A Study in a New Test Facility on Indoor Annoyance Caused by Sonic Booms

Jonathan Rathsam, Alexandra Loubeau, and Jacob Klos
Langley Research Center, Hampton, Virginia
Since its founding, NASA has been dedicated to the advancement of aeronautics and space science. The NASA scientific and technical information (STI) program plays a key part in helping NASA maintain this important role.

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

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

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

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

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

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

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

Specialized services also include creating custom thesauri, building customized databases, and organizing and publishing research results.

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

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

- E-mail your question via the Internet to help@sti.nasa.gov

- Fax your question to the NASA STI Help Desk at 443-757-5803

- Phone the NASA STI Help Desk at 443-757-5802

- Write to:
  NASA STI Help Desk
  NASA Center for AeroSpace Information
  7115 Standard Drive
  Hanover, MD 21076-1320
A Study in a New Test Facility on Indoor Annoyance Caused by Sonic Booms

Jonathan Rathsam, Alexandra Loubeau, and Jacob Klos
Langley Research Center, Hampton, Virginia
Acknowledgments

The authors thank Regina Johns for recruiting and scheduling test subjects, as well as administering hearing tests.

The use of trademarks or names of manufacturers in this report is for accurate reporting and does not constitute an official endorsement, either expressed or implied, of such products or manufacturers by the National Aeronautics and Space Administration.
Abstract

A sonic-boom simulator at NASA Langley Research Center has been constructed to research the indoor human response to low-amplitude sonic booms. The research goal is the development of a psychoacoustic model for individual sonic booms to be validated by future community studies. The study in this report assessed the suitability of existing noise metrics for predicting indoor human annoyance. The test signals included a wide range of synthesized and recorded sonic-boom waveforms. Results indicated that no noise metric predicts indoor annoyance to sonic-boom sounds better than Perceived Level, PL. During the study it became apparent that structural vibrations induced by the test signals were contributing to annoyance, so the relationship between sound and vibration at levels of equivalent annoyance has been quantified.
Contents

1 Introduction 3

2 Facility and Sound Generation Mechanism 4

3 Test Signals 5

4 Test Method 8

5 Analysis of Results 9
  5.1 Objective Data Analysis 9
    5.1.1 Acoustic Measurements 9
    5.1.2 Vibration Measurements 9
  5.2 Subjective Data Analysis 13
    5.2.1 Comparison of Noise Metrics 13
    5.2.2 Potential Imperfect Modeling of Loudness 14
    5.2.3 Multimodal Annoyance Response to Loudness and Vibration 15

6 Conclusions 19

A Loudness Levels of Signals 22

B Signal Processing Flow Chart 27
1 Introduction

Advances in aircraft design techniques suggest future aircraft with a sonic boom quiet enough to be acceptable for overland flight. NASA has set a goal to create a validated psychoacoustic model to assess the indoor annoyance caused by low-amplitude sonic booms. The model will inform the development of an outdoor, noise-based regulation for low-amplitude sonic boom exposure. The study reported here was conducted in an indoor sonic-boom simulator at NASA Langley Research Center. Its primary intent was to evaluate the performance of existing noise metrics in predicting indoor annoyance to a wide range of sonic-boom signals. During the study it became apparent that structural vibrations induced by the test signals were contributing to annoyance, so the relationship between sound and vibration at levels of equivalent annoyance has been quantified.

The link between structural vibration and annoyance is well-documented in field studies of community reactions to sonic booms. In one study, 93% of 1145 residents surveyed reported “interference with ordinary living activities” due to “shaking of the house”, and 38% reported annoyance due to “shaking of the house” [1]. A more recent field study indicated that 75% - 95% of surveyed residents reported noticing rattles or vibration associated with sonic booms [2]. A list of community surveys on reactions to sonic booms can be found in [3]. Although vibration has been linked with indoor annoyance in these field studies, it has not previously been investigated in controlled laboratory studies. Below are summaries of the main existing laboratory studies on indoor annoyance caused by sonic boom and blast noise.

A study in a previous laboratory facility was conducted on indoor annoyance caused by blast noise [4]. The blast noise was simulated by a large acoustic piston, produced by stretching a heavy rubber membrane across a large shake table of dimensions 12 ft by 12 ft. In the same large room that housed the acoustic piston, a house mock up was constructed for test subjects. The house mock up was vibration-isolated from the shake table, so vibration was either nonexistent or barely perceptible. A paired-comparison test method was used to quantify annoyance caused by blast noise with and without rattle noise. It was found that the reference signal had to be increased by 6 - 13 dB to yield the same annoyance as the test signal in the presence of rattle. The only metrics analyzed in the study were A-weighted and C-weighted Sound Exposure Level.

In another study, sound-reproduction systems were installed in private residences so annoyance caused by sonic-boom waveforms could be measured with high situational realism [5]. The test apparatus was not intended to control or measure vibration and rattle, so these variables were not considered in the analysis. A category scaling test method was used, and the data was analyzed using regression and correlation. Equivalent annoyance resulted when individual event level was adjusted by 3 dB per halving or doubling of number of events. This so-called “energy addition theory” was found applicable to sonic booms even when they occurred relatively infrequently during the day. Results indicated that Perceived Level Mark VII (PL) [6] was a significantly better predictor of annoyance than other noise metrics. However, no time-varying metrics were studied.
A previous study at NASA Langley Research Center was conducted using a small, loudspeaker-driven booth for reproducing outdoor sonic-boom waveforms. A filter was applied to outdoor waveforms to simulate the effects of transmission through structural walls [7]. A magnitude-estimation technique was used, and the data was analyzed via regression and correlation. Vibration and rattle were nonexistent in the test apparatus. The key finding in this study was that loudness and annoyance ratings were equivalent for outdoor signals, while annoyance ratings were higher than loudness ratings for indoor signals. Based on this result, subjects in the current study were asked to rate annoyance instead of loudness. Another key result was that Perceived Level (PL) was found to be the best predictor of annoyance for indoor and outdoor booms.

