NASA/TM–2013–217975

Laboratory Headphone Studies of Human Response to Low-Amplitude Sonic Booms and Rattle Heard Indoors

Alexandra Loubeau, Brenda M. Sullivan, Jacob Klos, and Jonathan Rathsam
Langley Research Center, Hampton, Virginia

Joseph R. Gavin
Gulfstream Aerospace Corporation, Savannah, Georgia

March 2013

NASA STI Program . . . in Profile

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 organizing and publishing research results, distributing specialized research announcements and feeds, providing information desk and personal search support, and enabling data exchange services.

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 to help@sti.nasa.gov

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

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

- Write to:
  STI Information Desk
  NASA Center for AeroSpace Information
  7115 Standard Drive
  Hanover, MD 21076-1320

NASA/TM–2013–217975

Laboratory Headphone Studies of Human Response to Low-Amplitude Sonic Booms and Rattle Heard Indoors

Alexandra Loubeau, Brenda M. Sullivan, Jacob Klos, and Jonathan Rathsam
Langley Research Center, Hampton, Virginia

Joseph R. Gavin
Gulfstream Aerospace Corporation, Savannah, Georgia

National Aeronautics and
Space Administration

Langley Research Center
Hampton, Virginia 23681-2199

March 2013

Acknowledgments

The authors would like to thank Patricia Davies, Andrew Marshall, Clyde G. Medley, Jr., and Regina D. Johns for their contributions to this research.

Available from:

NASA Center for AeroSpace Information
7115 Standard Drive
Hanover, MD 21076-1320
443-757-5802

Abstract

Human response to sonic booms heard indoors is affected by the generation of contact-induced rattle noise. The annoyance caused by sonic boom-induced rattle noise was studied in a series of psychoacoustics tests. In order to study response to effects beyond that of loudness, sounds were normalized to the same Perceived Level (PL) or set of PL in each test. Stimuli were divided into three categories and presented in three different studies: isolated rattles at the same calculated PL, sonic booms combined with rattles with the mixed sound at a single PL, and sonic booms combined with rattles with the mixed sound at three different PL. The low-amplitude sonic booms, both measured and synthesized, were filtered to simulate presentation inside structures with different transmission and reverberation properties. The rattle sounds due to sonic booms or direct impulsive mechanical loading on structures and objects were recorded in a residential home. Subjects listened to sounds over headphones and were asked to judge the level of a number of factors, including annoyance. Annoyance to different rattles was shown to vary significantly according to rattle object size, despite having set all rattle sounds to the same PL value. In addition, the combination of low-amplitude sonic booms and rattles can be more annoying than the sonic boom alone. Correlations of annoyance with metrics did not identify a sound quality metric capable of describing annoyance to rattle sounds beyond that explained by loudness level. Correlations and regression analyses for the combined sonic boom and rattle sounds identified the Moore and Glasberg Stationary Loudness (MGSL) metric as a primary predictor of annoyance for the tested sounds, despite its intended use for steady, not transient, sounds. Multiple linear regression models were developed to describe annoyance to the tested sounds, and simplifications for applicability to a wider range of sounds are presented.

Contents

1 Introduction

Civil supersonic flight over land is currently prohibited in the United States [1] because of the annoyance caused by sonic booms. New low-boom aircraft designs, however, aim to reduce the sonic boom noise to a level that is perceived to be acceptable. In order to assess the effectiveness of these low-boom designs, laboratory and field studies of human response to these low-amplitude sonic booms are needed. In the absence of low-boom aircraft, surrogate aircraft and simulation techniques are being used to advance the understanding of perception of low booms.

The metric Perceived Level (PL) has been found to be the best predictor of human annoyance to sonic booms, both outdoors and indoors [36, 37]. Although other metrics also correlate highly with subjective annoyance, PL most consistently accounts for loudness effects outdoors and additional annoyance effects both outdoors and indoors [36]. These results have been gathered for isolated sonic booms without the presence of contact-induced rattle noise, which is often caused indoors by sonic booms when they impact buildings. Comparisons of outdoor and indoor reactions in field studies have identified differences in perception, and rattle noise has been targeted as one likely contributor to elevated annoyance to booms experienced indoors [16, 56].

1.1 Background on Human Response to Rattles

Human response to impulsive noises, such as sonic booms, heard indoors is affected by the generation of contact-induced rattle noise. Understanding this indoor human response is important to determine acceptability of low-amplitude sonic booms from proposed low-boom aircraft designs. Therefore a facility at NASA Langley Research Center, the Interior Effects Room, has been constructed for subjective tests of sonic booms heard indoors [31, 32]. Rattle is one of the key parameters that affects human response indoors that will be investigated in this facility.

Before beginning tests in the facility, a better understanding of rattle was desired to aid in test design. Previous rattle tests can be mainly categorized into three groups. First, there have been several community studies of sonic booms and other impulsive noises. Secondly, there have also been controlled field tests where subjects were asked to rate annoyance to specific sonic booms from real flyovers. Lastly, there have been many laboratory tests of human response to sonic booms using simulators, where the rattle was sometimes controlled.

Community studies indicate that rattle and vibration are important to perception of sonic booms. Two field surveys in the 1960s of communities exposed to sonic booms over less than one year identified rattling and vibration as undesired effects that increased annoyance for booms experienced indoors [5,6,39,40]. Another field survey study on long-term sonic boom exposure also cited vibration and rattle as major contributors to disturbance or annoyance that may have caused mean annoyance to be higher indoors than outdoors [16]. The effects of structural vibration can be perceived through visual, auditory, and tactile cues, such as seeing windows moving, hearing windows rattling, and feeling the floor vibrating. Frequently, these effects are grouped together in surveys, and it is not possible to separate reactions resulting from each type of cue. However, Schomer [47] analyzed several studies of human response to large-amplitude impulsive sounds, both in the field and in the laboratory, and found that vibration was not a significant contributor to human response.

Only perception of the impulsive sound itself and secondary rattle noises determined the human response.

Controlled field studies of human response to sonic booms, other impulsive noises, and aircraft noise have investigated differences in annoyance between indoor and outdoor environments. An indoor "penalty" sometimes can be deduced from their data that quantifies the difference in level between outdoor and indoor listening that results in the same annoyance judgment. Johnson and Robinson [29] found an indoor penalty of 5 phons for a variety of sounds, including sonic booms, aircraft flyover noise, and explosion noise. Kryter *et al.* [34] investigated acceptability of sonic booms and aircraft flyover noise and were able to decouple the feeling of vibration from rattle by using a vibration isolator under the seating area for half the subjects. They found that vibration itself did not significantly affect acceptability responses, but the presence of secondary rattle sounds "substantially" affected acceptability ratings for the indoor environment. Schomer and Neathammer [49] found a rattle SEL penalty of $12 - 20$ dB for naturally occurring rattles in homes in response to actual helicopter flyover noise. A more recent study by Sullivan *et al.* [56] found that annoyance ratings to sonic booms were the same indoors and outdoors on average. However, a post-test questionnaire revealed that the subjects recalled feeling more annoyed indoors, due to rattle sounds, house vibrations, or startle effects.

Subjective laboratory tests of sonic booms and other sounds have been used to explore particular effects in an even more controlled environment. Pearsons and Kryter [44] used simulated outdoor booms, recorded indoor booms, and recorded aircraft flyovers in the laboratory to assess differences in acceptability for indoor and outdoor sounds. They found a 13 dB rattle penalty in Overall Sound Pressure Level (OASPL) for a window added to the simulator that rattled in response to the booms. This window rattle was not controlled. Pearsons *et al.* [45] presented simulated sonic booms and recorded transportation sounds to subjects both outside and inside a house simulator that included dishes that rattled. A 13 dB ASEL indoor boom penalty was found, although the rattle was not controlled and no quantitative data on the rattles was reported. Schomer and Averbuch [48] simulated blast sounds that impacted a test house that was outfitted with rattling windows, lights, bric-a-brac, doors, etc. A rattle penalty ranging from 6 to 13 dB in ASEL, depending on blast level, was reported. Introduction of a recorded rattle with simulated indoor booms was used by Fidell *et al.* [15], who found a rattle penalty of 5 dB for boom annoyance. In contrast, Cawthorn *et al.* [8] found no rattle penalty for low-level controlled recorded rattle sounds introduced in a living room simulator with recorded aircraft flyover noise.

The large range of rattle penalties reported in these studies ($0 - 20$ dB) is potentially due to differences in the character of the rattle sound sources and their levels, which often were not controlled. Many studies did not document the character or loudness of the rattle sounds, or other possible visual or tactile cues present, which makes it difficult to investigate causes of the disparity in reported rattle penalties. In addition, different psychophysical methods were employed in these tests, making it difficult to directly compare the different studies. Although results from these studies are inconsistent, the majority of the studies concluded that rattle has a measurable effect on human annoyance to sonic booms and other sounds.

1.2 New Boom and Rattle Test

The Gulfstream NASA Boom Rattle Test (GNBRT) performed at NASA Langley Research Center was designed to shed insight on human response to sonic-boom-induced rattle noise indoors for low-amplitude booms. Because loudness is a major contributor to human annoyance to noise, limiting the influence of loudness was desired to study the effects of other psychoacoustic factors. In order to control the effects of loudness, sounds were equalized to a fixed Perceived Level (PL) [55] in each test. The resulting human response would then be attributed to factors other than loudness. Stimuli were divided into three categories and presented in three different studies: isolated rattles at the same calculated PL, sonic booms combined with rattles and presented at the same calculated PL, and sonic booms combined with rattles and presented at three different PL.

2 Test Sounds

The sonic boom and rattle sounds, presented to subjects over high-fidelity headphones, were obtained from field measurements, laboratory measurements, and simulations. To enable control over the booms and rattles presented to subjects, separate rattle stimuli and boom stimuli were generated and then mixed together to simulate the indoor soundscape for a home ensonified by a sonic boom. The sources of these rattle and boom stimuli, and how they were mixed together, are discussed in the following sections.

2.1 Rattle Sounds

The rattle sounds were recorded binaurally in residential houses, which were subjected either to actual sonic booms or to direct impulsive mechanical loading on structures and objects, and in the Gulfstream Acoustic Test Facility (ATF), a transmission loss facility consisting of a hemi-anechoic room and a reverberation chamber separated by a transmission loss window. The sonic-boom-induced rattle sounds were recorded by NASA in a house on Edwards Air Force Base (EAFB) [30]. This house, shown in Fig. 1, is of a construction typical of homes in the American Southwest, with wood framing, plywood sheathing, metal lath, and stucco exterior. The low-amplitude sonic booms that ensonified the house were generated by F-18 aircraft executing a low-boom dive maneuver [20,21]. In order to generate window rattle sounds in the laboratory, the sonic booms measured outdoors away from the housing area [30] were played back in the hemi-anechoic chamber of the ATF, with a residential window mounted in the transmission loss window separating the hemi-anechoic and reverberant chambers; microphones in the reverberation room captured the resulting window rattle sounds, as shown in Fig. 2. In addition, recordings of window rattle resulting from playback of synthesized sonic boom waveforms were performed in the ATF, yielding a larger range of rattles than had been measured in the field. The remaining rattle sounds that were recorded were created by impulsive mechanical loading on various structures and objects in another home of typical American construction. All the rattle recordings were high-pass filtered to remove the boom or the impulsive source, while retaining the high-frequency rattle noise. In total, forty different binaural rattles were selected for use in the present studies from the recordings described above.

Rattles are complex sounds, with respect to both temporal and spectral components.

Figure 1. Exterior view of house on Edwards Air Force Base (EAFB) where rattles caused by low-amplitude sonic booms were recorded.

Figure 2. Recording of window rattles in the Gulfstream Acoustic Test Facility (ATF).

The example spectrograms in Figs. 3 and 4, which show frequency content of two representative rattles vs. time, demonstrate the complexity and diversity of these sounds. In Fig. 3, some high-frequency tonal components are evident. This rattle is of wine glasses clinking. In Fig. 4, the spectrogram shows impulsive behavior and more low-frequency content. This rattle is a recording of nuts and bolts rolling around in a metal jug. It is because of this complexity that it has been difficult to define rattles and to find a metric that can describe them.

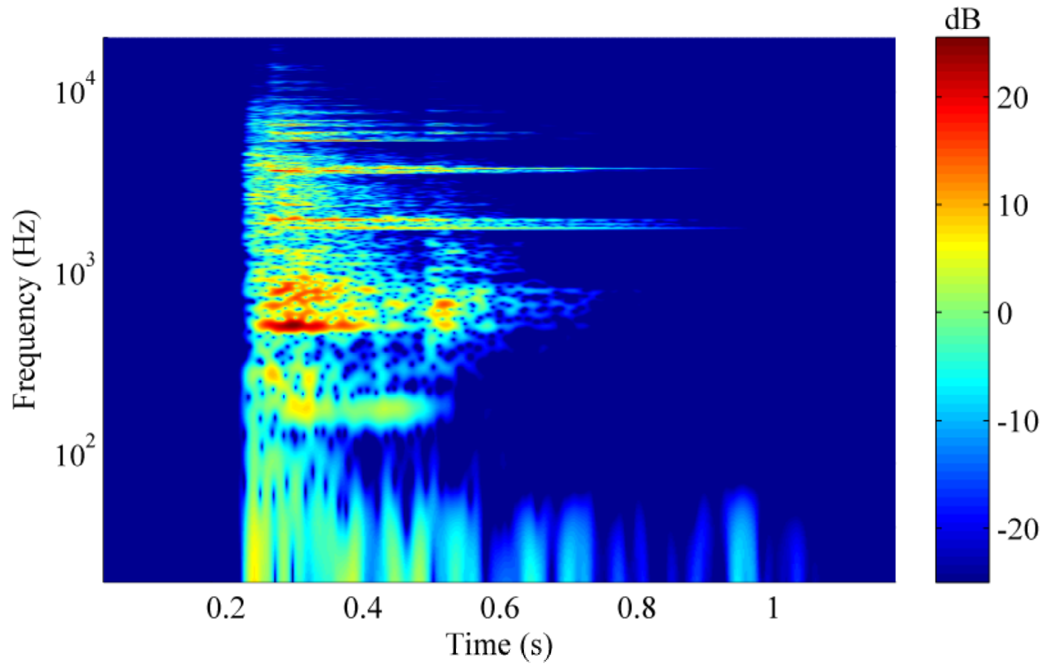

Figure 3. Example spectrogram of the Power Spectral Density (dB *re* $(20\mu\text{Pa})^2/\text{Hz}$) of a rattle sound with tonal components.

The rattle sounds were equalized to fixed PL values. The playback amplitude of the waveforms was determined through an iterative procedure described below.

- The sound was played back through a pair of headphones mounted on an artificial binaural head and measured by the microphones in the binaural head.
- The PL values for the left and right ears were calculated, and the decibel average of the two values was compared to the target PL.
- If the measured PL was within ± 0.2 dB of the target, the amplitude was retained as the final playback amplitude[1]. If the measured PL differed from the target by more than the tolerance value, the amplitude was adjusted, and the procedure was repeated.

[1] The equalization procedure was followed for only one pair of headphones. It was later found that the tolerance of ± 0.2 dB resulted in a maximum difference of approximately ± 1 dB across the different pairs of headphones.

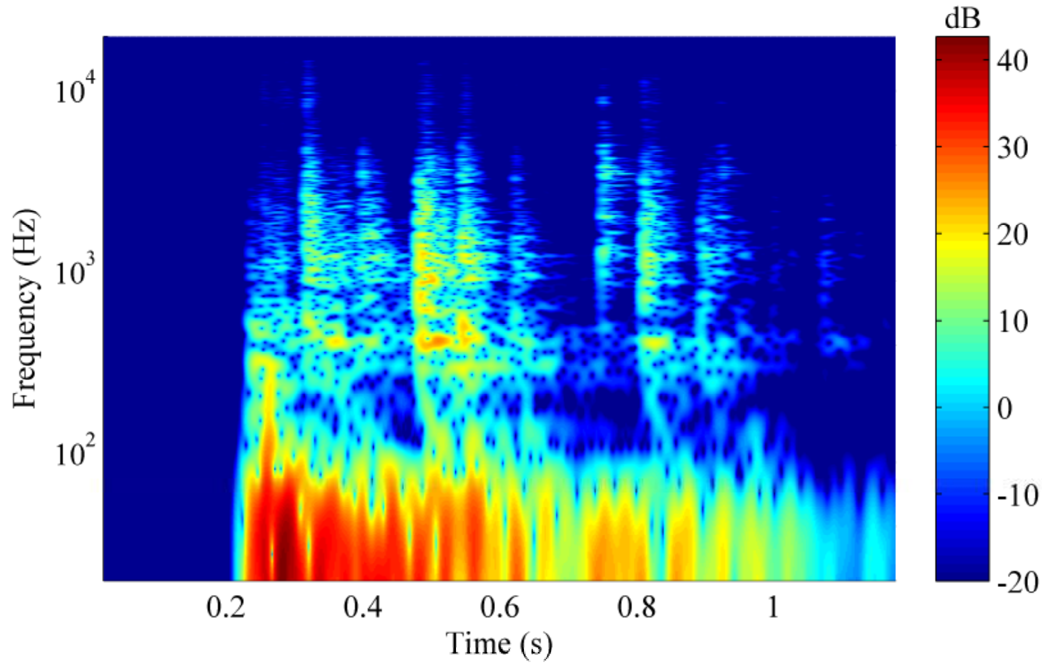

Figure 4. Example spectrogram of the Power Spectral Density (dB *re* $(20\mu\text{Pa})^2/\text{Hz}$) of a rattle sound with impulsive characteristics.

2.2 Indoor Sonic Booms

Outdoor sonic booms, both measured and synthesized, were filtered to simulate indoor booms that would result from transmission into different structure types. In addition, filters to account for sound radiation in rooms with different reverberation properties were applied. The measured outdoor sonic booms were recorded during the same tests at EAFB as the rattles. Synthesized outdoor booms considered for this study include tanh-thickened N-wave, ramp, flattop, and front-shock-minimized booms [35]. Several methods for filtering booms were investigated, including deconvolution, an empirical method, and a semi-empirical method.

In the deconvolution method, a digital transfer function is extracted by deconvolving the outdoor signature from an indoor recording. This method can potentially cover the full audio bandwidth, with all complexities inherently included. Deconvolution, however, requires new sonic boom measurements with different structures for modeling of each combination of construction, room size, and room absorption. In addition, the outdoor signature is idealized at a single point, and zeros in the spectrum cause numerical problems. Another weakness of this method is that the indoor recording must be rattle free, i. e. the transmission must be linear.

The empirical method employs low-dimensional FIR filters to create transmission loss (TL) spectra, and balloon pop measurements are used to create the room impulse response to characterize reverberation. This method separates the TL effect from reverberation and enables control of the frequency dependence of both effects. In contrast with the deconvolution method, the empirical method can be used to simulate transmission into different structure types. Additionally, the variety of reverberation models is limited only by the range of room sizes and room absorption tested in impulse response measurements. Unfortunately, it is difficult to measure high-quality room impulse response data at low frequencies that are important for sonic booms because they are often below the first room mode.

The semi-empirical filter method combines the empirical method for TL with numerical processing to handle reverberation filters and experimental Head-Related Transfer Functions (HRTF). The strengths of this method are that existing room acoustics models can already handle arbitrary geometry and absorption for creation of a variety of reverberation models. Arrays of numerical sources can be created to simulate a distributed source, such as a wall or window, by adjusting the phase. The main drawback of this method is that modeling of reverberant decay is stochastic, which may reduce the realism of synthesized signals.

The semi-empirical approach was chosen and applied to the booms to synthesize transmission through partitions of heavy, moderate, and light TL into rooms of small, medium, and large size. Filtering five booms with three TL options and three room size options generated 45 separate booms. Audition of these booms to determine realism and diversity of sounds led to a down-selection of four filtered booms. One EAFB recorded boom was selected with three different filters applied: large room with light TL, large room with moderate TL, and small room with moderate TL. A second synthesized ramp boom was chosen with filtering to simulate a small room with moderate TL.

2.3 Mixing Boom and Rattle Sounds

Three tests were performed in this study series. In the first test, subjects were presented with rattle sounds in the absence of indoor booms. In the second and third tests, controlled mixtures of indoor boom and rattle sounds were presented. The simulated indoor sonic booms and recorded rattle stimuli were combined to create the illusion that the rattles are caused by the booms. The time of maximum loudness was calculated for the boom and rattle sounds using the MGTVL metric (see Sec. 7.1). The rattle was shifted in time so that its maximum loudness occurs 10 ms after the time of the boom's maximum loudness.

For a given PL value, the relative levels of the boom and rattle sounds within a combination were varied. These relative variations were employed to determine whether the mixed boom and rattle sound is more annoying than the boom alone, at what level of rattle this increased annoyance may occur, and which rattle level results in the highest annoyance rating. Combinations range from the boom being the only audible sound to the rattle being the only sound, with seven intermediary combinations in Test 2 and five intermediary combinations in Test 3. The combinations are denoted by a dB decrease in the rattle level relative to the rattle only level. An iterative procedure similar to that described in Sec. 2.1 was followed to equalize the mixed sounds to fixed PL values in the second and third tests.

- The playback amplitude of the isolated rattle sound was determined for the target PL (see Sec. 2.1).

- The rattle sound level was decreased by a fixed amount (see Secs. 5.1 and 6.1) and then combined with a sonic boom during playback through a pair of headphones mounted on a binaural head.

- The mixed sound was measured with the binaural head, and the decibel average of the left and right PL values was compared to the target PL.
- If the measured PL was within a tolerance of ± 0.2 dB from the target, the boom amplitude was retained as the final boom playback amplitude. If the measured PL differed from the target by more than the tolerance value, the boom amplitude was adjusted (keeping the rattle amplitude constant), and the procedure was repeated.

3 Test Setup

Subjects listened to sounds over high-fidelity headphones and were asked to judge the level of annoyance or of other factors, depending on the test. The tests were conducted in a small anechoic chamber at NASA Langley Research Center. This facility, shown in Fig. 5, is structurally isolated from the rest of the building, thereby creating a quiet environment for the headphone testing. The tests were conducted with groups of three or four subjects

Figure 5. Headphone test setup in a small anechoic chamber at NASA Langley Research Center.

at a time. A playback and recording system was developed that uses a server computer for automated playlist playback and for prompting of subjects. The subjects use client netbook computers to make their judgments, which are sent back to the server in real time.

Ideally, a playback system capable of reproducing the full frequency content of sonic booms is desired. The headphone playback system used in the tests is capable of accurate sound reproduction from 10 Hz to 10 kHz, which is sufficient for accurate playback of rattle sounds. Signals were low-pass filtered to eliminate high-frequency ambient noise and noise associated with dynamic range limitations. Although the system cannot reproduce all the low-frequency energy of sonic booms, it does have a better frequency range than most headphone systems and faithfully reproduces frequencies in the audible range. To determine the importance of low-frequency energy for boom playback, informal listening tests were conducted using a high-pass filtered signal (at 20 Hz) for the headphones with and without the addition of a subwoofer for reproducing the low frequencies. The differences in low-frequency energy between the two configurations was not found to be perceptible. Thus the added complexity of subwoofer augmentation was considered unnecessary, and the headphone system with filtered stimuli was deemed acceptable for the boom and rattle tests.

Test subjects were recruited from the community and were compensated for their participation. Subjects received an audiometric test beforehand to confirm that their hearing was within 40 dB of reference hearing threshold levels [26]. See App. A for summary statistics of participant gender and age and the number of participants for each test.

Each test began with a familiarization session, where subjects listened to a few sounds to introduce the types of sounds they would hear in the test. Then subjects completed a practice session, where they became familiar with the test procedure and with entering judgments on the computers. Finally, subjects participated in the actual test. A random time delay was introduced between sounds in an attempt to avoid anticipation and to maximize startle. Sounds were presented in a different random order for each group of subjects.

