Spring 4-22-2016

The Effects of Tones in Noise on Human Annoyance and Performance

Joonhee Lee
University of Nebraska - Lincoln, joonhee.lee@huskers.unl.edu

THE EFFECTS OF TONES IN NOISE

ON HUMAN ANNOYANCE AND PERFORMANCE

by

Joonhee Lee

A DISSERTATION

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of Doctor of Philosophy

Major: Architectural Engineering

Under the Supervision of Professor Lily M. Wang

Lincoln, Nebraska

May, 2016

THE EFFECTS OF TONES IN NOISE

ON HUMAN ANNOYANCE AND PERFORMANCE

Joonhee Lee, Ph.D.

University of Nebraska, 2016

Advisor: Lily M. Wang

Building mechanical equipment often generates prominent tones because most systems include rotating parts like fans and pumps. These tonal noises can cause unpleasant user experiences in spaces and, in turn, lead to increased complaints by building occupants. Currently, architectural engineers can apply the noise criteria guidelines in standards or publications to achieve acceptable noise conditions for assorted types of spaces. However, these criteria do not apply well if the noise contains perceptible tones. The annoyance thresholds experienced by the general population with regards to the degree of tones in noise is a significant piece of knowledge that has not been well-established. Thus, this dissertation addresses the relationship between human perception and noises with tones in the built environment.

Four phases of subjective testing were conducted in an indoor acoustic testing chamber at the University of Nebraska to achieve the research objective. The results indicate that even the least prominent tones in noises can significantly decrease the cognitive performance of participants on a mentally demanding task. Factorial repeated-measures analysis of variance of test results have proven that tonality has a

crucial influence on working memory capacity of subjects, whereas loudness levels alone did not. A multidimensional annoyance model, incorporating psycho-acoustical attributes of noise in addition to loudness and tonality, has been proposed as a more accurate annoyance model.

Acknowledgements

I would like to express my deep appreciation and gratitude to my advisor, Dr. Lily Wang, for her continuous guidance and mentorship through my entire doctoral period. Dr. Wang's enthusiasm motivated me in all the time of research and writing of this dissertation. I feel so fortunate to have her as my advisor and mentor.

Besides my advisor, I would also like to thank my dissertation committee. Thank you, Dr. Walt Jesteadt, for valuable advice in the psychoacoustic area and research design. Thank you, Dr. Clarence Waters, for supporting me as an early career researcher. Thank you, Dr. James Bovaird, for teaching me statistical analyses. Thank you, Dr. Erica Ryherd, for reviewing my manuscript and providing insightful comments. I would like to acknowledge Mr. Jerry Lilly from JGL Acoustics for providing audio recordings to support my research.

I greatly appreciate the undergraduate and graduate students at the University of Nebraska who participated in my research, including Zhao Peng, Hyun Hong, Matt Blevins, Jenn Francis, Laura Brill, Madeline Davidson, Kristin Hanna, Brenna Boyd, and Rachel Obenland.

Last but not the least, I would like to thank my family. My parents and my sister for supporting me spiritually throughout writing this dissertation and my life. I cannot express my gratitude enough to my wife, Jahae Ri, for her sacrifice and support while I pursued my degree. I could not achieve my degree without her support.

Table of Contents

List of Figures

List of Tables

1. Chapter One

Introduction

1.1 Background

Building mechanical systems including HVAC (heating, ventilating, and air conditioning) equipment have become more energy-efficient nowadays, but less attention is being paid to the sound quality of the equipment. An assortment of building mechanical equipment generates prominent tones via rotating parts like fans and pumps. The tonal noises can cause unpleasant evaluation of spaces and potentially increased complaints by building occupants. So far, however, there has been limited research on the effects of tones on human annoyance that can be used to set objective guidelines or limits on tones in noise. Current indoor noise evaluation methods such as Noise Criteria and Room Criteria also do not directly account for tonal characteristics of noises (Ryherd and Wang, 2010).

Noise regulations about tonal components in a number of municipalities mostly reduce designated maximum allowable noise levels by 5 dB when a source of sound includes any pure tones (Los Angeles County, 1978; New York, 2006; Seattle, 2007; Minnesota, 2008). Pure tone components are often determined by a one-third octave band measurement according to ISO 1996-2 Annex D5 (ISO, 2007). However, the one-third octave band measurement technique is not always capable of detecting the tonal component, if the tone falls on the edge of two bands, and a 5 dBA penalty value is rather arbitrary because the value is not linked to accurate annoyance perception.

Thus, this dissertation describes subjective investigations on how exposure to tonal noise impacts human annoyance perception and task performance in the built environment, using a larger variety of signals than most previous studies. The dissertation addresses three complementary research objectives: 1) to examine the relationship between associated tonal noise metrics and annoyance perception, 2) to determine upper limits of acceptability for tonality with the goal of developing a dose-response relationship that can be used to set guidelines for tones in noise, and 3) to identify effect of tones on human task performance.

1.2 Dissertation Outline

Chapter 2 provides a comprehensive literature review regarding the noise metrics, noise-induced annoyance, factors impacting annoyance perception, and test methodologies for measuring the annoyance in the previous studies. Chapter 3 explains the test facilities used in the subjective studies and statistical analyses of this dissertation.

Four phases of subjective testing were conducted in an indoor acoustic testing chamber at the University of Nebraska to achieve the research objectives. Chapter 4 presents a subjective study that had participants complete Sudoku puzzles while being exposed to noise stimuli. In this study, relations between noise metrics and annoyance perception are investigated to develop an annoyance prediction model. Chapter 5 describes a similar subjective test with a digit span task. This study expands the number of noise signals and participants to develop a dose-response relationship for determining the upper limit of tonal components in noises. Chapter 6 introduces a subjective test to investigate multidimensional aspects of annoyance perception using actual building mechanical noise signals with tones. Assorted audio recordings from building mechanical equipment are used in contrast to the previous studies which used artificially synthesized noise stimuli. In Chapter 7, the annoyance perception of multi-tone complexes is investigated with a series of paired comparison tasks, and the results are used to improve the accuracy of the proposed model linking tonality metrics and annoyance perception.

A brief summary and conclusions of this dissertation research are presented in Chapter 8. The limitations of the research and future research directions are also discussed.

2. Chapter Two

Literature Review

2.1 Introduction

Audible tones in noises such as those generated from aircraft, wind turbines, and building mechanical systems have been recognized as a serious source of public noise pollution since the 1960s. Therefore, a considerable amount of literature has been published on the relationship between human annoyance and tones in noises. This chapter begins by reviewing noise metrics related to tonality and annoyance perception. It will then go on to summarize the previous studies concerning the definition of noise-induced annoyance, factors that influence noise annoyance, and subjective test methodologies to measure the annoyance perception by noises. While studies related to tonal noises are a priority for review, other studies are also examined if they contribute to this research.

2.2 Noise Parameters

The noise metrics introduced in this chapter fall under three categories. The first category encompasses noise metrics that were developed to quantify tonality perception. The second category deals with widely used noise metrics for loudness perception. The last category includes those metrics that have been proposed for quantifying annoyance by noises. The last type usually combines two or more perceptual attributes of the noise signals, such as loudness and tonality. Abbreviations for each of the noise metrics are introduced in square brackets and will be used throughout the dissertation.

2.2.1 Tonality Metrics

One of the most straightforward methods proposed for calculating tonality of a noise signal involves measuring one-third octave bands of background sound pressure levels, as described in ISO 1996-2 Annex D (ISO, 2007). The presence of prominent tones in background noises is determined by comparing a one-third octave band's SPL to the values in both adjacent bands. The tone decision criteria are: 15 dB level difference for low frequency one-third-octave bands (25 Hz to 125 Hz), 8 dB for middle frequency bands (160 Hz to 400 Hz), and 5 dB for high frequency bands (500Hz to 10,000 Hz). Many municipalities in the United States have adopted this method in their noise regulations and apply a 5 dB penalty if tones are detected when comparing against maximum allowed noise levels (Los Angeles County, 1978; New York, 2006; Seattle, 2007; Minnesota, 2008). However, the one-third octave band method may not detect tonal components, particularly if tones are located at the boundary frequencies of the one-third octave bands because sound energy from tones will be split into two adjacent octave bands. The 5 dB penalty value is also arbitrary since adding 5 dB does not necessarily or accurately reflect how annoyance perception is changed. Several other tonality metrics have been developed to overcome the deficiencies of the one-third octave band method.

2.2.1.1 Tone-to-Noise Ratio and Prominence Ratio

The most widely used metrics for tonality perception are Tone-to-Noise Ratio [ΔL_{tnr}] and Prominence Ratio [ΔL_{pr}], as standardized in ANSI/ASA S12.10/Part 1 Annex D (ANSI/ASA, 2010). These methods can identify tones between 89.1 Hz and 11,220 Hz. Caution is needed when using the methods for tones below or above the frequency range since the methods do not provide the psycho-acoustical evidence for those frequencies. The frequency spectra from Fast Fourier Transform (FFT) analysis without any weighting filters are used to calculate the Tone-to-Noise Ratio and the Prominence Ratio. Caution needs to be taken for the FFT analysis to ensure that the frequency resolution is less than 0.25% of the tone frequency for Tone-to-Noise Ratio and 1% for the Prominence Ratio.

As the name implies, Tone-to-Noise Ratio is the decibel level difference between the tonal noise energy and masking noise energy within the critical bandwidth centered on the tone frequency. The critical band is the frequency bandwidth where broadband noise contributes to the masking of tones near the tone. The formula for the Tone-to-Noise Ratio is:

$$L_{tnr} = 10 log \frac{X_t}{X_n}, \tag{2.1}$$

where X_t is the mean-square sound pressure of the tone and X_n is the mean-square sound pressure of the masking noise, which is the total mean-square sound pressure in the critical band without the tonal part. Tones are regarded as prominent if the Tone-to-Noise Ratio is greater than 8 dB above 1 kHz and the prominence criteria increase at lower frequencies:

$$\begin{aligned} &L_{tnr} \geq 8.0 dB \; for \; f_t \geq 1000 Hz, \\ &L_{tnr} \geq 8.0 + 8.33 \log\left(\frac{1000}{f_t}\right) dB \; for \; f_t < 1000 Hz, \end{aligned} \tag{2.2}$$

where f_t is the tone frequency under investigation.

The Prominence Ratio is the exceedance level of the critical band centered on the tone to the average level of the two adjacent critical bands. The concept of this methodology is similar to the one-third octave band method except that critical bands replace one-third octave bands. The equation for the Prominence Ratio is as follows:

$$\begin{aligned} L_{pr} &= 10log\frac{X_M}{(X_L + X_U)\times 0.5}\ for\ f_t > 171.4Hz, \\ L_{pr} &= 10log\frac{X_M}{\left[X_L\times\left(\frac{100}{\Delta f_L}\right)+X_U\right]\times 0.5}\ for\ f_t \le 171.4Hz, \end{aligned} \tag{2.3}$$

where X_M is the mean-square sound pressure of the middle critical band centered on a tone frequency; X_L,X_U are the mean-square sound pressures of the lower and upper critical bands; Δf_L is the bandwidth of the lower band; and f_t is the tone frequency under investigation. Tones are determined as prominent if the Prominence Ratio is greater than 9 dB for frequencies above 1 kHz, and the criteria increase at lower frequencies:

$$\begin{aligned} L_{pr} &\ge 9.0\,dB\ for\ f_t \ge 1000\,Hz, \\ L_{pr} &\ge 9.0 + 10\log\left(\frac{1000}{f_t}\right) dB\ for\ f_t < 1000\,Hz, \end{aligned} \tag{2.4}$$

where f_t is the tone frequency under investigation.

Tone-to-Noise Ratio and Prominence Ratio analyze tones independently unless multiple tones are sufficiently close. According to the ANSI standard, Tone-to-Noise Ratio may be more appropriate for multiple tones in adjacent critical bands whereas the Prominence Ratio is more accurate for multiple tones within the same critical band.

However, Hellweg et al. (2000, 2001) found through a round robin test that neither metric correlates well with subjective perception when multiple tones or harmonics exist.

2.2.1.2 Tonal Audibility

Tonal Audibility [ΔL_{ta}] is introduced in ISO 1996-2 Annex C (ISO, 2007). The metric is calculated based on the steady-state A-weighted frequency spectrum of a noise recording. In the standard, tones are technically defined as local maxima with a 3 dB bandwidth smaller than 10% of the bandwidth of the critical band.

There are two main differences between Tonal Audibility and the previous two metrics, Tone-to-Noise Ratio and Prominence Ratio. One major difference is that Tonal Audibility includes a frequency correction term in its calculation so that the prominence criteria of tones is constant across frequencies. The other difference is that it uses a linear regression line instead of actual noise components when calculating masking tonal levels within the critical bands. The equation is given by:

$$L_{ta} = L_{pt} - L_{pn} + 2dB + \log\left[1 + \left(\frac{f_c}{502}\right)^{2.5}\right], \qquad (2.5)$$

where L_{pt} is the total sound pressure level of the tones; L_{pn} is the total sound pressure level of the masking noise in the critical band; and f_c is the center frequency of the critical band. Based on the Tonal Audibility calculation, penalty factors between 0 to 6 dB are provided to adjust the overall A-weighted noise levels, rather than setting prominence criteria. It also requires separate analysis for each tone within a multi-tonal noise signal.

2.2.1.3 Aures' Tonality Model

Aures (1985) developed a tonality metric that includes the frequency, bandwidth and levels of all tonal components in a noise signal. This method first calculates weighting functions based on each tonal component's bandwidth, frequency and tonal level by Equation (2.6):

$$\begin{aligned} w_1(\Delta z_i) &= \frac{0.13}{\Delta z + 0.13}, \\ w_2(f_i) &= \left(\frac{1}{\sqrt{1+0.2(f/700+700/f)^2}}\right)^{0.29}, \\ w_3(\Delta L_i) &= \left(1-e^{\frac{-\Delta L}{15}}\right)^{0.29}, \end{aligned} \tag{2.6}$$

where Δz_i is the bandwidth of the tonal component in Bark; f_i is the frequency of the tone in Hz; and ΔL_i is the excess level of the tonal component above the broadband masking noise, as proposed by Terhardt et al. (1982). The Bark unit corresponds to the critical bandwidth of hearing. Then, these weighting functions are combined to derive an overall weighting function w_T for all tonal components by Equation (2.7):

$$w_T = \sqrt{\sum_{i=1}^{n}\left[w_1^{\frac{1}{0.29}}(\Delta z_i)\, w_2^{\frac{1}{0.29}}(\Delta f_i)\, w_3^{\frac{1}{0.29}}(\Delta L_i)\right]^2}. \tag{2.7}$$

Another weighting function w_{Gr} accounts for the overall loudness of tone to noise ratio and is expressed as:

$$w_{Gr} = 1 - \frac{N_{Gr}}{N}, \tag{2.8}$$

where N_{Gr} is the loudness of the broadband noise component, and N is the total loudness of the sound. Finally, Aures' tonality K [Aures] is calculated as:

$$K = c \times w_T^{0.29} \times w_{Gr}^{0.79}, \tag{2.9}$$

where c is a calibration constant to give a 1 kHz pure tone of 60 dB SPL a K value of one.