A fourth study was conducted using a simulator with a demountable partition, which allowed both exterior and interior sonic booms to be simulated [8]. The study investigated the increment in noise metric value by which reference signals had to be increased to maintain subjective equality with sonic booms at increased levels. The ratio of required increase in reference signal level to increase in sonic boom level is called the relative growth rate of annoyance. The key finding was that the relative growth rate of annoyance depended on the reference sound. Relative growth rates of annoyance were typically 2:1 between either broadband or mid-frequency references and sonic booms, but the relative growth rate was 1:1 when the reference was an octave-band noise burst centered at 63 Hz. This result suggests that the low-frequency content of sonic booms, rather than the impulsiveness, controls annoyance. In contrast to [4], results indicated that the reference sound had to be increased by only 5 dB to yield the same annoyance as the test signal in the presence of rattle.

The study reported in this document was the second study conducted in the facility. The first study analyzed indoor annoyance caused by sonic booms, but only three test signals were played at seven amplitudes each. Fewer test signals had to be used because each subject listened at all three listener locations to maximize the number of different subjects responding at each location. Although the study produced other useful results, the authors concluded that three test signals were insufficient for the purpose of comparing noise metrics. In the current study the subjects did not rotate seats, so more waveforms could be played back to each group of subjects. A NASA Technical Memorandum documenting the first study is currently being prepared.

2 Facility and Sound Generation Mechanism

The current test facility was built specifically for testing indoor annoyance caused by sonic booms. Sonic-boom waveforms can be played over two arrays adjacent to two exterior walls of the small room shown in Fig. 1. Only one array was used for this study. The room has dimensions 12 ft x 14 ft x 8.5 ft and conventional wood and drywall construction. The arrays of loudspeakers create a desired pressure loading on the exterior surfaces of the wall that drives the structure into vibration, and the structural vibrations couple with the interior volume. As a result the sound transmission from exterior to interior occurs much as it would in the field. The inside
of the facility is furnished as a living room and seats three test subjects for each test session, as shown in Fig. 2. Subjective judgments are logged on netbook computers and monitored in real time via wireless local area network. Sound-absorbing wedges were placed inside the facility’s two closets to minimize room reverberation time. A document detailing the facility is currently being prepared and will be available as a NASA Technical Memorandum.

Figure 1: A loudspeaker array is shown as attached to the exterior of the facility. The array used in the current study is not visible in this photo and contains 24 cells instead of 28 cells as shown here.

Figure 2: Listener locations within the facility.

3 Test Signals

A total of 23 test signals were used. The calculated loudness levels of these signals are listed in Tables A1 - A4 in App. A. A signal processing flow chart for the preparation of test signals is given in App. B. Five of the signals were low-amplitude
Table 1: Matrix of band-pass-filtered synthesized signals. Signals used in the study are designated with their signal number in Tables A1 - A4 in App. A. All signals are filtered with a fourth-order high-pass Butterworth filter.

N-waves measured with microphones on ground boards far from buildings and other vertical surfaces [9]. The duration of these N-waves was approximately 150 ms with rise times between 1 ms and 7 ms. These signals are numbered 16 – 20 in Tables A1 - A4. The remaining 18 test signals were synthesized sonic booms. Two of the synthesized booms were 200-ms-duration N-waves, with rise times of 6 ms and 9 ms. The rise time is the time elapsed as the shock grows from 10% to 90% of its maximum amplitude. The desired rise time of these two signals is incorporated into the synthesized waveform at the front and rear shocks according to Eq. 1 [10]

$$ p(t) = \frac{\Delta p}{2} \left[ 1 + \tanh \left( \frac{t}{T} \right) \right] $$

where $p(t)$ is the instantaneous pressure, $\Delta p$ is shock amplitude, $t$ is the time, and $T$ is the rise time. This synthetic procedure for introducing a rise time into a nearly-instantaneous shock is known as “tanh-thickening.” The two signals with rise times of 6 ms and 9 ms are numbered 22 and 23 in Tables A1 - A4 and belong to the family of signals used in the first study conducted in the facility. A third synthesized signal, number 21, is the ground signature predicted for a low-boom aircraft configuration.

The remaining 15 signals were constructed to systematically vary spectral content in order to identify differences among metrics, as described below. To motivate the method of constructing these 15 signals, it is necessary to consider first the range of noise metrics. The primary difference among the following metrics is their weighting function, particularly at low frequencies: Perceived Level (PL) [11], A- and C-weighted Sound Exposure Level (ASEL and CSEL) [12], Stationary Loudness (MGSL) [13], Time-Varying Loudness (MGTVL) [14], Zwicker Loudness Level–diffuse (LLZd) [15], and Perceived Noise Level (PNL) [16]. If signals are designed with varied low-frequency content, it is possible to expose these differences and reduce correlations among metrics.