4 Test 1: Subjective Tests of Rattles

The first test was developed to investigate human response to rattle sounds in the absence of a sonic boom to see if people respond differently to rattles of differing character. As described in Sec. 2.1, a variety of binaural rattle sounds were collected to explore the effects of a range in sound character. To reduce the effect of loudness on annoyance judgments, the amplitude of each rattle in Test 1 was adjusted to produce a uniform calculated Perceived Level (PL) of 70 ± 1 dB. Rattles were selected to emphasize different sound qualities based on the calculation of the sound quality metrics listed in Table D2.

Three separate subtests of subjective response to rattle were conducted with different methodologies: paired comparison, category line scaling, and semantic differential. Each test method was chosen to gather different data about the effects of rattle, but it was also desired to be able to compare the results from the different methods at the conclusion of the test series. Each subtest had 24 listeners who participated in groups of three.

4.1 Paired Comparison Subtest

The paired comparison (PC) subtest consisted of nine rattle sounds of equal PL, presented in pairs, and listeners were asked to judge which sound was more annoying in each pair. An example of the judgment screen presented to subjects is included in Fig. 6. The nine rattles were presented in all possible pair combinations, resulting in 36 pairs ($t(t-1)/2 = 36$, where $t = 9$). In addition, the ordering of sounds in each pair was also reversed, resulting in a total of 72 pairs presented to the subjects for judgment. This PC method [10,19] allows for a ranking of signals in terms of increasing annoyance.

The resulting proportion matrix in Table 1 includes the probability results for each sound pairing. This matrix presents the probability of a row element being chosen as more annoying than a column element. A value of 0.50 is entered where a sound would be compared to itself, denoting that the estimated annoyance would be the same. The values of each pairing and reverse pairing add up to 1. For example, the probability of rattle

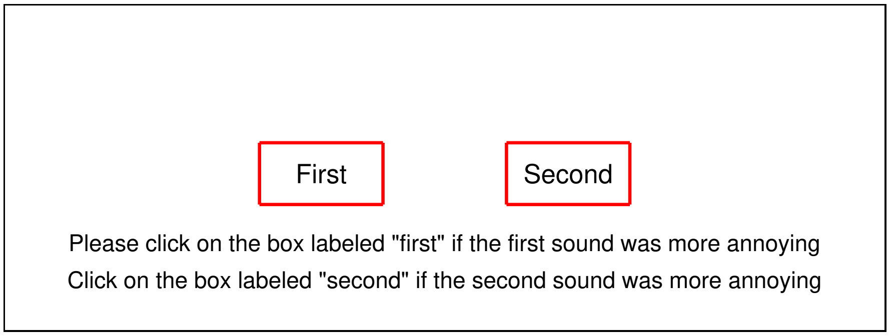

Figure 6. Example judgment screen for paired comparison subtest.

1 being more annoying than rattle 2 is 0.29, and the probability of rattle 2 being more annoying than rattle 1 is 0.71; these two values add up to 1 $(0.29 + 0.71 = 1)$. The score in the rightmost column is an addition of the probabilities for each sound in a row, and this score is used to rank the sounds in terms of annoyance, from the smallest score for the least annoying sound to the largest score denoting the most annoying sound. A short description of each rattle is included in each row. In general, small object rattles were judged to be less annoying than the rattle of structural objects, such as doors and windows.

Rattle Sound	1	2	3	4	5	6	7	8	9	Score
1. Wall art	0.50	0.29	0.35	0.13	0.10	0.17	0.00	0.04	0.10	1.68
2. Candle globe	0.71	0.50	0.42	0.33	0.33	0.29	0.19	0.17	0.19	3.13
3. Wine glass	0.65	0.58	0.50	0.44	0.52	0.38	0.33	0.31	0.33	4.05
4. Window	0.88	0.67	0.56	0.50	0.40	0.44	0.27	0.35	0.31	4.38
5. Door	0.90	0.67	0.48	0.60	0.50	0.50	0.31	0.40	0.27	4.63
6. Garage door	0.83	0.71	0.63	0.56	0.50	0.50	0.50	0.40	0.38	5.00
7. Bedroom door	1.00	0.81	0.67	0.73	0.69	0.50	0.50	0.44	0.46	5.79
8. Ceiling fan	0.96	0.83	0.69	0.65	0.60	0.60	0.56	0.50	0.46	5.85
9. Window	0.90	0.81	0.67	0.69	0.73	0.63	0.54	0.54	0.50	6.00

Table 1. Proportion matrix for nine rattle sounds judged in terms of annoyance in paired comparison subtest.

Once the ranking of sounds is determined, it is desired to know whether the differences in annoyance are statistically significant. Firstly, an overall test of equality [10] is performed for the desired significance level of 0.05. Given the score of the ith sound, $a_i = n(\text{score} - 0.5)$, the standardized sum of squares is given by

$$D_n = \frac{4\left[\sum_{i=1}^{t} a_i^2 - \frac{1}{4}tn^2\,(t-1)^2\right]}{nt}, \tag{1}$$

where $n = 24$ is the number of subjects and $t = 9$ is the number of sounds. The value of $D_n = 171.5$ is compared to the 5% significance level $(\alpha = 0.05)$ of the chi-square distribution with $t - 1$ degrees of freedom $(\chi^2(8) = 15.5)$, and it is found that a significant difference exists between the annoyance scores.

Two additional tests of statistical significance are performed on this data [10], and it is found that not all annoyance scores are significantly different from one another. The least significance difference method involves calculation of a critical value $m_c = 1.96\sqrt{\frac{1}{2}nt} + 0.5$, rounded to the next greatest integer, for a two-sided test on each pair of scores. The difference in a_i scores for each pair must be greater than or equal to m_c to be declared significantly different. This method groups sounds 3, 4, and 5; 4, 5, and 6; 6, 7, and 8; and 7, 8, and 9. Each of these groups represents annoyance scores that are not significantly different from each other. Although the analysis groups different rattle sounds, there are still significant differences between annoyance to small object rattles from an art frame or a candle globe and annoyance to larger object rattles from doors or windows.

A more conservative multiple comparison range test is performed that also involves calculation of a critical value for significant differences. The upper α significance point of the W_t distribution is found, and the critical value $R^* = 0.5W_{t,\alpha}\sqrt{nt}+0.25$ is rounded to the next greatest integer, R^+. The value of R^+ is found to be less than the factor $n(t-1)-0.5n$, so no further calculations are needed. The difference in a_i scores for each pair must be greater than or equal to R^+ to be declared significantly different. This method results in larger groupings of sounds: 2, 3, and 4; 3, 4, 5, and 6; and 5, 6, 7, 8, and 9. Even with this conservative method, it is certain that the score for sound 1 is significantly different from any other score. Sound 1, rattle from a wooden art frame hanging on the wall, is judged much less annoying than all the other sounds, and particularly less annoying than sound 9, rattle from a bedroom window.

4.2 Category Line Scaling Subtest

The category line scaling (CS) subtest consisted of 40 rattle sounds of equal PL, and listeners were asked to judge their annoyance to each rattle sound individually on a scale from Slightly Annoying to Very Annoying (see Fig. 7). Listeners were instructed to mark their annoyance judgment anywhere along the line, thereby using a continuous line instead of separated categories. The 40 rattle sounds in this subtest included the nine sounds used in the paired comparison subtest. Of the 40 rattle sounds, 35 sounds were presented once to the listeners and the remaining five sounds were presented twice, for a total of 45 judgments. The randomized playlist was then repeated in reverse order, for a total of 90 judgments per subject.

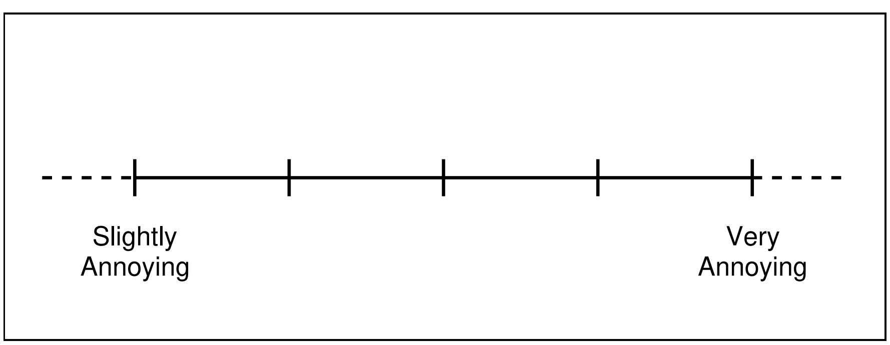

Figure 7. Example judgment screen for category line scaling subtest.

A General Linear Model (GLM) Repeated Measures analysis is performed for the judgments from the five sounds presented to each listener four times to test the effects of repeats on listener responses. Mauchly's test of sphericity gives a low significance value of 0.002, which indicates that the repetition data set is small and sphericity can not be assumed. Applying the Greenhouse-Geisser correction results in $F(2.718, 323.425) = 0.52$ with a significance value (p-value) of 0.651. This large p-value demonstrates that there are no significant effects of repeats, and subject responses are reliable.

Results from the CS subtest lead to a ranking of the 40 rattle sounds on a scale from 1 to 5, representing an increase in annoyance from "slightly" to "very". The mean annoyance and 95% confidence interval computed for each sound are arranged in order of increasing mean annoyance and are presented in Fig. 8. The means range from 2.1 to 3.6, which crosses the middle of the scale. It is shown that there is a difference in annoyance between rattles, even when the calculated PL is the same for each rattle. This test thus exposes variance in annoyance not accounted for by the PL metric.

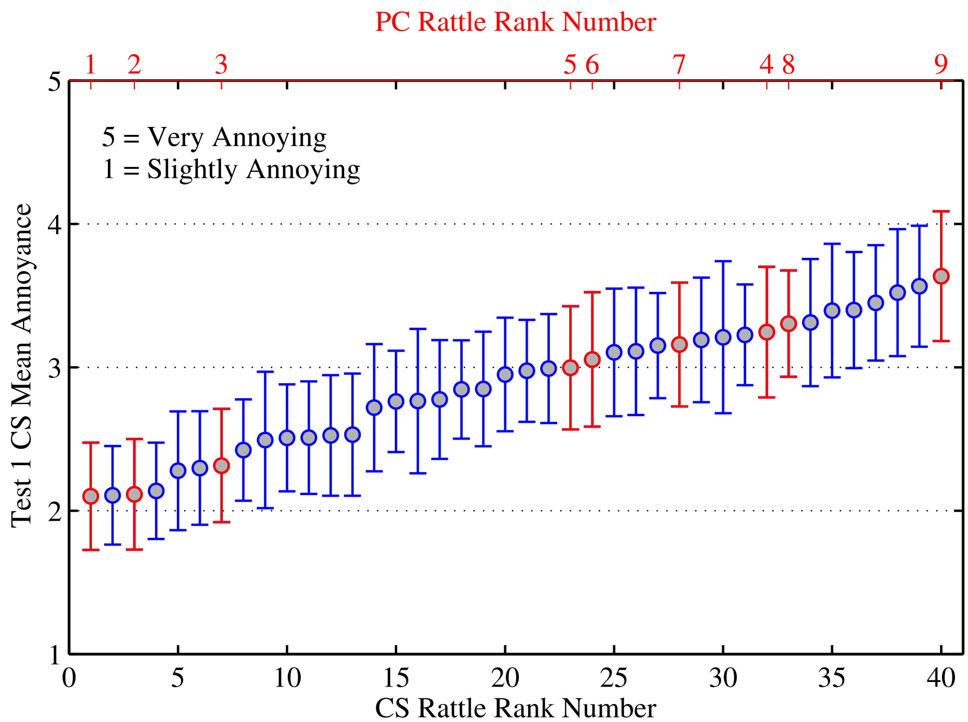

Figure 8. Mean annoyance and 95% confidence intervals for the category line scaling subtest, arranged in rank order from Slightly to Very Annoying. The nine sounds also used in the paired comparison test are highlighted in red, and the top axis presents the corresponding rank numbering from the paired comparison subtest.

The nine sounds also used in the paired comparison test are highlighted in red in Fig. 8. As shown in the top axis, the ranking of these nine sounds in order of increasing annoyance matches between the two subtests, with the exception of PC rattle 4. The PC rattle 4 (window) is ranked more annoying in the CS subtest; the CS mean score places rattle 4 between rattles 7 (bedroom door) and 8 (ceiling fan). The least and most annoying sounds from the PC subtest are also found to be the least and most annoying sounds, respectively, in the CS subtest, despite the inclusion of many more sounds in this second subtest.

A Repeated Measures Analysis of Variance (ANOVA) test is performed on the CS results to determine whether the annoyance means of the rattle sounds are statistically different. The F-test of difference in means with a Greenhouse-Geisser correction for sphericity violation gives $F(5.594, 128.663) = 10.085$ and $p < 0.001$, which shows that the mean annoyance does vary with rattle sound. Consequently there is less than a 0.1% probability that the differences are due to chance. This general result is followed by an analysis of pairwise comparisons that determine which rattles differ on annoyance at the 5% significance level. The mean annoyance to each rattle sound is compared to every other rattle sound. Three examples of these pairwise comparisons are shown in Fig. 9. Figure 9(a) shows the mean and 95% confidence intervals with the least annoying rattle circled in red. All rattle results shown in gray do not differ significantly in annoyance from the least annoying rattle, but the blue rattles are significantly different. These significantly different rattles are mostly ranked 33rd and above. In Fig. 9(b), the annoyance to the middle rattle (rank 20) is shown to not be significantly different from any other rattle. In Fig. 9(c), the most annoying rattle is shown to be significantly different from the rattles mostly ranked 13 and below.

Thus it is certain that the set of chosen rattles differ in annoyance, despite the fact that the Perceived Level is the same for each rattle. The number and variety of rattles tested indicate that this conclusion may be valid for other isolated rattles. Rattles from “large” objects such as windows, walls, and doors are found to be more annoying than rattles from “small” objects. The rank order in terms of annoyance obtained from the PC and CS methods is consistent across the tests, so it can be said that this general result is not dependent on the psychometric method. Category line scaling emerges as a preferred method for subsequent tests because it supports judgments of many more rattles. With the CS method, each sound is judged only once instead of in relation to each other sound. It is difficult, however, to distinguish the differences in response for several of the rattles given these data. Some rattles, despite having different sound qualities, do not elicit a significant difference in annoyance response. The PC method, however, does result in better annoyance discrimination between some sounds. Even with the conservative groupings of PC sounds discussed in Sec. 4.1, rattles ranked in the middle with medium annoyance are significantly different from the least and most annoying rattles, which cannot be concluded from the CS results.

4.3 Semantic Differential Subtest

The semantic differential (SD) subtest consisted of the same nine rattle sounds used in the paired comparison subtest, and listeners were asked to judge the rattles on 20 different continuous subjective scales representing different subjective factors. It is desired to observe variation of responses on different subjective scales to help explain the range in annoyance responses observed in the PC and CS subtests. A total of 180 judgments were required, and a shorter duration between sounds than that used for the first two subtests was implemented to keep the test length within one hour despite the larger number of judgments.

The subjective scales were devised to gather more information about people’s perception of rattle sounds in addition to annoyance. The scales were chosen from results of preliminary tests conducted at Purdue University [11]. In their test, a variety of the rattle sounds were

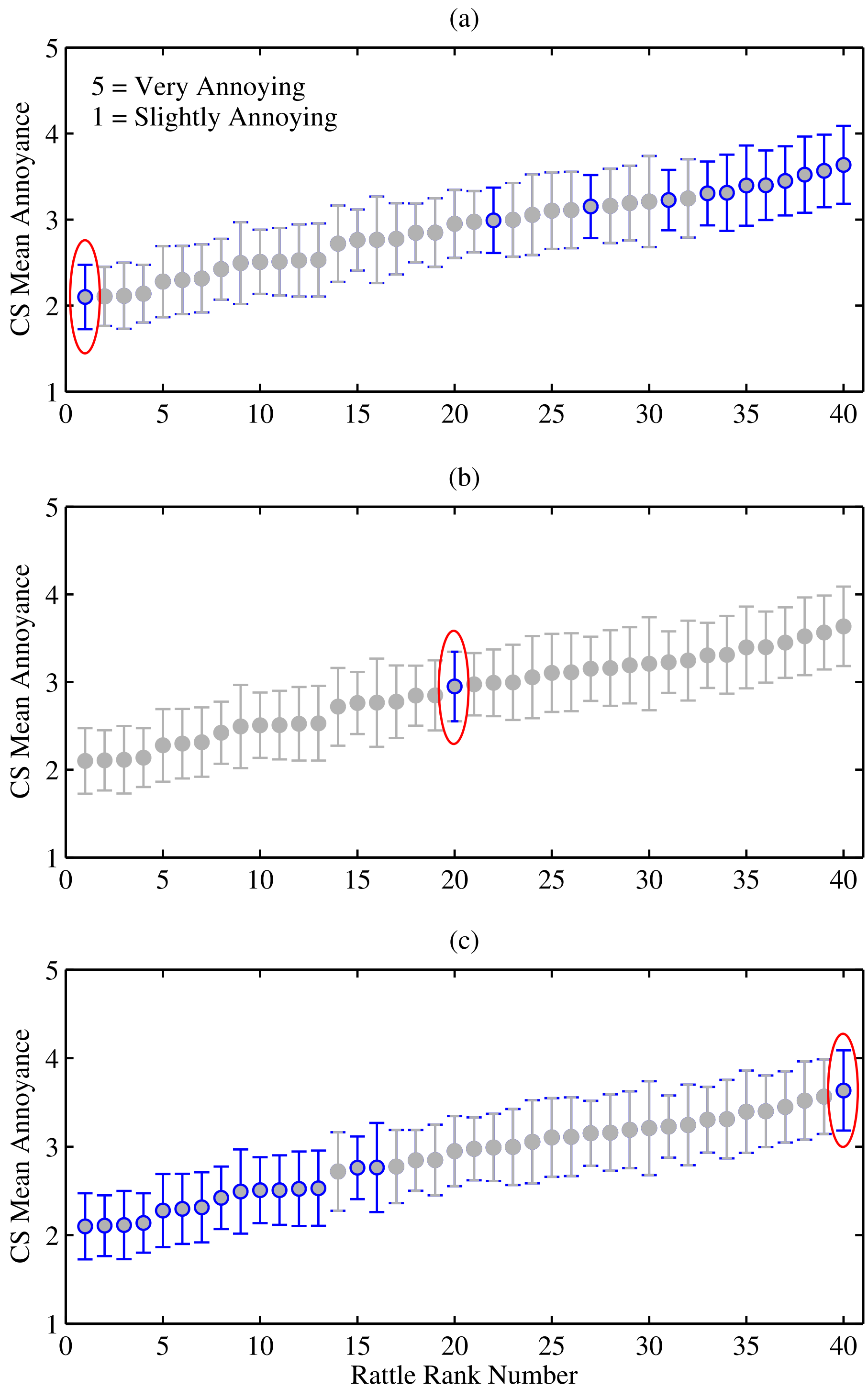

Figure 9. Mean annoyance and 95% confidence intervals for category line scaling subtest, arranged in increasing rank order from Slightly to Very Annoying. Pairwise comparison examples for the (a) least annoying rattle, (b) middle rattle, and (c) most annoying rattle.

presented to seven subjects, and each subject was asked to describe the sound attributes with adjectives and short phrases. Similar words from the responses were collected and ranked according to their frequency of appearance in the results. The words were also arranged into several categories according to what attributes they correspond to: impulsive, level, literal, literal/impulsive, literal/repetitive, spectral, temporal, and vibratory. In this context, literal attributes consist of similes or onomatopoetic descriptions. Potential scales using these words and corresponding antonyms were devised and reviewed by the research team. Scales deemed to be redundant were combined or replaced with a new single scale. The final 20 label pairs for the SD scales are included in Table 2, and each of these pairs label the end points on a 5-point scale. Scales are arranged in a bipolar configuration so that the left descriptor is nominally a positive reaction, with the corresponding negative reaction on the right, although this classification is not clear for all scale label pairs. The middle of the scale represents a natural zero point of indifference in most cases; the notable exception is for the unipolar scale Not Annoying–Annoying. An example screen for the Shallow–Deep scale is shown in Fig. 10. As with the CS subtest, subjects were asked to mark their judgment anywhere along the line.

Quiet	Loud	Continuous	Repetitive
Far	Close	Calm	Agitated
Isolated	Enveloping	Smooth	Rough
Simple	Complex	Steady	Vibrating
Familiar	Strange	Dull	Sharp
Not Annoying	Annoying	Low Pitch	High Pitch
Safe	Dangerous	Shallow	Deep
Brief	Sustained	Light	Heavy
Slow	Rapid	Soft	Hard
Soothing	Startling	Gradual	Abrupt

Table 2. Twenty label pairs for semantic differential subtest subjective scales.

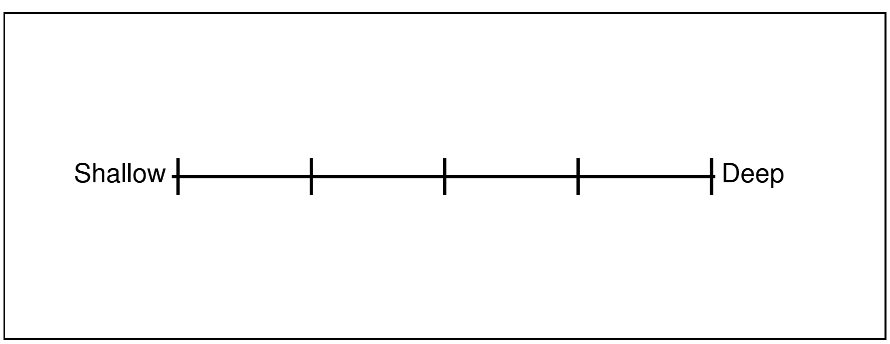

Figure 10. Example judgment screen for semantic differential subtest for the word pair Shallow–Deep.

The SD subtest was further subdivided into two tests: one where a randomized set of all sounds was judged on a particular scale before continuing to the next scale (SD1), and another where the order of all sound/scale pairings was randomized (SD2). These different presentation methods were investigated because it was unknown if the scale presentation order would affect the results. This subdivision resulted in twelve listeners per SD subtest.

The relationships between the twenty scale variables and the judgments for the nine rattle sounds in the two SD subtests are difficult to represent graphically. One visualization of the variability for different sounds on different scales is a box plot, given in Fig. 11, showing the medians with vertical lines, lower and upper quartile values (25%-75% of points) with boxes, whiskers (covering 99.3% of data) with dashed lines, and outliers with plus signs. The dashed line whiskers extend to the values that fall within approximately $\pm 2.7\sigma$, where σ is the standard deviation. The judgment data are given values from -2 to +2, representing the left- and right-hand ends of the scale, respectively. The values shown are based on the mean scores across all 12 subjects for each sound in each SD subtest. The boxplots therefore show the variation on different subjective scales for the nine chosen rattle sounds.

Results from SD1 and SD2 are similar. Some scales, such as Simple–Complex, Shallow—Deep, Safe–Dangerous, and Light–Heavy, show a large amount of variation across the nine sounds for both subtests. Other scales, such as Far–Close and Continuous–Repetitive, do not show much variation and as such do not contribute to explanations of the differences between sounds. Due to the impulsive nature of the sounds, nearly all sounds were judged to be abrupt, startling, and rapid (positive end of respective scales). In addition, nearly all sounds were judged to be close, as opposed to far. This is likely due to the nature of the rattle sounds that were presented over headphones. Although some listeners informally commented on the spaciousness perceived in some sounds, this does not appear to have led them to perceive sounds as coming from far away.

