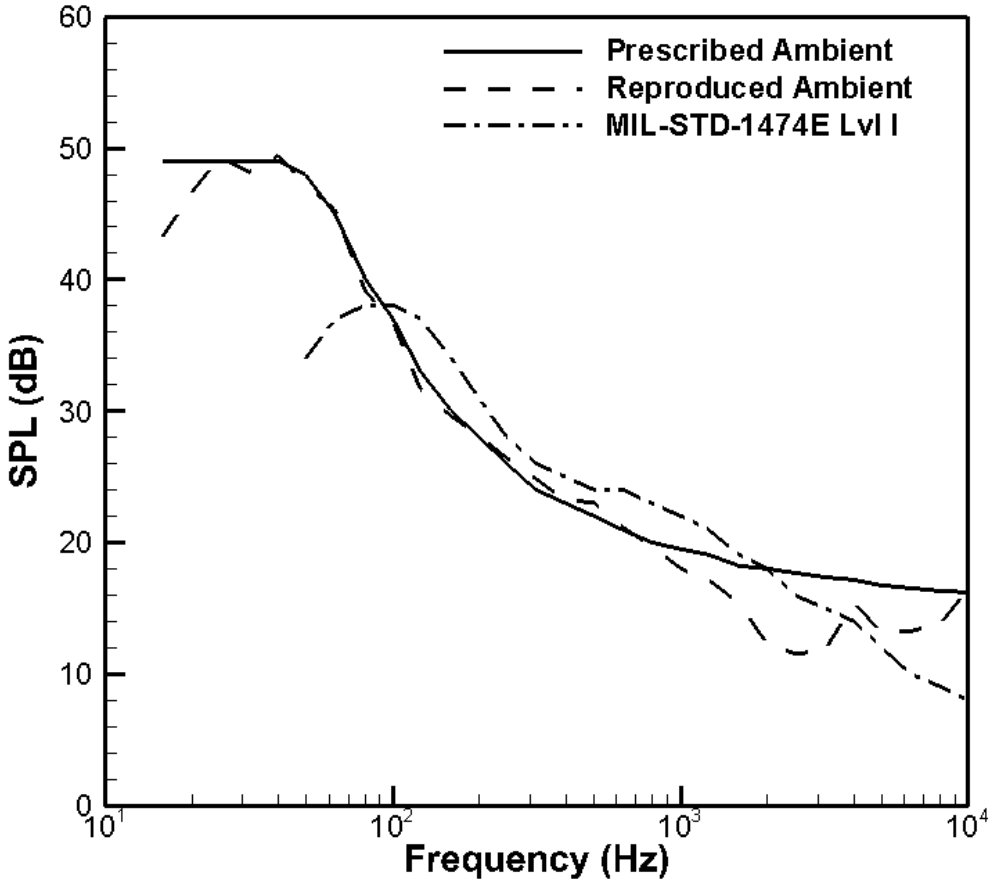
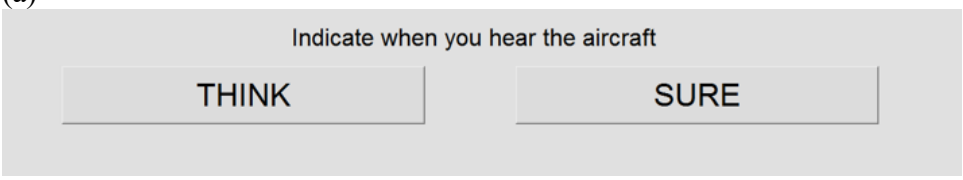


Figure 4. Prescribed and reproduced ambient noise used in the EER.