2.2.1.4 Others

There are a few other tonality metrics that have been developed by researchers but have not been widely adopted in the noise community. Spectral Contrast was developed by Berglund et al. (2002). In this study, similarity and preference ratings of environmental noises were measured. The authors found that the acoustic metric that best correlated with noise preferences was Spectral Contrast, which quantifies tonality of the noise by counting the number of local maxima within Zwicker's specific loudness critical-band spectra. The specific loudness can be calculated from the decibel level for each critical band, which is similar to one-third octave band spectra of A-weighting sound pressure levels. Susini et al. (2004) investigated the sound quality of indoor air-conditioning units and found that one of the dominating perceptual structures was highly correlated with Noise-to-Harmonic Ratio. NHR is the ratio of the broadband noise part and harmonic parts by resynthesizing the noise with digital signal processing techniques.

2.2.2 Loudness Metrics

Widely used loudness metrics are also investigated in this research due to the close relationship between loudness and annoyance perception. A variety of loudness levels have been used to assess noise-induced annoyance depending on the context of the studies. An A-weighted equivalent sound level (L_{Aeq}) is the most common noise metric for environmental noise assessment because it is easy and convenient to measure. Other widely used noise metrics are day-night average sound level (Kryter, 1982, 2007; Miedema & Vos, 1998) for steady community noises, statistical noise levels for time-fluctuating noises (Tang, 1997), loudness levels from the standard ISO532B (ISO, 1975) and ANSI S.3.4-2007 (ANSI, 2007) for more sophisticated loudness perception, and Perceived Noise Level (Kryter, 1960) specifically for aircraft noise nuisance. This dissertation does not present detailed procedures or formula to calculate theses loudness levels.

2.2.3 Combined Metrics

There are a few noise metrics that consider both loudness and tonality to quantify tonal noises. The primary idea of these combined metrics is that they add penalty values, derived from tonality, to loudness levels.

2.2.3.1 Tone-corrected Perceived Noise Level

Kryter (1960) developed a noise metric called Perceived Noise Level [PNL] for jet aircraft noise based on one-third octave band spectra. The metric utilizes equal 'noisiness' contours developed from subjective equal annoyance perception tests. However, Little's study (1961) found a weak relation between PNL and noises with tones. PNL was consequently revised with a tone-correction factor and named Tone-Corrected Perceived Noise Level [PNLT] (Kryter and Pearsons, 1965). The tone correction factor varies from 0 to 6.7 dB according to the frequency of tones and the level differences between one-third octave band values. In the 1970s, PNLT was adopted for use by the United States Federal Aviation Authority (FAA) in their regulations (FAA, 1969).

2.2.3.2 Joint Nordic Method

The Joint Nordic Method [JNM] (Pedersen et al., 2000) is standardized in ISO 1996-2, along with the simplified one-third octave band method. The penalty K values derived from Tonal Audibility are added to A-weighted sound pressure levels. Because adding a 5 dB penalty to the overall sound level is too drastic when using the simplified one-third octave band method, it increases inaccuracy to the perception of annoyance. Thus, this method varies penalty values from 0 dB to 6 dB according to the prominence of tones. Subjective tests with artificial and real recordings from industry and wind turbine noises were conducted for the noise annoyance assessment. The criteria are given by:

$$\begin{aligned} &K = 6\,dB \;\; for\; \Delta L_{ta} > 10\,dB, \\ &K = \Delta L_{ta} - 4\,dB \;\; for\; 4\,dB \le \Delta L_{ta} \le 10\,dB, \\ &K = 0\,dB \;\; for\; \Delta L_{ta} < 4\,dB. \end{aligned} \tag{2.10}$$

For signals with multiple tones, individual Tonal Audibility should be calculated for each, and then the highest value of Tonal Audibility is used to calculate the penalty, K.

2.2.3.3 Sound Quality Indicator

Sound Quality Indicator [SQI] has been recently implemented by the Air-conditioning, Heating and Refrigeration Institute (AHRI) to rate the sound quality of building mechanical product noise (AHRI, 2012). The metric is based on the Perceived

Noise Level procedure and ISO 532B loudness level proposed by Zwicker (ISO, 1975). The calculation begins with one-third octave band data. When any one-third octave band value exceeds the average of the two adjacent bands by more than 1.5 dB, the level of that band is arithmetically adjusted. Then, all one-third octave band sound levels are converted to rating indices according to a conversion table in the standard to calculate SQI. The formula for SQI is expressed by:

$$\begin{aligned} &\mathrm{SQI} = \mathrm{K} + 10 \times \log \sum_{i=100\,Hz}^{10{,}000\,Hz} N_i, \\ &K = 11.83888 - 4.94569\, lnX + 0.614812\left(lnX\right)^2, \\ &X = \frac{\Sigma N_i}{N_m}, \end{aligned} \tag{2.11}$$

where $\sum N_i$ is an arithmetic sum of rating indices for the one-third octave bands from 100 to 10,000 Hz and N_m is the maximum one-third octave band rating index. The metrics mainly aim to compare sound power data of HVAC products, but the usage can be extended to sound pressure data.

2.2.3.4 Psychoacoustic Annoyance

Fastl and Zwicker (2001) have proposed a Psychoacoustic Annoyance [PA] index based on their psychoacoustic studies by combining three different sound attributes: the loudness, the tone color, and the temporal fluctuation. The Psychoacoustic Annoyance is given by:

$$\begin{aligned}
&\mathrm{PA} = N_5\left(1+\sqrt{w_S^2 + w_{FR}^2}\right),\\
&w_s = (S - 1.75) * 0.25\log(N_5 + 10)\ for\ S > 1.75\,acum, \qquad (2.12)\\
&w_{FR} = \frac{2.18}{(N_5)^{0.4}}(0.4 * F + 0.6 * R),
\end{aligned}$$

where N_5 is the 5% percentile loudness in sones; w_S considers the effects of sharpness S in acum; and w_{FR} is the influence of fluctuation strength F in vacil and roughness R in asper. Sharpness is a measure of the high frequency perception of noise signals, roughness is a measure of the rapid amplitude fluctuation from 15 Hz to 300 Hz, and fluctuation strength is a measure of the slow amplitude fluctuation perception up to 30 Hz. The units of sone, acum, vacil and asper are psychological units of assorted acoustic perception metrics proposed by the authors. As expressed in Equation (2.12), the attribute of tonality is not explicitly included in the Psychoacoustic Annoyance.

2.2.4 Summary

This section has provided a description of the noise metrics related to tones in noises. The tonality metrics include Tone-to-Noise Ratio, Prominence Ratio, Tonal Audibility, Aures' Tonality, Spectral Contrast and Noise-to-Harmonic Ratio. Among these metrics the Spectral Contrast and Noise-to-Harmonic Ratio metrics in this section are not utilized in this dissertation mainly because these methods are not verified by other researchers yet except the authors. Among assorted loudness metrics, A-weighted sound level, Perceived Noise Level, and ANSI and ISO Loudness levels are only calculated for noise signals tested in the next subjective studies because of the noise signal characteristics and testing purposes. The combined parameters with loudness levels include Tone-corrected Perceived Noise Level, Joint Nordic Method, Sound Quality Indicator, and

Psychoacoustic Annoyance. The Psychoacoustic Annoyance is also excluded for the signal analysis in next chapters because the parameter does not include the tonality term.

2.3 Noise-induced Annoyance

2.3.1 Definitions

Noise-induced annoyance is a key factor in environmental noise assessment. However, there is some degree of uncertainty around the use of the term ‘annoyance’ by noise researchers. It is mainly because the aims of assorted noise annoyance studies vary according to the background contexts.

According to a definition provided by ISO/TS 15666 (ISO, 2003), noise-induced annoyance is “one person’s individual adverse reaction to noise in various ways including dissatisfaction, bother, annoyance and disturbance”. This standard aims to provide specifications for annoyance questionnaires mainly about community noises. The World Health Organization approaches noise annoyance as an adverse effect on health. In the WHO report, noise annoyance is defined as “the experience of a variety of negative responses, such as anger, disappointment, dissatisfaction, withdrawal, helplessness, depression, anxiety, distraction, agitation or exhaustion” (Kim, 2007). Noise annoyance can subsequently cause psychosocial symptoms such as tiredness, stomach discomfort, and stress. In the study by Guski et al. (1999), noise-induced annoyance refers to a multi-dimensional concept related to behavioral effects such as disturbance and interference and evaluative aspects like nuisance and unpleasantness.

Because the term annoyance embodies broad perceptual concepts, a variety of specific definitions have been suggested by some previous studies. The term ‘unbiased’ annoyance has been proposed to indicate the annoyance perception purely determined by noise characteristics (Guski et al., 1999). Pedersen (2007) divided noise annoyance into three types: global, specific, and potential annoyance. The global annoyance is about holistic noise experiences over time and location without specific incidents and contexts. The specific annoyance is the annoyance response to a specific stimulus in a specified context for specific persons. The potential annoyance is the annoyance response from laboratory or controlled field experiments.

Although differences of opinion continue to exist, there appears to be some agreement that annoyance perception is influenced by noise signal characteristics, the context of measurement, and personal attributes (Pedersen, 2007). In the sections that follow, assorted factors in each of these categories will be discussed.

2.3.2 Factors Influencing Annoyance

2.3.2.1 Noise Signal Characteristics

Noise signal characteristics are those that are physically measured from the noise signals only. There is general agreement in the field of noise annoyance that, among noise signal characteristics, the loudness of a noise signal is most significantly related to annoyance perception. A variety of loudness levels have been used to assess noise-induced annoyance depending on the context of the measurements. Although loudness is clearly the most reliable factor for determining annoyance perception, previous studies have found that loudness metrics alone only predict a small portion of annoyance perception. Brocolini et al. (2012) conclude that at most 30% of the annoyance is accountable by loudness because of other acoustic characteristics and non-acoustic factors.

There is consensus among noise researchers that impulsivity and tonality are two other main noise signal characteristics that must be considered when assessing annoyance (Brambilla and Pedrielli, 1996; Kerry et al., 1998; Sailer and Hassenzahl, 2000; Marquis-Favre et al., 2005; Pedersen, 2007). The impulsiveness of a noise signal is due to single bursts of short duration (Starck et al., 2003). Such impulse noise not only increases annoyance perception significantly but also can cause severe hearing loss. ANSI S12.10 (ANSI/ASA, 2010) specifies a measurement procedure to determine the impulsivity of noises by calculating the difference between time-averaged A-weighted impulse sound pressure level and A-weighted sound pressure level. Tones in noise is another dominant feature that influences annoyance perception as shown in several studies on aircraft, office equipment, HVAC noise, product noise, and wind-turbine noises (Lee et al., 2004; More and Davies, 2010).

One factor of interest related to tonality is the presence of harmonics in noise signals. Although there is no proposed procedure to quantify the annoyance perception by harmonic tones, some previous studies have found that the harmonic components in a noise signal affect overall annoyance perception. Lee et al. (2005) found that harmonic components besides the fundamental frequencies also affect tonality perception. The authors proposed a modification of the frequency weighting function of Aures' model based on their subjective results. Yanagisawa et al.(2011) investigated the emotional sound quality of a vacuum cleaner and argue that consonant tones increase pleasantness perception when compared to the original vacuum cleaner sound without peak tones. In contrast, dissonant peak tones decrease pleasantness of the cleaner's sound quality. Töpken et al. (2010) point out that the ratio of fundamental frequency and harmonics is a crucial factor influencing the pleasantness of the noise perception.

Other possible acoustic factors related to annoyance perception include fluctuation of noises (Fastl and Zwicker, 2001; Dittrich and Oberfeld, 2009), excessive spectral concentrations especially in low frequencies (Persson et al., 1985; Persson and Björkman, 1988), and vibration perception (Schomer, 2005).

2.3.2.2 Non-acoustic Factors

There are many non-acoustic factors that can affect annoyance perception besides noise signal characteristics. The non-acoustic factors may generally be classified into two groups: the context of measurement and personal attributes.

The context of measurement includes all environmental factors that have a potential to affect annoyance reactions. Pedersen (2007) listed the time of day, the location of measurement, and the activity during exposure to the major factors of the context. The author also argues that the subjective responses in a laboratory experiment cannot be identical with the responses in real situations because of their controlled environments. Due to the artificial context of being in the laboratory, the annoyance perception therein should vary from the actual situation. Hünerbein et al. (2010) suggested relative measures for laboratory studies. Kroesen et al. (2013) investigated the effects of survey context when measuring annoyance perception. The authors found that the annoyance rating responses can be affected by the preceding question items. The results showed that when subjects rate the annoyance of an aircraft noise in the context of other noise sources, the average ratings are significantly higher than when it is measured individually or in the context of the normal conversation. The authors explain that this is mainly because the subjects are framed in the particular definition by the context of the preceding questions, and thus, they change their definition of annoyance over time.

The personal attributes category includes noise sensitivity, fear from the noise source, age, and attitude towards the source, among others. Marquis-Faver et al. (2005) summarized non-acoustic factors through reviewing previous annoyance studies. They additionally mention the perception of neighborhood, cultural background, time spent at home, personal daily experience, and gender factors. The authors also argue that fear and noise sensitivity have been shown to have the most significant effects on annoyance perceptions among non-acoustic factors. Fear from the noise source by listeners is rarely experienced with building mechanical noise and, thus, noise sensitivity among the personal attributes is mainly of interest in this research. Noise sensitivity refers to "the internal states of any individual which increase their degree of reactivity to noise in general" (Job, 1999). Several noise sensitivity scales have been developed including those by Weinstein (1980), and Schutte et al. (2007).