The mean and 95% confidence intervals are computed for each rattle sound on each subjective scale for both the SD1 and SD2 methods. The SD annoyance ranking is similar to that from the PC subtest. One example of the results for the scale Light–Heavy is given in Fig. 12. The rattle sounds are arranged in the order of increasing annoyance from the paired comparison subtest. In this case and for several other subjective scales, the left side (or negative numbers) seems to correspond to a judgment of less annoying. A few scales, such as Light–Heavy, separate the sounds into two distinct groups. These scales are: Shallow–Deep, Safe–Dangerous, Soft–Hard, and Light–Heavy. As shown in Fig. 12, the first three sounds appear to group in the negative region, indicating a perception of Light, while most of the remaining sounds are grouped in the positive region, indicating a perception of Heavy. This grouping is apparent for both the SD1 and SD2 methods. The three “light” sounds are rattles from small objects, such as a wall hanging, candle globe, and wine glass, so the perception of Light is appropriate. Interestingly, these three rattles were also judged to be Shallow, Safe, and Soft. The rattles judged closer to the Deep, Dangerous, Hard, and Heavy ends of the scales are from a window, door, garage door, bedroom door, ceiling fan, and another window. The ceiling fan rattle, while resulting in a positive score, was actually judged to be close to the middle zero point on most of these scales.

The confidence interval width for each scale is calculated to determine which scales resulted in the most consistent answers, regardless of the mean score. For each scale, the 95% confidence interval widths are averaged across the nine sounds. Variations are similar across the SD1 and SD2 subtests, and these results are combined to give one average confidence interval width for each scale, as presented in Table 3. Analysis indicates that the smallest confidence interval widths correspond to the scales Soothing–Startling, Shallow–Deep, Light–Heavy, and Smooth–Rough. This indicates that it is probably easiest for

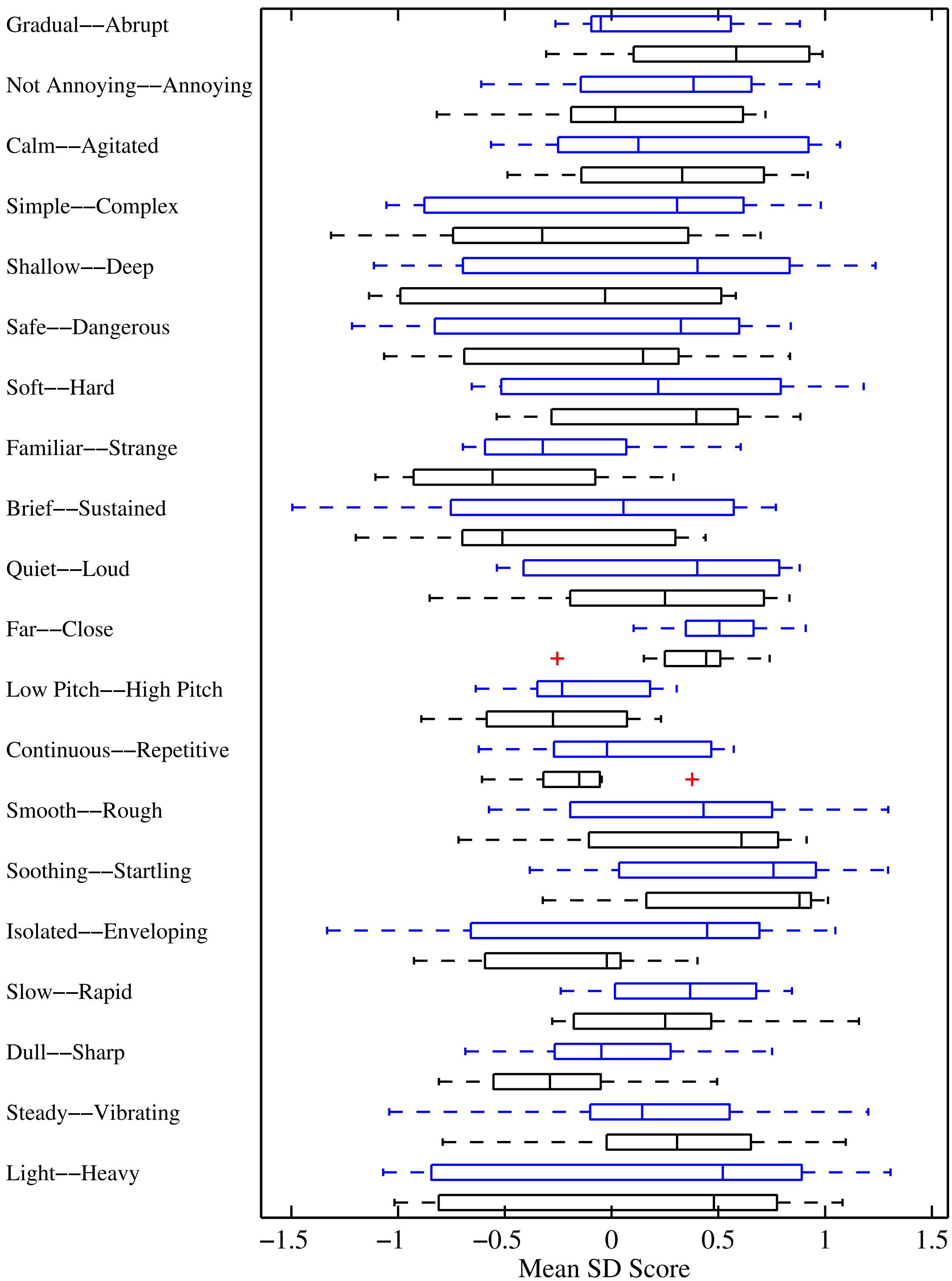

Figure 11. Box plot of mean judgment data across listeners for all scales from both semantic differential subtests: SD1 (——) and SD2 (——).

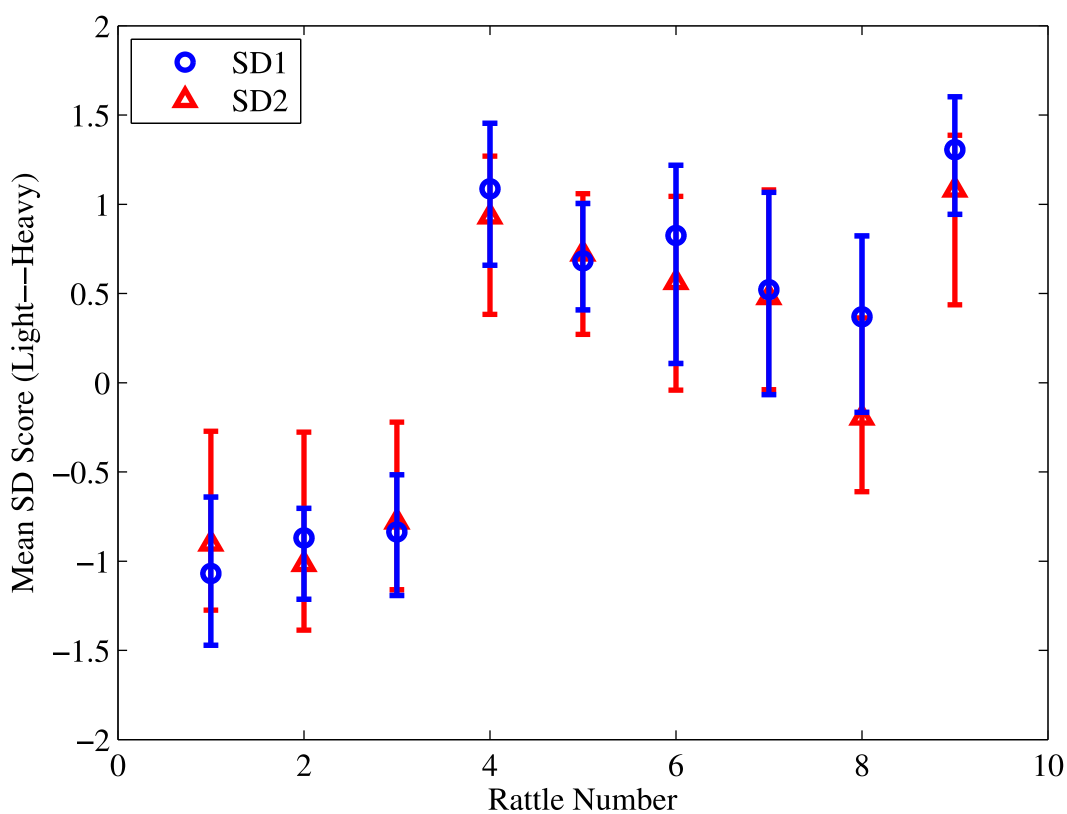

Figure 12. Mean scores and 95% confidence intervals for the Light–Heavy subjective scale for both SD1 and SD2 subtests. Rattle sounds are arranged in order of increasing annoyance from the paired comparison subtest.

subjects to judge the rattles on these scales. By contrast, the largest confidence interval widths correspond to the scales Dull–Sharp, Low Pitch–High Pitch, Continuous–Repetitive, and Far–Close. It is probably hardest to judge the rattles on these scales.

Subjective Scale	95% CI width
Soothing–Startling	0.79
Shallow–Deep	0.86
Light–Heavy	0.89
Smooth–Rough	0.90
Safe–Dangerous	0.92
Quiet–Loud	0.94
Simple–Complex	0.95
Not Annoying–Annoying	0.97
Familiar–Strange	0.99
Steady–Vibrating	0.99
Calm–Agitated	1.00
Gradual–Abrupt	1.01
Soft–Hard	1.03
Brief–Sustained	1.03
Slow–Rapid	1.04
Isolated–Enveloping	1.04
Far–Close	1.07
Continuous–Repetitive	1.08
Low Pitch–High Pitch	1.13
Dull–Sharp	1.15

Table 3. Average 95% confidence interval widths across all nine sounds for each subjective scale in both semantic differential subtests.

4.3.1 Semantic Differential Correlations

Correlations between judgments on the annoyance scale and the 19 other subjective scales are computed for both SD1 and SD2. Table 4 presents these correlation coefficients, with insignificant correlations ($p \geq 0.05$) shaded in gray. It is shown that judgments on the scale Far–Close do not have a significant correlation with annoyance in either subtest. The scales with larger confidence intervals identified above all exhibit low and insignificant correlation with annoyance in at least one of the semantic differential subtests. One interesting result is that judgments on the Continuous–Repetitive scale correlate well with annoyance in SD1, but not at all in SD2. Most correlations are similar between the two test methods, and it is unknown why the Continuous–Repetitive correlations are vastly different between the two methods. It might be difficult to rate the rattle sounds on this scale, as evidenced by the grouping of judgments about 0, the neutral point, as shown in Fig. 13(a). Perhaps a test with a larger set of sounds would provide a larger range of judgments on this scale for valid correlations. The scales with high and significant correlations in both subtests include Simple–Complex and Smooth–Rough. This indicates that sounds described as simple and smooth are less annoying than complex and rough sounds.

Subjective Scale	Correlation Coefficient r	
	SD1	SD2
Gradual–Abrupt	0.65	0.72
Calm–Agitated	0.82	0.90
Simple–Complex	0.92	0.92
Shallow–Deep	0.71	0.57
Safe–Dangerous	0.78	0.84
Soft–Hard	0.83	0.67
Familiar–Strange	0.65	0.72
Brief–Sustained	0.97	0.77
Quiet–Loud	0.88	0.83
Far–Close	0.10	0.50
Low Pitch–High Pitch	0.22	0.76
Continuous–Repetitive	0.88	0.04
Smooth–Rough	0.93	0.90
Soothing–Startling	0.89	0.84
Isolated–Enveloping	0.91	0.74
Slow–Rapid	0.76	0.76
Dull–Sharp	0.71	0.21
Steady–Vibrating	0.91	0.82
Light–Heavy	0.81	0.57

Table 4. Correlation coefficients (r) between judgments on the annoyance scale and the 19 other subjective scales for both semantic differential subtests.

Example plots of the Continuous–Repetitive and Simple–Complex ratings versus the annoyance ratings are presented in Fig. 13(a) and (b), respectively. The ratings and linear regression line in Fig. 13(a) for SD2 show that there is no correlation between this scale and annoyance, as described earlier. Regardless of annoyance, subjects rated the sounds near the zero neutral point in most cases. In contrast, the SD1 ratings do show a correlation with annoyance. In Fig. 13(b) a high correlation is demonstrated for the Simple–Complex scale in both subtests.

4.3.2 Semantic Differential Factor Analysis

Finally, a factor analysis is performed for the semantic differential data [50]. This analysis identifies common factors that can be used to describe the overlapping dependencies in the data. The goal is to use the large number of subjective scales to determine a smaller set of factors that explain the variation in subjective response. This dimension reduction technique can be used to identify underlying factors that are not directly observed. Data from the SD1 and SD2 subtests are combined for this analysis.

Factor loadings from a four-factor analysis of the combined data is presented in Fig. 14. It is found that four factors is the smallest number of factors to sufficiently explain the data; the p-value indicating whether to reject the null hypothesis of four common factors is 0.132,

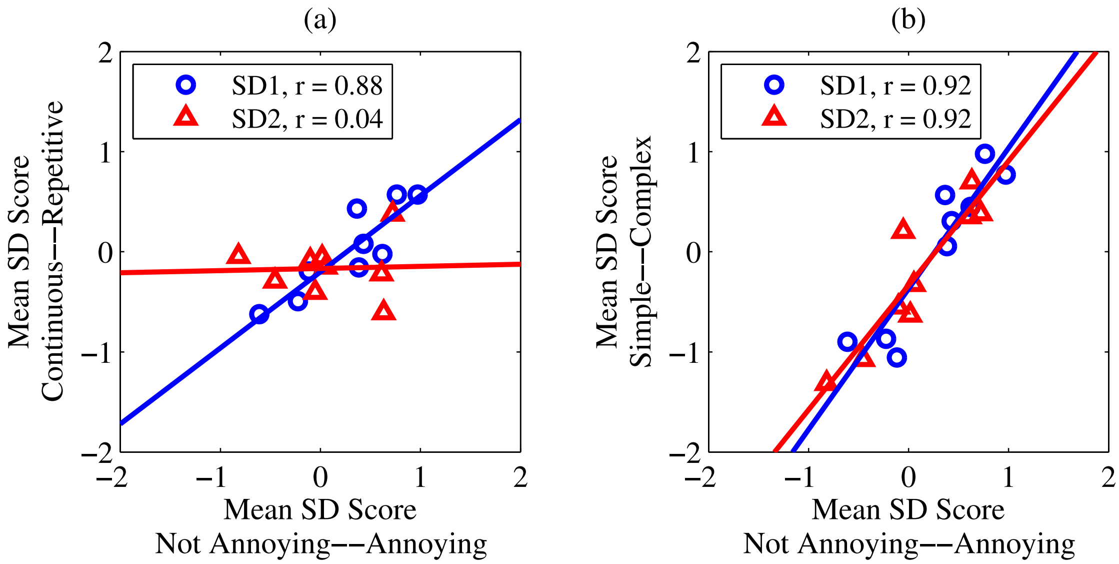

Figure 13. Subjective scale ratings versus annoyance ratings and linear fit lines for both SD1 and SD2 subtests. (a) Continuous–Repetitive scale. (b) Simple–Complex scale.

which is not significant at the 95% level. Several methods for rotating factor loadings, as well as no rotation, are investigated to find a solution in which the majority of subjective scales can each be described by a single factor. Promax rotation, an oblique method which does not constrain the factors to be orthogonal and thus assumes the factors are correlated, produces the most desirable result.

When interpreting factor analysis data, factor loadings greater than 0.6 are commonly used to determine what the factors represent [50]. In this case, rounded factor loadings greater than or equal to 0.5 that do not have high loading (≥ 0.4) on any other factor are used. Table 5 shows the numeric values for the factor loadings and sorts the scales according to these criteria. Thus the factor loadings in Fig. 14 and Table 5 can be interpreted by considering the following groupings of subjective scales:

1. Light–Heavy, Shallow–Deep, Safe–Dangerous, Quiet–Loud
2. Smooth–Rough, Calm–Agitated
3. Brief–Sustained, Simple–Complex, Isolated–Enveloping
4. Low Pitch–High Pitch, Dull–Sharp

Subjective Scale	Factor 1	Factor 2	Factor 3	Factor 4
Light–Heavy	0.796	-0.142	0.221	-0.050
Shallow–Deep	0.743	-0.144	0.186	-0.009
Safe–Dangerous	0.661	0.138	0.041	0.068
Quiet–Loud	0.520	0.245	0.022	0.113
Soft–Hard	0.493	0.467	-0.116	0.020
Gradual–Abrupt	0.418	0.382	-0.233	0.038
Smooth–Rough	0.197	0.764	-0.035	-0.046
Calm–Agitated	0.245	0.604	-0.073	0.118
Soothing–Startling	0.422	0.536	-0.007	-0.199
Steady–Vibrating	-0.029	0.453	0.391	-0.125
Far–Close	-0.115	0.370	-0.032	-0.126
Continuous–Repetitive	-0.068	0.356	0.071	-0.013
Slow–Rapid	-0.047	0.326	0.124	0.165
Brief–Sustained	0.059	-0.117	0.755	-0.002
Simple–Complex	0.048	0.032	0.681	0.044
Isolated–Enveloping	0.119	-0.023	0.619	-0.058
Familiar–Strange	-0.104	0.160	0.370	-0.084
Not Annoying–Annoying	0.192	0.254	0.286	0.194
Low Pitch–High Pitch	0.030	-0.126	-0.045	0.900
Dull–Sharp	-0.160	0.188	0.139	0.459

Table 5. Factor loadings from a four-factor analysis of the semantic differential subtest. Data from SD1 and SD2 are combined.

Considering these groupings, factor 1 is related to spectral balance and level. Spectral balance is defined here as the perceived balance of low- and high-frequency energy. For

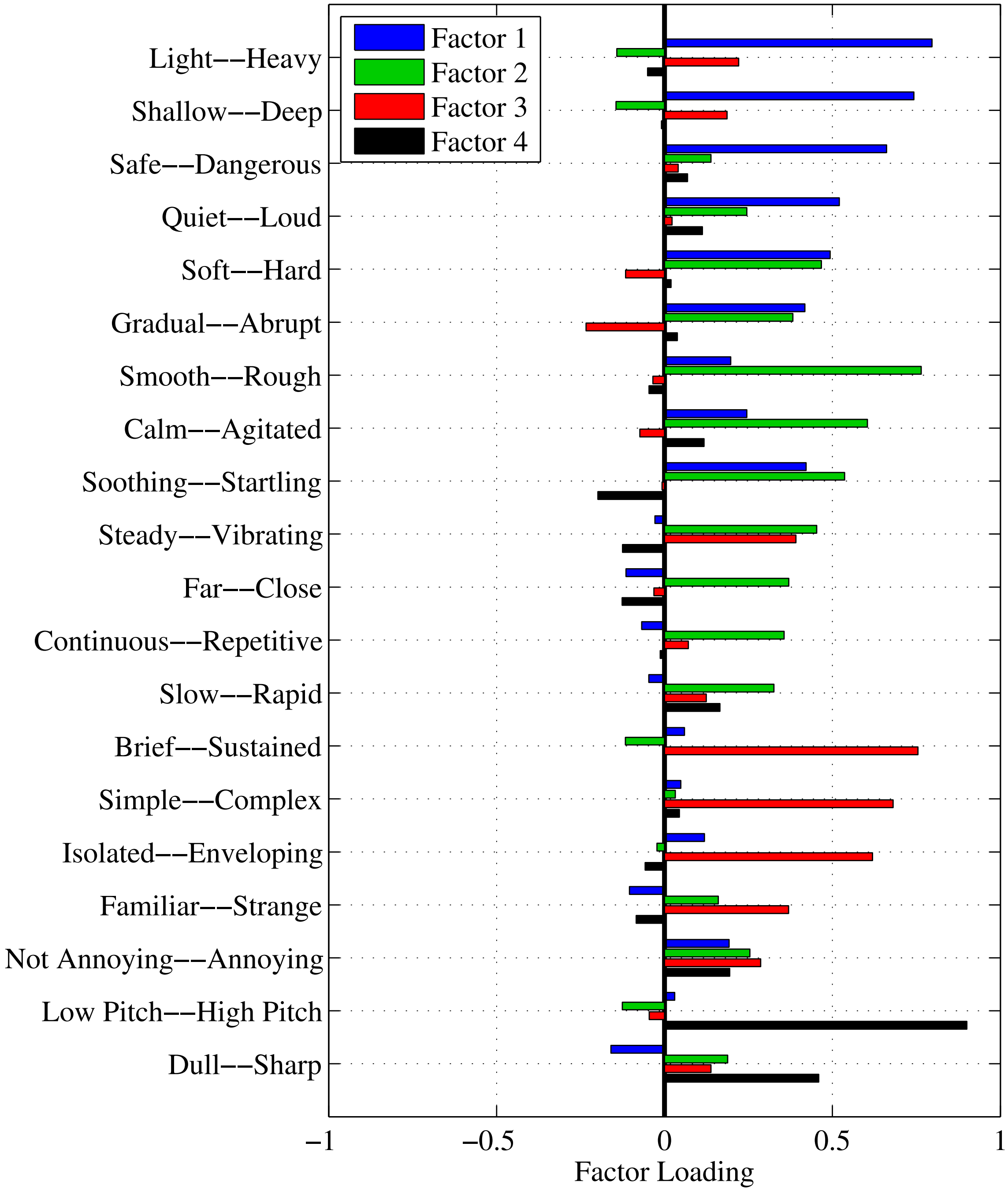

Figure 14. Factor loadings from a four-factor analysis of the semantic differential subtest. Data from SD1 and SD2 are combined.

example, sounds described as light and shallow may contain more high-frequency energy, while heavy and deep sounds may contain more low-frequency energy. Factor 2 is related to temporal variability. Factor 3 is more difficult to summarize and appears to be related to the impulsive nature, complexity, and envelopment of the sound. Factor 4, like factor 1, is related to spectral balance, but it is a separate factor. Only one scale, Low Pitch–High Pitch, contributes highly to this fourth factor; while Dull–Sharp exhibits its largest factor loading for factor 4, it is still not an extremely high loading. It is possible that subjects do not perceive these scales to be related to spectral balance, contrary to the authors' assumption. As stated above, however, this fourth factor cannot be ignored. Judgments on other scales describing firmness, abruptness, startle, vibration, distance, repetition, speed, and familiarity can be related to the four factors, although they do not contribute highly to one single factor.

Considering the factor loadings for the scale Not Annoying–Annoying, the four factors contribute to annoyance with similar factor loadings, although factor loadings for 2 and 3 are slightly higher than for 1 and 4. Annoyance to this limited set of rattle sounds is therefore related to the spectral balance, level, temporal variability, impulsivity, complexity, envelopment, and sharpness of the rattle sound. Thus it is found that level contributes to annoyance. Subjects apparently did perceive differences in loudness despite the PL normalization.

Taking into account the confidence intervals, correlations with annoyance, and factor analysis results, it is possible to identify which subjective scales would be best suited for additional tests with these rattle sounds. Scales that could be eliminated due to ambiguity in judgments, low correlation with annoyance, and low factor loadings include the Continuous–Repetitive and Far–Close scales. In addition, the scales Dull–Sharp and Low Pitch–High Pitch could be eliminated due to inconsistent judgments and low correlation with annoyance in at least one subtest. For the types of rattle sounds investigated, low variation and low factor loadings on the Slow–Rapid and Gradual–Abrupt scales results in little useful information. Fourteen scales remain after these eliminations, and more could be excluded due to similarities between scales. Increasing the variety of rattle sounds would allow for even greater insight into how people describe their perception of rattle sounds.