(a)



(b)

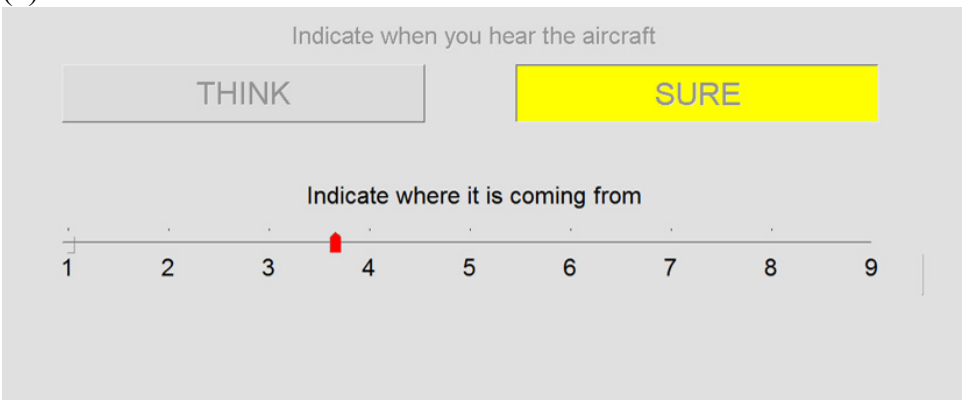


Figure 5. Images displayed on tablet asking subjects to indicate if they think or are sure they heard the aircraft (a), and where they thought it was coming from (b).

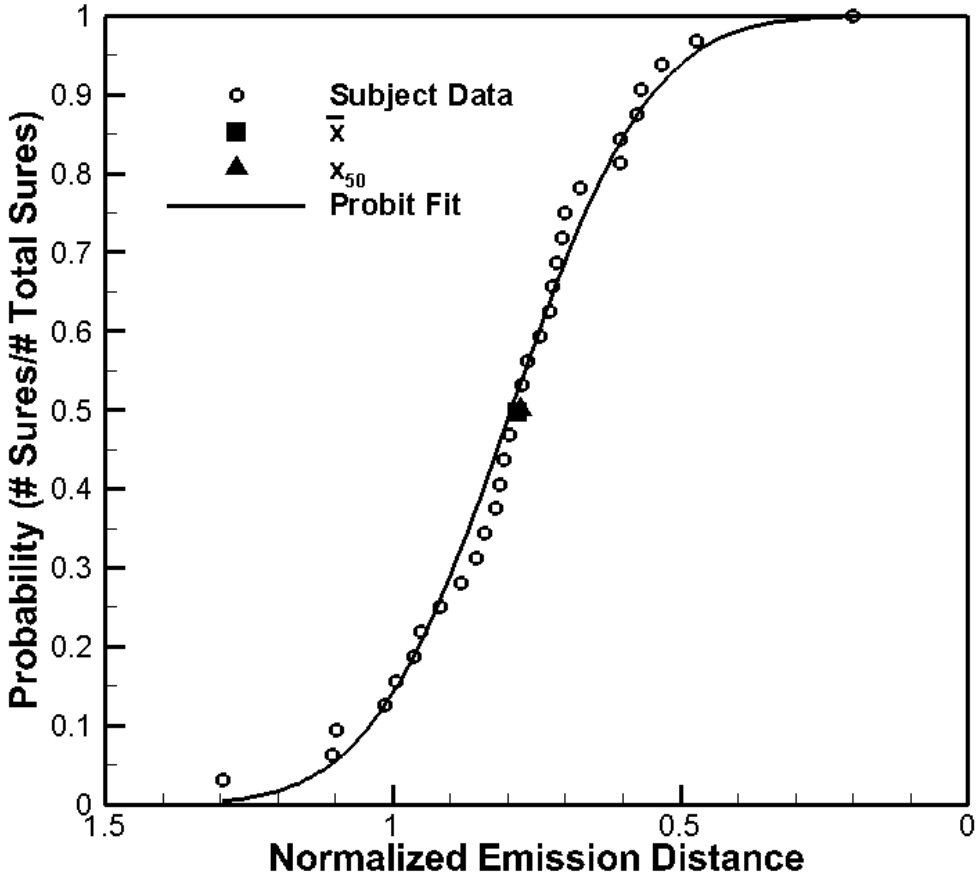


Figure 6. Subject data for sample ID 157.