Each synthetic signal began as two delta functions separated by 200 ms, as shown in Fig. 3(a). Separate low-pass and high-pass Butterworth filters with the same corner frequencies, $f_0$, were successively applied to yield a waveform as in Fig. 3(b). The resulting spectrum looked like a band-pass filter centered about

<table>
<thead>
<tr>
<th>Corner Frequency [Hz]</th>
<th>LP Filter Order</th>
<th>2</th>
<th>3</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>1</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>2</td>
<td>8</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>3</td>
<td>9</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>-</td>
<td>10</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>4</td>
<td>11</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>75</td>
<td>5</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>110</td>
<td>6</td>
<td>-</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 3: Method for constructing filtered-impulse test signals.

Figure 4: Correlations between PL and other noise metrics for listener locations 1-3.
frequency $f_0$, as shown in Fig. 3(c). This procedure was repeated at various corner frequencies, indicated with vertical lines in Fig. 3(d). All signals were filtered with a fourth-order high-pass Butterworth filter, and a low-pass Butterworth filter of either second, third, or fifth order. The procedure resulted in a full matrix of 21 signals shown in Table 1. These signals were subsequently down-selected to the test sounds numbered 1 – 15 in Table 1. The signals were down-selected to keep the test length within the attention span of an average subject.

Each test signal was played at six amplitudes. The amplitudes were intended to be evenly distributed above and below the equality point. A pilot test with 12 subjects was conducted to ensure that the range of amplitudes for each test signal was roughly centered about the amplitude of subjective equality with the reference. The range of amplitudes was chosen to make signals loud enough to be audible, but not so loud that they caused audible creaks or rattles in the test facility. A separate pilot test with 12 subjects showed that equality point remained consistent for two different reference sounds.

After the study it became apparent that test signals were mistakenly played with inverted polarity. The error does not affect calculated noise metrics because all metrics are based on signal energy, or pressure-squared quantities. Also, the authors listened to boom signals with intended and inverted polarity and concluded that there were no audible differences in the test signals as heard inside the facility. In the future, the polarity of the desired waveforms will be confirmed.

In summary, a suite of signals was designed with the requirement of maintaining a sonic-boom-type character of a front shock and a rear shock in the time domain. The correlation of several metrics with PL across all 23 test signals is shown in Fig. 4. A-weighted Sound Exposure Level (ASEL) has a high correlation with PL, but for most other noise metrics the correlation is not strong. As a result, differences are expected among metrics in their capacity to predict annoyance.

4 Test Method

The test method was paired-comparison with a reference sound. The reference sound was the tanh-thickened N-wave of duration 200 ms and rise time of 6 ms, i.e. signal 22 at amplitude 3 in App. A. This sound was also used as a test sound. The method requires that subjects hear a pair of sounds and choose which sound is more annoying. Each pair consisted of the reference sound at a constant amplitude and a test sound at one of six amplitudes. Subjects heard each pair in both orders, A-B and B-A. The 23 test sounds, each replayed at 6 amplitudes, paired with the reference in both orders, yielded 276 pairs, which took approximately 1 hour and 15 minutes to judge. Each group of test subjects heard the pairs presented in a different random order. The simulator seats three subjects, and ten groups were tested for a total of 30 subjects. Subjects remained in the same seat throughout the test.

Thirty subjects (10 male, 20 female) participated in the study. The subjects were obtained from a subject pool of local residents and were paid for their participation in the study. Ages of the test subjects ranged from 18 to 64 years with a median
All subjects were audiometrically screened prior to the test to demonstrate auditory acuity within 40 dB of audibility thresholds for tones from 500 Hz to 6 kHz.

5 Analysis of Results

5.1 Objective Data Analysis

5.1.1 Acoustic Measurements

To estimate the range of exposure, the test signals were measured at six locations within the simulator: current listener locations 1 - 3, and proposed listener locations, 4 - 6. The coordinates of each microphone location are given in Table 2, and shown in Fig. 5. Single half-inch microphones were placed approximately at seated ear height. Head and torso effects were therefore not taken into account. Each acquired sound represents the time-domain average of 10 repeats to reduce effects of background noise and microphone thermal noise inherent in any single measurement. A fading window was applied to each measured waveform prior to Fourier Analysis to minimize spectral leakage. The window function was 2 seconds in length, as shown in Fig. 6. It consisted of the first half of a 1-sec-duration Hanning window, followed by one second of unity gain, followed by the second half of a 1-sec-duration Hanning window. The window was aligned as needed to position the entire boom signal within the unity portion of the window.

Analysis of subjective response data is least complicated when each signal causes the same loudness level at all listener locations. However, the variation in loudness levels among listener locations 1 - 3 for a single test sound was as much as 10 PL dB due to varied distance from the wall with the active loudspeaker array, as shown in Fig. 5, as well as due to room acoustics. For the current single-array configuration, an analysis of sound levels measured at the six locations was performed to determine if listener locations could be found with more uniform exposure. A repeated measures ANOVA test was conducted on the loudness level (PL) of all 138 signals at all six locations. Mauchly’s Test of Sphericity [17] yielded a significant result, indicating that the covariance in loudness level among the locations is not uniform. However, the $F$-statistic in the ANOVA test was invariant whether or not it was corrected for violations of sphericity. The results indicated a significant difference in loudness level among all six locations, $F(5, 685) = 783.12, p \leq 0.0001$. Post-hoc Bonferroni comparisons across the listener locations indicated that the mean loudness level of signals at locations 1 and 5 is not significantly different at the 95% confidence level. By contrast, the mean loudness level of signals at all other locations is different. Subsequent measurements have shown that loudness exposure is more uniform in the room with both arrays active.