2.4 Subjective Testing Methodology

Assorted subjective testing methods have been utilized to measure noise-induced annoyance, some of which have been focused on tones in noise. The methods may be classified into four main sub-groups on the basis of the aims of the studies.

2.4.1 Annoyance Questionnaire Studies

The most widely used method to measure annoyance perception is to use a questionnaire with absolute judgment scales. ISO 15666 (ISO, 2003) specifies two standard questions and scales for annoyance ratings: verbal rating scale and numerical rating scale. The verbal rating scale consists of five choices: Not at all, Slightly, Moderately, Very, and Extremely. In the numerical rating scale, the subject is asked to choose a number between 0 and 10. Even though it is recommended to use multiple items on an annoyance questionnaire to achieve higher reliability (Job et al., 1996), many studies still use a single-item question of annoyance. The responses to scale items are usually analyzed with statistical analyses such as analysis of variance, correlation, and linear regression.

In the 1980s, Hellman (1984, 1985) found that that tonal components in broadband spectra impact ratings of annoyance, loudness and noisiness, and that the number of tones and frequency differences between tones as well as the frequency of the tone itself influence annoyance perception.

Landström et al. (1995) investigated the noise annoyance of signals with different spectral shapes. They found that the relation between individual annoyance ratings and sound levels was weak because of tonal components in the noise. The tonal components raised annoyance ratings equally about 3-6 dB in pressure levels. Miedema and Vos (1998b) also suggested extra correction factors for impulsive or tonal components when predicting total annoyance for transportation and industrial noises.

Researchers have examined the association between annoyance questionnaires and known noise metrics. Hastings et al. (2003) investigated the assorted tonality metrics for predicting tonality and annoyance perception of noises. They proposed modifications in calculating the existing metrics and suggested that the bandwidth and roll-off rate of tones should be included for accurate tonality perception for aircraft noise. Ryherd and Wang (2008, 2010) investigated assorted building mechanical noise samples and showed that current indoor noise criteria were not accurately reflecting subjective annoyance perception because the criteria do not typically account for tonal characteristics in assessment. More and Davies (2010) investigated the relation between tonal aircraft noises and human annoyance. The subjects were asked to rate their annoyance after listening to simulated aircraft noises over headphones. The authors found that the modified Joint Nordic Method rating for tonality and a linear regression model with Zwicker's loudness and Aures' Tonality were the most accurate noise metrics among the utilized parameters.

Trolle et al. (2014) investigated short-term annoyance due to tramway noise through multilevel regression analysis. The authors found that three acoustic measures of A-weighted noise level, the variance of time-varying A-weighted pressure (VAP), and total energy of the tonal components in high frequencies (TETC) were highly correlated with annoyance scale responses. TETC reflects the high frequency piercing character of squeal noise. A multi-level regression model with A-weighted noise level, TETC and noise sensitivity was proposed in this study.

2.4.2 Paired Comparison Methodology

An alternative method for measuring perception is the paired comparison method. This method involves comparing a pair of sound stimuli. Then subjects are asked to judge which in the pair is more annoying or preferred or to adjust one in the pair until it is equally annoying (or preferred) to the other stimulus. Kahn et al. (1996) argued that the paired comparison method is more reliable than the questionnaire scale based methods because of its consistency for both trained and untrained participants.

Laux et al. (1993) investigated the relationship between low frequency modulated noise and annoyance perception using paired comparisons. The subjects were asked to select which noise was more annoying than the other. The responses were transformed into a relative scaled annoyance rating. The authors found that Zwicker's annoyance model was highly correlated with annoyance rating, and the correlation coefficient was increased when a modulating factor was included in the model. Lee et al. (2005) investigated the tonality perceptions of harmonic complex tones in machinery noise using the paired comparison method. The subjects were asked to adjust the tonality of a single tone to equalize the perceived tonality of the complex tones. The complex tone stimuli varied with fundamental frequency, a number of harmonics, signal-to-noise ratio, first harmonic order, and roll-off rate of harmonic tones. Aures' tonality model was used to quantify the tonality of the noises. They found that perception of tonality was a function of the pitch strength of the harmonic components. Also, they indicated that Aures' tonality model overestimated perceived tonality concerning complex tones.

Perceptual weight analysis can be categorized as a paired comparison method, but it has specific methodology and purpose. This method provides the relative weights of each component of perceptual features such as loudness from a trial-by-trial analysis. While the level or magnitude of some components varies randomly, subjects are usually asked to choose the noise stimulus from a pair based on loudness or preference perceptions.

Correlation analysis between variations of each component and responses provides the relative weighting of components. Perceptual weight analysis has often been used to investigate spectral components (Leibold et al., 2007; Jesteadt et al., 2014) or temporal components (Oberfeld et al., 2012) of complex noises contributing to overall loudness.

Perceptual weight analysis has not been widely used in noise annoyance studies. Dittrich and Oberfeld (2009) adopted this method in their investigation on annoyance and loudness perception of temporally varying stimuli. They found that temporal weighting improved the prediction of loudness and annoyance, and that the annoyance responses were significantly different from the loudness responses.

One of the biggest challenges in investigating tonal noise is to include the effects of harmonics on overall tonality and annoyance perception. This methodology is the ideal method to explore this research question.

2.4.3 Dose-Response Model for Annoyance Perception

One of the main aims of environmental noise studies is to propose acceptable noise levels. Dose-response relationships (or noise-exposure models) between noise levels and annoyance have been developed and introduced to suggest maximum allowable noise levels. Generally, the percentage proportion of highly annoyed (%HA) or annoyed (%A) persons is predicted by the model with related noise metrics like the A-weighted sound pressure level. The percentage of the annoyance responses collected from verbal scales is recommended, but many previous studies also used numerical scales on the annoyance surveys. In this case, ratings above 72 out of 100 for the highly annoyed and ratings above 50 out of 100 for the annoyed are commonly-used categorization methods (Pedersen, 2007).

Dose-response relationships have typically been developed with a logistic regression model or a quadratic ordinary least squares regression (Miedema and Vos, 1998a). The logistic regression model is a multiple regression with a categorical outcome variable. The equation is given by:

$$\%\text{ of Highly Annoyed} = \frac{1}{1+e^{-(C_0+C_1X_1\cdots C_nX_n)}} \tag{2.13}$$

where $C_0, C_{1,} \cdots C_n$ are coefficients of the model and $X_{1,} \cdots X_n$ are prediction variables, which are typically noise metrics for noise studies. Maximum-likelihood estimation is used to estimate the coefficients of the logistic regression model (Field, 2013a).

Due to significant differences between noise sources, dose-response models for noise exposure relationship have been developed for a number of specific noise source types: wind turbines (Pedersen et al., 2009; Janssen et al., 2011), aircraft, road traffic, and railway noise (Schultz, 1978; Fidell et al., 1991; Miedema and Vos, 1998b). Even if the models are based on huge data sets from field measurements, there remains some uncertainty and a wide confidence interval in these dose-response relationships (Schomer, 2001, 2005). To date, a dose-response model for tonal building noises has not been developed. This paper uses annoyance ratings and likelihood-to-complain responses as the outcome variable for such a model, against a number of the noise metrics described in the previous section.

2.4.4 Multi-dimensional Scaling Studies

The multidimensional scaling (MDS) analysis technique has been used especially in sound quality research. This method can be utilized to identify how subjects evaluate noise signals with a number of unknown perceptual dimensions (Wickelmaier, 2003). These unknown psychological dimensions form the latent basis for a person to evaluate the sound quality of noises (Woodcock et al., 2014). Subjects are usually asked to judge how similar a pair of sound stimuli are or how preferable one of the pair is over the other. The proximity data from the similarity question or the dominance data from the preference question are organized in matrix form for all pairs of stimuli. Then, the number of dimensions can be determined by measuring the goodness-of-fit of a solution applied to the response matrix. MDS analysis is beneficial for investigating the relation between sound stimuli and unidentified perceptions, but one of the challenges with the MDS technique is interpreting what each dimension is. Usually, additional correlational analyses are required for this work.

The MDS technique has been used in psychoacoustic and noise research areas to investigate the annoyance perception by sound quality of car interior noises (Bisping, 1997; Choe, 2001), HVAC noises (Berglund et al., 2002; Susini et al., 2004), concert hall acoustics (Bradley, 2006) and railway noises (Woodcock et al., 2014). Two studies of MDS related to tonality are highlighted in this section. Berglund et al. (2002) investigated perception of environmental noises including ventilation-like noise spectra with the multidimensional scaling methodology, and concluded that Spectral Contrast, which is related to the tonality, is the best acoustic index for predicting the preference rating of noises. Susini et al. (2004) analyzed indoor air-conditioning unit sound quality by MDS analysis. They found that the sound quality of the air-conditioning units was based on three perceptual dimensions; these were significantly correlated with loudness, Spectral Centroid, and Noise-to-Harmonic Ratio (NHR). The spectral centroid is related to the "brightness" of the sound perception, and can be determined by the distribution of harmonics in the spectrum. The NHR is, as explained in Section 2.2.1, the ratio of broadband noise components to harmonic components, and is related to the tonal strength of the noise signal. Listeners' preferences significantly changed as these parameters varied.

2.5 Relation of Noises with Task Performance

The effects of noise on task performance is a major area of interest within the field of acoustics. A variety of cognitive tasks have been implemented in previous studies when exposing subjects to noise signals: digit span tasks involving memorization of numbers in order (Saeki et al., 2004; Haka et al., 2009; Ebissou et al., 2013), free recall tasks involving memorization of words (Lee and Jeon, 2013), crossword puzzles (Frank et al., 2007), proofreading tasks (Holmberg et al., 1993), concurrent multi-tasks (Bailey and Konstan, 2006), and comprehensive multiple tasks like typing, reasoning and math test (Ryherd and Wang, 2008) or operation span task, dot series task, reading, and proofreading tasks (Haka et al., 2009). Even though there is some evidence for effects of tones on task performance, the generalizability of these studies has been limited, mainly due to the diverse types of noise sources and noise levels.

Moreover, research on the effects of tonal strength on task performance has produced conflicting results, and there is no general agreement to date. Laird (1933) found that complex tones increased the error rates on tasks of laboratory experiments. The author argued that tones above 512 Hz have a greater effect on performance than lower frequency tones. Grjmaldi (1958) also found tendencies of slower response times and increasing error rates in coordinated movement performance for tones in the range of 2400 Hz to 4800 Hz. Ryherd and Wang (2008) investigated the influences of discrete tones in background noise on task performance. Six different background noise conditions with assorted tonal levels were used in the subjective test. Although the results showed a trend for the annoyance and distraction perception ratings to be higher for the prominent tonal noises, there was no significant relationship between task performance and noise conditions. They recommended that a wider range of tonal signals be tested in future research.

2.6 Summary

This chapter has provided a description of the noise metrics related to tones in noises. The definitions and methodologies for testing noise annoyance in previous studies also have been discussed in this chapter. The annoyance response data must be interpreted with caution because there are influential non-acoustic factors such as individual noise sensitivity. A review of previous research on noise-induced annoyance and investigating the effects of noises on task performance is presented. The evidence suggests that tonality in noise is one of the primary factors in annoyance perception, but more research is necessary to determine acceptable levels of tones in noises. The impact on task performance of tonal signals that are commonly found in the built environment is not as clear. The effects of the presence of complex tones or harmonics are also poorly understood to date.

3. Chapter Three

Research Facilities and Statistical Methods

3.1 Introduction

This chapter describes the research methods used for the subjective tests presented in Chapters 4, 5, 6, and 7. Information about the test facility and equipment are presented. The statistical methods used to analyze the subjective test data are also discussed.

3.2 Facilities

All subjective tests were completed in an acoustic testing chamber at the University of Nebraska. Figure 3.1 illustrates a schematic plan of the testing chamber, which has a volume of approximately 27.8 m^3. The chamber is acoustically isolated from a monitor room and nearby spaces. Materials in the room include carpeted floor, gypsum board walls with additional absorptive panels, acoustic bass traps, and acoustical ceiling tiles. The average mid-frequency reverberation time is 0.31 seconds, and the ambient background noise level is 37 dBA when the air-conditioning in the chamber is turned off.

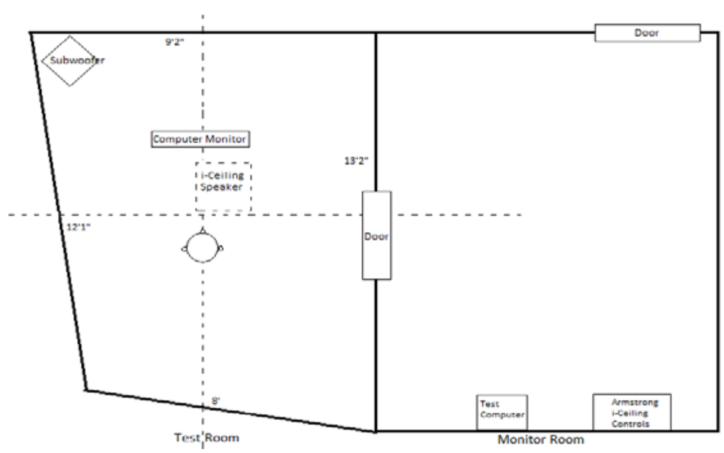


Figure 3.1 Schematic plan of the Acoustic Testing Chamber at the University of Nebraska

Figure 3.2 presents the ambient background noise levels in the chamber across octave bands. The tonal test signals were generated through a ceiling-mounted Armstrong i-ceiling speaker and a sub-woofer in a corner. The i-ceiling speaker appears as other ceiling tiles in the ceiling grid, so that participants cannot visually identify the location of the sound source. Participants sat in the middle of the chamber and were advised not to move their location during the experiment.

