Summer 2014

Assessment of noise metrics for application to rotorcraft

Andrew L. McMullen
Purdue University

Graduate School ETD Form 9
(Revised 01/14)

PURDUE UNIVERSITY
GRADUATE SCHOOL
Thesis/Dissertation Acceptance

This is to certify that the thesis/dissertation prepared

By Andrew McMullen

Entitled Assessment of Noise Metrics for Application to Rotorcraft

For the degree of Master of Science in Mechanical Engineering

Is approved by the final examining committee:

Prof. Patricia Davies

Prof. J. Stuart Bolton

Prof. Hong Z. Tan

Prof. George T. Chiu

To the best of my knowledge and as understood by the student in the *Thesis/Dissertation Agreement. Publication Delay, and Certification/Disclaimer (Graduate School Form 32)*, this thesis/dissertation adheres to the provisions of Purdue University's "Policy on Integrity in Research" and the use of copyrighted material.

Approved by Major Professor(s): Prof. Patricia Davies

Approved by: Prof. David C. Anderson 07/29/2014

Head of the Department Graduate Program Date

ASSESSMENT OF NOISE METRICS

FOR APPLICATION TO ROTORCRAFT

A Thesis

Submitted to the Faculty

of

Purdue University

by

Andrew L. McMullen

In Partial Fulfillment of the

Requirements for the Degree

of

Master of Science in Mechanical Engineering

August 2014

Purdue University

West Lafayette, Indiana

ACKNOWLEDGMENTS

I would like to thank NASA for their financial support of this research through the Subsonic Rotary Wing project of the NASA Fundamental Aeronautics Program. I would like to thank Dr. Stephen A. Rizzi (contract monitor, NASA Langley Research Center) for his ongoing support and guidance with this work. I would also like to thank Aric Aumann (AS&M, Inc), Dr. Selen Okcu (NIA), and Andrew Christian (NIA) for their assistance throughout this project. Additional assistance was provided by Dr. Jeff Mendoza, Dr. Dan Shannon, and Dr. Arthur Blanc from United Technologies Research Center regarding signal modification and test design review.

Most importantly, I would like to thank my advisor, Dr. Davies, for her constant support and guidance. Her direction and encouragement allowed me to truly achieve a deeper understanding of this work and its purpose.

I would like to thank my parents and family for their ongoing support of my work and chosen career path. Last but not least I would like to thank my fellow graduate students, especially Nicholas Sakamoto and Brandon Sobecki, for their help and advice along the way.

TABLE OF CONTENTS

Page

Page

LIST OF TABLES

Table Page

Table | Page

Table Page

LIST OF FIGURES

Figure Page

Figure Page

Figure Page

Figure Page

Figure Page

Figure Page

ABSTRACT

McMullen, Andrew L. MSME, Purdue University, August 2014. Assessment of Noise Metrics For Application to Rotorcraft. Major Professor: Patricia Davies, School of Mechanical Engineering.

It is anticipated that the use of rotorcraft passenger vehicles for shorter journeys will increase because their use can reduce the time between boarding and take-off. The characteristics of rotorcraft noise are very different to that of fixed wing aircraft. There can be strong tonal components, fluctuations that can also make the noise sound impulsive, and future rotorcraft may produce proportionally more low frequency noise content. Most metrics that are used today to predict noise impact on communities around airports (e.g., L_{dn}) are just functions of A-weighted sound pressure level. To build a better noise annoyance model that can be applied to assess impact of future and current rotorcraft, it is important to understand the perceived sound attributes and how they influence annoyance. A series of psychoacoustic tests were designed and performed to further our understanding of how rotorcraft sound characteristics affect annoyance as well as evaluate the applicability of existing noise metrics as predictors of annoyance due to rotorcraft noise. The effect of the method used to reproduce sounds in the psychoacoustics tests was also investigated, and so tests were conducted in the NASA Langley Exterior Effects Room using loudspeaker arrays to simulate flyovers and in a double walled sound booth using earphones for playback. A semantic differential test was performed, and analysis of subject responses showed the presence of several independent perceptual factors relating to: loudness, sharpness, roughness, tonality, and impulsiveness. A simulation method was developed to alter tonal components in existing rotorcraft flyover recordings to change the impulsiveness and tonality of the sounds. Flyover recordings and simulations with varied attributes were used as stimuli in an annoyance test. Results showed that EPNL and SELA performed well as predictors of annoyance, but outliers to generate trends have tonal related characteristics that could be contributing to annoyance. General trends in results were similar for both test environments, though differences were greater for the annoyance tests than the semantic differential tests.

1. INTRODUCTION

Environmental noise has become part of our daily lives, with transportation being a key source. Protective noise levels were investigated by the U.S. Environmental Protection Agency, where acceptable noise levels for a variety of conditions were defined [1]. Additionally, activity interference and annoyance were studied through social surveys, and it was found that up to 20.6% of people were highly disturbed by aircraft noise for certain activities. However, noise levels are not always sufficient for predicting annoyance. Many other factors are present in human reaction to noise, such as attitude toward the noise source, presence of tones or impulses, and duration of the noise [1]. For example, aircraft, rail, and road-traffic noise can cause different reaction, even when levels are the same [2]. ISO 1996-1:2003, used to assess environmental noise, accounts for these differences with a 3 dB penalty for aircraft noise and a 6 dB bonus for train noise [3]. These effects on human perception are referred to as "aircraft malus" and "railway bonus" [2]. However, the "aircraft malus" may be caused by other, non-acoustic, opinions of the source such as crash possibilities and loss of privacy [4].

Noise can also have various health effects, including hearing damage. Previous studies have shown a relationship between aircraft noise exposure and the risk for respiratory, digestive, mental instability, depression, and nervousness [5]. The various effects and number of people affected provide the motivation for ongoing research regarding aircraft.

1.1 Background and Motivation for Research

It has been shown that the implementation of vertical takeoff and landing (VTOL) aircraft as a replacement for conventional aircraft on shorter flights can help reduce airport congestion and flight delays [6]. Currently, 26% of commercial operations from the 64 major airports consist of a length less than 500 miles, which would allow for the use of aircraft not requiring runways. Current concept VTOL aircraft for completing these short flight operations include large civil tilt rotors (LCTRs). These vehicles use proprotors mounted on rotating engine pods, which allows for the vehicle to takeoff vertically. Then the engine pods rotate while in-flight so that the rotors can act as propellers. However, the introduction of LCTRs and/or other rotorcraft would have many impacts on the airport community, with the most significant being noise. It is important to assess how the noise generated by these vehicles would effect the community. In order to predict this effect, the noise source(s) and propagation path must be characterized, as well as the perception of the sound as heard by those in the community. The corresponding impact of the sound, e.g., annoyance, sleep disturbance, cognitive impairment, and long term health effects. In this research perception of rotorcraft noise and annoyance are studied.

1.2 Problem Statement

Rotorcraft noise can have attributes that are significantly different to those of conventional fixed-wing aircraft noise. The noise can contain strong tonal components, strong fluctuations, and the spectral balance may be different to that of conventional aircraft. Many currently used aircraft certification noise metrics may not accurately account for the effects of these noise characteristics. In order to evaluate the applicability of these metrics and/or develop new evaluation methods, the various attributes of rotorcraft must be identified and analyzed and the relationship between the attributes and overall judgments of the sounds identified and quantified. Three components are necessary: (1) source characterization (trajectory, directivity, spectral content); (2) a model of propagation from source to receiver; and (3) a model of perception of the noise and judgments, such as annoyance. Multiple flight tests have been conducted by NASA in order to collect acoustic data and characterize various rotorcraft noise sources [7–9]. These tests covered six vehicles performing various operations.

A large concern with the implementation of rotorcraft is that currently used aircraft noise metrics may not accurately account for the impact of rotorcraft noise. Many environmental noise metrics employ A-weighting, which may inappropriately attenuate high level, low frequency components. Effective Perceived Noise Level ($EPNL$), the current aircraft certification noise metric used by the Federal Aviation Administration, does not use information below the 50 Hz one-third octave band. The $EPNL$ calculation includes a tone correction, however the correction may not adequately account for strong low frequency tones. Tonal metric calculations can be complex and have issues with non-stationary sounds, such as when frequencies undergo a Doppler shift during a flyover. If a person can track the pitch of a sound from the tonal content, the tonalness of the sound will be almost as annoying as when the sound is stationary. However, the calculated tonal prominence (a component of all tonal metrics) may be less when the sound is nonstationary [10] due to spectral estimation limitations.

The fluctuations in rotorcraft noise may also add to annoyance. The low fundamental frequency generated by the main rotor of a helicopter will not sound tonal (contain discernable pitches), but the harmonic structure can generate rapid variations in loudness that are trackable, which adds an impulsive character to the sound. This character is commonly present in noise generated by rotary wing aircraft [11], as well as machines operating with diesel engines [12,13], such as motorcycles, trucks, drills, etc. These varying sounds can be harder to acclimatize to because of the impulsive characteristic. Fluctuation strength calculations exist but can be difficult to calculate [2]. Sutherland and Burke (1979) suggest that up to a 6 dB penalty be applied to sounds with the presence of blade slap when assessing annoyance.

While many previous studies have shown correlation between the perceived level of a sound and the corresponding annoyance, other factors will have an effect. Context is a key component when attempting to predict annoyance. For instance, the presence of tones in music is considered pleasing, while tones in environmental or machinery noise is found annoying. Unbiased annoyance was developed through research at the Technical University of Munich, where Zwicker began work on predicting annoyance purely based on sound characteristics. From this work, the Psychoacoustic Annoyance model was developed by Fastl and colleagues. This model includes a loudness term, as well as roughness and fluctuation strength. The roughness calculation is based on fast loudness fluctuations (>30 oscillations/second), while the fluctuation strength calculation is based on slow loudness fluctuations (<20 oscillations/second). Proposed models by Hastings [14] and More [15] add a tonal component to the annoyance calculation for application to diesel noise and aircraft noise, respectively. More found that the inclusion of tone metrics or tone penalties in annoyance models generally improved predictability.

The overall goal of this research was to determine what characteristics were present in human response to rotorcraft noise, and whether or not currently used metrics are sufficient for quantifying annoyance.

1.3 Objectives

The following are the objectives of this research:

1. To determine which attributes of rotorcraft noise affect annoyance
2. To quantify the strength of those attributes
3. To understand how to combine attribute strengths to predict annoyance levels
4. To assess the performance of noise metrics currently used in aircraft applications, such as EPNL and statistics of weighted sound pressure levels

1.4 Approach and Thesis Outline

Figure 1.1 contains a simplified schematic of the research approach. The approach involves first gathering a comprehensive set of rotorcraft recordings and doing three types of analysis on them, summarized in (i) to (iii) below:

i. Recordings → Signal Analysis → Physical Spectral and Temporal Sound Characteristics → Physical Sound Components (time-varying harmonic families, time varying filtered noise, transients, etc.)

ii. Recordings → Sound Metrics ($EPNL, SELA, PL, SELC$, Tonality, Sharpness, Loudness, etc.)

iii. Recordings → Playback & Evaluation of Attributes → Response Analysis → Independent Perceived Attributes (these may or may not be associated with sound quality metrics)

Second, we would like to know the relationships between the physical sound components, sound metrics and perceived attributes. For this we:

i. Need the ability to modify sound components and recombine them to produce sounds with a range of perceived attribute strengths. We need to be able to do this without creating sounds that would be perceived as being artificial. This sound manipulation capability is key to human response (annoyance) model development.

ii. Want to examine the effect of signal modifications on sound metrics. Questions: Do metrics track strength of perceived attributes? How should metrics be modified to do a better job of tracking perceived attributes?

Finally we would like to know the relationship between perceived attributes and annoyance, and develop an annoyance model that is a function of sound metrics.

i. Generate sets of sounds where the strengths of individual perceived attributes are, ideally, varied independently (this is not always straightforward and may not be possible, e.g., independent variation of loudness and tonality often leads to variations in roughness). These would be attributes that we hypothesize would affect annoyance. For example, for rotorcraft these might be loudness, tonalness and impulsiveness.

ii. Design subjective tests where people evaluate how annoying these sounds are. Analyze results and test the annoyance model hypothesis. Compare annoyance model performance with performance of currently used aircraft noise metrics. Identify deficiencies in annoyance predictions.

While this is set out sequentially, the process is usually iterative with identified deficiencies in the annoyance models being used to guide further analysis. This may lead to additional signal analysis, additional metric development and improvements to the sound simulation and modification. This will enable development of new sets of sounds with independent variation of additional variables, which will be used in further annoyance model development and validation.

The research described in this thesis contains elements of all of the above, but is not as comprehensive as described above. What remains to be done is described in the last chapter of this thesis.

An initial collection of over 100 rotorcraft flyover recordings were gathered and analyzed to determine ranges of various sound attributes present. The details of the signals and the analysis are given in Chapter 2. The original signals were used in a semantic differential test to determine how many independent rotorcraft noise characteristics are perceived in this set of sounds; this is described in Chapter 3. A simulation method (described in Chapter 2) was then developed to characterize and modify the tonal components of selected recordings in order to generate signals with varied attributes. These modified signals, as well as original recordings, were used in an annoyance test (described in Chapter 4) so that the effect of the variations on annoyance could be observed. Noise metrics were evaluated for their performance in each test. Recommendations for the use of metrics and future work are presented in Chapter 5.

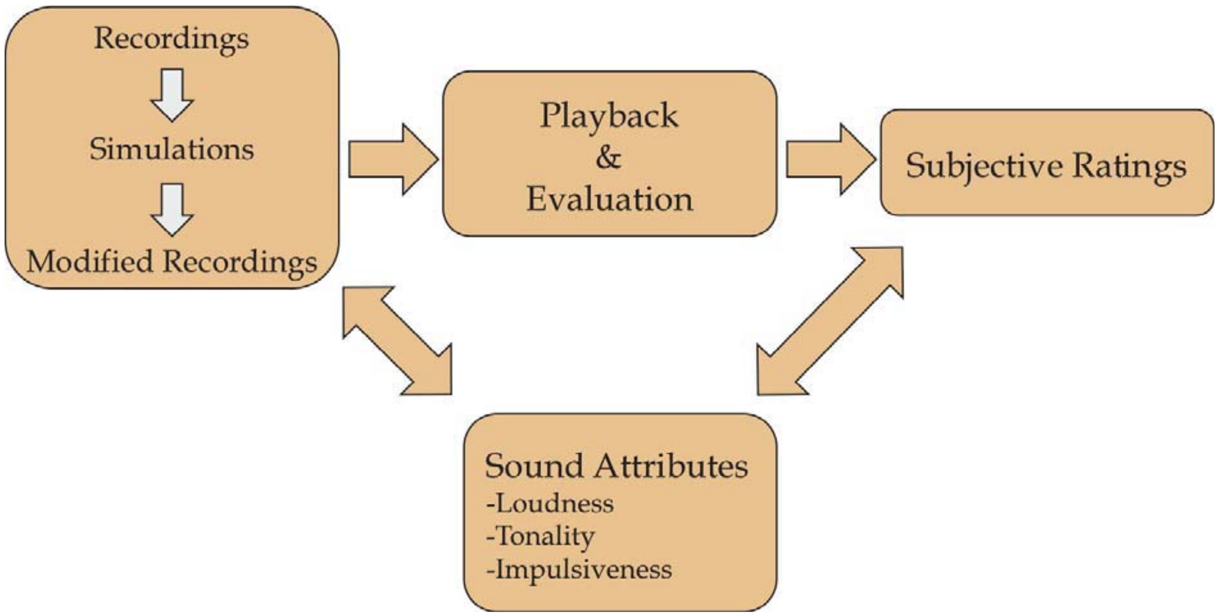

Figure 1.1. Research Approach.

2. BASELINE SOUNDS, ANALYSIS, METRICS AND MODIFICATIONS

A large set of recordings was gathered from several previously performed acoustic flight tests. These recordings were first analyzed, where level-based and psychoacoustic sound quality metrics were calculated. A subset of these signals was chosen to be used in subjective testing, and a simulation method was developed to modify these recordings in order to generate signals with varied attributes. The recordings, analysis, and modifications are described in the following sections.

2.1 Measurements from NASA Tests

Recordings were gathered from previous acoustic flight tests of various helicopters at Eglin Air Force Base and a test of the XV-15 Tiltrotor in Waxahachie, Texas [7–9]. The vehicles evaluated in these tests included the Bell 206, BO105, MD520N, MD902, Mi-8M, and XV-15. Table 2.1 contains a description of some of the features and characteristics of these vehicles.

Table 2.1. Blade passage frequencies (BPF) and maximum speeds of vehicles used to generate test stimuli.

Vehicle	Main Rotor BPF (Hz)	Tail Rotor/NOTAR* BPF (Hz)	Max speed (knots)
Bell 206	13	85	112
MBB BO105	28	74	131
MD520N	40	1167*	124
MD902	32.7	1100*	152
Mil Mi-8M	16	56	135
XV-15	32	N/A	220

The Bell 206, BO105, and Mi-8M operate as traditional helicopters, using a main rotor for lift and tail rotor to counter the torque from the main rotor. The MD520N and MD902 operate using the no tail rotor, or NOTAR system. This employs an exhaust fan in the tail boom to generate the necessary countering torque. While the traditional helicopter rotors generate blade passage frequencies at lower frequencies (<100 Hz), the NOTAR vehicles generate fan noise with a blade passage frequency above 1000 Hz. The remaining vehicle in this group, the XV-15, is an experimental tiltrotor aircraft, designed to show the advantages of vertical takeoff and landing (VTOL) aircraft over traditional helicopters. This vehicle uses rotating pods that contain both the engines and rotors, and does not use any type of tail rotor.

Figures 2.1 - 2.6 show spectrograms of the overhead sections of flyover recordings from each of the vehicles listed in Table 2.1. For those readers familiar with aircraft flyover sounds, it will be obvious that these recordings were taken on the ground because no ground reflection effects (cancellation and enhancement at different frequencies through time) can be seen in the spectrograms. A program was developed to simulate ground reflection effects by modifying the ground level recording, but after some consideration it was decided not to implement this modification to the sounds used in the subjective tests.

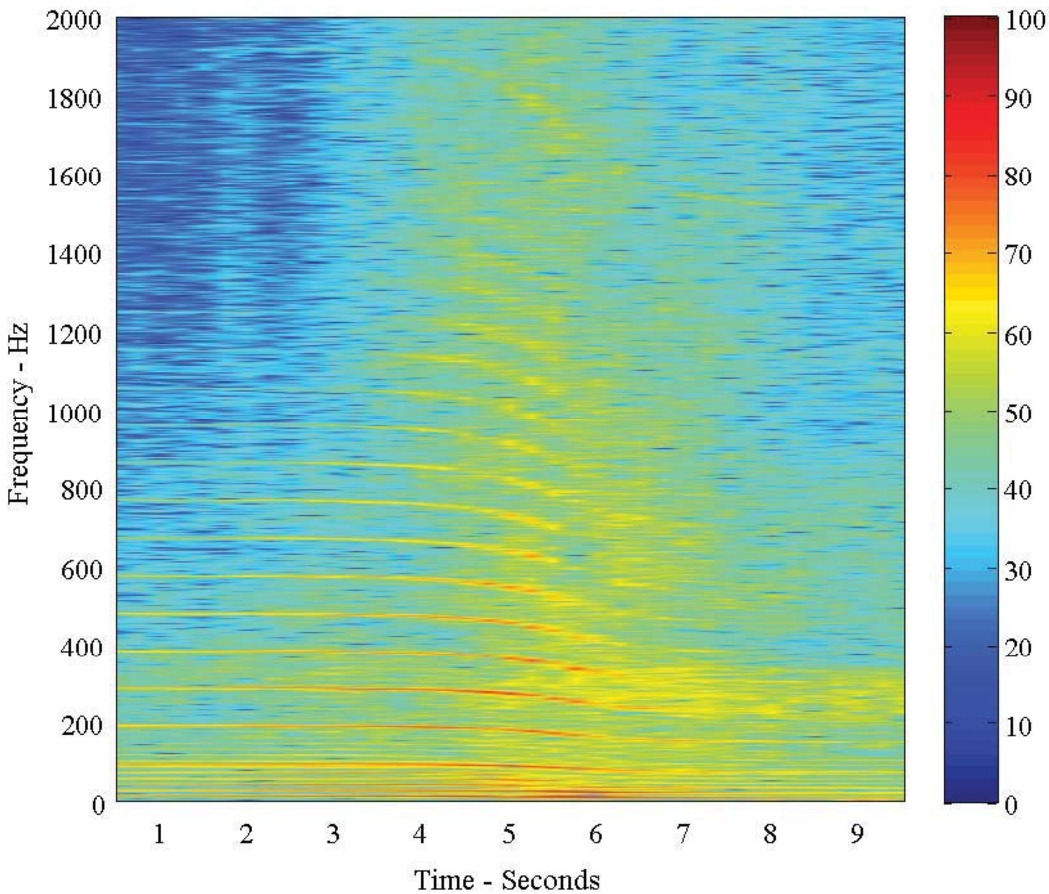

Figure 2.1. Spectrogram of the overhead section of a Bell 206 flyover, 80 knots, 150 ft. BPFs - 13 Hz and 85 Hz.

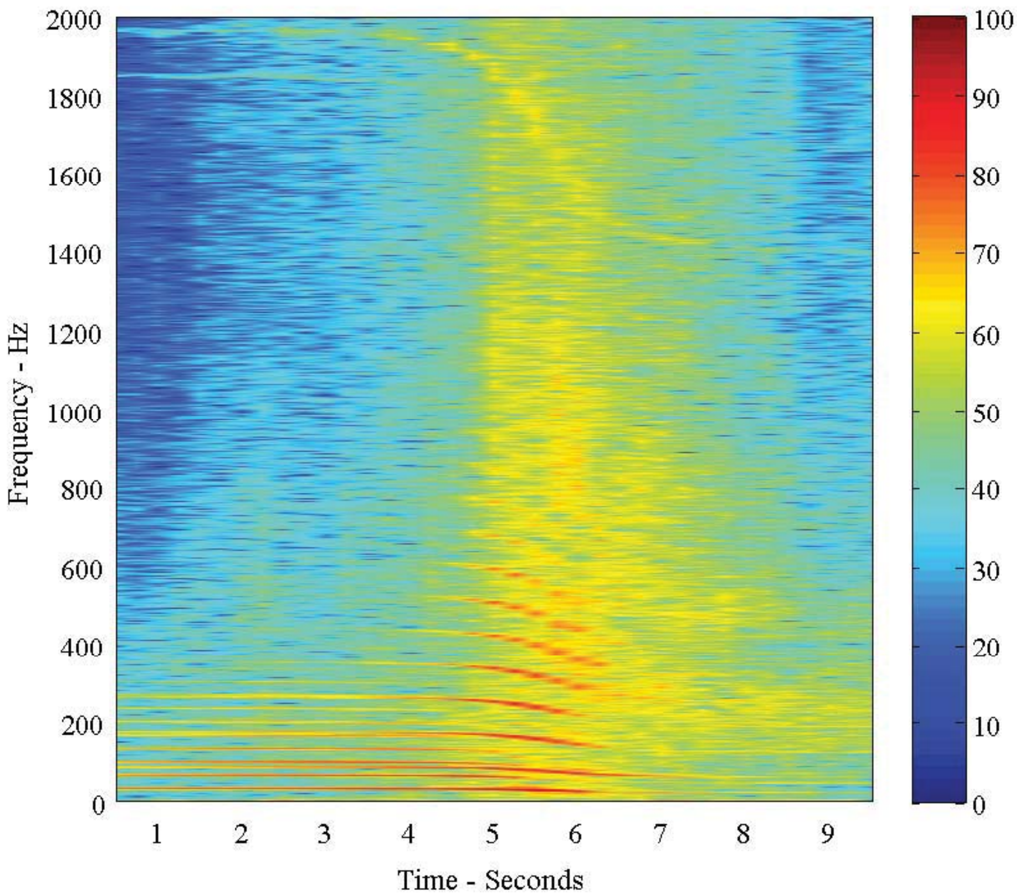

Figure 2.2. Spectrogram of the overhead section of a BO105 flyover, 115 knots, 150 ft. BPFs - 28 Hz and 74 Hz.

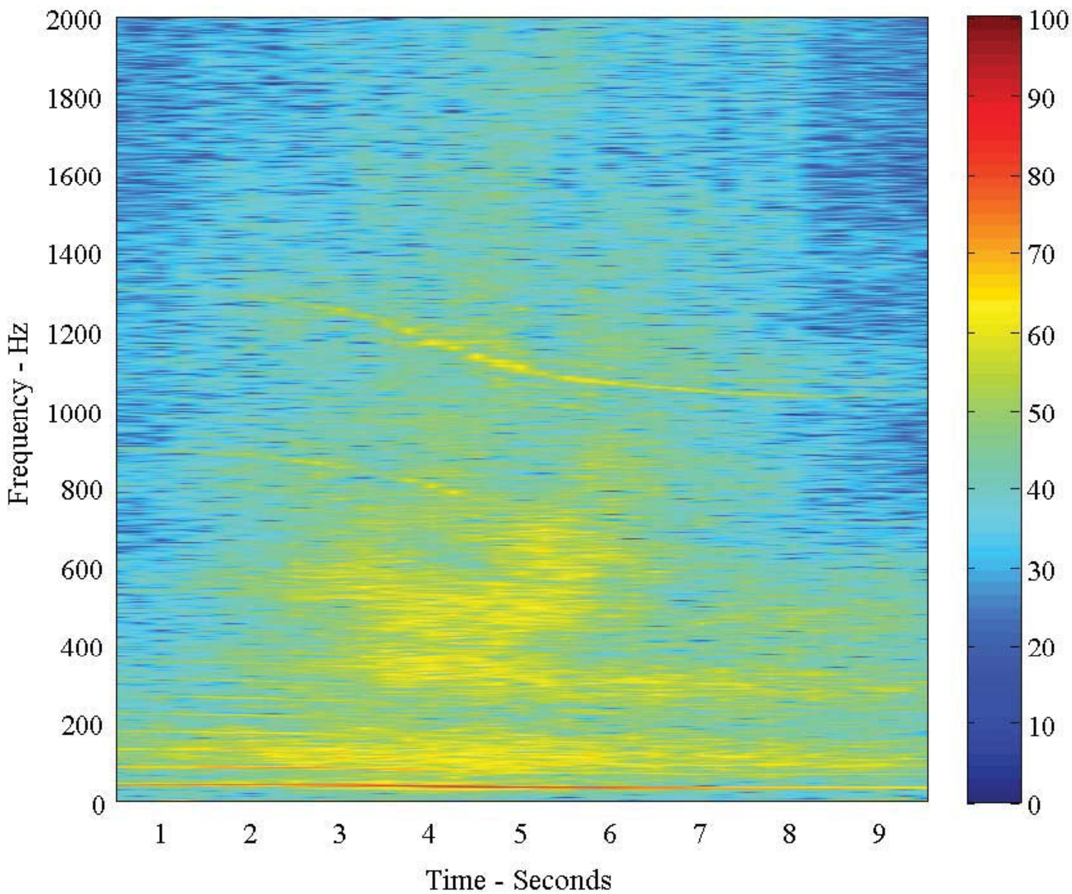

Figure 2.3. Spectrogram of the overhead section of a MD520N flyover, 80 knots, 250 ft. BPFs - 40 Hz and 1167 Hz.

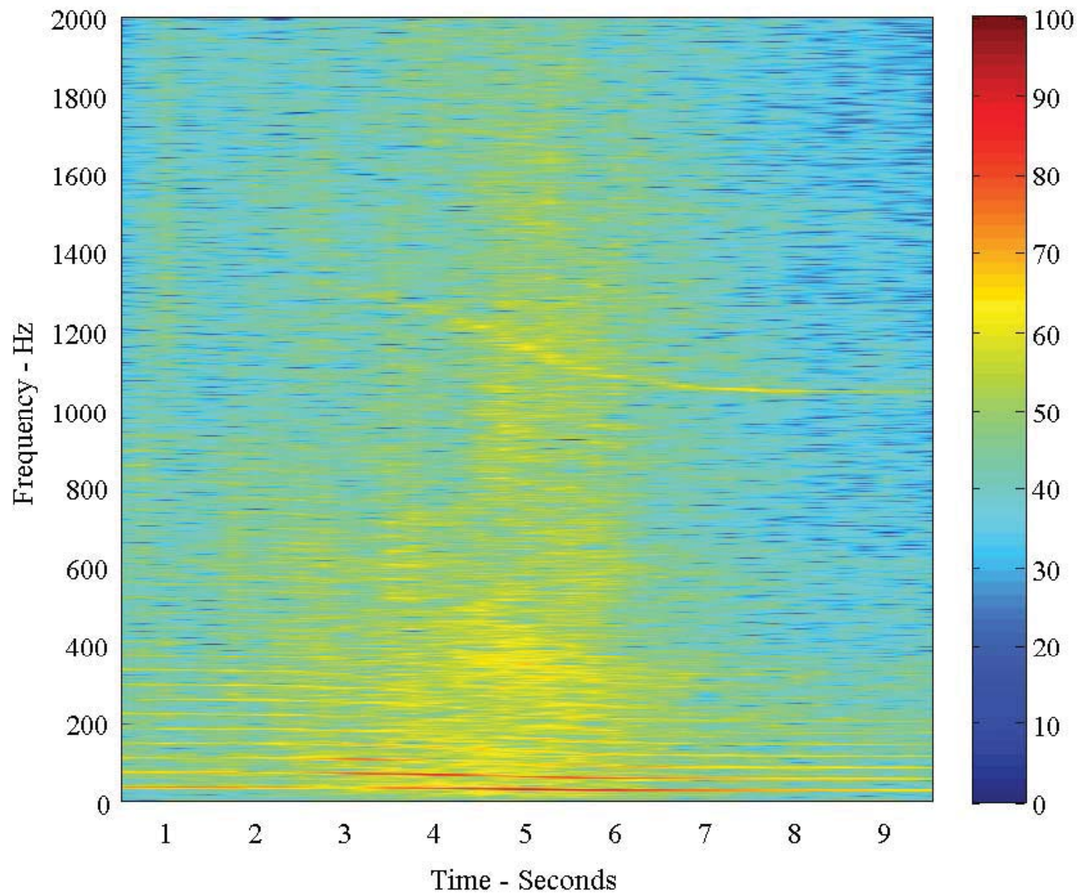

Figure 2.4. Spectrogram of the overhead section of a MD902 flyover, 84 knots, 208 ft. BPFs - 32.7 Hz and 1100 Hz.

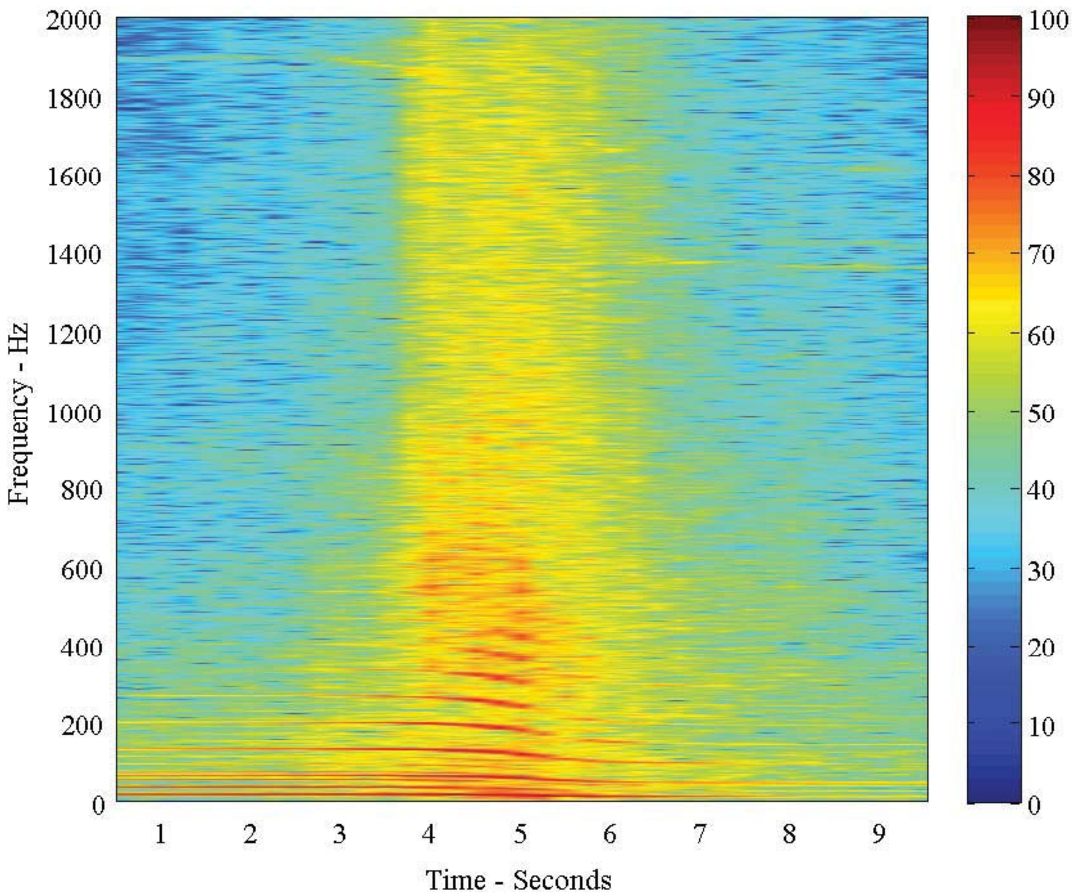

Figure 2.5. Spectrogram of the overhead section of a Mi8 flyover, 112 knots, 144 ft. BPFs - 16 Hz and 56 Hz.

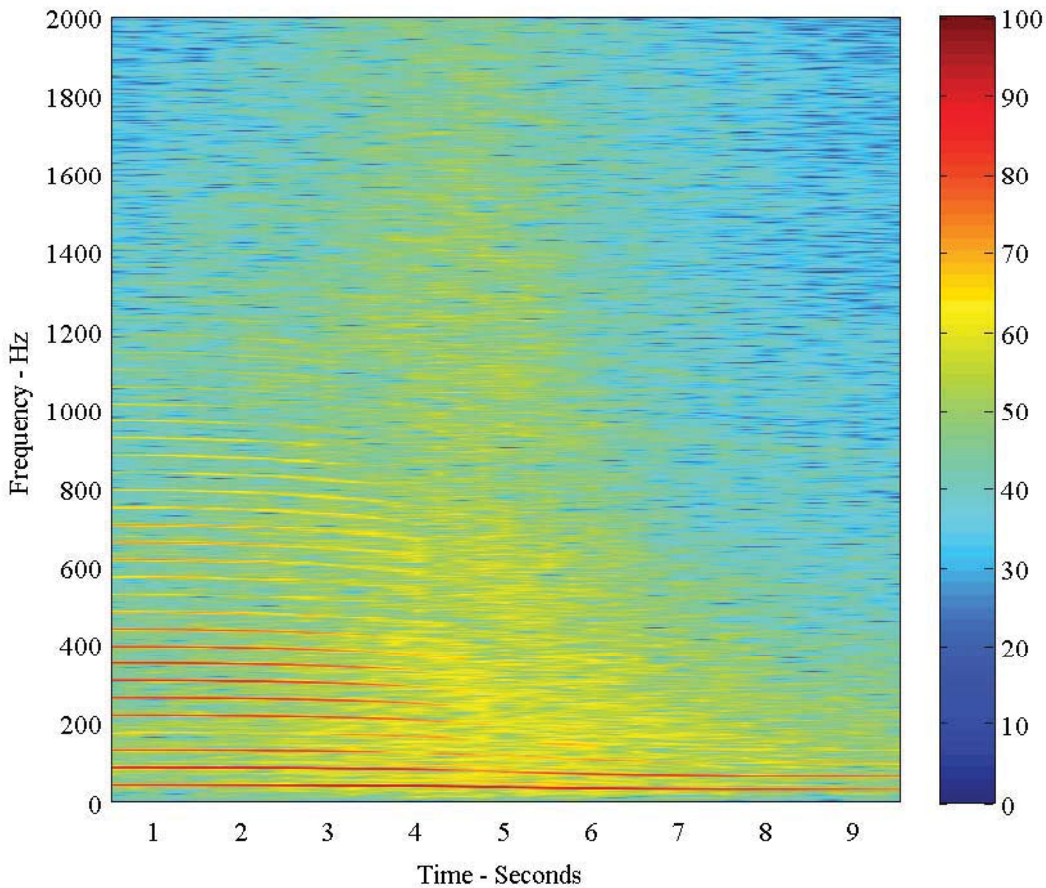

Figure 2.6. Spectrogram of the overhead section of a XV-15 flyover, 110 knots, 394 ft. BPF - 32 Hz.

2.2 Characterization and Modification of Tonal Components

A method was developed to characterize and modify the tonal components present in the collected recordings of rotorcraft flyovers. The method was designed so that perceptual attributes of the flyover (fundamental frequency, tonality, and impulsiveness) could be easily manipulated in order to develop stimuli for subjective tests. The simulation procedure is outlined in the following sections.

2.2.1 Tone Component Frequency Determination

The fundamental frequency of each harmonic series through time was predicted using Equation (2.1) with known blade passage frequencies (BPFs) and vehicle flight tracking data,

$$f_o = \frac{cf_s}{c + dv_s/dt} \ , \tag{2.1}$$

where c is the speed of sound, f_s is the frequency at the source (vehicle), f_o is the frequency at the observer (microphone), and v_s is the radial velocity of the source relative to the observer.

2.2.2 Tone Component Amplitude and Phase Estimation

The fundamental frequency at the microphone was integrated to determine phase using Equation (2.2),

$$\phi(t) = \int 2\pi f_h(t)dt \ , \tag{2.2}$$

where h is the harmonic series of interest. The amplitude of the sine (A_k) and cosine (B_k) components for each harmonic through time were estimated using Equation (2.3),

$$\begin{bmatrix} p(t) \\ \downarrow \\ t = n\Delta \end{bmatrix} = \begin{bmatrix} ... & sin(k\phi_h(t)) & cos(k\phi_h(t)) & ... \\ & ... & ... & \\ & \downarrow & \downarrow & \\ & t = n\Delta & t = n\Delta & \end{bmatrix} \begin{bmatrix} A_k \\ B_k \\ ... \\ ... \end{bmatrix} + \begin{bmatrix} n(t) \\ ... \\ \downarrow \\ t = n\Delta \end{bmatrix}, \quad (2.3)$$

where the $p(n\Delta)$ used are from a section of the time history of pressure, $r(t)$ is the distance from where the sound was emitted to the receiver location, and $n(t)$ is the contribution of non-tonal components to the sound.

To estimate the coefficients of the sines and cosines as a function of time it was found necessary to use relatively short segments of the pressure time history to achieve the best results. If the segment is too long, faster variations in the tone magnitudes are smoothed, preventing full removal of the tones (see Figure 2.7). Shorter segments also gave rise to higher variance estimates because of the more limited averaging in the normal equations used in the solution of Equation (2.3) (see Figure 2.8). A segment length equivalent to a small number of periods (2-6) of the lowest frequency being fitted to the data was found to give the best results for the signals used in this study. Additionally, if an insufficient number of harmonics were included, harmonic content not associated with the tones of interest were modeled and gave rise to additional amplitude variations of the tones modeled. This could be controlled by including a high number of harmonics (≥ 40) in the set of tones being fit to the data, essentially all the harmonics in the signal that are above the noise floor.

2.3 Signal Modifications

Having estimated the time-varying frequency, amplitudes, and phases, the tonal component is reconstructed and removed from the original sound. A small amount of this reconstruction is added back into the sound because the removal causes dips in the spectrum to appear and this adjustment puts the background levels back at the noise floor level.

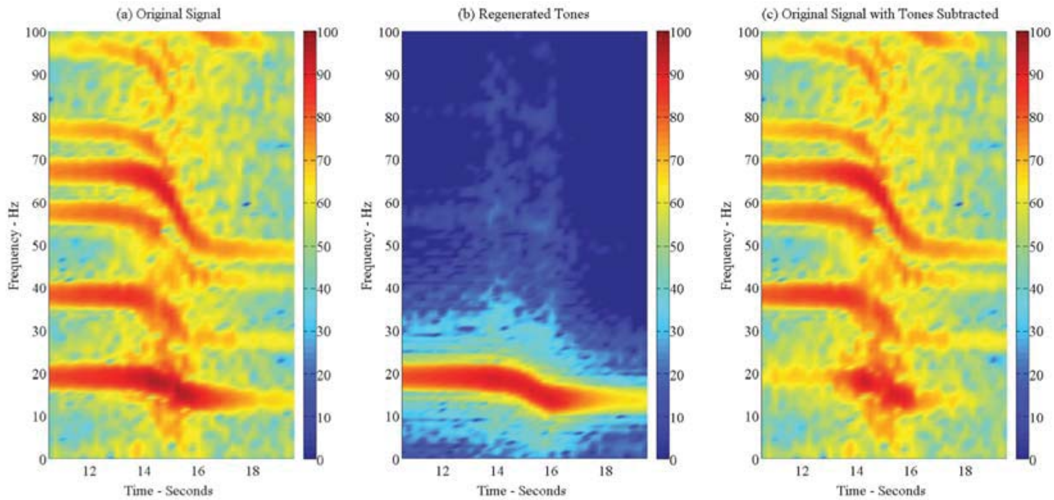

Figure 2.7. Example of tonal component estimation with too long of a window (20 periods of lowest frequency, 32000 data points), Mil Mi-8M helicopter; (a) Original Signal; (b) Regenerated matched tonal components; (c) Original Signal with matched tonal components removed.

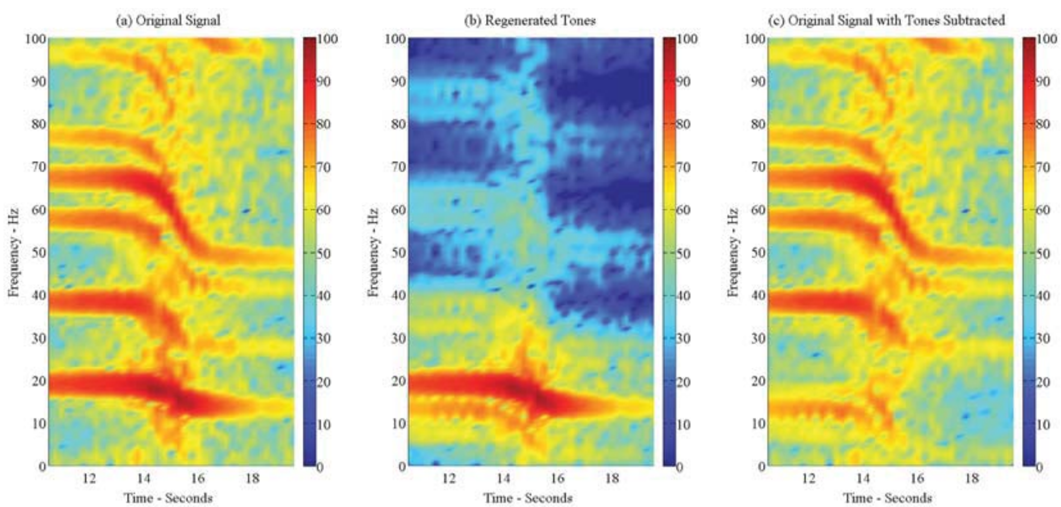

Figure 2.8. Example of tonal component estimation with too short of a window (1 period of lowest frequency, 1600 data points), Mil Mi-8M helicopter; (a) Original Signal; (b) Regenerated match tonal components; (c) Original Signal with matched tonal components removed.

2.3.1 Tonal Component Modification

The harmonic series was modified in one or more of the following ways:

1. Adjust fundamental frequency (simulates different blade passage frequency)
2. Adjust magnitude (simulates different tonal component strength)
3. Align or misalign the phase of harmonics (simulates more or less impulsive sound)

The modified harmonic series was then added to the signal from which the tonal components had been removed. In Figure 2.9 are shown segments of time histories from (1) the original recording, (2) the increased impulsiveness simulation, and (3) the decreased impulsiveness simulation.

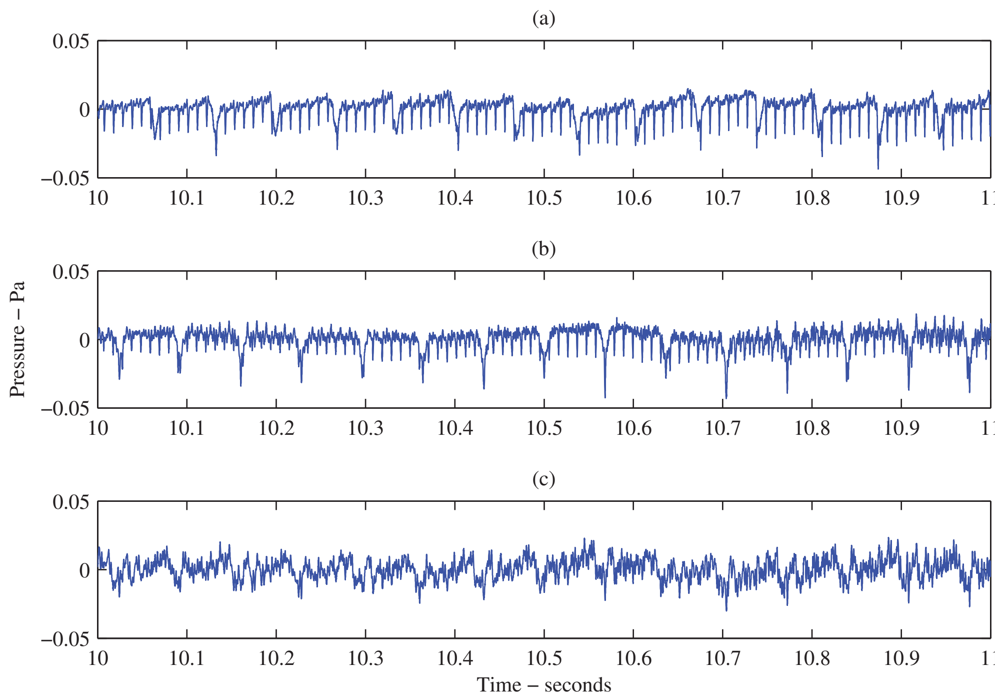

Figure 2.9. Example of tonal component modification to change impulsiveness, Bell 206 helicopter; (a) Original Signal; (b) Phase aligned tonal components; (c) Phase randomized tonal components.

2.3.2 Ground Reflections

As noted earlier, a method to modify ground level recordings to include ground reflection effects was developed and successfully implemented. In this method, the signal from the ground recording was split into two parts: the direct and the reflected. At the ground these would be the same. Based on the vehicle speed and location and the height of the persons ears above the ground, the delay and spherical spreading attenuation of each part was adjusted through time and this time-varying adjustment was based on the calculated distances and delays of when the direct signal and the ground reflection would arrive at the listeners ears. Additional atmospheric absorption effects were not simulated. The fractional delays associated with the reflected path (relative to the direct path) were managed by using a time-varying finite impulse response filter that is a modification of the sinc function in Shannon's Sampling Theorem) after the signal had been resampled to a sampling rate 10 times that of the original signal. The fractional delay method works well up to about one-tenth of the sampling rate and hence the prior up sampling was performed to account for the limitation of the fractional delay filter.

While, as illustrated in Figures 2.10 and 2.11, this worked well, (much better than simply rounding to location of the nearest point in the acquired time history), it was not used with the stimuli used in this research. This was because we did not want to add another parameter to vary beyond tone family modification at this stage in the research.