5.1.2 Vibration Measurements

Subjects were asked in post-test conversations if certain types of signals were consistently more annoying than others. Subjects responded that they were annoyed if they “felt the room shake” or when it “sounded like someone had fallen in the
The authors interpreted these comments as references to vibration. Based on this feedback, the vibration caused by the 23 test signals at 6 levels was measured on the ground at the foot position of each listener location. The accelerometer locations are listed in Table 2 and shown in Fig. 5. Each test sound was repeated 10 times and averaged to reduce effects of background noise inherent in any single measurement. The measured vibration data was filtered by a weighting function, $w_k$, shown in Fig. 7. The perception threshold for whole-body vibrations is a $w_k$-weighted peak amplitude of $1.5 \text{ cm/s}^2$ [18].

The vibration at the floor was measured with and without approximately 150-lb weights in all seats to determine whether the weight of human subjects had an observable effect on floor vibrations. The $w_k$-weighted peak accelerations measured with and without loading at all three listener locations were compared using a two-factor repeated-measures ANOVA analysis. The two-factors were listener location and loading condition. The repeated-measures ANOVA technique was chosen because the same test signals were analyzed at each listener location. The analysis simultaneously tested three hypotheses: 1) measured vibration did not vary with

---

**Figure 5:** Interior of simulator indicating location of measurement transducers. Location coordinates are listed in Table 2.
Table 2: Location coordinates of measurement sensors. Microphones were placed at ear height of seated subjects and accelerometers were attached to the hardwood floor. The coordinate origin is shown in Fig. 5.

<table>
<thead>
<tr>
<th>Sensor</th>
<th>x [m]</th>
<th>y [m]</th>
<th>z [m]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mic 1</td>
<td>3.47 (11' 5&quot;)</td>
<td>1.89 (6' 2&quot;)</td>
<td>1.13 (3' 8&quot;)</td>
</tr>
<tr>
<td>Mic 2</td>
<td>2.33 (7' 8&quot;)</td>
<td>2.71 (8' 11&quot;)</td>
<td>1.13 (3' 8&quot;)</td>
</tr>
<tr>
<td>Mic 3</td>
<td>1.21 (4' 0&quot;)</td>
<td>2.77 (9' 1&quot;)</td>
<td>1.11 (3' 8&quot;)</td>
</tr>
<tr>
<td>Mic 4</td>
<td>2.37 (7' 9&quot;)</td>
<td>0.57 (1' 10&quot;)</td>
<td>1.10 (3' 7&quot;)</td>
</tr>
<tr>
<td>Mic 5</td>
<td>1.78 (5' 10&quot;)</td>
<td>0.57 (1' 10&quot;)</td>
<td>1.11 (3' 8&quot;)</td>
</tr>
<tr>
<td>Mic 6</td>
<td>1.19 (3' 11&quot;)</td>
<td>0.52 (1' 8&quot;)</td>
<td>1.10 (3' 7&quot;)</td>
</tr>
<tr>
<td>Accel 1</td>
<td>2.95 (9' 8&quot;)</td>
<td>1.63 (5' 4&quot;)</td>
<td>0 (0' 0&quot;)</td>
</tr>
<tr>
<td>Accel 2</td>
<td>2.13 (7' 0&quot;)</td>
<td>2.13 (7' 0&quot;)</td>
<td>0 (0' 0&quot;)</td>
</tr>
<tr>
<td>Accel 3</td>
<td>1.22 (4' 0&quot;)</td>
<td>2.13 (7' 0&quot;)</td>
<td>0 (0' 0&quot;)</td>
</tr>
</tbody>
</table>

Figure 6: Time-domain fader window applied to an interior microphone measurement.

The result of the first hypothesis test indicates a significant difference in measured vibration at different listener locations, $F(2, 274) = 96.380, p \leq 0.0001$. The result of the second hypothesis test indicates a significant difference in measured vibration with and without loading, $F(1, 137) = 61.496, p \leq 0.0001$. The result
from the third hypothesis test indicates an interaction effect between listener location and loading condition, $F(2, 274) = 18.439, p \leq 0.0001$. Fig. 8 illustrates this interaction effect and other trends in the measured data. The loading had almost no effect on vibration at listener location 2, but the loading did cause a measurable difference in vibration at locations 1 and 3. Accelerometer 2 was located at the center of the room in the $x$-direction, so it probably lies on a nodal line for significant vibrational modes. The results indicate that vibration measurements should be made with loading to approximate the weight of human subjects. The results also indicate that vibration cannot be considered uniform across listener locations.
5.2 Subjective Data Analysis

5.2.1 Comparison of Noise Metrics

The paired-comparison test method results in the proportion of subjects more annoyed by each test signal than by the reference signal, or simply the “percent more annoyed”. Each test signal is played at six amplitudes. Generally the percent more annoyed increases with signal amplitude. When percent more annoyed is plotted against test-signal level, the data can be fitted by a logistic regression curve, which has an $S$-shape, as shown in Fig 9. The point on the curve at which percent more annoyed equals 50% is called the equality point, or EP. It represents the level of subjective equality between the test signal and the reference signal.