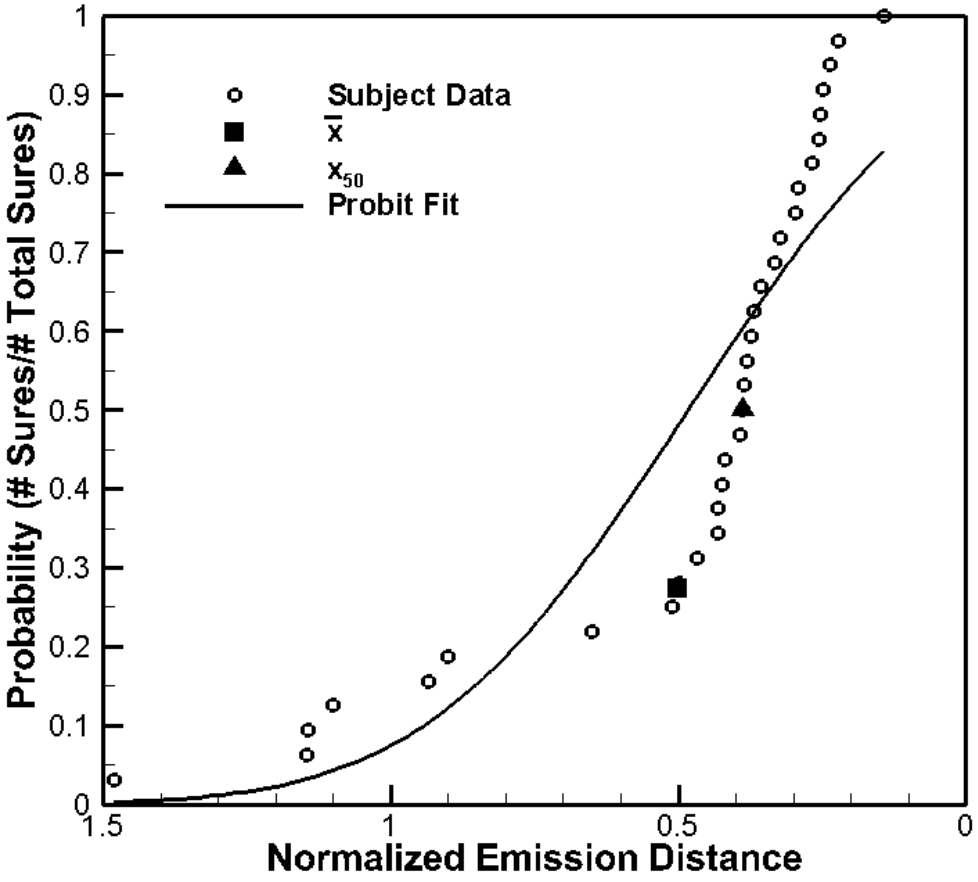


Figure 7. Subject data for sample ID 159.