4.3.3 Comparison of Semantic Differential Methods

A comparison of multi-dimensional test methods was performed by Parizet and Nosulenko [43] for pairs of sounds from idling diesel cars. The first "conventional" method involved judging each pair of sounds according to a set of parameters before judging the next pair of sounds. The second method, similar to SD1 in the current study, involved judging all the pairs of sounds for one parameter before judging the next parameter. Subjects also completed a questionnaire on the perceived length and difficulty of the test. It was found that results from the two methods were equivalent. The questionnaires did not indicate a preference for either method based on perceived test length or difficulty. However, the second method was chosen as the preferred method because results were more consistent and the actual length of the test was shorter.

In terms of different methods for this study's semantic differential presentation, SD1 is the preferred method, in which all sounds are judged on a particular scale before continuing to the next scale. Confidence intervals are slightly smaller and most correlations with annoyance are higher for SD1. Subjects are likely able to concentrate on a particular scale while listening to different sounds. General conclusions are the same for each subtest, and the average time required for completion of the subtests is almost the same. This preference is in agreement with the findings by Parizet and Nosulenko, although the alternate method with random sound/scale pairings in SD2 is not the same as the conventional method used by them.

4.4 Test 1 Summary

Three subjective subtests of human reactions to rattle sounds were conducted using three different psychometric methodologies. The rattles were presented in the absence of sonic boom sounds, and all rattles were normalized to the same Perceived Level. In the paired comparison subtest, nine rattles are ranked in order of increasing annoyance. Rattles from small objects, such as wall art and a wine glass, are found to be less annoying than rattles from larger structural elements, such as doors and windows. Since each sound must be compared to every other sound, a disadvantage of this method is the small number of sounds that can be tested in a typical 1-hour test.

In the category line scaling subtest, a much larger number of rattles (forty) was presented for annoyance judgments. Despite all sounds having the same PL, significant differences in mean annoyance are observed. Annoyance ranking of sounds is very similar to that observed during the paired comparison subtest for the nine common sounds. This subtest confirms the paired comparison conclusion that "large" rattles are found to be more annoying than "small" rattles, which indicates that this result is not dependent on the psychometric method. The number and variety of rattles tested also indicate that this conclusion may be valid for other rattles. Category line scaling emerges as a preferred method for subsequent tests because it enables judgments of many more sounds than the paired comparison method.

The semantic differential subtest explored subjects' reactions to nine rattle sounds, which were the same as those in the PC subtest, on a variety of subjective scales. A comparison of two different ordering methods indicates a preference for judging all sounds on a particular subjective scale before continuing to the next scale. Analyses identify which scales result in consistent judgments across subjects, indicated by the confidence interval about the mean rating on each scale, and which subjective factors correlate the best with annoyance. Agitation, complexity, duration, and roughness are the subjective factors that correlate the highest with annoyance. A factor analysis represents the twenty scales with four common factors that can be interpreted as spectral balance and level; temporal variability; impulsivity, complexity, and envelopment; and sharpness. These four factors contribute to annoyance with similar weightings.

All three subtests indicate a difference in annoyance between rattle sounds of the same calculated PL. Human response to these impulsive sounds therefore reflects a sensitivity to other factors not accounted for in loudness level. The metric PL is not sufficient to describe reactions to the rattles, and additional factors such as temporal variability and complexity may need to be considered.

5 Test 2: Subjective Test of Booms and Rattles at Same PL

5.1 Test 2 Description

After exploring human reactions to isolated rattle sounds, a second test was conducted to explore annoyance to the combination of sonic booms and rattle sounds. The main objective was to investigate whether differences in annoyance still exist between rattles once they are combined with sonic booms. In addition, the response to the combination was compared to that for a sonic boom alone. The category line scaling method was chosen for Tests 2 and 3 based on analysis of the different methods performed in Test 1.

Using the method described in Sec. 2.3, four indoor sonic booms were combined with five rattles selected from Test 1. The rattles were chosen to span the annoyance range and included the least and most annoying rattles from Test 1. The separate rattle and boom stimuli were mixed together to investigate how people react to the combined sound. The amplitudes of the constituent sounds were systematically varied so that the amount of rattle in the combined sound spanned the range from rattle only (with no boom) to boom only (with no rattle). The combined sounds presented to the subjects were all normalized to the same total PL value of 65 ± 1 dB. Nine combinations for each pairing were created, including an isolated boom, seven boom and rattle combinations with differing relative levels, and an isolated rattle. The combinations are denoted by a dB decrease in the rattle level relative to the isolated rattle level for a PL of 65 dB. The nine rattle levels are 0, -2.4, -4.9, -7.3, -12.1, -17.0, -21.9, -30, and $-\infty$ dB relative to the isolated rattle level. The two ends of the scale, 0 and $-\infty$ dB, represent the rattle alone and boom alone sounds, respectively. The middle levels represent the mixed sounds with a decreasing contribution of rattle to the loudness level of the overall mixed sound. The increments were selected for a finer resolution in rattle amplitude near the rattle only end of the scale (0 dB), where annoyance was predicted to vary more widely. The level -30 dB was determined by the investigators to be slightly above the just-audible rattle level and was selected to anchor the opposite end of the scale of mixed sounds.

A total of 169 sounds were presented to the subjects for judgment, including 149 different sounds and an additional 20 sounds that were repetitions of some sounds. Different random orders of sounds were used for each group of three or four subjects to eliminate any ordering bias. A total of 55 subjects were tested. An example screen for the annoyance judgment scale in Test 2 is shown in Fig. 15. This scale is anchored at both ends and in the middle by word descriptors. However, subjects were asked to mark their judgments anywhere along the line. The scale anchors encompass a larger range of annoyance than the scale in the Test 1 category line scaling subtest, because it was believed that annoyance would vary more for this more diverse set of boom and rattle sounds.

5.2 Test 2 Annoyance to Boom and Rattle Sounds at the Same PL

Figure 16 presents results for the mean annoyance for one boom (a recorded outdoor sonic boom filtered to simulate low transmission loss and reception in a large room) and all five rattles tested. The mean annoyance is shown vs. the rattle level, which is relative to the rattle only level, as discussed in Sec. 5.1. At the left is the mean annoyance to the boom alone. This is the same for all rattle cases because it represents annoyance to a single boom sound without rattle. Moving to the right, the ratio of rattle to boom increases,

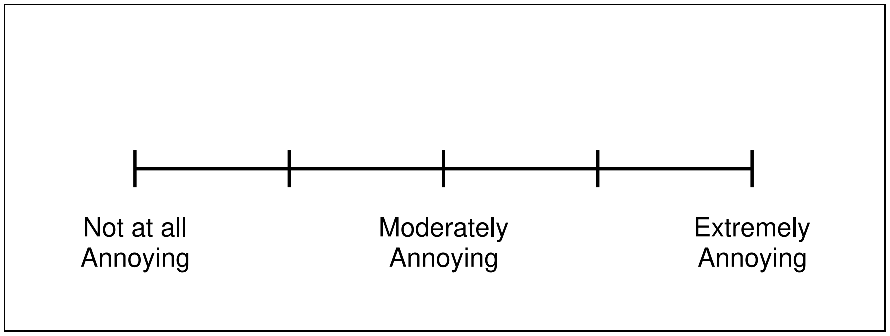

Figure 15. Example judgment screen for Test 2.

and the mean annoyance generally increases. This suggests that the additional presence of rattle increases human annoyance to sonic booms, even though the Perceived Level is held constant at 65 dB. The mean annoyance at a rattle level of 0 dB corresponds to the rattles presented alone. A range in isolated rattle annoyance is present in these data from Test 2, and the rank order of the rattle sounds presented alone is similar to that observed in Test 1. In several instances, the maximum annoyance is observed at a rattle level of −5 dB, and annoyance decreases as the rattle level is further increased and the boom level is correspondingly decreased.

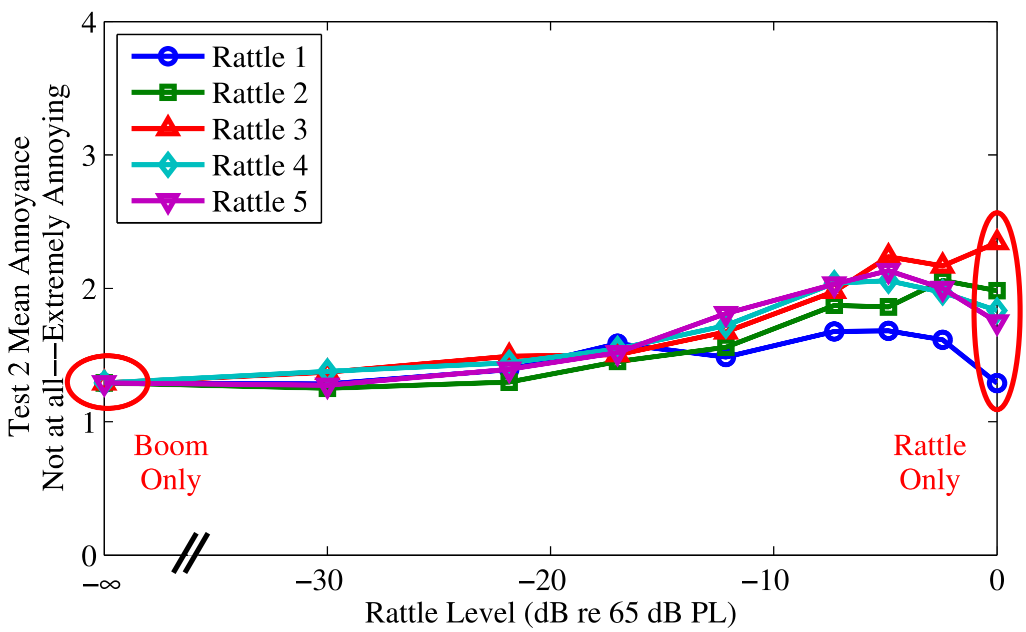

Figure 16. Test 2 mean annoyance for Boom 1 (a recorded outdoor sonic boom filtered to simulate low transmission loss and reception in a large room) and five rattles as a function of rattle level relative to the isolated rattle level.

As shown for Rattle 4 in Fig. 17, mean annoyance to a rattle mixed with different booms is similar, despite the differences in boom characteristics (see App. B for more examples). As explained in Sec. 2.2, the four booms tested differ in either origin or the applied transmission loss and reverberation filters. These results suggest that the presence of rattle is a more important contributor to annoyance than differences in characteristics of the four filtered booms in this test, when the sounds are presented at the same loudness level. Differences in annoyance to different rattles do not disappear when rattles are combined with different booms. Note that since these sounds are presented over headphones, very low frequencies, which could affect human reactions through tactile response or whole body vibration, are not present in the signals. These sounds could be studied in a test environment with more realistic low-frequency reproduction to confirm the above conclusions.

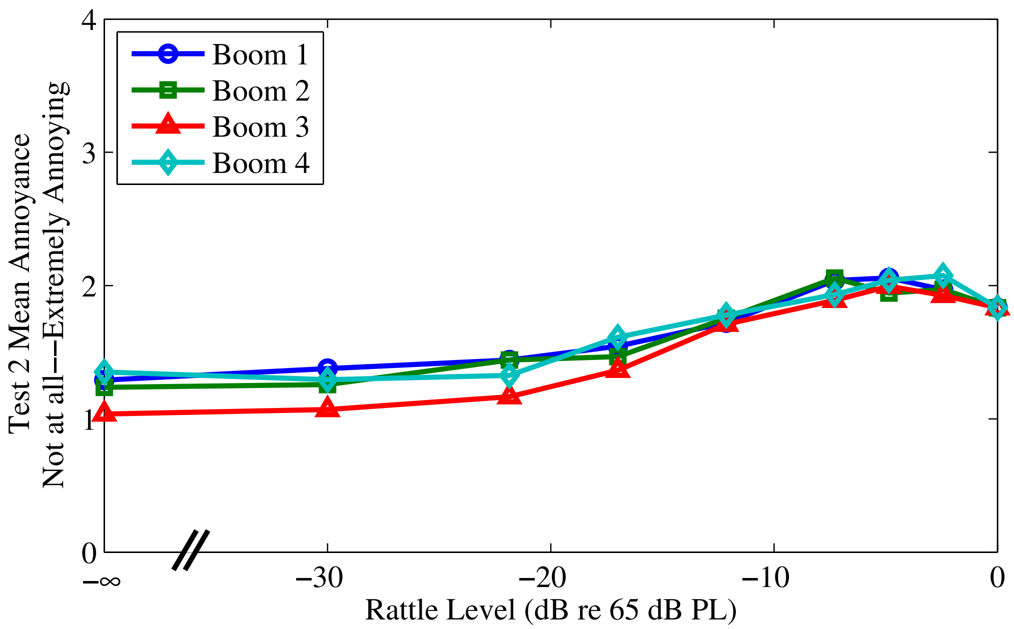

Figure 17. Test 2 mean annoyance for Rattle 4 and four booms as a function of rattle level relative to the isolated rattle level.

5.3 Test 2 Statistical Analysis

Several statistical tests were performed to test the observations noted above. A one-way repeated measures Analysis of Variance (ANOVA) test was performed to test for differences in annoyance among the nine combinations of booms and rattles for each pairing[2]. The ANOVA test is performed 20 times, once for each boom and rattle pairing. In each case, Mauchly's test of sphericity shows that sphericity is violated, and the Greenhouse-Geisser method is used to adjust the degrees of freedom and correct the results. If sphericity were assumed, the between levels degrees of freedom would be 8, and the within levels degrees of freedom would be 432. Here "levels" refers to the nine rattle levels defined in Sec. 5.1.

[2]Recall that each of the four booms was paired with each of the five rattle sounds, resulting in 20 pairs. Each pairing resulted in nine sounds, including the boom alone, rattle alone, and seven mixed boom and rattle sounds.

The adjusted degrees of freedom, values for the F-statistic, and p-values for the 20 ANOVA tests are included in Table 6. For example, the results for Boom 1 and Rattle 1 indicate a difference in annoyance with $F(3.889, 210.027) = 4.057$ and $p = 0.004$. All tests are significant, indicating that there is a significant difference in annoyance among the nine rattle levels for each pairing.

Sonic Boom	Rattle	Between Levels Degrees of Freedom	Within Levels Degrees of Freedom	F-statistic	Significance (p-value)
B1	R1	3.889	210.027	4.057	0.004
	R2	3.859	208.412	17.094	< 0.001
	R3	3.365	181.712	21.787	< 0.001
	R4	3.731	201.464	12.077	< 0.001
	R5	4.013	216.681	14.796	< 0.001
B2	R1	3.751	202.577	4.379	0.003
	R2	3.453	186.462	13.156	< 0.001
	R3	3.585	193.602	26.330	< 0.001
	R4	3.499	188.937	11.245	< 0.001
	R5	3.421	184.722	13.417	< 0.001
B3	R1	5.091	274.896	3.511	0.004
	R2	4.125	222.762	18.077	< 0.001
	R3	3.158	170.558	35.266	< 0.001
	R4	3.160	170.653	14.796	< 0.001
	R5	4.180	225.739	12.385	< 0.001
B4	R1	4.411	238.214	6.813	< 0.001
	R2	4.695	253.529	19.179	< 0.001
	R3	3.381	182.555	33.554	< 0.001
	R4	3.447	186.123	15.420	< 0.001
	R5	4.151	224.158	14.784	< 0.001

Table 6. One-way repeated measures Analysis of Variance (ANOVA) results for Test 2. Corrections for violations of sphericity are performed using the Greenhouse-Geisser method.

Post-hoc pairwise comparisons were performed with the Bonferroni method to determine which rattle levels differ significantly on annoyance, and whether mixed boom and rattle signals are more annoying than either boom or rattle signals alone. It is found that the combination of boom and rattle is more annoying than the boom alone in 18 out of 20 cases ($p \leq 0.05$). However, the boom and rattle combinations are not significantly more annoying than the rattle alone in any case (except for Boom 4 and Rattle 1). This last point is contrary to what may be inferred from simply observing the plots of data, such as that shown in Fig. 16. The annoyance response to the combination of boom and rattle sounds is thus governed by the response to the rattle sounds.

A threshold for rattle annoyance is estimated from the pairwise comparison data for the 18 cases where the combination of boom and rattle is more annoying than the boom alone. This threshold is the rattle level at which subjects became significantly more annoyed by the combined sound than by the boom alone. A histogram of the rattle level threshold for mean annoyance is given in Fig. 18. The maximum frequency of occurrence, about one-third of all cases, is shown for a threshold of −12.1 dB relative to the isolated rattle PL of 65 dB. In other words, subjects indicated an initial increase in annoyance relative to the isolated boom when the rattle level was −12.1 dB for 30% of the cases. Although not shown here, the annoyance continues to increase beyond this threshold up to a maximum near a rattle level of −5 dB. Note that this threshold cannot be translated into a rattle penalty *per se* because the sonic boom level was also adjusted in order to retain a total PL of 65 dB.

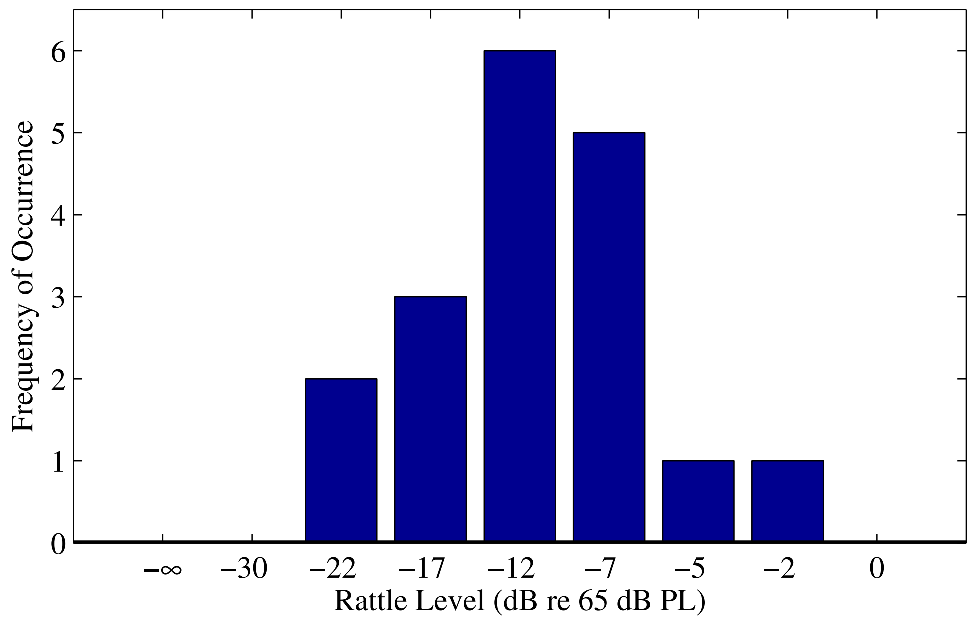

Figure 18. Test 2 histogram of rattle level threshold for mean annoyance.

6 Test 3: Subjective Test of Booms and Rattles at Different PL

6.1 Test 3 Description

While Test 2 identifies a significant increase in annoyance due to the inclusion of a rattle sound with a sonic boom, this effect is only demonstrated for one loudness level. In Test 3, the mixed boom and rattle sounds were normalized to three different PL values of 61.5, 65, and 68.5 dB. In order to accommodate this larger set of sounds, seven rattle levels were down-selected from the original nine levels, and two booms and three rattles were chosen from the original four and five, respectively. Eliminating rattle levels −30 and −17 dB, the seven rattle levels in Test 3 are 0, −2.4, −4.9, −7.3, −12.1, −21.9, and $-\infty$ dB relative to the isolated rattle level. Booms 1 and 4 were chosen because of differences in their character, and they represent a recorded boom received in a large room with low TL and a synthesized ramp boom received in a small room with moderate TL, respectively. Rattles 1, 3, and 4 were chosen based on differences in annoyance response found in Test 2.

A total of 149 sounds were presented to the subjects for judgment on a category line scale identical to that from Test 2 (see Fig. 15). Of the 149 sounds, only 105 unique sounds are used in the analysis, while 20 sounds were repetitions of some sounds, and 24 extra sounds were included to introduce more variety to the set of presented sounds. Different random orders of sounds were used for each group of four subjects to eliminate any ordering bias. A total of 40 subjects were tested.

6.2 Comparison of Tests 2 and 3

Of the 105 test sounds in Test 3, there are 35 sounds in common with Test 2 (corresponding to the middle PL of 65 dB). A comparison of the mean annoyance to these 35 sounds in Tests 2 and 3 shows a high correlation with a correlation coefficient of $r = 0.955$ ($p < 0.001$). As presented in Fig. 19, the geometric mean regression line [42], which accounts for error in both x and y, exhibits a nearly $y = x$ relationship, and the slope of the line is 1.15. The

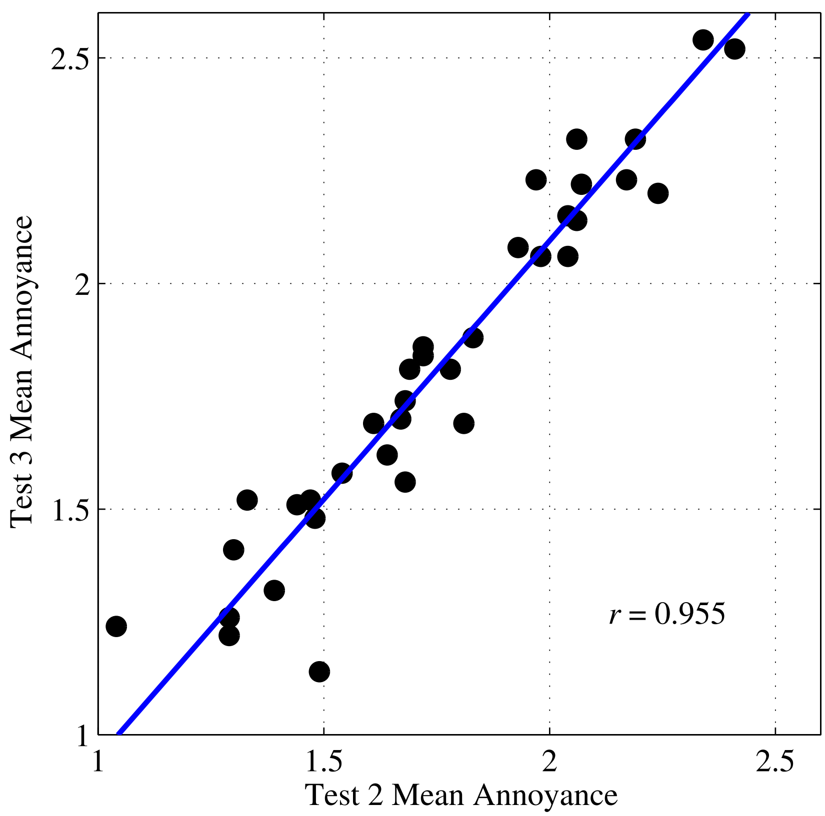

Figure 19. Comparison of mean annoyance in Tests 2 and 3 with geometric mean regression line.

annoyance to boom and rattle sounds at a PL of 65 dB therefore matches between the tests, confirming that the tests are repeatable with different subjects and when presented within a larger set of sounds that include variations in loudness.