![Figure 9: The subjective responses were fitted by a logistic regression curve with an $S$-shape. The equality point, EP, is the level at which 50% of the respondents are more annoyed by the test signal.](image)

![Figure 10: The notional metric groups the equality points B and C close together, but not equality point A. The figure of merit, $M$, quantifies the spread in equality points.](image)
A perfect noise metric would yield the same EP for all signals judged equally annoying. The predictive capability of the metric is, therefore, quantified by the numeric spread of equality points for the 23 test sounds. The narrower the spread in equality points, as illustrated in Fig. 10, the better the predictive capability of each noise metric.

As mentioned earlier, the levels from a single test sound varied by up to 10 PL dB across listener locations due to room acoustics and distance from the array. The variation in levels prevents direct comparison of equality points across listener locations. Instead equality points are compared based on their deviation from the reference equality point. The reference sound was used as a test sound, compared with itself, and the resulting equality point was recorded. The reference equality point at each location was then subtracted from the equality point of each test signal at that location. This normalization by the reference equality point enabled direct comparison of equality points across listener location.

The normalized differences were averaged across all listener locations to produce a figure of merit $M$, as shown in Eq. 2. In the equation, $EP$ is the equality point for the reference signal or for a test signal 1 - 23. The lower the value of $M$, the better the predictive capacity of the metric. The results for some common metrics are shown in Fig. 11. The metric PLmod is explained in Sec. 5.2.2. The error bars represent 95% confidence intervals about the figure of merit, $M$, and were constructed using a bootstrap resampling technique [19]. However, the error bars cannot be used to compare the metrics directly.

$$M = \sum_{location=1}^{3} \sum_{signal=1}^{23} \left| EP_{signal} - EP_{ref} \right|_{location} \text{ [dB]}$$  \hspace{1cm} (2)

An ANOVA procedure was used to compare the figures of merit, $M$. A repeated-measures technique is not chosen in this case because the analysis is conducted across different metrics. There is a significant difference across figures of merit, $M$, $F(7, 520) = 6.47, p \leq 0.0001$. Post-hoc Bonferroni comparisons indicate that the difference in $M$ is not significant among ASEL, PL, PLmod, and MGSL. In other words, these metrics were the best predictors of equivalent annoyance for these test signals. The value of $M$ for all other metrics is significantly greater than for ASEL. The similarity between ASEL and PL is expected due to the high correlation between ASEL and PL as shown in Fig. 4.

5.2.2 Potential Imperfect Modeling of Loudness

The equality point is examined separately for each of the filtered-impulse test signals indicated with a numeral in Table 1. If the PL metric were predicting annoyance with no error, each equality point would occur at the same loudness level for all signals. As shown in Fig. 12(a), for a fifth-order low-pass filter, the equality point increases monotonically with center frequency at each listener location, indicating a systematic error in the performance of PL. This trend suggests that the weighting function may overemphasize low frequencies. The trend is also present to a lesser extent for the third-order and second-order low-pass filtered sounds. Signals pro-
duced with lower-order filters contain more high-frequency energy, as shown in 3(d), which tends to obscure the differences.

The authors developed two hypotheses to explain the observation that equality point increases with center frequency. The first hypothesis is that the PL metric is improperly modeling low-frequency annoyance. The PL metric is based on the Sone curves which are designed to model the sensitivity of the human ear. Stevens, who developed PL, extended the Sone curves in a calculation procedure called Mark VII so all curves approach a value of 160 dB at 1 Hz [11], shown in Fig. 13(a). It seems unlikely that this is a true representation of the human hearing system. As an experiment, the limiting level for the Sone curves was reduced from 160 dB to 115 dB at 1 Hz, as shown in Fig. 13(b). There was no experimental evidence to suggest a value of 115 dB. Calculating the PL using the modified Sone curves is referred to as modified PL, or PLmod, throughout this document. The loudness levels calculated by Modified PL are shown in Fig. 12(b). Interestingly, the monotonic relationship between equality point and filter center frequency observed in Fig. 12(a) does not exist for Modified PL. The authors recognize that any improvement in loudness modeling achieved by modifying the Sone curves was not captured by the figure of merit, $M$, which was used to compare the metrics. While these results suggest imperfect modeling of low-frequency loudness by PL Mark VII, results from a more focused test design would be needed to confirm this claim.

![Figure 11: Figure of merit, $M$, for each noise metric. The error bars indicate a 95% confidence interval about the mean based on a bootstrap resampling technique [19] using 10,000 resampled points.](image)

### 5.2.3 Multimodal Annoyance Response to Loudness and Vibration

A second hypothesis to explain the relationship between equality point and filter center frequency arises when annoyance is considered as a multimodal response. That is, annoyance is affected not only by sound, but also vibration, as demonstrated previously for railway noise [20] as well as helicopter over-flight noise [21]. If annoyance is influenced by both sound and vibration, then test signals producing
Figure 12: The equality point in PL for signals with fifth-order low-pass filtering. Panel (a) shows PL, and Panel (b) shows modified PL. The error bars indicate a 95% confidence interval about the mean based on the logistic regression curve fit. Horizontal lines represent the equality point of the reference signal at each location.
Figure 13: Existing [11] and modified Sone curves. Sone index is indicated on each curve.
vibration could be judged equivalent at lower sound levels than signals not causing perceptible vibration. This trend is observed in the experimental data plotted in Fig. 14.