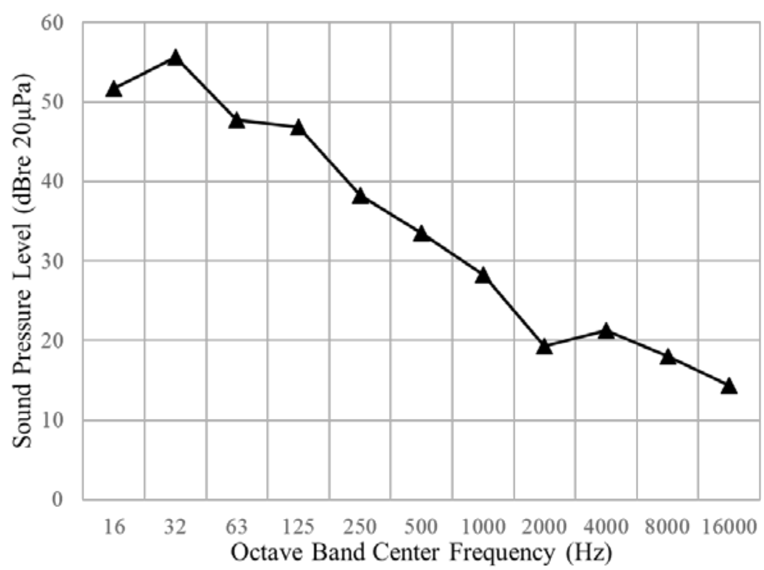


Figure 3.2 Measured octave band spectra for the ambient background noise in the test chamber when air-conditioning is off

3.3 Statistical Analysis

3.3.1 Parametric Test Assumptions

When using statistics to assess a model, there are assumptions for the data that should be met. The assumptions depend on the types of statistics, but most of the following assumptions are essential for the parametric tests like linear regression and analysis of variance (ANOVA) used in this dissertation.

One of the most fundamental assumptions in parametric statistical test is that the data follow normal distribution. Caution is needed in assessing test statistics when data violate the normality assumption. The Shapiro-Wilks test and the modified Kolmogorov-Smirnov test are commonly used to test for normality in addition to graphical investigation of data distribution.

Homoscedasticity (or homogeneity of variance) means that dependent variables should have equal levels of variance across the range of the predictor variable. The Levene test is the most common one for testing this assumption.

Linearity refers to the outcome variable being linearly related to prediction variables (Field, 2013b). If this linearity assumption is not met, all parametric test statistics are useless because the model should be analyzed with nonlinear models.

Independence means that the errors of individual observations should be not related to each other. If this assumption is not met, multilevel statistical modeling should be used instead. The multilevel models use clustered structures to take account of correlations between individual data.

3.3.2 Correlation

The most widely used method to investigate the relation between variables is a statistical correlation. Correlation is a standardized way to measure covariance between two variables. Covariance is an indicator of when deviation of one variable is associated with another variable positively or negatively. Covariance can be measured by the following equation:

$$\mathrm{cov}_{xy} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{(N-1)} \tag{3.1}$$

where $\overline{x}, \overline{y}$ are the means of the variables; x_i, y_i is the individual sample value; and N is the number of observations. The limitation of covariance is that it depends on the scale of the measurement and is not standardized (Field, 2013b). To overcome this limitation, Pearson product-moment correlation coefficient r is used to compare the coefficients between variables.

The correlation coefficient can be calculated by the following equation:

$$r = \frac{\mathrm{cov}_{xy}}{s_x s_y} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{(N-1)s_x s_y} \tag{3.2}$$

where s_x, s_y are the standard deviations of the variables. By standardizing the covariance, the Pearson's correlation coefficient ranges from -1 to +1. A coefficient value of +1 indicates a perfect positive relation, while a coefficient value of -1 indicates a perfect negative relation. A value of zero means there is no linear relationship between two variables. The squared value of the correlation coefficient value, R^2, is widely used as a quantifier of correlation along with the coefficient itself. R^2 is called the coefficient of determination, which is a measure of the amount of variances shared between two variables. For example, a R^2 value of .42 means that 42% of variability in one variable is shared by the other variable.

When the samples do not meet the assumptions discussed in the previous section, Spearman's correlation coefficient, r_s (Spearman's rho), can be used. Spearman's coefficient is a nonparametric statistic based correlation based on ranked data (Field, 2013b). The method basically converts the scale variables to ranked data first and then applies Pearson's correlation equation.

The significance of a correlation is generally evaluated by using t-statistics. The t-statistics analysis tests the hypothesis that the correlation coefficient is significantly different from zero.

3.3.3 Analysis of Variance

Analysis of Variance (ANOVA) is a statistical method to compare the means of variables at two or more different experimental conditions. ANOVA is widely adopted to compare the effect of treatments in an experimental study by calculating the ratio of systematic variances and unsystematic variances. In an experimental study, the means of the treatment groups are usually compared to the mean of the control group to reveal the effect of treatment. ANOVA shows if each of the group means is significantly better than the overall mean across the groups in predicting an outcome variable.

In an independent design, different participants are recruited for different experimental conditions; a repeated measures (within-subjects) design recruits the same participants across different experimental conditions. When both scenarios are used together in an experiment, it is called a mixed design.

When there are several independent variables to investigate, it is called factorial design. In many cases, the number of independent variables are specified by referring to the analyses as being two-way or three-way ANOVA. In this dissertation, repeated-measure factorial ANOVAs are used.

The F-ratio is used for testing the null hypothesis that the group means are all the same in ANOVA, meaning there is no effect of treatment across groups. The F-ratio is a ratio of the variation accounted for by the model and the variation not explained by the model (residual). As shown in Equation (3.3), the F-ratio is calculated by dividing the mean squares of the model by the mean squares of the residual. The calculated F-ratio can be compared against the value one can obtain in the F-distribution with the null hypothesis that the group means are all equal (Field, 2013b).

$$F = \frac{\text{mean squares of model}\,(\mathrm{MS_m})}{\text{mean squares of residual}\,(\mathrm{MS_r})},$$
$$MS_m = \frac{SS_m}{df_m} = \frac{\sum_{k=1}^{k} n_k (\bar{x}_k - \bar{x}_{grand})^2}{df_m}, \tag{3.3}$$
$$MS_r = \frac{SS_r}{df_r} = \frac{\sum_{k=1}^{k} s_k^2 (n_k - 1)}{df_r}$$

where:

SS_m = Model sum of squares, SS_r = Residual sum of squares

df_m = Model degree of freedom, df_r = Residual degree of freedom

$\bar{x}_k$ = Mean of each group k, $\bar{x}_{grand}$ = grand mean

n_k = Number of groups, s_k^2 = variance of each group k.

While the F-test identifies if there are differences between group means, it does not provide any information about which group is affected. There are two additional analyses in ANOVA to evaluate the difference between specific groups: planned contrasts and post hoc tests. Planned contrasts are used when there is specific hypothesis to test and post hoc tests can be used when there is no specific hypothesis.

When factorial repeated-measure ANOVA are used to evaluate experimental data, one additional assumption is required for the data: the assumption of sphericity. This assumption requires that variances of the differences between treatment groups should be approximately equal. The assumption can be tested by Mauchly's test. If Mauchly's test statistic is significant indicating that the assumption of sphericity is violated, the F-test should be analyzed with more restrictive ways such as Greenhouse-Geisser or Huynh-Feldt estimate (Field, 2013b).

3.3.4 Linear Regression

Linear regression is a way of predicting an outcome variable (or dependent variable) with predictor variables (or independent variable) by fitting a straight line between them. When using only one predictor variable, it is referred to as simple regression, and when there are more than one predictor variables, it is called multiple regression.

The multiple regression model can be expressed as:

$$Y = (b_0 + b_1 X_1 + b_2 X_2 + \cdots + b_n X_n) + \varepsilon \tag{3.4}$$

where:

Y= outcome variable

b_n = coefficient of the nth predictor variable, b_0 = intercept

X_n = nth predictor variable

ε = residual.

The ordinary least square method is used to determine coefficients of the regression model by minimizing the residual term in the model. To assess the goodness of fit of the regression model, R^2 is used. R^2 represents the amount of variance in the outcome variable explained by the model (Field, 2013b). The statistical significance of R^2 can be assessed by the F-ratio, as with ANOVA. When assessing statistical significance of individual predictors, a t-statistic test is used with the null hypothesis that the coefficient value of the predictor is zero. The t-test gives insight as to whether or not the model with the individual predictor performs better without the predictor.

Because the coefficient values of the regression model depend on the measurement units of each predictor variable, the values are not comparable to each other for their effect on the outcome variable. To compare the performance of predictor variables, standardized beta (β) should be used. The beta values quantify the number of standard deviation changes caused by changing the predictor variable by one standard deviation.

3.3.5 Logistic Regression

The logistic regression model is one of the regression models for a categorical outcome variable. If the outcome variable contains only two cases like ‘Yes’ or ‘No’, a binary logistic regression model can be used. If the outcome variable contains more than two cases, a multinomial logistic regression model should be used. For one of the noise studies in this dissertation, 'annoyed' or 'not annoyed' is used so that a binary logistic model is adequate for analysis.

One of the main aims of environmental noise studies is to propose acceptable noise levels based on human responses. Dose-response relationships (or noise-exposure model) between noise levels and annoyance have been introduced in the noise community to suggest maximum allowable noise levels. Dose-response relationships have typically been developed with a logistic regression model or a quadratic ordinary least squares regression model (Miedema and Vos, 1998).

The binary logistic regression equation is expressed by:

$$P(Y) = \frac{1}{1+e^{-(C_0+C_1*X_1+L+C_n*X_n)}} \tag{3.5}$$

where P(Y) is the possibility of outcome Y occurring; C_0, C_1, ⋯, C_n are coefficients of the model; and X_1, ⋯, X_n are prediction variables, which are typically noise levels for noise studies. Figure 3.3 illustrates a typical form of the binary logistic regression model.

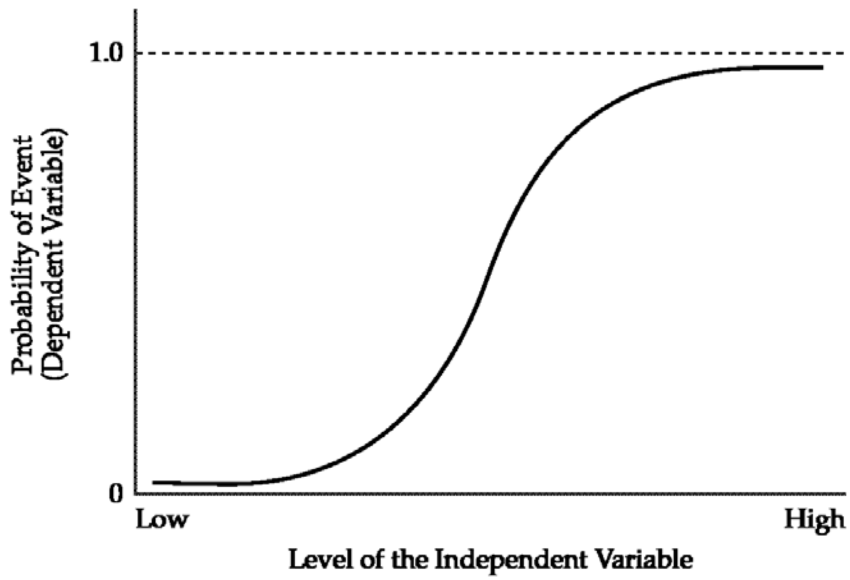


Figure 3.3 Graphical form of the logistic regression model (figure from Hair et al. (2009))

For the logistic regression model, maximum-likelihood estimation is used to estimate the coefficients of the model (Field, 2013). By using an iterative fitting model, maximum-likelihood estimation determines the closest coefficient value for the observed data. The goodness-of-fit of the developed model can be assessed by using R^2, similar to the R^2 of the linear regression. It is a measure of how well the prediction model fits the response data, derived from the chi-square ($\chi 2$) and deviance (-2LL) values. The deviance is calculated from the 'deviance' of the expected probability from the observed values. The deviance can be calculated as follows:

$$-2LL = -2 * \sum_{i=1}^{N} \left[Y_i \ln(P(Y_i)) + (1 - Y_i) \ln(1 - P(Y_i)) \right] \qquad (3.6)$$

in which Y_i, $P(Y_i)$ indicate an individual observed outcome and its expected probability. The deviance represents how much information is not explained by the model. The chi-square can be derived from the deviance as follows:

$$\begin{aligned}\chi^2 &= (-2LL(null)-(-2LL(model)) \\ &= 2LL(model)-2LL(null)\end{aligned} \tag{3.7}$$

in which -2LL(null) and -2LL(model) mean the deviance values when no predictor is included and when predictors are included, respectively. The chi-square value indicates how much the model prediction is improved against the model with no predictor. Several ways have been proposed to calculate R^2 for logistic regression, derived from the chi-square and the deviance, and in this dissertation, Homser & Lemeshow's (2004) , Cox & Snell's (1989), and Nagelkerke's R^2 (1991) are calculated (Field, 2013b).

The odd ratio is the exponential of each coefficient in the logistic regression model. The ratio indicates how the 'odds' of the outcome occurring will change when a unit of predictor changes. If the ratio is greater than one, there would be positive relation between the predictor and the odds of the outcome. For example, if the coefficient value is .7, then the odds ratio of that predictor variable is 2, which is $e^{\wedge.7}$. An odds ratio of 2 indicates that the possibility that the outcome will occur increases by two when the predictor variable is increased by one unit.

3.4 Summary

In this chapter, the acoustic facilities used for subjective tests are introduced. Also, statistical methods applied in this dissertation to analyze the test responses are described. Table 3.1 summarizes the four subjective studies to be presented in the dissertation.

Table 3.1 Brief summary of four subjective studies

Chapter No.	Chapter 4	Chapter 5	Chapter 6	Chapter 7
Purpose of Study	Investigate relation between noise metrics and annoyance	Develop dose-response relationship	Explore multi-dimensional aspects of annoyance	Improve annoyance prediction of multi-tone complexes
Methodology	Sudoku puzzle	Digit span	Multidimensional Scaling	Perceptual weighting
Participants No.	10	20	20	10
Signals	20 artificially generated	40 artificially generated	18 actual building noises	25 artificially generated
Statistics	Correlation, multivariate regression, repeated measure ANOVA			

4. Chapter Four

Relations between Noise Metrics and Human Annoyance and Performance

4.1 Introduction

In this chapter, the relationships between current noise metrics, human annoyance perception, and task performance under tonal noise conditions have been examined through subjective testing. In this study, subjects were asked to complete Sudoku puzzle while exposed to artificially synthesized noise with a tone. The subjects filled out a subjective rating questionnaire on the noise they had just experienced. The results have been used to identify significant noise metrics and develop a multivariate regression model to predict annoyance perception.

4.2 Methods

4.2.1 Noise Stimuli and Equipment

A total of twenty-two noise signals were generated for use in this study by the program Test Tone Generator from EsserAudio (Esser, 2014). Two levels of broadband noise without any tonal components were used: either 40 dBA or 55 dBA overall, following a -5 dB/octave Room Criteria (RC) contour (Blazier, 1981).