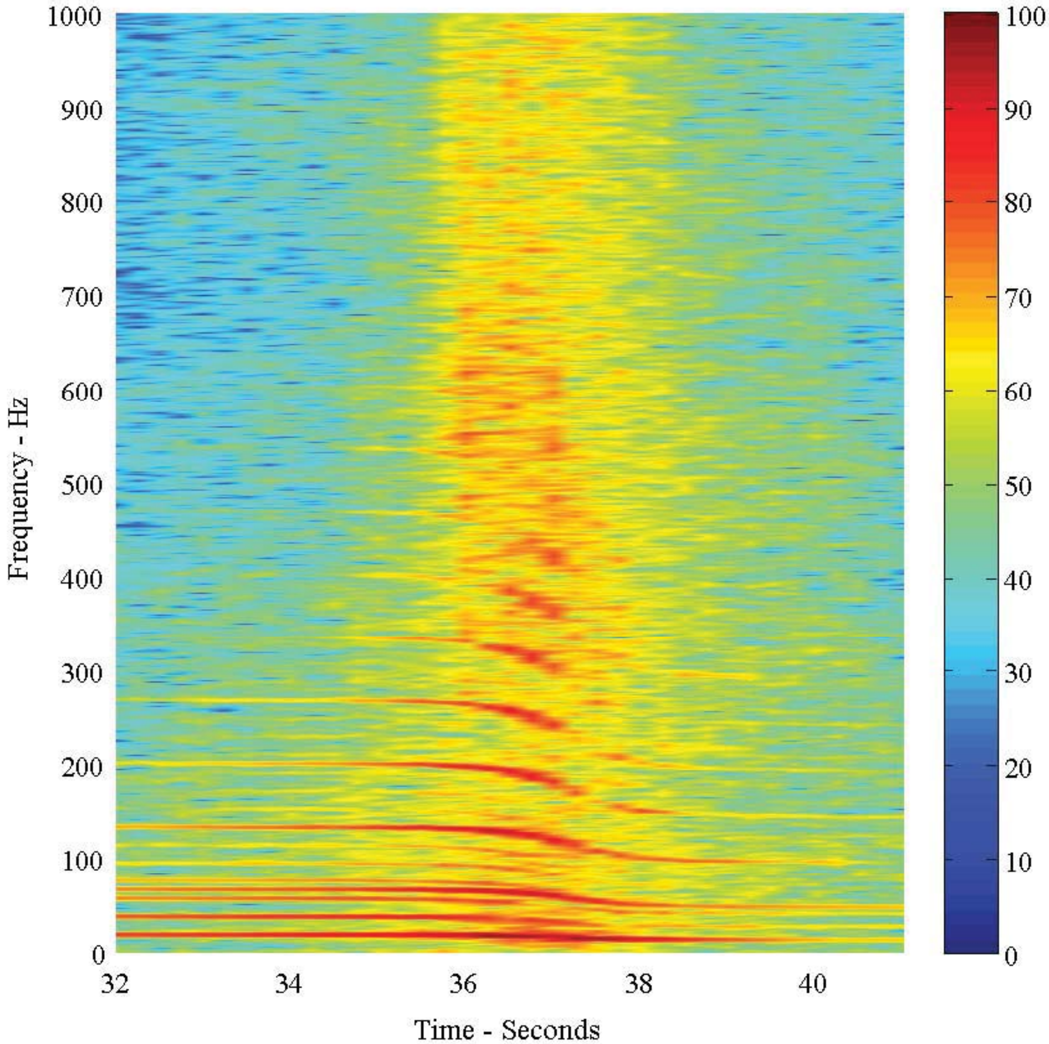

Figure 2.10. Spectrogram of a Mil Mi-8M flyover recording.

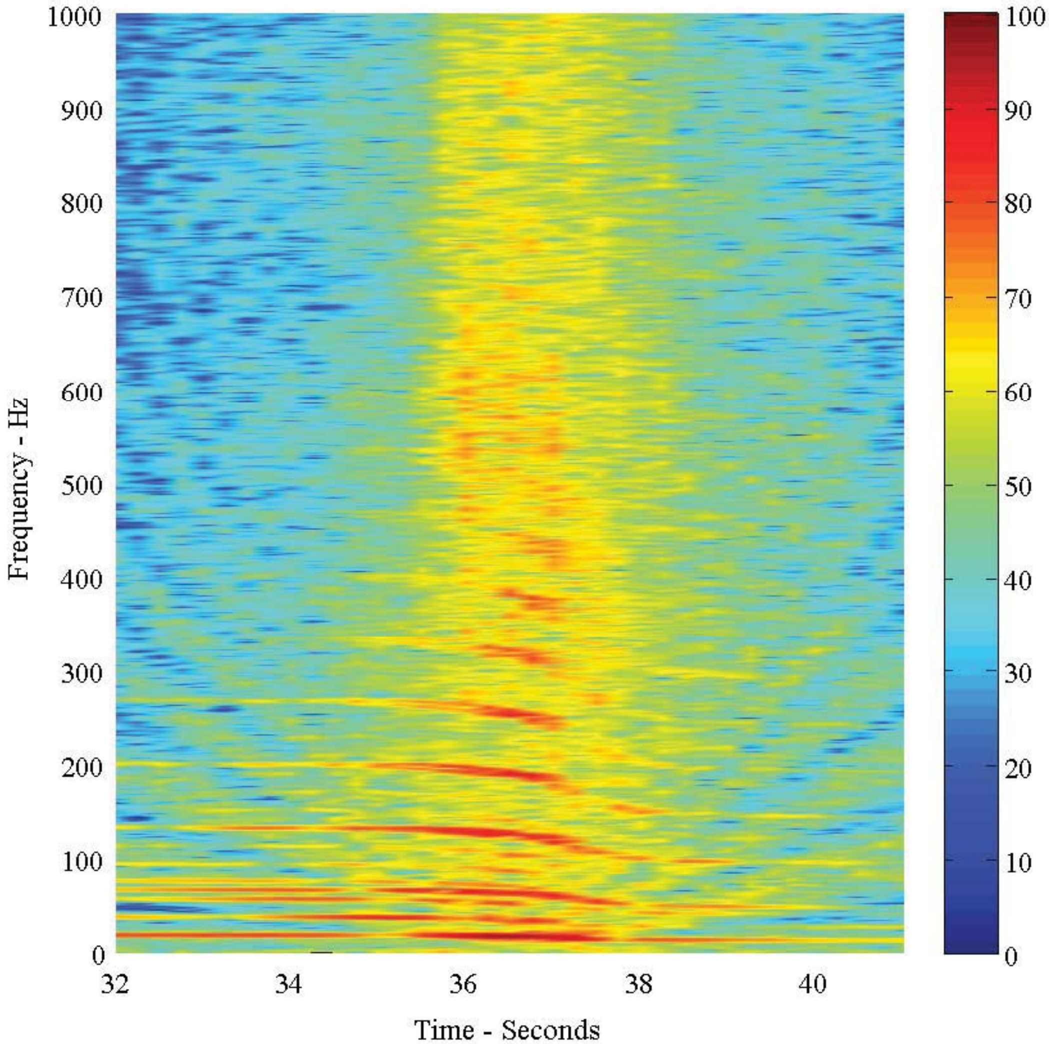

Figure 2.11. Spectrogram of a Mil Mi-8M flyover recording with ground effects added.

2.4 Metrics and Models for Time Varying Sounds

Many metrics and models exist for the evaluation of various types of noise. These include simple, level-based metrics, as well as complicated annoyance models. The following sections detail metrics and models relevant to the evaluation of general transportation noise as well as those designed specifically for the evaluation of aircraft noise.

2.4.1 SEL, L_{max}

Sound Exposure Level (SEL) is generally used to evaluate environmental noise. SEL is calculated using Equation (C.4),

$$SEL = \int_{t_1}^{t_2} \frac{p^2(t)}{p_{ref}^2} dt \ , \tag{2.4}$$

where p is sound pressure, p_{ref} is the reference pressure (20 μ Pa), and t_1 and t_2 are the instances in time where the level is 10 dB down from the maximum. SEL can also be calculated using A-weighted or C-weighted sound pressure to produce $SELA$ or $SELC$.

Other level-based metrics used for the evaluation of noise include the maximum sound pressure level, L_{max}, which can also be calculating used A-weighted or C-weighted sound pressure to produce LA_{max} or LC_{max}.

2.4.2 Loudness: Perceived Loudness, Zwicker, Moore and Glasberg

Past studies have shown that numerical estimates of human perception are proportional to the magnitude of a stimulus. This relationship is referred to as the "power law" [16]. Experimental results led to the development of Stevens' Loudness model, where the perceived loudness L is proportional to the power function of the stimuli intensity I. This model is shown in Equation (2.5),

$$L = kI^p \ , \tag{2.5}$$

where the constant k depends on the units used, and the constant p depends on the type of stimulus. For uniform noise, p is chosen to be 0.23 [2], and for a 1 kHz tone p is chosen to be 0.3. For the second case a doubling of loudness corresponds to a tenfold increase in intensity. After multiple iterations, Stevens Mark VI Loudness [17] was standardized as ISO 532-A-1975 [18] and ANSI S3.4-1980 [19]. This method calculates octave band sound pressure levels and compares their loudness to that of critical band noise at 1 kHz. Partial loudness values are then compiled into total loudness, using Equation (2.6),

$$s_t = s_m + F\left(\sum (s - s_m)\right) \ , \tag{2.6}$$

where s_m is the greatest of the loudness indices (sones), s_n are the individual loudnesses, and F are the fractional loudness factors, which take into account masking effects. F depends on the type of octave measurement (0.15 for one-third octaves, 0.3 for octaves). Stevens Mark VII Loudness [20] uses a more refined partial loudness calculation, and F is calculated based on level. Mark VII loudness gives perceived level of loudness or Perceived Loudness (PLdB).

More recent loudness models, such as Zwicker's [21] and Moore and Glasberg's [22] take into account frequency sensitivity and masking. Zwicker's model, incorporated into the standards ISO 532B [18] and DIN 45631 [23], is considered appropriate for predicting the loudness of complex, broadband noise. The biggest difference between the two loudness models is in the definition of the critical bands at low frequencies. Glasberg and Moore's bands continue to get smaller as frequencies decrease but in Zwicker's model the low frequency bandwidth is constant.

2.4.3 Sound Quality Metrics

Many sound characteristics other than loudness are involved in the perception of a sound. Sound quality metrics, such as Sharpness, Tonality, Roughness, and Fluctuation Strength, are used to quantify some of the more well known attributes so that perception of a sound can be quantified. The algorithms for calculating these metrics work well for relatively simple signals but Roughness and Fluctuation Strength algorithms are not straightforward to implement for more complicated time-varying sounds. As with all sound metrics, it should always be kept in mind that the value calculated may not reflect a person's perception of roughness or fluctuation because the algorithm may not be appropriate for the signal being analyzed. Typically, compromises are made in the calculation and these may result in poorer prediction of that sound attribute's strength.

Sharpness is a measure of spectral balance, meaning a sound is sharper when it has more high frequency content than low frequency content. A sharpness model developed by von Bismarck [24] sharpness is a function of the centroid of the loudness spectrum, with higher frequency bands weighted higher than lower frequency bands. Sharpness uses the unit acum, and is calculated using Equation (2.7),

$$S = c\,\frac{\int_0^{24} N'(z)g(z)z\,dz}{N} \quad acum, \tag{2.7}$$

where c is a normalization constant, N' is the specific loudness at critical band, and g is a weighting factor, calculated as shown below:

$$g(z) = \begin{cases} 1 & \text{for } z \leq 16 \text{ ,} \\ 0.066e^{0.171z} & \text{for } z > 16 \text{ ,} \end{cases} \tag{2.8}$$

where z is the critical band rate in Bark. Narrow band noise with 1 kHz center frequency, 160 Hz bandwidth, and a sound pressure level of 60 dB would produce a Sharpness of 1 acum.

Roughness is a measure of fast fluctuations in loudness. This metric is largest when loudness fluctuations are approximately 60 to 70 cycles per second. Zwicker and Fastl [2] developed a model that calculates roughness using Equation (2.9),

$$R = 0.3 f_{mod} \int_0^{24} \Delta L(z) dz \quad asper, \tag{2.9}$$

where z is the critical band rate in Bark, f_{mod} is the modulation frequency in kHz, and $\Delta L(z)$ is the modulation depth of the specific loudness after temporal filtering. For complex signals, $\Delta L(z)$ can be approximated by

$$\Delta L(z) = 20 \log_{10} \left(\frac{F_{N'_{max}}(z)}{F_{N'_{min}}(z)} \right) \text{ or } \Delta L(z) = 20 \log_{10} \left(\frac{F_{N'_1}(z)}{F_{N'_{99}}(z)} \right) \ , \tag{2.10}$$

where N_{max}, N_{min}, N_1, and N_{99} correspond to maximum specific loudness, minimum specific loudness, and specific loudness exceeded 1% and 99% of the time, respectively.

A sound with a 1 kHz center frequency, sound pressure of 60 dB and 100%, 70 Hz amplitude modulation produces a roughness value of 1 asper. The calculation is simple for amplitude modulated tones, but for complex signals the modulation frequency can be difficult to determine. Research on the quantification of perceived roughness is ongoing.

Slow fluctuations in loudness are quantified by fluctuation strength, which is largest for fluctuation around 4 cycles per second. Zwicker and Fastl proposed separate fluctuation strength models for broadband noise and for pure tones [2]. The broadband noise model is shown below:

$$F_{BBN} = \frac{5.8(1.25m - 0.25)(0.05L_{BBN} - 1)}{(f_{mod}/5)^2 + (4/f_{mod}) + 1.5} \quad vacil, \tag{2.11}$$

where m is the modulation factor, L_{BBN} is the broadband noise level, and f_{mod} is the modulation frequency. The pure tone model is shown below,

$$F = \frac{0.008 \int_0^{24} (\Delta L(z) dz)}{(f_{mod}/4) + (4/f_{mod})} \quad vacil, \tag{2.12}$$

which integrates modulation depth ($\Delta L(z)$) across critical band rate. The reference sound that produces a fluctuation strength of 1 vacil is a tone with a 1 kHz center frequency that is 100% amplitude modulated at 4 Hz, and has a sound pressure of 60 dB.

Many models exist that attempt to quantify the tonality or tonalness of a sound. Two such models are included in the ANSI S1.13-1995 standard - Tone-to-noise ratio and Prominence ratio. In each model, tone locations or critical bands are determined using both narrow band and critical band spectrum data. Usually only the strongest tone is used in the final value calculation. In a different model developed by Aures [25], all tonality components are summed to produce the final value. The challenge with the calculation in all of these methods lies in the dependence on the spectrum, which can cause problems when sounds are non-stationary and involve tones that changes frequency through time, and also when there are random components along with the tonal components in the signal. Window size used in the calculation can greatly affect the result because of spectral smoothing. This usually leads to an underestimation of the tonality of a sound.

Quantifying the impulsiveness of a sound has been an ongoing challenge for many researchers. Many impulse (and other) penalties have been proposed to adjust level-based metrics including those listed in [11] for DNL adjustment, also given by Schomer in [26]. In previous research studies it has often been found, that the predicted loudness exceeded 5% of the time (N_5) is highly correlated to people's perception of the loudness of an event, but the percentile that yields the highest correlation to people's responses is smaller (loudness exceeded 2% or 3% of the time) when the sound is impulsive [27].

2.4.4 Combined Models

Several environmental noise and equipment noise assessment methods combine A-weighted sound pressure level with a tone correction to improve the correlation with annoyance cause by the noise. Tone corrections vary with application (e.g., refrigeration equipment, wind turbines, aircraft) and are not always adopted, so their use is country and locality dependent as well as company and industry dependent. Tone corrections typically vary from 0 to 9 dB and are based on the prominence of the tonal component relative to levels in surrounding frequency bands (tonal prominence). Tonal prominence, which is calculated from an estimated spectrum, can be difficult to assess for complex sounds that include both random noise and tonal components, for sounds that contain multiple tones within one and adjoining critical bands, and with sounds that are varying with time.

Linear and nonlinear regression models of multiple metrics are also often used and these may work well for a defined range of operating conditions and for specific applications, but are not usually appropriate outside these ranges or for other applications. Zwicker and his group have also developed, based on a body of transportation research, a more general annoyance model which is a nonlinear function of several metrics. This is described below.

In aircraft certification Effective Perceived Noise Level ($EPNL$) is used. This is a metric that is based on Perceived Level ($PLdB$), described earlier, and incorporates both a tone correction and an event correction. For transient time-varying sounds, the summative judgment of the sound is of interest. Depending on the sound, the loudness statistic that is most correlated with peoples judgments of loudness varies. For more impulsive sounds this might be the maximum loudness, the loudness exceeded 1 to 3 percent of the time. For less impulsive sounds like those from aircraft flyovers or passing traffic people have typically used loudness exceeded 5 or 10% of the time. A problem with these is the definition of the time, typically from where the sound first exceeds the background level to when it goes below it for the last time. $SELA$ and $EPNL$ (described below) avoid this of the time by quantifying the event effect by integrating around the peak level. It should be noted that loudness perception and annoyance are not always equivalent, particularly when the class of sounds being investigated have a ranges of multiple characteristics that make sounds more or less annoying. When only the level varies and all other characteristics scale with loudness, then a loudness metric will track changes in annoyance due to noise, but care must be taken generalizing this relationship to other applications the offset and gradient of linear single loudness metric annoyance models may both change with application due to other sound characteristics as well as the relationship between the receiver of the noise and the source of the noise.

Described below are just two types of combined models that were examined in this research. The reader is referred to [28, 29] for an overview of other combined models, particularly those related to tone corrections.

2.4.4.1 Effective Perceived Noise Level

This is used in aircraft certification and is described in detail FAA Part 36 Appendix A Section 4 [30]. It is based on the Perceived Loudness ($PLdB$) metric described earlier in Section 2.4.2. It includes tone corrections and an event correction. A summary of the main calculation steps is given below.

i. Perceived Noise Level (PNL) is typically calculated every $T = 0.5$ seconds giving a time history $PNL(t)$ where $t = nT$.

ii. The prominence of individual bands in the third-octave $PNL(t, f)$ loudness spectra are examined to determine if tone corrections are needed. These corrections (between 0 and 6 dB) are added to the $PNL(t)$ calculation to give $PNLT(t)$.

iii. The maximum $PNLT(t)$ value is found and the time at which the threshold $PNLTMAX$ - 10 dB is first exceeded and the last time $PNLT(t)$ drops below this threshold are determined.

iv. $PNLT(t)$ is converted from dB to energy and this is integrated between these times and converted back to decibels to give $EPNL$.

A more detailed discussion of this metric and its applicability to shorter signals is in Appendix C.

2.4.4.2 Psychoacoustic Annoyance Models

The Psychoacoustic Annoyance model is an attempt to quantify noise annoyance caused by a variety of noise sources and thus is a function of metrics that quantify the strengths of various noise attributes including loudness. The first version of this model, called Unbiased Annoyance, was developed by Zwicker in the 1980s and its basis was the results of a number of noise studies, mostly transportation noise [2]. The intent was to predict annoyance caused by noise, independent of context and listeners former experiences, recognizing that many factors do affect annoyance serving to intensify or attenuate annoyance reactions. Such unbiased models should be useful to noise control engineers and machinery designers who can modify noise sources and control propagation and to people who manage and control environmental noise exposure through changes in operations. The models should be useful for predicting trends in annoyance as sound characteristics change not necessarily predicting actual annoyance levels which are dependent on context and the experiences and expectations of the populations exposed to the noise.

The Psychoacoustic Annoyance model developed by [2] includes measures of sharpness, fluctuation strength, roughness, and loudness in an attempt to quantify annoyance. The model is calculated using Equation (2.13):

$$PA = N_5 \left[1 + \sqrt{w_s^2 + w_{FR}^2} \right] \ , \tag{2.13}$$

where,

$$w_s = \begin{cases} 0.25(S - 1.75)\log_{10}(N_5 + 10) & \text{for } S > 1.75 \ , \\ 0 & \text{for } S < 1.75 \ , \end{cases} \tag{2.14}$$

and,

$$w_{FR} = \frac{2.18}{(N_5)^{0.4}}(0.4F + 0.6R). \tag{2.15}$$

This model is not commonly used, but has effectively explained responses to transportation noise [31]. When a sound contains little to no significant fluctuations and a low sharpness value, the model becomes equivalent to N_5. This model does not take into account tonality.

Additional work has been done to modify this model to attempt to include the tonality of the sound. A modified model developed by More and Davies [15] includes a tonality term based on Aures Tonality. The modified annoyance model is given in Equation (2.16),

$$PA_{mod} = N_5(1 + \sqrt{\gamma_0 + \gamma_1 w_s^2 + \gamma_2 w_{FR}^2 + \gamma_3 w_T^2}) \; , \tag{2.16}$$

where the tonality term is quantified by:

$$w_T^2 = [(1 - e^{-\gamma_4 N_5})^2 (1 - e^{-\gamma_5 K_5})^2]. \tag{2.17}$$

The coefficients for this model were developed based on the responses from multiple subjective tests where different parameters were varied.

2.4.4.3 Cumulative Exposure to Noise

While $EPNL$ is used in aircraft certification, it should be noted that loudness models with or without tone corrections are not generally used in measurements of environmental noise due to aircraft. Those, such as Day-Night Level (DNL) or Day-Evening-Night Level (DEN), are based on long term averages of A-weighted sound pressure level with penalties only based on time of exposure (day, evening, night). The cumulative effects of multiple exposures to rotorcraft is not addressed in this research and should be the subject of future research. While penalties (including those for tonal and impulsiveness) were a part of the proposed application of DNL [11,26] it is somewhat surprising that they are not used widely today to quantify effects of noise exposure around airports. If needed for individual aircraft noise certification, it would seem that cumulative exposure to noise measures should also include some form of penalties. This may be particularly important when trying to assess responses to both fixed wing aircraft and rotorcraft using the same noise assessment methodology.

2.5 Summary

In this chapter, the measurements gathered, simulation method developed, and various metrics and models were described. In the following chapters the metrics described above were calculated as part of the analysis process. Metrics were calculated using HEAD Acoustics Artemis software, Bruel and Kjær Type 7698 Sound Quality software, the ANOPP2 Acoustic Analysis API [32], or software developed by graduate researchers at Purdue.

3. SEMANTIC DIFFERENTIAL TESTS

Described in this chapter are the test setup, procedure, and results for the semantic differential tests that were performed using: (1) earphones for playback and (2) using a loudspeaker array for playback that can simulate aircraft flyovers. The goals of this set of tests were: (1) to determine the number of independent sound characteristics perceived when listening to a broad set of rotorcraft, and (2) to determine whether the playback method affected the responses and thus the results of the subsequent analysis.

3.1 Pre-Test Analysis

Preparation for the semantic differential tests included generation of test stimuli and a set of words pairs to be used at the ends of the scales. Ten second segments were extracted from each recording collected (as described in Section 2.1). Psychoacoustics based sound quality metrics (e.g., Loudness, Sharpness, and Tonality) were calculated for these segments [2]. The twenty signals used in the test were chosen so that the metric values spanned the same range as the full set of recordings. Sixteen of the twenty signals chosen for the tests were centered in time around the point when the craft was directly above the microphone. The remaining four signals were created by amplifying a section of the flyover that was distant (between 1500 and 4500 feet from the source along the flight path) from the overhead point. The signals were amplified through time so that they would have a time-varying loudness profile (Zwicker Loudness) similar to the corresponding profile of the overhead section of the same flight, but without the presence of a strong Doppler shift.

3.1.1 Lexicon

An informal listening test was performed at Purdue University to gather words used to describe rotorcraft noise. Volunteers listened to various segments of rotorcraft fly-over recordings that were played over loudspeakers with good low frequency response characteristics. While listening they wrote down words to describe the noise.

3.1.2 Development of Word Pairs for Ends of Scales

The words listed by the lexicon volunteers were used to generate word pairs that could be used at the end of scales for the evaluation of rotorcraft noise. These word pairs, along with word pairs used in previous semantic differential tests focused on transportation noise [33–35], were combined into a list of 88 word pairs. This list of pairs was presented to colleagues who were asked to evaluate them based on how comfortable they felt using the scales to rate rotorcraft noise. The results from this evaluation were used to narrow the list down to 19 pairs that cover a range of rotorcraft noise attributes. The final set of words pairs chosen is shown in Table 3.1. The scales are separated into four groups, which can be classified as: basic sound attributes, source characteristics, effects, and summative judgments.

3.1.3 Overview of the Two Semantic Differential Tests

The semantic differential test was designed to be performed at two locations: (1) in a Sound Quality Booth at Purdue University, Herrick Laboratories, and (2) in the Exterior Effects Room (EER) at NASA Langley Research Center. The test signals and word scales used at each location were identical. The test procedure was similar between locations with a few variations. Upon completion of the informed consent forms and hearing tests, the subjects started the test. For a listening scenario, subjects were told to think of yourself hearing these sounds several times throughout the day while you are outside, around your house or in your community. They then completed

Table 3.1. Word pairs used at the ends of the semantic differential scales.

Left word	Right word	Left word	Right word
Clean	Rumbling	Distant	Close
Dull	Sharp	Slow	Fast
Expected	Surprising	Weak	Powerful
Low Frequency	High Frequency	—	—
Gentle	Harsh	Harmless	Threatening
Smooth	Rough	Easily Ignored	Distracting
Soft	Loud	Soothing	Agitating
Not Squeaky	Very Squeaky	—	—
Steady	Irregular	Acceptable	Not Acceptable
Not Tonal	Very Tonal	Not Annoying	Very Annoying
Gently Varying	Thumping		

a familiarization section, which consisted of listening to 10 of the 20 signals without being required to make any type of response. Next, they completed a practice test, where they evaluated 2 signals on all 19 scales. Finally, the subjects completed the actual test, which consisted of evaluating the complete set 20 signals on all 19 scales. Signals were presented in a different random order for each test session, and scales were presented in a different random order for each signal-session combination.

For the practice test and actual test, signals were repeated until the subjects completed an evaluation sheet consisting of the 19 scales for the corresponding signal. Three seconds of silence were included between each repetition. Upon completion of the test, subjects were given a comment sheet and asked for feedback and comments concerning the test and signals. Subjects were compensated for participation in the test.

In the following sections the set of test signals and the variations between the tests held at the two locations are described.

3.1.4 Test Signals

The final set of semantic differential test signals are described in Table 3.2. This set included six vehicles performing a variety of operations. The start and end of the each test signal was shaped to go gently to zero but care was taken not to make the start and end have very obvious features. To accomplish this, two one-half second tapers were combined with a 9 second long rectangular window and applied to the 10-second signals. This window function affected only the first and last half-second of the signals, and did not add any noticeable artifacts.

The earphones are limited at low frequencies and the loudspeaker playback was level limited, so the signals were played at different levels in the two tests. The sound metrics for the signals used in both tests are given in Table 3.3. These were calculated using HEAD Acoustics Artemis Sound Quality Software (Version 12) based upon the raw signals described above, not the reproduced signals measured in the laboratory tests.

Table 3.2. Attributes of flight recordings used to develop stimuli used in the subjective tests. Superscript 1 denotes that a distant rather than overhead part of recording was used to generate test signal. See Table 2.1 for additional vehicle details.

Signal	Vehicle	Avg Speed (kts)	Operation
1[1]	XV-15	70	Nacelle Angle 90°
2[1]	XV-15	110	Nacelle Angle 80°
3	Bell 206	100	N/A
4	XV-15	220	Nacelle Angle 0°
5	MD902	83.6	Approach
6	Bell 206	80	N/A
7	BO105	115	N/A
8	MD520N	100	N/A
9	Mi-8M	63	Approach
10	Mi-8M	45	Approach
11	Bell 206	100	N/A
12[1]	Mi-8M	104	Level Flight
13	Mi-8M	112	Level Flight
14	MD520N	80	N/A
15	XV-15	50	Nacelle Angle 80°
16	Bell 206	100	N/A
17	BO105	80	N/A
18[1]	XV-15	110	Nacelle Angle 60°
19	MD902	65.7	Approach
20	Mi-8M	106	Level Flight

Table 3.3. Metric values for each signal played in the earphone test (1) and the loudspeaker test (2). Psychoacoustics-based metrics exceeded 5% of the time: Loudness (sones) (N_5), Sharpness (acum) (S_5), Tonality (K_5), and other level metrics: $EPNL$ (dB), $SELA$ (dB), $PLdB$.

Metric/Signal	1	2	3	4	5	6	7	8	9	10
N_5 (1)	46.5	43.7	42.3	18.9	34.6	37.8	54.5	28.2	53.0	63.7
N_5 (2)	20.5	19.3	19.0	7.93	15.4	16.9	24.8	12.4	23.8	25.1
S_5 (1)	0.80	0.67	1.32	0.94	1.40	1.33	1.17	1.13	1.06	1.27
S_5 (2)	0.78	0.64	1.31	0.93	1.38	1.32	1.16	1.11	1.06	1.27
K_5 (1)	0.13	0.15	0.42	0.12	0.12	0.43	0.21	0.10	0.02	0.04
K_5 (2)	0.12	0.15	0.48	0.13	0.12	0.50	0.24	0.11	0.03	0.04
$EPNL$ (1)	88.8	88.9	83.8	69.8	81.7	82.2	86.2	77.0	91.9	94.3
$EPNL$ (2)	74.9	75.4	70.5	55.9	68.6	68.7	73.1	63.7	78.5	78.8
$SELA$ (1)	88.3	89.1	82.4	71.0	79.9	79.9	80.5	77.0	90.0	92.6
$SELA$ (2)	75.7	76.5	69.8	58.5	67.3	67.9	72.8	64.4	77.4	77.8
$PLdB$ (1)	101.0	101.2	94.9	84.4	91.8	93.1	97.5	88.5	103.2	106.1
$PLdB$ (2)	87.8	87.9	82.3	73.1	79.9	80.9	84.7	76.8	89.9	90.2
...	...	...	...	...	...	...	...	...	...	...

Table 3.3 continued

Metric/Signal	11	12	13	14	15	16	17	18	19	20
...	...	...	...	...	...	...	...	...	...	...
N_5 (1)	28.5	65.9	72.3	27.8	42.8	28.1	32.6	45.6	48.2	69.2
N_5 (2)	12.6	29.6	33.1	12.1	18.9	12.4	14.5	19.6	21.9	31.6
S_5 (1)	1.23	1.29	1.41	1.13	1.09	1.23	1.13	0.74	1.43	1.44
S_5 (2)	1.20	1.30	1.42	1.10	1.08	1.19	1.11	0.69	1.43	1.45
K_5 (1)	0.46	0.11	0.16	0.10	0.03	0.46	0.25	0.41	0.06	0.19
K_5 (2)	0.54	0.11	0.18	0.11	0.04	0.53	0.29	0.43	0.07	0.21
$EPNL$ (1)	78.6	91.4	91.3	78.0	86.4	78.4	80.6	86.9	86.6	90.8
$EPNL$ (2)	65.2	78.1	78.4	64.8	72.3	64.9	67.1	73.4	73.5	77.8
$SELA$ (1)	78.0	87.0	89.2	77.7	84.6	77.7	79.6	86.4	85.7	88.1
$SELA$ (2)	65.4	74.4	76.6	65.2	72.1	65.1	67.0	73.8	73.1	75.5
$PLdB$ (1)	89.6	102.0	101.9	89.0	98.9	89.4	91.9	99.6	98.0	101.1
$PLdB$ (2)	77.9	88.6	89.0	77.3	85.9	77.8	79.8	86.3	85.0	88.3

3.2 Test with Sound Played Back Using Earphones

The semantic differential test was first performed using earphones for playback in a Sound Quality Booth at Purdue University, Herrick Laboratories. The setup, subjects, and variations from the baseline test are described. This is followed by the results and analysis.

3.2.1 Earphone Test Setup

The Semantic Differential Test was performed in an Acoustic Systems IAC double walled sound booth, where subjects were tested on an individual basis. For this test, signals were played back through a high quality LynxOne sound card, Tucker-Davis HB7 amplifier, and a set of Etymotic Research ER-2 tube earphones. The sound card and earphones were chosen because of their flat frequency responses and low noise floors. Disposable foam eartips (ER-14A) were used with the earphones, which add an additional 25-30 dB of background noise attenuation.

3.2.1.1 Procedures used in the Earphone Test

For this earphone test, subjects provided responses by making marks on an evaluation sheet with a pencil. Subjects were instructed to raise their hand when they completed an evaluation sheet. The Test Operator monitored the subject through a window in the Sound Booth, and advanced the test to the next signal once the subject raised their hand. The playback program was written such that when prompted to advance to the next signal, the current signal would finish playing before advancing. Subjects were compensated $10 for participation in the test.

3.2.1.2 Calibration of Signals in the Earphone Test

Two 1 kHz calibration tones were created (70 and 90 dB) using the same calibration factor as the test signals. These calibration tones were used to calibrate the left and right channels of the system playback prior to the testing of each subject. Additionally, the signal set was played back through the system and the maximum A-weighted Sound Pressure Level was recorded.

An outer ear filter was applied to all signals prior to playback to account for the transition from outdoor recordings to playback at the inner ear. This filter is based on the one described by Moore and Glasberg in their calculation of Time-Varying Loudness [36].

3.2.2 Semantic Differential Earphone Test Subjects

Subjects were recruited from Purdue Universitys campus via flyers, and consisted of students and staff. They were given a description of the test and completed a consent form and background information questionnaire. After this, a hearing test was conducted. Subjects were required to pass a hearing screening requiring no more than 20 dB of hearing loss in either ear in a range of frequencies from 125 to 8000 Hz. A total of 37 people volunteered to participate in this study, but one failed the hearing test, so only results for the remaining 36 are reported. Ten of the subjects were male, 22 were female, and four declined to respond to that question. The subjects ranged in age from 18 to 30.

3.2.3 Results

Numbers were assigned to each response based on the position of the mark on the scale. The numbers for each scale ranged from -9 (left most point) to +9 (right most point). The average of subject responses and standard deviations were calculated for each scale-signal combination. These results are shown in Table A.2 in Appendix A.2. Standard deviations ranged from 2.25 to 5.38. The minimum, maximum, and average standard deviation across all 20 signals for each scale is shown below in Table 3.4. The smallest average standard deviations were for the Soft-Loud scale (3.17 average) and the largest average standard deviations were for the Not Tonal-

Very Tonal (4.07 average) and the Acceptable-Not Acceptable (4.09 average) scales. Several subjects asked about the tonal scale during the instruction and learning part of the test indicating that there was some difficulty in using that scale.

3.2.3.1 Statistics of Responses on Different Scales

Correlations between responses on the Not Annoying-Very Annoying scale and the other 18 subjective scales were calculated. These correlation coefficients are shown in Table 3.5. When compared to the annoyance scale ratings, 12 of the 18 remaining scale ratings had a correlation coefficient above 0.9. The scale ratings most highly correlated with annoyance were from the Easily Ignored-Distracting (0.99), Acceptable-Not Acceptable (0.98) and Soft-Loud (0.98) scales. The scale ratings least correlated with annoyance were from the Not Tonal-Very Tonal (0.68), Low Frequency-High Frequency (0.66), Steady-Irregular (0.59) and Not Squeaky-Very Squeaky (0.56) scales. The standard deviation of the estimate mean (SEM) was also calculated for each scale-signal combination. These values describe the variation in the mean estimates.

The average ratings ($\pm$SEM values) for some of the scales are plotted in Figure 3.1. As expected, linear correlations are seen between specific rating pairs such as annoyance/acceptability, loudness/annoyance, and loudness/threatening. The red crosses in Figure 3.1 represent the distant amplified signals, which deviate from the linear pattern in plots (d), (e), and (f) due to their unusual character (when compared to the other sounds presented). There is also a clear identification of the squeak sound heard in signals from the Mi-8M, shown by the separation of gray crosses in (f).

3.2.3.2 Factor Analysis

A factor analysis was performed on the subject responses to determine the independent factors across the various scale ratings. Using this method, the subjective responses can be explained using a smaller number of underlying factors, some of which may not be directly observed. Factor analysis was performed on the full set

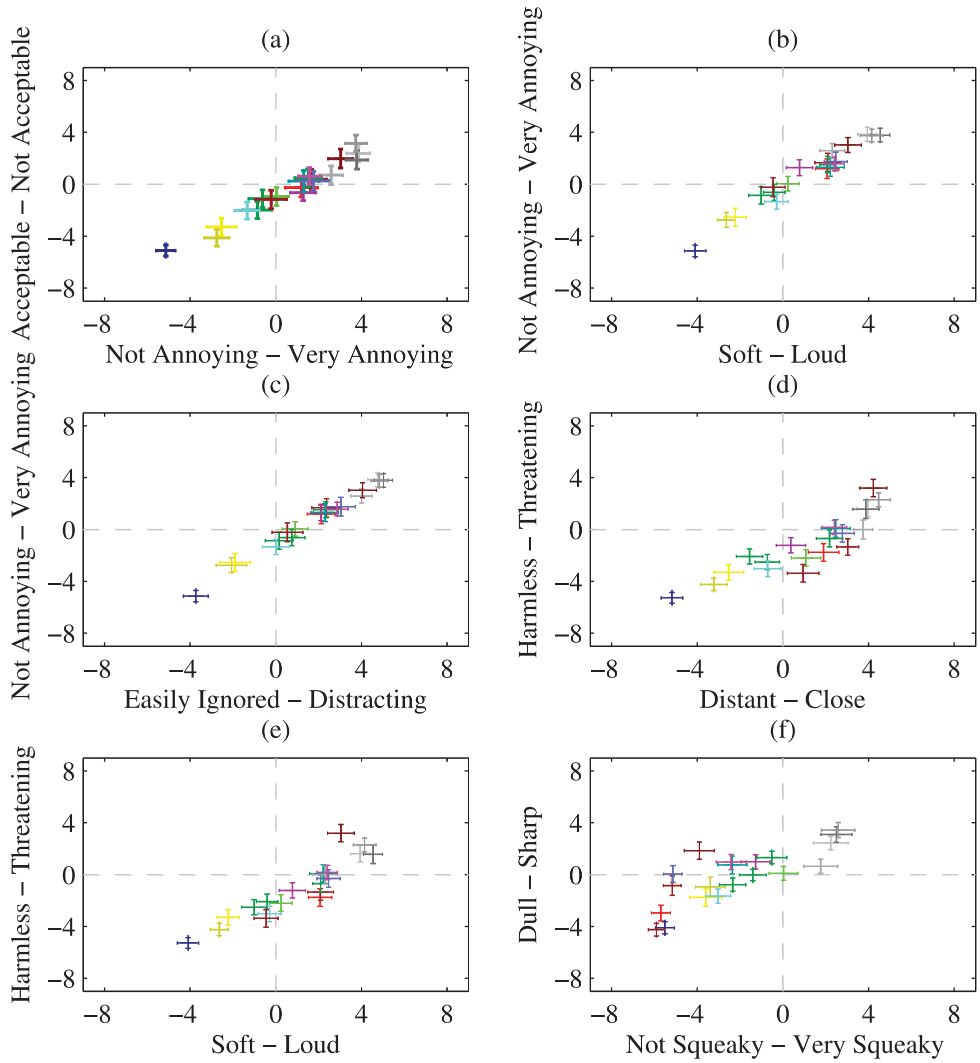

Figure 3.1. Various mean scale ratings from the semantic differential test using earphones plotted against each other. Error bars on each plot correspond to the standard deviation of the estimated mean. Points are color-coded by vehicle, and range from light to dark based on craft speed. The amplified distant signals are grouped separately. Bell 206 - green (2 speeds), BO105 - cyan (2 speeds), Mi-8M - grey (4 speeds), MD902 - magenta (2 speeds), MD520N - yellow (2 speeds), XV-15 - blue (2 speeds), amplified distant signals - red (3 speeds).

Table 3.4. For each scale-signal combination, subjects' responses were averaged and the standard deviation was calculated. The minimum, maximum, and average of the standard deviation values across all 20 signals are shown ordered by average value. Scale extremes ranged from -9 to +9. The numbers in parentheses denote which signal the minimum or maximum standard deviation refers to.

Subjective Scale	Minimum	Maximum	Average
Soft-Loud	2.22 (3)	3.88 (7)	3.17
Soothing-Agitating	2.34 (20)	4.02 (1)	3.22
Gentle-Harsh	2.29 (20)	4.12 (2)	3.27
Smooth-Rough	2.36 (4)	5.30 (12)	3.45
Weak-Powerful	2.82 (10)	4.45 (2)	3.49
Dull-Sharp	2.62 (11)	4.59 (8)	3.49
Clean-Rumbling	2.30 (13)	4.95 (18)	3.51
Easily Ignored-Distracting	2.35 (20)	4.33 (14)	3.56
Distant-Close	2.65 (9)	4.40 (18)	3.64
Not Annoying-Annoying	2.63 (4)	4.53 (1)	3.64
Harmless-Threatening	2.46 (4)	4.38 (9)	3.68
Gently Varying-Thumping	2.83 (4)	5.16 (1)	3.69
Steady-Irregular	2.93 (1)	4.99 (12)	3.73
Low Frequency-High Frequency	3.02 (7)	4.67 (12)	3.74
Slow-Fast	3.18 (5)	4.31 (12)	3.77
Not Squeaky-Squeaky	2.33 (18)	4.89 (10)	3.79
Expected-Surprising	3.44 (6)	4.63 (7)	3.98
Not Tonal-Very Tonal	3.43 (5)	4.98 (2)	4.07
Acceptable-Not Acceptable	2.54 (4)	4.93 (7)	4.09

of scales with two factors and repeated with an increasing number of factors until the data could be explained at a 95% significance level. For the full set of scales,

Table 3.5. Correlation coefficients (ρ) between annoyance scale ratings and the 18 other scale ratings in the earphone test.

Subjective Scale	Correlation Coefficient ρ	p-value
Easily Ignored-Distracting	0.99	<0.0001
Acceptable-Not Acceptable	0.98	<0.0001
Soft-Loud	0.98	<0.0001
Soothing-Agitating	0.97	<0.0001
Distant-Close	0.97	<0.0001
Smooth-Rough	0.97	<0.0001
Gentle-Harsh	0.97	<0.0001
Weak-Powerful	0.96	<0.0001
Gently Varying-Thumping	0.93	<0.0001
Expected-Surprising	0.93	<0.0001
Harmless-Threatening	0.93	<0.0001
Clean-Rumbling	0.93	<0.0001
Dull-Sharp	0.78	0.0001
Slow-Fast	0.74	0.0002
Not Tonal-Very Tonal	0.68	0.0009
Low Frequency-High Frequency	0.66	0.0015
Steady-Irregular	0.59	0.0062
Not Squeaky-Squeaky	0.56	0.0106

the p-value did not exceed 0.05 until 9 factors were used, resulting in a p-value of 0.1592. To simplify interpretation of factor analysis results, this process was repeated on a subset of 11 scales that described sound characteristics, as shown on the left half of Table 3.1. For this scale subset, the p-value remained below 0.05 until 4 factors were used, resulting in a p-value of 0.0745. The factor loading matrices were rotated using the promax rotation algorithm, which first rotates to the orthogonal varimax solution, then relaxes restrictions on orthogonality to fit a simpler structure.

While nine factors were necessary to explain the variance present in the results including all 19 scales, it should be noted that only five of these factors have individual loadings above 0.6, which is a commonly used cutoff for significance in factor analysis [37, 38]. For this analysis, factor loadings of 0.6 or greater, along with the highest loading for each scale that does not have a loading above 0.6, will be deemed significant. Factor loadings produced from the nine-factor analysis of the full set of scales are presented in Figure 3.2, and significant loadings are sorted by factor in Table 3.6.

Sorting by factor and requiring each scale to be represented once forces groupings that may help understand the meaning of each factor. The groupings shown in Table 3.6 show the possibility of multiple strong factors. Factor 1 contributes strongly to scales relating to impressions of the sound, and is also the factor with the highest loading on the Soft-Loud scale. Factor 2 mainly has high loadings for scales relating to spectral balance, and Factor 3 has high loadings for scales relating to roughness. The rest of the factors (4-9) each have significant loadings for one scale. These groupings are a starting point for determining the true number of independent factors present in human response, and will help guide the design of future subjective tests.

This grouping process was repeated on the results from the four-factor analysis of the sound characteristic scales as shown on the left half of Table 3.1. The factor loadings for the four-factor analysis of the subset of scales and the significant loadings are shown in Figure 3.3 and Table 3.7.

In the four-factor analysis of sound characteristic scale ratings, all four factors had at least one loading exceeding 0.6. Some similar groupings are present in the subset factor analysis, such as the spectral balance grouping seen in Factor 3 and the pairing between the Clean-Rumbling and Smooth-Rough scales in Factor 1. However, this factor grouping puts the Steady-Irregular scale separate from the spectral balance scales, which shows the possibility of an Irregularity factor.

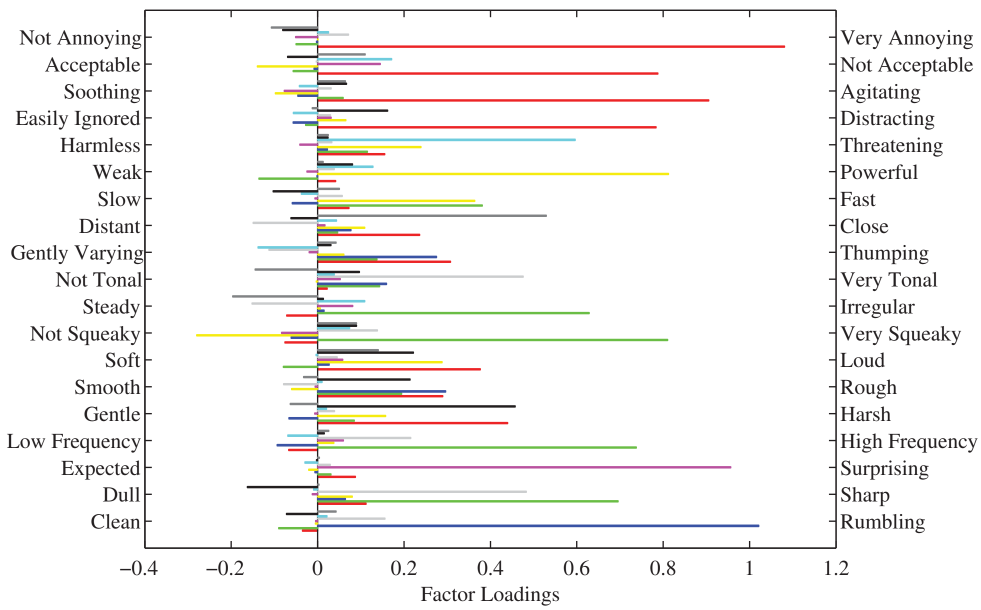

Figure 3.2. Factor loadings from a nine-factor analysis of ratings on all scales in the earphone semantic differential test. Factors 1 through 9 are coded as follows: red, green, blue, yellow, pink, light gray, cyan, black, and dark gray, respectively.

Table 3.6. Significant factor loadings from a nine-factor analysis of ratings on all scales in the earphone semantic differential test. Colors from Figure 3.2: Factor 1 - Red, Factor 2 - Green, Factor 3 - Blue, Factor 4 - Yellow. Factor 5 - Pink, Factor 6 - Light gray, Factor 7 - Cyan, Factor 8 - Black, Factor 9 - Dark gray.

Factor	Scale	Loading
1	Easily Ignored-Distracting	0.78
1	Soothing-Agitating	0.91
1	Acceptable-Not Acceptable	0.79
1	Not Annoying-Very Annoying	1.08
1	Soft-Loud	0.38
1	Gently Varying-Thumping	0.31
2	Dull-Sharp	0.70
2	Low Frequency-High Frequency	0.74
2	Not Squeaky-Very Squeaky	0.81
2	Steady-Irregular	0.63
2	Slow-Fast	0.38
3	Clean-Rumbling	1.02
3	Smooth-Rough	0.30
4	Weak-Powerful	0.81
5	Expected-Surprising	0.96
6	Not Tonal-Very Tonal	0.48
7	Harmless-Threatening	0.60
8	Gentle-Harsh	0.46
9	Distant-Close	0.53

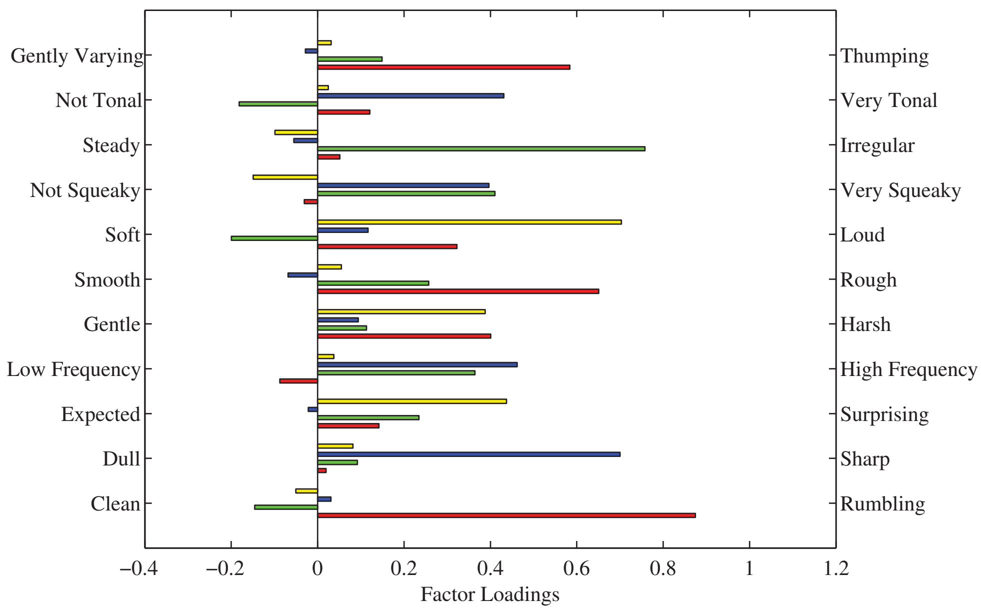

Figure 3.3. Factor loadings from a four-factor analysis of ratings on sound characteristic scales in the earphone semantic differential test. Factors 1 through 4 are coded as follows: red, green, blue, and yellow, respectively.

Table 3.7. Significant factor loadings from a four-factor analysis of ratings on sound characteristic scales in the earphone semantic differential test. Colors from Figure 3.3: Factor 1 - Red, Factor 2 - Green, Factor 3 - Blue, Factor 4 - Yellow.