The plot of equality points, expressed in PL dB, against $w_k$-weighted peak vibration, is shown in Fig. 14. All data points for each listener location represent levels of equivalent annoyance caused by the test signals. Each color represents a different listener location. The triangular data points represent the reference signal at each listener location. The loudness level at the equality point, on the $y$-axis, is determined from the logistic regression procedure described in Sec. 5.2.1. No signal was played exactly at the equality level so vibration values are determined using interpolation. Linear interpolation of the peak acceleration is appropriate because peak acceleration increases nearly linearly with sound level in this range. The perception threshold of 1.5 cm/s^2 [18] for full-body vibration is indicated with a dotted vertical line. At vibration levels below threshold, the PL values for subjectively equivalent sounds vary less than 5 dB in general from the reference loudness level, indicated by triangular data points. Sounds above the vibration threshold are subjectively equivalent at lower PL values, as much as 10 - 14 dB below the reference level.

All data points above the vibration perception threshold represent combinations of vibration and loudness level that may produce a multimodal annoyance response. Therefore, a best-fit line through these points estimates the trade-off between vibration and loudness level as causes of annoyance. The unshaded blue point is an outlier on the loudness level scale. It was not used in computing the best-fit line.
Table 3: Slopes of the regression lines in Fig. 14, indicating the magnitude of the trade-off between vibration and sound level. The confidence interval is calculated as the slope ± 1.96 × standard error. The units are dB/cm².

A calculation procedure was used that accounts for error in both the vibration and loudness level estimates [22]. The estimated slope of each line is given in Table 3. As indicated, the uncertainty in the measured results yields a trade-off across all three listener locations with 95% confidence of −1.80 to −4.01 dB/cm². The fact that all three correlations are negative suggests that test signals producing vibration are indeed judged equivalent at lower levels than signals not causing vibration.

6 Conclusions

A study has been conducted to evaluate the performance of noise metrics in predicting indoor annoyance to a suite of recorded sonic booms and synthesized sonic-boom-type waveforms. Most synthesized signals were designed to have the impulsive character of a sonic boom but contain energy only within a desired spectral region. This mixture of recorded and synthesized test signals provided enough spectral variation to distinguish the performance of several common noise metrics. The results indicate that no metric predicts equivalent annoyance more reliably at the 95% confidence level than Perceived Level (PL). This result agrees with previous results in the literature [5, 7], although the cited studies analyzed neither Stationary Loudness nor Time-Varying Loudness. For the sounds tested in this study, A-weighted Sound Exposure Level, Stationary Loudness, and Time-Varying Loudness all predict equivalent annoyance as well as, but not better than, PL.

Because all test signals were compared with a reference at a single level, the results do not indicate how well the metrics describe the growth rate of annoyance. Absence of a growth-rate assessment may be viewed as a disadvantage of testing many signals against a single reference sound. The advantage of the technique was an in-depth investigation of a single point on the loudness function. The error in predicting subjective annoyance using PL was not random but related to the signal’s low-frequency pass-band. Signals with lower-frequency content were found equivalently annoying at lower loudness levels. Analysis indicated two possible causes of this phenomenon. The first possible cause is improper modeling of loudness at low frequencies. When the slope of the Sone curves at low frequencies was dramatically reduced, the monotonic relationship between equality point and filter center frequency ceased to exist. The second possible cause was the presence of vibration, which added an additional mode to the annoyance response of test
subjects. For these test sounds a trade-off was found of $-1.80$ to $-4.01 \, \text{dB cm}^2 \text{s}^{-2}$ at the 95% confidence level for equivalent annoyance. In other words, above the vibration perception threshold, 1 PL dB was equivalent to between $1.80 \, \text{cm}^2 \text{s}^{-2}$ and $4.01 \, \text{cm}^2 \text{s}^{-2}$ of $w_k$-weighted peak vibration. Focused follow-on tests are needed to clarify the influence of vibration on annoyance versus possible improper modeling of low-frequency loudness.

It was observed that the exposure was not consistent across listener locations in the indoor simulator when the test signals were played over a single loudspeaker array. Differences in loudness level were due to room acoustics as well as varied receiver distance from the wall with the active array. Since the completion of this study a second loudspeaker array adjacent to a second exterior wall has been activated for use in future tests. Subsequent measurements have shown that loudness exposure is more uniform in the room with both arrays active. Seating locations for future tests will be changed as necessary to achieve the most uniform exposure. If exposure is uniform across listener locations then subjective responses can be compared directly among test locations without the need to normalize by the reference signal level.