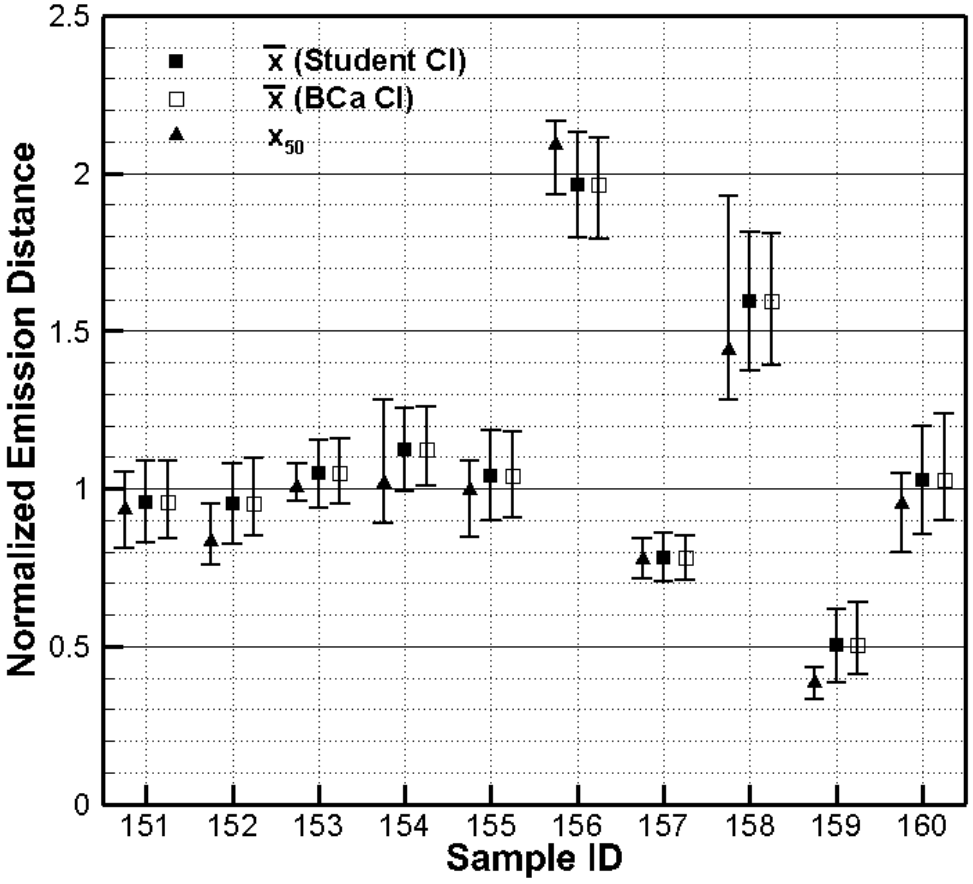


Figure 8. Comparison of normalized mean and median emission distances and their 95% CIs.