Factor	Scale	Loading
1	Clean-Rumbling	0.87
1	Smooth-Rough	0.65
1	Gentle-Harsh	0.40
1	Gently Varying-Thumping	0.58
2	Steady-Irregular	0.76
2	Not Squeaky-Very Squeaky	0.41
3	Dull-Sharp	0.70
3	Low Frequency-High Frequency	0.46
3	Not Tonal-Very Tonal	0.43
4	Soft-Loud	0.70
4	Expected-Surprising	0.44

3.2.3.3 Relationship between Metrics and Responses

Average loudness and annoyance ratings were compared to various level metrics and the results are shown in Figures 3.4 and 3.5 and Table 3.8. Level metric calculations were performed on source signals at sound levels corresponding to playback levels, i.e., not from recordings in the playback environment. Statistics of Zwicker Time-Varying Loudness were calculated using the HEAD Acoustics ArtemiS (Version 12). A-weighted Sound Exposure Level ($SELA$) and Effective Perceived Noise Level ($EPNL$) were calculated using the ANOPP2 Acoustic Analysis API [32]. $EPNL$ uses a 10 dB down from peak level cutoff for calculation. This 10 dB drop may not occur in test signals at the same time as in the full flyover, giving different values for $EPNL$ for full flyovers and test signals. Shown in this section are the $EPNL$ values as calculated for the test signals only. A discussion of the $EPNL$ metric and its applicability to shorter signals is included in Appendix C.

For this test, Effective Perceived Noise Level ($EPNL$) when fitted to the average response data resulted in the highest coefficient of determination (R^2) for the Soft-Loud scale (R^2 = 0.8479) and for the Not Annoying-Very Annoying Scale (R^2 = 0.8770). Coefficient of determination values were recalculated for every possible combination of 18 of the 20 signals in order to understand the variability present in the correlation between metrics and responses. Using subsets of 18 generates a set of 190 R^2 values, and the means and standard deviations of these sets are presented in Table 3.8 in parentheses, respectively, next to the corresponding values.

3.3 Test with Sound Played Back Using Loudspeakers

The Semantic Differential Test was also performed using loudspeakers for playback in the Exterior Effects Room (EER) at NASA Langley Research Center. The setup, subjects, and test specific procedures are described, followed by a presentation of the results and analysis.

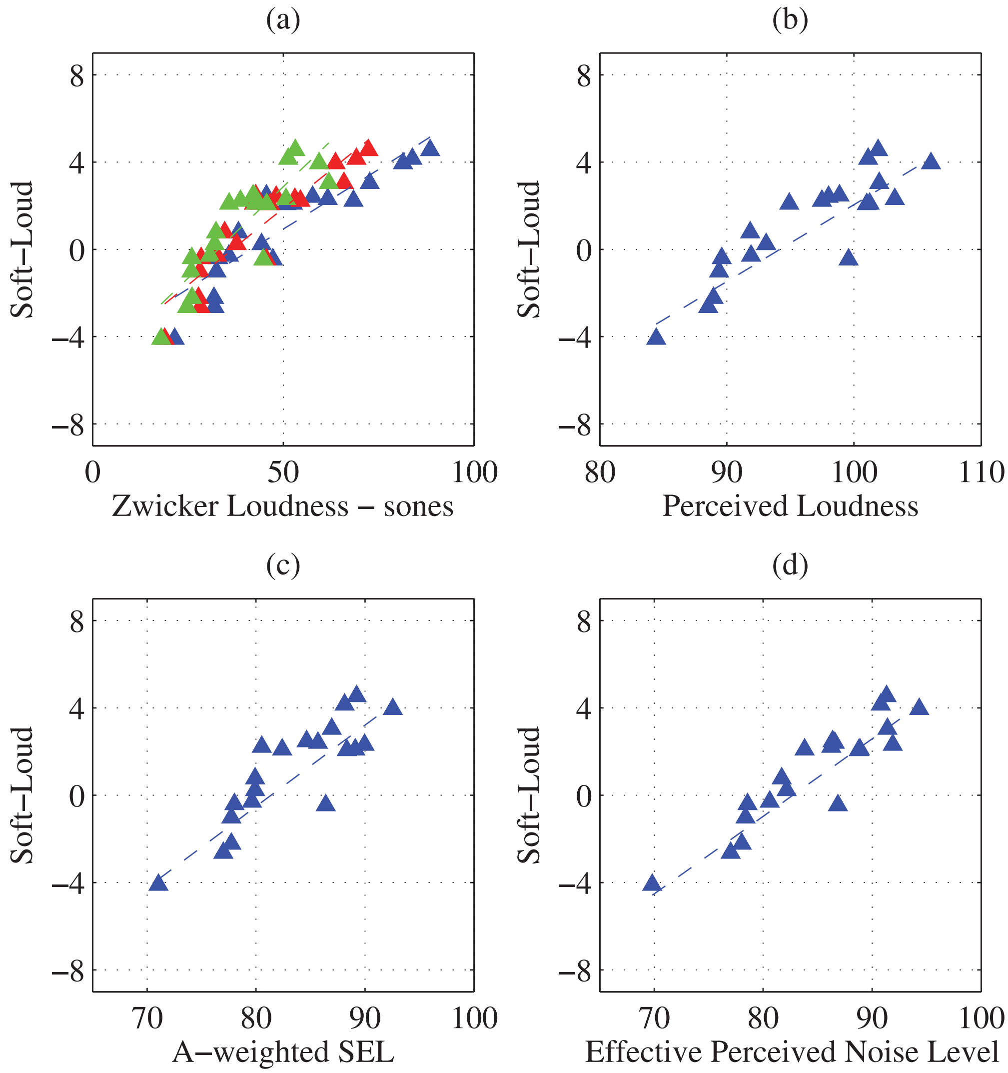

Figure 3.4. Average of the loudness ratings in the earphone test plotted against various metrics: (a) statistics of Zwicker Time-Varying Loudness calculated using HEAD Acoustics Artemis Sound Quality Software (Version 12) (DIN 45631): maximum (blue triangles, R^2 = 0.8039), level exceeded 5% of the time (red triangles, R^2 = 0.8249), and level exceeded 10% of the time (green, R^2 = 0.7627). (b) Perceived Loudness ($PLdB$), R^2 = 0.7859; (c) A-weighted Sound Exposure Level ($SELA$), R^2 = 0.7422; and (d) Effective Perceived Noise Level ($EPNL$), R^2 =0.8479.

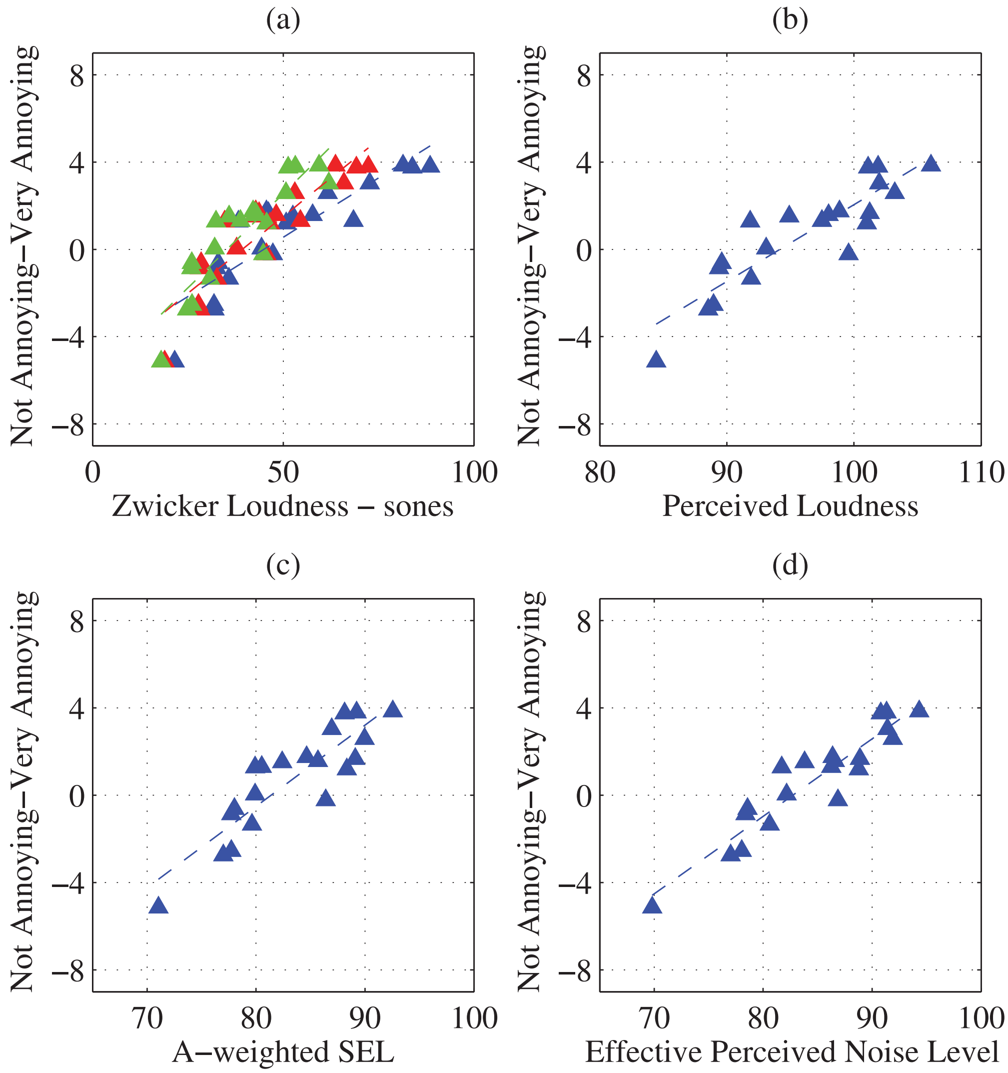

Figure 3.5. Average of the annoyance ratings from the earphone test plotted against various metrics: (a) statistics of Zwicker Time-Varying Loudness calculated using HEAD Acoustics Artemis Sound Quality Software (Version 12) (DIN 45631): maximum (blue triangles, $R^2 = 0.7770$), level exceeded 5% of the time (red triangles, $R^2 = 0.8092$), and level exceeded 10% of the time (green, $R^2 = 0.7894$). (b) Perceived Loudness ($PLdB$), $R^2 = 0.7945$; (c) A-weighted Sound Exposure Level ($SELA$), $R^2 = 0.7713$; and (d) Effective Perceived Noise Level ($EPNL$), R^2 =0.8770.

Table 3.8. Coefficients of determination values (R^2) for single factor models of individual metrics as predictors of average loudness and annoyance responses in the earphone test. N_{max}, N_5, and N_{10} denote statistics of Zwicker Time-Varying Loudness. Values in parentheses denote mean and standard deviations of R^2 values calculated as described in Section 3.2.3.3.

Metric/Scale	Soft-Loud (R^2)	Not Annoying-Very Annoying (R^2)
N_{max}	0.8039 (0.8040, 0.0201)	0.7770 (0.7780, 0.0216)
N_5	0.8249 (0.8248, 0.0188)	0.8092 (0.8099, 0.0177)
N_{10}	0.7627 (0.7628, 0.0301)	0.7894 (0.7900, 0.0225)
$PLdB$	0.7859 (0.7850, 0.0328)	0.7945 (0.7936, 0.0291)
$SELA$	0.7422 (0.7405, 0.0399)	0.7713 (0.7692, 0.0353)
$EPNL$	0.8479 (0.8467, 0.0261)	0.8770 (0.8753, 0.0230)

3.3.1 Specific Test Setup and Variations from Baseline Test

This test was performed in the Exterior Effects Room (EER) at NASA Langley Research Center. For this test, signals were played back through a set of 27 K&H O300 mid and high frequency satellite speakers and 4 K&H O900 subwoofers [39]. Playback was controlled by a real-time audio server using an implementation of Vector-Base Amplitude Panning. This form of playback allows for the simulation of flyovers with any given flight path. However, playback was limited to a lower sound level than the earphone due to the capabilities of the loudspeakers. Metrics were recalculated for these lower level signals, and results were evaluated accordingly.

3.3.1.1 Procedures used in the Loudspeaker Test

Similar to the earphone test, subjects began by completing a familiarization section and practice test before the actual test. The signals and scales used in this test were the same as the earphone test, except for the adjusted level. For this test, subjects completed evaluations on tablet PCs. Subjects used either their finger or a stylus to move a marker on each scale. Scales were presented across two pages, and buttons at the bottom of the page allowed subjects to go back and forth between the two pages, and mark their completion of each evaluation. The signals were repeated until all subjects had completed the corresponding evaluation. This test was broken up into three sessions of 6, 7, and 7 sound evaluations with short breaks (a maximum of 5 minutes) between the sessions. Upon completion of the test, subjects were given an optional comment sheet and asked for feedback and comments on the test and signals. Each subject's hearing was tested again and they were compensated $50 in addition to travel mileage compensation.

3.3.1.2 Calibration of Signals in the Loudspeaker Test

The same calibration tones used in the earphone test (1000 Hz, 70 dB and 90 dB) were used for calibration. These two test tones were used to calibrate the system playback prior to each test group.

Two 1 kHz calibration tones were created (70 and 90 dB) using the same calibration factor as the test signals. These calibration tones were used to calibrate the system playback prior to the testing of each subject. However, due to the level limitations of the system, the 70 dB and 90 dB tones were calibrated to levels in the EER of 58 dB and 78 dB, respectively. Additionally, the three highest-level signals were played back through the system and the maximum A-weighted Sound Pressure Levels were recorded.

3.3.2 Semantic Differential Loudspeaker Test Subjects

Subjects were recruited from the general public around the NASA Langley area. Subjects were required to pass a hearing screening requiring no more than 30 dB of hearing loss over the 125 to 8000 Hz range. This differs from the Purdue Testing

Figure 3.6. Picture illustrating use of cloth screens to separate subjects. People shown did not participate as subjects in the test.

requirement of a max of 20 dB hearing loss. The 30 dB value was chosen as a compromise between 20 dB and 40 dB, the latter being the levels typically used by NASA Langley researchers in other subjective tests. Regina D. Johns, a Certified Occupational Hearing Conservationist (COHC), performed recruitment and hearing screenings. As part of the recruitment process, each subject was required to meet with the COHC before the beginning of testing and complete their hearing screening. At this point they were given the scheduled day and time of their test session. Subjects participated in groups of 4, and test sessions were held twice a day for five days. Subjects were separated by cloth screens, illustrated in Figure 3.6 and instructed to not make any sounds or motions that would affect the responses of any other subjects. A total of 40 people volunteered to participate in the test, however one subject was unable to perform the test so the number of subjects was 39. Subjects ranged in age from 18-67, 15 were male, and 24 were female.

3.3.3 Results

Numbers were assigned to each response based on the position of the mark on the scale by the computer software on the subjects tablet PCs. The software assigned a number between 1 (left most point) and 5 (right most point) to a precision of two decimal places. These values were translated to a scale of -9 to 9 so that results would be comparable between the two tests. Similar to the results from the earphone test, the average of the subjects' responses and standard deviations were calculated for each scale-signal combination. These results are shown in Table A.4 in Appendix A.3. The minimum, maximum, and average standard deviation across all 20 signals for each scale is shown below in Table 3.9. The smallest average standard deviations were for the Soft-Loud scale (3.21 average) and the largest average standard deviations were for the Not Tonal-Very Tonal scale (4.36 average). Similar to the earphone test, many subjects asked about the tonal scale, indicating difficulty in using that scale.

3.3.3.1 Statistics of Responses on Different Scales

Correlations between responses on the Not Annoying-Very Annoying scale and the other 18 subjective scales were calculated. These correlation coefficients are shown in Table 3.10.

When compared to the annoyance scale, 13 of the 18 remaining scales had a correlation coefficient above 0.9. The scales most highly correlated with annoyance were the Easily Ignored-Distracting, Soothing-Agitating, Gentle-Harsh, Acceptable-Not Acceptable scales, all with a correlation coefficient of 0.99. The scales with the least amount of correlation with annoyance were the Not Squeaky-Very Squeaky (0.61) and Slow-Fast (0.57) scales. The standard deviation of the estimate mean (SEM) was also calculated for each signal-scale combination. These values describe the variance of the mean estimates. The average ratings ($\pm$ SEM values) for some of the scales are plotted in Figure 3.7. Similar to the results shown in Figure 3.1, there is a strong linear trend present in these plots, and the responses to the Mi-8M

Table 3.9. For each scale-signal combination in the loudspeaker test, subjects' responses were averaged and the standard deviation was calculated. The minimum, maximum, and average of the standard deviation values across all 20 signals are shown ordered by average value. Scale extremes ranged from -9 to +9. The numbers in parentheses denote which signal the minimum or maximum standard deviation refers to.

Subjective Scale	Minimum	Maximum	Average
Soft-Loud	2.37 (4)	4.10 (12)	3.21
Gentle-Harsh	1.86 (4)	4.31 (2)	3.25
Soothing-Agitating	2.80 (4)	4.24 (12)	3.41
Weak-Powerful	2.65 (13)	4.28 (8)	3.48
Harmless-Threatening	2.57 (16)	4.63 (12)	3.55
Distant-Close	2.68 (20)	4.29 (18)	3.59
Expected-Surprising	2.28 (4)	4.60 (2)	3.61
Smooth-Rough	2.70 (8)	4.76 (12)	3.61
Not Annoying-Annoying	2.33 (4)	4.65 (12)	3.63
Acceptable-Not Acceptable	2.21 (4)	4.71 (10)	3.66
Steady-Irregular	2.30 (4)	5.81 (1)	3.71
Not Squeaky-Squeaky	1.81 (4)	5.14 (10)	3.72
Gently Varying-Thumping	3.00 (8)	4.84 (2)	3.77
Dull-Sharp	2.72 (14)	4.91 (2)	3.78
Slow-Fast	2.93 (3)	4.41 (1)	3.81
Low Frequency-High Frequency	2.27 (4)	4.74 (20)	3.88
Easily Ignored-Distracting	2.55 (8)	4.69 (11)	3.92
Clean-Rumbling	2.39 (1)	5.07 (2)	4.10
Not Tonal-Very Tonal	3.70 (1)	5.38 (10)	4.36

Table 3.10. Correlation coefficients (ρ) between annoyance scale ratings and the 18 other scale ratings in the loudspeaker test.

Subjective Scale	Correlation Coefficient ρ	p-value
Easily Ignored-Distracting	0.99	<0.0001
Soothing-Agitating	0.99	<0.0001
Gentle-Harsh	0.99	<0.0001
Acceptable-Not Acceptable	0.99	<0.0001
Soft-Loud	0.98	<0.0001
Expected-Surprising	0.98	<0.0001
Clean-Rumbling	0.97	<0.0001
Smooth-Rough	0.97	<0.0001
Harmless-Threatening	0.96	<0.0001
Distant-Close	0.95	<0.0001
Weak-Powerful	0.95	<0.0001
Gently Varying-Thumping	0.95	<0.0001
Not Tonal-Very Tonal	0.94	<0.0001
Low Frequency-High Frequency	0.85	<0.0001
Dull-Sharp	0.83	<0.0001
Steady-Irregular	0.80	<0.0001
Not Squeaky-Squeaky	0.61	0.0045
Slow-Fast	0.57	0.0082

signals in plot (f) deviate most from the linear trends, but not as much as Figure 3.1 (Purdue Test).

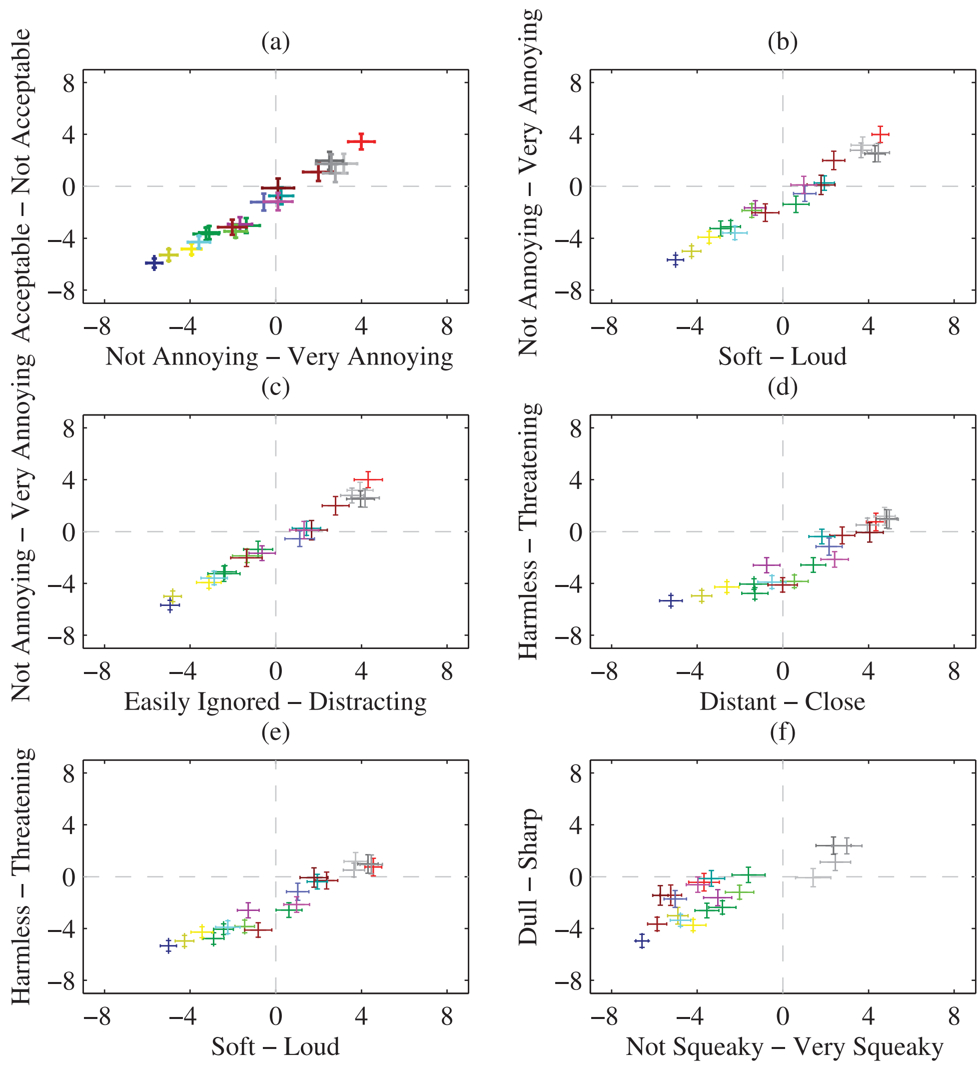

Figure 3.7. Various mean scale ratings from the semantic differential test using loudspeakers plotted against each other. Error bars on each plot correspond to the standard deviation of the estimated mean. Points are color-coded by vehicle, and range from light to dark based on craft speed. The amplified distant signals are grouped separately. Bell 206 - green (2 speeds), BO105 - cyan (2 speeds), Mi-8M - grey (4 speeds), MD902 - magenta (2 speeds), MD520N - yellow (2 speeds), XV-15 - blue (2 speeds), amplified distant signals - red (3 speeds).

3.3.3.2 Factor Analysis

As with the earphone test data, a factor analysis was performed on the subject responses from the loudspeaker test to determine common factors. The analysis was performed on the full set of scales, as well as the subset of sound characteristic scales (shown on the left half of Table 3.1), and repeated with an increasing number of factors until a p-value of 0.05 of greater was obtained. For the full set of scales, 8 factors were necessary to sufficiently explain the variance present, resulting in a p-value of 0.0575. For the subset of sound characteristic scales, a p-value of 0.30 was obtained once the number of factors was increased to 5. To allow for comparison between factor analysis results from the two tests, the scale grouping procedure from section 3.2.3.2 was repeated. The factor loadings for the full set of scales and subset of scales are presented first in Figure 3.8 and Table 3.11, followed by the loadings and groupings for the subset of scales in Figure 3.9 and Table 3.12.

Similar to the factor analysis performed on the results from the earphone test, there appears to be a factor relating to impressions of the sound, a spectral balance factor, and a roughness factor. Many of the individually grouped scales, such as Expected-Surprising, Not Tonal-Very Tonal, and Harmless-Threatening, were also grouped individually in the earphone test factor analysis.

Again, similar groupings are seen between the factor analyses from the earphone and loudspeaker tests. Throughout all of the factor analyses, the Dull-Sharp and Low Frequency-High Frequency scales were always in the same grouping. Similarly, the Clean-Rumbling and Smooth-Rough scales were usually grouped together. This indicates the presence of a spectral and a roughness factor. More comparisons of the factor analyses results from both tests are discussed in Section 3.4.

3.3.3.3 Relationship between Metrics and Responses

Average loudness and annoyance ratings were compared to various level metrics and the results are shown in Figures 3.10 and 3.11 and Table 3.13. Level metric calcu-

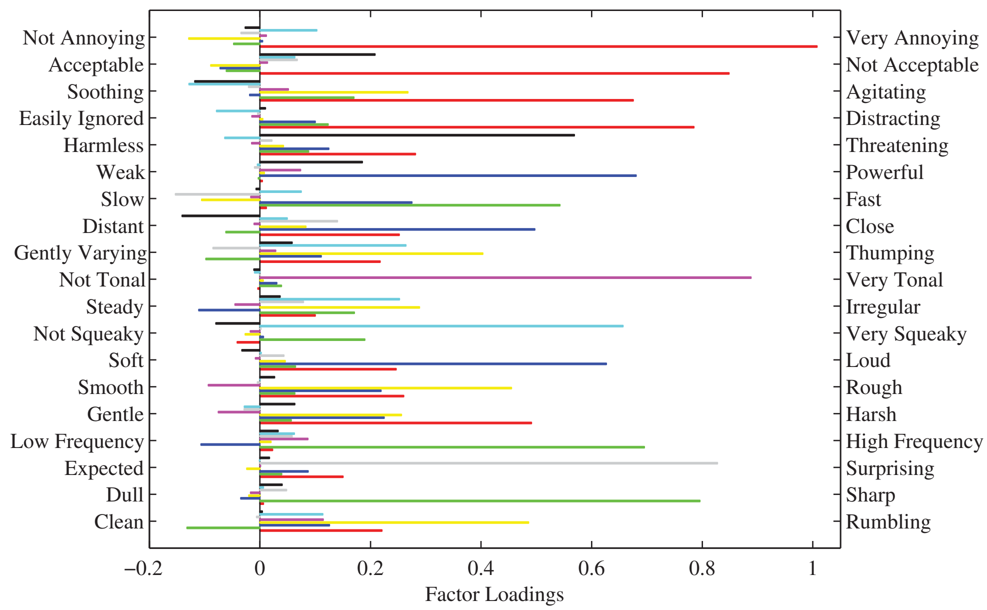

Figure 3.8. Factor loadings from a eight-factor analysis of ratings on all scales in the semantic differential test using loudspeakers. Factors 1 through 8 are coded as follows: red, green, blue, yellow, pink, gray, cyan, and black, respectively.

Table 3.11. Significant factor loadings from a eight-factor analysis of ratings on all scales in the semantic differential test using loudspeakers. Colors from Figure 3.8: Factor 1 - Red, Factor 2 - Green, Factor 3 - Blue, Factor 4 - Yellow. Factor 5 - Pink, Factor 6 - Light gray, Factor 7 - Cyan, Factor 8 - Black.

Factor	Scale	Loading
1	Easily Ignored-Distracting	0.78
1	Soothing-Agitating	0.68
1	Acceptable-Not Acceptable	0.85
1	Not Annoying-Very Annoying	1.01
1	Gentle-Harsh	0.49
2	Dull-Sharp	0.80
2	Low Frequency-High Frequency	0.70
2	Slow-Fast	0.54
3	Soft-Loud	0.63
3	Weak-Powerful	0.68
3	Distant-Close	0.50
4	Clean-Rumbling	0.49
4	Smooth-Rough	0.45
4	Steady-Irregular	0.29
4	Gently Varying-Thumping	0.40
5	Not Tonal-Very Tonal	0.89
6	Expected-Surprising	0.83
7	Not Squeaky-Very Squeaky	0.66
8	Harmless-Threatening	0.57

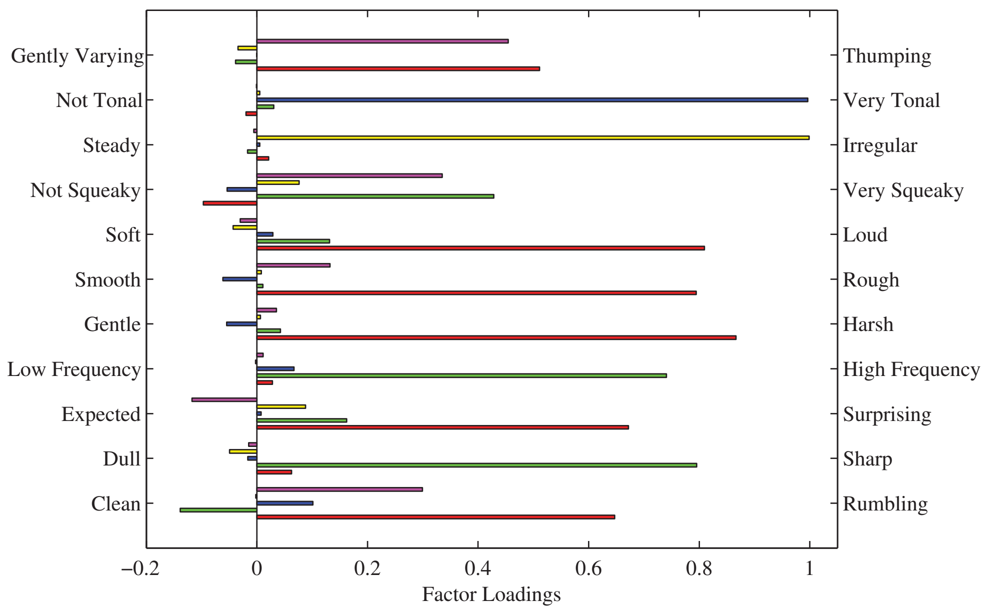

Figure 3.9. Factor loadings from a five-factor analysis of ratings on all scales in the semantic differential test using loudspeakers. Factors 1 through 5 are coded as follows: red, green, blue, yellow, and pink, respectively.

Table 3.12. Significant factor loadings from a five-factor analysis of ratings on all scales in the semantic differential test using loudspeakers. Factor 5 had no loadings over 0.6, and was not the strongest factor for any scale. For this factor, the three highest loadings are listed in place of significant factors. Colors from Figure 3.9: Factor 1 - Red, Factor 2 - Green, Factor 3 - Blue, Factor 4 - Yellow. Factor 5 - Pink.

Factor	Scale	Loading
1	Clean-Rumbling	0.65
1	Expected-Surprising	0.67
1	Gentle-Harsh	0.87
1	Smooth-Rough	0.79
1	Soft-Loud	0.81
1	Gently Varying-Thumping	0.51
2	Dull-Sharp	0.80
2	Low Frequency-High Frequency	0.74
2	Not Squeaky-Very Squeaky	0.43
3	Not Tonal-Very Tonal	1.00
4	Steady-Irregular	1.00
5	Clean-Rumbling	0.30
5	Not Squeaky-Very Squeaky	0.34
5	Gently Varying-Thumping	0.45

lations were performed on source signals at sound levels corresponding to playback levels, i.e., not on recordings in the playback environment. For this test, Perceived Loudness ($PLdB$), Effective Perceived Loudness ($EPNL$) and $SELA$ are the mostly highly correlated with subject responses on the Soft-Loud scale and the Not Annoying-Very Annoying Scale. Coefficient of determination values were recalculated for every possible combination of 18 of the 20 signals in order to understand the variability present in the correlation between metrics and responses. Using subsets of 18 generates a set of 190 R^2 values, and the means and standard deviations of these sets are presented in Table 3.13 in parentheses next to the corresponding values. From this it can be seen that the differences in the R^2 values for several of the metrics are probably not significant.

Table 3.13. Coefficients of determination values for single factor models of individual metrics as predictors of average loudness and annoyance responses in the earphone test. N_{max}, N_5, and N_{10} refer to statistics of Zwicker Time-Varying Loudness. Values in parentheses denote mean and standard deviations of R^2 values calculated as described in Section 3.2.3.3.

Metric	Soft-Loud	Not Annoying-Very Annoying
N_{max}	0.7390 (0.7396, 0.0391)	0.6271 (0.6286, 0.0494)
N_5	0.7748 (0.7754, 0.0332)	0.6608 (0.6622, 0.0433)
N_{10}	0.7891 (0.7907, 0.0422)	0.7198 (0.7227, 0.0524)
$PLdB$	0.8889 (0.8888, 0.0174)	0.8566 (0.8565, 0.0238)
$SELA$	0.8876 (0.8881, 0.0157)	0.8731 (0.8736, 0.0201)
$EPNL$	0.8723 (0.8735, 0.0158)	0.8327 (0.8340, 0.0243)

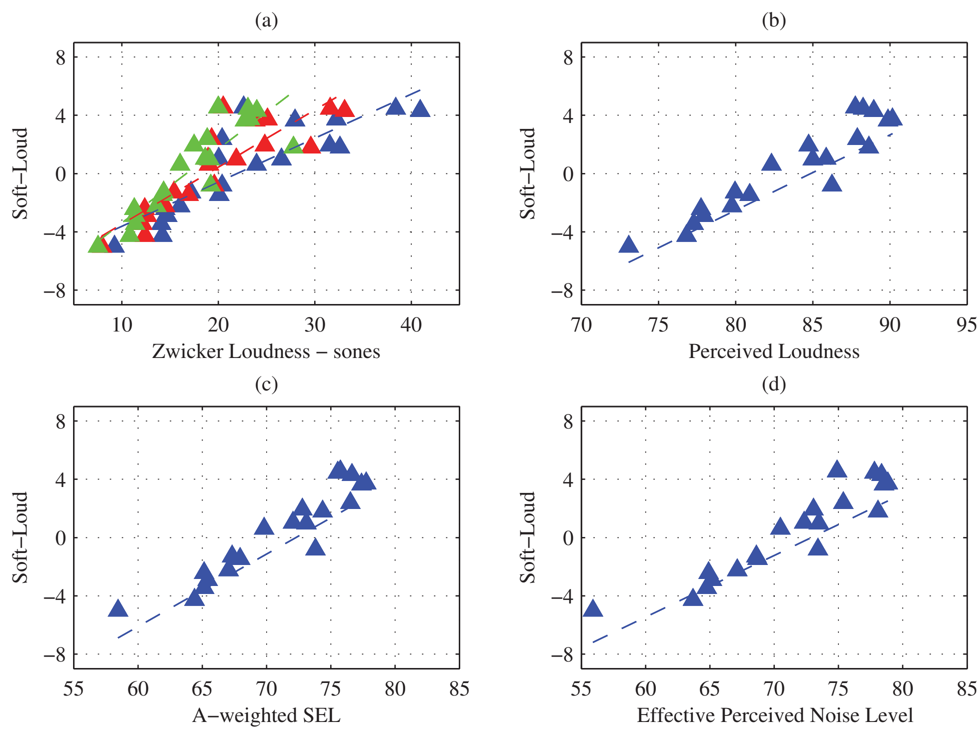

Figure 3.10. Average of the loudness ratings plotted against various metrics: (a) statistics of Zwicker Time-Varying Loudness calculated using HEAD Acoustics Artemis Sound Quality Software (Version 12) (DIN 45631): maximum (blue triangles, $R^2 = 0.7390$), level exceeded 5% of the time (red triangles, $R^2 = 0.7748$), and level exceeded 10% of the time (green, $R^2 = 0.7891$). (b) Perceived Loudness ($PLdB$), $R^2 = 0.8889$; (c) A-weighted Sound Exposure Level ($SELA$), $R^2 = 0.8876$; and (d) Effective Perceived Noise Level ($EPNL$), R^2 =0.8723.

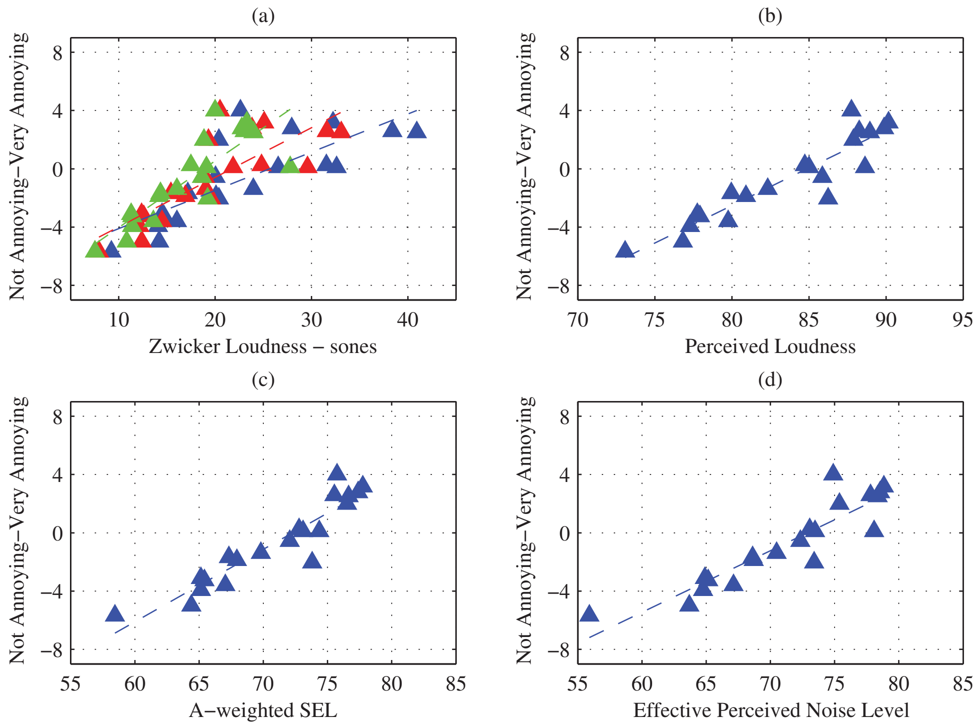

Figure 3.11. Average of the annoyance ratings from Test 2 plotted against various metrics: (a) statistics of Zwicker Time-Varying Loudness calculated by using HEAD Acoustics Artemis Sound Quality Software (Version 12) (DIN 45631): maximum value (blue triangles, $R^2 = 0.6271$), level exceeded 5% of the time (red triangles, $R^2 = 0.6608$), and level exceeded 10% of the time (green triangles, $R^2 = 0.7198$). (b) Perceived Loudness ($PLdB$), $R^2 = 0.8566$; (c) A-weighted Sound Exposure Level ($SELA$), $R^2 = 0.8731$; and (d) Effective Perceived Noise Level ($EPNL$), $R^2 = 0.8327$.

3.4 Comparison of Results from the Two Tests

While the two tests performed were intended to be as similar as possible, it is important to recognize the differences between them when attempting to compare results. The same semantic differential scales were used in both tests, but the presentation and data entry method were different. For the earphone test, subjects made marks with a pencil on paper, and for the loudspeaker test subjects moved markings on scales on a tablet PC. Examples of completed evaluation sheets for the earphone test and the loudspeaker test are included in Appendices A.21 and A.22, respectively. Subjects for each test also came from different backgrounds. For the earphone test, subjects were recruited from the Purdue University campus, whereas for the loudspeaker test subjects were recruited from around the NASA Langley Research Center and surrounding area. Langley Research Center borders Langley Air Force Base and military aircraft noise is often heard when outdoors in the area. Purdue University also has a nearby airport, but the aircraft flying in the area are very different in size and noise signature, mostly consisting of single- and twin-propeller planes.

However, the main difference between the tests was due to the playback method and playback level. The earphone test was conducted one subject at a time in a small room (roughly an 8 ft cube) and the loudspeaker test was performed in the EER (a 39-seat auditorium) with four subjects at a time. In the earphone test the same signal was presented through both earphones, while in the loudspeaker test Vector-Base Amplitude Panning was used to simulate flyovers based on actual flight paths. In the earphone test signals were played at a maximum level of 90 dBA, while in the loudspeaker test signals were limited to a maximum level of 78 dBA. Due to these differences, a comparison of test results is limited to observation, and no strong conclusions about the effect of playback method on subject response can be drawn. However, similar trends are present in the analysis of the data from the two tests, i.e. in the scale ratings to scale ratings correlation, metrics analysis, and factor analysis. These trends can help guide future subjective testing by providing starting points for determining and characterizing factors present in people's responses to rotorcraft noise.

It should be noted that results presented are a function of the signal set used. People may notice multiple characteristics but if the characteristics follow the same type of variation from signal to signal, the factor analysis would reveal only one independent factor. While a wide range of rotorcraft sounds were presented, no attempt to vary specific characteristics independently was made. This limitation should be kept in mind when examining the results of the analysis.

To provide a direct visual comparison of subject responses on all the scales, plots of average ratings for each signal from both tests were generated. Examples are shown in Figures A.13 and A.11 for signals 13 and 11, respectively, which illustrate a case where the average ratings in both tests are close and a case where the ratings are further apart, respectively. Recall that all of the sounds were played back at a lower overall sound pressure level in the EER with some signals additionally attenuated to address some amplifier saturation issues. Signal 11 was one of the quieter sounds and not one that was further attenuated. The full set of responses on the scales to each of the 20 signals in both tests is shown in Appendix A.

Similar trends are seen between the standard deviations for scale responses in both tests. The average and range of standard deviations for each scale did not vary a large amount between the two tests, with the largest difference in averages being 0.59 (scales were scored from -9 to 9) for the Clean-Rumbling scale. These small differences show that the ability to rate signals on each scale was mostly maintained between the two tests. However, the scale correlations with annoyance were different between the two tests. Four of the 18 scale correlations ρ with annoyance changed by greater than 0.15 between the two tests. By subtracting the annoyance correlations for the earphone test from the correlations for the loudspeaker test, it can be seen that the differences are largest for the Not Tonal-Very Tonal (0.26), Steady-Irregular (0.21), Low Frequency-High Frequency (0.19), and Slow-Fast (-0.17) scales. Positive numbers mean that the correlation is stronger for the loudspeaker test. At first, the

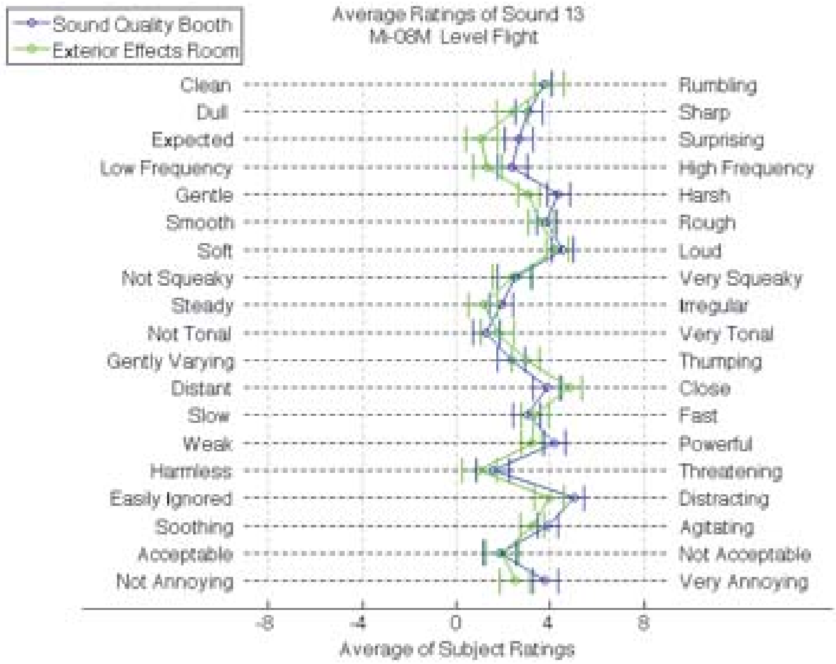

Figure 3.12. Average scale ratings (+/- SEM) for Sound 13 from both the earphone test (blue) and the loudspeaker test (green).

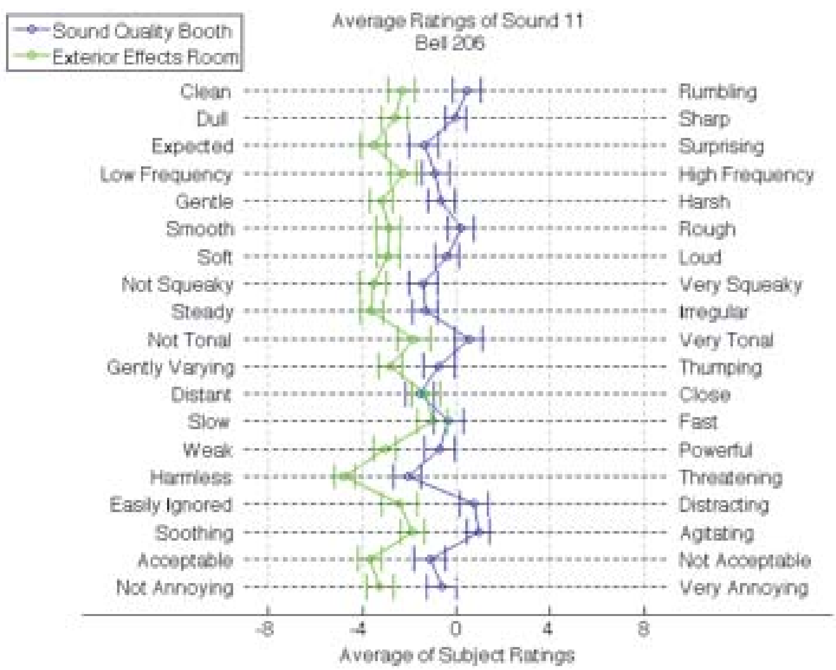

Figure 3.13. Average scale ratings (+/- SEM) for Sound 11 from both the earphone test (blue) and the loudspeaker test (green).

lower correlation between Slow-Fast ratings and annoyance ratings in the loudspeaker test seemed surprising. However, "fast" sounds in the EER may not be found by the loudspeaker test subjects to be more annoying because the presentation is more natural and it is a noise source that this group is somewhat used to, so results are perhaps to be expected.

Mean scale ratings from both tests, for selected scales, are plotted against one another in Figure 3.14. Points are color-coded by vehicle, and range from light to dark based on craft speed. The amplified distant signals are grouped separately. Bell 206 - green (2 speeds), BO105 - cyan (2 speeds), Mi-8M - grey (4 speeds), MD902 - magenta (2 speeds), MD520N - yellow (2 speeds), XV-15 - blue (2 speeds), amplified distant signals - red (3 speeds). For many of the scales, ratings of signals were lower in the loudspeaker test than the earphone test when averaged across all subjects, but followed a linear relationship. The exception is the Not Tonal - Very Tonal scale ratings (Figure 3.14 (d)), which do not follow a strong linear trend. As noted earlier, several subjects had difficulty understanding the meaning of the words on this scale. Certain signals stand out from the linear trend, such as Signal 1 (brightest red dot). This signal was one of the four generated by amplifying a distant section of a flyover, and appears to be an outlier in plots (a), (b), (e), and (f) of Figure 3.14. In the playback of the amplified distant signals (1, 2, 12, and 18) in the loudspeaker test, their original distant flight path was used, so the sound remained in front of the listeners, and did not travel overhead as in the playback of the other signals. This may explain some of the difference in responses for these signals between the two tests.

Predictions of loudness and annoyance based on metrics varied between the two sets of test results. $EPNL$ has the highest R^2 value for both subjective loudness and annoyance ratings in both tests while for other metrics the R^2 values are as low as 0.6271 (Zwicker Loudness - N_{max} as a predictor of annoyance - $R^2 = 0.7770$ in the earphone test, $R^2 = 0.6271$ in the loudspeaker test).