References


Appendix A

Loudness Levels of Signals

Tables A1, A2, and A3 contain the Perceived Level [dB] and subjective judgments for all signals at listener locations 1, 2, and 3, respectively. Table A4 contains the average loudness level of each signal at the facility exterior. Acquired exterior signals were averaged across 18 microphone locations, and a fading window was applied prior to calculating loudness level, as described in Sec. 5.1.1. The descriptions of signals 1 – 15 are given in Table 1. Signals 16 – 20 were field measurements. Signal 21 is the predicted ground signature from a supersonic vehicle configuration, and signals 22 and 23 are tanh-thickened N-waves with rise times of 6 and 9 ms, respectively.
Table A1: Listener Location 1. Each of the 23 test signals was played at 6 amplitudes. In each cell, the first number is the Perceived Level [dB] of the signal. The second number is the number of responses out of 20 indicating the test signal was more annoying than the reference signal. Test signal 22 at amplitude 3 was also used as the reference sound.
<table>
<thead>
<tr>
<th>Test Signal</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>50.8—8</td>
<td>52.9—11</td>
<td>55.0—13</td>
<td>56.8—15</td>
<td>59.0—19</td>
<td>60.6—18</td>
</tr>
<tr>
<td>2</td>
<td>46.6—2</td>
<td>48.9—4</td>
<td>51.1—9</td>
<td>53.4—12</td>
<td>55.4—10</td>
<td>57.3—18</td>
</tr>
<tr>
<td>3</td>
<td>45.8—4</td>
<td>48.1—3</td>
<td>51.4—7</td>
<td>53.6—6</td>
<td>55.5—11</td>
<td>57.5—11</td>
</tr>
<tr>
<td>4</td>
<td>48.6—6</td>
<td>51.9—5</td>
<td>53.8—11</td>
<td>55.7—10</td>
<td>57.7—14</td>
<td>59.6—13</td>
</tr>
<tr>
<td>5</td>
<td>47.1—7</td>
<td>49.1—6</td>
<td>52.3—9</td>
<td>54.2—6</td>
<td>56.1—13</td>
<td>57.8—12</td>
</tr>
<tr>
<td>6</td>
<td>48.4—5</td>
<td>50.4—4</td>
<td>52.4—10</td>
<td>54.3—9</td>
<td>56.2—13</td>
<td>58.0—13</td>
</tr>
<tr>
<td>7</td>
<td>43.5—5</td>
<td>44.3—3</td>
<td>46.2—7</td>
<td>49.0—12</td>
<td>51.1—16</td>
<td>52.4—18</td>
</tr>
<tr>
<td>8</td>
<td>47.9—7</td>
<td>50.0—14</td>
<td>51.9—15</td>
<td>54.1—17</td>
<td>55.7—19</td>
<td>57.9—19</td>
</tr>
<tr>
<td>9</td>
<td>46.6—2</td>
<td>48.7—3</td>
<td>51.6—8</td>
<td>53.5—9</td>
<td>55.5—10</td>
<td>57.4—14</td>
</tr>
<tr>
<td>10</td>
<td>51.7—3</td>
<td>53.7—6</td>
<td>55.6—4</td>
<td>57.6—12</td>
<td>59.7—18</td>
<td>61.8—18</td>
</tr>
<tr>
<td>11</td>
<td>52.1—6</td>
<td>54.0—3</td>
<td>56.0—7</td>
<td>58.0—9</td>
<td>60.0—14</td>
<td>62.0—18</td>
</tr>
<tr>
<td>12</td>
<td>40.9—8</td>
<td>43.0—9</td>
<td>45.7—16</td>
<td>47.4—17</td>
<td>49.5—16</td>
<td>51.0—20</td>
</tr>
<tr>
<td>13</td>
<td>45.6—7</td>
<td>47.2—9</td>
<td>50.1—15</td>
<td>52.3—18</td>
<td>54.5—15</td>
<td>56.8—20</td>
</tr>
<tr>
<td>14</td>
<td>50.1—10</td>
<td>52.0—9</td>
<td>54.2—9</td>
<td>56.2—14</td>
<td>58.3—18</td>
<td>60.5—17</td>
</tr>
<tr>
<td>15</td>
<td>51.3—5</td>
<td>53.2—7</td>
<td>55.4—4</td>
<td>57.5—9</td>
<td>59.7—10</td>
<td>61.9—17</td>
</tr>
<tr>
<td>16</td>
<td>49.9—4</td>
<td>52.6—3</td>
<td>54.5—7</td>
<td>56.4—9</td>
<td>58.2—12</td>
<td>60.2—16</td>
</tr>
<tr>
<td>17</td>
<td>49.2—4</td>
<td>51.8—8</td>
<td>53.7—7</td>
<td>55.6—9</td>
<td>57.6—15</td>
<td>59.4—19</td>
</tr>
<tr>
<td>18</td>
<td>45.1—5</td>
<td>47.4—2</td>
<td>50.8—8</td>
<td>53.0—9</td>
<td>55.1—12</td>
<td>57.1—14</td>
</tr>
<tr>
<td>19</td>
<td>43.7—6</td>
<td>46.1—7</td>
<td>49.6—7</td>
<td>51.9—5</td>
<td>54.2—10</td>
<td>56.2—12</td>
</tr>
<tr>
<td>20</td>
<td>45.0—4</td>
<td>47.3—5</td>
<td>50.7—3</td>
<td>52.9—8</td>
<td>55.0—5</td>
<td>57.0—15</td>
</tr>
<tr>
<td>21</td>
<td>46.7—5</td>
<td>48.6—6</td>
<td>51.3—8</td>
<td>53.2—13</td>
<td>55.0—14</td>
<td>56.7—17</td>
</tr>
<tr>
<td>22</td>
<td>52.7—5</td>
<td>54.6—4</td>
<td>56.5—9</td>
<td>59.5—13</td>
<td>61.5—18</td>
<td>63.6—19</td>
</tr>
<tr>
<td>23</td>
<td>48.8—5</td>
<td>50.9—4</td>
<td>52.9—7</td>
<td>55.8—17</td>
<td>57.6—14</td>
<td>59.5—20</td>
</tr>
</tbody>
</table>