6.3 Test 3 Annoyance to Boom and Rattle Sounds at Different PL

Returning to the full data set of Test 3, an example of the mean subjective annoyance results with rattle level for Boom 1 and Rattle 1 for the three PL values is shown in Fig. 20. The corresponding results from Test 2 at a PL of 65 dB are included as a dashed line for reference. These Test 2 results are very similar to results corresponding to the middle PL value from Test 3, as explained in Sec. 6.2. It is shown that the trends in mean annoyance are similar for the three PL groups and that the higher PL sounds are more annoying, as expected. Consistent with Test 2, some combinations of boom and rattle are more annoying than the boom alone, and this effect is independent of PL. Figures for all six boom and rattle combinations are included in App. C.

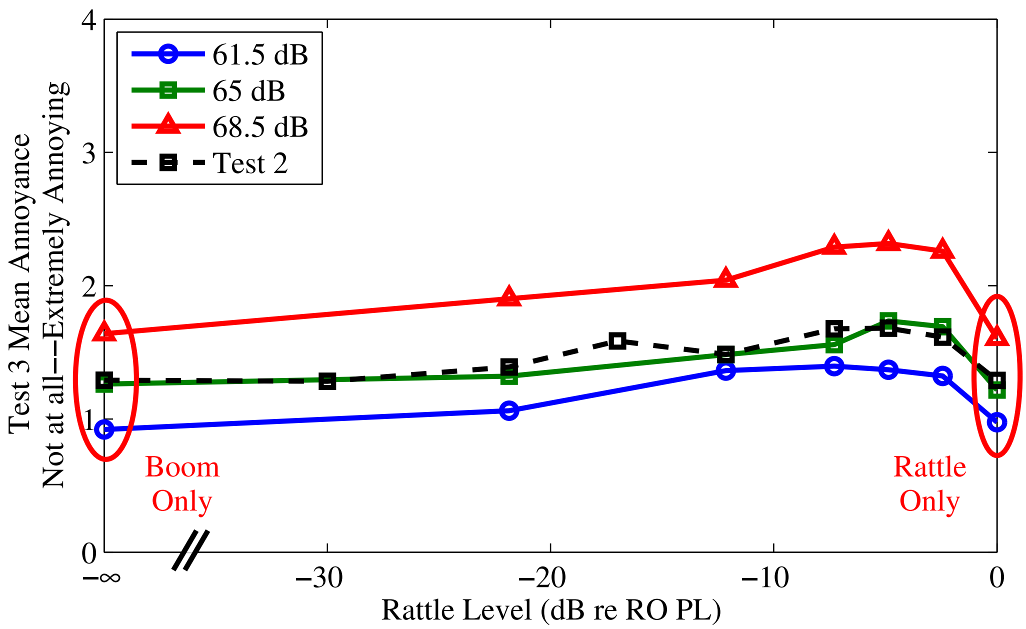

Figure 20. Test 3 mean annoyance for Boom 1 and Rattle 1 as a function of rattle level relative to the isolated rattle only (RO) level.

6.4 Test 3 Statistical Analysis

A series of one-way repeated measures ANOVA tests are conducted to test whether a difference in annoyance exists among the different rattle levels. A series of 18 ANOVA tests are conducted, with one analysis for each boom, rattle, and PL combination. Each test includes annoyance to the isolated boom at all three PL values. In each ANOVA case, Mauchly's test of sphericity shows that sphericity is violated, and the Greenhouse-Geisser method is used to adjust the degrees of freedom and correct the results. If sphericity were assumed, the between levels degrees of freedom would be 7, and the within levels degrees of freedom would be 273. The adjusted degrees of freedom, values for the F-statistic, and p-values for the 18 ANOVA tests are included in Table 7. For example, the results for Boom 1 and Rattle 1 indicate a difference in annoyance with $F(3.624, 141.333) = 4.920$ and $p = 0.001$. All tests are significant beyond the 5% level, which leads to the conclusion that there is a significant difference in annoyance among the rattle levels for each pairing.

Sonic Boom	Rattle	PL	Between Levels Deg. of Freedom	Within Levels Deg. of Freedom	F-statistic	Significance (p-value)
B1	R1	61.5	3.624	141.333	4.920	0.001
		65.0	4.332	168.967	7.195	< 0.001
		68.5	3.384	131.970	24.210	< 0.001
	R3	61.5	3.404	132.742	9.924	< 0.001
		65.0	4.080	159.127	24.072	< 0.001
		68.5	3.297	128.601	47.862	< 0.001
	R4	61.5	3.493	136.208	10.153	< 0.001
		65.0	3.578	139.526	17.743	< 0.001
		68.5	3.888	151.618	33.791	< 0.001
B4	R1	61.5	4.229	164.924	4.033	0.003
		65.0	4.629	180.533	5.520	< 0.001
		68.5	4.487	174.975	15.427	< 0.001
	R3	61.5	4.343	169.384	9.620	< 0.001
		65.0	4.717	183.966	26.837	< 0.001
		68.5	4.613	179.911	51.167	< 0.001
	R4	61.5	3.694	144.068	6.663	< 0.001
		65.0	3.979	155.196	15.848	< 0.001
		68.5	3.973	154.956	25.190	< 0.001

Table 7. One-way repeated measures Analysis of Variance (ANOVA) results for Test 3. Corrections for violations of sphericity are performed using the Greenhouse-Geisser method.

Next, post-hoc pairwise comparisons were performed with the Bonferroni method to determine which rattle levels differ significantly on annoyance, and it is found that the combination of boom and rattle is more annoying than the boom alone in 15 out of 18 cases ($p \leq 0.05$).

An additional dummy variable regression analysis was performed to estimate rattle penalties. The objective of this analysis is to model subjective annoyance as it varies with total PL for four cases: boom only, boom and Rattle 1, boom and Rattle 3, and boom and Rattle 4. Annoyance to each boom alone and annoyance to each boom and rattle combination with the rattle at a level of $-2.4\,\text{dB}$ (relative to the isolated rattle level) for the three different PL cases are included in the regression with the total PL value of each sound. This rattle level is chosen because it is the highest rattle level tested for sounds that are mixtures of boom and rattle. It is expected that the data can be modeled by linear relationships in this test's limited PL range and that the regression lines for the boom and rattle mixtures will be different from the boom only regression line.

It is found that a regression model that includes interaction of total PL with rattle type does not significantly differ from a simpler additive model with no interactions ($p > 0.05$). This means that the relative increase in annoyance for a given increase in total PL is the same regardless of rattle type (including the category of no rattle). The simpler model is therefore chosen, which results in parallel regression lines with equal slope, as shown in Fig. 21. The predicted annoyance from this simple model is represented by

$$Y' = A + B_1 D_1 + B_2 D_2 + B_3 D_3 + B_4 X, \tag{2}$$

where Y' is the predicted annoyance; A, B_1, B_2, B_3, and B_4 are the regression coefficients; D_1, D_2, and D_3 are dummy variables representing rattle categories; and X is the total PL. It is shown that combining booms with Rattle 3 results in the largest difference in annoyance from the boom only case, despite the sounds having the same PL value. This annoyance difference can be expressed in equivalent units of PL by utilizing the slope of the regression lines. The dB difference in total PL between the boom only case and each boom and rattle combination for equal annoyance is found to be 3.62, 8.85, and 6.38 dB for Rattles 1, 3, and 4, respectively. In other words, a combination of Rattle 1 and a boom is as annoying as a boom alone that is 3.62 dB louder in PL. This illustrates that the metric PL is not adequately accounting for the added annoyance of introducing a rattle noise.

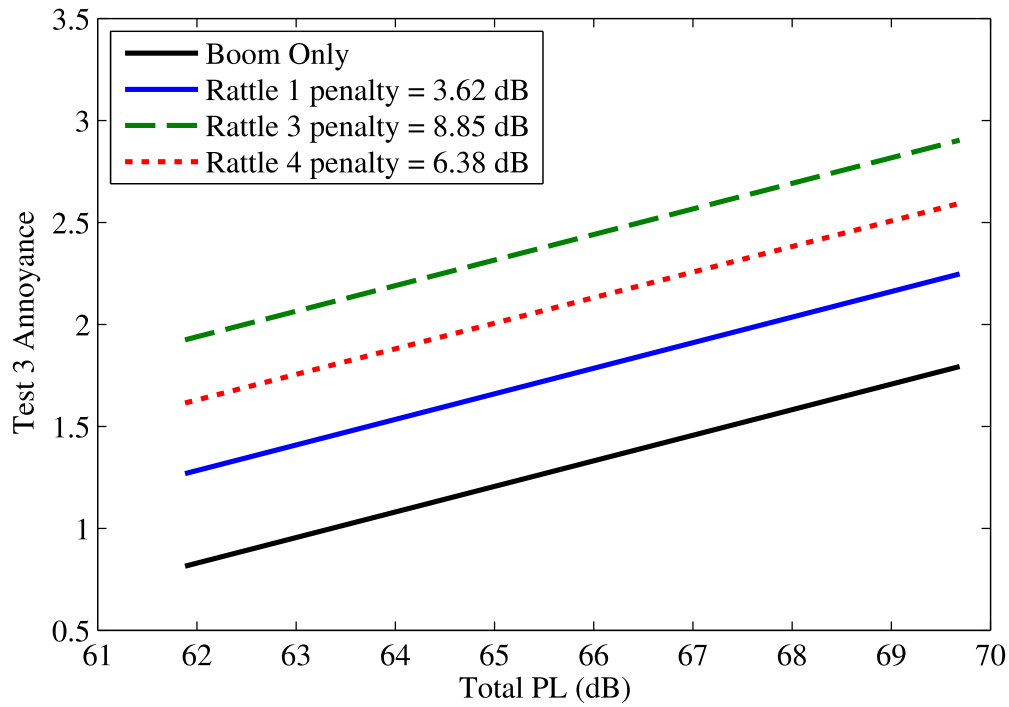

Figure 21. Test 3 dummy regression analysis for a simple additive model with no interactions.

The rattle penalties calculated here fall within the lower part of the range reported in the literature, as summarized in Sec. 1.1. The values are reasonably consistent with the 5 dB rattle penalty found by Fidell *et al.* [15] for a recorded rattle with simulated indoor booms. The benefit of the current research and calculated rattle penalties is in the variety of rattles tested, the control and definition of these rattles, and application to low-amplitude sonic booms.

As in Tests 1 and 2, it is certain that normalizing the sounds to have the same PL value still results in differences in annoyance. The metric PL therefore does not sufficiently describe annoyance to the tested rattle sounds or to combinations of sonic booms and rattle sounds. PL should not be used alone to predict annoyance to these sounds.

7 Objective Metrics Analysis and Subjective Annoyance Predictions

This section reports the relationship between objective metrics and subjective annoyance for all signals used in Tests 1-3. Each test is treated separately due to the different nature and objective of each test. Correlations between objective metrics and subjective response are given, and regression analyses for constructing human response models are presented for Tests 2 and 3.

7.1 Objective Metrics

The metrics chosen for analysis belong to two psychoacoustic categories: loudness and sound quality. Loudness metrics are selected because louder sounds are generally rated as more annoying than quieter sounds. Although Perceived Level was used to normalize the sounds, other loudness metrics still detect differences between the sounds. Different loudness metrics use different frequency weightings, and investigation of a variety of these metrics may help explain the spectral balance factor that was identified in Test 1 as being important. Sound quality metrics are selected to further characterize sounds. Sound quality metrics use models of human hearing to quantify characteristics of sound signals above and beyond loudness. The following metrics are selected for analysis (see App. D for descriptions of the metrics):

- Sound Exposure Level
  - A-weighted (ASEL)
  - C-weighted (CSEL)
  - Unweighted (ZSEL)
- Perceived Level (PL)
- Perceived Noise Level (PNL)
- Zwicker Loudness Level
  - Frontal incidence (LLZ_f)
  - Diffuse incidence (LLZ_d)
- Moore and Glasberg Stationary Loudness (MGSL)
- Moore and Glasberg Time-Varying Loudness (MGTVL)
- Loudness DIN45631

- Loudness HEAD
- Loudness ISO532A
- Loudness ISO532B
- Relative Approach
- Roughness
- Hearing Model Roughness
- Hearing Model Impulsiveness
- Kurtosis
- Tonality

With the exception of the Moore and Glasberg metrics, these objective metrics are intended to be calculated on monaural signals. However, the test sounds in all three tests were recorded and played back binaurally across several pairs of headphones, used by subjects and also attached to a binaural head. The following procedure is followed to yield a single, average metric value across the individual channels. First the metrics are calculated individually for each channel. The higher metric value from each headset pair is retained, and the median of these values is reported. It is found that correlations change by only 0.01 if an alternate method is used, such as taking the mean metric value across headphones of the mean metric value between binaural channels.

For time-varying metrics, there is no standard to prescribe whether the maximum metric value or the time-integrated value of the metric is to be reported. For some time-varying metrics the correlation changes markedly depending on whether the maximum or time-integrated values are used. However, there is no systematic pattern in the variation, and the metrics with the highest correlation remain the highest regardless of whether maximum or time-integrated values are used. The correlations reported here are calculated using the maximum metric value.

7.2 Correlations Between Metrics and Subjective Annoyance for Test 1

Objective metrics are calculated for all signals in Test 1; only the subjective data from the category line scaling subtest is considered here. Pearson product-moment correlation coefficients (r) [22] are calculated to demonstrate the strength of linear dependence between values of each metric and average annoyance for each signal in the test. The results are displayed in Table 8. The square of the correlation coefficients, the coefficients of determination (r^2), are also given to illustrate the proportion of variability in annoyance explained by each metric.

For the rattle sounds in Test 1, there are significant correlations ($p < 0.0001$) for about one third of the objective metrics, as denoted by asterisks in Table 8. More significant correlations or a higher degree of correlation would be expected if the signals spanned a greater range of metric values [22]. The small range of some metric values is an artifact of having normalized signals to the same PL value (70±1 dB) in this study. In fact, correlations

Metric	Test 1 r	Test 1 r^2
ASEL	-0.78^{**}	0.61^{**}
CSEL	0.73^{**}	0.53^{**}
ZSEL	0.72^{**}	0.52^{**}
PNL	-0.67^{**}	0.45^{**}
$\mathrm{LLZ_f}$	0.62^{**}	0.38^{**}
$\mathrm{LLZ_d}$	0.62^{**}	0.38^{**}
MGTVL	0.47	0.22
Tonality	−0.36	0.13
Relative Approach	0.33	0.11
Roughness	−0.33	0.11
Loudness ISO532A	−0.31	0.10
MGSL	0.21	0.04
Loudness DIN45631	−0.19	0.04
HM Roughness	0.18	0.03
Kurtosis	−0.10	0.01
Loudness ISO532B	−0.10	0.01
HM Impulsiveness	−0.09	0.01
Loudness HEAD	0.07	0.00

Table 8. Test 1 correlation coefficients (r) and coefficients of determination (r^2) for objective metrics and subjective annoyance ($N = 40$). $^{**}\,p < 0.0001$

with PL are not listed because the PL values are nominally the same for all signals, and a correlation would not be valid.

Significant correlations with annoyance to rattle sounds are found for traditional loudness metrics, such as SEL, PNL, and LLZ. It is worth noting that ASEL and PNL show a high negative correlation with annoyance. This is probably another result of the PL normalization. Some rattle sounds with more low-frequency content are found to result in higher annoyance. ASEL and PNL apply a steeper low-frequency rolloff than PL and consequently assign lower metric values for these sounds, while PL remains constant. ASEL and PNL therefore do not adequately account for the low-frequency effects that may cause higher annoyance. Finally, more advanced loudness metrics and the chosen sound quality metrics do not describe annoyance to these rattles well.

7.3 Test 2 Metrics Analysis

7.3.1 Correlations Between Metrics and Subjective Annoyance for Test 2

The signals in Test 2 contain sonic boom and rattle sounds mixed together. All signals are normalized to a PL of $65 \pm 1\,\mathrm{dB}$, which implies a limited loudness level range, as in Test 1. The addition of sonic booms results in very different correlations from Test 1. As shown in Table 9, the metrics with the highest correlation with subjective annoyance are MGSL, Loudness ISO532A, and Roughness, all of which have low correlations to isolated rattles in Test 1. It therefore appears that these metrics correlate highly with annoyance due to the presence of sonic booms.

As in Test 1, correlations with PL are not reported because of invalidity due to a trivial range of values. Range restriction in several metrics is also present due to the PL normalization. ASEL, PNL, and several other metrics show a negative correlation due to the peculiarities of PL normalization, similar to Test 1.

Metric	Test 2 r	Test 2 r^2
MGSL	0.87**	0.76**
Loudness ISO532A	0.82**	0.67**
Roughness	0.76**	0.58**
ASEL	−0.74**	0.55**
Loudness DIN45631	0.70**	0.50**
HM Impulsiveness	0.63**	0.40**
Relative Approach	−0.61**	0.37**
PNL	−0.50**	0.25**
Loudness HEAD	0.47**	0.22**
CSEL	−0.44**	0.19**
ZSEL	−0.39**	0.15**
MGTVL	−0.30*	0.09*
Loudness ISO532B	0.27*	0.07*
HM Roughness	0.23	0.05
Tonality	−0.13	0.02
LLZ_f	−0.10	0.01
Kurtosis	0.08	0.01
LLZ_d	−0.06	0.00

Table 9. Test 2 correlation coefficients (r) and coefficients of determination (r^2) for objective metrics and subjective annoyance ($N = 149$). $^* p < 0.001$, $^{**} p < 0.0001$

7.3.2 Human Response Model for Test 2

While correlation analysis indicates the strength of the relationship between annoyance and objective metrics, multiple linear regression can be used to construct a model that estimates annoyance from linear combinations of the noise metrics. The best relationship between annoyance and metrics is sought that also uses a minimum number of metrics in the prediction.

Based on the correlations presented in Sec. 7.3.1, each metric's correlation strength is assigned using Cohen's effect size criteria [9]. All metrics with a trivial ($|r| < 0.1$) or small ($0.1 \leq |r| < 0.3$) effect size are eliminated from consideration in the multiple regression model. For Test 2, Loudness ISO532B, HM Roughness, Tonality, LLZ_f, Kurtosis, and LLZ_d are eliminated.

Additionally, metrics containing only a small range of values are eliminated, because a restriction in range can invalidate use of the regression model for predictions beyond the current sample of signals [22]. Generally, range restriction tends to decrease the degree of correlation, and a correction formula has been developed to estimate the correlation for the case of an unrestricted range [59]. This correction, however, is not always valid [59], and in the current analysis the range-restricted metrics are simply removed from the analysis. After consideration of effect size in Test 2, ASEL is the only remaining metric that exhibits range restriction, and thus it is eliminated.

The remaining eleven metrics are included in the development of a multiple linear regression model using a 'stepwise' method. This technique for screening variables is useful when linear dependence exists between the metrics, as is the case here. A minimum tolerance criterion of 0.1 is set to avoid multicollinearity [41] and to only allow inclusion of metrics that result in a significant increase in the explained regression, denoted by the coefficient of multiple determination R^2.

The resulting optimum multiple regression equation contains five metric variables for Test 2: MGSL, HM Impulsiveness, Loudness DIN45631, CSEL, and Loudness HEAD. A linear combination of these five metrics results in the best estimate of the observed annoyance from Test 2. It is desired, however, to use these metrics to predict annoyance to other sounds. Overfitting the model to current data can occur with a large number of variables. Instead, a more efficient model with fewer variables is sought to establish a general relationship. Examining the change in R^2 with the addition of each metric to the model is used to accomplish this objective. If the change in R^2 for inclusion of a metric is less than 0.05, then the preceding model without the last metric is chosen as the final model. For Test 2, this results in a final reduced multiple regression model including only MGSL and HM Impulsiveness, as given by the following equation:

$$Annoyance = -3.819 + 0.116 * MGSL - 0.310 * HMImpulsiveness. \tag{3}$$

The simplest model that includes only MGSL would account for 75.8% of the variation in annoyance in Test 2, and the above model that additionally includes HM Impulsiveness accounts for 82% of the variation; this change of 6.2% is considered significant enough to warrant inclusion of the extra metric in the model. A plot of the predicted annoyance vs. actual annoyance for Test 2 is shown in Fig. 22. The correlation between predicted and reported annoyance is shown both for the initial regression model that includes five variables (MGSL, HM Impulsiveness, Loudness DIN45631, CSEL, and Loudness HEAD) and the final reduced model that includes only two variables (MGSL and HM Impulsiveness).

7.4 Test 3 Metrics Analysis

7.4.1 Correlations Between Metrics and Subjective Annoyance for Test 3

The signals in Test 3 contain both sonic boom and rattle sounds, and they are normalized to PL values of 61.5, 65, and 68.5 dB. In contrast to Tests 1 and 2, the increased variation in PL causes almost all loudness metrics to correlate highly with subjective annoyance, as shown in Table 10. Test 3 may be the only test for which there is enough variation in loudness level to exercise each metric to a satisfactory degree. With this variation in PL, all correlations are also positive in Test 3, except for Tonality, which has an extremely low, insignificant correlation and can be ignored. Additionally, the majority of the correlations are significant beyond the 0.0001 level in Test 3. The metric with the highest correlation to annoyance is MGSL, followed closely by Loudness HEAD and Loudness DIN45631.

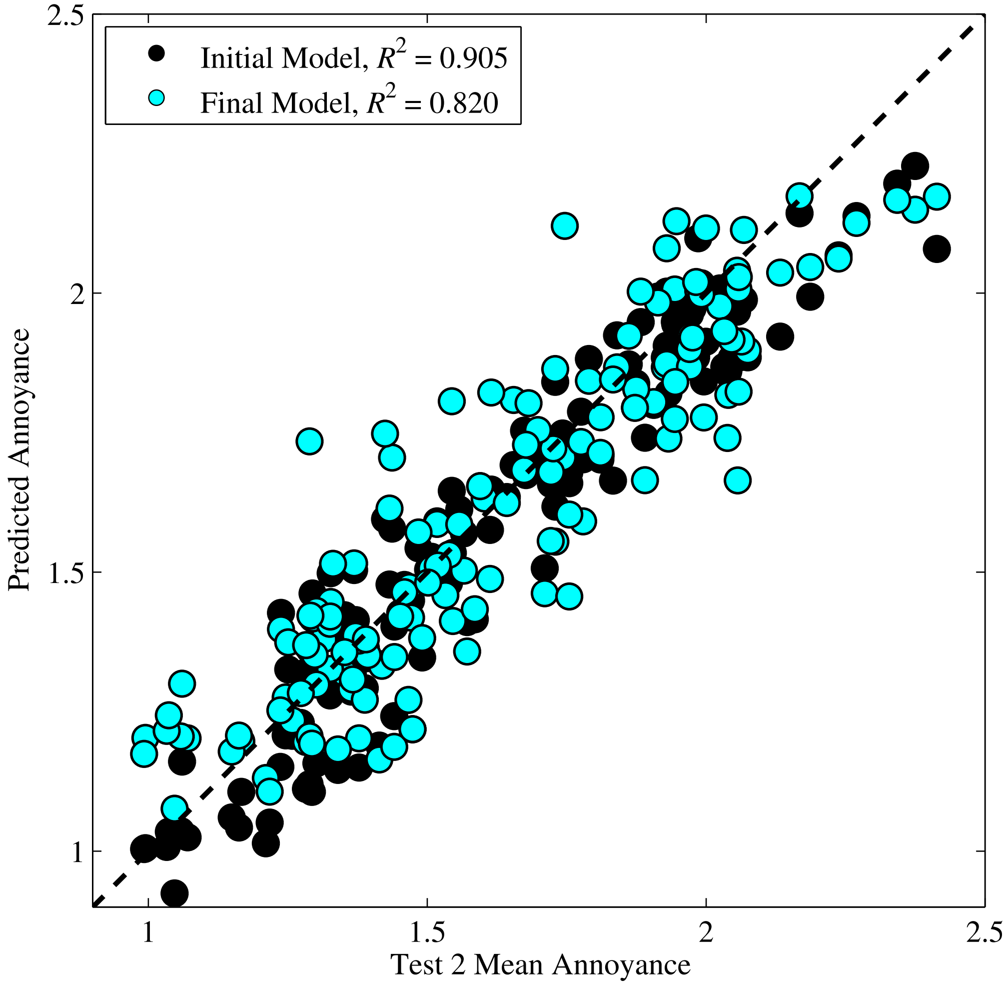

Figure 22. Predicted annoyance vs. reported annoyance for both the initial and the reduced regression equations for Test 2. The final reduced model accounts for 82% of the variation in reported annoyance.