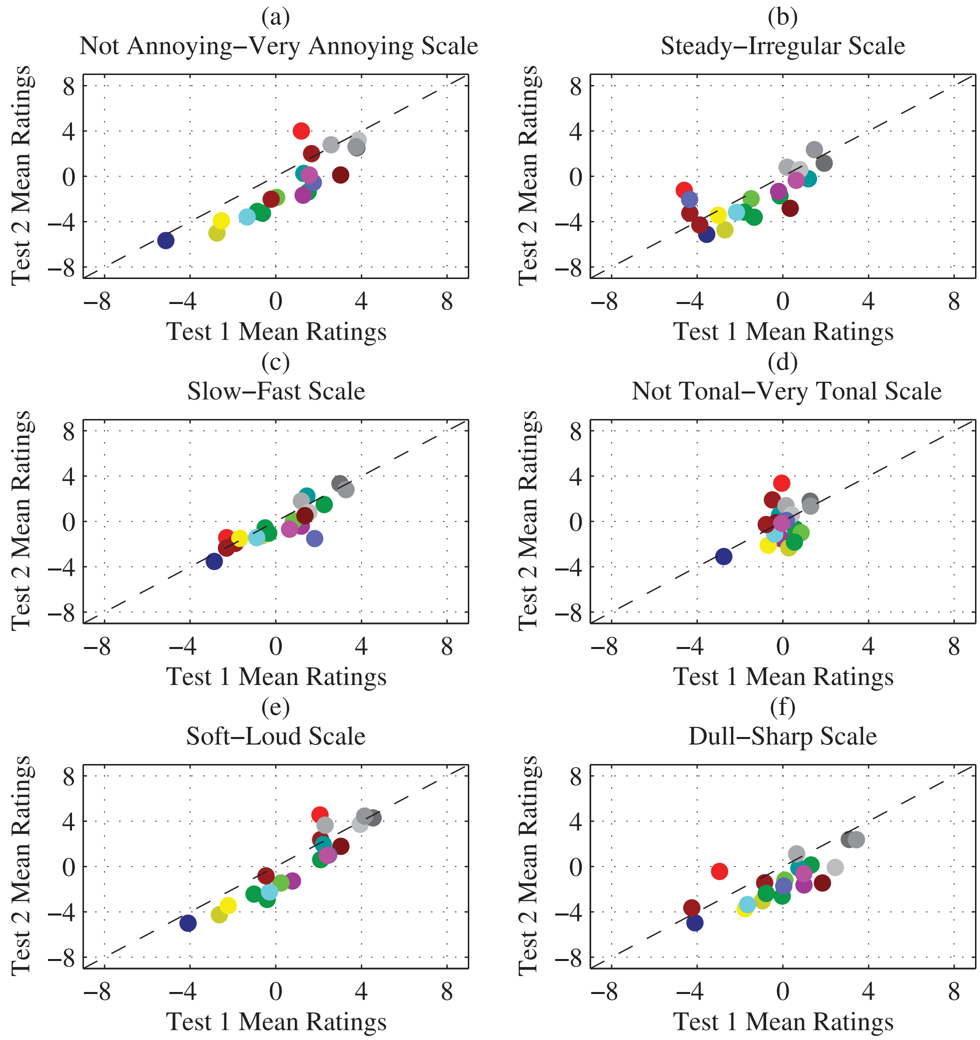

Figure 3.14. Mean scale ratings from the earphone test (Test 1) and loudspeaker test (Test 2) plotted against each other for a few selected scales: (a) Not Annoying - Very Annoying; (b) Steady - Irregular; (c) Slow - Fast; (d) Not Tonal - Very Tonal; (e) Soft - Loud; and (f) Dull - Sharp.

The factor analyses performed on the results from the two tests yielded many similar results. Based on the groupings shown in Tables 3.6, 3.7, 3.11, and 3.12, there are a few strong common factors present in these data sets. In all factor analyses, the Dull-Sharp and Low Frequency-High Frequency scales, as well as the Clean-Rumbling and Smooth-Rough scales, were usually paired together. These show the presence of a spectral balance factor and a roughness factor. The Not Tonal-Very Tonal scale was grouped separately in 3 out of the 4 analyses, twice with a factor loading of 0.89 or greater. This shows the presence of a tonality factor. The Steady-Irregular scale had the highest loading in its grouping twice, and twice was paired (earphone test) with the Not Squeaky-Very Squeaky scale. These scales show the presence of an irregularity factor in the data. The Soft-Loud scale was often grouped with scales relating to impressions of the sound. Loudness is most likely its own separate factor, as it is known that loudness plays a large role in annoyance [2].

3.5 Summary of Findings

A set of two semantic differential tests were designed and performed as a first step in the characterization of rotorcraft noise. A group of 20 rotorcraft flyover recordings varying in a wide range of attributes were evaluated by over 70 subjects across the two tests. Subjects responses, as well as their relationship with metrics, were analyzed in order to identify important factors present in human response to rotorcraft noise. It was determined that while some commonly used metrics were correlated with subjects ratings of annoyance, multiple factors were present in the response data obtained. Factors relating to loudness, irregularity, roughness, spectral balance, and tonality were observed and may have an impact on judgments of annoyance. It is important to note that the results of these analysis of this test are signal set dependent, and only recordings were used in this test (no simulated signals). Signal characteristics were not manipulated to avoid co-variation of characteristics, which is important for understanding the role of each characteristic in annoyance and the development of an annoyance model.

4. ANNOYANCE TESTS

Described in this chapter are the test setup, procedure, and results for the Annoyance Test performed: (1) using earphones for playback and (2) using loudspeakers for playback. The goals of this test are: to gather data to examine how tone-family characteristics affect annoyance, and to determine which metrics or combinations of metrics produce the best predictions of the average of subjects ratings. Additionally, the combined effect of different test environments and different subject pools is of interest and so the responses from the two tests were compared.

4.1 Test Sounds

Seven base signals were chosen from a large collection of rotorcraft flyover recordings of different crafts operating at different speeds and elevations. The details of the set of base signals are described below in Table 4.1. The vehicle characteristics (blade passage frequencies and maximum operating speed) are listed in Table 2.1. This collection includes recordings from acoustic flight tests at Eglin Air Force Base [7, 9] and an acoustic flight test of the XV-15 tiltrotor aircraft held in Waxahachie, Texas [8]. The base signals were chosen so that as many different crafts and operating conditions could be included. For six of the seven base signals, fifteen-second segments of the original recordings were extracted, centered on the point in time at which the vehicle was directly above the microphone. The remaining base signal was generated by extracting a segment in the recording where the vehicle was far from the microphone (denoted by * in Table 4.1). This signal was amplified through time such that the resulting signal had a predicted time-varying Zwicker loudness profile similar to that of the corresponding overhead recording of the same flight, but the tonal components did not undergo the corresponding Doppler shift. All of the base signals were then amplified through time to account for altitude and speed changes between the flight tests, as an attempt to avoid response bias due to altitude and flight speed changes during the recording, while still maintaining the overall differences between sounds from different vehicles.

Table 4.1. Details of the seven base signals used to generate the annoyance test stimuli. The * denotes a distant amplified signal.

Vehicle	Operation	BPFs (Hz)
Bell 206	Level Flight	13, 85
BO105	Level Flight	28, 74
MD520N	Level Flight	40, 1167
MD902	Approach	32.7, 1100
XV-15	Nacelle Angle 80°	32
XV-15	Nacelle Angle 0°	32
XV-15*	Nacelle Angle 60°	32

4.1.1 Recordings and Simulations

For each of the seven base signals, a set of five simulations was generated for use in the parametric annoyance test, including (1) the original recording, (2) lowered tone magnitudes (50% of original), (3) raised tone magnitudes (150% of original), (4) raised tone magnitudes (150% of original) and raised impulsiveness (phase aligned), and (5) original signal with lowered impulsiveness (harmonic families of randomized phase). Additionally, two XV-15 flyovers were simulated with lowered fundamental frequencies, at one-half and one-third of the original (32 Hz). This common set was high pass filtered using a 4th order Butterworth high pass filter with a cutoff frequency of 25 Hz to account for the common reproduction capabilities of both test environments. The details of the common set are shown in Tables B.1 - B.3. When the

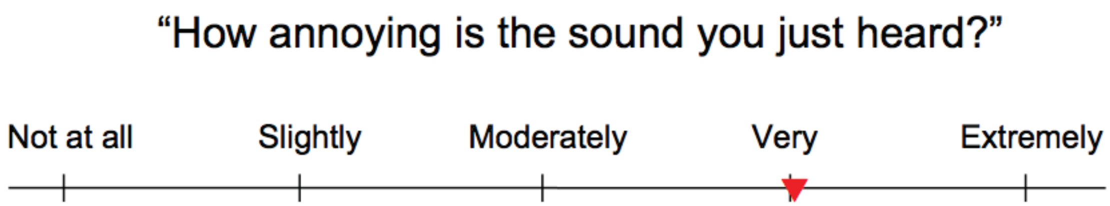

Figure 4.1. Example of a scale similar to the scales used by subjects evaluate the sounds in the annoyance test at both locations.

test was performed in each location the common set of signals was used. However, at each location, an additional different set of 55 signals was included in the test. Those additional signal sets are described below in the corresponding test specific sections. The scales subjects used to rate signals in each test was similar to that shown in Figure 4.1.

4.2 Baseline Test Setup and Procedures for Both Tests

After a brief introduction and description of the test, subjects completed a pre-approved consent form and questionnaire (Purdue IRB approval # 1209012637 and NASA LaRC IRB MPA Code NASA3082281305HR), and had their hearing tested to determine if they met the criteria to participate in the test using the same criteria as the Semantic Differential Test (see sections 3.2.2 and 3.3.2 for criteria). The subject was then informed of the detailed test procedures. Then the subject listened to 8 sounds to become familiar with the types of sounds that they would be hearing. Next, they completed a practice test, where they evaluated 6 sounds by making a mark on a parametric annoyance scale on a computer or tablet PC. Then in the main test subjects evaluated 110 sounds by making marks on the parametric annoyance scale as they did in the practice test. The subjects were instructed to think of themselves hearing these sounds several times throughout the day while they were outside, around their house or in their community. The sounds were played in a different random order for each subject. On completion of the test, the subjects hearing was tested, and the subjects were compensated for participation in the test.

4.3 Annoyance Test with Earphone Playback (Purdue Test)

This test was performed at Purdue University, at the Ray W. Herrick Laboratories in a quiet room. The specific test setup, subjects, procedure, and results are given in this section.

4.3.1 Specific Test Setup and Additional Stimuli

In this test, signals were played back through a high quality LynxOne sound card, a Tucker-Davis HB7 amplifier and Etymotic Research ER-2 tube earphones. Testing was performed in a double walled IAC sound booth. The sound card and the earphones were chosen because of the flat frequency response and low noise floor. The disposable foam eartips used on the earphones (ER-14A) add an additional 25-30 dB of attenuation of the background noise. Prior to playback, the signals were filtered to account for the change from an outdoor recording (free field) to presentation at the eardrum (ER-2 earphones present sound directly to the eardrum).

For this test, in addition to the common set of signals, two sets of eleven loudness-normalized signals were generated, and 33 signals were repeated to bring the total number of signals to 110. The loudness normalization was based on Zwicker's model of time-varying loudness exceeded 5% of the time (N_5), where each set of signals had the same N_5 value. This allowed for the observation of changes in annoyance between a set of sounds with the same loudness, as well as a study of repeatability in subjects responses. The details of this signal set are given in Table B.8 - B.10.

4.3.2 Subjects

A total of 40 subjects participated in this test. Subjects were students and staff at Purdue University. The subjects ranged in age from 18-58 years, 19 were male and 21 were female. All of the subjects passed a hearing screening requiring no more than 20 dB of hearing loss in either ear in a range of frequencies from 125 to 8000 Hz. Subjects were compensated $10 for their participation in the test.

4.3.3 Results

Numbers were assigned based on the position of the mark the subject made on the parametric scale, ranging from 1 (left end of scale) to 9 (right end of scale). Numbers corresponding to the five verbal scale markings shown in Figure 4.1 are 2, 3.5, 5, 6.5, and 8, respectively. The average of the subjects responses and the standard deviation of the estimated mean (SEM) were calculated for each signal. The SEM for all signals was lower than 0.26, which corresponds to 3.3% of the entire scale, or 17% of the distance between two words on the scale. In the following subsections the test results and their relationship with vehicles, modifications, and metrics are described.

4.3.3.1 Results by Vehicle and Modification

In Figure 4.2 the average annoyance ratings sorted by vehicle are shown with different symbols for each modification type. Two trends are seen within the rating of the XV-15 simulations. Annoyance ratings tended decrease with lowered fundamental frequencies (Heli /2 and Heli /3 tended to be rated as being of lower annoyance signal sets than signal set Heli, and similarly Plane /2 and Plane /3 to Plane). The XV-15 in “airplane mode” (Plane sets) was rated less annoying than in “helicopter mode” (Heli sets). For the XV-15 simulations with fundamental frequencies at one-half (16 Hz) and one-third (10.7 Hz) of the original, the first harmonic component in the series has shifted to a frequency region that is outside the sound reproduction capabilities

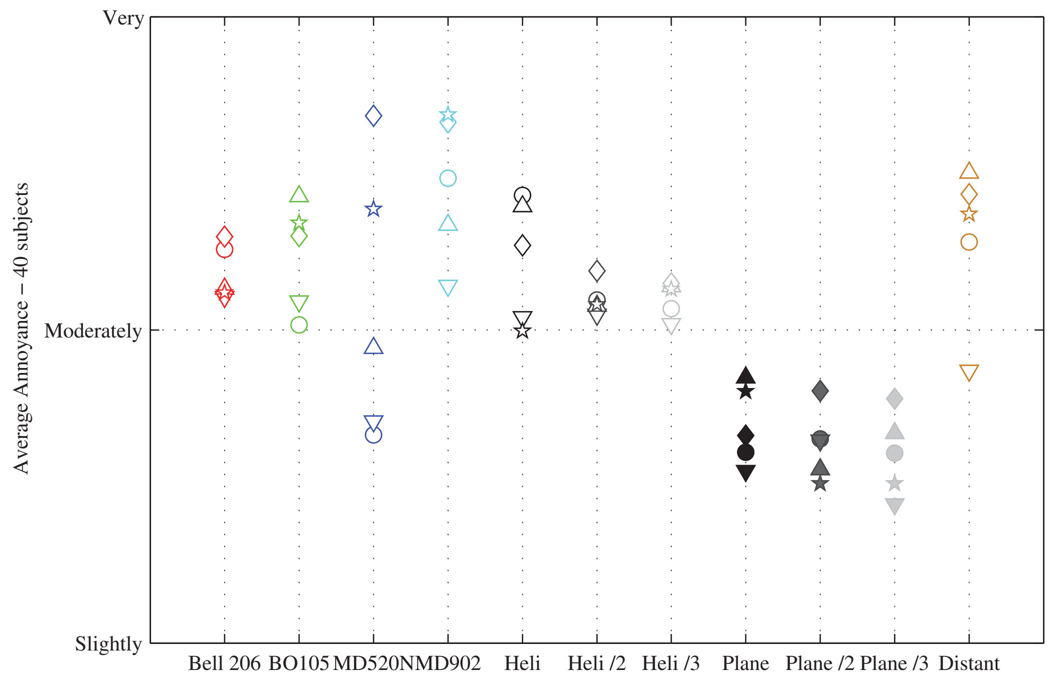

Figure 4.2. Average annoyance ratings plotted against vehicle type. Original signal (circle), lowered tones (downward triangle), raised tones (upward triangle), raised tones and phase aligned (diamond), randomized phase (star). The changes to the XV-15 fundamental frequency are treated as their own vehicle and /2 and /3 correspond to one-half and one-third the original fundamental frequency, respectively. The labels Heli and Plane correspond to the helicopter mode (Nacelle Angle 80°) and airplane mode (Nacelle Angle 0°) of the XV-15.

of both playback systems used although the other harmonics in the series are above the low frequency reproduction limits. Also, sound energy has been shifted to lower frequencies where the human hearing system is less sensitive. One might expect that both of these would lead to a decrease in annoyance. At the same time, by lowering the fundamental frequencies, the period of the temporal variation has shifted towards a region where people are more sensitive to fluctuations [2], but that would be expected to increase annoyance. On average, subjects clearly showed an increase in annoyance with increased tonal levels, as most downward triangles are near the bottom of the range of responses, and most upward triangles are near the top of the rating range for each vehicle. Differences between the simulations with raised tones and those with raised tones and impulsiveness are present in some vehicle groups, but not all. This may be due to differences in the impulsiveness of the original signals, as some of the original signals already sounded highly impulsive.

4.3.3.2 Relationship between Metrics and Responses

Level-based metrics were calculated for each input signal: Effective Perceived Noise Level ($EPNL$), A-weighted Sound Exposure Level ($SELA$), Perceived Loudness ($PLdB$), and statistics of Zwicker (DIN 45631), and Moore and Glasberg Time-Varying Loudness such as level exceeded 5% (N_5) or 10% (N_{10}) of the duration of the signal. $EPNL$ and $SELA$ were calculated using ANOPP2 Acoustic Analysis API [32]. Zwicker Loudness was calculated by using HEAD Acoustics Artemis Sound Quality Software (Version 12) (DIN 45631), while Moore and Glasberg Time-Varying Loudness [36] was calculated using software developed by a previous graduate student at Purdue University [40], that incorporates relevant revisions in ANSI S3.4-2007 [41].

The average annoyance ratings were compared to the various metrics values for each signal and the results are shown in Figure 4.3. The points that are the largest outliers in (b), (c), and (d) are mainly from the MD902 (cyan). For these signals, $EPNL$ was the most highly correlated to the average of the subjects responses (R^2 = 0.637). Correlation with loudness (N_5) was much lower than expected based on previous tests. Correlations with all percentiles of loudness were investigated, finding the highest for this data set to be with N_{45}, with an R^2 value of 0.546. In Figure 4.4 (a) is shown the R^2 values as a function of percentile exceeded. Annoyance responses plotted against N_{45} are shown in 4.3 (b), and $EPNL$ values converted to sones plotted against N_{45} in (c).

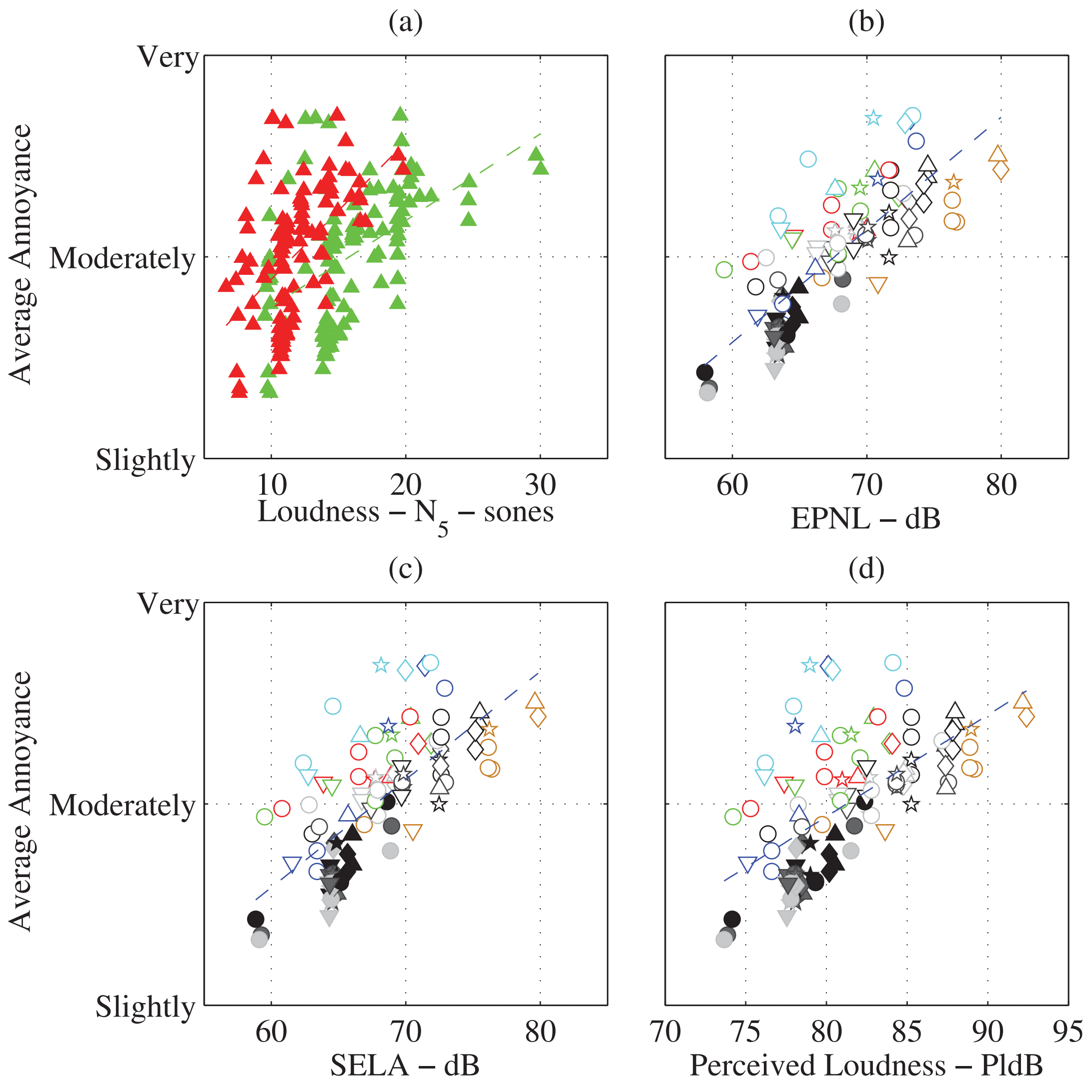

Figure 4.3. Average annoyance ratings plotted against various metrics for the earphone test. Plotted against (a) Zwicker Loudness: level exceeded 5% of the time (green triangles, $R^2 = 0.276$) and Moore and Glasberg Short-Term Loudness: level exceeded 5% of the time (red triangles, $R^2 = 0.263$); (b) Effective Perceived Noise Level, $R^2 = 0.637$; (c) A-weighted Sound Exposure Level, $R^2 = 0.504$; and (d) Perceived Loudness, $R^2 = 0.400$.

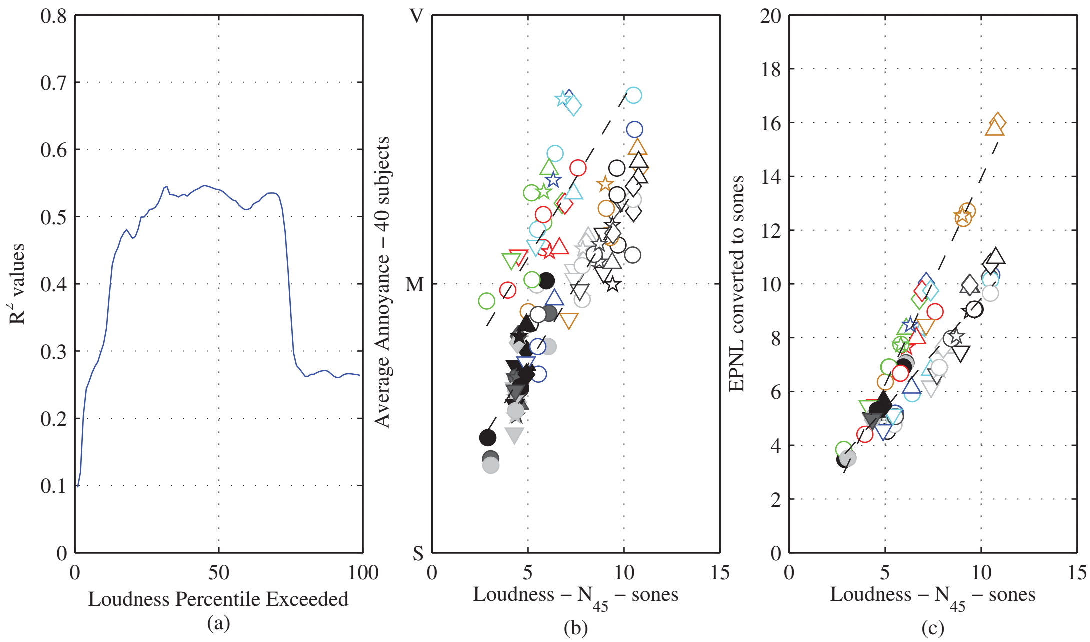

Figure 4.4. Evaluation of the annoyance prediction performance of level-based metrics for the earphone test. (a) Coefficient of determination values (R^2) as a function of percentile of loudness exceeded; (b) Average annoyance ratings plotted against Moore and Glasberg Short-Term Loudness: level exceeded 45% of the time ($R^2 = 0.546$); (c) $EPNL$ values converted to sones plotted against Moore and Glasberg Short-Term Loudness: Level exceeded 45% of the time ($R^2 = 0.750$).

Figure 4.4 (b) is used to illustrate differences in the annoyance predictions produced by the N_{45} metric, and (c) is used to understand differences between N_{45} and $EPNL$. Plot (b) shows a separation between the trend for the XV-15 signals (black and gray) and the trend for traditional helicopters (Bell 206 - Red and BO105 - Green). This trend separation is less obvious in $EPNL$ predictions , hence in plot (c) there are still two trend lines, although the distant amplified signals (brown) are now more clearly aligned with the trend for traditional helicopters rather than the trend for the XV-15 aircraft. So there appears to be attributes of sounds that lead to increases in annoyance not captured by loudness alone and the tone and event corrections improve the $EPNL$ performance. The loudness model used in $EPNL$ is an earlier loudness model.

Other psychoacoustic-based sound quality metrics, including Sharpness, Roughness, Tonality, and Fluctuation Strength, were calculated for these signals. The left plot in Figures 4.5 - 4.8 shows average annoyance ratings plotted against these metrics. A linear model from N_{45} was fit to the data, and the residuals of the that model are compared to each of the metrics. These results are shown in the right plot in Figures 4.5 - 4.8.

From these results it is not clear how to include these metrics in an annoyance model. There appear to be trends with some aircraft types (slight negative trend with sharpness for XV-15). Signals with very low sharpness levels can be annoying because they can sound "heavy". The sharpness metric was not designed to capture this characteristic. To capture this researchers have used the difference between C-weighted and A-weighted Sound Pressure Level [42]. There appears to be an opposite weak trend for the distant signals (brown) - an increase in annoyance with sharpness.

There is a weak trend with Roughness once loudness effects have been removed. Similarly with tonality, most noticeable for the XV-15 sounds. Some of the tone-family changes also caused changes in loudness (increased/decreased levels) making it difficult to differentiate between loudness and tonality as drivers for response changes. Note the tonality metric produced 0 values for the signals with very low fundamental frequencies. This is because of the fast roll-off of tonality perception at lower frequencies that is captured in Aures' model. The fluctuation strength metric has a weak trend with annoyance except for two distant signals (brown) with larger fluctuation strength values (3.2 and 3.7) that are not shown in Figure 4.8.

From these results it is clear that if we wish to understand the impact of individual sound characteristics (and the corresponding metric measurements of their strength) on annoyance, they will need to be varied individually while keeping other character-

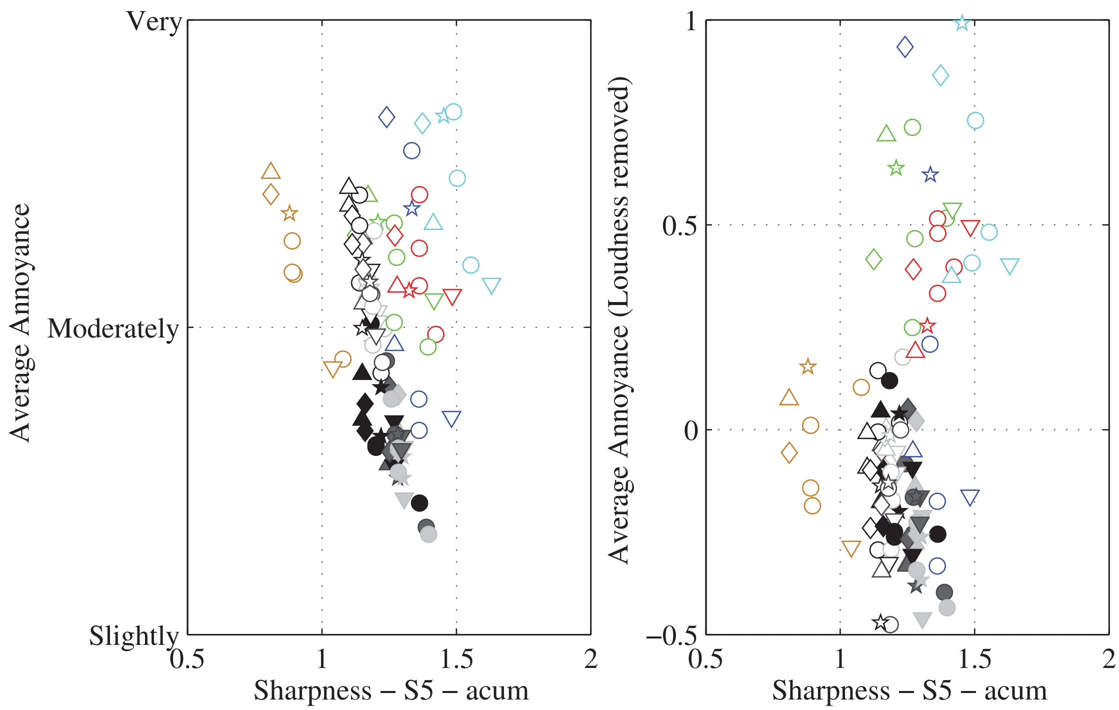

Figure 4.5. (a) Average annoyance ratings from the earphone test plotted against Sharpness exceeded 5% of the time. (b) Average "loudness removed" annoyance plotted against Sharpness exceeded 5% of the time. See Figure 4.2 for color and symbol coding.

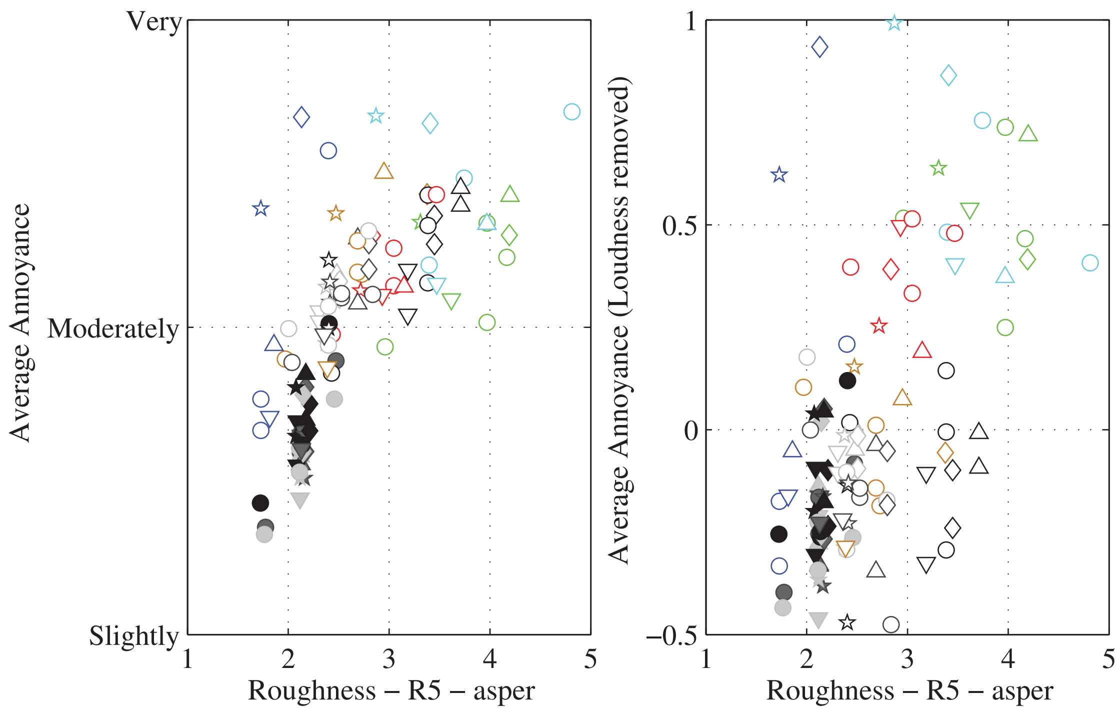

Figure 4.6. (a) Average annoyance ratings from the earphone test plotted against Roughness exceeded 5% of the time. (b) Average "loudness removed" annoyance plotted against Roughness exceeded 5% of the time. See Figure 4.2 for color and symbol coding.

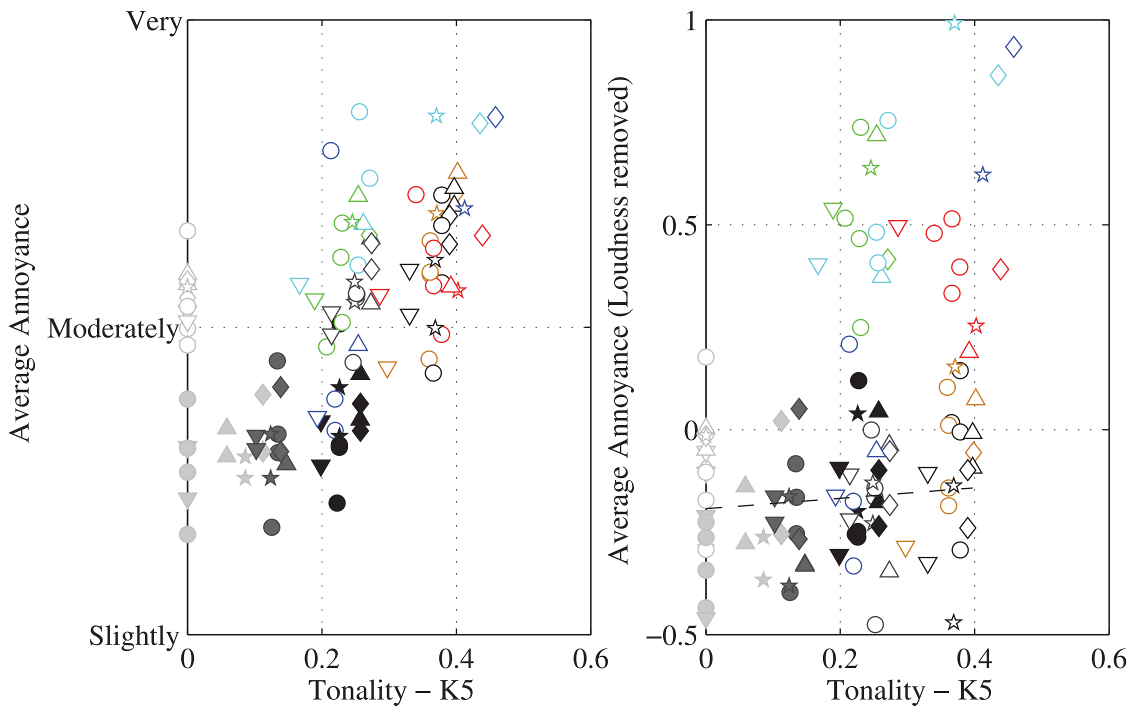

Figure 4.7. (a) Average annoyance ratings from the earphone test plotted against Aures Tonality exceeded 5% of the time. (b) Average "loudness removed" annoyance plotted against Aures Tonality exceeded 5% of the time. See Figure 4.2 for color and symbol coding.

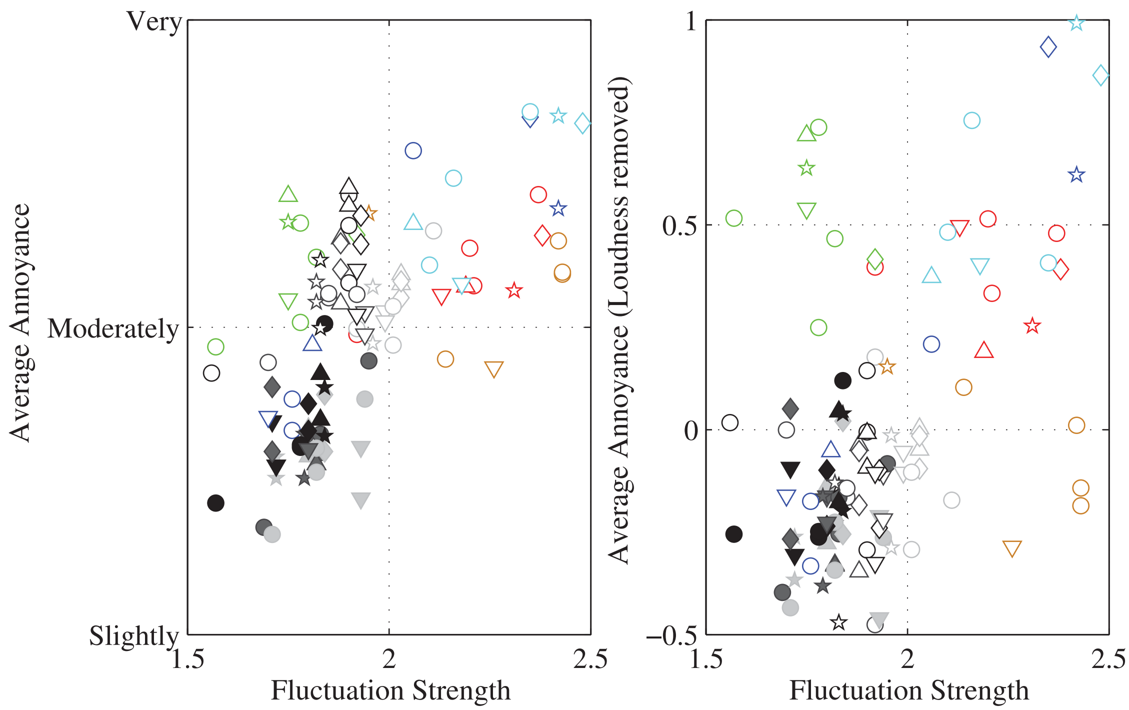

Figure 4.8. (a) Average annoyance ratings from the earphone test plotted against Fluctuation Strength. (b) Average "loudness removed" annoyance plotted against Fluctuation Strength. See Figure 4.2 for color and symbol coding.

istics constant. This is challenging with sounds that are combinations of tones and noise. For example, increasing tonality while keeping loudness constant creates a less rough sound because the noise component has to be reduced to keep loudness the same.

In addition, an impulsiveness metric was not found that captured the increased impulsiveness perceived with some sounds when the tonal components were phase aligned. This also requires further investigation.

4.3.3.3 Repeatability and Loudness Normalization

In Figure 4.9 are shown the average of the subject responses to signals plotted against subjects responses to the repeated versions of those signals. The error bars for each point correspond to the standard deviation of the estimated mean. The correlation between the responses to repeated signals ($\rho = 0.81$) is an indication consistency in subjects responses.

The set of seven base signals (as described in Table 4.1), along with the two XV-15 simulations at one-half and one-third the original blade passage frequency, were used to create two loudness-normalized sets of eleven signals. The first set was normalized to the lowest N_5 of the group (9.8 sones), and the second set was normalized to the highest N_5 of the group (19.5 sones). Figures 4.10 and 4.11 show the average annoyance ratings for these signals compared to various metrics. It is clear from Figure 4.10 that normalizing N_5 values is not necessarily the same as normalizing other level metrics.

In Figure 4.11 are average annoyance ratings for loudness normalized signals plotted against various sound quality metrics. Plots (b) and (d) show that when N_5 is held constant there are strong linear strends with annoyance for both Roughness and Fluctuation Strength. No significant trends are seen for this signal set between Sharpness or Tonality with annoyance ratings.

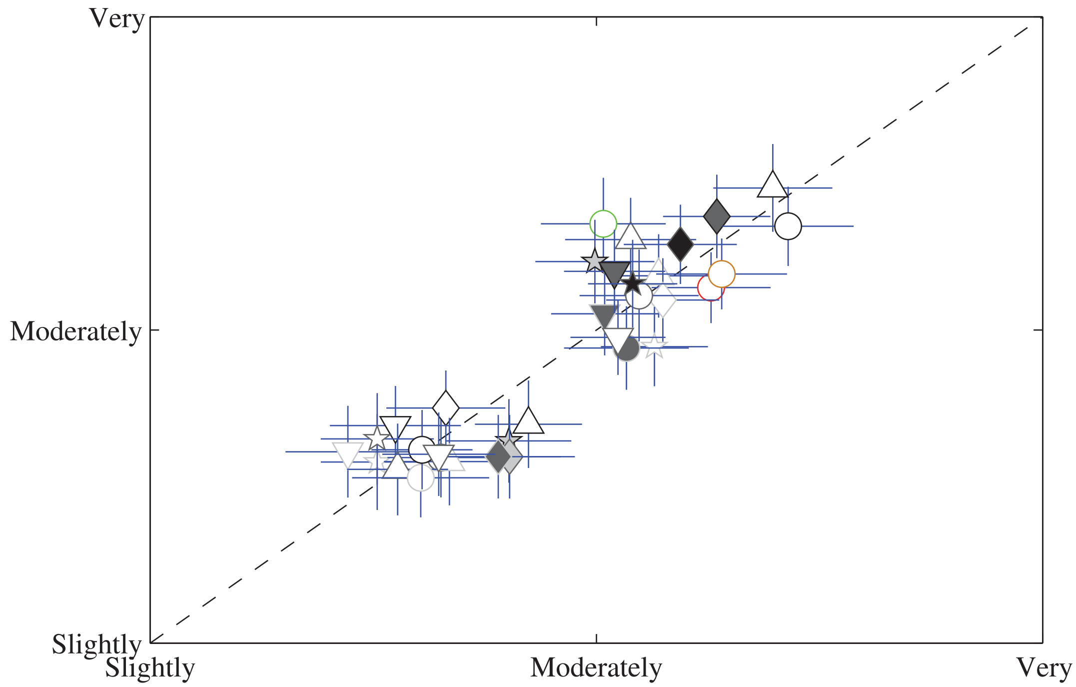

Figure 4.9. Average annoyance responses for repeated signals in the earphone test. Error bars correspond to standard deviations of the estimated means.

In Figure 4.12 sound quality metric values for the loudness normalized signals are plotted against one another. Plot (a) shows a slight trend between Roughness and Fluctuation Strength values, while plots (b) and (c) show little to no trends between other pairs of metrics. The calculations for Roughness and Fluctuation Strength are similar, where the main difference is that Roughness is focused on faster fluctuations in loudness (largest values are for sounds with loudness fluctuations of 60 to 70 cycles per second) and Fluctuation Strength is focused on slower fluctuations in loudness (largest values are for sounds with loudness fluctuations of around 4 cycles per second).

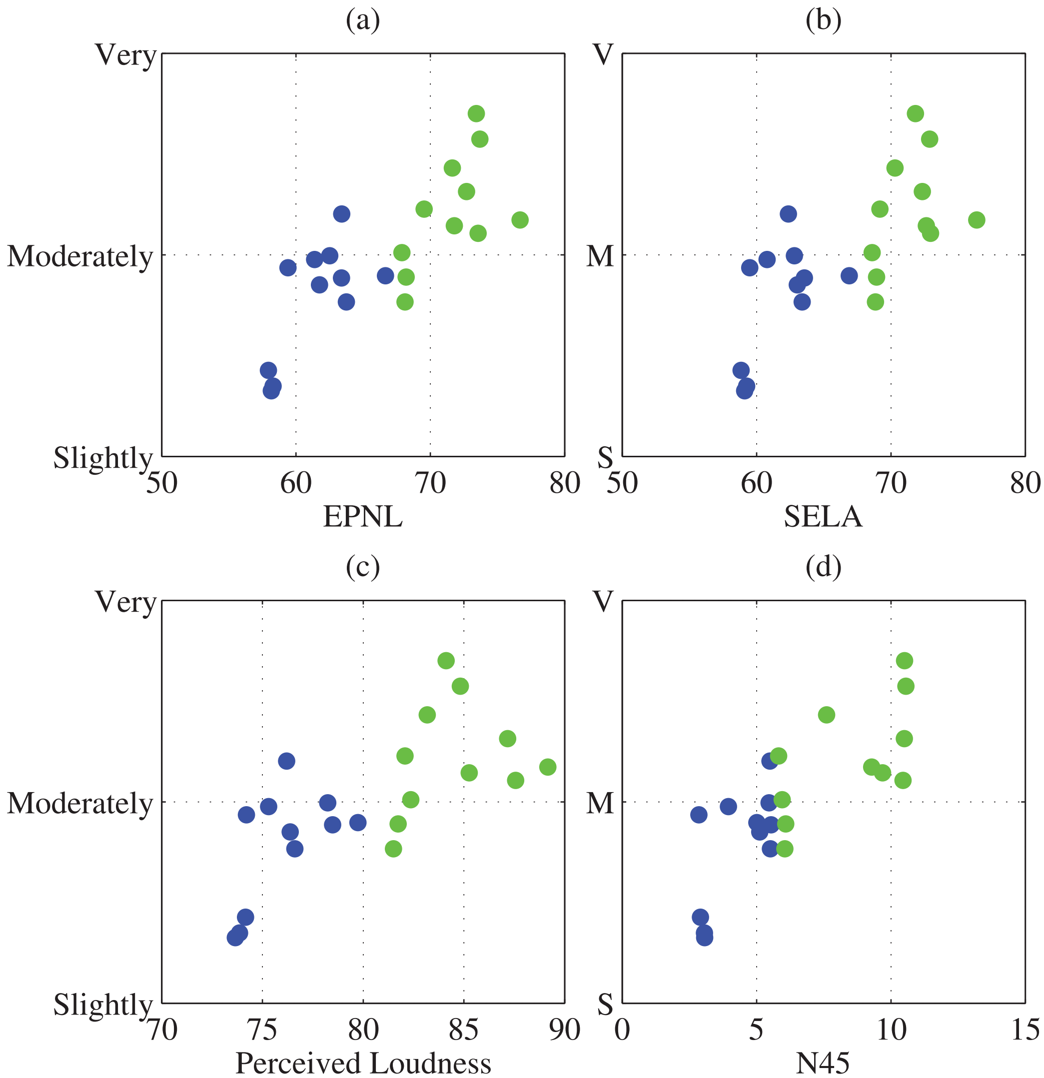

Figure 4.10. Average annoyance responses plotted against various metrics for the loudness (N_5) normalized signals sets from the earphone test: low loudness normalization (blue) and high loudness normalization (green).

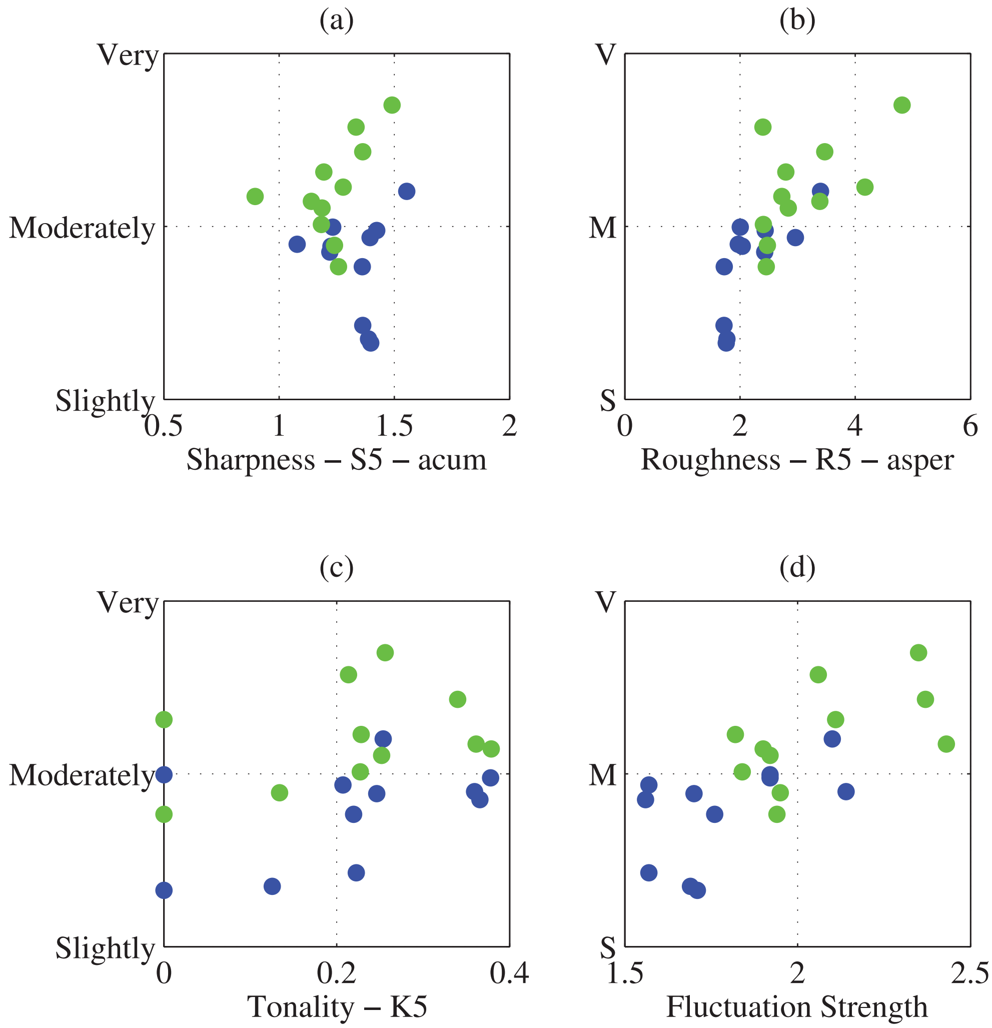

Figure 4.11. Average annoyance responses plotted against various sound quality metrics for loudness normalized signals sets from the earphone test: low loudness normalization (blue) and high loudness normalization (green).