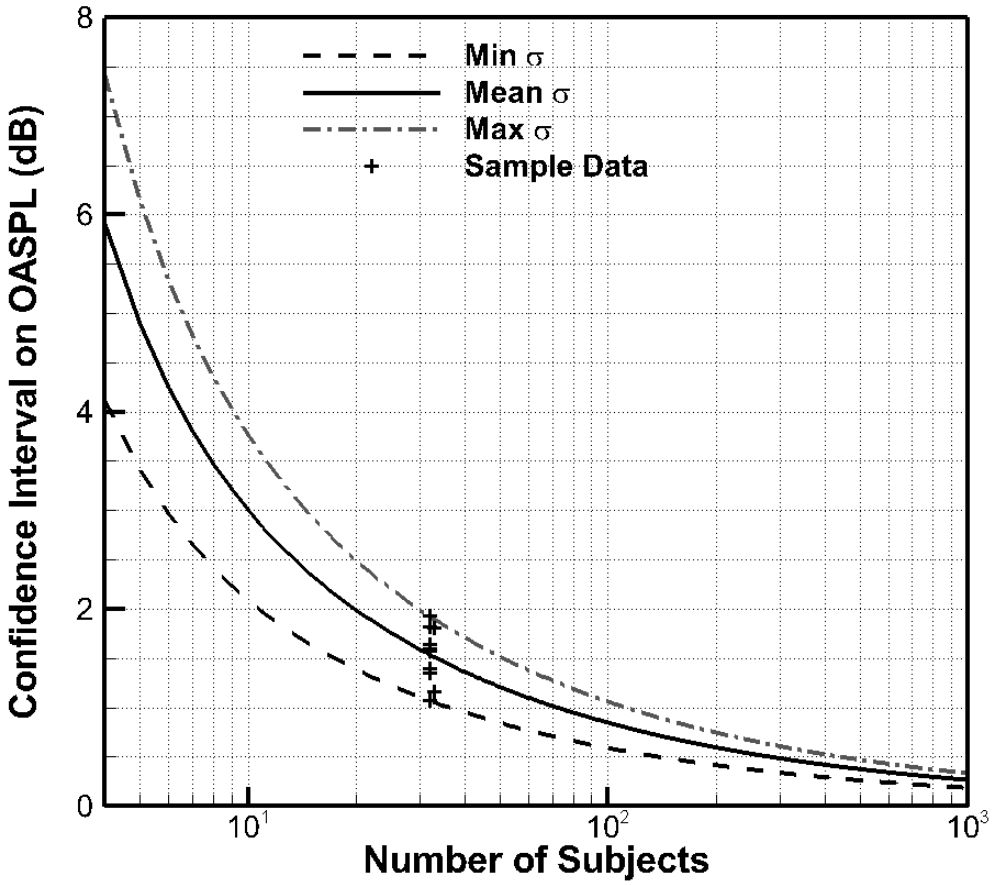


Figure 9. Confidence intervals as a function of the number of test subjects (based on test data).

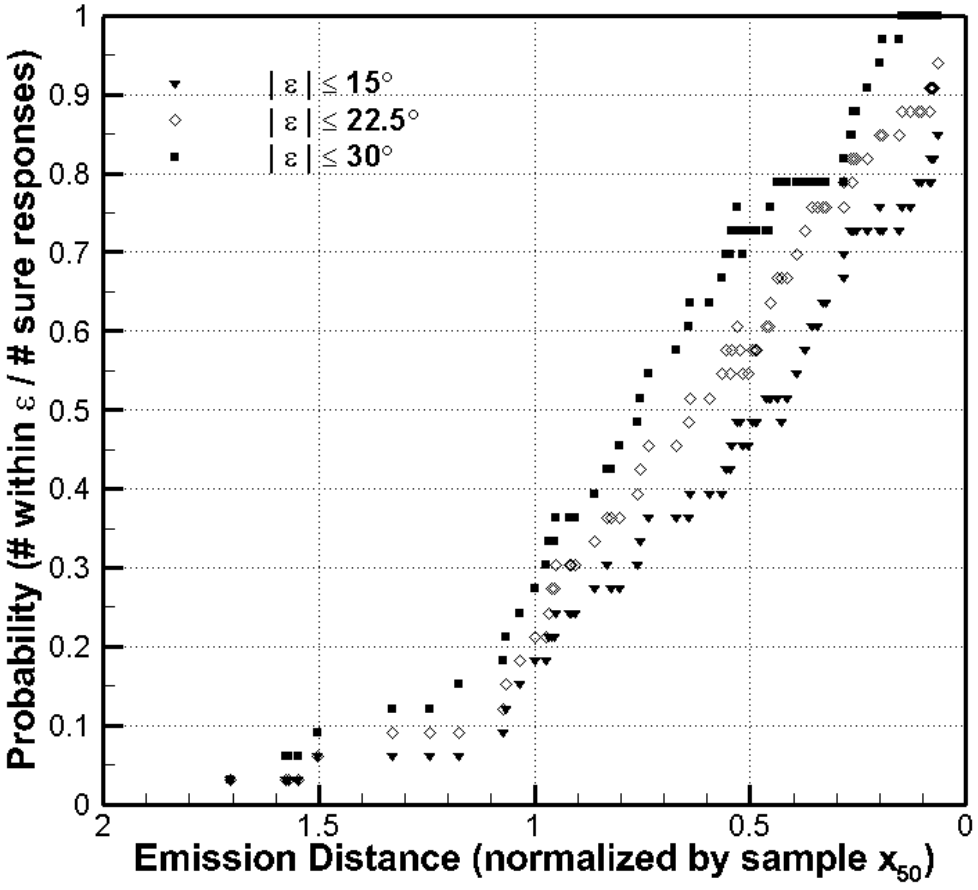


Figure 10. Localization results for sample ID 155.