Metric	Test 3 r	Test 3 r^2
MGSL	0.88^{**}	0.77^{**}
Loudness HEAD	0.86^{**}	0.74^{**}
Loudness DIN45631	0.83^{**}	0.68^{**}
Loudness ISO532B	0.80^{**}	0.64^{*}
Loudness ISO532A	0.79^{**}	0.63^{**}
MGTVL	0.74^{**}	0.55^{**}
PL	0.69^{**}	0.48^{**}
Roughness	0.68^{**}	0.47^{**}
LLZ_d	0.67^{**}	0.45^{**}
LLZ_f	0.65^{**}	0.42^{**}
PNL	0.47^{**}	0.23^{**}
ASEL	0.46^{**}	0.21^{**}
HM Roughness	0.44^{**}	0.20^{**}
HM Impulsiveness	0.42^{**}	0.17^{**}
ZSEL	0.41^{**}	0.17^{**}
Relative Approach	0.40^{**}	0.16^{**}
CSEL	0.37^{**}	0.14^{**}
Kurtosis	0.26	0.07
Tonality	-0.09	0.01

Table 10. Test 3 correlation coefficients (r) and coefficients of determination (r^2) for objective metrics and subjective annoyance ($N = 105$). $^{**}\, p < 0.0001$

As shown in Fig. 23, the Moore and Glasberg Stationary Loudness (MGSL) metric predictions exhibit a linear relationship and relatively high correlation with mean annoyance in both Tests 2 and 3. It is interesting to note that the MGSL metric was devised for steady sounds, not transient sounds as studied here, yet it still predicts the mean annoyance better than all the other metrics calculated for these tests.

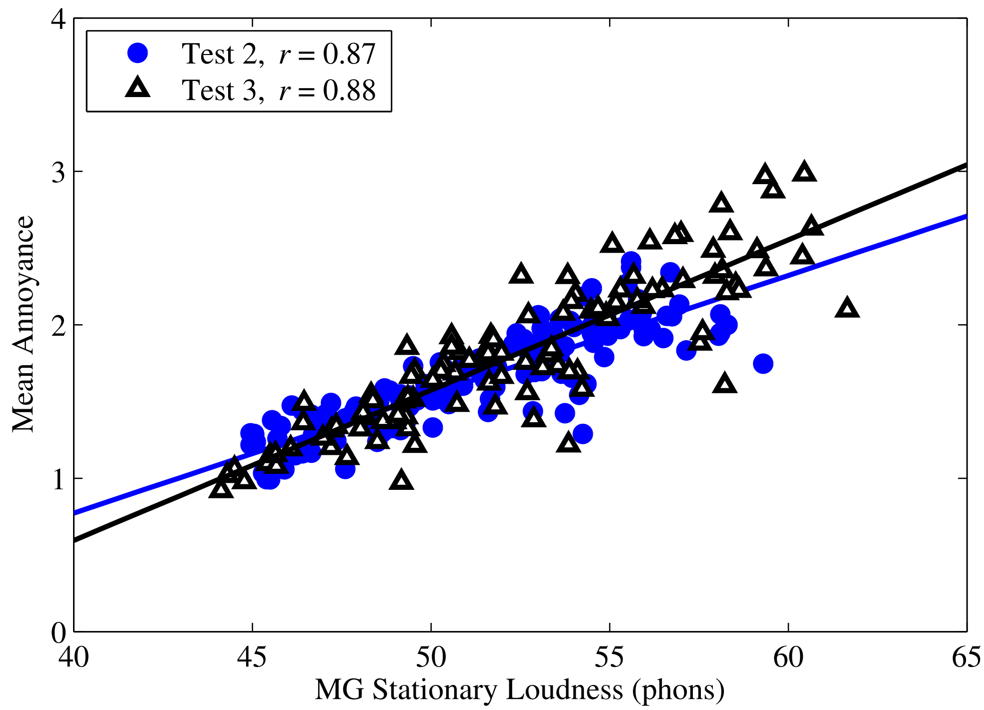

Figure 23. Mean annoyance correlation with Moore and Glasberg Stationary Loudness predictions for Tests 2 and 3 ($p < 0.0001$).

It is worth noting that the correlation of annoyance with certain metrics for boom and rattle mixtures varies according to the particular rattle present in the mixture. An example of this is shown below in Fig. 24, where annoyance has a much higher correlation with roughness for boom mixtures with Rattle 3 ($r = 0.91$) than for boom mixtures with Rattle 1 ($r = 0.60$) or Rattle 4 ($r = 0.79$). Regardless, only total correlations with all the sounds are reported in Table 10. In the example case, the overall correlation of annoyance with roughness is $r = 0.68$.

7.4.2 Human Response Model for Test 3

A multiple linear regression model is also developed for Test 3. The same considerations for effect size from Test 2 are applied, which eliminates Kurtosis and Tonality. Because of the increased variation in PL, none of the metrics exhibit a restricted range, and no other metrics can be eliminated from the analysis.

The remaining seventeen metrics are used in a stepwise regression procedure, and the resulting multiple regression equation contains four metrics for Test 3: MGSL, CSEL, ASEL,

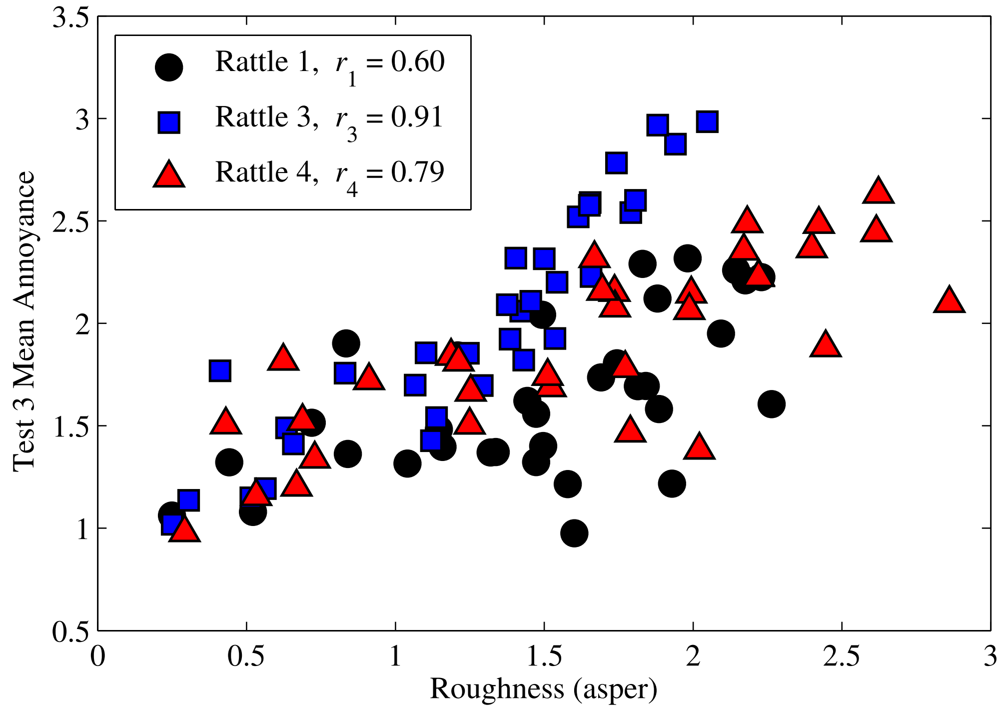

Figure 24. Mean annoyance vs. roughness for Test 3 boom and rattle sounds plotted by constituent rattle sounds. The overall correlation of annoyance with roughness is $r = 0.68$ ($p < 0.0001$).

and HM Imuplsiveness. A linear combination of these four metrics results in the best estimate of the observed annoyance from Test 3. For application of the prediction model to other signals, an efficient model with the smallest number of metric variables that can still describe the data adequately is desired. Therefore, as in Test 2, the change in R^2 is examined to quantify the utility of each metric added to the model. This analysis results in a final reduced multiple regression model that includes MGSL, CSEL, and ASEL, as given by the following equation:

$$Annoyance = -4.386 + 0.119 * MGSL + 0.038 * CSEL - 0.058 * ASEL. \tag{4}$$

The simplest model that includes only MGSL would account for 76.7% of the variation in annoyance in Test 3, and the above final model that additionally includes CSEL and ASEL accounts for 92.1% of the variation. This large change of 15.4% justifies inclusion of the two extra metrics in the model. The predicted annoyance versus actual annoyance for Test 3 is shown in Fig. 25. The correlation between predicted and measured annoyance is shown both for the initial regression model that includes four variables (MGSL, CSEL, ASEL, and HM Imuplsiveness) and the final reduced model that includes only three variables (MGSL, CSEL, and ASEL).

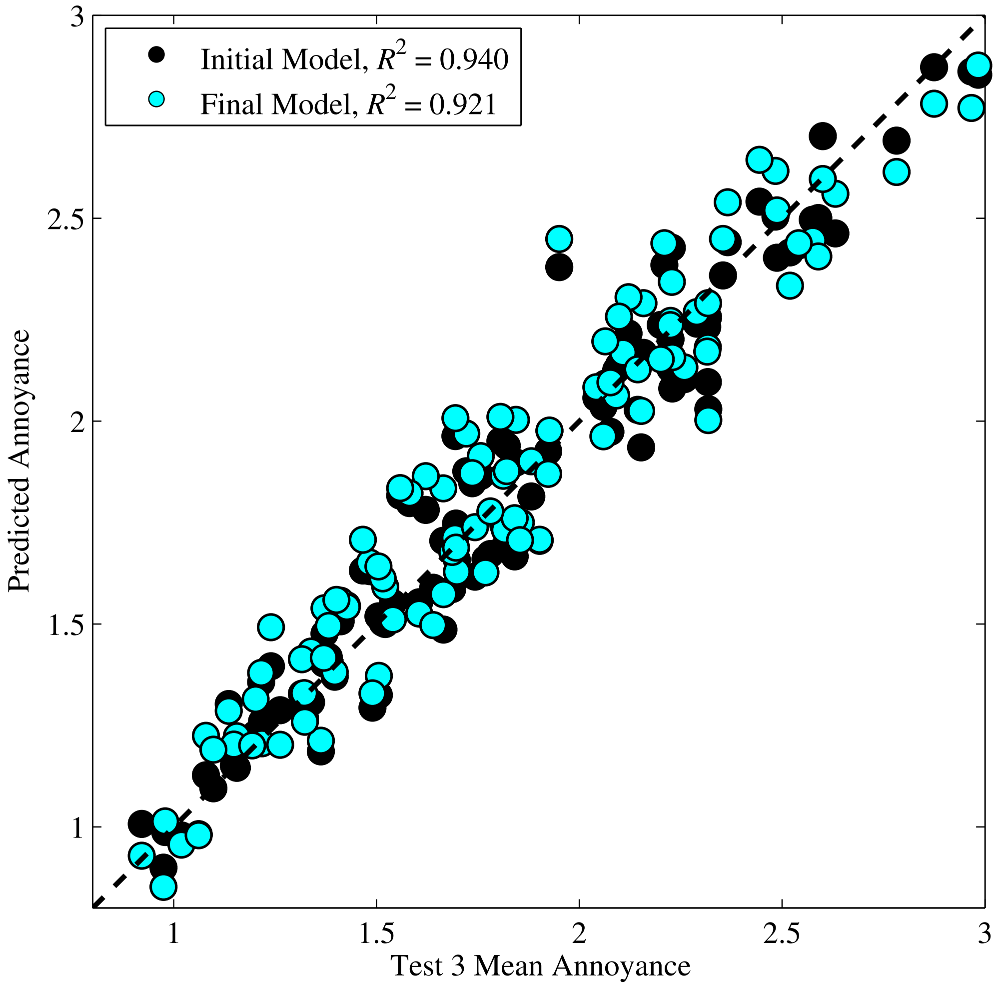

Figure 25. Predicted annoyance vs. reported annoyance for both the initial and the reduced regression equations for Test 3. The final reduced model accounts for 92.1% of the variation in reported annoyance.

7.5 Summary of Metrics Analysis

Correlations of metrics with annoyance for rattle sounds in Test 1 fail to identify a sound quality metric that can describe annoyance beyond that explained by loudness level. It is found that low-frequency content in rattle sounds leads to a higher annoyance, and some traditional metrics such as ASEL and PNL do not account for this effect. On the other hand, the metrics CSEL and ZSEL, which have less or no low-frequency rolloff, respectively, do exhibit a reasonably high amount of correlation with annoyance.

The human response models developed for Tests 2 and 3 differ due to the particular types of signals tested. Holding PL constant or varying PL over a modest range has an effect on the derived models for predicting annoyance to sonic boom and rattle sounds. For example, normalization to nominally a single PL value in Test 2 does not allow for meaningful correlations with certain metrics, notably ASEL. The MGSL metric, however, while highly correlated to many of the other loudness metrics, appears as the first choice in building a human response model for both tests. The explained variance can be increased by including additional metrics as linear terms in the models. A measure of impulsiveness appears to account for a small, but significant, portion of variation in annoyance when PL is held constant. When PL is varied over a range of 7 dB, CSEL and ASEL are chosen as the significant additional contributors to the model.

8 Summary and Conclusions

A series of five subjective tests was conducted to explore annoyance to sonic boom-induced rattle sounds in an indoor environment. A collection of 40 binaural rattle sounds with varying temporal and spectral properties were studied. A subset of these sounds was also combined with up to four low-amplitude indoor sonic booms for further study.

Annoyance to different rattles is shown to vary significantly, and an annoyance rank ordering of the rattles was performed. Of the different psychometric methods employed, category line scaling was chosen as the preferred method for follow-on studies because of its efficiency and favorable comparison with paired comparison results. A difference in annoyance exists between rattle sounds despite their presentation to subjects at the same Perceived Level (PL). In general, annoyance increases as the size of the rattling object increases. For example, rattles emanating from structural components of a house were found to be more annoying than rattles from bric-a-brac. An increase in low-frequency content of the sound with larger objects appears to explain this effect. An investigation into the different factors that contribute to annoyance to these rattle sounds found that the most important characteristics are spectral balance, level, temporal variability, impulsivity, complexity, and envelopment of the sound.

It is found that the combination of sonic booms and rattles is often more annoying than the sonic boom alone at equal PL, at any of the three PL values tested. Because sounds were normalized to the same PL values, these results show that the PL metric does not fully predict human annoyance to the selected indoor sonic boom and rattle sounds. In order to quantify the effect of rattle on annoyance to low-amplitude sonic booms in an indoor environment, a rattle penalty analysis was performed. The rattle penalty ranges from 3.6 to 8.9 dB, depending on the rattle sound. In other words, an increase in boom PL of 3.6 to 8.9 dB would result in an increase in annoyance equivalent to that due to the additional presence of rattle with a boom.

Analysis of metrics shows that most sound quality metrics and traditional loudness metrics, such as SEL and PNL, are poor predictors of annoyance to the sonic boom and rattle sounds. One advanced metric that does correlate well with mean annoyance is Moore and Glasberg Stationary Loudness (MGSL), which accounts for transmission through the outer and middle ear and considers the absolute hearing threshold spectrum. Linear combinations of metrics are shown to result in human response models that are able to predict annoyance more accurately. These models identify psychoacoustic metrics to describe annoyance beyond that explained by PL. For the sounds studied, a successful annoyance model includes MGSL in combination with HM Impulsiveness when PL is held constant. When a modest variation of 7 dB in PL is introduced, the annoyance model includes MGSL in combination with CSEL and ASEL. The models should be applied to other sonic boom and rattle signals for validation.

These studies indicate that the presence of rattle is an important contributor to annoyance of low-amplitude sonic booms heard indoors. A large library of rattle sounds for controlled studies has been created which spans a range of psychoacoustic metrics, and a subset of these rattles has been identified as applicable for more detailed experiments. These tests, however, were performed with sounds presented binaurally over headphones, which have a limited low-frequency response and thus cannot produce the full spectrum of sonic booms. In addition, filtering of the sonic booms to simulate structural transmission and indoor reception was approximate. Despite these limitations, results can be used to design future sonic boom and rattle studies in a facility capable of accurately reproducing the indoor boom, such as the Interior Effects Room at NASA Langley Research Center [31,32].

References

1. *Civil Aircraft Sonic Boom*. Code of Federal Regulations, Title 14, Pt. 91.817, 2011.

2. ANSI. *American National Standard acoustical terminology*. ANSI S1.1-1994, 1994.

3. ANSI. *American National Standard procedure for the computation of loudness of steady sounds*. ANSI S3.4-2007, 2007.

4. R. L. Bennett and K. S. Pearsons. Handbook of aircraft noise metrics. Technical Report N81-21871, U.S. Department of Commerce, National Technical Information Service, 1981.

5. P. N. Borsky. Community reactions to sonic booms in the Oklahoma City area: Vol II: Data on community reactions and interpretations. Technical Report AMRL-TR-65-37, Aerospace Medical Research Laboratories, Wright-Patterson Air Force Base, 1965.

6. P. N. Borsky. Sonic boom exposure effects II.4: Annoyance reactions. *J. Sound Vib.*, 20(4):527–530, 1972.

7. W. R. Bray. The "Relative Approach" for direct measurement of noise patterns. *Sound and Vibration*, 20–22, September 2004.

8. J. M. Cawthorn, T. K. Dempsey, and R. DeLoach. Human response to aircraft-noise-induced building vibration. Technical Report CP-2052, NASA, 1978.

9. J. Cohen. A power primer. *Psychological Bulletin*, 112(1):155–159, 1992.

10. H. A. David. *The Method of Paired Comparisons*. Charles Griffin & Company, 1988.

11. P. Davies and A. J. Marshall. Personal communication, November 2008.

12. DIN. *Procedure for calculating loudness level and loudness*. DIN 45631, 1991.

13. DIN. *Calculation of loudness level and loudness from the sound spectrum - Zwicker method - Amendment 1: Calculation of the loudness of time-variant sound*. DIN 45631/A1, 2010.

14. H. Fastl, F. Völk, and M. Straubinger. Standards for calculating loudness of stationary or time-varying sounds. 1435–1440. Internoise 2009, Ottawa, Canada, 2009.

15. S. Fidell, L. Silvati, and K. Pearsons. Relative rates of growth of annoyance of impulsive and non-impulsive noises. *J. Acoust. Soc. Am.*, 111(1):576–585, 2002.

16. J. M. Fields. Reactions of residents to long-term sonic boom noise environments. Technical Report CR-201704, NASA, 1997.

17. K. Genuit. Objective evaluation of acoustic quality based on a relative approach. 3233–3238. Internoise 1996, Liverpool, UK, 1996.

18. B. R. Glasberg and B. C. J. Moore. A model of loudness applicable to time-varying sounds. *J. Audio Eng. Soc.*, 50(5):331–342, 2002.

19. J. Guilford. *Psychometric Methods*. McGraw-Hill Book Company, 1954.

20. E. A. Haering, Jr., L. J. Cliatt, T. J. Bunce, T. B. Gabrielson, V. W. Sparrow, and L. L. Locey. Initial results from the variable intensity sonic boom propagation database. In *14th AIAA/CEAS Aeroacoustics Conference (29th AIAA Aeroacoustics Conference)*, 1–58, 2008.

21. E. A. Haering, Jr., J. W. Smolka, J. E. Murray, and K. J. Plotkin. Flight demonstration of low overpressure N-wave sonic booms and evanescent waves. In A. A. Atchley, V. W. Sparrow, and R. M. Keolian, editors, *Innovations in Nonlinear Acoustics, ISNA17, 17th International Symposium on Nonlinear Acoustics*, volume 838, 647–650. AIP, 2006.

22. W. L. Hays. *Statistics*. Wadsworth, 5th edition, 1994.

23. HEAD Acoustics. *Psychoacoustic analyses in ArtermiS I*. HEAD Application Note, 2010.

24. HEAD Acoustics. *Psychoacoustic analyses in ArtermiS II*. HEAD Application Note, 2010.

25. ISO. *Acoustics - Method for calculating loudness level*. ISO 532-1975, 1975.

26. ISO. *Acoustics - Reference zero for the calibration of audiometric equipment - Part 1: Reference equivalent threshold sound pressure levels for pure tones and supra-aural earphones*. ISO 389-1:1998(E), 1998.

27. ISO. *Acoustics - Method for calculating loudness - Part 1: Stationary sounds*. Draft ISO/DIS 532-1, 2011.

28. G. M. Jackson and H. G. Leventhall. Calculation of the perceived level of noise (PLdB) using Stevens' method (Mark VII). *Applied Acoustics*, 6:23–34, 1973.

29. D. R. Johnson and D. W. Robinson. The subjective evaluation of sonic bangs. *Acustica*, 18(5):241–258, 1967.

30. J. Klos. Vibro-acoustic response of buildings due to sonic boom exposure: July 2007 field test. Technical Report TM-2008-215349, NASA, 2008.

31. J. Klos. Overview of an indoor sonic boom simulator at NASA Langley Research Center. In *INTER-NOISE 2012 Proc.*, 8973–8983, 2012.

32. J. Klos, B. M. Sullivan, and K. P. Shepherd. Design of an indoor sonic boom simulator at NASA Langley Research Center. In *Noise-Con Proc.*, volume 117, 535–546. Note: Some facility details have changed since this publication, 2008.

33. K. D. Kryter. Scaling human reactions to the sound from aircraft. *J. Acoust. Soc. Am.*, 31(11):1415–1429, 1959.

34. K. D. Kryter, P. J. Johnson, and J. R. Young. Psychological experiments on sonic booms conducted at Edwards Air Force Base. Technical Report Contract AF49(638)-1758, Stanford Research Institute, Menlo Park, CA, 1968.

35. J. D. Leatherwood and B. M. Sullivan. Laboratory study of effects of sonic boom shaping on subjective loudness and acceptability. Technical Report TP-3269, NASA, 1992.

36. J. D. Leatherwood and B. M. Sullivan. Loudness and annoyance response to simulated outdoor and indoor sonic booms. Technical Report TM-107756, NASA, 1993.

37. J. D. Leatherwood, B. M. Sullivan, K. P. Shepherd, D. A. McCurdy, and S. A. Brown. Summary of recent NASA studies of human response to sonic booms. *J. Acoust. Soc. Am.*, 111(1):586–598, 2002.

38. B. C. J. Moore, B. R. Glasberg, and T. Baer. A model for the prediction of thresholds, loudness, and partial loudness. *J. Audio Eng. Soc.*, 45(4):224–240, 1997.

39. C. W. Nixon and P. N. Borsky. Effects of sonic boom on people: St. Louis, Missouri, 1961-1962. *J. Acoust. Soc. Am.*, 39(5, Pt. 2):S51–S58, 1966.

40. C. W. Nixon and H. H. Hubbard. Results of USAF-NASA-FAA flight program to study community response to sonic booms in the Greater St. Louis area. Technical Report TN D-2705, NASA, 1965.

41. R. M. O'Brien. A caution regarding rules of thumb for variance inflation factors. *Quality and Quantity*, 41:673–690, 2007.

42. J. B. Ollerhead. An evaluation of methods for scaling aircraft noise perception. Technical Report CR-1883, NASA, 1971.