Table A2: Listener Location 2. Each of the 23 test signals was played at 6 amplitudes. In each cell, the first number is the Perceived Level [dB] of the signal. The second number is the number of responses out of 20 indicating the test signal was more annoying than the reference signal. Test signal 22 at amplitude 3 was also used as the reference sound.
Table A3: Listener Location 3. Each of the 23 test signals was played at 6 amplitudes. In each cell, the first number is the Perceived Level [dB] of the signal. The second number is the number of responses out of 20 indicating the test signal was more annoying than the reference signal. Test signal 22 at amplitude 3 was also used as the reference sound.
Table A4: Perceived Level [dB] of test signals averaged across 18 microphones at facility exterior. Test signal 22 at amplitude 3 was also used as the reference sound.
Appendix B

Signal Processing Flow Chart

1. "inflection points, $f_s$"
2. " BoomGen"
3. " $t_{\text{rise}} - t_{10-90} \leq tol1$"
4. "adjust $p_{\text{max}}$"
5. "normalize max value to 1"
6. " $PL - PL_{\text{desired}} \leq tol2$"
7. "6-Hz HPF"
8. "2+ signals?"
9. "combine signals"
10. "time delay, second waveform, second multiplier"
11. "normalize to prevent quantization noise"
12. "normalized equalized waveform + final multiplier"
• **Create a tanh-thickened sonic boom waveform (A-E):** The inflection points of the desired boom signature, as ordered pairs of time and overpressure, as well as the sampling frequency $f_s$ (A), are entered into the Matlab code BoomGen (B). BoomGen produces a sampled sonic boom waveform of desired peak overpressure with tanh thickening. The rise time of a tanh-thickened waveform cannot be specified directly, but a desired rise time can be achieved by iteratively adjusting $p_{\text{max}}$. The difference between the rise time of the resulting waveform, $t_{10-90}$, and a desired rise time, $t_{\text{rise}}$, is compared with a user-specified tolerance, $tol_1$ (C). If the resulting rise time is within the tolerance of the desired rise time, the waveform is normalized to a maximum value of unity (E). Otherwise, the maximum overpressure is adjusted (D), and the waveform is regenerated in BoomGen (B).

• **Scale the waveform to a desired loudness level (F-I):** If no tanh-thickened waveform is desired, then the processing begins with this stage. The waveform is entered into the Matlab code NOM (F), which filters the waveform into third-octave band sound pressure levels and calculates noise metrics. A multiplier (G) is applied to the waveform, which converts the waveform unit to pressure in pascals at the exterior array microphones. The Perceived Loudness (PL) of the waveform is calculated, and the result is compared with the desired PL value. If the difference between the desired and predicted PL values is less than a user-specified tolerance, $tol_2$ (H), the waveform and the multiplier are passed on. Otherwise, the multiplier is adjusted iteratively (I) until the predicted PL matches the desired PL within the tolerance, $tol_2$.

• **High-pass filter (J):** The waveform is filtered by a 6-Hz high-pass filter (J) to remove spectral content that cannot be reproduced in the facility.

• **Combine multiple signals (K-M):** Applying the equalization filter (N) adds several seconds to the length of each signal. If multiple signals need to be played in quick succession, the signals are combined into a single waveform prior to equalization (L). In the combined waveform, the component signals are separated by a short, user-specified time delay. Each waveform originally had its own multiplier. The greatest multiplier is retained, and the waveform originally associated with the lesser multiplier is scaled in amplitude by the ratio of the lesser multiplier to the greater multiplier. The result of this manipulation is that each component waveform is scaled to its desired amplitude although the entire waveform has only one multiplier.

• **Apply equalization filter (N):** The equalization filter (N) is applied to account for unavoidable filtering of the desired signal by the sound reproduction chain.

• **Normalize to maximize dynamic range (O):** Although this step (O) was not used in the current study, it will be used in future studies. The equalization filter reduces the dynamic range of the waveform considerably. For some waveforms the reduced dynamic range produces audible quantization noise. The quantization noise does not typically transmit into the facility, but
it can be audible at the facility exterior. To solve this problem, the waveform is renormalized \((O)\) to a maximum value of 0.99, as shown in Eq. B1. A value of 0.99 is chosen to ensure no clipping. The multiplier is also adjusted, as shown in Eq. B2, so the resulting overpressure remains unchanged. As a double check of the scaling, when the normalized waveform is multiplied by the final multiplier, the result is the same as multiplying the equalized waveform by the prior multiplier.

\[
\text{normalized waveform} = \frac{0.99 \times \text{waveform}}{\max(\text{|waveform|})} \quad \text{(B1)}
\]

\[
\text{final multiplier} = \frac{\text{multiplier} \times \max(\text{|waveform|})}{0.99} \quad \text{(B2)}
\]

- **Output:** The processing technique produces an equalized waveform with an associated multiplier. The multiplier corresponds approximately to the waveform’s peak overpressure in pascals at the facility exterior.
A sonic-boom simulator at NASA Langley Research Center has been constructed to research the indoor human response to low-amplitude sonic booms. The research goal is the development of a psychoacoustic model for individual sonic booms to be validated by future community studies. The study in this report assessed the suitability of existing noise metrics for predicting indoor annoyance. The test signals included a wide range of synthesized and recorded sonic-boom waveforms. Results indicated that no noise metric predicts indoor annoyance to sonic-boom sounds better than Perceived Level, PL. During the study it became apparent that structural vibrations induced by the test signals were contributing to annoyance, so the relationship between sound and vibration at levels of equivalent annoyance has been quantified.