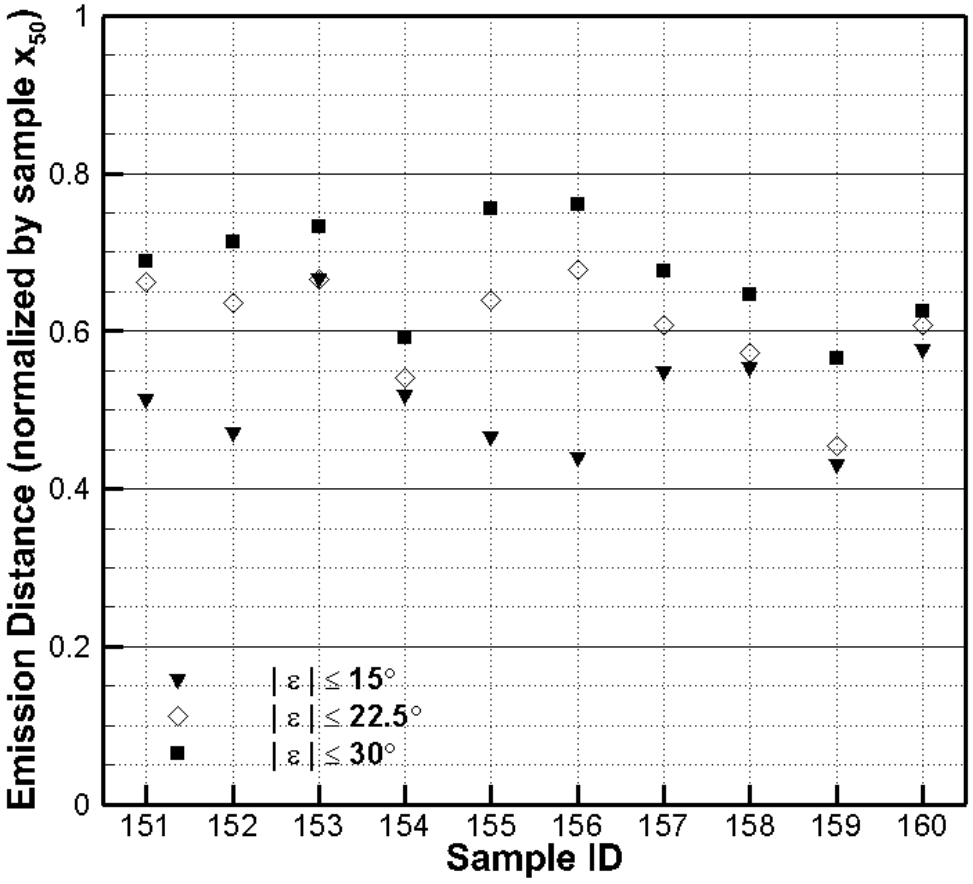


Figure 11. Comparison of localization data for all sample IDs at the 50% probability level.