43. E. Parizet and V. N. Nosulenko. Multi-dimensional listening test: Selection of sound descriptors and design of the experiment. *Noise Control Eng. J.*, 47(6):227–232, 1999.

44. K. S. Pearsons and K. D. Kryter. Laboratory tests of subjective reactions to sonic boom. Technical Report CR-187, NASA, 1965.

45. K. S. Pearsons, B. Tabachnick, R. Howe, K. K. Ahuja, and J. C. Stevens. A study of the effects of sonic boom waveform modification on annoyance. NASA CR (Contract NAS1-19061), unpublished, 1993.

46. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in FORTRAN: The Art of Scientific Computing*. Cambridge University Press, New York, NY, 2nd edition, 1992.

47. P. D. Schomer. Human response to house vibrations caused by sonic booms or air blasts. *J. Acoust. Soc. Am.*, 64(1):328–330, 1978.

48. P. D. Schomer and A. Averbuch. Indoor human response to blast sounds that generate rattles. *J. Acoust. Soc. Am.*, 86(2):665–673, 1989.

49. P. D. Schomer and R. D. Neathammer. The role of helicopter noise-induced vibration and rattle in human response. *J. Acoust. Soc. Am.*, 81(4):966–976, 1987.

50. S. Sharma. *Applied Multivariate Techniques*. John Wiley & Sons, 1996.

51. K. P. Shepherd and B. M. Sullivan. A loudness calculation procedure applied to shaped sonic booms. Technical Report TP-3134, NASA, 1991.

52. R. Sottek and K. Genuit. Models of signal processing in human hearing. *Int. J. Electron. Commun.*, 59:157–165, 2005.

53. R. Sottek and K. Genuit. Sound quality evaluation of fan noise based on advanced hearing-related parameters. *Noise Control Eng. J.*, 57(4):384–390, 2009.

54. S. S. Stevens. Procedure for calculating loudness: Mark VI. *J. Acoust. Soc. Am.*, 33(11):1577–1585, 1961.

55. S. S. Stevens. Perceived level of noise by Mark VII and decibels (E). *J. Acoust. Soc. Am.*, 51(2):575–601, 1972.

56. B. M. Sullivan, J. Klos, R. D. Buehrle, D. A. McCurdy, and E. A. Haering, Jr. Human response to low-intensity sonic booms heard indoors and outdoors. Technical Report TM-2010-216685, NASA, 2010.

57. E. Terhardt. Acoustic roughness and fluctuation strength. Technical Translation NASA-TT-F-13941. [Trans. from Acustica, 20:215-224, 1968], 1971.

58. E. Terhardt, G. Stoll, and M. Seewann. Algorithm for extraction of pitch and pitch salience from complex tonal signals. *J. Acoust. Soc. Am.*, 71(3):679–688, 1982.

59. D. W. Zimmerman and R. H. Williams. Restriction of range and correlation in outlier-prone distributions. *Applied Psychological Measurement*, 24(3):267–280, 2000.

60. E. Zwicker and H. Fastl. *Psychoacoustics: Facts and Models*. Springer Verlag, Heidelberg, Germany, 1990.

Appendix A

Participants in Tests 1-3

Test subjects were recruited from the local Hampton Roads community and were compensated for their participation. Subjects received an audiometric test beforehand to confirm that their hearing was within 40 dB of reference hearing threshold levels [26]. The following table lists the number of participants, gender classification, and mean age for each of the tests conducted.

Test	Number of Participants	Gender		Mean Age
		Male (%)	Female (%)	
Test 1 Paired Comparison	24	45.8%	54.2%	40
Test 1 Category Line Scaling	24	25.0%	75.0%	43
Test 1 Semantic Differential	24	37.5%	62.5%	51
Test 2	55	38.2%	61.8%	44
Test 3	40	50.0%	50.0%	30

Table A1. Information on participants in Tests 1-3.

Appendix B

Test 2 Annoyance Figures

This appendix includes figures of the Test 2 mean annoyance as a function of rattle level. Figures B1-B4 each present the mean annoyance for a single boom and five rattles. Figures B5-B9 each present the mean annoyance for a single rattle and four booms.

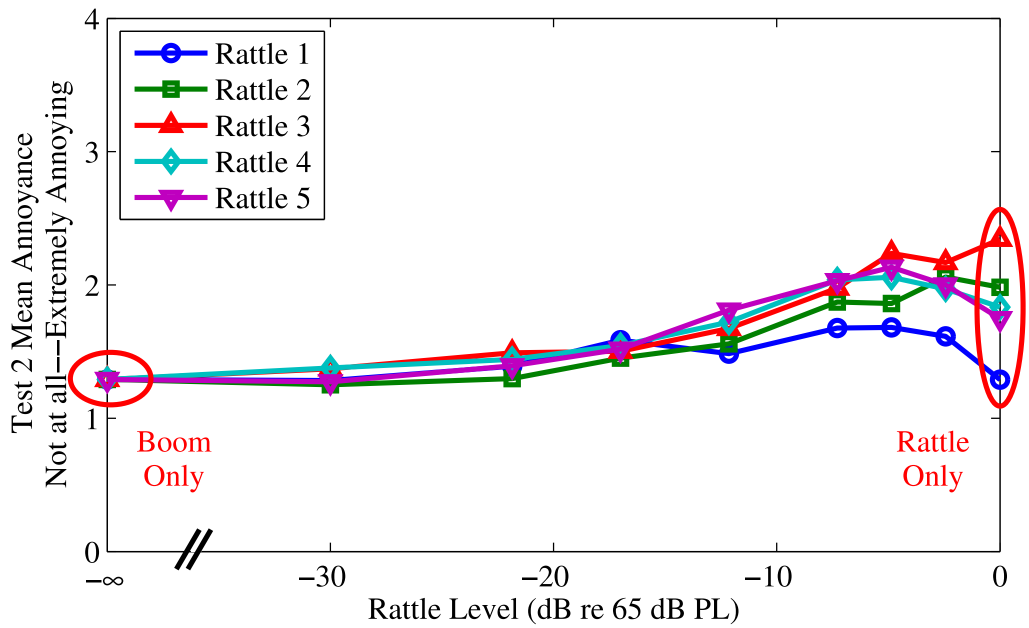

Figure B1. Test 2 mean annoyance for Boom 1 (recorded boom filtered to simulate reception in a large room with light transmission loss) and five rattles as a function of rattle level relative to the isolated rattle level.

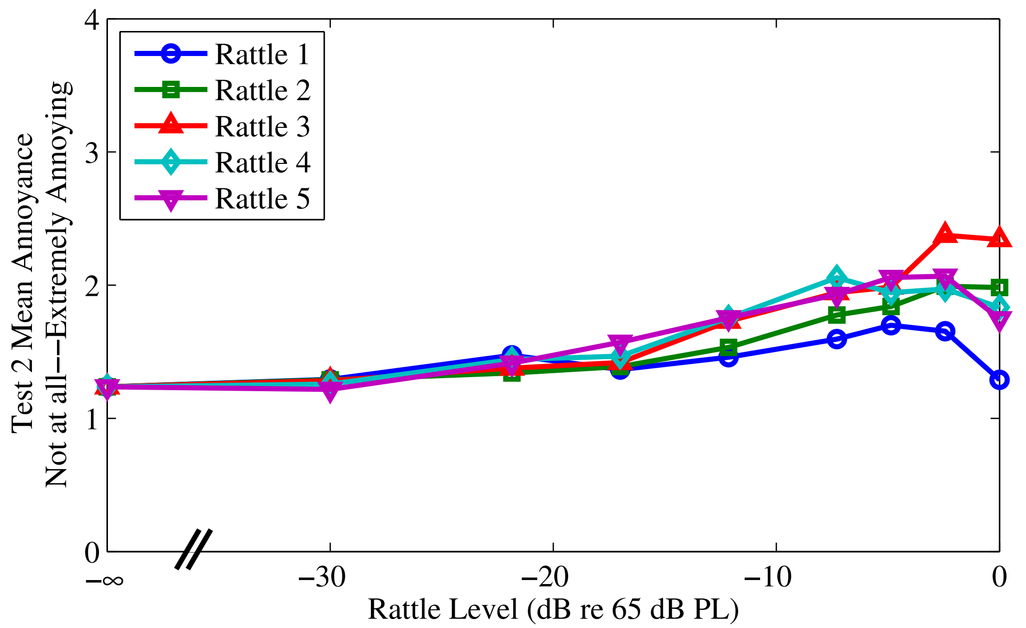

Figure B2. Test 2 mean annoyance for Boom 2 (recorded boom filtered to simulate reception in a large room with moderate transmission loss) and five rattles as a function of rattle level relative to the isolated rattle level.

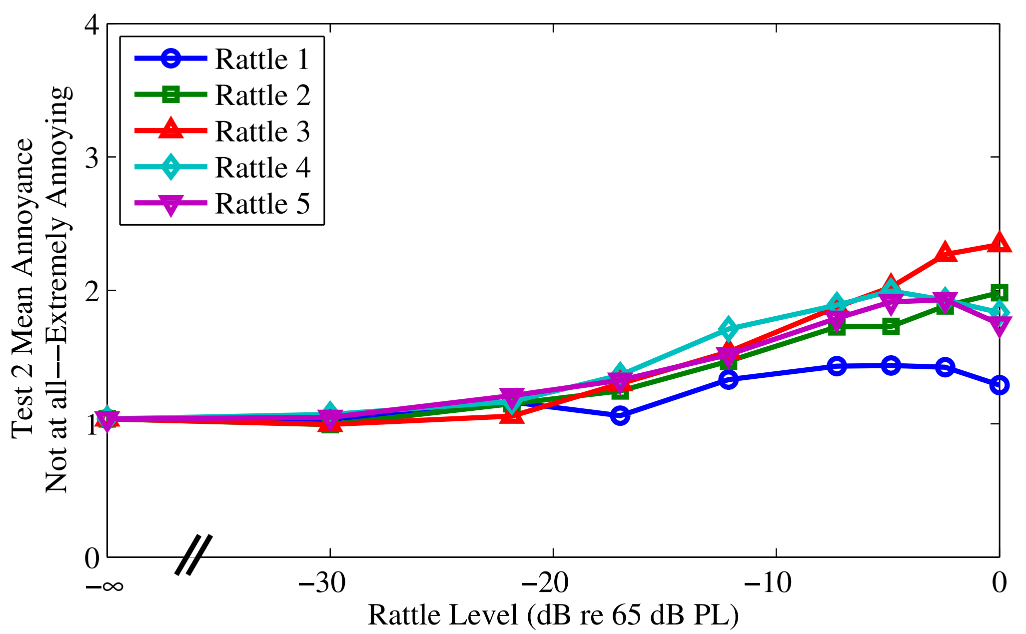

Figure B3. Test 2 mean annoyance for Boom 3 (recorded boom filtered to simulate reception in a small room with moderate transmission loss) and five rattles as a function of rattle level relative to the isolated rattle level.

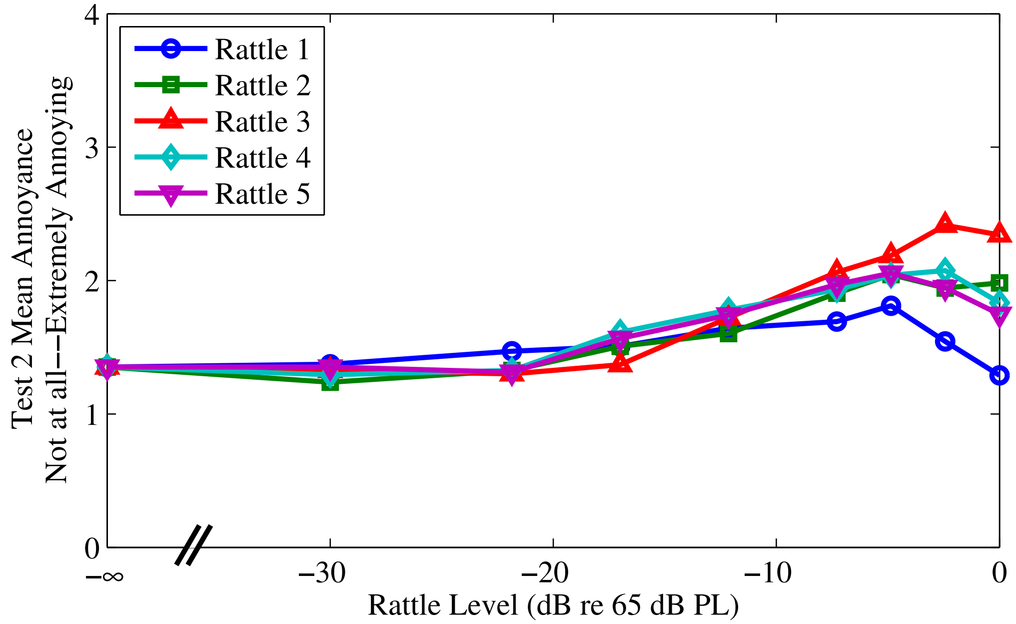

Figure B4. Test 2 mean annoyance for Boom 4 (synthesized ramp boom filtered to simulate reception in a small room with moderate transmission loss) and five rattles as a function of rattle level relative to the isolated rattle level.

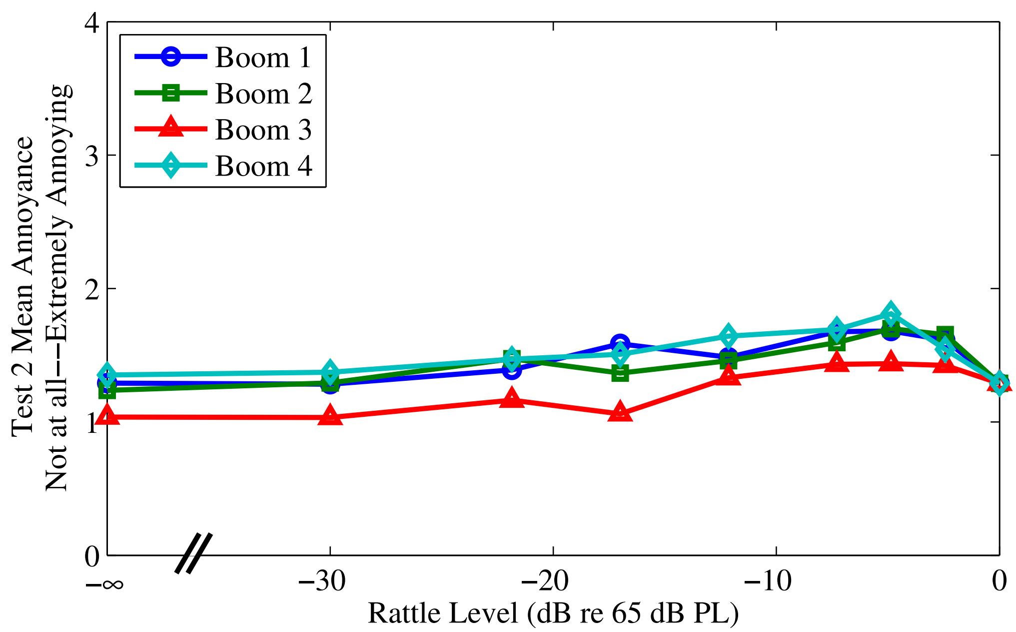

Figure B5. Test 2 mean annoyance for Rattle 1 and four booms as a function of rattle level relative to the isolated rattle level.

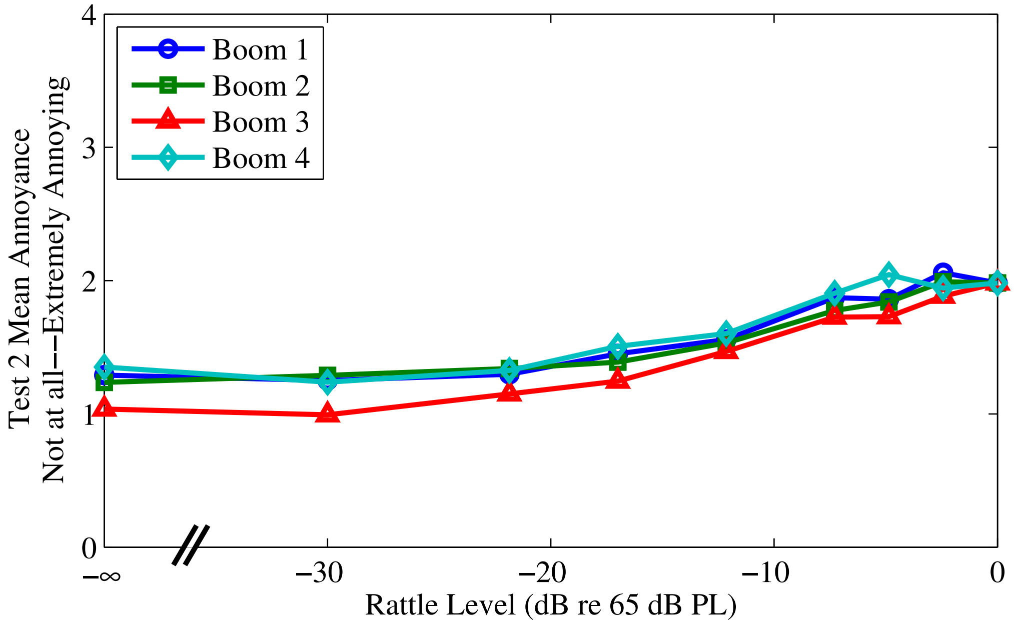

Figure B6. Test 2 mean annoyance for Rattle 2 and four booms as a function of rattle level relative to the isolated rattle level.

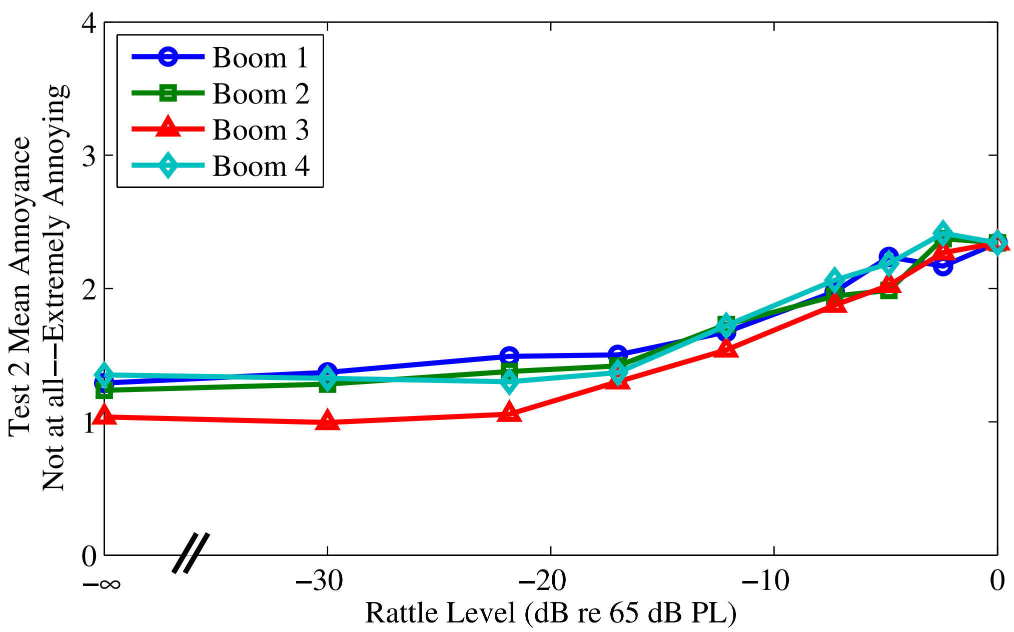

Figure B7. Test 2 mean annoyance for Rattle 3 and four booms as a function of rattle level relative to the isolated rattle level.

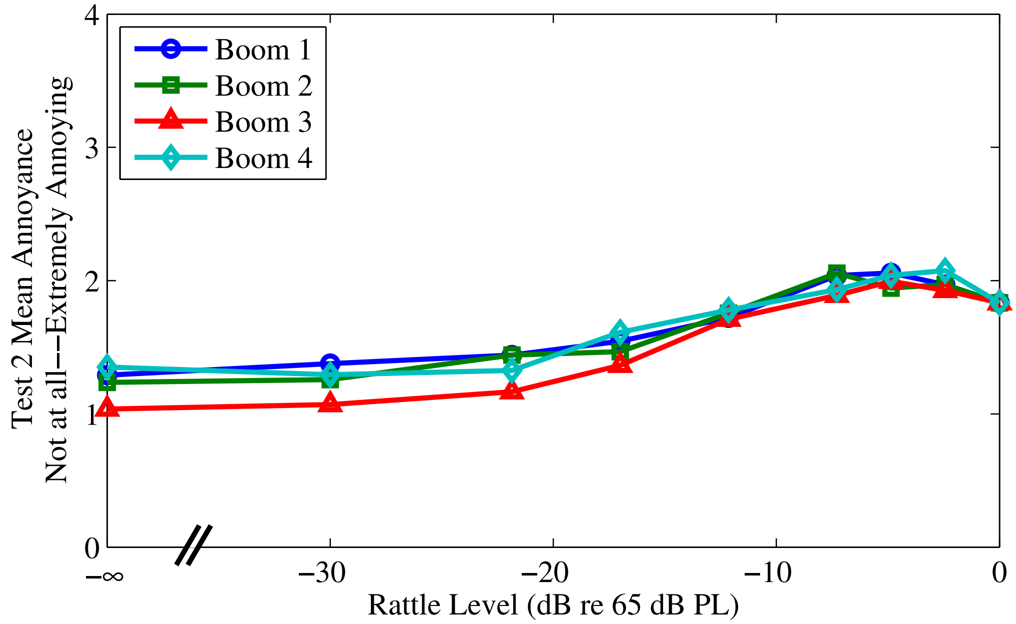

Figure B8. Test 2 mean annoyance for Rattle 4 and four booms as a function of rattle level relative to the isolated rattle level.

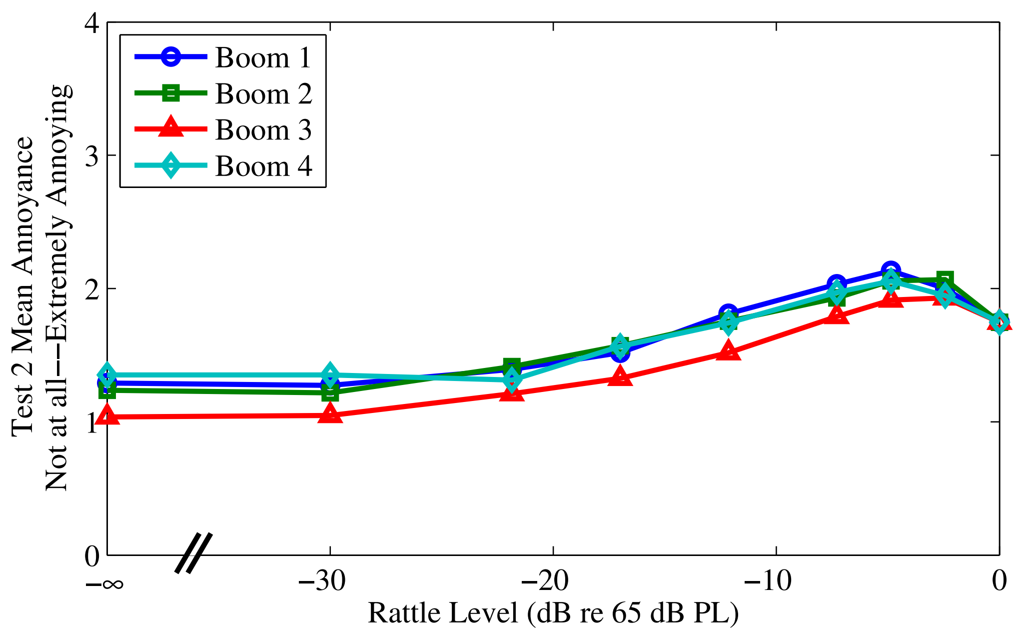

Figure B9. Test 2 mean annoyance for Rattle 5 and four booms as a function of rattle level relative to the isolated rattle level.