A tone at one of two frequencies (125 Hz or 500 Hz) and at one of five prominence levels was added separately to the broadband noise signals, to create the other twenty noise signals. The five tone levels were selected to range from below to above the prominence thresholds listed in ANSI S12.10 (ANSI/ASA, 2010): PR=18 dB for 125 Hz and PR=12 dB for 500 Hz. Table 4.1 presents the Prominence Ratio values for each test signal.

Table 4.1 Prominence Ratios for the tones in the noise stimuli used in the subjective testing as listed by tonal frequency, background noise level, and tone level.

Frequency (Hz)	BNL (dBA)	Prominence Ratio (dB)				
		Tone Level 1	Tone Level 2	Tone Level 3	Tone Level 4	Tone Level 5
125	40	15	18	21	24	27
	55	13	15	18	21	24
500	40	9	12	15	18	21
	55	6	9	12	15	18

Figure 4.1 illustrates the one-third octave band spectra of the test signals. All tonal signals were measured using a B&K 4189-A microphone through the B&K PULSE system at the listener's ear position in the testing chamber, and averaged over a minute for calculation of noise metrics. The related metrics, introduced in Chapter 2, were calculated using Matlab (MathWorks, 2013).

4.2.2 Test Participants

Ten participants, four females and six males, were recruited from the University of Nebraska at Omaha community and paid to complete this study, ranging in age from 25 to 43 years old. All participants completed an orientation session including a hearing screening test before participation and demonstrated normal hearing with thresholds below 25 dB hearing level (HL) from 125 Hz to 8 kHz. The noise sensitivity of each participant was also measured by a reduced version (13 items only) of the Noise-Sensitivity-Questionnaire (NoiSeQ) by Schutte et al. (2007) during the orientation session.

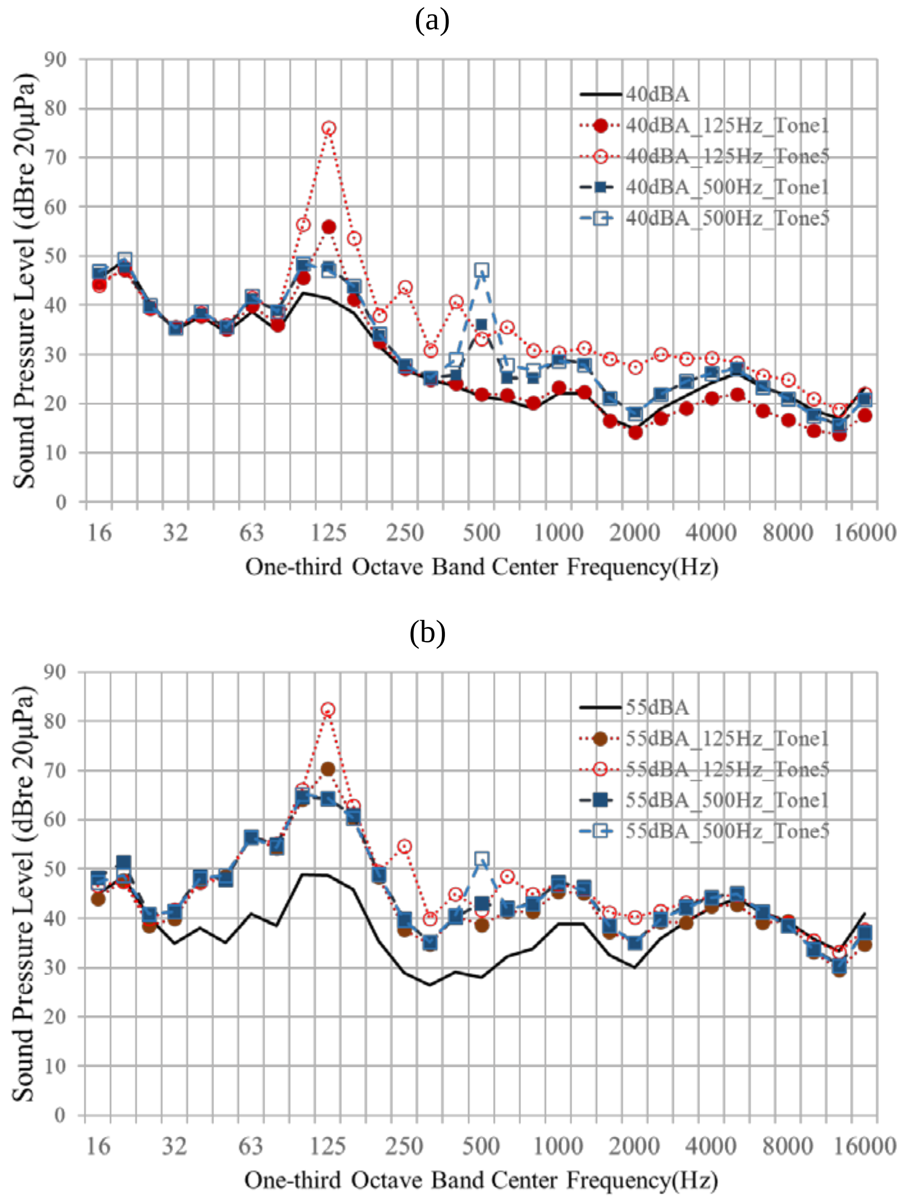


Figure 4.1 Measured one-third octave band spectra for a few of the test noise signals: (a) broadband 40 dBA signals, and (b) broadband 55 dBA signals. Tones were either at 125Hz or 500Hz; for clarity, only the lowest and highest tonal strengths are presented.

4.2.3 Subjective Testing Procedure

The main test consisted of two parts: a direct assessment with task (part A) and a magnitude adjustment test (part B). The results of part B have been presented in another master student's thesis (Francis, 2014) and hence are not included in this dissertation. In part A, participants were asked to complete as many Sudoku number puzzles as possible for ten minutes while exposed to noise signals, some with assorted tonal components. The Sudoku puzzle is a logic puzzle where one completes a 9 by 9 grid with numbers so that each column, each row, and each of the nine 3 by 3 sub-grids contains all digits from 1 to 9.

All participants practiced solving Sudoku puzzles during the orientation session before participating in the main test, and the difficulty of all Sudoku puzzles in the main test was held constant. After spending ten minutes solving the Sudoku puzzles, the subjects answered five questions on a subjective questionnaire about the noise they had just heard. The questionnaire was a modified version of the NASA task load index (Hart, 2006). The original NASA task load index is divided into six subscales: mental demand, physical demand, temporal demand, performance, effort, and frustration. In this study, the questions on physical demand, temporal demand, and frustration were not included; instead questions were added on rating loudness and annoyance incurred by noise as shown in Table 4.2. Participants were asked to respond to each question based on a 21-point scale on a paper form.

Table 4.2 Items from the subjective questionnaire, as modified from the NASA task load index.

Description	Questions
Mental Demand	1. How mentally demanding was the task?
Overall Performance	2. How successful were you in accomplishing what you were asked to do?
Effort	3. How hard did you have to work to accomplish your level of performance?
Loudness	4. How loud was the noise?
Annoyance	5. How annoying was the noise?

Part A consisted of ten 30-minute sessions that were completed by each subject individually on different days. Within each 30-minute session, subjects were exposed to three noise signals (each for ten minutes) and thus completed three sequences of Sudoku puzzles (different puzzles each time) followed by the questionnaire. To minimize the influence of back-to-back comparisons of tonal noise conditions, a neutral background noise condition without any tonal components was used as the second signal within each 30-minute test session. Within a single 30-minute test session, the background noise level of the signals remained at a constant level. The presentation order of the background noise levels and tonal test signals was carefully balanced across subjects using a Latin square design.

Two task performance measures were gathered from (1) counting the number of Sudoku puzzles a subject fully completed within a ten-minute trial and (2) quantifying the accuracy of the puzzle answers in terms of number of correct numbers placed across all completed puzzles.

4.3 Results and Discussions

4.3.1 Relating Subjective Responses and Task Performance

Correlation analysis was conducted on the participants' subjective responses to the modified NASA task load index questionnaires and the task performance outcomes related to the Sudoku puzzles (Table 4.3). Two subjects' responses were excluded from all analyses because they rated responses randomly regardless of sound characteristics.

Table 4.3 Spearman's correlation analysis of the subjective responses and Sudoku puzzle task performance. TLX-avg is the average value of the responses to all five questions on the modified task load index questionnaire. '# of completed' refers to the number of completed puzzles for each trial and 'accuracy' indicates accuracy rates of participants' puzzle answers.

	Mental Demand	Performance	Effort	Loudness	Annoyance	TLX-avg	# of completed	Accuracy
Mental Demand	-							
Performance	.260	-						
Effort	.610**	.496*	-					
Loudness	.501*	.105	.230	-				
Annoyance	.528*	.162	.398	.948**	-			
TLX-avg	.631**	.374	.601**	.880**	.956**	-		
# of puzzles completed	-.317	-.438	-.394	.074	-.020	-.171	-	
Accuracy	-.105	-.483*	-.071	-.289	-.252	-.330	.080	-

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

A "TLX-avg" score was calculated as the averaged value of all five items from the modified survey to represent an overall rating of subjective task load perception induced by noise exposure. Since the task difficulty was held constant with equivalently difficult Sudoku puzzles throughout the experiment, the variations in subjective ratings observed within subjects can be considered as the result of varied background noise conditions. Job et al. (2001) have recommended against using a single question item about annoyance because of its reduced validity; consequently, the composite modified Noise TLX rating is proposed as an alternative in this laboratory study. With a Cronbach's alpha coefficient for reliability of .82, and a test-retest correlation of Noise TLX measure for stability of .77, the "TLX-avg" questionnaire was found to be internally consistent and stable over time and thus suitable for the purpose of this test.

Spearman's correlation (ρ) was utilized because not all of the variables met the assumption of having normal distribution and the sample size was small. As Table 4.3 indicates, most of the subjective responses were significantly correlated to each other. Specifically of interest, the mental demand showed strong correlations with perceptions of loudness and annoyance of the noise, and as expected, loudness and annoyance ratings were significantly correlated with each other (ρ=.948). The task performance results of '# of completed' (the number of completed puzzles for each trial) and 'accuracy' (accuracy rates of participants' puzzle answers) showed non-significant correlation with subjective responses except subjective performance responses (ρ=-.483 with the accuracy).

4.3.2 Relating Noise Attributes to Annoyance Perception

To understand how the physical aspects of the noise signals (background noise level, tone frequency, and tonal strength) related to annoyance perception, a three-way repeated measure ANOVA (analysis of variance) was conducted. Mauchly's test indicated that the assumption of sphericity had been met for the main effects of tonal strength and its interactions with frequency and background noise level. The analysis indicates a significant main effect of background noise level [$F(1,7)=82.606$, $p<.001$], tone frequency [$F(1,7)=20.006$, $p=.003$], and tonal strength [$F(4,28)=4.758$, $p=.005$] on annoyance perception. The main analysis shows that the 55 dBA based tonal signals were significantly more annoying than 40 dBA based tonal signals, and that the 125 Hz tonal signals were significantly more annoying than 500 Hz tonal signals. Contrast comparisons reveal that the 4th highest in prominence tonal signals, [$F(1,7)=10.420$, $p=.014$] and 5th highest in prominence tonal signals [$F(1,7)=12.069$, $p=.010$] were perceived as more annoying than the least (1st) prominent tonal signals. Figure 4.2 illustrates the mean annoyance ratings against background noise level, tonal frequency and tone strengths. Summarizing these results, the overall background noise level does impact annoyance, with louder levels leading to greater annoyance. The lower frequency tone generated greater annoyance ratings, but one should note that the prominence levels of the 125 Hz tone versus those of the 500 Hz tone used in the study were (or were not) the same even though the relative differences from the threshold of tones presented in ISO 1996-2 are the same. The data on tonal strength shows that higher tone levels increase annoyance.

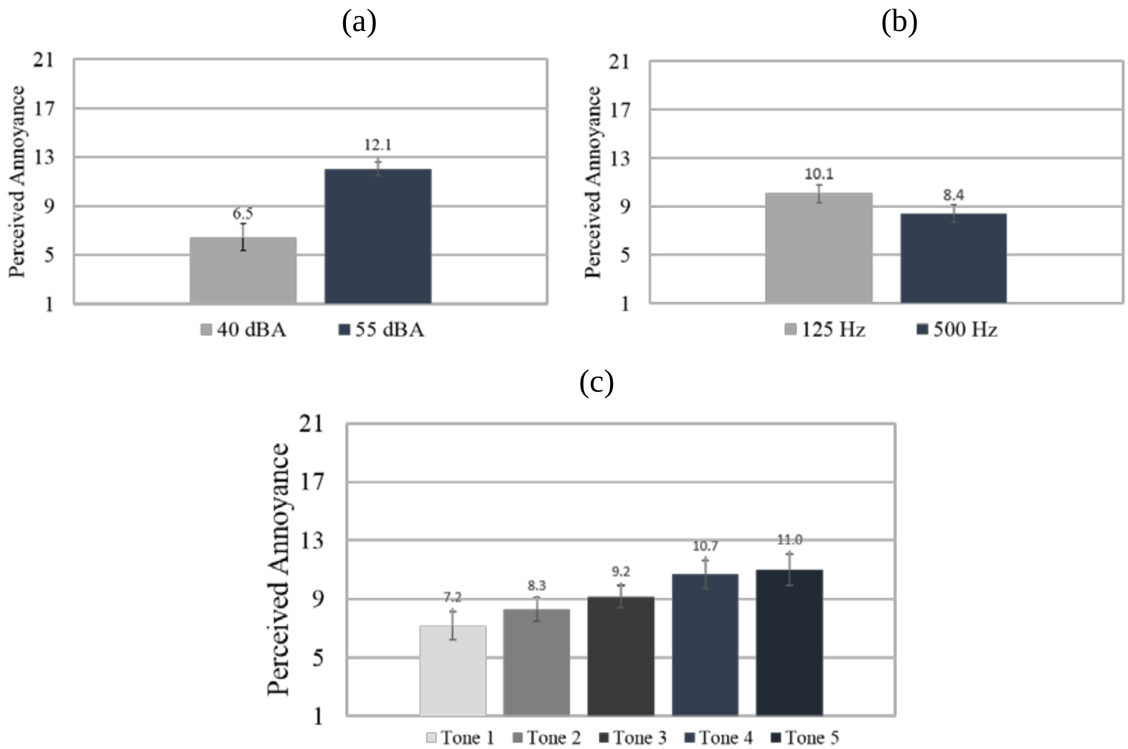


Figure 4.2 Mean annoyance perception ratings plotted against (a) background noise level, (b) tonal frequency, and (c) strength of the tones, where Tone 1 indicates the least prominent tone and Tone 5 indicates the most prominent tone. Error bars indicate standard error.