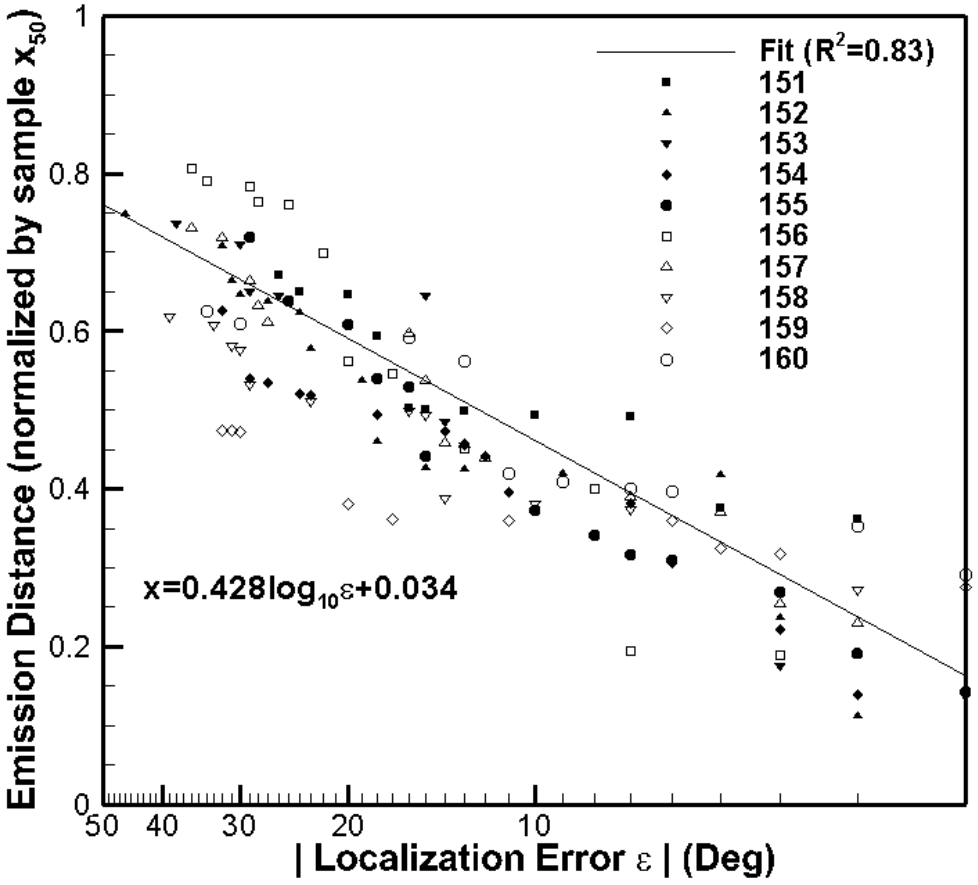


Figure 12. Distance at which 50% of respondents evaluated the source azimuthal angle to be within specified error.

Table 1. List of sample IDs and simulated conditions.

Sample ID	Azimuth	Surface	Propagation Model	Speed	Subject Group	No. Subjects
151	Nose	Pk. Sandy Silt	Spherical corr.	1x	a	32
152	Nose	Pk. Sandy Silt	Spherical corr.	1x	b	32
153	Nose	Pk. Sandy Silt	Spherical corr.	1x	c	33
154	Nose	Pk. Sandy Silt	Spherical corr.	1x	c	33
155	Nose	Pk. Sandy Silt	Spherical corr.	1x	c	33
156	Nose	New Asphalt	Spherical corr.	1x	a	32
157	Nose	Grass	Spherical corr.	1x	a	32
158	Nose	Pk. Sandy Silt	Plane wave	1x	b	32
159	45° Adv.	Pk. Sandy Silt	Spherical corr.	1x	b	32
160	Nose	Pk. Sandy Silt	Spherical corr.	2x	a	32

Table 2. Normality tests of audibility data (based on emission distance).

Sample ID	K-S test (*p*-value)	Lilliefors test (*p*-value)
151	0.6770	0.2476
152	0.0383	0.0010
153	0.4359	0.0675
154	0.5555	0.1365
155	0.3173	0.0275
156	0.3407	0.0335
157	0.9715	0.5000
158	0.7406	0.3196
159	0.0142	0.0010
160	0.0601	0.0010

Table 3. One-way ANOVA results (based on emission distance).

Sample ID 1	Sample ID 2	Description	F	$F_{critical}$	p-value
151	152	Same stimuli, diff. subject groups	0.004	3.996	0.949
151	153		1.189	3.993	0.280
152	153		1.360	3.993	0.247
151	156	Different ground, same subject group	94.61	3.996	4.30E-14
151	157		5.756	3.996	1.95E-02
156	157		173.0	3.996	1.30E-19
152	158	Different propagation, same subject group	26.83	3.998	2.64E-06
152	159	Different source angle, same subject group	27.94	3.996	1.72E-06
151	160	Alt. test method, same subject group	0.420	3.996	0.519

Table 4. Statistics of the localization results across all sample IDs.

Normalized Emission Distance	Localization Error		
	$\lvert\varepsilon\rvert \leq 30°$	$\lvert\varepsilon\rvert \leq 22.5°$	$\lvert\varepsilon\rvert \leq 15°$
Sample mean	0.676	0.607	0.518
Sample std. dev.	0.067	0.069	0.072