Appendix C

Test 3 Annoyance Figures

This appendix includes figures of the Test 3 mean annoyance as a function of rattle level. Figures C1-C6 each present the mean annoyance for a single boom and rattle at seven rattle levels and at three PL values. The corresponding results from Test 2 at a PL of 65 dB are included as a dashed line for reference.

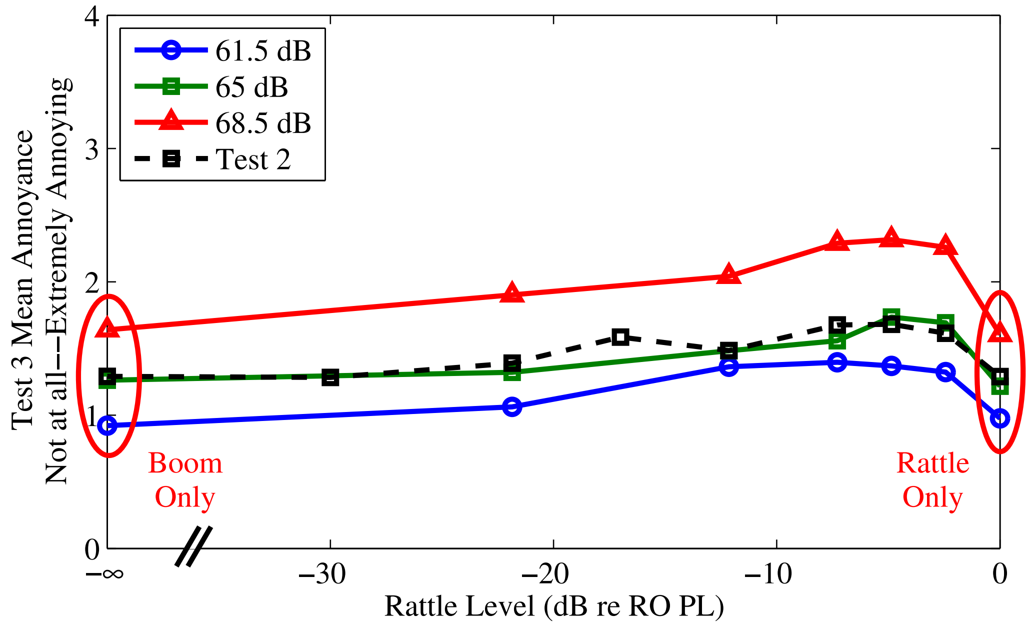

Figure C1. Test 3 mean annoyance for Boom 1 and Rattle 1 as a function of rattle level relative to the isolated rattle only (RO) level.

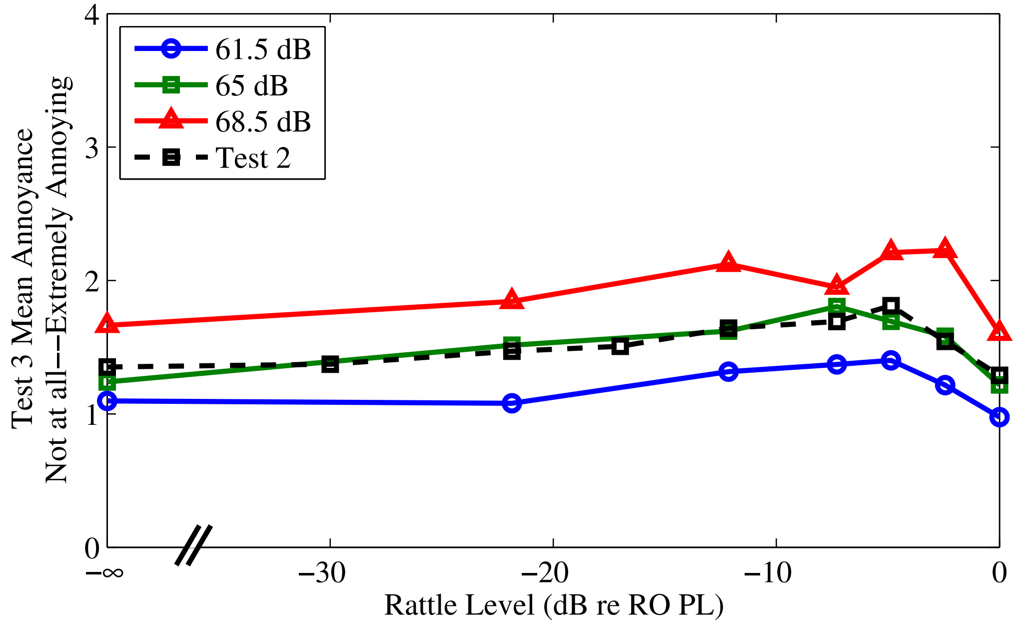

Figure C2. Test 3 mean annoyance for Boom 4 and Rattle 1 as a function of rattle level relative to the isolated rattle only (RO) level.

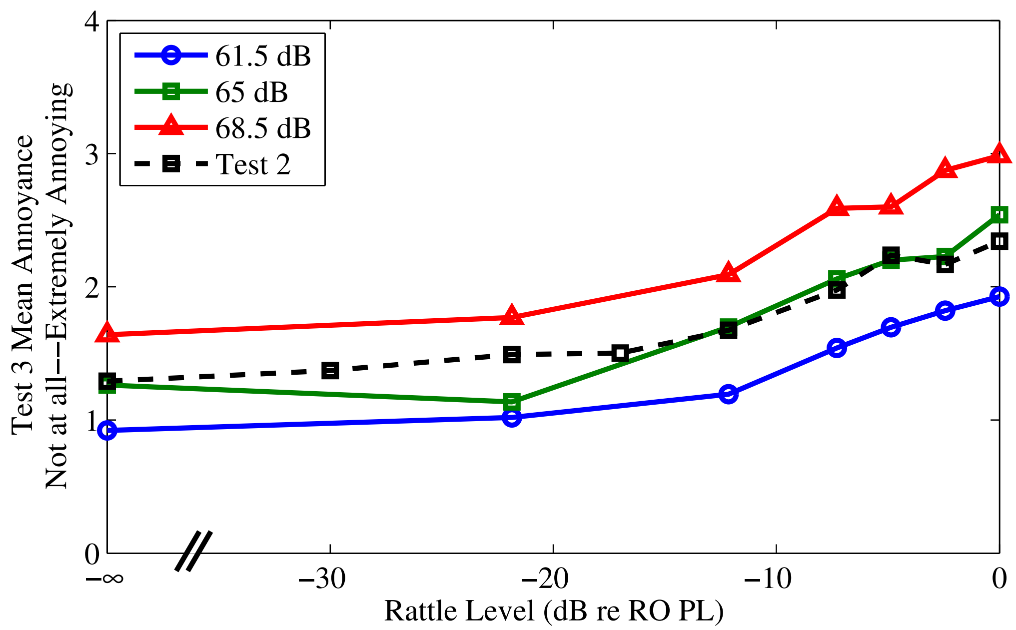

Figure C3. Test 3 mean annoyance for Boom 1 and Rattle 3 as a function of rattle level relative to the isolated rattle only (RO) level.

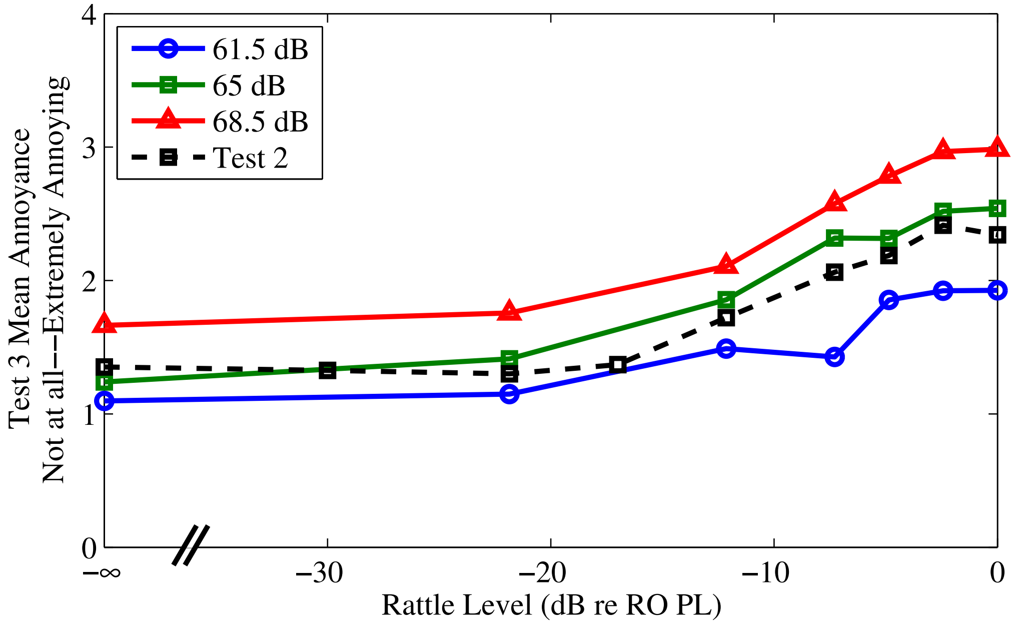

Figure C4. Test 3 mean annoyance for Boom 4 and Rattle 3 as a function of rattle level relative to the isolated rattle only (RO) level.

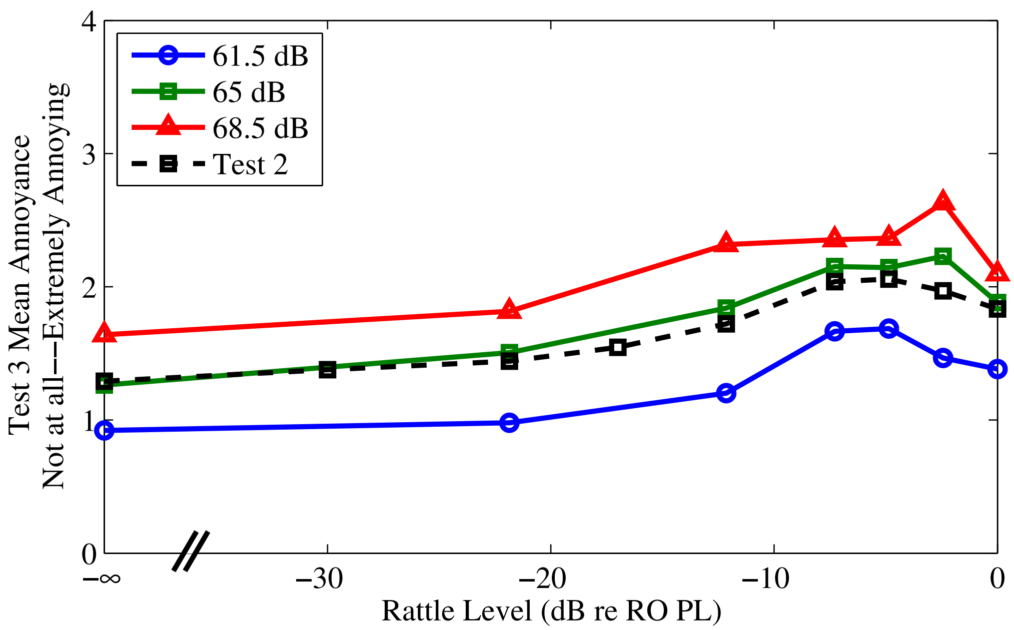

Figure C5. Test 3 mean annoyance for Boom 1 and Rattle 4 as a function of rattle level relative to the isolated rattle only (RO) level.

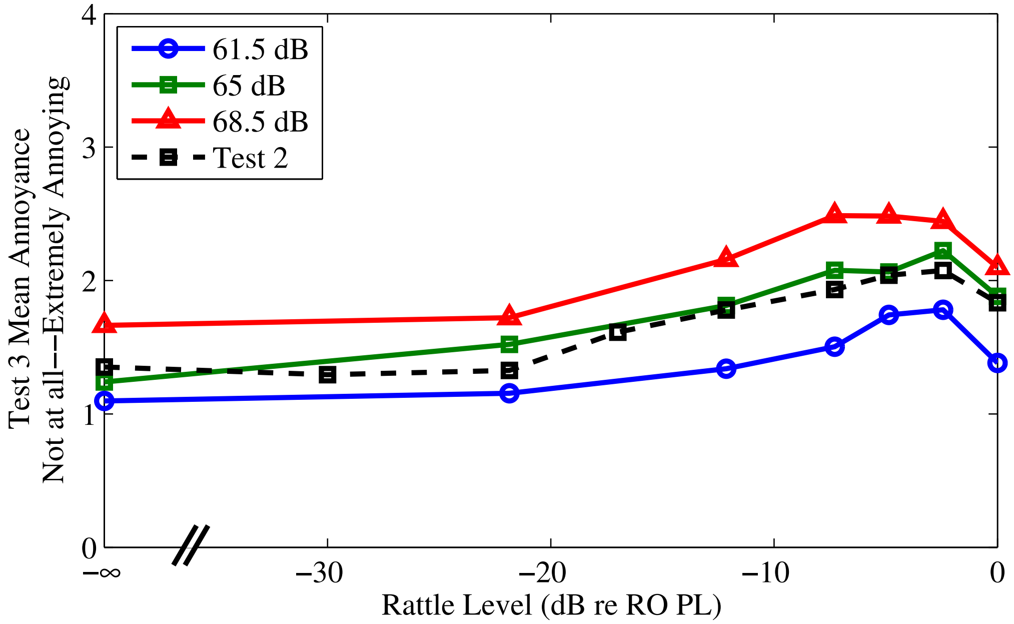

Figure C6. Test 3 mean annoyance for Boom 4 and Rattle 4 as a function of rattle level relative to the isolated rattle only (RO) level.

Appendix D

List of Objective Metrics

The loudness and sound quality metrics selected for analysis are described in Tables D1 and D2 below. The procedure for calculation of one-third octave spectra used for PL, PNL, LLZ, and MGSL is given in Ref. [51].

Table D1: Names and descriptions of loudness metrics

Metric Name (symbol)	Description
Sound Exposure Level: A-weighted (ASEL) C-weighted (CSEL) Unweighted (ZSEL)	Sound Exposure Level is the energy-averaged sound level over a specified length of time, with a reference duration of 1 s [2,4] and allows for the application of different weighting functions . In the expression below, the integral is performed over the period T of the squared pressure signal $p^2(t)$, the reference time $t_0 = 1\,\mathrm{s}$, and the reference pressure $p_0 = 20\,\mu\mathrm{Pa}$. $$\mathrm{SEL} = 10\log_{10}\left\{\frac{\int_0^T p^2(t)\,\mathrm{d}t}{p_0^2 t_0}\right\}$$ The implementation for an A-weighted pressure spectrum is given as $$\mathrm{ASEL} = 10\log_{10}\left\{\frac{T\sum_1^N p_{An}^2}{p_0^2 t_0}\right\},$$ where N is the number of frequency samples in the spectrum, p_{An} is the A-weighted spectral level at the nth frequency, and T is the period in seconds. Alternatively, the C-weighted or unweighted spectral levels can be used to calculate CSEL or ZSEL, respectively.

Continued on Next Page...

Table D1 – Continued

Metric Name (symbol)	Description
Perceived Level (PL)	The PL metric used in the present study is an updated version of the previous Stevens Loudness Level, Mark VI [54], which is standardized in Method A of ISO 532-1975 [25]. The signal spectrum is filtered into one-third octave bands, and each band is converted by a rule to a perceived value in sones. A summation procedure is used to determine the total loudness in sones, which is then converted to PL in dB. $$S_t = S_m + F\left(\sum S - S_m\right)$$ In the expression, S_t is the total loudness, S_m is the greatest loudness across the bands, $\sum S$ is the sum of the loudness of all bands, and F is a fractional factor (set to 0.15 in Mark VI[D1]) that determines the contributions of weaker bands to the total loudness. The Perceived Level used in this study follows the updated Stevens Loudness Level Mark VII calculation [28, 55]. The frequency-weighting contours were updated in Mark VII to match an average of 25 experimental contours fitted with 5 line segments instead of the simpler 3 segments used in Mark VI. In Mark VII the contours are also extended down to 1 Hz for use with sonic booms. The loudness summation procedure remains the same, although the value of F is no longer fixed and is determined by the loudness of the loudest band.
Perceived Noise Level (PNL)	PNL was developed to provide a rating of the noisiness of a sound. The PNL of a sound is the sound pressure level in dB of an octave band of noise centered at 1 kHz that is judged to be as noisy as the sound. As with PL, the signal spectrum is first filtered into one-third octave bands. The contours of perceived noisiness, the "noy" curves, are used instead of the equal loudness index, or "sone" curves [4, 33]. The noy curves were developed based on subjective noisiness and annoyance rather than subjective loudness. The summation procedure for PNL is identical to that for PL.

Continued on Next Page...

[D1]The ISO 532-1975 standard [25] incorrectly lists F as 1.15 for one-third octave bands.

Table D1 – Continued

Metric Name (symbol)	Description
Zwicker Loudness Level: Frontal incidence (LLZ_f) Diffuse incidence (LLZ_d)	Zwicker Loudness Level, standardized in Method B of ISO 532-1975 [25], is also calculated from the signal spectrum filtered into one-third octave bands. If the level of a sound in one frequency band significantly exceeds the level of the sound in the next highest frequency band, the loudness level in the latter band is increased according to predefined graphical curves. The shape of these curves depends on the sound level, frequency band, and whether the sound field is free or diffuse. The DIN 45631 [12,13], HEAD [23], and ISO532B [25] loudness methods used in the present study are all based on the Zwicker Loudness Level. The differences between them are enumerated in Ref. [23].
Moore and Glasberg Stationary Loudness (MGSL)	MGSL, standardized in ANSI S3.4-2007 [3,38], is based on the signal spectrum, which can be specified in one-third octave bands from 50 to 16,000 Hz. As used in this study, the stages of this loudness model for steady sounds are: 1. a filter corresponding to transfer through the outer ear 2. a filter corresponding to transfer through the middle ear 3. excitation pattern calculation from the physical spectrum 4. transformation of excitation pattern to specific loudness pattern 5. determination of overall loudness from specific loudness A comparison of loudness calculations using the MGSL and DIN 45631 methods shows that MGSL gives systematically higher loudness values for broadband signals [23]; specifically, a difference by a factor of 1.27-2.31 is found for pink noise of different levels [14]. The "ISO532A" loudness method used in the present study is an MGSL procedure [23] and is based on the method defined in the draft standard ISO/DIS 532-1 [27].
Moore and Glasberg Time-Varying Loudness (MGTVL)	In contrast to most of the above metrics, the MGTVL model uses the signal waveform input to calculate loudness level variations with time. It is similar to MGSL except that the excitation pattern in step 3 is calculated from a short-term FFT [18]. The resulting "instantaneous" loudness is calculated and then converted to a short-term loudness using an averaging technique.

Continued on Next Page…

Table D1 – Continued

Metric Name (symbol)	Description
Relative Approach	Relative Approach is based on the notion that for a single signal human hearing is more sensitive to differences in temporal structures or spectral patterns than differences in level. A reference value is formulated as an average of the signal in the time and frequency domains. Relative approach quantifies the degree of deviations from this average [7, 17, 52, 53].

Table D2: Names and descriptions of sound quality metrics

Metric Name	Description
Roughness	Roughness results from temporal fluctuations in the signal spectrum for which modulation frequency is between 20 and 300 Hz. In this range, the subjective impression is one of roughness. The roughness unit, the asper, is referenced to the roughness impression of a 1 kHz sine tone with a level of 60 dB, amplitude-modulated at a rate of 70 Hz with a modulation depth of 1 [24, 57, 60].
Hearing Model Roughness	Roughness alone has been found to over-predict the subjective response to unmodulated noise. In response, a roughness calculation procedure was developed based on Sottek's Hearing Model [24, 52]. This so-called "Hearing Model Roughness" has been shown to outperform "Roughness" in predicting subject response to real-world sounds [24].
Hearing Model Impulsiveness	Impulsiveness describes repeated short-duration increases in amplitude. The peak repetition frequency for impulsiveness is 10 Hz. The Hearing Model Impulsiveness metric is also based on Sottek's Hearing Model [24, 52].
Kurtosis	Kurtosis is a statistical term that quantifies the "peakedness" of a distribution [46]. Kurtosis is used in this study to quantify the peakedness of the signal's time history.
Tonality	Tonality quantifies the degree to which the signal is comprised of tonal components versus broadband noise. The contribution of individual tones to the overall tonality depends on the frequency range; specifically, a 700 Hz tone will result in a maximum tonality impression. The value of 1 tu (tonality unit) is defined for a 1 kHz sine tone at a level of 60 dB [24, 58].

REPORT DOCUMENTATION PAGE							*Form Approved* *OMB No. 0704-0188*
The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**							
1. REPORT DATE *(DD-MM-YYYY)* 01- 03 - 2013		**2. REPORT TYPE** Technical Memorandum					**3. DATES COVERED** *(From - To)*
4. TITLE AND SUBTITLE Laboratory Headphone Studies of Human Response to Low-Amplitude Sonic Booms and Rattle Heard Indoors							**5a. CONTRACT NUMBER**
							5b. GRANT NUMBER
							5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) Loubeau, Alexandra; Sullivan, Brenda M.; Klos, Jacob; Rathsam, Jonathan; Gavin, Joseph R.							**5d. PROJECT NUMBER**
							5e. TASK NUMBER
							5f. WORK UNIT NUMBER 984754.02.07.07.18.03
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) NASA Langley Research Center Hampton, VA 23681-2199							**8. PERFORMING ORGANIZATION REPORT NUMBER** L-20217
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) National Aeronautics and Space Administration Washington, DC 20546-0001							**10. SPONSOR/MONITOR'S ACRONYM(S)** NASA
							11. SPONSOR/MONITOR'S REPORT NUMBER(S) NASA/TM-2013-217975
12. DISTRIBUTION/AVAILABILITY STATEMENT Unclassified - Unlimited Subject Category 71 Availability: NASA CASI (443) 757-5802							
13. SUPPLEMENTARY NOTES							
14. ABSTRACT Human response to sonic booms heard indoors is affected by the generation of contact-induced rattle noise. The annoyance caused by sonic boom-induced rattle noise was studied in a series of psychoacoustics tests. Stimuli were divided into three categories and presented in three different studies: isolated rattles at the same calculated Perceived Level (PL), sonic booms combined with rattles with the mixed sound at a single PL, and sonic booms combined with rattles with the mixed sound at three different PL. Subjects listened to sounds over headphones and were asked to report their annoyance. Annoyance to different rattles was shown to vary significantly according to rattle object size. In addition, the combination of low-amplitude sonic booms and rattles can be more annoying than the sonic boom alone. Correlations and regression analyses for the combined sonic boom and rattle sounds identified the Moore and Glasberg Stationary Loudness (MGSL) metric as a primary predictor of annoyance for the tested sounds. Multiple linear regression models were developed to describe annoyance to the tested sounds, and simplifications for applicability to a wider range of sounds are presented.							
15. SUBJECT TERMS Annoyance; Headphone; Human response; Indoor boom; Loudness; Perceived level; Psychoacoustics; Rattle; Sonic boom;							
16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** STI Help Desk (email: help@sti.nasa.gov)		
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER** *(Include area code)*		
U	U	U	UU	74	(443) 757-5802		

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18