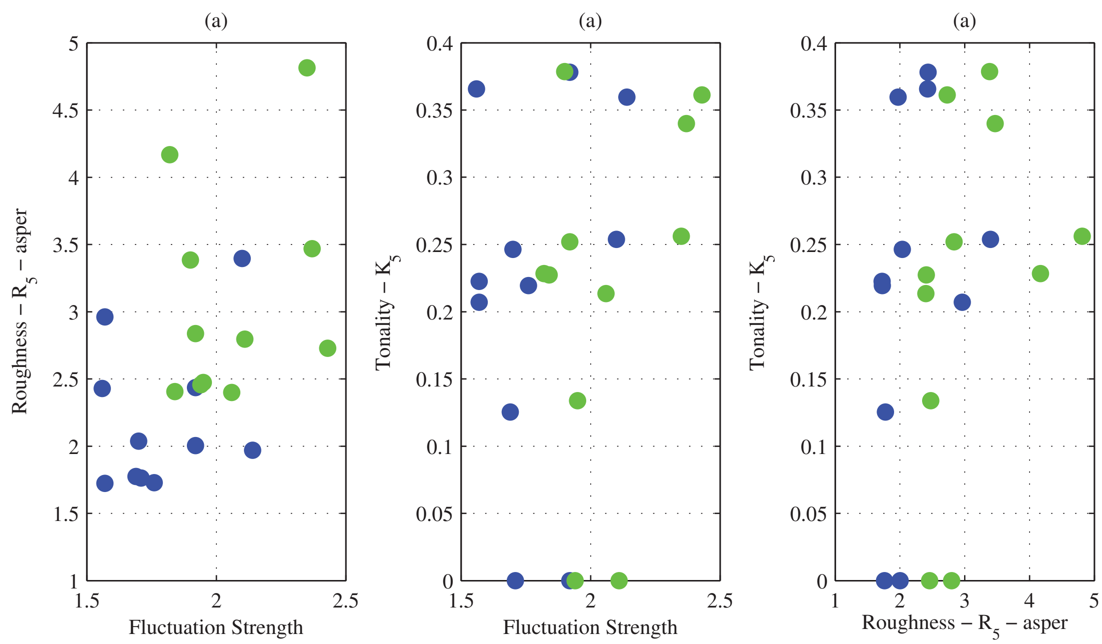

Figure 4.12. Various sound quality metrics plotted against one another for loudness normalized signals sets from the earphone test: low loudness normalization (blue) and high loudness normalization (green).

4.4 Annoyance Test with Loudspeakers Playback

This test was performed at NASA Langley Research Center in the Exterior Effects Room (EER). The specific test setup, subjects, procedure, and results are detailed in this section.

4.4.1 Specific Test Setup

For this test, signals were played back through a set of 27 K&H O300 mid and high frequency satellite speakers and 4 K&H O900 subwoofers [39]. Playback was controlled by a real-time audio server using an implementation of Vector-Base Amplitude Panning.

For this test, a second set of signals was created using a 4th order high pass Butterworth filter with a cutoff frequency of 17 Hz from the common set. This was an attempt to fully utilize the low-frequency reproduction capability of the Exterior Effects Room, as the system is calibrated to have a near-flat response down to 17 Hz. The total number of signals used in this test was 110, 55 high pass filtered with a cutoff frequency of 17 Hz and 55 high pass filtered with a cutoff frequency of 25 Hz.

4.4.2 Subjects

Forty subjects participated in this test. Subjects were recruited from the general public around the NASA Langley area. The subjects ranged in age from 19-58 years, 13 were male and 27 were female. All of the subjects passed a hearing screening requiring no more than 30 dB of hearing loss in either ear in a range of frequencies from 125 to 8000 Hz. This differs from the Purdue testing requirement of a max of 20 dB hearing loss. The 30 dB value was chosen as a compromise between 20 dB and 40 dB, the latter being the levels typically used by NASA Langley researchers in other subjective tests. Subjects were compensated $50 in addition to travel mileage compensation.

4.4.3 Results

Numbers were assigned to the scale, ranging from 1 (left end of scale) to 11 (right end of scale). (Not at All, Slightly, Moderately, Very, and Extremely) corresponded to the numbers (2, 4, 6, 8, and 10), respectively. The average of the subjects responses and the standard deviation of the estimated mean (SEM) were calculated for each signal. The SEM for all signals was lower than 0.36, which corresponds to 3.6% of the entire scale, or 18% of the distance between two words on the scale.

The differences in ratings of sounds that were high-pass filtered with a cut-off frequency at 17 Hz and those with a cut-off at 25 Hz were usually small with no significant trends related to vehicle or type of modification. Figure 4.13 illustrates the high correlation ($R^2 = 0.883$) between responses to these signals.

The Exterior Effects Room is calibrated for playback at a microphone near the center of the room. This microphone is located between two seats where subjects sat during the test. Subjects sat in this row (will be referred to as the front row) and the row directly behind it (will be referred to as the back row). Distance between subject seat locations and the center calibration microphone is not the same for both rows, so some small differences are present in stimuli presentation. Each signal was recorded in each of the four locations where subjects sat during the test. Recording was performed using a National Instruments PXI-1042Q acquisition system, four GRAS 26CA preamplifiers, and GRAS 40AQ microphones. $EPNL$ values were calculated for these recordings and shown in Figure 4.14 are $EPNL$ values averaged for each row. Most signals have higher $EPNL$ values for the front row recordings, with the exception being the MD520N (blue) and MD902 (cyan), which have higher values for the back row recordings. These two vehicles use the NOTAR system, which employs a rear facing high speed fan. This fan generates a strong tone that becomes more prevalent after the vehicle has passed the overhead point, which means that the back seat recordings could have higher tone corrections in the $EPNL$ metric, causing higher $EPNL$ values. Average annoyance ratings from subjects in each row are plotted against one another in Figure 4.15, where almost all signals had higher ratings from subjects who sat in the front row than from those who sat in the back row.

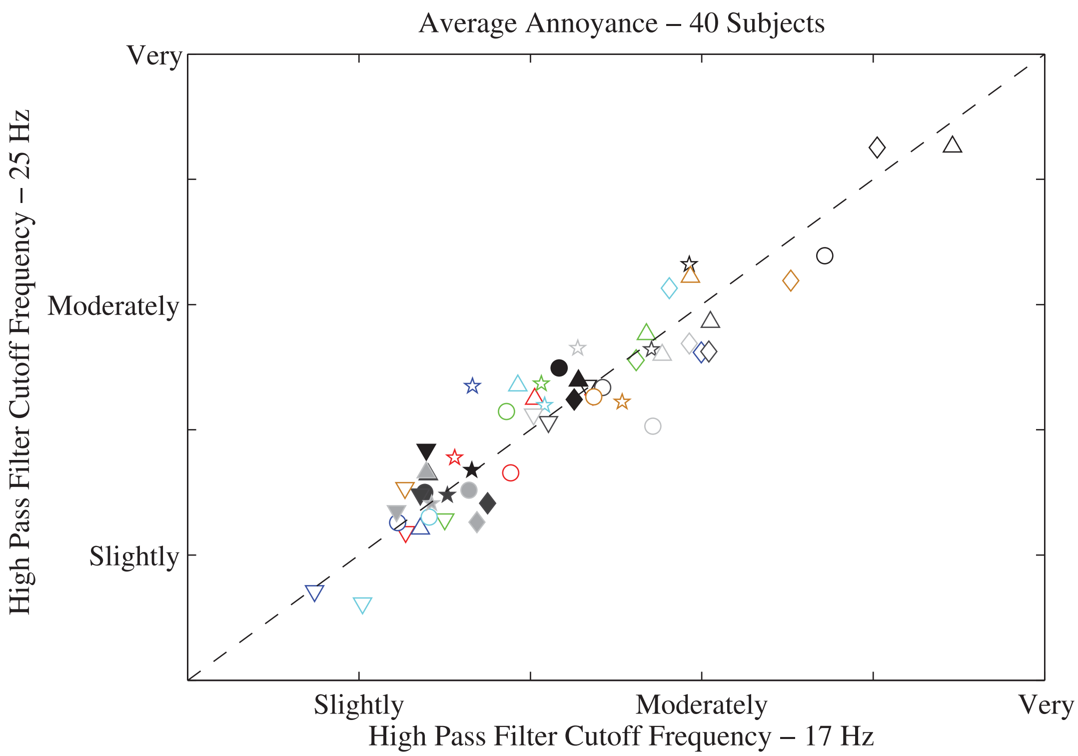

Figure 4.13. Average annoyance ratings for signals high pass filtered with a 17 Hz cutoff frequency plotted against average annoyance ratings for identical signals high pass filtered with a 25 Hz cutoff frequency.

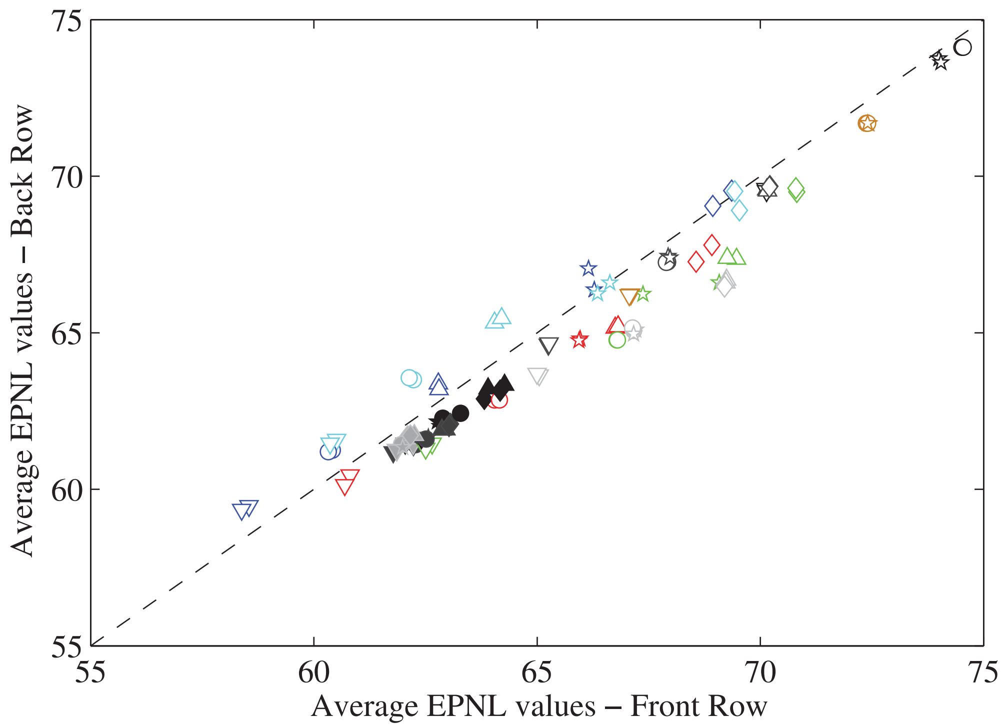

Figure 4.14. Average $EPNL$ values for the two recording locations in the front row plotted against average $EPNL$ values for the two recording locations in the back row.

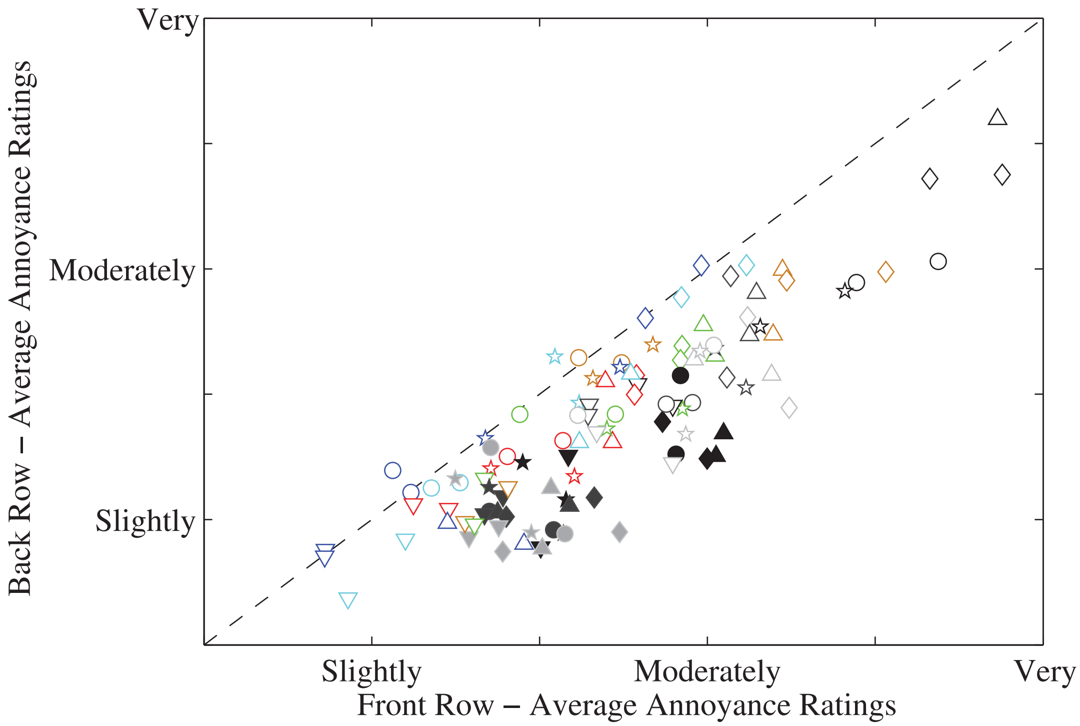

Figure 4.15. Average annoyance ratings from subjects seated in the front row plotted against average annoyance ratings from subjects seated in the back row.

4.4.3.1 Results by Vehicle and Modification

Average annoyance ratings sorted by vehicle with symbols for each modification type are shown in Figure 4.16. Similar to the results from the earphone test, subjects rated the XV-15 in plane mode to be less annoying than the XV-15 in helicopter mode, and ratings for simulations with lowered fundamental frequencies decreased with decreases in frequency. For most vehicles, the highest average annoyance ratings were for those signals with raised tone magnitudes and those with both raised tone magnitudes and impulsiveness, and the lowest average annoyance ratings were for those signals with lowered tone magnitudes.

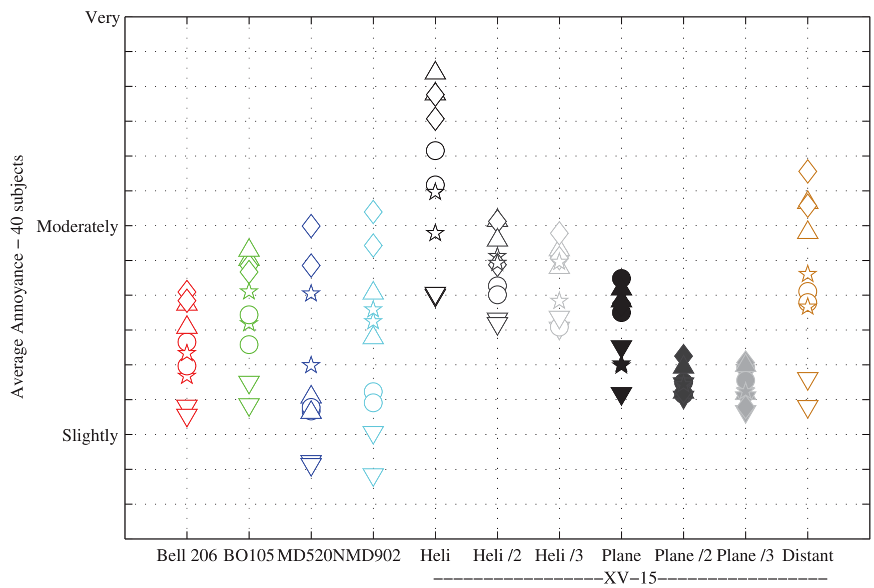

Figure 4.16. Average annoyance ratings plotted against vehicle type. Original signal (circle), lowered tones (downward triangle), raised tones (upward triangle), raised tones and phase aligned (diamond), lowered phase not aligned (star). The changes to the XV-15 fundamental frequency are treated as their own vehicle, where /2 and /3 correspond to one-half and one-third the original fundamental frequency. The labels Heli and Plane correspond to the helicopter mode (Nacelle Angle 80°) and airplane mode (Nacelle Angle 0°) of the XV-15.

4.4.3.2 Relationship between Metrics and Responses

Level-based metrics were calculated for all four recordings of each signal, and averaged across the four locations. The metrics calculated were Effective Perceived Noise Level ($EPNL$), A-weighted Sound Exposure Level ($SELA$), Perceived Loudness ($PLdB$), and statistics of Zwicker (DIN 45631), and Moore and Glasberg Time-Varying Loudness such as level exceeded 5% (N_5) or 10% (N_{10}) of the duration of the signal. $EPNL$ and $SELA$ were calculated using ANOPP2 Acoustic Analysis API [32]. Zwicker Loudness was calculated using HEAD Acoustics Artemis Sound Quality Software (Version 12) (DIN 45631), while Moore and Glasberg Time-Varying Loudness [36] was calculated using software developed by a previous graduate student at Purdue University [40], that incorporates relevant revisions in ANSI S3.4-2007 [41]. The average annoyance ratings were compared to these metrics values for each signal and the results are shown in Figure 4.17. For these signals, $EPNL$ was most highly correlated to the average of the subjects responses ($R^2 = 0.805$). The two types of loudness models produced highly correlated metric values for N_5 ($R^2 = 0.99$) and had a similar correlation with annoyance ratings (Moore and Glasberg N_5, $R^2 = 0.55$; Zwicker N_5, $R^2 = 0.61$).

One of the differences between the plots shown in Figure 4.17 is that there are more groups of signals of nearly the same Loudness metric values that have very different average annoyance ratings (see arrows on plot), whereas this is not so evident in the $EPNL$ and $SELA$ results. The group of points in a vertical line around 11 sones largely consists of simulations of the XV-15 aircraft in airplane mode including those with lowered fundamental frequencies. The points in vertical lines around 16 and 19 sones mainly consist of XV-15 signals, where the symbols with lower annoyance ratings denote simulations based on the segment recorded far from the microphone, and the symbols with higher annoyance ratings denote simulations based on the XV-15 in helicopter mode. The points with $EPNL$ and $SELA$ values around 66 dB and 65 dB include signals from a variety of vehicles that range in annoyance ratings with

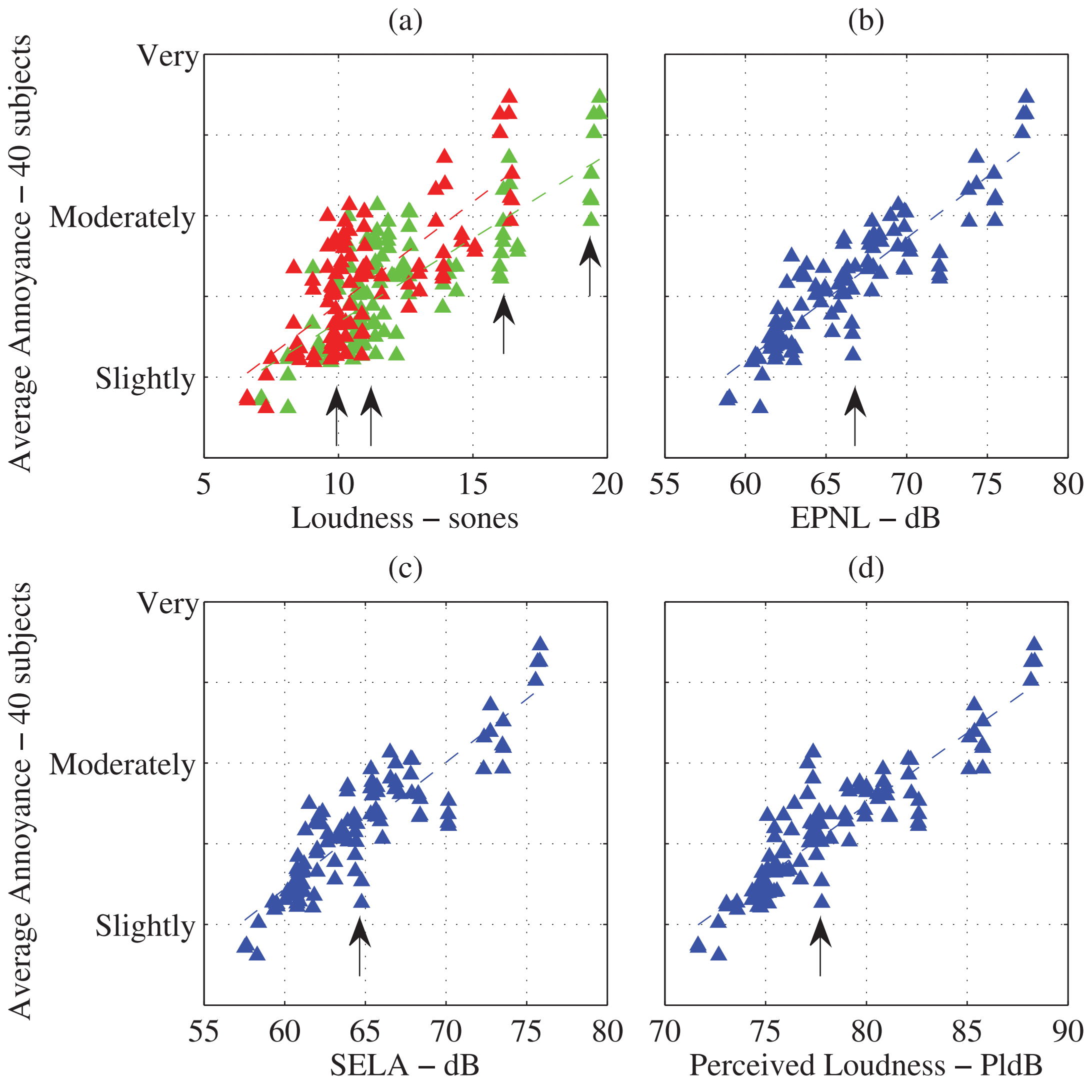

Figure 4.17. Average annoyance ratings for the loudspeaker test plotted against various metrics. Plotted against (a) Zwicker Loudness: level exceeded 5% of the time (green triangles, $R^2 = 0.609$) and Moore and Glasberg Short-Term Loudness: level exceeded 5% of the time (red triangles, $R^2 = 0.551$); (b) Effective Perceived Noise Level, $R^2 = 0.805$; and (c) A-weighted Sound Exposure Level, $R^2 = 0.798$. Arrows indicate locations where there are signals with very similar metric values but with a large range of annoyance responses.

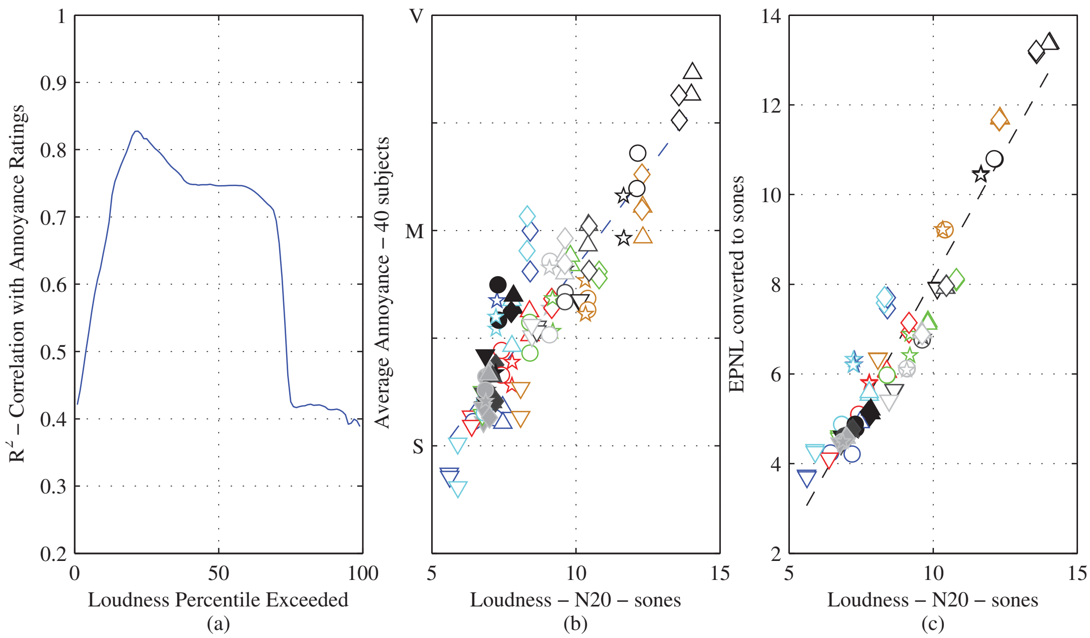

Figure 4.18. Evaluation of the annoyance prediction performance of level-based metrics for the loudspeaker test. (a) Coefficient of determination (R^2) as a function of percentile of loudness exceeded; (b) Average annoyance ratings plotted against Moore and Glasberg Short-Term Loudness: Level exceeded 20% of the time ($R^2 = 0.821$); (c) $EPNL$ values converted to sones plotted against Moore and Glasberg Short-Term Loudness: Level exceeded 20% of the time ($R^2 = 0.921$).

the simulation type. For these groups, the symbols with lower annoyance ratings are those with lowered tones and tones that are not phase aligned, and the symbols with higher annoyance ratings have raised tones and/or tones that are phase aligned.

The two metrics that performed best involve an integration around the peak value of the metric being calculated (A-weighted sound pressure level or $PNLT(t)$) from when the metric value first goes above 10 dB below peak level to when it goes below this level for the last time). The correlation between percentiles of loudness other than level exceeded 5% of the time and annoyance were also examined. The R^2 values as a function of the percentile are shown in Figure 4.18 (a). The highest correlation (R^2 = 0.821) occurs for loudness exceeded 20% of the time and the annoyance scores are plotted versus this metric in Figure 4.18 (b). The sounds deviating most from the general trend (around 8 sones) are from the two NOTAR vehicles, the MD520N and MD902. In Figure 4.18 (c) the $EPNL$ values converted to sones are plotted against Loudness exceed 20% of the time, with symbol coding as in Figure 4.16. When $EPNL$ (in sones) is compared to Loudness exceeded 20 of the time, some of the sounds from the two NOTAR vehicles (blue and cyan symbols) are above the best-fit line relating the two metrics (around 7 to 8 sones). This may be due to the tone correction in the $EPNL$ calculation. Note that these signals are not obvious outliers in Figure 4.17 (b), and thus this tone correction could be accounting for some of the impact that the high frequency tone in the NOTAR sounds has on annoyance.

Other psychoacoustic-based sound quality metrics, including Sharpness, Roughness, Tonality, and Fluctuation Strength, were calculated for these signals. The left plot in Figures 4.19 - 4.22 show average annoyance ratings plotted against these metrics. The effect of these metrics on annoyance separate from that which can be predicted from loudness was also examined. For this comparison a linear annoyance model from N_{20} was created, and the residuals of the that model were compared to each of the metrics. These results are shown in the right plot in Figures 4.19 - 4.22. Similar to the earphone test the results do not show clear trends, only even weaker trends with Roughness and Fluctuation Strength, and points towards responses being mainly loudness driven.

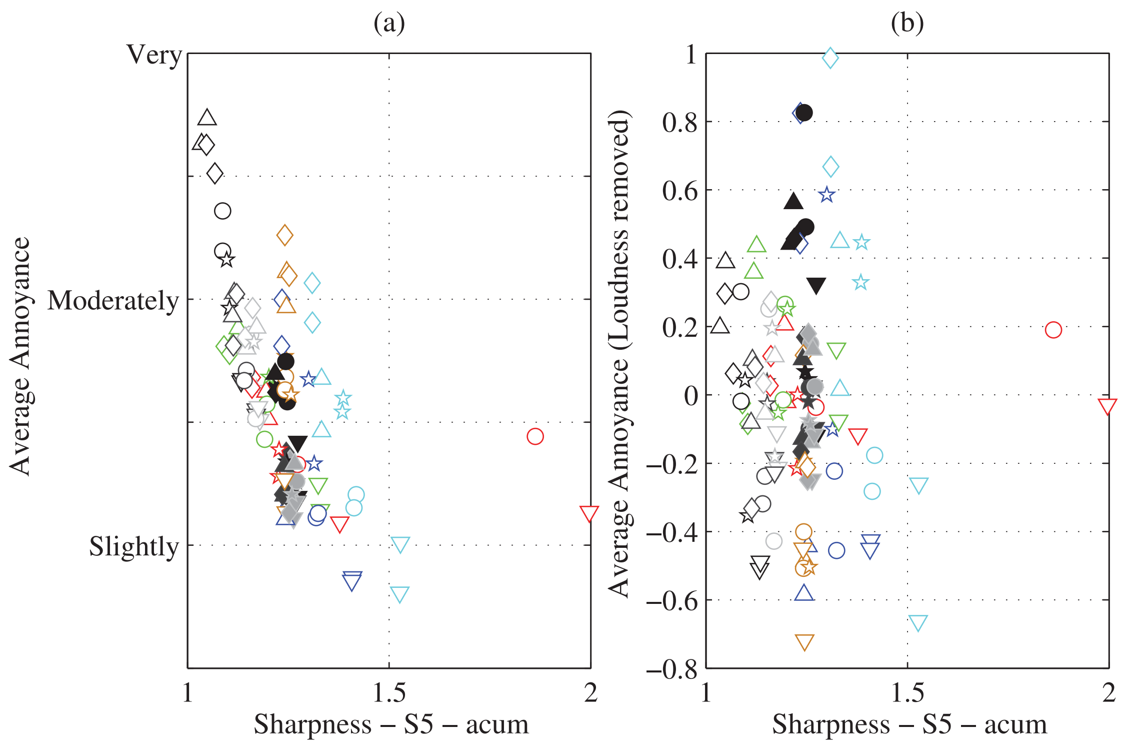

Figure 4.19. (a) Average annoyance ratings plotted against Sharpness exceeded 5% of the time; (b) Average "loudness removed" annoyance plotted against Sharpness exceeded 5% of the time.

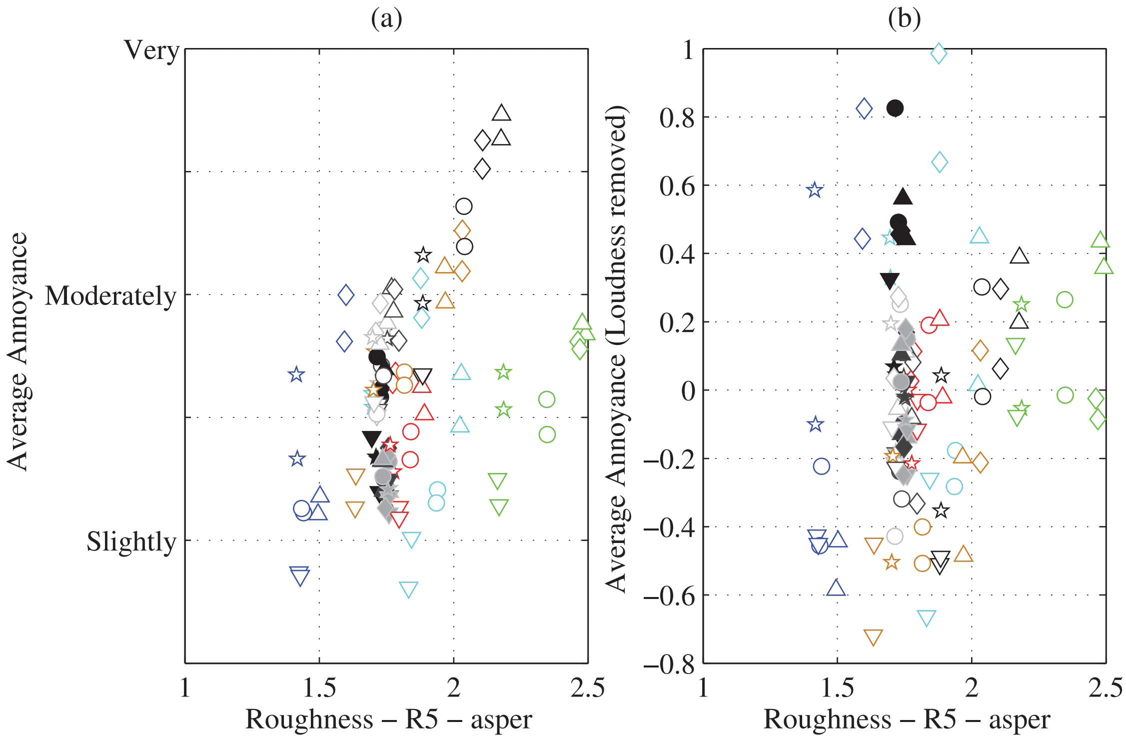

Figure 4.20. (a) Average annoyance ratings plotted against Roughness exceeded 5% of the time; (b) Average "loudness removed" annoyance plotted against Roughness exceeded 5% of the time.

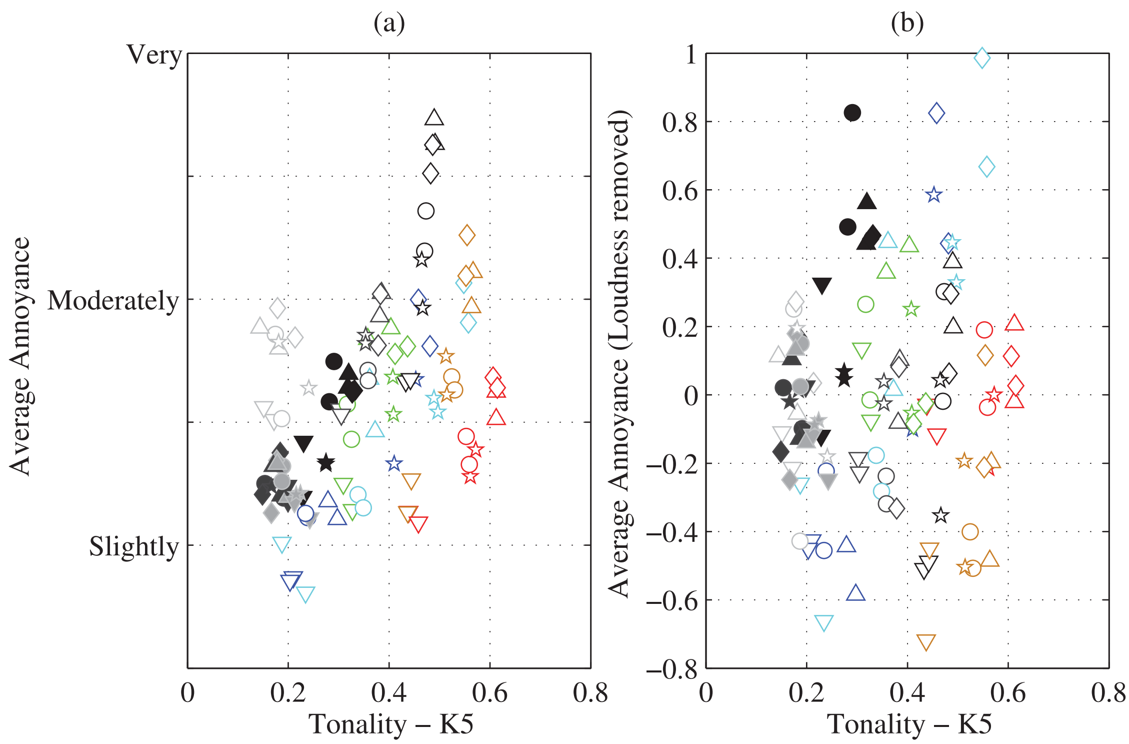

Figure 4.21. (a) Average annoyance ratings plotted against Aures Tonality exceeded 5% of the time; (b) Average "loudness removed" annoyance plotted against Aures Tonality exceeded 5% of the time.

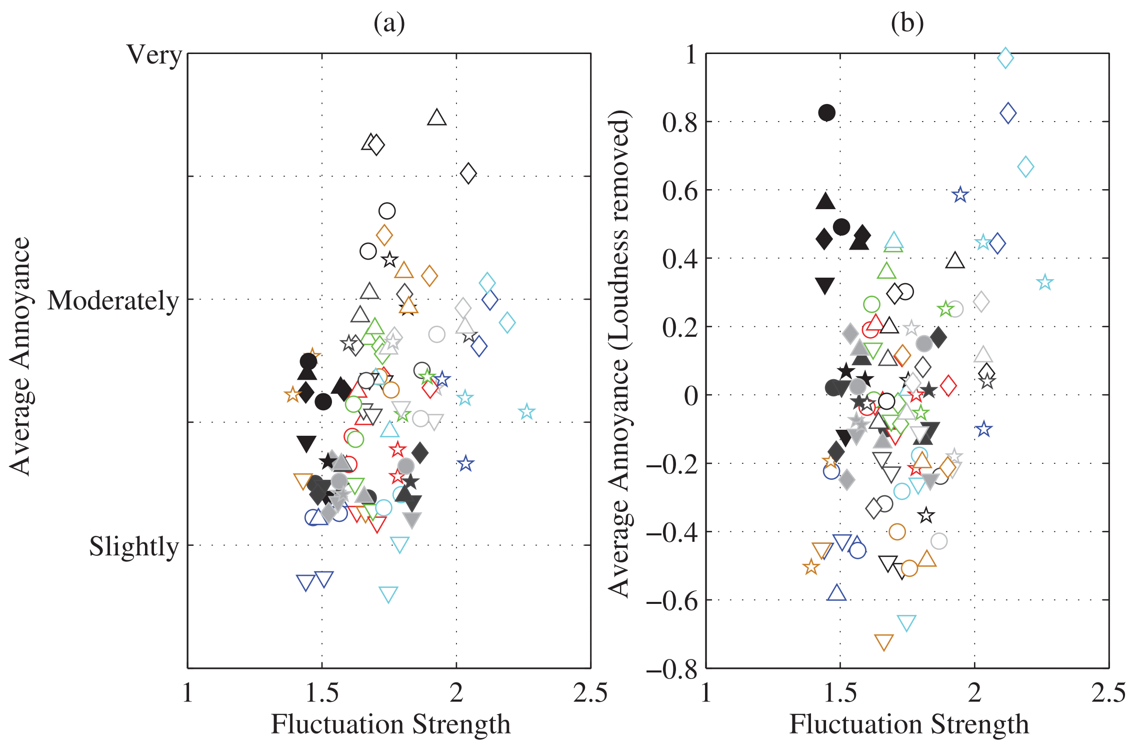

Figure 4.22. (a) Average annoyance ratings plotted against Fluctuation Strength; (b) Average "loudness removed" annoyance plotted against Fluctuation Strength.

4.5 Comparison of Annoyance Results from the Two Tests

It is important to point out that there are other differences between these two tests other than the playback method. These differences include the test environment (small booth; large room), subject population (differences in age, gender breakdown, and community noise exposure). The test performed using earphones was conducted at Purdue University, which is near a small airport. This airport mainly has operations of small single- and twin-propeller aircraft which are audible throughout the day. The test performed using loudspeakers was conducted at NASA Langley Research Center, which borders Langley Air Force base, where military aircraft and rotorcraft are present on a usual basis. While both test environments are nearby sources of aircraft noise, the noise signatures produced by the vehicles examined in this research are quite different to those the subjects regularly hear.

Similar trends for the two tests are seen in the XV-15 simulation group, and between simulation types involving lowered and raised tone magnitudes. For the loudspeaker test, the four helicopters (Bell 206, BO105, MD520N, and MD902) are rated more similarly to the XV-15 in "plane" mode, whereas the earphone test results showed ratings for these vehicles to be more similar to the XV-15 in "helicopter" mode.

In Figure 4.23 subject responses from the tests in both environments for the set of 55 common signals are plotted against each other. Calibration was performed to match maximum slow A-weighted sound pressure levels in the two environments. On average subjects in the earphone testing environment rated sounds as more annoying, as shown by the placement of the majority of the points above the equal response line. The points furthest from this line are mainly from the two NOTAR vehicles, while the points closest to this line include XV-15 signals both in "plane" mode and "helicopter" mode.

In Figure 4.24 average responses from the earphone test plotted against average responses from the front row (a) and back row (b) of the loudspeaker test are shown. Nearly every point moves to the left when going from (a) to (b), showing that the back

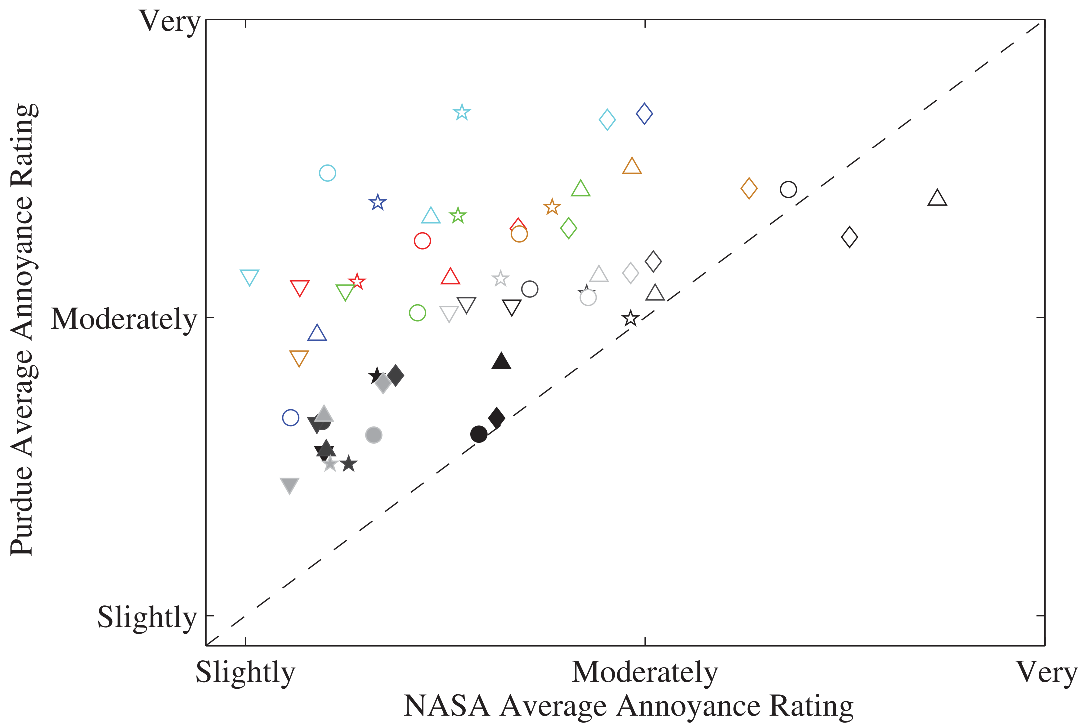

Figure 4.23. Average annoyance ratings from the two annoyance tests plotted against one another. See Figure 4.2 or 4.16 for color and symbol coding.

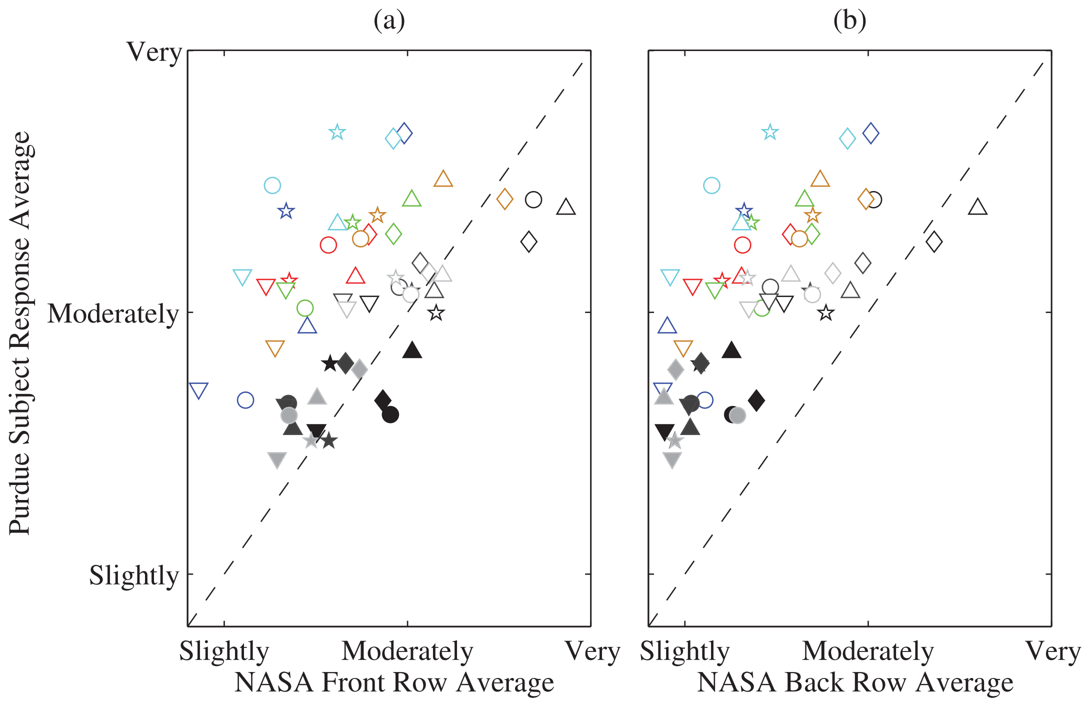

Figure 4.24. Average annoyance ratings from the earphone test plotted against front row ratings (a) and back row ratings (b) from the loudspeaker test. See Figure 4.2 or 4.16 for color and symbol coding.

row (on average) rated signals as less annoying, and that the earphone test responses were more closely related to responses of subjects who sat in the front row for the loudspeaker test. While the a large number of points in (a) are well above the line, the XV-15 signals (black and gray) are fairly compact around this line, meaning that these two subjects groups (subject who sat in the front row at NASA and subjects at Purdue) rated this set of signals similarly.

Table 4.2. Coefficients of determination between subject responses and metrics.

	Earphones	Loudspeakers		
		Front Row	Back Row	All
$EPNL$	0.637	0.721	0.795	0.805
$SELA$	0.504	0.724	0.776	0.798
$PLdB$	0.400	0.783	0.740	0.813
Loudness - N_5	0.263	0.541	0.502	0.551
Loudness - best	0.546	0.772	0.763	0.821

The coefficient of determination values for different metrics for each of the subject groups are given in Table 4.2. For all subject groups from the loudspeaker test, little to no improvement was seen for using the $EPNL$ metric rather than $SELA$. Subject responses in the earphone test were much more correlated with $EPNL$ values than $SELA$ values. This may indicate that subjects in the earphone test responded more to the presence of tones, and the tonal correction in the $EPNL$ calculation aided in predicting responses. However, the $EPNL$ metric consistently performs well across all subject groupings.

4.6 Summary of Findings from the Two Annoyance Tests

A parametric annoyance test was performed to examine the relationship between annoyance ratings and variations in rotorcraft noise characteristics, in particular characteristics controlled by the tonal content of the signal, and to assess the applicability of different noise metrics for evaluation of rotorcraft noise. In general, more impulsive sounds (tone families phase aligned) and more tonal sounds were found to be more annoying than sounds with lower level tones or sounds with tones that were not phase aligned. In both tests, $EPNL$ and $SELA$ both performed reasonably well, and similarly, as predictors of average annoyance ratings. Both $EPNL$ and $SELA$ are event metrics and involve integration around peak levels, perhaps a similar approach with loudness should be considered, rather than using the percentile loudness predictions, which are signal duration dependent. While the results of this experiment showed that the $EPNL$ metric performed reasonably well, there may be room for improvement. There are outliers from the general trend lines that have characteristics possibly contributing to annoyance.

5. CONCLUDING COMMENTS AND RECOMMENDATIONS FOR FUTURE WORK

The goal of this research was to determine what attributes of rotorcraft noise have an effect on annoyance, and to assess the performance of various noise metrics for their applicability to rotorcraft. Due to the differences between the noise signatures of traditional fixed-wing aircraft and those of rotorcraft, current aircraft noise metrics needed to be investigated for their applicability to rotorcraft. The hypothesis was that the strong, low frequency tonal components, and impulsiveness of rotorcraft noise would create the need for an additional penalty or metric for accurate prediction of annoyance.

5.1 Summary

A collection of rotorcraft flyover recordings were gathered and analyzed as a first step in this assessment. These recordings were used to generate a set of test stimuli to be used in two semantic differential tests - one conducted using earphones and the other using loudspeaker arrays for playback. These tests were performed to determine which attributes were present in human response and to determine if the playback method affected the results. A factor analysis of subjective responses showed the presence of factors related to loudness, sharpness/spectral balance, roughness, impulsiveness, and tonality. These results led to the development of a signal modification method in order to change tonal components in signals. This was accomplished by first characterizing the time-varying characteristics of the tonal components of a signal; then removing them; followed by regenerating a new set of tonal components with adjusted magnitudes, relative phases, and/or frequencies; and finally adding these to the tones-removed signal. This method was used to generate a set of stimuli including original recordings, as well as signals with altered tonality and impulsiveness. A parametric test was performed using this set of stimuli (again in the two playback environments) so that the effect of these altered attributes on annoyance ratings could be observed.

It was found from analysis of the annoyance test results that $EPNL$ and $SELA$, which are commonly used for the evaluation of aircraft noise, were the metrics most highly correlated with annoyance ratings. Additionally, it was found that for some subject groups there was little to no improvement through the use of $EPNL$ over use of $SELA$.