4.3.3 Correlation and Regression Analysis with Noise Metrics

Spearman’s nonparametric correlation coefficients were calculated between all noise metrics and average participants’ perception ratings. The results have been analyzed in three scenarios: first with all twenty signals included, then with the average ratings for ten signals grouped separately by background noise level (40 dBA or 55 dBA). Table 4.4 presents correlation coefficients between all noise metrics with subjective perceptions.

For the group of all signals, the noise metric that demonstrates highest correlation coefficients with the perceived loudness, annoyance, and TLX-avg ratings is ANSI

Loudness Level. Other loudness metrics were also significantly correlated to the perception ratings, but the tonality metrics such as Prominence Ratio, Tone-to-Noise Ratio, Tonality Audibility and Aures's Tonality did not statistically correlate or had lower coefficients than loudness metrics. This confirms that loudness is the most dominant factor in determining subjective perceptions of noise.

When the signals are grouped by background noise levels, though, tonality metrics did show higher correlations with subjective ratings than loudness metrics. The coefficient values for the assorted tonality metrics are all very similar with no particular metric clearly performing better than others. The results suggest that when background noise level is controlled or comparable, tonality becomes a more influencing factor on annoyance evaluation. Figure 4.3 illustrates average ratings and standard deviation of the annoyance ratings across eight participants for each of the twenty noise stimuli (a) with the ANSI Loudness Levels across the entire group and (b) with Tonal Audibility ratings separated by two background noise levels of 40 dBA and 55 dBA.

For all cases, combined metrics such as the Joint Nordic Method, Tone-corrected Perceived Noise Level and Sound Quality Indicator did not show remarkably better performance than loudness metrics, even though they were significantly related with annoyance ratings. The results suggest that imposing penalty values to loudness levels based on tonal strength may not be the most effective way to quantify overall subjective annoyance of tonal noise. Instead, tonality and loudness of building mechanical noises should be considered as separate metrics.

Table 4.4 Spearman's correlation analysis of noise metrics against subjective responses and Sudoku puzzle task performance. The results are analyzed first with all signals included, and then in two groups separated by background noise level (40 dBA or 55 dBA). Bolded values indicate metric with highest statistically significant correlation values.

All signals(40dBA & 55dBA BNL)			
	Loudness	Annoyance	TLX-avg
PR	.150	.186	.147
TNR	-.123	-.081	-.095
ΔLta	.006	.056	.019
Aures	.297	.359	.314
dB	.805**	.824**	.772**
dBA	.866**	.887**	.842**
ANSI Loudness	**.946****	**.950****	**.926****
ISO Loudness	.938**	.952**	.925**
PNL	.892**	.920**	.886**
PNLT	.869**	.877**	.826**
JNM	.840**	.869**	.818**
SQI	.904**	.899**	.856**
40dBA BNL only			
PR	.794**	.867**	.782**
TNR	.794**	.867**	.782**
ΔL_{ta}	**.778****	**.888****	**.815****
Aures	.673*	.709*	.697*
dB	.806**	.939**	.855**
dBA	.794**	.927**	.830**
ANSI Loudness	.685*	.745*	.697*
ISO Loudness	.685*	.745*	.697*
PNL	.685*	.842**	.867**
PNLT	.794**	.830**	.758*
JNM	.794**	.927**	.830**
SQI	.806**	.806**	.709*
55dBA BNL only			
PR	.799**	.867**	.758*
TNR	.709*	.845**	.845**
ΔL_{ta}	**.787****	**.891****	**.818****
Aures	.781**	.903**	.782**
dB	.715*	.756*	.530
dBA	.707*	.770**	.564
ANSI Loudness	.878**	.855**	.709*
ISO Loudness	.817**	.867**	.697*
PNL	.720*	.806**	.539
PNLT	.744*	.782**	.527
JNM	.707*	.770**	.564
SQI	.689*	.663*	.444

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).

(a)

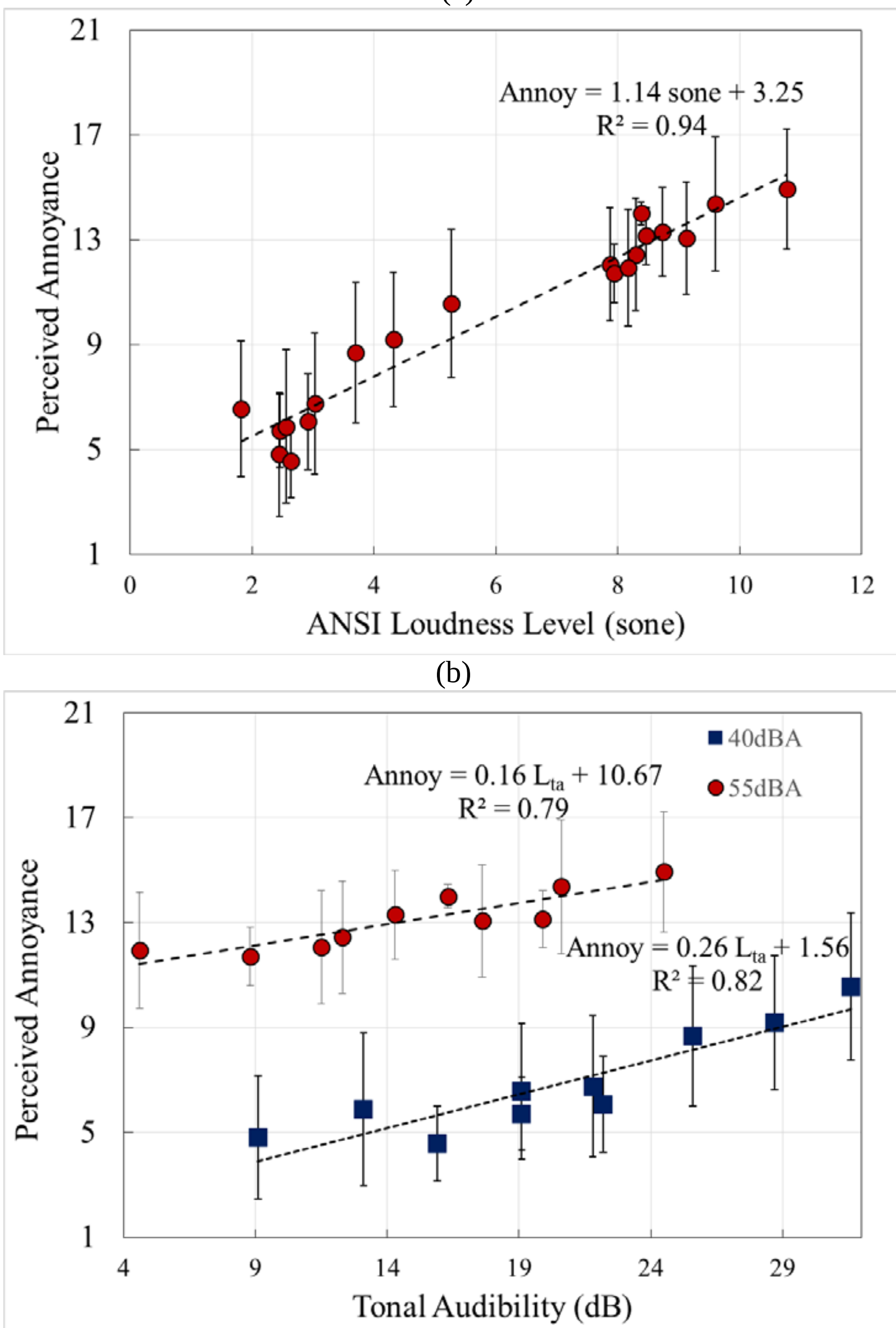


Figure 4.3 Averages (mark) and standard deviations (error bar) of the annoyance ratings across 8 participants for each noise stimulus plotted against (a) ANSI Loudness Level for all stimuli and (b) Tonal Audibility for 40 dBA and 55 dBA BNL separately. Dashed lines indicate regression lines of annoyance rating prediction with regard to each metric.

Based on the results in Table 4.4, ANSI Loudness Level and Tonal Audibility were selected to be used as predictors for a linear multiple regression model for annoyance perception, because these two metrics resulted in the strongest correlation with annoyance perception among other noise metrics. Equation (4.1) presents the multivariate regression model with ANSI Loudness Level and Tonal Audibility.

$$Annoyance = 1.806 + 1.164 * \left[ANSI\ Loudness\left(sone \right) \right] + .072 * \left[Tonal\ Audibility\left(dB \right) \right] \qquad (4.1)$$

Table 4.5 also presents standard error of coefficients, standardized coefficients and statistical significance when ANSI Loudness Level was only used (in step 1) and when Tonal Audibility was also included (in step 2), in addition to the coefficient values for each predictor. Standardized beta values indicate the number of standard deviations that the outcome annoyance will change as a result of one standard deviation change in the predictor. The R^2 value for this model is .943, which is a measure of goodness-of-fit of linear regression, indicating that 94.3% of the annoyance rating variance can be explained by the ANSI Loudness model only; the regression line is plotted in Figure 4.3(a). When including Tonal Audibility as a second predictor, the R^2 value increased to .962. Even though this increase is small, the multivariate regression model does significantly predict more variation in annoyance perception when including Tonal Audibility as a second predictor; for step 2, the ANSI Loudness Level [$t(17)=20.796$, $p < .001$] and Tonal Audibility [$t(17)=2.943$, $p=.009$] are both significant predictors of annoyance perception.

Table 4.5 Linear regression model of predictors for annoyance perception, with 95% bias corrected and accelerated confidence intervals reported in parentheses. Confidence intervals and standard errors are based on 1000 bootstrap samples. Standardized β values indicate the number of standard deviations that the outcome annoyance will change as a result of one standard deviation change in the predictor.

	b	SE B	β	p
	Step1			
Constant	3.254 (2.305, 4.310)	.512		p=.001
ANSI Loudness(Phon)	1.137 (1.004, 1.263)	.066	.971	p=.001
	Step 2			
Constant	1.806 (.498, 3.187)	.683		p=.020
ANSI Loudness(Phon)	1.164 (1.043, .1.308)	.069	.994	p=.001
Tonal Audibility(dB)	.072 (.027, .1111)	.021	.141	p=.004

Note. $R^2 = .40$ *for Step 1;* $\Delta R2 = .02$ *for Step 2 (ps = .011).*

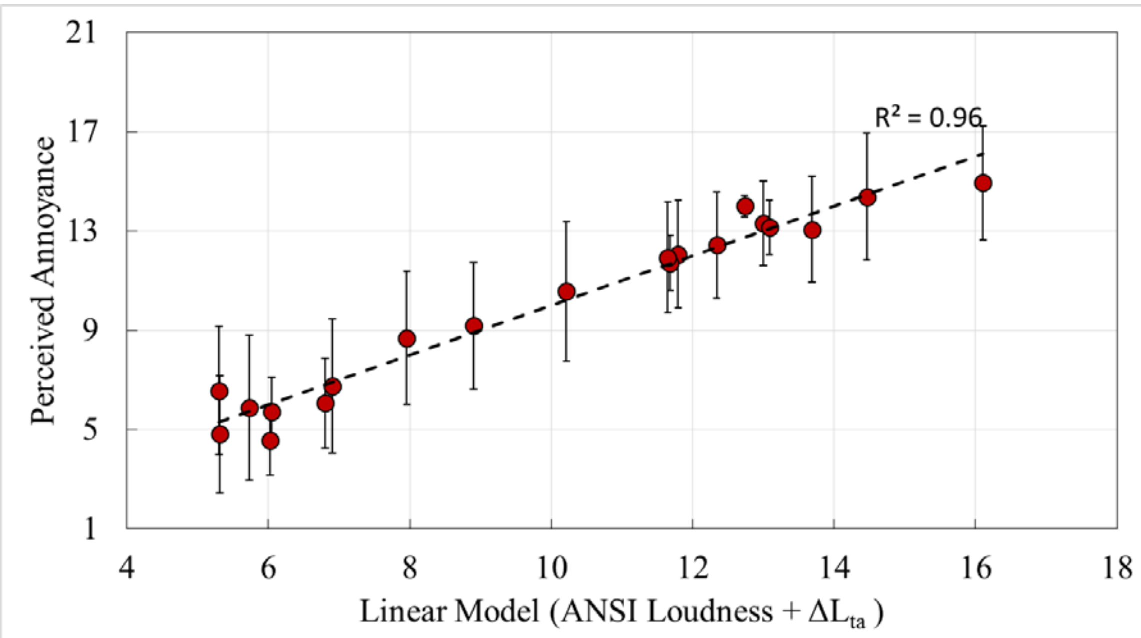


Figure 4.4 Averages (mark) and standard deviations (error bar) of the annoyance ratings across 8 participants for each noise stimulus plotted against the proposed linear regression model of annoyance perception from Equation 4.1 (dashed) based on ANSI Loudness level and Tonal Audibility (R^2=.96).

Figure 4.4 illustrates the regression line with the calculated linear model based on Equation 4.1.

4.4 Summary

The purpose of this study was to investigate how noise signals with varying degrees of single prominent tones affect subjective annoyance perception and task performance and to develop a prediction model of annoyance using current noise metrics. Subjects completed Sudoku puzzles and a questionnaire modified from the NASA task load index to quantify the overall work load caused by building mechanical noise in this study. The validity of the modified questionnaire is high based on its reliability coefficient and test-rest coefficient, and the average response from the questionnaire is found to correlate significantly with perceived annoyance and loudness of the background noise signals. A factorial repeated measure ANOVA reveals that participants feel more annoyed with increasing background noise level, lower tone frequency and stronger prominence of the tone strength. Correlation analysis with noise metrics and subjective perception ratings suggest that ANSI Loudness Level among the tested loudness metrics correlates most strongly with annoyance perception, while assorted tonality metrics showed relatively weaker but still statistically significant correlations with annoyance. A statistically significant multivariate regression model with ANSI Loudness Level and Tonal Audibility has been developed, which demonstrates a R^2 value of .962.

The results in this study, however, do not provide any criteria about a threshold of acceptability for tonality. The next subjective testing is designed with the findings and outcomes from this study with an increased number of signals to investigate the dose response relationship.