There were variations in average annoyance responses not predicted by $EPNL$ and $SELA$ that were related to signal modifications. In general, more impulsive sounds were rated as more annoying, and signals with higher (or lower) tonal levels were rated as more (or less) annoying. Use of additional metrics that quantify tonality did not improve predictions of annoyance, even though signal modifications were related to changes in tone-family characteristics. It is hypothesized that the tone correction in $EPNL$ was somewhat helpful but for some sets of sounds it did not quantify the changing tone characteristics that appeared to be affecting people's responses. Zwicker's and Moore and Glasberg's time varying loudness models were also examined as predictors of annoyance but did not generally perform as well as $EPNL$. While loudness exceeded 5% of the time is often used as an event loudness metrics, it was found that loudness exceeded 10-20% of the time performed better for these tests. Two-metric models incorporating loudness and tonality did not perform much better than a one-metric (loudness-based) model. Correlations for the earphone test were significantly lower than for the EER (loudspeaker) tests and this needs further investigation. There were also differences in responses that were a function of aircraft type. These were less pronounced when responses were plotted against $EPNL$ rather than N_{20}, but were still present. The responses for some aircraft followed difference trend lines. Typically distant aircraft sounds with a less pronounced Doppler shift were below the trend line predicted by $EPNL$, meaning that the $EPNL$ metric may be overestimating the annoyance as a result of these sounds. The XV-15 was generally rated as less annoying when in “plane” mode rather than “helicopter” mode, and also rated as less annoying when simulated with lowered blade passage frequencies.

Many similar trends were seen between the two playback locations in both the semantic differential tests (most factors present were seen in results from both tests) and annoyance tests (trends within the XV-15 signals and lowered blade passage frequency simulations). However, some differences were apparent in the results from the annoyance test between test environments. Signals generated from flyover recordings of the two NOTAR vehicles were rated very differently (much higher at Purdue). The ability of metrics to predict annoyance ratings also varied between test locations. The difference between the prediction ability of $EPNL$ and that of $SELA$ was very small for the results from the loudspeaker test (difference in R^2 values of 0.007) but noticeable for the results from the earphone test (difference in R^2 values of 0.133). Results from both annoyance tests showed changes in subject responses when signals included phase aligned tonal components (more impulsive) but little to no changes were produced in metric values. Design of an impulsiveness metric is included in recommendations for future work.

5.2 Contributions

While this work was limited to results from subjective tests using segments of flyover recordings played back via earphones or loudspeakers, the results produced provide many contributions towards a deeper understanding of rotorcraft noise, as well as the applicability of noise metrics. The major contributions are as follows.

1. A lexicon was performed to gather words used to describe rotorcraft noise. This produced many common words such as: helicopter (volunteers recognized the noise source), whine (recognition of high frequency components), and choppy (describing the impulsiveness of the noise). The results of this lexicon are included in Appendix D.

2. The subject responses from the semantic differential test (described in Chapter 3) were analyzed to determine the number of independent perceptual factors. The factors identified are related to loudness, irregularity, roughness, spectral balance, and tonality. The presence of these attributes is vital to our understanding of human response to rotorcraft noise, and led to the development of the the simulation method described in Chapter 2.

3. A simulation method was developed to characterize and modify tonal components of a rotorcraft flyover recording. This method can be used to alter the strength, phase alignment, and fundamental frequency of harmonic series within the recording, all of which vary through time. This allows for the generation of rotorcraft signals with different levels of tonality and impulsiveness, as well as lowered blade passage frequencies.

4. The results from the annoyance test showed that the currently used certification metric ($EPNL$) performs well, but could be improved. Changes to harmonic series resulted in changes in the average annoyance ratings but the metric did not always predict these changes. There is evidence that the tone correction in $EPNL$ is helpful but for some sets of signals it did not fully account for the variation in tone strengths.

5.3 Recommendations for future work

Research on the noise impact of various vertical takeoff and landing (VTOL) vehicles is ongoing. It is important that these noise sources continue to be investigated so that if implemented there will be no effect on the community. Suggestions for future work include:

1. Subjective test using signals where important attributes (loudness, tonality, impulsiveness) are varied independently.

2. Development of a metric to quantify impulsiveness.

3. Investigation of annoyance to a combination of traditional aircraft and rotorcraft, i.e., a subjective test containing signals of both aircraft and rotorcraft flyovers.

4. Verification of results with surveys in communities with rotorcraft noise.

5. Further investigate the differences between the results of the earphone and loudspeaker annoyance tests.

The first three of these items were included in the original project proposal, but the scope of project has since been limited due to budget constraints. Additional subjective testing would allow for further assurance of the performance of metrics, as well as the possibility of developing an improved annoyance model addressing inclusion of impulsiveness and improvements to tone corrections. While results of subjective testing gives indications of preferences to reproduced signals, it will be important to verify these results with field testing and live aircraft. In a true outdoor flyover setting, the presence of rotorcraft noise can be very different to that of a signal played back over headphones. The signals presented to subjects in this research were at most 15 seconds, where flyovers can last over a minute long. At close distances rotorcraft may produce tactile effects which earphones are not capable of reproducing. While the results of this research show the capability of $EPNL$ and other noise metrics to predict annoyance from rotorcraft noise, it is the opinion of this author that additional subjective testing (via signal playback and in-field) is necessary before the implementation of these vehicles can begin.

LIST OF REFERENCES

LIST OF REFERENCES

[1] EPA (U.S. Environmental Protection Agency). Information on levels of environmental noise requisite to protect public health and welfare with an adequate margin of safety, Report No. 550-9-74-004. March 1974.

[2] H. Fastl and E. Zwicker. *Psychoacoustics Facts and Models*. Springer, Berlin New York, 2007.

[3] ISO. Acoustics - Description, Measurement and Assessment of Environmental Noise - Part 1: Basic Quantities and Assessment Procedures. ISO 1996-1:2003, International Organization for Standardization, Geneva, Switzerland, 1996.

[4] B. Berglund and T. Lindvall. Community noise,Archives of the Center for Sensory Research. *Stockholm University and Korlinska Institute*, 2:1–195, 1995.

[5] T. Miyakita, T. Matsui, A. Ito, T. Tokuyama, K. Hiramatsu, Y. Osada, and T. Yamamoto. Population-based questionnaire survey on health effects of aircraft noise on residents living around u.s. airfields in the ryukyus—part i: An analysis of 12 scale scores. *Journal of Sound and Vibration*, 250:129–137, 2002.

[6] Civil Tiltrotor Development Advisory Committee Report to Congress (CTRDAC-1995-REPORT) in accordance with PL102-581, December 1995.

[7] M.E. Watts, D.A. Conner, and C.D. Smith. Joint Eglin Acoustic Week III Data Report. Technical Report NASA/TM-2010-216206, National Aeronautics and Space Administration, March 2010.

[8] D. Conner, M. Marcolini, B. Edwards, and J. Brieger. XV-15 Tiltrotor Low Noise Terminal Area Operations. In *American Helicopter Society 53rd Annual Forum*, Virginia Beach, VA, April 1997. American Helicopter Society.

[9] C. Burley, C. Smith, and D. Conner. Flight acoustic testing and data acquisition for the rotorcraft noise model (rnm). In *American Helicopter Society 62nd Annual Forum*, Phoenix, AZ, May 2006. American Helicopter Society.

[10] A.H. Hastings, K.H. Lee, P. Davies, and A.M. Surprenant. Measurement of the attributes of complex tonal components commonly found in product sound. *Noise Control Engineering Journal*, 51:195–209, 2003.

[11] L.C. Sutherland and R.E. Burke. Annoyance, loudness and measurement of repetitive type impulsive noise sources. EPA Environmental Protection Agency Report No. EPA 550/9-79-103, Environmental Protection Agency, November 1979.

[12] M.F. Russell, S.A. Worley, and C.D. Young. Towards an objective estimate of the subjective reaction to diesel engine noise. In *Proceedings of the 1987 Noise and Vibration Conference*, Traverse City, Michigan, May 1987.

[13] H. Schiffbanker, F.K. Brandl, and G.E. Thien. Development and application of an evaluation technique to assess the subjective character of engine noise. In *Proceedings of the 1991 Noise and Vibration Conference*, Traverse City, Michigan, May 1991.

[14] A.H. Hastings. *Sound Quality of Diesel Engines, Ph.D Thesis*. Purdue University, August 2004.

[15] S. More and P. Davies. Human responses to the tonalness of aircraft noise. *Noise Control Engineering Journal*, 58:420–440, 2010.

[16] S.S. Stevens. Concerning the form of the loudness function. *Journal of the Acoustical Society of America*, 29:603–606, 1957.

[17] S.S. Stevens. Procedure for calculating loudness: Mark VI. *Journal of the Acoustical Society of America*, 33:1577, 1961.

[18] ISO532B. Acoustics - method for calculating loudness level. Iso, International Organization for Standardization, Geneva, Switzerland, 1975.

[19] ANSI S3.4-1980. Procedure for the computation of loudness of noise. Ansi, American National Standards Institute, New York, 1980.

[20] S.S. Stevens. Perceived level of noise by Mark VII and decibels. *Journal of the Acoustical Society of America*, 51:575, 1972.

[21] E. Zwicker. Procedure for calculating loudness of temporally variable sounds. *Journal of the Acoustical Society of America*, 62:675–682, 1977.

[22] B. Moore and B. Glasberg. A revision of Zwicker's loudness model. *Acustica - Acta Acustica*, 82:335–345, 1996.

[23] DIN45631/A1. Acoustics - method for calculating loudness level. Din, Deutsches Institut fur Normung, Berlin, Germany, 2010.

[24] G. von Bismarck. Sharpness as an attribute of the timbre of steady sounds. *Acustica*, 30:159–172, 1974.

[25] W. von Aures. The sensory euphony as a function of auditory sensations. *Acustica*, 59:130–141, 1985.

[26] P.D. Schomer. Criteria for assessment of noise annoyance. *Noise Control Engineering Journal*, 53:132–144, 2005.

[27] B. Berry and E. Zwicker. Comparison of subjective evaluation of impulsive noise with objective measurements of the loudness-time function given by loudness meter. In *Proceedings of Inter-Noise '86*, Cambridge, MA, July 1986.

[28] S. More. *Aircraft Noise Characteristics and Metrics, Ph.D Thesis*. Purdue University, 2010.

[29] K. Hoon. *Perception of tones in machinery noise and its influence on annoyance, Ph.D Thesis*. Purdue University, 2006.

[30] FAA Federal Aviation Regulations (FARS, 14 CFR). Calculation of Effective Perceived Noise Level From Measured Data. Federal Aviation Administration Code of Federal Regulations., 14 CFR Part 36, Appendix A2, Section A36.4 2006.

[31] S. More and P. Davies. An examination of the influence of spectral balance on annoyance ratings of aircraft sounds. In *Proceedings of Inter-Noise 2007*, Istanbul, Turkey, August 2007.

[32] L. Lopes and C. Burley. Design of the next generation aircraft noise prediction program: ANOPP2. In *17th AIAA/CEAS Aeroacoustics Conference*, Portland, OR, June 2011. American Institute of Aeronautics and Astronautics. AIAA 2011-2854.

[33] S. Namba, S. Kuwano, and M. Koyasu. The measurement of temporal stream of hearing by continuous judgments-in the case of the evaluation of helicopter noise. *Journal of the Acoustical Society of Japan*, 14:341–352, 1993.

[34] V. Quaranta, I. Dimino, M. d'Ischia, and F. Cenedese. 3D internal noise simulation for vibro-acoustic comfort assessment: an application proposal for AW-109 helicopter. *The Journal of Aerospace Science, Technology and Systems*, 90:119–130, 2011.

[35] J.S. Kerrick, D.C. Nagel, and R.L. Bennett. Multiple ratings of sound stimuli. *Journal of the Acoustical Society of America*, 45:1014–1017, 1969.

[36] B. Moore and B. Glasberg. A model of loudness applicable to time-varying sounds. *Journal of the Audio Engineering Society*, 50:331–342, 2002.

[37] S. Sharma. *Applied Multivariate Techniques*. John Wiley & Sons, New York, 1996.

[38] J. Hair. *Multivariate Data Analysis*. Prentice Hall, Upper Saddle River, NJ, 2005.

[39] K. Faller II, S. Rizzi, and A. Aumann. Acoustic performance of a real-time three-dimensional sound-reproduction system. Technical Report NASA/TM-2013-218004, National Aeronautics and Space Administration, June 2013.

[40] A.J. Marshall. *Development of a Model of Startle Resulting form Exposure to Sonic Booms, Ph.D Thesis*. Purdue University, 2012.

[41] ANSI S3.4-2007. Procedure for the computation of loudness of steady-state sounds. Ansi, American National Standards Institute, New York, 2007.

[42] J. Vos. On the annoyance caused by impulse sounds produced by small, medium large, and large firearms. *Journal of the Acoustical Society of America*, 109:244–253, 2001.

APPENDICES

A. SEMANTIC DIFFERENTIAL TEST - METRICS, SCALE RATINGS, AND EXAMPLE EVALUATION SHEETS

This Appendix contains metric values, average scale ratings, and example evaluation sheets for the semantic differential test performed using: (1) earphones and (2) using loudspeakers. The test and main analysis results are described in Chapter 3.

A.1 Metrics and Scale Responses for Both Tests

The metric values and scale responses for the semantic differential tests are shown here. The semantic differential tests are discussed in Chapter 3.

Table A.1. Metric values for each signal played in the test using: (1) earphones and (2) loudspeakers. Psychoacoustics-based metrics exceeded 5% of the time: Loudness (N_5), Sharpness (S_5), Tonality (K_5), and other level metrics: $EPNL$, $SELA$, $PLdB$.

Signal / Metric	1	2	3	4	5	6	7	8	9	10
***N5* (1)**	46.53	43.68	42.33	18.92	34.61	37.77	54.47	28.20	53.01	63.66
***N5* (2)**	20.50	19.31	19.02	7.93	15.41	16.90	24.82	12.38	23.83	25.10
***S5* (1)**	0.80	0.67	1.32	0.94	1.40	1.33	1.17	1.13	1.06	1.27
***S5* (2)**	0.78	0.64	1.31	0.93	1.38	1.32	1.16	1.11	1.06	1.27
***K5* (1)**	0.13	0.15	0.42	0.12	0.12	0.43	0.21	0.10	0.02	0.04
***K5* (2)**	0.12	0.15	0.48	0.13	0.12	0.50	0.24	0.11	0.03	0.04
***EPNL* (1)**	88.77	88.88	83.81	69.83	81.70	82.16	86.24	77.01	91.90	94.33
***EPNL* (2)**	74.90	75.38	70.48	55.91	68.61	68.70	73.06	63.69	78.54	78.83
***SELA* (1)**	88.33	89.11	82.41	71.04	79.90	79.90	80.53	76.99	89.97	92.55
***SELA* (2)**	75.73	76.52	69.81	58.45	67.31	67.94	72.79	64.40	77.38	77.75
***PL* (1)**	101.01	101.24	94.92	84.44	91.84	93.10	97.47	88.54	103.22	106.06
***PL* (2)**	87.75	87.89	82.32	73.07	79.96	80.91	84.73	76.82	89.87	90.17

Signal / Metric	11	12	13	14	15	16	17	18	19	20
***N5* (1)**	28.48	65.90	72.28	27.77	42.79	28.07	32.57	45.56	48.15	69.18
***N5* (2)**	12.59	29.58	33.10	12.14	18.94	12.37	14.49	19.59	21.86	31.58
***S5* (1)**	1.23	1.29	1.41	1.13	1.09	1.23	1.13	0.74	1.43	1.44
***S5* (2)**	1.20	1.30	1.42	1.10	1.08	1.19	1.11	0.69	1.43	1.45
***K5* (1)**	0.46	0.11	0.16	0.10	0.03	0.46	0.25	0.41	0.06	0.19
***K5* (2)**	0.54	0.11	0.18	0.11	0.04	0.53	0.29	0.43	0.07	0.21
***EPNL* (1)**	78.57	91.41	91.32	78.03	86.37	78.38	80.60	86.86	86.58	90.79
***EPNL* (2)**	65.16	78.08	78.35	64.76	72.33	64.90	67.13	73.41	73.48	77.81
***SELA* (1)**	78.01	86.95	89.23	77.74	84.64	77.73	79.63	86.39	85.68	88.12
***SELA* (2)**	65.41	74.35	76.64	65.15	72.05	65.14	67.04	73.80	73.09	75.53
***PL* (1)**	89.61	101.99	101.90	88.96	98.87	89.40	91.90	99.58	97.99	101.11
***PL* (2)**	77.93	88.64	88.96	77.29	85.87	77.77	79.77	86.25	85.01	88.27

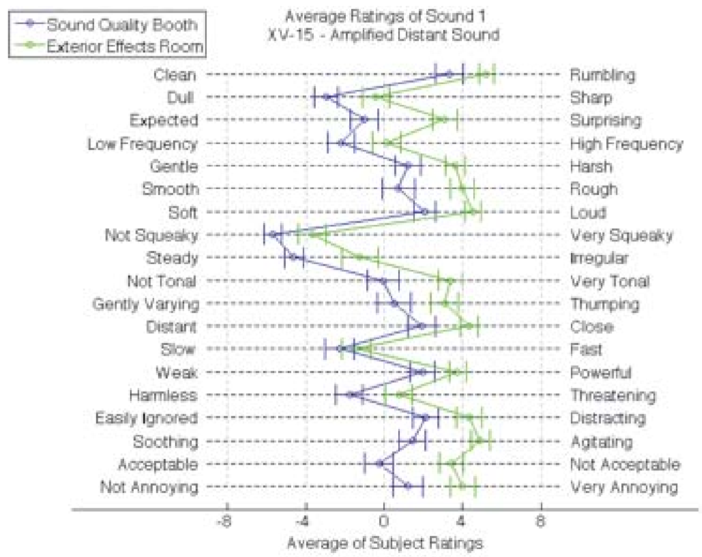

Figure A.1. Average scale ratings (+/- SEM) for Sound 1 from both the earphone test (blue) and the loudspeaker test (green).

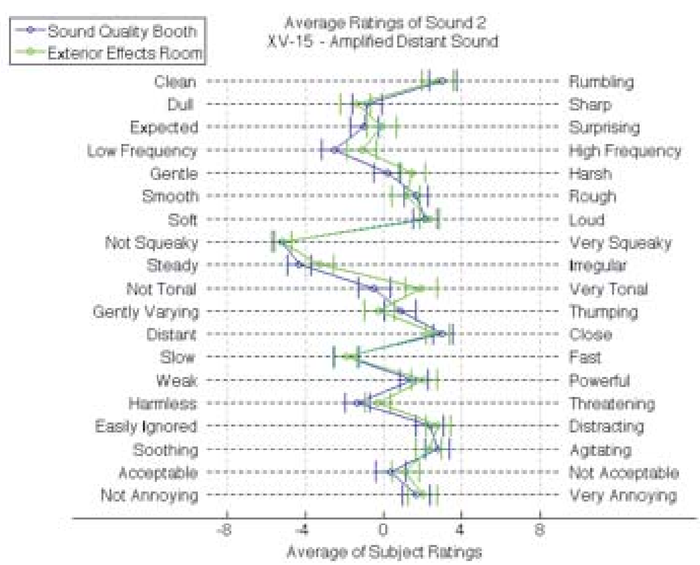

Figure A.2. Average scale ratings (+/- SEM) for Sound 2 from both the earphone test (blue) and the loudspeaker test (green).

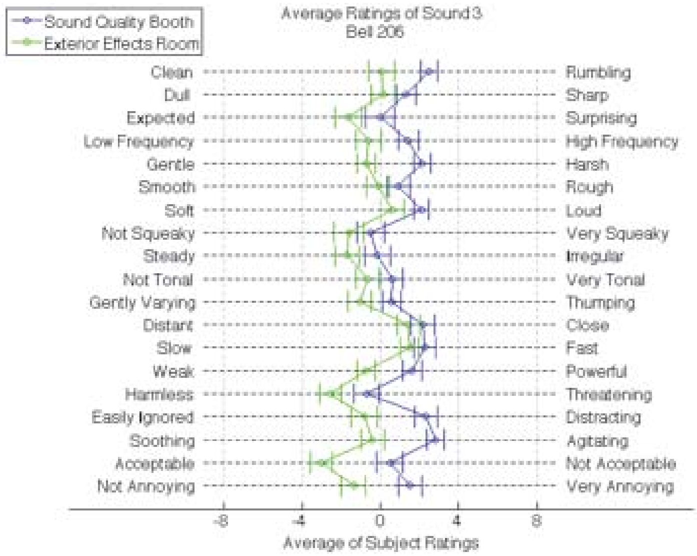

Figure A.3. Average scale ratings (+/- SEM) for Sound 3 from both the earphone test (blue) and the loudspeaker test (green).

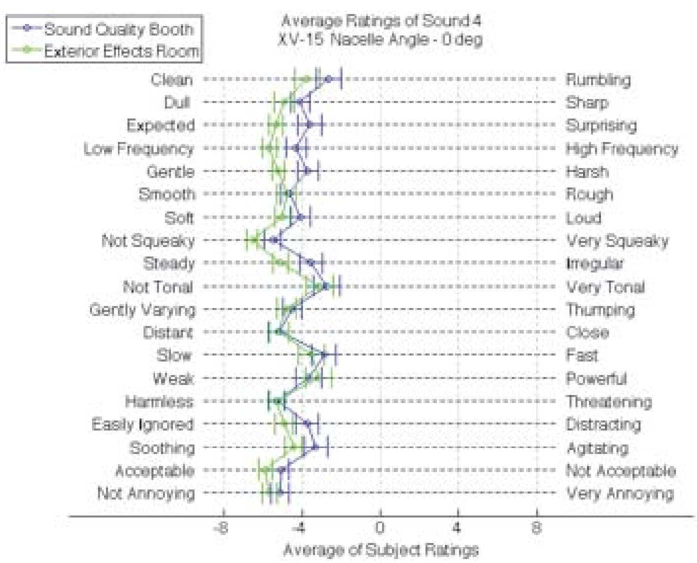

Figure A.4. Average scale ratings (+/- SEM) for Sound 4 from both the earphone test (blue) and the loudspeaker test (green).

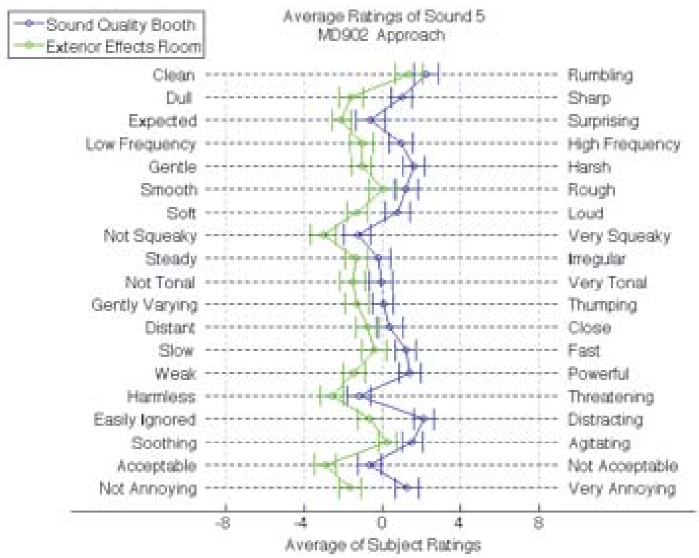

Figure A.5. Average scale ratings (+/- SEM) for Sound 5 from both the earphone test (blue) and the loudspeaker test (green).

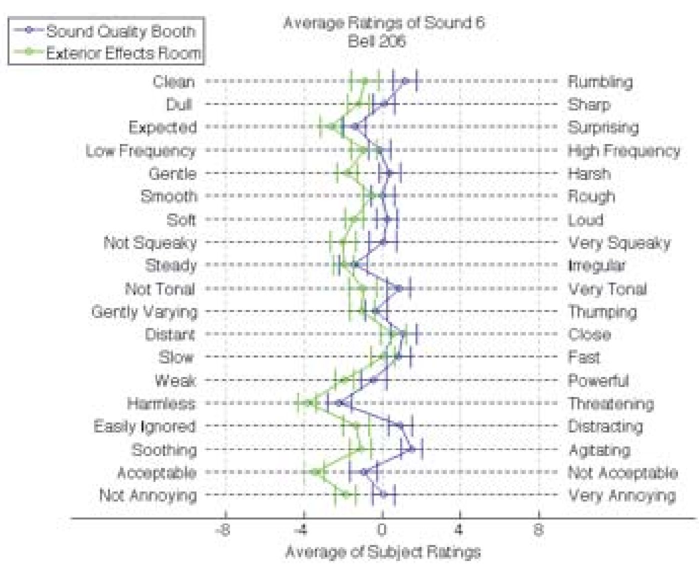

Figure A.6. Average scale ratings (+/- SEM) for Sound 6 from both the earphone test (blue) and the loudspeaker test (green).

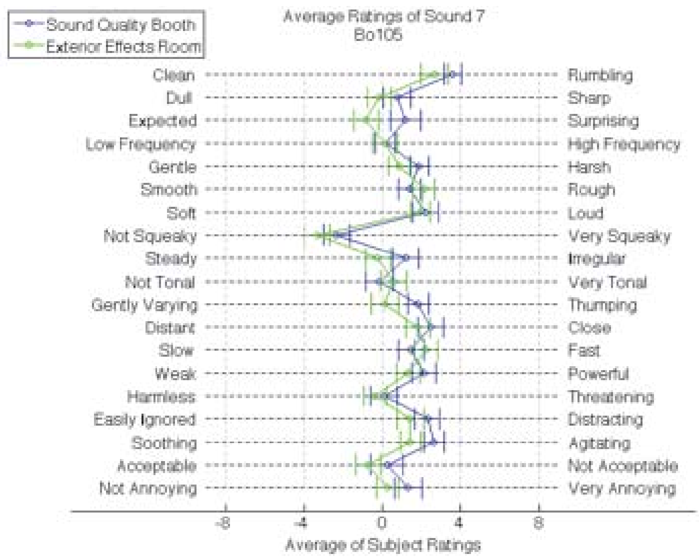

Figure A.7. Average scale ratings (+/- SEM) for Sound 7 from both the earphone test (blue) and the loudspeaker test (green).

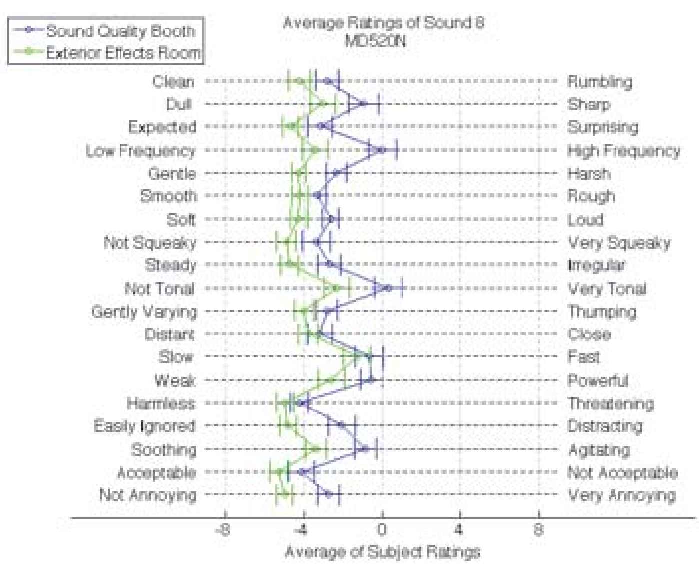

Figure A.8. Average scale ratings (+/- SEM) for Sound 8 from both the earphone test (blue) and the loudspeaker test (green).

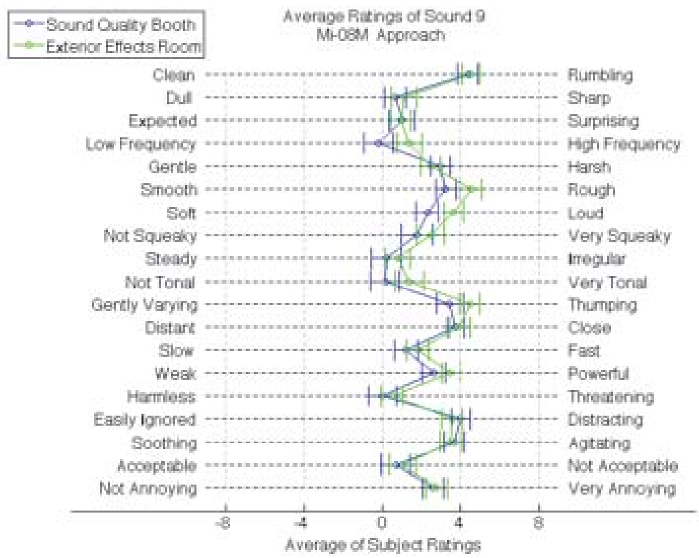

Figure A.9. Average scale ratings (+/- SEM) for Sound 9 from both the earphone test (blue) and the loudspeaker test (green).

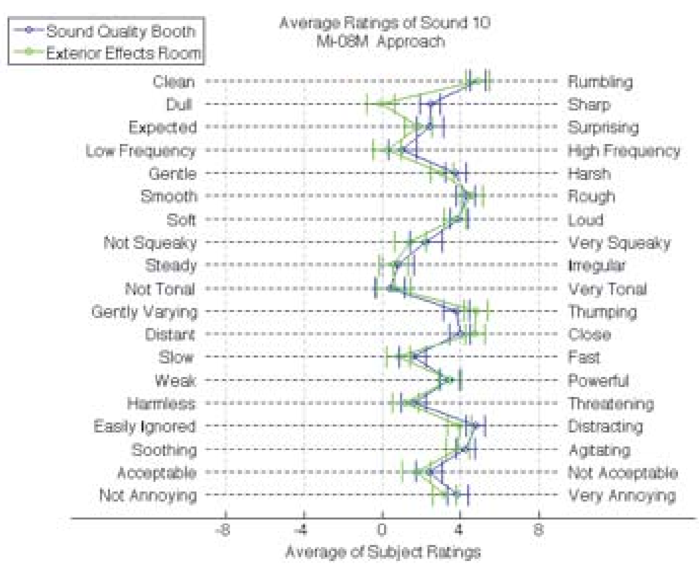

Figure A.10. Average scale ratings (+/- SEM) for Sound 10 from both the earphone test (blue) and the loudspeaker test (green).

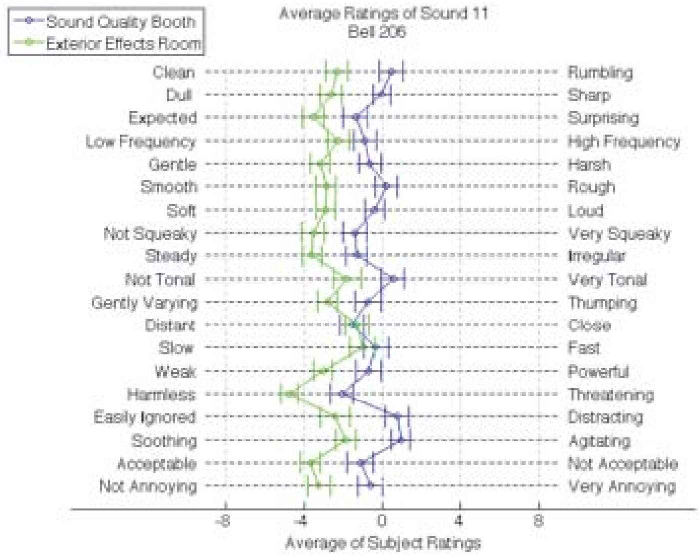

Figure A.11. Average scale ratings (+/- SEM) for Sound 11 from both the earphone test (blue) and the loudspeaker test (green).

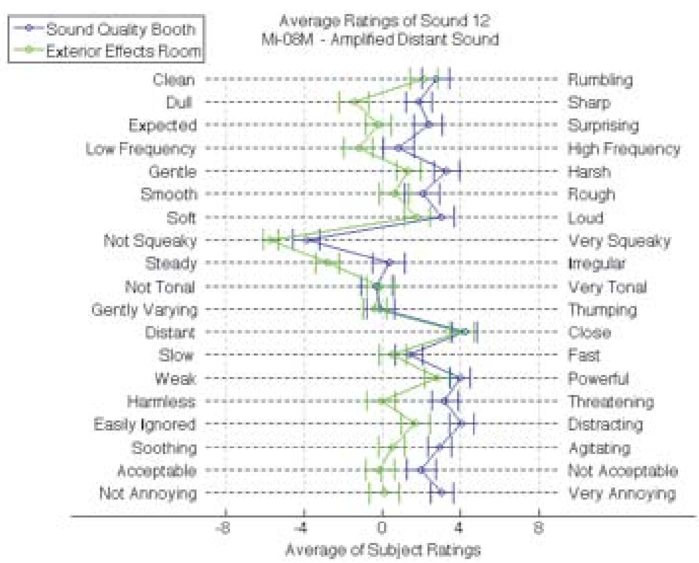

Figure A.12. Average scale ratings (+/- SEM) for Sound 12 from both the earphone test (blue) and the loudspeaker test (green).

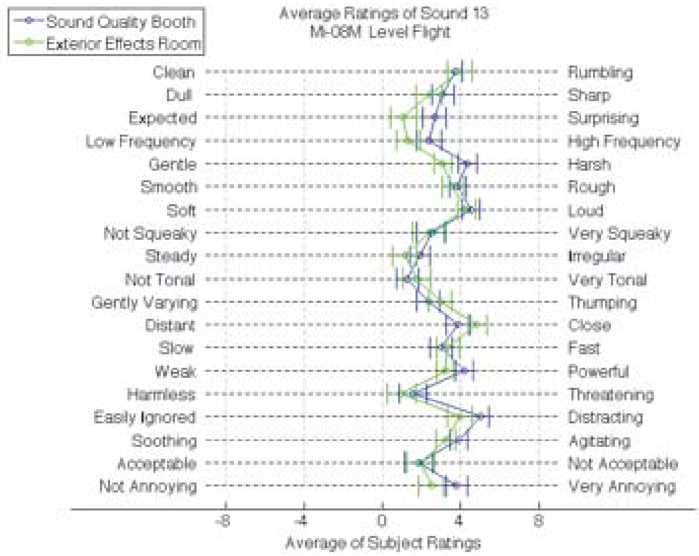

Figure A.13. Average scale ratings (+/- SEM) for Sound 13 from both the earphone test (blue) and the loudspeaker test (green).

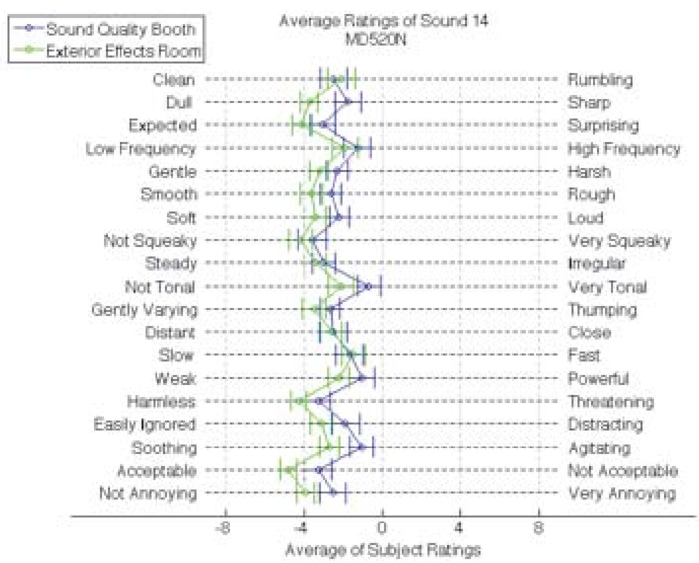

Figure A.14. Average scale ratings (+/- SEM) for Sound 14 from both the earphone test (blue) and the loudspeaker test (green).

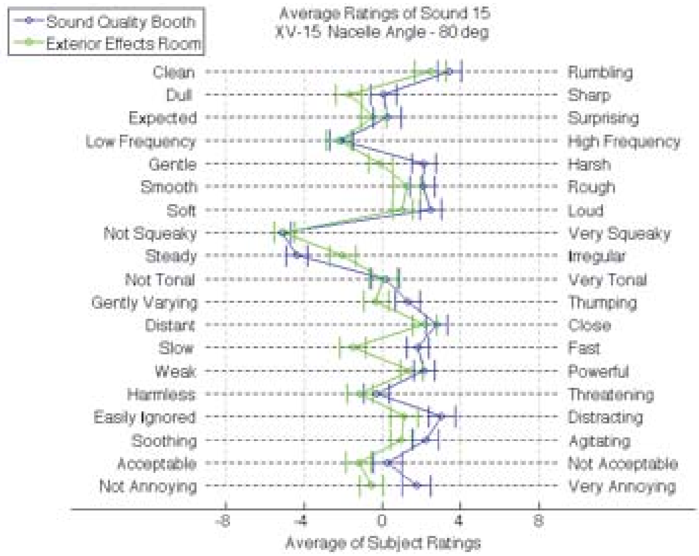

Figure A.15. Average scale ratings (+/- SEM) for Sound 15 from both the earphone test (blue) and the loudspeaker test (green).

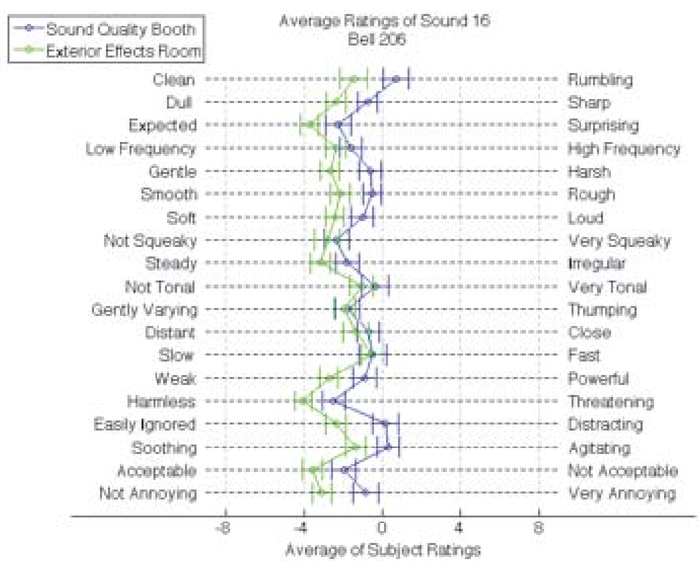

Figure A.16. Average scale ratings (+/- SEM) for Sound 16 from both the earphone test (blue) and the loudspeaker test (green).

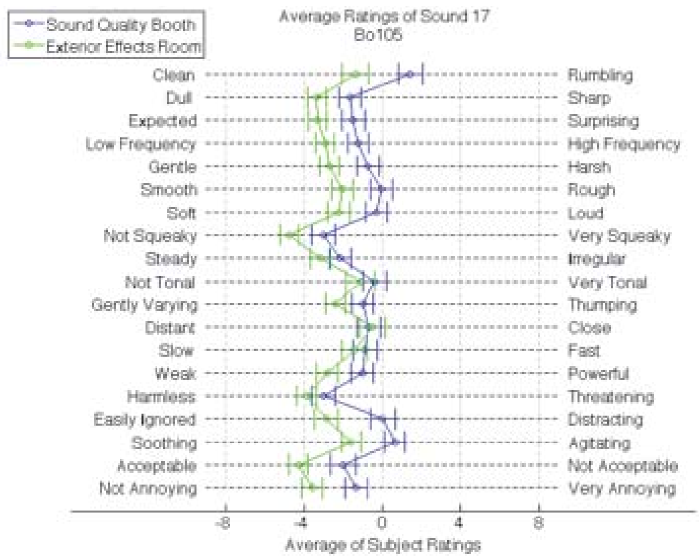

Figure A.17. Average scale ratings (+/- SEM) for Sound 17 from both the earphone test (blue) and the loudspeaker test (green).

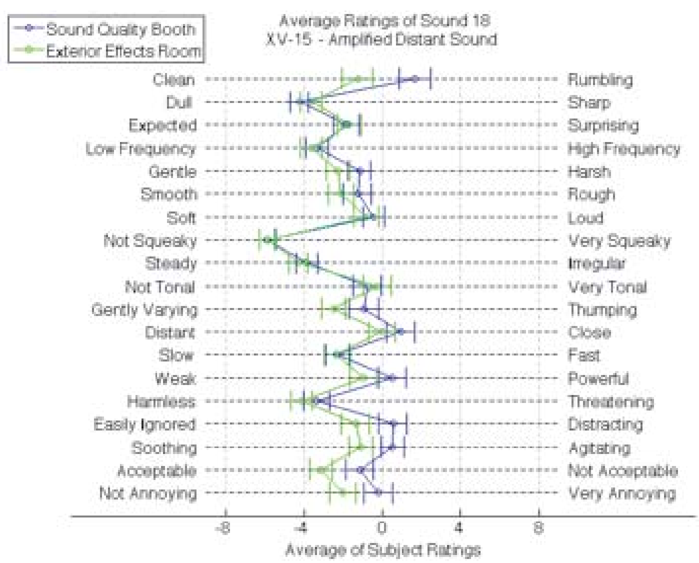

Figure A.18. Average scale ratings (+/- SEM) for Sound 18 from both the earphone test (blue) and the loudspeaker test (green).

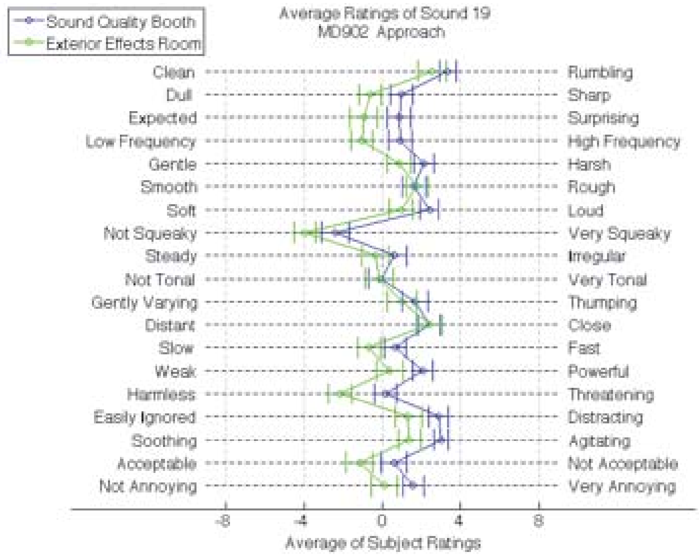

Figure A.19. Average scale ratings (+/- SEM) for Sound 19 from both the earphone test (blue) and the loudspeaker test (green).

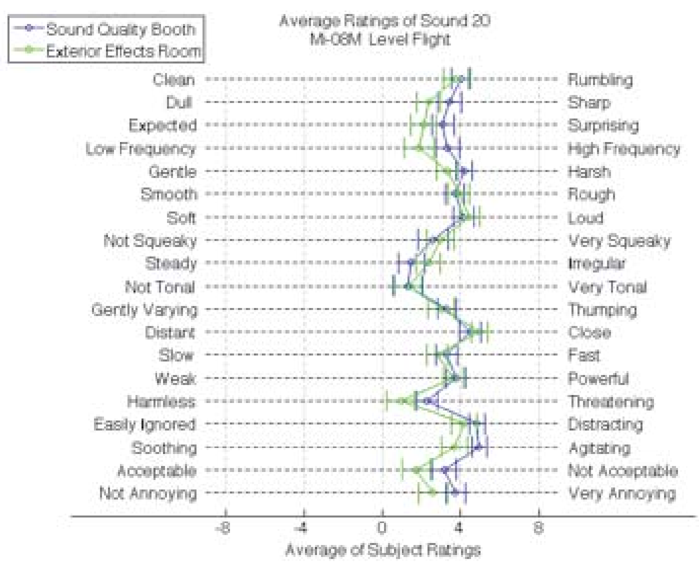

Figure A.20. Average scale ratings (+/- SEM) for Sound 20 from both the earphone test (blue) and the loudspeaker test (green).

A.2 Results From the Semantic Differential Test Using Earphones

A sample rating sheet is shown in Figure A.21. Means and standard deviations for all scale-signal combinations are shown in TableA.2. This earphone test is described in section 3.2.

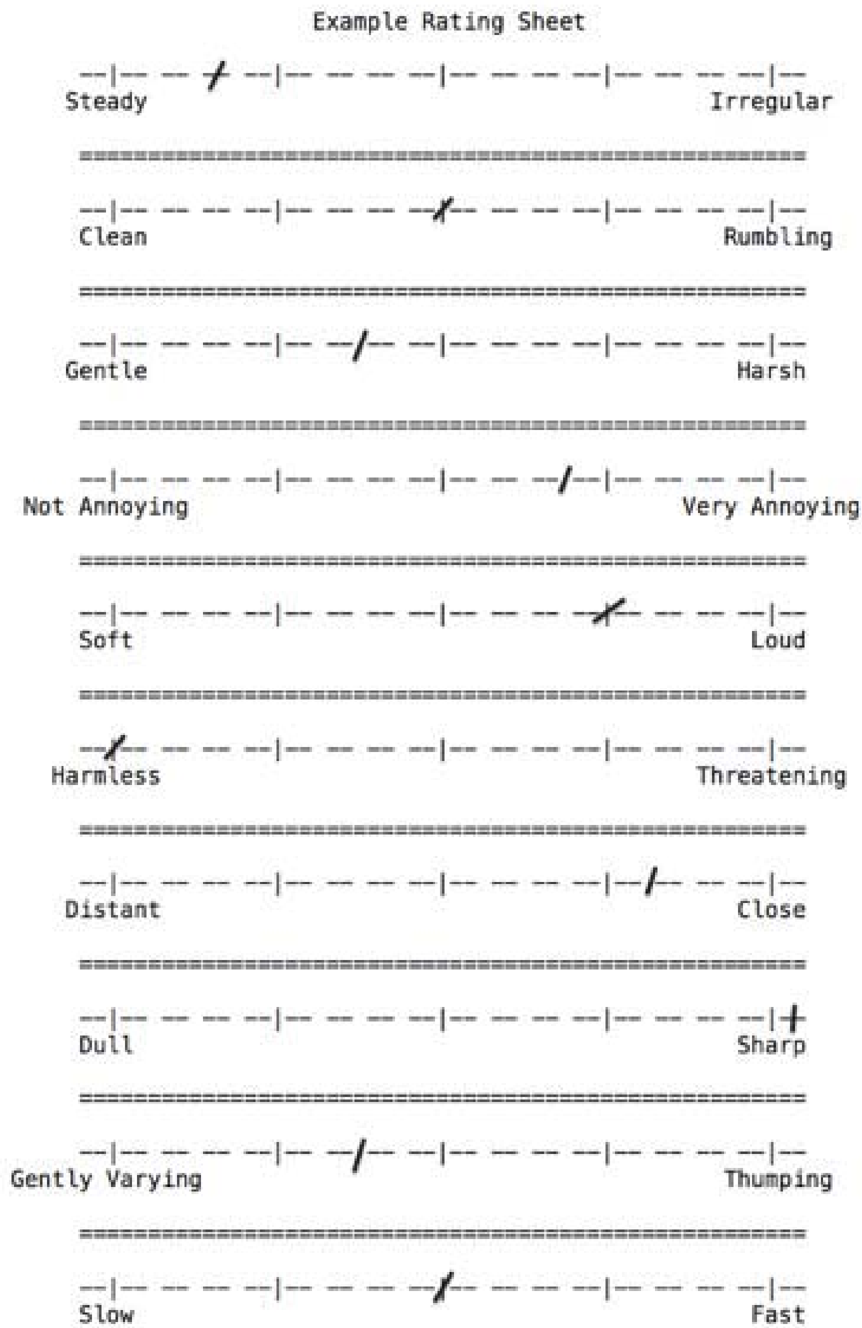

Figure A.21. Example of a completed evaluation sheet for the semantic differential earphone test. Continued on next page.

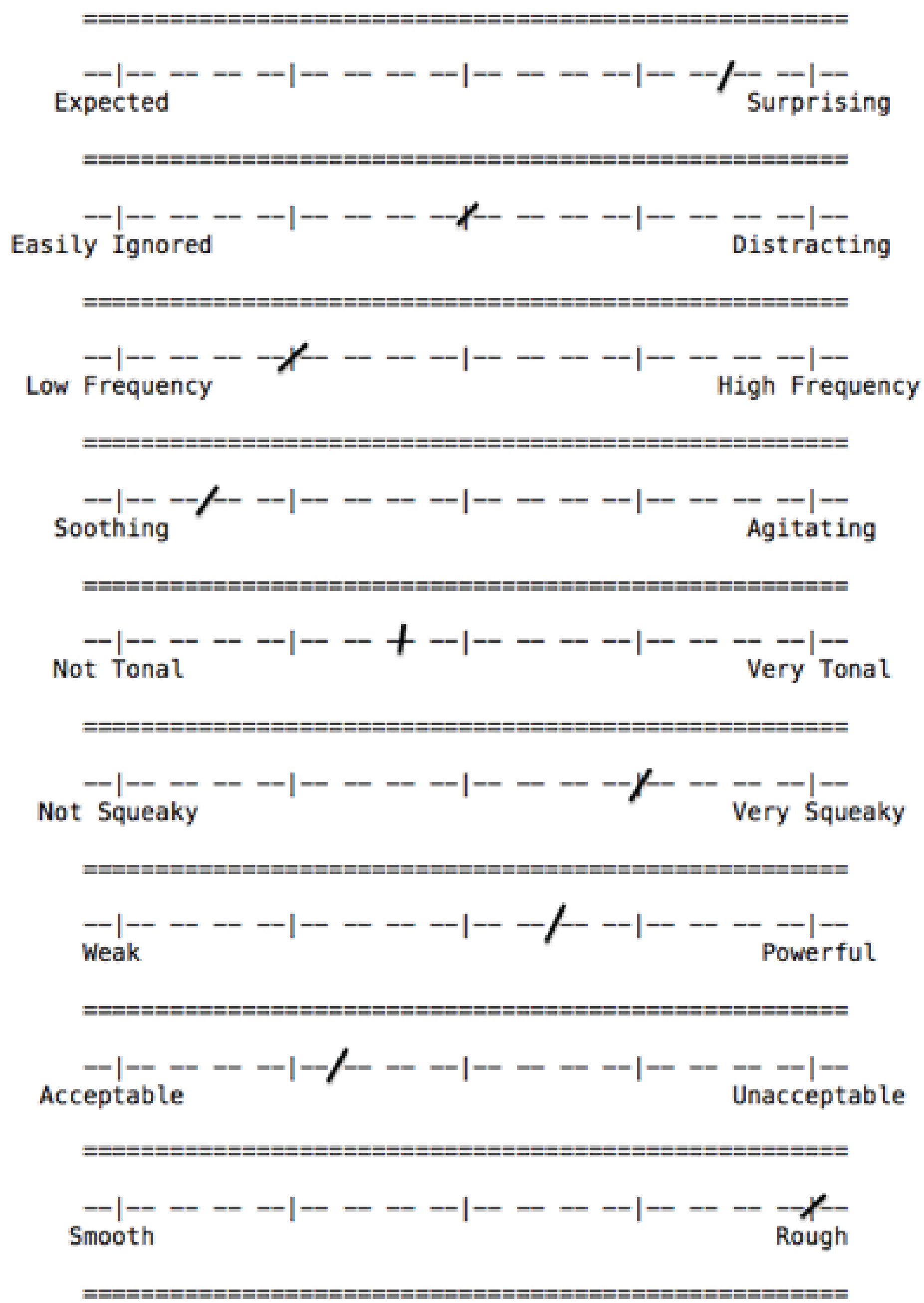
Expected
Surprising
Easily Ignored
Distracting
Low Frequency
High Frequency
Soothing
Agitating
Not Tonal
Very Tonal
Not Squeaky
Very Squeaky
Weak
Powerful
Acceptable
Unacceptable
Smooth
Rough

Table A.2. Estimated means (gray) and standard deviations (white) for scale-signal combinations for the earphone test for signals 1-10. Number of subjects = 36. Scale endpoints are -9 and +9.

Scale \ Signal	1	2	3	4	5	6	7	8	9	10
Acceptable-Not Acceptable	-0.26	0.38	0.49	-5.11	-0.65	-0.94	0.24	-4.13	0.69	2.39
	4.32	4.60	4.13	2.54	3.72	4.21	4.93	3.93	4.38	4.04
Clean-Rumbling	3.33	3.03	2.49	-2.65	2.24	1.17	3.60	-2.82	4.36	4.90
	4.30	4.11	2.54	3.84	3.82	3.58	2.59	3.71	3.23	2.48
Distant-Close	1.90	3.03	2.18	-5.18	0.38	1.08	2.47	-3.22	3.75	3.97
	4.28	3.05	3.67	3.01	4.09	4.11	4.06	3.78	2.65	3.19
Dull-Sharp	-2.96	-0.85	1.32	-4.11	0.99	0.10	0.75	-0.96	0.64	2.44
	3.64	4.55	2.97	2.79	3.29	3.16	4.12	4.59	3.28	3.09
Easily Ignored-Distracting	2.13	2.38	2.33	-3.74	2.11	0.90	2.31	-2.07	4.00	4.76
	3.93	4.19	3.42	3.47	2.98	3.58	3.75	4.26	2.98	2.82
Expected-Surprising	-1.00	-0.99	-0.01	-3.64	-0.60	-1.43	1.18	-3.19	1.01	2.43
	4.32	4.34	4.40	3.67	4.62	3.44	4.63	3.78	4.04	4.06
Gentle-Harsh	1.21	0.19	2.14	-3.72	1.61	0.38	1.88	-2.35	2.97	3.75
	3.84	4.12	2.47	3.18	3.41	3.23	2.83	3.18	3.13	3.15
Gently Varying-Thumping	0.50	0.88	0.57	-4.51	0.03	-0.33	1.82	-2.83	3.44	3.78
	5.16	4.82	2.87	2.83	3.18	3.48	3.20	3.30	4.04	4.01
Harmless-Threatening	-1.76	-1.35	-0.69	-5.28	-1.22	-2.19	0.08	-4.24	0.00	1.60
	4.08	3.85	3.87	2.46	3.56	3.89	4.04	2.85	4.38	3.81
Low Frequency-High Frequency	-2.21	-2.53	1.42	-4.32	0.96	-0.14	0.17	0.00	-0.18	1.03
	4.04	3.96	3.20	3.17	3.67	3.28	3.02	4.27	4.54	4.35
Not Annoying-Very Annoying	1.19	1.67	1.51	-5.14	1.28	0.04	1.31	-2.75	2.58	3.85
	4.53	4.25	3.72	2.63	3.62	3.32	4.18	3.42	3.25	3.22
Not Squeaky-Squeaky	-5.69	-5.17	-0.51	-5.51	-1.28	0.03	-2.36	-3.39	1.76	2.24
	2.69	2.52	4.22	2.66	4.05	4.03	4.09	4.18	4.76	4.89
Not Tonal-Very Tonal	-0.06	-0.49	0.58	-2.76	-0.07	0.85	-0.15	0.28	0.15	0.40
	4.92	4.98	3.52	4.08	3.43	3.45	4.30	4.23	4.39	4.53
Slow-Fast	-2.29	-1.90	2.26	-2.88	1.19	0.82	1.46	-0.65	1.19	1.54
	4.31	3.98	3.36	3.78	3.18	3.72	3.88	4.10	3.69	4.04
Smooth-Rough	0.72	1.67	0.93	-4.71	1.21	0.01	1.40	-3.33	3.24	4.27
	4.92	3.72	3.58	2.36	3.51	3.53	3.44	2.64	3.21	2.84
Soft-Loud	2.07	2.10	2.10	-4.10	0.78	0.24	2.22	-2.64	2.31	3.94
	3.25	3.60	2.22	2.99	3.68	3.15	3.88	2.52	3.51	2.68
Soothing-Agitating	1.44	2.75	2.82	-3.32	1.53	1.50	2.64	-0.85	3.69	4.25
	4.02	3.78	2.75	3.45	3.17	3.15	3.22	3.39	3.02	2.81
Steady-Irregular	-4.61	-4.35	-0.14	-3.56	-0.19	-1.49	1.19	-2.71	0.19	0.79
	2.93	3.60	3.92	3.48	3.46	4.30	4.00	3.58	4.47	4.90
Weak-Powerful	1.96	1.53	1.63	-3.67	1.39	-0.46	2.11	-0.57	2.61	3.44
	3.76	4.45	2.93	3.87	3.34	3.89	3.53	3.38	3.57	2.82

Table A.3. Estimated means (gray) and standard deviations (white) for scale-signal combinations for the earphone test for signals 11-20. Number of subjects = 36. Scale endpoints are -9 and +9.

Scale \ Signal	11	12	13	14	15	16	17	18	19	20
Acceptable-Not Acceptable	-1.11	1.97	1.88	-3.29	0.26	-1.99	-2.03	-1.17	0.63	3.15
	4.09	4.37	4.28	4.02	4.40	3.86	3.96	4.26	3.96	3.78
Clean-Rumbling	0.44	2.72	3.72	-2.50	3.43	0.71	1.43	1.67	3.35	4.03
	3.69	4.26	2.30	4.15	3.61	3.94	3.71	4.95	2.47	2.88
Distant-Close	-1.56	4.22	3.88	-2.53	2.78	-0.72	-0.71	0.94	2.40	4.47
	3.67	3.80	3.70	4.17	3.38	3.42	3.75	4.40	3.42	3.31
Dull-Sharp	-0.03	1.85	3.10	-1.76	0.04	-0.78	-1.65	-4.25	0.97	3.43
	2.62	3.96	3.58	4.03	3.90	3.04	3.28	2.97	3.48	3.50
Easily Ignored-Distracting	0.75	4.06	5.03	-1.90	3.04	0.15	0.01	0.54	2.86	4.83
	3.64	3.88	2.58	4.33	4.02	3.87	3.80	4.31	3.04	2.35
Expected-Surprising	-1.38	2.35	2.67	-3.07	0.24	-2.26	-1.50	-1.88	0.86	3.10
	3.80	4.39	3.69	3.76	4.22	3.89	3.46	3.97	3.61	3.59
Gentle-Harsh	-0.64	3.29	4.34	-2.32	2.13	-0.63	-0.74	-1.18	2.11	4.18
	3.43	3.91	2.85	3.34	3.61	3.39	3.45	3.59	2.97	2.29
Gently Varying-Thumping	-0.75	-0.08	2.35	-2.65	1.29	-1.79	-1.01	-0.97	1.64	3.29
	4.10	4.26	3.35	3.06	3.78	3.77	3.28	4.52	3.93	2.84
Harmless-Threatening	-2.08	3.19	1.57	-3.29	-0.31	-2.51	-3.03	-3.38	0.17	2.29
	3.45	3.99	4.28	3.60	4.02	3.50	3.64	4.04	3.31	3.22
Low Frequency-High Frequency	-0.89	0.82	2.39	-1.28	-2.11	-1.63	-1.24	-3.35	0.90	3.35
	3.57	4.67	3.75	4.13	3.70	3.35	3.21	3.57	3.65	3.62
Not Annoying-Very Annoying	-0.63	3.03	3.79	-2.54	1.75	-0.86	-1.34	-0.22	1.57	3.75
	3.80	3.46	3.11	4.15	4.32	4.02	3.44	4.35	3.09	2.97
Not Squeaky-Squeaky	-1.40	-3.90	2.50	-3.60	-5.14	-2.35	-3.03	-5.90	-2.42	2.58
	3.73	4.21	4.36	4.44	2.57	3.70	3.52	2.33	4.18	4.68
Not Tonal-Very Tonal	0.54	-0.32	1.28	-0.68	0.17	-0.36	-0.39	-0.78	-0.06	1.29
	3.49	4.79	3.54	3.50	4.19	4.04	3.69	4.26	3.76	4.31
Slow-Fast	-0.33	1.36	3.00	-1.68	1.82	-0.50	-0.89	-2.31	0.64	3.28
	3.83	4.31	3.54	4.03	3.42	4.18	3.80	3.74	3.36	3.22
Smooth-Rough	0.22	2.06	3.85	-2.64	2.04	-0.53	-0.04	-1.28	1.63	3.72
	3.35	5.30	2.53	3.20	3.64	2.92	3.34	4.40	3.68	2.84
Soft-Loud	-0.40	3.04	4.54	-2.22	2.47	-1.01	-0.29	-0.46	2.40	4.15
	2.99	3.72	2.68	3.08	3.23	3.48	3.27	3.37	2.84	3.19
Soothing-Agitating	0.94	2.92	3.90	-1.06	2.21	0.31	0.65	0.51	3.00	4.96
	2.91	3.72	2.59	3.71	3.99	3.41	3.17	3.47	2.37	2.34
Steady-Irregular	-1.33	0.35	1.93	-3.01	-4.36	-1.79	-2.17	-3.88	0.63	1.47
	3.23	4.99	3.08	3.47	3.33	3.43	3.18	3.34	3.88	4.09
Weak-Powerful	-0.71	3.97	4.18	-1.07	2.13	-0.90	-1.03	0.51	2.07	3.77
	3.80	3.11	2.84	3.97	3.22	3.61	3.59	4.29	2.98	2.83

A.3 Results From the Semantic Differential Test Using Loudspeakers

A sample rating sheet is shown in Figure A.22. Means and standard deviations for all scale-signal combinations are shown in A.4. This loudspeaker test is described in section 3.3.

Table A.4. Estimated means (gray) and standard deviations (white) for scale-signal combinations for the loudspeaker test for signals 1-10. Number of subjects = 39. Scale endpoints are -9 and +9.

Scale \ Signal	1	2	3	4	5	6	7	8	9	10
Acceptable-Not Acceptable	3.44	1.10	-3.02	-5.91	-2.92	-3.47	-0.74	-5.30	1.01	1.74
	3.75	4.29	3.56	2.21	3.36	3.07	4.03	2.79	4.25	4.71
Clean-Rumbling	5.21	2.75	0.04	-3.75	1.37	-0.90	2.67	-4.24	4.54	4.90
	2.39	5.07	4.07	3.92	4.34	4.41	4.29	3.48	2.88	3.71
Distant-Close	4.34	2.77	1.42	-5.23	-0.76	0.54	1.82	-3.79	3.95	4.76
	2.76	3.67	3.62	3.32	3.89	4.02	3.79	2.96	3.25	3.11
Dull-Sharp	-0.43	-1.43	0.14	-4.96	-1.62	-1.20	-0.14	-3.02	1.12	-0.08
	4.22	4.91	3.70	3.15	3.83	3.41	3.83	4.05	4.08	4.38
Easily Ignored-Distracting	4.32	2.79	-0.83	-4.93	-0.64	-1.34	1.43	-4.82	3.56	3.94
	4.12	3.92	4.25	2.72	3.87	4.27	4.15	2.55	3.34	3.82
Expected-Surprising	3.08	-0.10	-1.65	-5.35	-2.10	-2.61	-0.84	-4.69	0.95	1.84
	3.95	4.60	4.23	2.28	3.02	3.49	3.97	2.57	3.24	4.16
Gentle-Harsh	3.62	1.45	-0.72	-5.25	-1.06	-1.81	0.88	-4.29	2.47	3.03
	3.19	4.31	3.04	1.86	3.19	3.06	3.67	2.27	3.18	3.61
Gently Varying-Thumping	3.06	-0.23	-1.07	-4.83	-1.32	-1.05	0.11	-4.03	4.48	4.80
	4.37	4.84	3.77	3.01	3.83	4.01	4.23	3.00	3.34	3.76
Harmless-Threatening	0.74	-0.30	-2.58	-5.34	-2.59	-3.84	-0.38	-4.96	0.50	1.17
	4.22	4.08	3.50	2.58	3.67	3.02	3.54	2.64	3.36	4.26
Low Frequency-High Frequency	0.12	-1.10	-0.63	-5.67	-1.07	-0.98	0.13	-3.45	1.38	0.26
	4.56	4.72	4.10	2.27	3.76	4.07	3.97	4.11	3.92	4.46
Not Annoying-Very Annoying	4.00	1.99	-1.38	-5.68	-1.67	-1.87	0.25	-5.00	2.78	3.18
	3.88	4.45	3.95	2.33	3.51	3.25	3.46	2.55	3.54	3.87
Not Squeaky-Squeaky	-3.68	-5.24	-1.60	-6.57	-3.04	-2.02	-3.34	-4.91	2.45	1.42
	4.48	3.19	4.82	1.81	4.11	4.15	3.87	2.96	4.42	5.14
Not Tonal-Very Tonal	3.38	1.90	-0.64	-3.11	-1.53	-1.00	0.57	-2.34	1.39	0.58
	3.70	5.00	3.76	4.23	3.96	4.35	4.31	4.09	4.63	5.38
Slow-Fast	-1.44	-1.93	1.49	-3.53	-0.41	0.03	2.24	-1.28	1.81	0.81
	4.41	3.68	2.93	4.11	3.90	3.52	3.68	4.35	3.57	3.83
Smooth-Rough	3.98	1.15	-0.12	-4.79	0.03	-0.50	2.13	-4.23	4.54	4.58
	3.79	4.54	3.58	3.00	4.29	3.06	3.53	2.70	3.52	3.49
Soft-Loud	4.55	2.38	0.62	-5.01	-1.29	-1.45	1.94	-4.26	3.65	3.73
	2.48	3.26	3.77	2.37	3.18	2.84	2.94	2.74	3.13	3.42
Soothing-Agitating	4.89	2.31	-0.40	-4.46	0.25	-1.11	1.43	-3.40	3.49	3.87
	2.98	4.08	3.65	2.80	2.97	3.50	3.23	3.31	3.37	3.67
Steady-Irregular	-1.24	-3.27	-1.73	-5.12	-1.34	-1.98	-0.23	-4.73	0.80	0.57
	5.81	4.43	3.78	2.30	3.54	3.31	3.84	2.85	3.92	4.52
Weak-Powerful	3.73	2.05	-0.74	-3.17	-1.46	-1.94	1.29	-2.60	3.49	3.47
	2.69	4.02	3.02	4.17	3.57	2.89	3.79	4.28	2.93	3.62

Table A.5. Estimated means (gray) and standard deviations (white) for scale-signal combinations for the loudspeaker test for signals 11-20. Number of subjects = 39. Scale endpoints are -9 and +9.

Scale \ Signal	11	12	13	14	15	16	17	18	19	20
Acceptable-Not Acceptable	-3.67	-0.13	1.96	-4.84	-1.22	-3.57	-4.31	-3.15	-1.18	1.73
	3.03	4.59	4.30	2.68	4.09	3.12	2.98	3.63	4.13	4.55
Clean-Rumbling	-2.33	2.13	3.94	-2.13	2.46	-1.46	-1.38	-1.28	2.54	3.75
	3.66	4.42	3.96	4.36	4.93	4.53	4.36	4.86	4.51	3.85
Distant-Close	-1.31	4.05	4.85	-2.62	2.16	-1.35	-0.51	-0.01	2.42	4.97
	3.83	3.96	3.03	3.54	3.82	4.11	4.12	4.29	4.01	2.68
Dull-Sharp	-2.61	-1.45	2.40	-3.74	-1.72	-2.37	-3.37	-3.65	-0.62	2.38
	3.50	4.70	4.16	2.72	4.03	3.24	2.92	3.24	3.65	3.88
Easily Ignored-Distracting	-2.42	1.67	3.96	-3.13	1.12	-2.37	-2.88	-1.37	1.32	4.17
	4.69	4.56	4.05	3.61	4.32	3.32	3.83	4.58	4.24	4.13
Expected-Surprising	-3.55	-0.18	1.08	-4.12	-0.46	-3.72	-3.33	-1.72	-0.97	2.10
	3.22	4.13	4.25	3.06	4.03	2.78	2.94	3.88	4.32	4.00
Gentle-Harsh	-3.19	1.32	3.07	-3.24	-0.09	-2.68	-2.72	-2.26	0.86	3.31
	3.07	3.93	2.89	3.08	3.79	3.00	3.06	3.66	3.81	3.36
Gently Varying-Thumping	-2.83	-0.40	2.95	-3.50	-0.33	-1.95	-2.43	-2.52	1.01	3.00
	3.20	3.68	3.60	3.58	4.07	3.32	3.14	3.67	4.77	4.18
Harmless-Threatening	-4.77	-0.07	0.98	-4.28	-1.16	-4.06	-3.90	-4.12	-2.15	0.94
	2.78	4.63	4.48	2.61	4.14	2.57	3.12	3.47	3.75	4.56
Low Frequency-High Frequency	-2.29	-1.23	1.32	-1.89	-2.24	-2.39	-2.95	-3.64	-1.06	1.85
	3.48	4.45	3.78	4.08	3.82	3.37	3.02	3.27	3.57	4.74
Not Annoying-Very Annoying	-3.26	0.11	2.51	-3.93	-0.56	-3.11	-3.59	-2.03	0.11	2.59
	3.65	4.65	3.89	2.85	3.72	2.95	3.26	4.19	4.15	4.45
Not Squeaky-Squeaky	-3.55	-5.72	2.37	-4.19	-5.03	-2.83	-4.78	-5.87	-3.97	2.99
	3.51	2.24	5.10	3.72	3.30	4.00	2.97	2.76	3.37	4.43
Not Tonal-Very Tonal	-1.82	-0.12	1.79	-2.10	0.07	-1.05	-1.12	-0.27	-0.16	1.35
	4.29	4.40	4.41	4.13	4.40	3.86	4.48	4.58	4.55	4.63
Slow-Fast	-1.03	0.52	3.34	-1.49	-1.51	-0.57	-1.43	-2.36	-0.68	2.80
	4.01	4.28	3.68	3.68	3.98	3.47	4.04	3.95	3.72	3.37
Smooth-Rough	-2.89	0.60	3.60	-3.63	1.26	-2.16	-2.07	-2.11	1.77	3.91
	3.25	4.76	3.21	3.40	4.51	3.20	3.39	4.19	3.52	3.31
Soft-Loud	-2.91	1.79	4.31	-3.45	1.03	-2.44	-2.24	-0.82	0.98	4.46
	3.07	4.10	3.00	3.18	3.27	2.88	3.54	3.94	3.76	3.26
Soothing-Agitating	-1.88	0.49	3.23	-2.73	0.96	-1.38	-1.59	-1.11	1.37	3.67
	3.12	4.24	3.22	3.10	3.55	3.11	2.92	3.76	3.59	3.98
Steady-Irregular	-3.62	-2.83	1.15	-3.44	-2.06	-3.17	-3.18	-4.31	-0.36	2.35
	3.02	3.79	3.96	3.40	4.29	3.15	3.17	3.04	4.28	3.88
Weak-Powerful	-3.04	2.80	3.21	-2.23	1.47	-2.74	-2.82	-0.97	0.37	3.64
	3.05	4.03	2.65	3.31	3.66	2.86	3.45	4.17	4.17	3.19

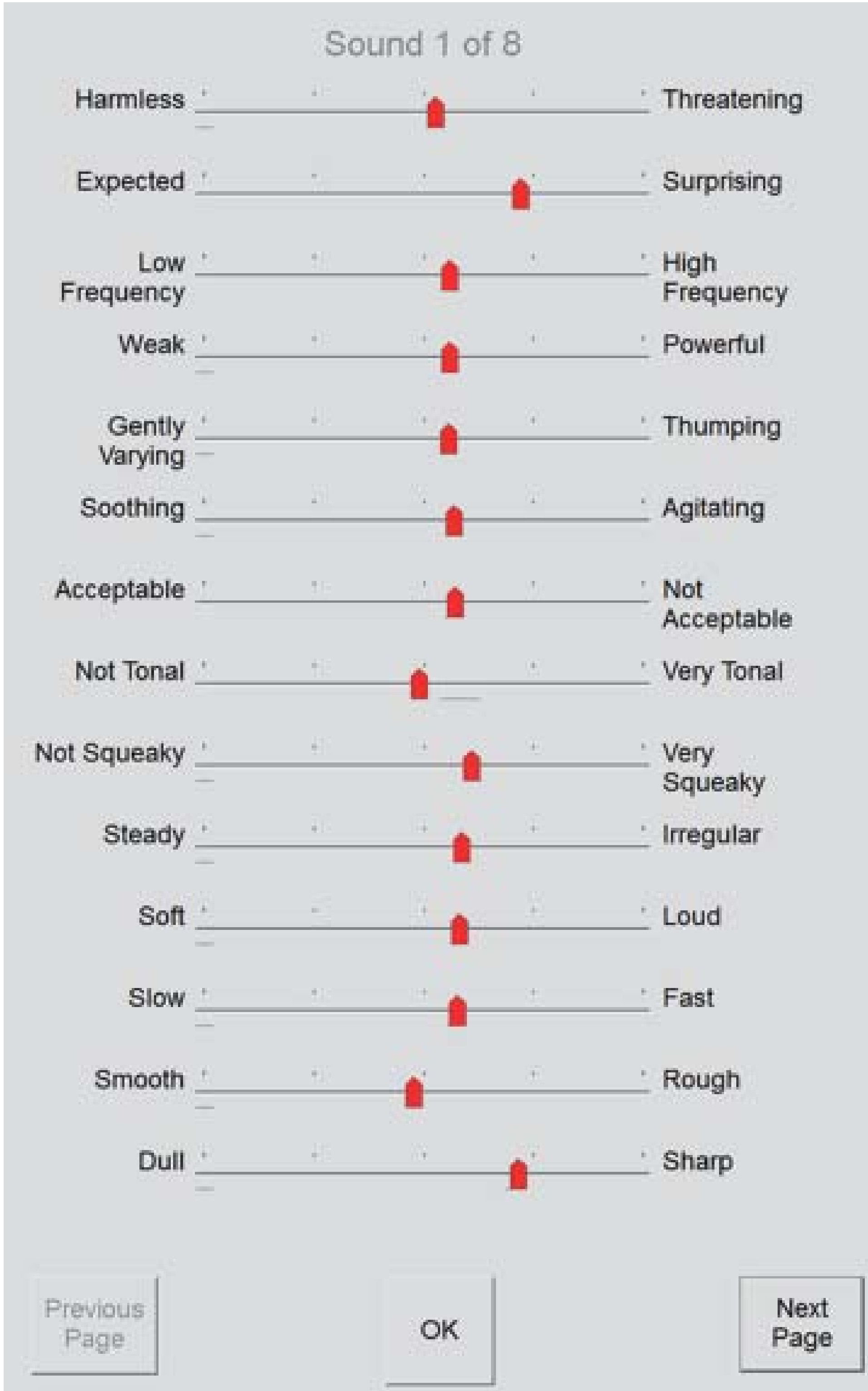

Figure A.22. Example of completed evaluation sheet for the semantic differential loudspeaker test (1 of 2).

Figure A.23. Example of completed evaluation sheet for the semantic differential loudspeaker test (2 of 2).

B. ANNOYANCE TEST - SIGNAL DETAILS, METRICS, AND SCALE RATINGS

This Appendix contains signal details, metric values, and scale ratings for the annoyance test performed using: (1) earphones and (2) using loudspeakers. These tests are discussed in Chapter 4.

Table B.1. Details of the common set of signals (1-18 of 55) for the Annoyance Test.

Signal Number	Vehicle	Tone Magnitude	Phase Change
1	Bell 206	Original	Original
2	Bell 206	50%	Original
3	Bell 206	150%	Original
4	Bell 206	150%	Aligned
5	Bell 206	Original	Random
6	BO105	Original	Original
7	BO105	50%	Original
8	BO105	150%	Original
9	BO105	150%	Aligned
10	BO105	Original	Random
11	MD520N	Original	Original
12	MD520N	50%	Original
13	MD520N	150%	Original
14	MD520N	150%	Aligned
15	MD520N	Original	Random
16	MD902	Original	Original
17	MD902	50%	Original
18	MD902	150%	Original

Table B.2. Details of the common set of signals (19-36 of 55) for the Annoyance Test.

Signal Number	Vehicle	Tone Magnitude	Phase Change
19	MD902	150%	Aligned
20	MD902	Original	Random
21	XV-15 (heli)	Original	Original
22	XV-15 (heli)	50%	Original
23	XV-15 (heli)	150%	Original
24	XV-15 (heli)	150%	Aligned
25	XV-15 (heli)	Original	Random
26	XV-15 (plane)	Original	Original
27	XV-15 (plane)	50%	Original
28	XV-15 (plane)	150%	Original
29	XV-15 (plane)	150%	Aligned
30	XV-15 (plane)	Original	Random
31	XV-15 (heli) 1/2	Original	Original
32	XV-15 (heli) 1/2	50%	Original
33	XV-15 (heli) 1/2	150%	Original
34	XV-15 (heli) 1/2	150%	Aligned
35	XV-15 (heli) 1/2	Original	Random
36	XV-15 (heli) 1/3	Original	Original

Table B.3. Details of the common set of signals (37-55 of 55) for the Annoyance Test.

Signal Number	Vehicle	Tone Magnitude	Phase Change
37	XV-15 (heli) 1/3	50%	Original
38	XV-15 (heli) 1/3	150%	Original
39	XV-15 (heli) 1/3	150%	Aligned
40	XV-15 (heli) 1/3	Original	Random
41	XV-15 (plane) 1/2	Original	Original
42	XV-15 (plane) 1/2	50%	Original
43	XV-15 (plane) 1/2	150%	Original
44	XV-15 (plane) 1/2	150%	Aligned
45	XV-15 (plane) 1/2	Original	Random
46	XV-15 (plane) 1/3	Original	Original
47	XV-15 (plane) 1/3	50%	Original
48	XV-15 (plane) 1/3	150%	Original
49	XV-15 (plane) 1/3	150%	Aligned
50	XV-15 (plane) 1/3	Original	Random
51	XV-15 (distant)	Original	Original
52	XV-15 (distant)	50%	Original
53	XV-15 (distant)	150%	Original
54	XV-15 (distant)	150%	Aligned
55	XV-15 (distant)	Original	Random

Table B.4. Calculated level metric and loudness values for the common set of signals (1-28 of 55) for the Annoyance Test. Tabulated as: EER Front Row (F), EER Back Row (B), and Purdue (P).

	EPNL			*SELA*			*PLdB*			N_5		
Signal	F	B	P	F	B	P	F	B	P	F	B	P
1	64.0	62.9	67.4	62.6	61.4	66.5	76.3	75.4	79.9	11.7	10.9	16.2
2	60.8	60.4	64.6	59.5	59.0	63.9	73.9	73.3	77.4	9.9	9.5	13.6
3	66.8	65.2	70.0	65.2	63.7	68.9	78.4	77.2	81.9	13.2	12.1	18.3
4	68.6	67.3	72.8	66.5	65.2	70.9	79.4	78.5	84.1	14.9	13.3	21.0
5	66.0	64.8	69.3	63.6	62.7	67.7	77.1	76.4	81.0	12.1	11.3	17.3
6	66.8	64.8	67.9	64.8	63.9	67.7	77.9	77.1	80.9	14.5	13.2	18.0
7	62.6	61.5	64.5	61.5	60.7	64.5	75.1	74.6	78.0	11.9	10.9	14.6
8	69.5	67.4	70.6	67.3	66.4	70.4	80.1	79.1	82.9	16.9	15.3	20.8
9	70.8	69.5	72.4	68.5	68.1	71.9	81.0	80.1	83.9	17.3	16.0	21.9
10	69.1	66.6	69.5	66.9	65.2	69.0	78.5	77.9	81.5	15.0	13.8	19.1
11	60.4	61.2	63.7	59.3	59.7	63.4	72.9	73.2	76.6	8.2	8.1	11.1
12	58.5	59.5	61.9	57.4	57.8	61.6	71.4	71.9	75.1	7.1	7.1	9.5
13	62.8	63.4	66.2	61.7	62.0	65.7	74.6	74.9	78.3	9.3	9.2	12.6
14	68.9	69.1	73.3	66.8	67.0	71.4	76.8	77.4	80.1	10.4	10.4	12.5
15	66.2	67.0	70.8	64.1	64.6	68.7	74.8	75.4	78.1	9.0	9.1	11.3
16	62.2	63.5	65.6	60.0	60.4	64.6	74.3	74.4	78.0	9.3	9.2	12.5
17	60.5	61.6	63.6	58.2	58.6	62.7	72.6	72.8	76.2	8.1	8.1	10.8
18	64.0	65.3	67.6	61.9	62.2	66.6	75.9	76.0	79.7	10.5	10.4	14.1
19	69.5	68.9	72.8	66.8	66.4	70.0	77.4	77.3	80.4	11.4	11.5	14.3
20	66.6	66.6	70.5	63.8	63.8	68.2	75.5	75.4	79.0	10.0	10.0	13.3
21	74.5	74.1	71.8	72.8	72.7	72.6	85.6	85.1	85.3	16.4	16.3	17.3
22	70.1	69.6	69.0	68.5	68.2	69.7	81.4	80.9	82.5	12.7	12.1	14.6
23	77.5	77.3	74.5	75.7	76.0	75.5	88.5	88.1	88.0	19.6	19.8	20.4
24	77.2	77.1	74.2	75.3	75.8	75.2	88.3	88.1	87.8	19.4	19.6	20.3
25	74.0	73.8	71.7	72.3	72.4	72.5	85.2	85.0	85.3	16.3	16.0	17.7
26	62.9	62.3	64.0	61.6	60.9	65.1	76.6	76.0	79.3	12.1	11.0	15.1
27	62.0	61.5	63.3	60.9	60.2	64.4	75.4	74.9	78.1	11.4	10.4	14.1
28	63.9	63.2	65.0	62.6	61.8	66.0	77.9	77.1	80.5	12.9	11.8	16.0

Table B.5. Calculated level metric and loudness values for the common set of signals (29-55 of 55) for the Annoyance Test. Tabulated as: EER Front Row (F), EER Back Row (B), and Purdue (P).

	EPNL			*SELA*			*PLdB*			N_5		
Signal	F	B	P	F	B	P	F	B	P	F	B	P
29	63.8	62.9	64.5	62.3	61.5	65.7	77.7	76.8	80.2	12.7	11.3	15.4
30	62.8	62.0	63.8	61.5	60.6	64.9	76.4	75.7	79.0	11.9	10.7	14.6
31	67.9	67.3	69.9	65.9	65.4	69.7	80.1	79.7	84.3	11.9	10.9	16.0
32	65.3	64.6	67.3	63.7	62.9	67.4	78.0	77.3	81.3	11.4	10.3	14.2
33	70.1	69.6	73.0	68.0	67.7	72.5	82.1	82.1	87.4	13.0	12.3	19.3
34	70.2	69.7	73.2	68.1	67.7	72.6	82.3	82.1	87.4	12.9	12.4	19.4
35	67.9	67.4	70.1	66.0	65.5	69.8	80.2	79.8	84.4	11.9	10.9	16.1
36	67.2	65.1	67.8	64.7	63.1	67.9	80.1	78.0	82.8	11.7	10.4	14.9
37	65.0	63.6	66.2	63.4	61.9	66.7	77.9	76.4	80.7	11.3	10.2	14.1
38	69.2	66.7	69.3	66.4	64.4	69.3	82.3	79.4	84.9	12.9	10.8	16.4
39	69.2	66.5	69.3	66.4	64.3	69.2	82.3	79.3	84.8	12.8	10.9	16.3
40	67.2	65.1	67.6	64.8	63.0	67.8	80.1	77.9	82.7	11.8	10.5	14.7
41	62.3	61.4	63.5	61.1	60.2	64.6	75.2	74.7	78.0	11.1	10.1	14.1
42	62.2	61.4	63.1	61.1	60.4	64.3	74.9	74.4	77.6	11.0	10.0	13.9
43	63.0	61.9	63.9	61.6	60.8	64.9	75.7	75.1	78.5	11.3	10.4	14.4
44	63.1	62.1	63.7	61.6	60.8	64.7	75.8	75.2	78.5	11.5	10.5	14.4
45	62.6	61.7	63.5	61.3	60.6	64.6	75.3	74.8	78.1	11.2	10.2	14.1
46	62.4	61.6	63.3	61.2	60.5	64.4	75.0	74.6	77.8	11.1	10.1	14.0
47	62.2	61.4	63.1	61.1	60.4	64.3	74.8	74.4	77.5	11.0	10.1	13.8
48	62.2	61.6	63.6	61.0	60.5	64.7	75.3	74.9	78.2	11.2	10.3	14.3
49	62.1	61.7	63.5	61.1	60.4	64.6	75.3	74.9	78.1	11.3	10.4	14.2
50	61.8	61.3	63.2	60.9	60.2	64.4	75.0	74.6	77.7	11.1	10.1	14.0
51	72.4	71.7	76.4	70.4	69.9	76.1	83.0	82.2	88.9	16.4	15.6	24.7
52	67.1	66.2	70.8	65.0	64.5	70.5	78.2	77.4	83.6	12.5	11.8	18.2
53	75.8	75.1	79.8	73.8	73.3	79.6	86.2	85.4	92.2	20.0	18.8	29.7
54	75.8	75.1	80.0	73.8	73.2	79.8	86.2	85.3	92.4	20.0	18.8	30.0
55	72.4	71.7	76.5	70.5	69.9	76.2	83.0	82.2	89.0	16.5	15.6	24.7

Table B.6. Calculated sound quality metric values for the common set of signals (1-28 of 55) for the Annoyance Test. Tabulated as: EER Front Row (F), EER Back Row (B), and Purdue (P).

	S_5			R_5			K_5			FS		
Signal	F	B	P	F	B	P	F	B	P	F	B	P
1	1.83	1.90	1.36	1.93	1.75	3.05	0.57	0.54	0.37	1.64	1.59	2.20
2	1.97	2.03	1.48	1.89	1.71	2.93	0.44	0.44	0.29	1.64	1.63	2.13
3	1.18	1.22	1.28	1.97	1.81	3.15	0.63	0.60	0.39	1.65	1.67	2.19
4	1.15	1.17	1.27	1.85	1.72	2.84	0.61	0.61	0.44	1.75	1.72	2.38
5	1.21	1.25	1.32	1.87	1.68	2.72	0.56	0.57	0.40	1.79	1.78	2.31
6	1.18	1.21	1.27	2.45	2.25	3.97	0.41	0.24	0.23	1.64	1.61	1.78
7	1.31	1.34	1.42	2.27	2.06	3.62	0.34	0.28	0.19	1.62	1.63	1.75
8	1.10	1.14	1.17	2.60	2.39	4.20	0.42	0.29	0.25	1.71	1.64	1.75
9	1.08	1.10	1.12	2.56	2.36	4.19	0.43	0.44	0.27	1.70	1.74	1.92
10	1.18	1.18	1.21	2.26	2.11	3.31	0.40	0.42	0.25	1.77	1.83	1.75
11	1.31	1.33	1.36	1.47	1.41	1.73	0.25	0.22	0.22	1.46	1.48	1.76
12	1.41	1.41	1.48	1.45	1.40	1.82	0.24	0.18	0.19	1.53	1.49	1.70
13	1.24	1.27	1.27	1.53	1.47	1.86	0.27	0.29	0.25	1.51	1.62	1.81
14	1.24	1.23	1.24	1.64	1.56	2.13	0.40	0.52	0.46	2.11	2.14	2.35
15	1.32	1.31	1.34	1.44	1.39	1.73	0.36	0.46	0.41	1.99	2.08	2.42
16	1.41	1.43	1.50	2.07	1.81	3.74	0.37	0.30	0.27	1.83	1.77	2.16
17	1.52	1.54	1.63	1.93	1.75	3.47	0.31	0.07	0.17	1.85	1.73	2.18
18	1.33	1.34	1.41	2.18	1.87	3.97	0.38	0.36	0.26	1.73	1.78	2.06
19	1.30	1.32	1.37	2.04	1.73	3.41	0.59	0.53	0.44	2.14	2.24	2.48
20	1.38	1.39	1.45	1.79	1.60	2.87	0.52	0.48	0.37	2.23	2.30	2.42
21	1.08	1.10	1.14	2.30	1.78	3.39	0.50	0.44	0.38	1.85	1.64	1.90
22	1.13	1.14	1.18	2.05	1.71	3.19	0.46	0.41	0.33	1.86	1.61	1.92
23	1.03	1.06	1.10	2.48	1.87	3.71	0.52	0.46	0.40	2.02	1.84	1.90
24	1.06	1.08	1.11	2.37	1.85	3.45	0.51	0.46	0.39	2.01	2.09	1.93
25	1.10	1.11	1.15	2.00	1.77	2.40	0.49	0.45	0.37	1.87	1.77	1.83
26	1.22	1.27	1.20	1.81	1.65	2.14	0.28	0.28	0.23	1.51	1.51	1.78
27	1.26	1.29	1.27	1.79	1.66	2.09	0.24	0.21	0.20	1.46	1.58	1.72
28	1.18	1.24	1.15	1.83	1.69	2.17	0.31	0.32	0.26	1.55	1.59	1.83

Table B.7. Calculated sound quality metric values for the common set of signals (29-55 of 55) for the Annoyance Test. Tabulated as: EER Front Row (F), EER Back Row (B), and Purdue (P).

	S_5			R_5			K_5			FS		
Signal	F	B	P	F	B	P	F	B	P	F	B	P
29	1.20	1.25	1.16	1.82	1.66	2.21	0.34	0.33	0.26	1.63	1.54	1.80
30	1.24	1.27	1.22	1.80	1.63	2.08	0.28	0.27	0.23	1.65	1.54	1.84
31	1.15	1.14	1.18	1.78	1.68	2.53	0.30	0.41	0.25	1.94	1.81	1.85
32	1.17	1.17	1.20	1.77	1.67	2.36	0.25	0.36	0.21	1.68	1.64	1.94
33	1.13	1.10	1.15	1.83	1.70	2.69	0.33	0.44	0.27	1.71	1.65	1.88
34	1.14	1.10	1.15	1.84	1.72	2.80	0.33	0.44	0.27	1.83	1.79	1.88
35	1.16	1.14	1.18	1.79	1.71	2.41	0.30	0.40	0.25	2.06	2.04	1.82
36	1.16	1.15	1.19	1.78	1.68	2.40	0.16	0.19	0.00	1.88	1.98	2.01
37	1.19	1.17	1.21	1.76	1.67	2.31	0.20	0.15	0.00	1.84	2.00	1.99
38	1.20	1.14	1.17	1.80	1.70	2.48	0.08	0.21	0.00	1.95	2.12	2.03
39	1.19	1.13	1.17	1.80	1.66	2.51	0.14	0.22	0.00	1.93	2.12	2.03
40	1.19	1.15	1.19	1.77	1.66	2.38	0.27	0.22	0.00	1.86	2.00	1.96
41	1.24	1.28	1.27	1.81	1.68	2.12	0.17	0.22	0.13	1.63	1.72	1.83
42	1.25	1.29	1.30	1.82	1.69	2.13	0.19	0.21	0.10	1.84	1.83	1.80
43	1.22	1.26	1.25	1.80	1.68	2.12	0.18	0.20	0.15	1.81	1.81	1.82
44	1.23	1.26	1.25	1.82	1.69	2.17	0.17	0.19	0.14	1.84	1.90	1.71
45	1.25	1.29	1.28	1.84	1.68	2.16	0.15	0.21	0.12	1.84	1.82	1.79
46	1.24	1.28	1.29	1.84	1.68	2.11	0.12	0.26	0.00	1.79	1.84	1.82
47	1.24	1.29	1.31	1.83	1.69	2.12	0.20	0.28	0.00	1.86	1.81	1.93
48	1.24	1.28	1.28	1.84	1.68	2.12	0.17	0.23	0.06	1.69	1.63	1.80
49	1.23	1.28	1.28	1.83	1.68	2.15	0.16	0.19	0.11	1.51	1.57	1.84
50	1.23	1.27	1.30	1.83	1.69	2.12	0.21	0.24	0.09	1.57	1.55	1.72
51	1.30	1.18	0.89	1.79	1.84	2.69	0.54	0.51	0.36	1.57	1.86	2.42
52	1.30	1.19	1.04	1.63	1.64	2.39	0.44	0.43	0.30	1.54	1.79	2.26
53	1.31	1.18	0.81	1.94	2.00	2.95	0.58	0.55	0.40	1.79	1.86	3.19
54	1.30	1.18	0.81	2.00	2.07	3.38	0.57	0.54	0.40	1.59	1.88	3.67
55	1.31	1.18	0.88	1.72	1.69	2.47	0.52	0.51	0.37	1.43	1.50	1.95

B.1 Signal Details and Results From the Annoyance Test Using Earphones

The details, metric values, and subject responses for the signals specific to the earphone annoyance test are shown below. This test is discussed in Section 4.3.

Table B.8. Details of the earphone test specific set of signals (1-18 of 55) for the Annoyance Test.

Signal Number	Vehicle	Tone Magnitude	Phase Change	Other
1	XV-15 (heli)	Original	Original	Repeated
2	XV-15 (heli)	50%	Original	Repeated
3	XV-15 (heli)	150%	Original	Repeated
4	XV-15 (heli)	150%	Aligned	Repeated
5	XV-15 (heli)	Original	Random	Repeated
6	XV-15 (heli) 1/2	Original	Original	Repeated
7	XV-15 (heli) 1/2	50%	Original	Repeated
8	XV-15 (heli) 1/2	150%	Original	Repeated
9	XV-15 (heli) 1/2	150%	Aligned	Repeated
10	XV-15 (heli) 1/2	Original	Random	Repeated
11	XV-15 (heli) 1/3	Original	Original	Repeated
12	XV-15 (heli) 1/3	50%	Original	Repeated
13	XV-15 (heli) 1/3	150%	Original	Repeated
14	XV-15 (heli) 1/3	150%	Aligned	Repeated
15	XV-15 (heli) 1/3	Original	Random	Repeated
16	XV-15 (plane)	Original	Original	Repeated
17	XV-15 (plane)	50%	Original	Repeated
18	XV-15 (plane)	150%	Original	Repeated

Table B.9. Details of the earphone test specific set of signals (19-36 of 55) for the Annoyance Test.

Signal Number	Vehicle	Tone Magnitude	Phase Change	Other
19	XV-15 (plane)	150%	Aligned	Repeated
20	XV-15 (plane)	Original	Random	Repeated
21	XV-15 (plane) 1/2	Original	Original	Repeated
22	XV-15 (plane) 1/2	50%	Original	Repeated
23	XV-15 (plane) 1/2	150%	Original	Repeated
24	XV-15 (plane) 1/2	150%	Aligned	Repeated
25	XV-15 (plane) 1/2	Original	Random	Repeated
26	XV-15 (plane) 1/3	Original	Original	Repeated
27	XV-15 (plane) 1/3	50%	Original	Repeated
28	XV-15 (plane) 1/3	150%	Original	Repeated
29	XV-15 (plane) 1/3	150%	Aligned	Repeated
30	XV-15 (plane) 1/3	Original	Random	Repeated
31	Bell 206	Original	Original	LN - max
32	BO105	Original	Original	LN - max
33	MD520N	Original	Original	LN - max
34	MD902	Original	Original	LN - max
35	XV-15 (heli)	Original	Original	LN - max
36	XV-15 (heli) 1/2	Original	Original	LN - max

Table B.10. Details of the earphone test specific set of signals (37-55 of 55) for the Annoyance Test.

Signal Number	Vehicle	Tone Magnitude	Phase Change	Other
37	XV-15 (heli) 1/3	Original	Original	LN - max
38	XV-15 (plane)	Original	Original	LN - max
39	XV-15 (plane) 1/2	Original	Original	LN - max
40	XV-15 (plane) 1/3	Original	Original	LN - max
41	XV-15 (distant)	Original	Original	LN - max
42	Bell 206	Original	Original	LN - min
43	BO105	Original	Original	LN - min
44	MD520N	Original	Original	LN - min
45	MD902	Original	Original	LN - min
46	XV-15 (heli)	Original	Original	LN - min
47	XV-15 (heli) 1/2	Original	Original	LN - min
48	XV-15 (heli) 1/3	Original	Original	LN - min
49	XV-15 (plane)	Original	Original	LN - min
50	XV-15 (plane) 1/2	Original	Original	LN - min
51	XV-15 (plane) 1/3	Original	Original	LN - min
52	XV-15 (distant)	Original	Original	LN - min
53	XV-15 (distant)	Original	Original	Repeated
54	Bell 206	Original	Original	Repeated
55	BO105	Original	Original	Repeated

Table B.11. Calculated level metric and sound quality metric values for the Purdue specific set of signals (1-28 of 55) for the Annoyance Test.

Signal	$EPNL$	$SELA$	$PLdB$	N_5	S_5	R_5	K_5	FS
1	71.8	72.6	85.3	17.3	1.14	3.39	0.38	1.90
2	69.0	69.7	82.5	14.6	1.18	3.19	0.33	1.92
3	74.5	75.5	88.0	20.4	1.10	3.71	0.40	1.90
4	74.2	75.2	87.8	20.3	1.11	3.45	0.39	1.93
5	71.7	72.5	85.3	17.7	1.15	2.40	0.37	1.83
6	69.9	69.7	84.3	16.0	1.18	2.53	0.25	1.85
7	67.3	67.4	81.3	14.2	1.20	2.36	0.21	1.94
8	73.0	72.5	87.4	19.3	1.15	2.69	0.27	1.88
9	73.2	72.6	87.4	19.4	1.15	2.80	0.27	1.88
10	70.1	69.8	84.4	16.1	1.18	2.41	0.25	1.82
11	67.8	67.9	82.8	14.9	1.19	2.40	0.00	2.01
12	66.3	66.7	80.7	14.1	1.21	2.31	0.00	1.99
13	69.3	69.3	84.9	16.4	1.17	2.48	0.00	2.03
14	69.3	69.2	84.8	16.3	1.17	2.51	0.00	2.03
15	67.6	67.8	82.7	14.7	1.19	2.38	0.00	1.96
16	64.0	65.1	79.3	15.1	1.20	2.14	0.23	1.78
17	63.3	64.4	78.1	14.1	1.27	2.09	0.20	1.71
18	64.9	66.0	80.5	16.0	1.15	2.17	0.26	1.83
19	64.5	65.7	80.2	15.4	1.16	2.21	0.26	1.80
20	63.8	64.9	79.0	14.6	1.22	2.08	0.23	1.84
21	63.5	64.6	78.0	14.1	1.27	2.12	0.13	1.83
22	63.1	64.3	77.6	13.9	1.30	2.13	0.10	1.80
23	63.9	64.9	78.5	14.4	1.25	2.12	0.15	1.82
24	63.7	64.7	78.5	14.4	1.25	2.17	0.14	1.71
25	63.5	64.6	78.1	14.1	1.28	2.16	0.12	1.79
26	63.3	64.4	77.8	14.0	1.29	2.11	0.00	1.82
27	63.1	64.3	77.5	13.8	1.31	2.12	0.00	1.93
28	63.6	64.7	78.2	14.3	1.28	2.12	0.06	1.80

Table B.12. Calculated level metric and sound quality metric values for the Purdue specific set of signals (29-55 of 55) for the Annoyance Test.