5. Chapter Five

Dose-response Relationship between Tonality Perception and Noise-induced Complaint by Tones for Building Mechanical System

5.1 Introduction

While the study presented in Chapter 4 showed that annoyance can be impacted by background noise level, tonal frequency, and tonal strength, it used a limited number of signals and test subjects. The research described in this chapter utilizes more signals and more listeners with a goal of developing a dose-response relationship between tonality perception and noise-induced annoyance by using a logistic regression model. Such a dose-response model can then be used to determine an upper limit of acceptability of tonality according to the corresponding background noise level. The results can help to develop a noise guideline including tone criteria for buildings.

5.2 Methods

5.2.1 Participants

Twenty listeners (9 females, 11 males) were paid to participate in the subjective test. The participants were recruited by using fliers distributed on the University of Nebraska at Omaha campus. The average age of all participants was 24.9 years with a standard deviation of 4.9. Most participants were university students or staff members. All listeners participated in an orientation session including a hearing sensitivity test to confirm that they had hearing thresholds below 25 dB HL from 125 Hz to 8 kHz for both ears.

Four subjects' responses were excluded for all analysis in the results sections since they submitted the same minimum rating across all signals; these four are shown as subjects 17 through 20 in Figure 5.3.

5.2.2 Noise Stimuli and Equipment

The stimuli were 40 broadband noise signals with tonal components and no obviously time-fluctuating components. Two broadband noise spectra were used, matching the room criteria RC-30 and RC-38 neutral contours. Neutral spectra were selected to eliminate perceptual impacts caused by spectral elements other than by the tones. Recordings of the noise signals were averaged over a minute at the listener's ear position (3'4" to 3'7") in the test chamber using a B&K 4189-A microphone through the B&K PULSE system. The overall sound pressure levels of the two broadband signals at the listener position were 57 dB (re 20 μPa) and 63 dB (re 20 μPa) respectively. Five levels of tones at one of four specific tonal frequencies (125, 250, 500, 1000 Hz) were added separately to the broadband background noises. The broadband signals and tones were generated using the program Test Tone Generator (Esser Audio) and digitally synthesized using the program Audacity 2.1.1. The tonal levels ranged from barely observable to prominent for each tonal frequency, with Tonal Audibility values ranging from 5 dB to 19 dB. Table 5.1 summarizes the tonality values of the stimuli. The overall sound pressure level of the 40 tonal signals ranged from 57.3 dB (re 20 μPa) to 70.7 dB (re 20 μPa).

Table 5.1 Tonality of noise stimuli used in the subjective test. The same level of tones are added to both RC-30 & RC-38 neutral spectra broadband noises.

Frequency (Hz)		125Hz	250Hz	500Hz	1kHz
Tonal Audibility (dB)	Tone level 1	5.4	5.7	5.2	5.1
	Tone level 2	7.2	7.7	7.5	7.8
	Tone level 3	9.4	9.7	9.9	10.1
	Tone level 4	13.2	12.7	12.1	13.5
	Tone level 5	19.4	19.5	19.0	19.2

5.2.3 Subjective Testing Procedure

The subjective test consisted of one orientation session and six main testing sessions. After the hearing threshold screening test, participants were informed about how annoyance is defined in this study and the purpose of the study. The participants also familiarized themselves with the main task by practicing for 10 minutes at the end of the orientation session.

The participants were next asked to attend four 30-minute sessions, each of which included ten test trials. For each trial, participants were asked to perform a digit span task in which they memorized a series of numbers in the reverse order of presentation while exposed to assorted tonal signals. Trials using only RC-30 neutral background noise were inserted between trials with tonal noise conditions to eliminate back-to-back comparisons of tonal noise conditions. The order of tonal noise signals was randomized by Latin-square design for all participants. Figure 5.1 illustrates the procedure of subjective testing in (a) a session and (b) a trial.

(a)

	One session (~30 min)					
Background noise	Tone 1	RC-30(N)	Tone 2	RC-30(N)	Tone 3	…
Task	Q	Q	Q	Q	Q	

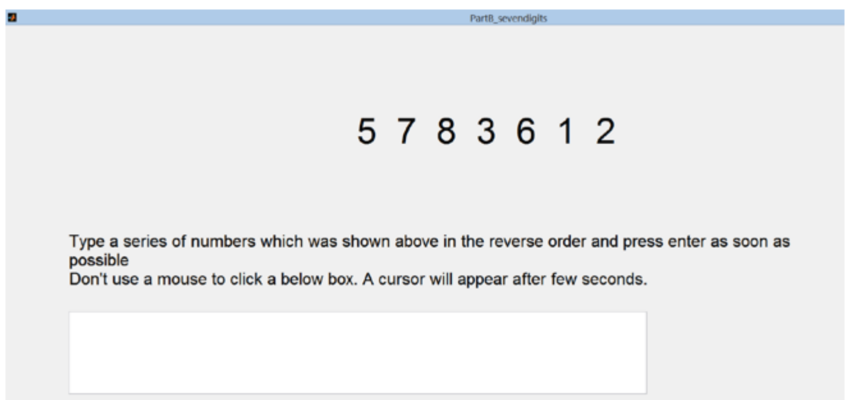
(b)

	One series of digit span tasks (~2 min)										
Background noise											
Digit span	4	4	5	5	6	6	7	7	8	8	Q

Figure 5.1 Description of subjective testing (a) in one session (b) in one trial.

The digit span task is a measure of short-term working memory commonly used in psychology experiments (Mølhave et al., 1986; Jahanshahi et al., 2008). The length of each digit span task increased from 4 digits up to 8 digits over a duration of approximately 2 minutes as illustrated in Figure 5.1 (b), while being exposed to a noise signal. There were two attempts at each digit span. The digits are displayed for 3 seconds and disappeared before answering. The digits were displayed and disappeared at once rather than one-by-one. When the digits disappeared, participants were asked to type the same digits in the reverse order with the given keypad. For example, when 42863 were the displayed digits, participants should type 36824. Conventionally the digit span task is completed when subjects fail to answer two consecutive questions correctly, but in this study, the maximum lengths was manually set to eight digits regardless of participants' answers to fix the duration time under a noise stimulus. A custom-written graphical user interface in Matlab controlled the presentation of all of the trials and noise signals; the program also measured accuracy of answers and reaction time of responses. After each trial, the participants were asked to fill out a subjective questionnaire with two items, indicating how annoyed they were by the noise, and whether or not they would complain about the noise. The annoyance question was answered on an 11-point continuous scale, and the complaint question was dichotomous choice. Figure 5.2 illustrates the Matlab graphic user interface of the subjective testing for the digit span task and subjective questionnaire.

(a)



(b)

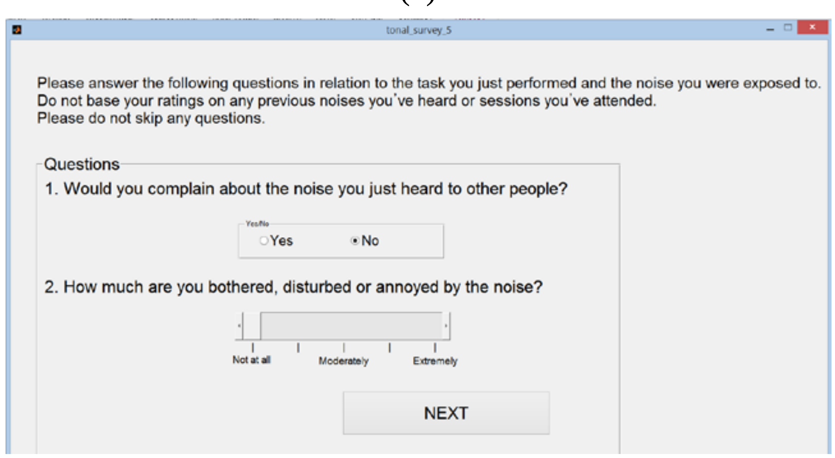


Figure 5.2 Test program interface implemented by MATLAB Graphic User Interface

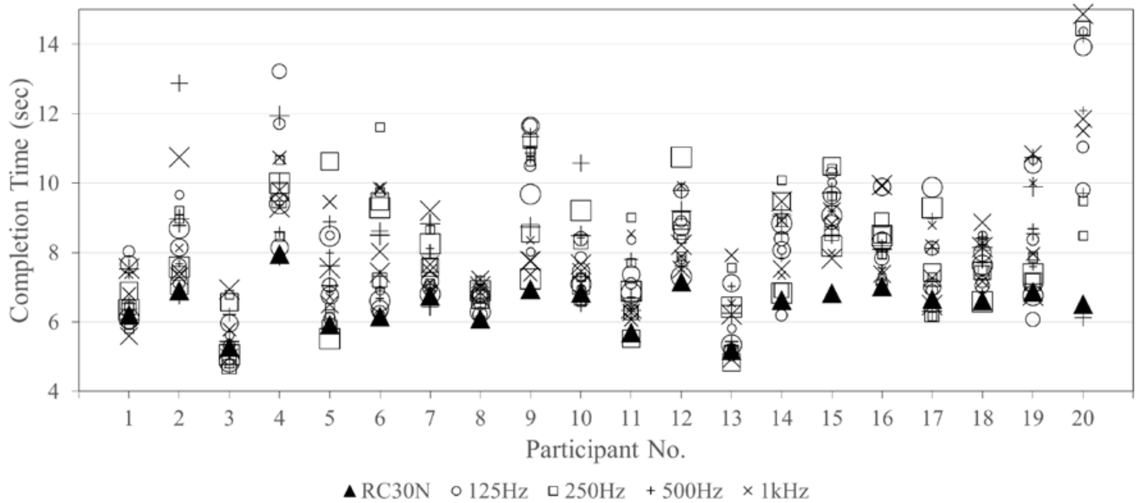
5.3 Results

5.3.1 Task Performance

To investigate the effects of tonal noise signals on task performance, the two outcome variables of (1) maximum number of correct digits provided for a single digit span test and (2) the time it took for the participant to complete a single digit span test were statistically analyzed. The first outcome variable related to identifying correctly the digit span sequence did not show any statistically significant differences between noise stimuli at all. Repeated-measure analysis of variance (ANOVA) , though, showed that all completion times under tonal noise conditions were significantly longer than completion times under broadband noise conditions with RC-30N only ($F(40,600)=2.78$, $p<.001$, $\eta_p^2=0.16$). The completion time with the RC-30N noise condition without any tonal component was measured in normal trials like other tonal noise conditions. The in-between trials of RC-30N noise condition as described in Figure 5.1 were not used to calculate the completion time due to influences from prior task conditions.

Figure 5.3 illustrates each of the individual's completion times for each tonal noise condition over the RC-30 and RC-38 background levels.

(a)



(b)

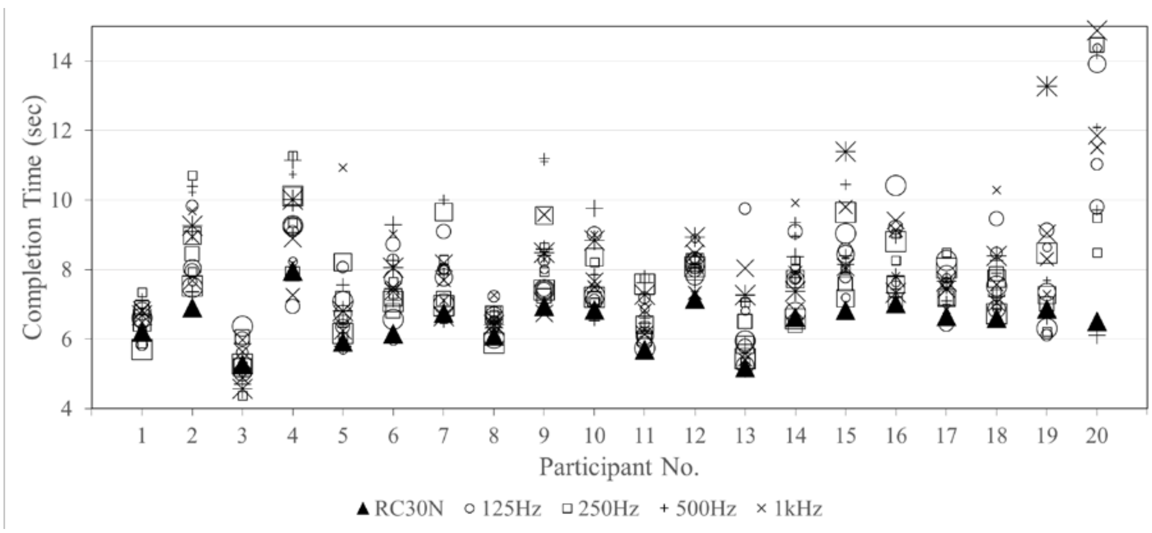


Figure 5.3 Measured completion times for the digit span task under assorted tonal noise conditions (a) above the RC-30 background noise and (b) the RC-38 background noise across participants. The size of each marker corresponds to the tone level of each frequency, with larger markers indicating higher tone levels.

A three-way repeated-measures ANOVA was conducted to investigate the relationships between background noise level, tone frequency, and tone level on completion time. Figure 5.4 compares the completion times between the two different background noise levels, four different tone frequencies, and two tone levels. Only two tone levels (the least and highest levels) are shown to illustrate the maximum difference. Statistical analyses indicate that the effects of background noise level and tone frequency on completion time are not statistically significant even though a trend of longer completion times with higher frequency tones is observed. The only significant factor was tone level ($F(4,60)=2.95$, $p=.027$, $\eta_p^2=0.16$), with higher tone strength resulting in longer completion times. These results indicate that the perception of tonality by participants can affect performance on a digit span task in terms of time taken, but not accuracy.