Signal	$EPNL$	$SELA$	$PLdB$	N_5	S_5	R_5	K_5	FS
29	63.5	64.6	78.1	14.2	1.28	2.15	0.11	1.84
30	63.2	64.4	77.7	14.0	1.30	2.12	0.09	1.72
31	71.6	70.3	83.2	19.7	1.36	3.47	0.34	2.37
32	69.5	69.2	82.1	19.7	1.28	4.17	0.23	1.82
33	73.7	72.9	84.8	19.7	1.33	2.40	0.21	2.06
34	73.4	71.8	84.1	19.6	1.49	4.81	0.26	2.35
35	71.8	72.6	85.3	19.3	1.14	3.38	0.38	1.90
36	73.6	72.9	87.6	19.4	1.19	2.84	0.25	1.92
37	72.7	72.3	87.2	19.5	1.19	2.80	0.00	2.11
38	67.9	68.6	82.4	19.4	1.18	2.41	0.23	1.84
39	68.2	68.9	81.7	19.4	1.24	2.47	0.13	1.95
40	68.1	68.9	81.5	19.4	1.26	2.46	0.00	1.94
41	76.7	76.4	89.2	19.7	0.90	2.73	0.36	2.43
42	61.4	60.8	75.3	9.8	1.42	2.44	0.38	1.92
43	59.4	59.5	74.2	9.8	1.39	2.96	0.21	1.57
44	63.7	63.4	76.6	9.9	1.36	1.73	0.22	1.76
45	63.4	62.4	76.2	9.9	1.55	3.40	0.25	2.10
46	61.7	63.0	76.4	9.6	1.22	2.43	0.37	1.56
47	63.4	63.6	78.5	9.8	1.23	2.04	0.25	1.70
48	62.5	62.8	78.2	9.9	1.23	2.00	0.00	1.92
49	57.9	58.8	74.2	9.8	1.36	1.72	0.22	1.57
50	58.3	59.3	73.9	9.7	1.39	1.77	0.13	1.69
51	58.2	59.1	73.7	9.9	1.40	1.76	0.00	1.71
52	66.7	66.9	79.7	9.9	1.08	1.97	0.36	2.14
53	76.4	76.1	88.9	24.7	0.89	2.69	0.36	2.43
54	67.4	66.5	79.9	16.2	1.36	3.05	0.37	2.21
55	67.9	67.7	80.9	18.0	1.27	3.97	0.23	1.78

Table B.13. Average and standard deviation of responses to signals 1-40 from the earphone annoyance test.

Signal	Average	Standard Deviation	Signal	Average	Standard Deviation
1	5.22	1.20	21	5.26	1.24
2	5.26	1.27	22	5.45	1.35
3	4.50	1.17	23	4.84	1.42
4	4.96	1.29	24	4.78	1.50
5	5.43	1.24	25	5.14	1.24
6	5.45	1.11	26	4.42	1.38
7	5.07	1.04	27	4.39	1.20
8	5.65	1.26	28	4.47	1.20
9	4.02	1.07	29	5.08	1.24
10	5.20	1.09	30	6.03	1.63
11	5.73	1.33	31	4.63	1.17
12	5.51	1.36	32	5.16	1.17
13	4.14	1.43	33	5.14	1.29
14	4.51	1.40	34	5.18	1.44
15	5.86	1.41	35	5.64	1.29
16	6.03	1.46	36	5.16	1.30
17	4.33	1.06	37	4.54	1.23
18	4.90	1.60	38	5.12	0.98
19	4.37	1.47	39	4.99	1.42
20	4.84	1.18	40	5.20	1.43

Table B.14. Average and standard deviation of responses to signals 41-80 from the earphone annoyance test.

Signal	Average	Standard Deviation	Signal	Average	Standard Deviation
41	4.71	1.39	61	4.33	1.40
42	4.48	1.12	62	4.32	1.40
43	4.37	1.30	63	5.42	1.39
44	5.34	1.50	64	4.55	1.31
45	5.64	1.38	65	4.65	1.45
46	4.67	1.32	66	3.99	1.40
47	4.26	1.20	67	5.56	1.43
48	5.41	1.21	68	4.91	1.37
49	4.47	1.28	69	5.58	1.28
50	4.48	1.39	70	4.77	1.11
51	4.71	1.31	71	5.02	1.43
52	4.92	1.18	72	4.41	1.09
53	4.91	1.27	73	6.05	1.47
54	5.50	1.30	74	5.40	1.15
55	5.51	1.28	75	5.30	1.45
56	4.26	1.22	76	5.21	1.10
57	5.27	1.06	77	5.06	1.08
58	5.03	1.14	78	5.00	1.25
59	5.33	1.25	79	4.96	1.15
60	5.59	1.26	80	4.16	1.35

Table B.15. Average and standard deviation of responses to signals 81-110 from the earphone annoyance test.

Signal	Average	Standard Deviation	Signal	Average	Standard Deviation
81	5.22	1.20	96	4.56	1.37
82	5.47	1.33	97	5.68	1.32
83	5.54	1.30	98	4.41	1.46
84	5.50	1.20	99	5.39	1.25
85	4.65	1.33	100	5.10	1.31
86	5.12	1.36	101	5.21	1.39
87	4.29	1.19	102	5.65	1.17
88	4.39	1.26	103	5.28	1.21
89	5.20	1.12	104	4.80	1.17
90	5.28	1.26	105	6.00	1.50
91	5.14	1.29	106	5.16	1.39
92	5.02	1.32	107	4.49	1.25
93	5.75	1.27	108	5.22	1.35
94	4.39	1.29	109	4.43	1.19
95	4.83	1.23	110	4.40	1.24

B.2 Results From the Annoyance Test Using Loudspeakers

The details, metric values, and subject responses for the signals specific to the loudspeaker annoyance test are shown below. This test is discussed in Section 4.4.

Table B.16. Calculated level metric and sound quality metric values for the NASA specific set of signals (1-28 of 55) for the Annoyance Test. Tabulated as EER Front Row (F) and EER Back Row (B).

	$EPNL$		$SELA$		$PLdB$		N_5		S_5		R_5		K_5		FS	
Signal	F	B	F	B	F	B	F	B	F	B	F	B	F	B	F	B
1	64.2	62.9	62.7	61.4	76.3	75.4	11.7	10.8	1.26	1.29	1.93	1.75	0.58	0.54	1.62	1.59
2	60.7	60.1	59.7	59.0	73.9	73.3	9.9	9.5	1.37	1.39	1.89	1.71	0.46	0.46	1.75	1.67
3	66.8	65.2	65.3	63.7	78.4	77.1	13.2	12.0	1.18	1.21	1.96	1.80	0.63	0.59	1.63	1.64
4	68.9	67.8	66.3	65.6	79.4	78.5	14.9	13.3	1.14	1.18	1.84	1.71	0.60	0.63	2.03	1.78
5	65.9	64.7	63.6	62.6	77.1	76.4	12.1	11.2	1.21	1.25	1.86	1.67	0.57	0.57	1.79	1.78
6	66.8	64.8	64.8	63.9	78.0	77.1	14.6	13.2	1.19	1.21	2.45	2.24	0.41	0.23	1.62	1.62
7	62.5	61.3	61.3	60.6	75.1	74.6	11.9	10.9	1.31	1.35	2.28	2.06	0.37	0.29	1.66	1.73
8	69.3	67.4	67.4	66.4	80.2	79.2	16.9	15.3	1.11	1.14	2.59	2.36	0.42	0.39	1.68	1.72
9	70.8	69.6	68.5	68.3	81.0	80.1	17.4	16.0	1.09	1.12	2.58	2.36	0.40	0.42	1.72	1.74
10	67.4	66.2	65.5	65.2	78.5	77.9	14.9	13.8	1.21	1.20	2.26	2.11	0.40	0.42	1.81	1.98
11	60.3	61.2	59.3	59.7	72.9	73.2	8.2	8.1	1.31	1.33	1.46	1.40	0.25	0.22	1.53	1.60
12	58.4	59.3	57.3	57.8	71.4	71.9	7.2	7.2	1.41	1.40	1.46	1.40	0.23	0.18	1.43	1.45
13	62.8	63.2	61.6	61.9	74.6	74.9	9.3	9.2	1.22	1.26	1.52	1.47	0.29	0.30	1.45	1.53
14	69.4	69.5	67.4	66.9	76.8	77.4	10.4	10.4	1.24	1.23	1.63	1.55	0.44	0.52	1.99	2.18
15	66.3	66.4	64.3	64.3	74.8	75.4	9.0	9.1	1.31	1.29	1.44	1.39	0.43	0.48	1.85	2.05
16	62.1	63.6	59.8	60.3	74.2	74.4	9.3	9.3	1.41	1.42	2.06	1.81	0.38	0.32	1.74	1.72
17	60.4	61.5	58.1	58.5	72.6	72.8	8.1	8.1	1.52	1.53	1.92	1.75	0.30	0.17	1.81	1.69
18	64.2	65.5	61.9	62.3	75.8	76.0	10.5	10.4	1.33	1.34	2.18	1.87	0.38	0.34	1.77	1.64
19	69.4	69.5	66.8	66.3	77.4	77.3	11.4	11.5	1.31	1.31	2.03	1.72	0.57	0.52	2.17	2.06
20	66.4	66.2	63.6	63.5	75.5	75.4	10.0	10.0	1.38	1.39	1.79	1.61	0.53	0.45	2.09	1.98
21	74.5	74.1	72.8	72.7	85.6	85.1	16.5	16.3	1.08	1.09	2.30	1.78	0.50	0.44	1.76	1.59
22	70.1	69.5	68.5	68.2	81.4	80.9	12.7	12.1	1.14	1.13	2.06	1.71	0.46	0.42	1.77	1.59
23	77.5	77.3	75.6	76.0	88.6	88.1	19.6	19.8	1.02	1.05	2.48	1.87	0.52	0.46	1.78	1.59
24	77.4	77.1	75.6	75.8	88.3	88.1	19.4	19.6	1.04	1.05	2.36	1.85	0.51	0.46	1.79	1.62
25	74.0	73.6	72.4	72.4	85.2	85.0	16.3	16.0	1.10	1.10	2.00	1.77	0.49	0.44	1.81	1.70
26	63.3	62.4	61.9	61.2	76.7	76.1	12.2	11.1	1.21	1.27	1.79	1.64	0.28	0.30	1.47	1.43
27	62.4	61.6	61.1	60.5	75.4	74.9	11.3	10.3	1.25	1.30	1.77	1.62	0.23	0.23	1.45	1.44
28	64.3	63.3	62.7	62.0	78.0	77.3	13.0	11.9	1.19	1.24	1.81	1.67	0.34	0.30	1.45	1.44

Table B.17. Calculated level metric and sound quality metric values for the NASA specific set of signals (29-55 of 55) for the Annoyance Test. Tabulated as EER Front Row (F) and EER Back Row (B).

	EPNL		*SELA*		*PLdB*		N_5		S_5		R_5		K_5		*FS*	
Signal	F	B	F	B	F	B	F	B	F	B	F	B	F	B	F	B
29	64.2	63.1	62.6	61.8	77.8	77.0	12.8	11.4	1.19	1.25	1.81	1.65	0.33	0.32	1.46	1.43
30	62.8	62.2	61.5	60.8	76.5	75.9	11.9	10.7	1.22	1.27	1.80	1.63	0.28	0.27	1.52	1.53
31	67.9	67.2	65.9	65.4	80.1	79.8	11.9	10.9	1.15	1.13	1.79	1.69	0.30	0.41	1.72	1.62
32	65.3	64.7	63.7	62.9	78.0	77.3	11.4	10.3	1.17	1.17	1.76	1.67	0.26	0.36	1.70	1.68
33	70.2	69.5	68.0	67.7	82.1	82.1	12.9	12.3	1.12	1.10	1.83	1.72	0.32	0.44	1.67	1.62
34	70.2	69.7	68.1	67.7	82.3	82.2	12.9	12.4	1.13	1.10	1.84	1.75	0.32	0.43	1.65	1.60
35	68.0	67.4	66.0	65.5	80.2	79.8	11.8	10.9	1.16	1.14	1.79	1.72	0.30	0.41	1.61	1.60
36	67.1	65.2	64.7	63.1	80.2	78.1	11.7	10.4	1.17	1.16	1.77	1.66	0.15	0.22	1.88	1.86
37	65.0	63.7	63.3	62.0	77.9	76.6	11.3	10.2	1.18	1.18	1.76	1.65	0.13	0.17	1.84	1.76
38	69.2	66.6	66.4	64.5	82.4	79.6	12.9	10.8	1.15	1.14	1.79	1.67	0.10	0.27	1.74	1.76
39	69.2	66.5	66.3	64.4	82.4	79.6	12.8	10.8	1.16	1.13	1.76	1.66	0.16	0.26	1.77	1.77
40	67.2	65.0	64.7	63.0	80.2	78.1	11.8	10.5	1.17	1.16	1.76	1.63	0.13	0.24	1.83	1.71
41	62.5	61.6	61.2	60.5	75.3	74.7	11.1	10.1	1.23	1.28	1.80	1.68	0.15	0.16	1.50	1.46
42	61.8	61.2	60.9	60.1	74.9	74.4	11.0	10.0	1.24	1.29	1.81	1.69	0.20	0.20	1.52	1.50
43	62.9	61.9	61.4	60.7	75.8	75.2	11.3	10.4	1.22	1.26	1.80	1.68	0.16	0.18	1.48	1.68
44	63.0	62.0	61.6	60.8	75.8	75.3	11.5	10.5	1.22	1.25	1.82	1.67	0.15	0.15	1.50	1.47
45	62.2	61.7	61.1	60.4	75.3	74.8	11.2	10.3	1.24	1.27	1.83	1.67	0.17	0.16	1.54	1.61
46	62.0	61.4	61.0	60.2	75.0	74.6	11.1	10.1	1.25	1.29	1.82	1.65	0.14	0.24	1.60	1.54
47	61.9	61.2	60.9	60.1	74.8	74.4	11.0	10.1	1.25	1.29	1.83	1.69	0.18	0.24	1.58	1.55
48	62.3	61.7	61.2	60.4	75.3	74.9	11.2	10.3	1.24	1.30	1.83	1.65	0.14	0.22	1.59	1.56
49	62.2	61.7	61.1	60.4	75.4	74.9	11.3	10.4	1.23	1.27	1.81	1.68	0.13	0.20	1.58	1.48
50	62.0	61.3	61.0	60.1	75.0	74.6	11.1	10.1	1.25	1.27	1.83	1.68	0.18	0.26	1.61	1.55
51	72.4	71.7	70.4	69.9	83.0	82.2	16.4	15.6	1.30	1.18	1.79	1.84	0.55	0.51	1.87	1.65
52	67.1	66.2	65.0	64.5	78.2	77.4	12.5	11.8	1.30	1.19	1.62	1.65	0.45	0.44	1.38	1.49
53	75.9	75.1	73.8	73.3	86.2	85.3	20.0	18.8	1.31	1.18	1.93	2.00	0.58	0.55	1.80	1.81
54	75.8	75.1	73.9	73.2	86.2	85.3	20.0	18.8	1.31	1.19	2.00	2.06	0.56	0.54	1.96	1.85
55	72.4	71.7	70.5	69.8	83.0	82.1	16.5	15.6	1.33	1.19	1.71	1.69	0.52	0.51	1.35	1.44

Table B.18. Average and standard deviation of responses to signals 1-40 from the loudspeaker annoyance test.

Signal	Average	Standard Deviation	Signal	Average	Standard Deviation
1	4.39	1.94	21	5.93	1.58
2	5.36	1.91	22	5.93	1.91
3	4.41	1.95	23	5.65	1.97
4	5.26	1.53	24	4.88	1.76
5	4.29	1.76	25	5.03	1.60
6	5.15	1.78	26	4.36	1.84
7	5.02	1.67	27	5.49	2.17
8	4.64	1.96	28	5.34	1.80
9	4.36	1.82	29	6.32	1.72
10	5.06	1.82	30	5.54	1.70
11	5.64	2.02	31	5.62	1.76
12	4.75	2.19	32	5.42	1.83
13	4.39	1.91	33	5.06	1.76
14	4.64	2.23	34	4.02	1.78
15	5.68	1.87	35	5.39	2.04
16	3.74	1.64	36	4.66	1.89
17	4.41	1.73	37	4.38	2.04
18	4.54	1.83	38	7.26	1.83
19	4.30	1.72	39	5.62	2.17
20	4.56	1.66	40	4.86	1.75

Table B.19. Average and standard deviation of responses to signals 41-80 from the loudspeaker annoyance test.

Signal	Average	Standard Deviation	Signal	Average	Standard Deviation
41	4.66	1.93	61	5.81	2.07
42	5.86	1.74	62	6.05	1.79
43	5.25	1.42	63	4.19	1.95
44	4.27	1.61	64	7.46	1.89
45	6.52	1.74	65	4.50	2.19
46	6.04	1.76	66	4.93	1.71
47	5.56	1.68	67	6.72	1.83
48	5.37	1.84	68	6.00	2.22
49	5.69	2.19	69	5.33	1.62
50	5.03	1.66	70	5.37	1.51
51	7.25	1.62	71	5.71	1.89
52	4.40	1.87	72	4.52	1.90
53	5.60	2.08	73	4.42	1.65
54	3.71	1.47	74	4.41	1.86
55	5.77	1.94	75	4.48	1.98
56	4.26	2.00	76	6.19	1.45
57	4.48	1.96	77	5.93	1.59
58	5.35	2.08	78	5.62	1.84
59	6.39	1.55	79	5.28	2.27
60	5.35	1.81	80	4.22	2.17

Table B.20. Average and standard deviation of responses to signals 81-110 from the loudspeaker annoyance test.

Signal	Average	Standard Deviation	Signal	Average	Standard Deviation
81	6.13	1.62	96	4.68	2.16
82	5.26	2.05	97	5.28	1.88
83	5.20	1.69	98	3.62	1.72
84	5.24	2.06	99	5.17	2.17
85	5.08	1.69	100	4.66	1.81
86	4.23	1.77	101	6.22	1.96
87	5.11	1.58	102	4.21	1.53
88	5.71	1.82	103	4.26	1.77
89	4.69	2.24	104	7.02	1.71
90	5.76	1.96	105	4.78	1.67
91	4.50	1.68	106	5.35	1.76
92	4.66	1.76	107	4.84	2.04
93	5.12	1.87	108	4.52	1.99
94	4.27	1.76	109	5.28	1.93
95	4.35	1.76	110	5.22	1.97

C. EPNL DISCUSSION

Metrics analyses were performed on the two sets of semantic differential test data are and presented in sections 3.2.3.3 and 3.3.3.3. In those sections, EPNL values are presented that were calculated from the test signals only, and not the full flyovers. Calculated EPNL values can vary depending on length due to the method of calculation. This process is explained below, as specified in FAA CFR Part 36 Appendix A2 Section A36.4.

The EPNL calculation is based on the values through time of PNLT (Tone Corrected Perceived Noise Level). The calculation is:

$$EPNL = PNLTM + D \tag{C.1}$$

where PNLTM is the maximum of the PNLT time history, and D is the correction factor for duration, which is calculated by:

$$D = 10log\left[\left(\frac{1}{T}\right)\int_{t(1)}^{t(2)} 10^{\frac{PNLT}{10}}\right] - PNLTM \tag{C.2}$$

where T is the normalizing time constant, t(1) is the first point of time after which PNLT becomes greater than PNLTM-10, and t(2) is the point of time after which PNLT remains constantly less than PNLTM-10.

The discrete version of the calculation is:

$$D = 10log\left[\sum_{k=0}^{2d} 10^{\frac{PNLT(k)}{10}} dt\right] - PNLTM - 13 \tag{C.3}$$

where d is the duration time defined by the points corresponding to the values PNLTM-10.

Substituting D back into the EPNL equation gives:

$$EPNL = average\,(PNLT(t(1):t(2))) + 10log(T) - 13 \tag{C.4}$$

When T = 20, 10log(T) is close to 13, and the equation simplifies to the average. T >20 produces a higher EPNL value and T <20 produces a lower EPNL value. This pattern around T = 20 explains why a 10 second long signal produces different results from the calculation as opposed to a longer section of the same event.

Presented below are plots comparing EPNL and SELA values for some of the test signals with the values calculated for the corresponding full flyovers. Numbers on the plot correspond to signal numbers for both tests (see Table 3.2 for flyover details). Signal numbers 1, 2, 12, and 18 (distant amplified signals) are excluded from this analysis due to the lack of a comparable full flyover signal.

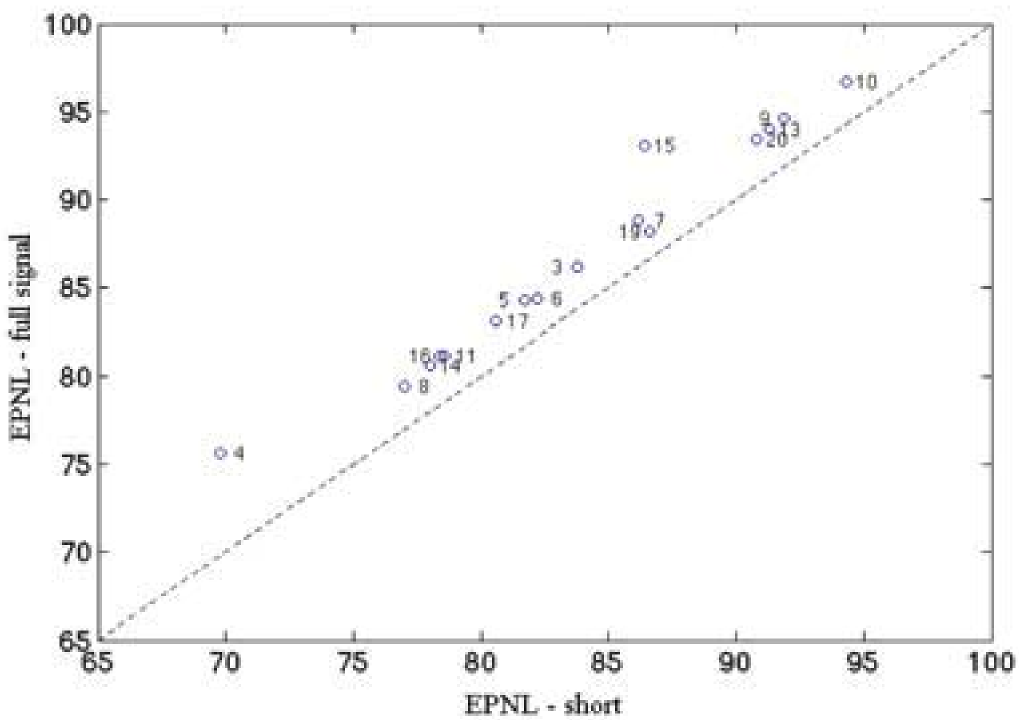

Figure C.1. EPNL values for full flyovers plotted against EPNL values for the 10 seconds duration signals used in the tests (EPNL-short).

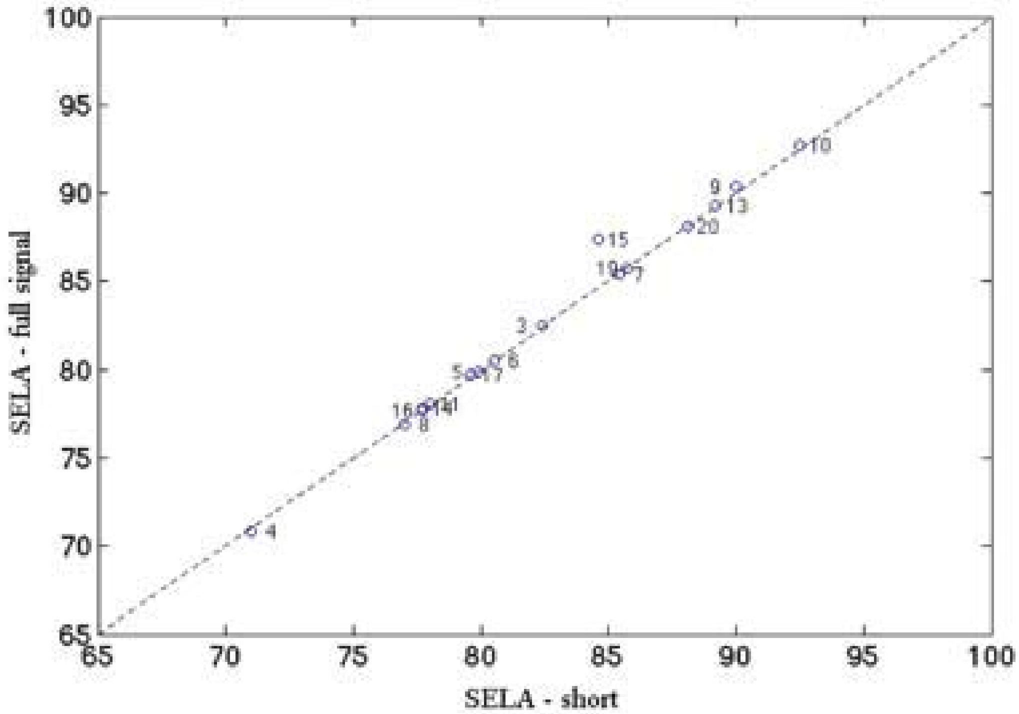

Figure C.2. A-weighted sound exposure (SELA) values for full flyovers plotted against SELA values for the 10 seconds duration signals used in the tests.

For almost all signals there is approximately a 5 dB difference in EPNL values for the entire recording and the corresponding signal used in the test. For the 161102 recording the difference is closer to 10 dB. The PNLT time histories of the full flyover and the test signal are shown in Figure E3. The dashed black line in the figure indicates the 10 dB down mark from the maximum PNLT value (PNLTM). The 10 dB down level is only reached at the ends of the test signal because of the 0.5 second ramps applied at the ends of the signal applied to make the signal go to zero at the ends. The duration that the full flyover is above the 10 dB down point is 29.5 seconds.

The EPNL calculation also excludes any information below the 50 Hz 1/3 octave band, which has a lower limit of 44.7 Hz. Main rotor blade passage frequencies for

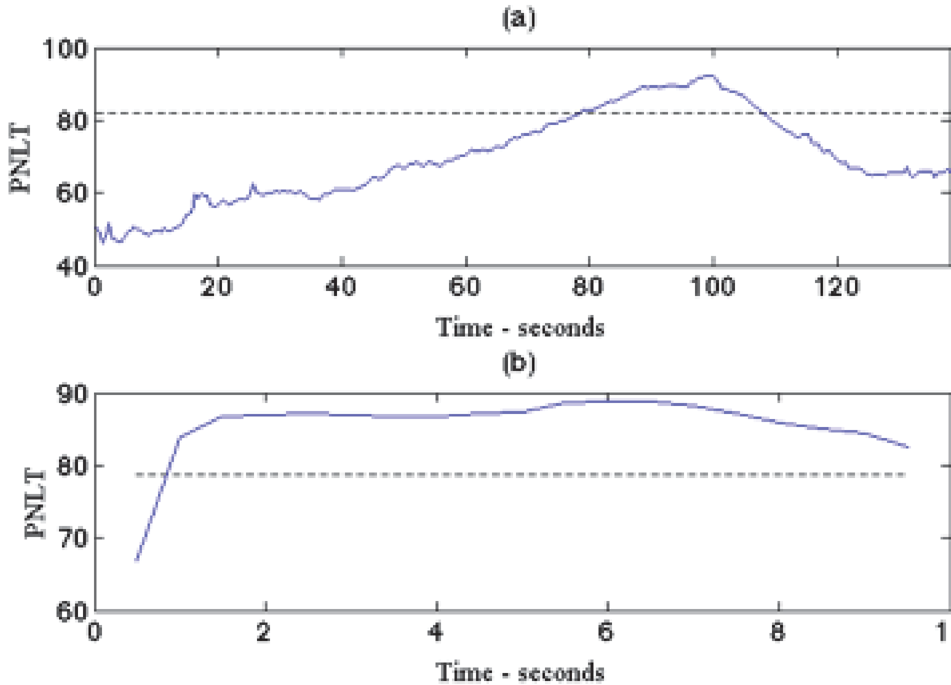

Figure C.3. PNLT time histories for the (a) recording of an XV-15 flyover (test ID 161102), and (b) of the 10 seconds around the peak value that was used as a test stimulus. The test stimulus was attenuated over the first and last half seconds to have zero pressure at the start and end of the signal.

all vehicles used in this test (Bell 206 $\sim$ 13 Hz, BO105 $\sim$ 28 Hz, MD520N $\sim$ 40 Hz, MD902 $\sim$ 39 Hz, Mi-08 $\sim$ 16 Hz, XV-15 $\sim$ 32 Hz) are below that lower limit.

D. LEXICON RESULTS

A lexicon was performed to gather words/phrases used to describe rotorcraft noise. These results are discussed in section 3.1.1. All words collected from the lexicon are shown in eight groups below in Tables D.1 and D.2. These groups are called: Source Descriptions, Impulsiveness, High Pitch Related, Reactions, Low Pitch Related, Moving Source, Roughness, and Other. Other refers to words that did not fit into any of the other 7 categories.

Table D.1. Partial results from the lexicon performed prior to the semantic differential test (1 of 2).

Source Descriptions				
helicopter	motorcycle	jet-like	engines	heavy machinery
lawnmover	planes	missile	propellers	diesels
wind-like	airplanes	bubbles	mechanical	noise on concrete
Impulsive				
choppy	flappy	buffeting	diesels	floppy
fluttery	garbled	impulsive	irregular	oscillatory
sputter	wobbly	shuttering		
High Pitch Related				
whine	high-pitched	squeaking	squealy	chirp
birds chirping	buzzy	piercing	screechy	wailing
whistling				
Reactions				
annoying	disruptive	disturbing	jarring	unsettling
irritating	thundering	mean	ominous	overpowering

Table D.2. Partial results from the lexicon performed prior to the semantic differential test (2 of 2).

Low Pitch Related				
rumbling	drone	foghorn	very low freq	
Moving Source				
flyby	abrupt	flyovers	doppler effect	touching down
pitch-varied				
Roughness				
harsh	rough	grating	raspy	
Other				
hum	chattery	distorted	hollow	jittery
lingering	loose	monotonic	tonal	through a phone
swishing				

E. GROUND EFFECTS SIMULATION

A method was developed to add ground effects to a flyover recording that was taken at the ground level. This was accomplished by adding reflections to simulate the pressure time history at a specified height above the ground. A time-varying filter was employed to add delayed reflections. Time delays were calculated using flight tracking data. These delays were not necessarily aligned with data points, so a Farrow Linear Fractional Delay filter was implemented using MATLAB to determine pressure values between data points. This filter's frequency response begins to decay when more than 0.1 samples from an existing data point, so the pressure time histories were upsampled by a factor of 10 before implementing the filter. Time delays were calculated using Equations E.1 - E.3:

$$d_{dir} = \sqrt{x^2 + y^2 + (z - h)^2} \tag{E.1}$$

$$d_{rfl} = \sqrt{x^2 + y^2 + (z + h)^2} \tag{E.2}$$

where x, y, and z are the distances on the direct path from the noise source to the receiver in the three coordinate directions, and h is the height of the receiver, and

$$t_{delay}(n) = \frac{d_{dir}(n) - d_{rfl}(n)}{c} \tag{E.3}$$

where c is the speed of sound.

The time delays were used to calculate the pressure of the reflections at the corresponding delayed time points which were added on to the original recording. The reflections were added to the original recording using the time-varying filter shown in Equation E.4:

$$y_n = \frac{1}{1+\alpha} x_n + \frac{\alpha}{1+\alpha} x_{n-t_{delay}(n)} \tag{E.4}$$

where y_n is the pressure time history with added reflections, x_n is the recorded pressure time history, and α is the ground reflection coefficient. An example of the implementation of this method is shown below in Figures E.1 (the original recording) and E.2 (ground effects added).

The MATLAB program used to achieve these results is included in section F.2.

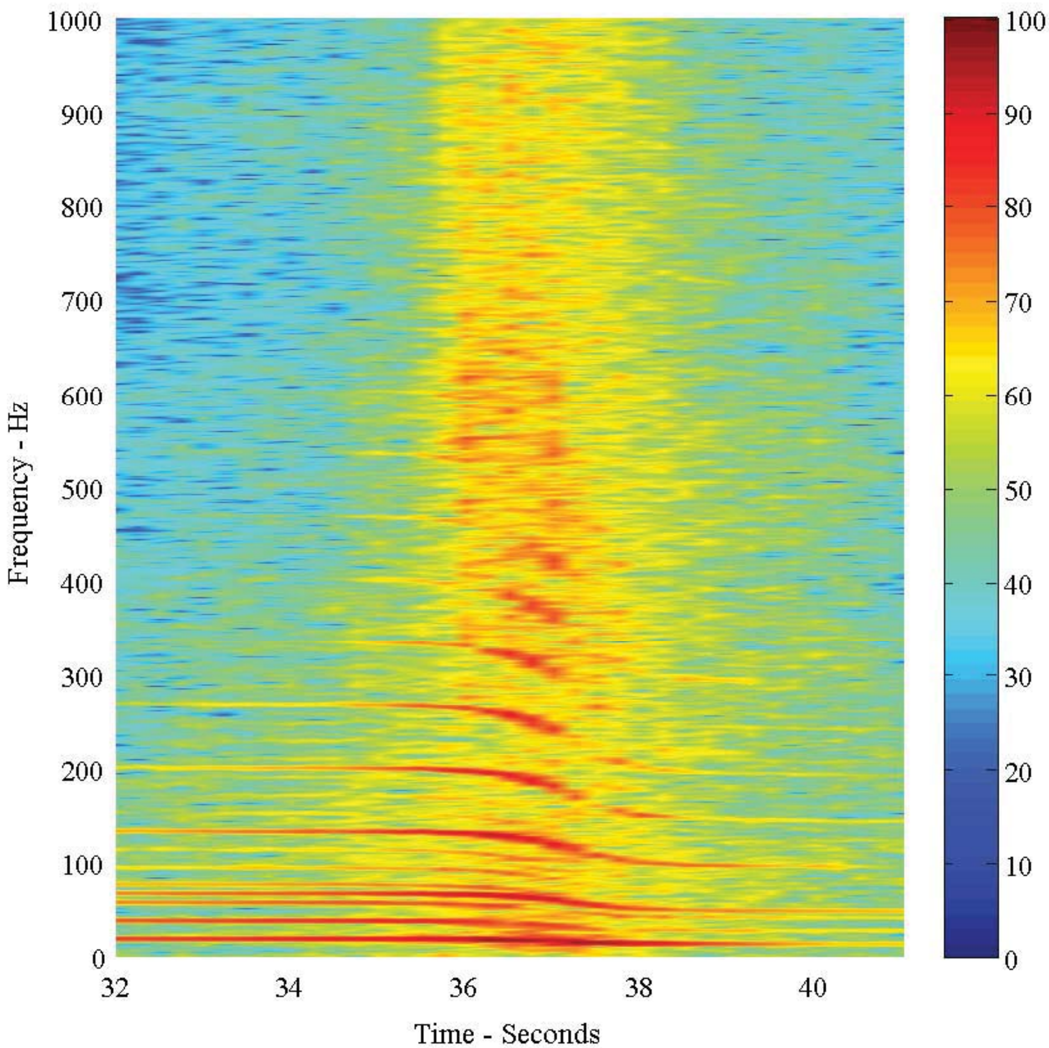

Figure E.1. Spectrogram of a Mil Mi-8M flyover recording.

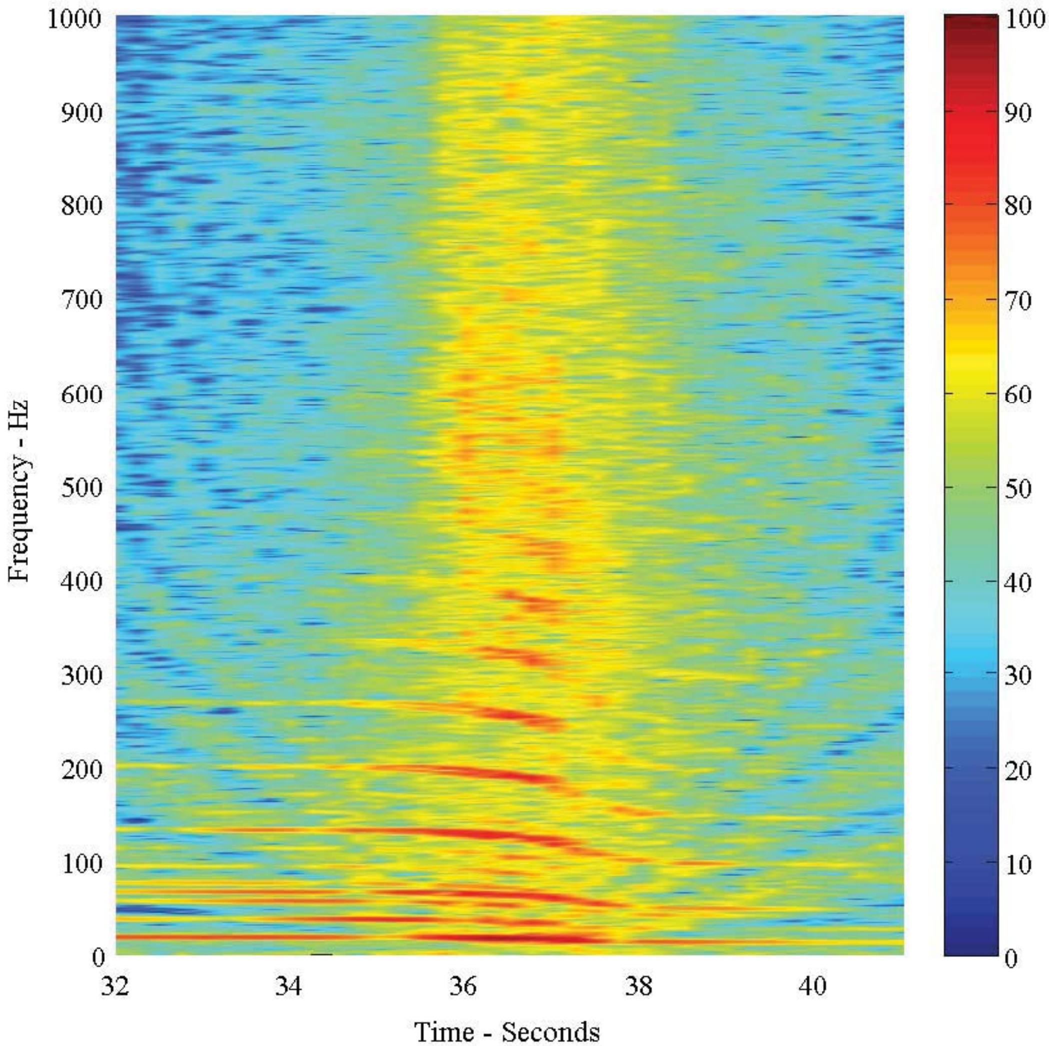

Figure E.2. Spectrogram of a Mil Mi-8M flyover recording with ground effects added.

F. SOFTWARE PROGRAMS

This Appendix contains programs written in MATLAB for characterization and modification of rotorcraft flyovers.

F.1 Characterization and Modification of Tonal Components

Software programs were written based on the methods described in Chapter 2, and are given below. The programs included here are written to: (1) characterize the tonal components of a signal using known frequencies through time and vehicle tracking data (tonefit.m), and (2) regenerate tonal components with varied magnitudes and phase alignment (tonegen.m).

Tonal Component Characterization Program

```
function [coefs_int]=tonefit(time1,dist1,pres1,freqtime,fs,stepsize,np)

% This function takes in time, pressure, and distance vectors
% as well as a frequency through time matrix
% where each column represents a frequency through time

% The function outputs the fitted coefficients (coefs_int), which
% can be used with tonegen.m to regenerate the fitted tones

% Other inputs:
%
% fs - sampling frequency
%
```

```
% stepsize - to be used between calculation points - odd number
%
% np - number of periods of lowest frequency in freqmat to be used in
% fitting window

%% 1 - Integrate frequencies through time to generate phi

nf=size(freqtime,2);

phitime=zeros(size(freqtime));
temp1=zeros(size(freqtime(:,1)));

c=2;

for ii=1:nf
    temp1=2*pi*integ_AM(freqtime(:,ii),fs,time1,c);
    phitime(:,ii)=[ones(1,c-1)*temp1(1).'; temp1(:)];
end
clear temp1

%% 2 - Fit through time

Tp=1/min(min(freqtime));
wl=round(np*Tp*fs)+mod(round(np*Tp*fs),2)+1;
% number of points - should be odd to include center point
wl2=floor(wl/2); % size of half window - to be used in loop
stpt=wl2+1;
etpt=length(pres1)-wl2;
```

```
cc=0;

for jj=stpt:stepsize:etpt
    cc=cc+1;

    philoop=phitime((jj-wl2):(jj+wl2),:);
    distloop=dist1((jj-wl2):(jj+wl2));

    Aloop=freqmat2(philoop,distloop,nf);

    if rank(Aloop)<size(Aloop,2)
        warning off
        rank1(cc)=rank(Aloop);
    end
    ploop=pres1((jj-wl2):(jj+wl2));

    alpha1=0.01;
    coefs(cc,:)=regress(ploop,Aloop,alpha1);

    coefs_time(cc)=time1(jj);
    coefs_r(cc)=dist1(jj);

end

%% 3 - Interpolate coefficients
coefs_int=zeros(length(pres1),nf*2);
for kk=1:nf*2
    coefs_int(stpt:etpt,kk)= ...
    ... interp1(coefs_time,coefs(:,kk),time1(stpt:etpt));
```

```
    coefs_int(1:stpt,kk)=coefs_int(stpt,kk);
    coefs_int(etpt:length(pres1),kk)=coefs_int(etpt,kk);
    end
end
```

Functions called by tonefit.m

```
Integration program

%       x-signal you want to integrate
%       fs-sampling frequency
%       T-time
%       c-point number of signal you want to start the integration,
%           2 < c < length(x)
% Designed with help from Jelena Parapovic

function [xi]=integ_AM(x,fs,T,c);

if length(T)==1
    t=(0:1:fs*T-1)/fs;
else
    t=T;
end
xi=zeros(length(t),1);
delta=1/fs;


for i=c:1:length(t);
    xi(i)=xi(i-1)+.5*(x(i)+x(i-1))*delta;
end
```

```
xi=xi(c:end);
```

Matrix Generation Program

```
function A=freqmat2(phi,dist1,nf)
A=zeros(length(dist1),2*nf);
for ii=1:nf
        A(:,2*ii-1)=(1./dist1(:)).*sin(phi(:,ii));
        A(:,2*ii)=  (1./dist1(:)).*cos(phi(:,ii));
end
```

Tone Regeneration Program

```
function tonestot=tonegen(coefs,phi,dist1,nf,mag,phase1)

% This function takes inputs of:
% coefs - output of tonefit.m
% phi - integrated frequency through time data
% nf - number of frequencies
% mag - magnitude adjustment (multiplication)
% phase1 - phase adjustment:
% 0 = original, 1 = sines, 2 = cosines, 3 = rand

% adjust magnitude
coefs=coefs.*mag;

tonestot=zeros(size(coefs(:,1)));

% start loop
```

```
for nn=1:nf

    coefs_sin_loop=coefs(:,2*nn-1);
    coefs_cos_loop=coefs(:,2*nn);
    coefs_mag_loop=sqrt(coefs_sin_loop.^2+coefs_cos_loop.^2);

    Aloop(:,1)=(1./dist1(:)).*sin(phi(:,nn));
    Aloop(:,2)=(1./dist1(:)).*cos(phi(:,nn));

% adjust phase
    if phase1==1
        % all sines
        coefs_sin_loop=coefs_mag_loop;
        coefs_cos_loop=zeros(size(coefs_mag_loop));
        % set to start at a multiple of 2*pi
        off1=mod(phi(1,nn),2*pi);
        phi(:,nn)=phi(:,nn)-off1;
        Aloop(:,1)=(1./dist1(:)).*sin(phi(:,nn));
        Aloop(:,2)=(1./dist1(:)).*cos(phi(:,nn));
    elseif phase1==2
        % all cosines
        coefs_cos_loop=coefs_mag_loop;
        coefs_sin_loop=zeros(size(coefs_mag_loop));
        % set to start at a multiple of 2*pi
        off1=mod(phi(1,nn),2*pi);
        phi(:,nn)=phi(:,nn)-off1;
        Aloop(:,1)=(1./dist1(:)).*sin(phi(:,nn));
        Aloop(:,2)=(1./dist1(:)).*cos(phi(:,nn));
    elseif phase1==3
```

```
        % random phase
        coefs_sin_loop=coefs_mag_loop;
        coefs_cos_loop=zeros(size(coefs_mag_loop));
        Aloop(:,1)=(1./dist1(:)).*sin(phi(:,nn)+2*pi*rand);
        Aloop(:,2)=(1./dist1(:)).*cos(phi(:,nn));
    end

    for ii=1:length(Aloop(:,1))
        tonestot(ii)=tonestot(ii)+Aloop(ii,1)*coefs_sin_loop(ii);
        tonestot(ii)=tonestot(ii)+Aloop(ii,2)*coefs_cos_loop(ii);
    end
end
```

F.2 Ground Effects Simulation

The following program was written to add simulated ground effects to a groundboard microphone recording of a flyover using tracking data.

Ground Effects Simulation Program

```
% function [pnew]=GroundEffects(p,xs,ys,zs,xr,yr,zr,fs,alphar,J)
% This function takes inputs of:
% p - pressure time history of groundboard recording
% xs - x-direction time history of moving source
% ys - y-direction time history of moving source
% zs - z-direction time history of moving source
% xr - x-direction position of receiver (single value)
% yr - y-direction position of receiver (single value)
% zr - z-direction position of receiver (single value)
% J  - upsampling multiple
```

```
% where x=0, y=0, z=0 refers to microphone location

% fs - sampling frequency - must be same for p and xs, ys, zs

% alphar - reflection coefficient of ground

% ALL DISTANCES MUST BE IN FEET

% set reflection parameters
alphaa=1; % actual

%% Determine time delays between direct and reflected path
% speed of sound
c=1115; % ft/s

% direct path
dist_dir=sqrt((xs-xr).^2+(ys-yr).^2+(zs-zr).^2);
% reflected path
dist_rfl=sqrt((xs-xr).^2+(ys-yr).^2+(zs+zr).^2);
% time delays
t_delays=(dist_rfl-dist_dir)/c;

%% Upsample
newfs=J*fs;

p_zeros=zeros(1,length(p)*J);
t_delays_zeros=zeros(1,length(p)*J);
```

```
p_zeros(1:J:end)=p;
t_delays_zeros(1:J:end)=t_delays;

% design and implement Butterworth low pass filter
% create filter
NA=7;
fc=(0.9*fs)/(newfs);
[B,A]=butter(NA,fc,'low');

% implement filter using filtfilt
p_up=J*filtfilt(B,A,p_zeros);

% design and implement Butterworth lp filter for tracking data
% create filter
NA=15;
fc=0.1;
[B2,A2]=butter(NA,fc,'low');

t_delays_up=J*filtfilt(B2,A2,t_delays_zeros);

% new time vector
t_up=0:1/newfs:(1/newfs)*(length(p)*J-1);

% create point delay vector
ptdelays=t_delays_up*newfs;
wholedelays=floor(ptdelays);
fracdelays=ptdelays-wholedelays;

% create fractional delay filter
```

```
h=dfilt.farrowlinearfd;
h.PersistentMemory=true;

% determine where to start filter
stpt=max(ptdelays)+1;
% round up to multiple of J
stpt=mod(-stpt,J)+stpt;

% preallocate pnew with zeros
pnew=zeros(1,(length(p_up)-stpt)/J);

% find time delayed vector
t_delayed=t_up-t_delays_up;

% start counter
cc=0;

refl=zeros(1,length(pnew));

for ii=(1+stpt):J:length(p_up)
    cc=cc+1;
    h.FracDelay=fracdelays(ii);
    h.States=p_up(ii-wholedelays(ii)-1);
    refl(cc)=filter(h,p_up(ii-wholedelays(ii)));
    t_delayed_loop(cc)=t_up(ii)-t_delays_up(ii);
    pnew(cc)=(1/(1+alphaa))*p_up(ii)+(alphar/(1+alphaa))*refl(cc);
end
```