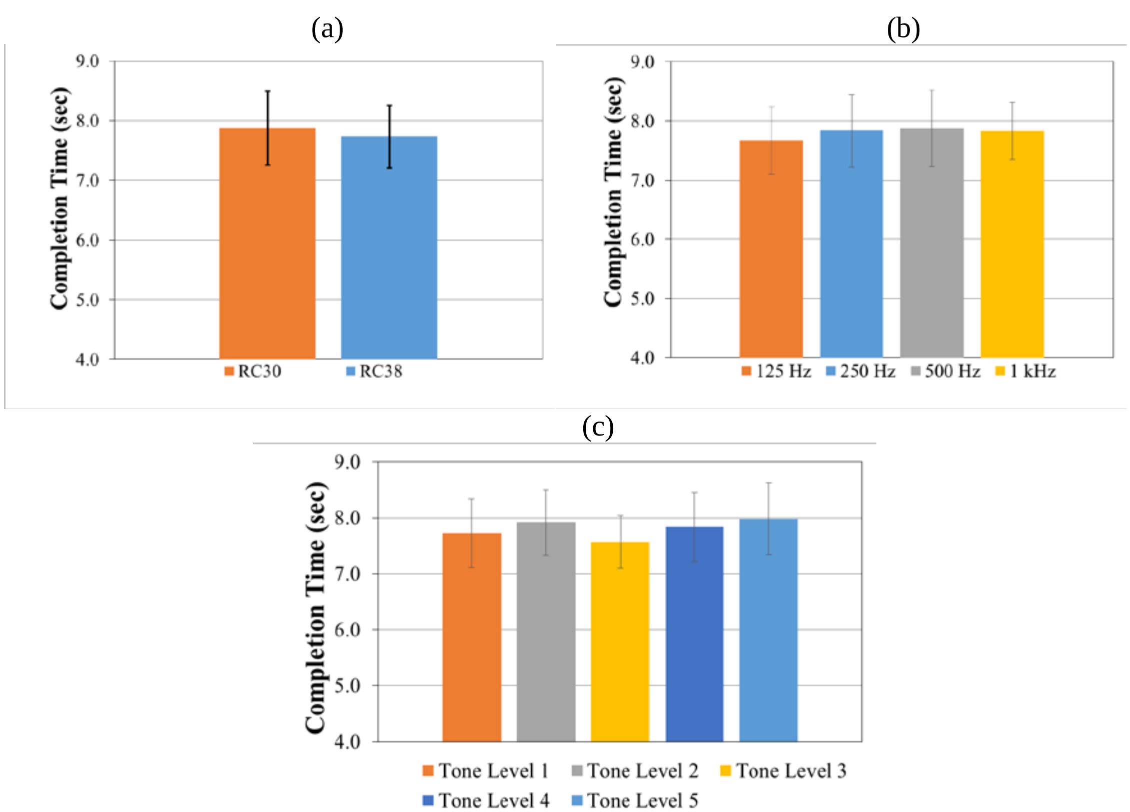


Figure 5.4 Effect of (a) background noise level, (b) tone frequency, and (c) tone level on completion time, where where Tone Level 1 indicates the least prominent tone and Tone Level 5 indicates the most prominent tone. Only tone level was found to be statistically significant. Error bars indicate on 95% confidence intervals.

5.3.2 Relationship Between Noise Metrics and Annoyance

To compare the relation between noise metrics and annoyance perception, Spearman's nonparametric correlation coefficients were calculated because the annoyance responses did not meet the normality assumption. Among the noise parameters previously introduced in Chapter 2, the following were chosen and calculated with the noise stimuli: Prominence Ratio (PR), Tone-to-Noise Ratio (TNR) and Tonal Audibility (ΔLta) for tonality parameters; un-weighted sound pressure level (SPLz), A-weighted sound pressure level (SPLa), ANSI Loudness level (ANSI Loudness) and ISO532B Loudness level (ISO Loudness) for loudness parameters; and Tone-corrected Perceived Noise Level (PNLT), Joint Nordic Method (JNM) and Sound Quality Indicator (SQI) for combined parameters.

The results are analyzed in three groups separately: first with all signals included and then with each base background noise level (Room Criteria 30 or 38) separately. Table 5.2 presents all correlation coefficients for each analysis. ANSI Loudness Level shows the highest correlation coefficients with annoyance ratings across all signals. When separating signals into the two background noise levels, though, tonality metrics show on par or slightly higher correlation with annoyance perception than loudness metrics. Among tonality metrics, Tonal Audibility demonstrates slightly better correlation than Tone-to-Noise Ratio and Prominence Ratio for all analyses. The results indicate that loudness is the most important feature of noise to predict annoyance perception, but then tonality of noise also should be included for the annoyance model, especially when background noise levels are kept constant. Combined metrics such as the Joint Nordic Method and Tone Corrected Perceived Noise Level and Sound Quality Indicator did not show better performance than loudness metrics, even though they were significantly related to annoyance ratings. The results show that imposing penalty values to loudness levels may not be the most effective way to quantify overall annoyance of the noise. These results confirm the same findings as the previous study in Chapter 4, even though it was conducted with different participants and context.

Table 5.2 Nonparametric Spearman correlation coefficients between noise metrics and annoyance perception (two-tailed, **p<0.01, *p< 0.05)

	Tonality			Loudness				Combined		
	PR	TNR	ΔL_{ta}	SPL_z (dB)	SPL_a (dBA)	ANSI Loud	ISO Loud	PNLT	JNM	SQI
All	.243	.277	.362*	.782**	.898**	.948**	.909**	.889**	.884**	.899**
RC-30N	.626**	.526*	.687**	.063	.709**	.718**	.559**	.650**	.732**	.754**
RC-38N	.470*	.875**	.895**	.412	.623**	.891**	.737**	.711**	.763**	.566*

A three-way repeated measure ANOVA was conducted as with the previous study in Section 4.3.2 to identify effects of background noise level, tone frequency, and tone strength on annoyance ratings. Mauchly's test indicated that the assumption of sphericity had been violated for the main effects of tonal strength [$\chi^2(9) = 28.51$, $p = 0.001$] and tone frequencies [$\chi^2(5) = 12.94$, $p = 0.024$]. Thus, degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity (ε=.49 for the main effect of tone strength and .70 for the main effect of tone frequencies). The analysis indicates that there was a significant main effect of background noise level [$F(1,15)=62.477$, $p<.001$]. RC-38N based noise signals were significantly more annoying than RC-30N based noise signals. There was also a significant main effect of the tone strength on annoyance perception [F(1.97, 29.57) = 78.82, p<.001]. Contrast reveals that the annoyance ratings of Tone Level 3 and above were significantly higher than the rating of Tone Level 1. Contrary to the finding in the previous study in Section 4.3.2, the tone frequency did not affect annoyance ratings significantly even though the annoyance ratings increased slightly as frequency increases. Figure 5.5 illustrates the mean annoyance ratings against background noise level, tonal frequency and tone strengths.

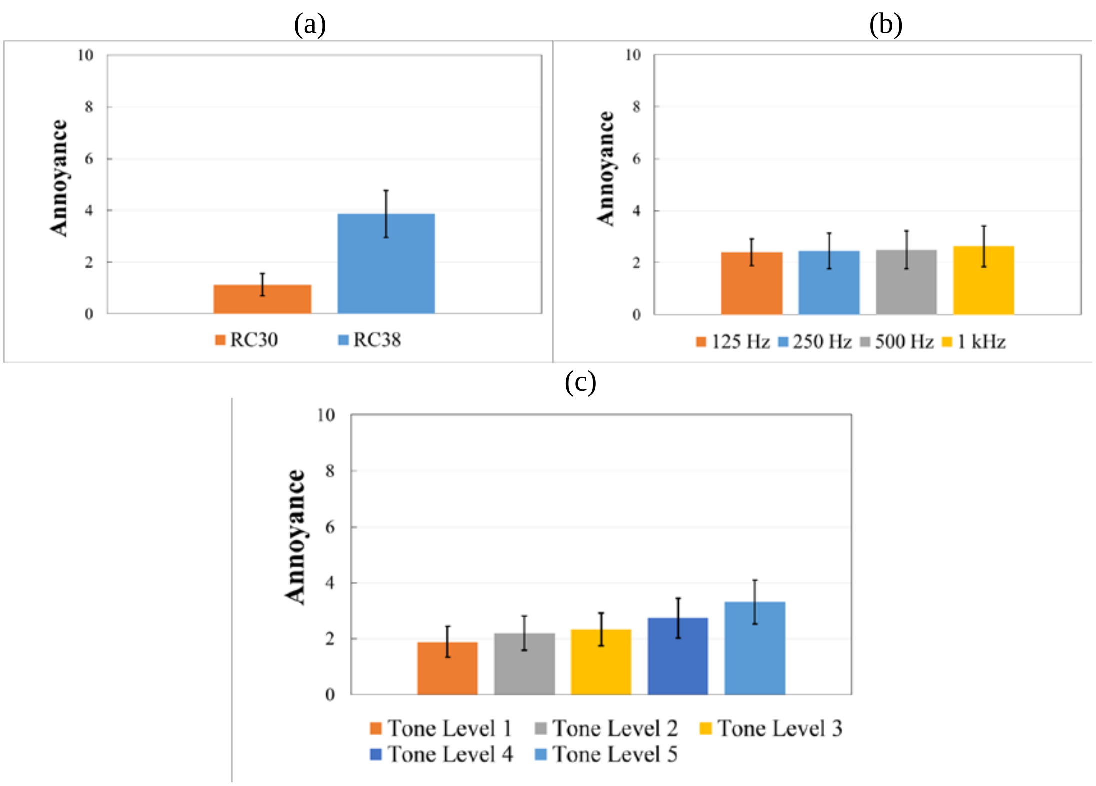


Figure 5.5 Mean annoyance perception ratings across participants plotted against (a) background noise level, (b) tonal frequency, and (c) strength of the tones, where Tone 1 indicates the least prominent tone and Tone 5 indicates the most prominent tone. Error bars indicate on 95% confidence intervals.

5.3.3 Dose-response Model

A dose-response model has been developed from the gathered complaint responses to determine thresholds of acceptability for tonality, using a binary logistic regression model. Based on the correlation analysis in Section 5.3.2, ANSI Loudness Level and Tonal Audibility are chosen as two prediction variables for the regression model. To compare the performance by using the dichotomous complaint responses, the same logistic regression models were calculated with % annoyed and % highly annoyed. The break-points to convert the continuous scale data to the categorical data were set to 5.0 and 7.2 respectively (Pedersen, 2007) for the percentage of annoyed and highly annoyed persons. Table 5.3 presents coefficient values and statistics for all three models. The chi-square (χ^2) value indicates how much the model prediction is improved against the model with no predictor and the R^2 is a measure of how well the prediction model fits the response data. The ratio indicates how the 'odds' of the outcome occurrence will change with a unit of predictor change.

The logistic regression equations for % Complaint, % Annoyed and % Highly Annoyed can be expressed as:

$$\%\ \text{Complaint} = \frac{1}{1+e^{(20.4-.29[\text{ANSI Loudness level}]\ -\ .04[\Delta L_{ta}])}}, \tag{5.1}$$

$$\%\ \text{Annoyed} = \frac{1}{1+e^{(21.8-.30[\text{ANSI Loudness level}]\ -\ .05[\Delta L_{ta}])}}, \tag{5.2}$$

$$\%\ \text{Highly Annoyed} = \frac{1}{1+e^{(23.8-.29[\text{ANSI Loudness level}]\ -\ .12[\Delta L_{ta}])}} \tag{5.3}$$

where % complaint is the percentage of possibility that complaints would be lodged against a particular tonal noise condition.

Table 5.3 Coefficients of the logistic regression model predicting whether a participant would (a) complain [95% BCa bootstrap confidence intervals based on 1000 samples], (b) be annoyed, or (c) be highly annoyed.

% Complaint				
	b	95% CI for Odds Ratio		
		Lower	Odds	Upper
Constant	-20.35 [-24.36, -16.96]			
ANSI Loudness (phon)*	.29 [.24, .35]	1.27	1.34	1.41
ΔL_{ta} (dB)**	.04 [.00, .09]	1.00	1.04	1.08

Note. R^2 = .24 (Hosmer & Lemeshow) .26 (Cox & Snell) .36 (Nagelkerke). Model $\chi^2(2)$=189.00, p<0.001. * p < .001. ** p = .05.

% Annoyed				
	b	95% CI for Odds Ratio		
		Lower	Odds	Upper
Constant	-21.77 [-26.89, -17.70]			
ANSI Loudness (phon)*	.30 [.24, .38]	1.27	1.35	1.45
ΔL_{ta} (dB)**	.05 [.00, .10]	1.00	1.05	1.10

Note. R^2 = .22(Hosmer & Lemeshow) .20 (Cox & Snell) .32 (Nagelkerke). Model $\chi^2(2)$=141.85, p<0.001. * p < .001. ** p = .03.

% Highly Annoyed				
	b	95% CI for Odds Ratio		
		Lower	Odds	Upper
Constant	-23.84 [-50.82, -15.65]			
ANSI Loudness (phon)*	.29 [.16, .69]	1.14	1.34	1.56
ΔL_{ta} (dB)**	.12 [.02 .22]	1.03	1.13	1.23

Note. R^2 =.18 (Hosmer & Lemeshow) .06 (Cox & Snell) .20 (Nagelkerke). Model $\chi^2(2)$= 38.61, p<0.001. * p < .001. ** p < .01

All models of % Complaint, % Annoyed and % Highly Annoyed are statistically significant (p<.001) and both predictors (ANSI Loudness Level and Tonal Audibility) significantly improve the model fit to complaint responses based on chi-square statistics. For instance, the % Complaint model yields a chi-square (χ^2) of 189.00, which is highly significant ($p < .001$). The accuracy of the model's prediction against observed responses was 76.4%. Figure 5.6 illustrates the logistic regression lines with actual responses which are expressed with dots. The dots represent calculated percentages of complaints, annoyed and highly annoyed for each of 40 noise signals.

The result shows that the % Complaint model is more similar to the % Annoyed model rather than % Highly Annoyed. The % Complaint model also showed better performance with regards to chi-square and R-squared statistics (χ^2 =189.00, R2=.24) than % Annoyed (χ^2 =141.85, R^2=.22) and % Highly Annoyed (χ^2 =38.61, R^2=.18) models. Current guidelines suggest dividing the continuous scale into certain breakpoints for the % Annoyed or % Highly Annoyed dose response models. However, the results from this study show that these dose-response models show lower chi-square statistics and wider confidence intervals. One reason for this is that subjects may still feel confused about meaning of the annoyance, even though they are informed about the definition of annoyance in the orientation session. The question on whether they are going to complain or not may feel easier for the subjects to answer, because it is a more behaviorally-based question. Another reason is that setting the breakpoint at 72 (or 50) points and over implies a very distinct difference for responses near the breakpoint; 73 points will be counted as annoyed but 71 points will be counted as not annoyed, even though the actual response difference is small. Thus, developing a dose-response model based on % Complaint, rather than % Annoyed or % Highly Annoyed, is recommended.

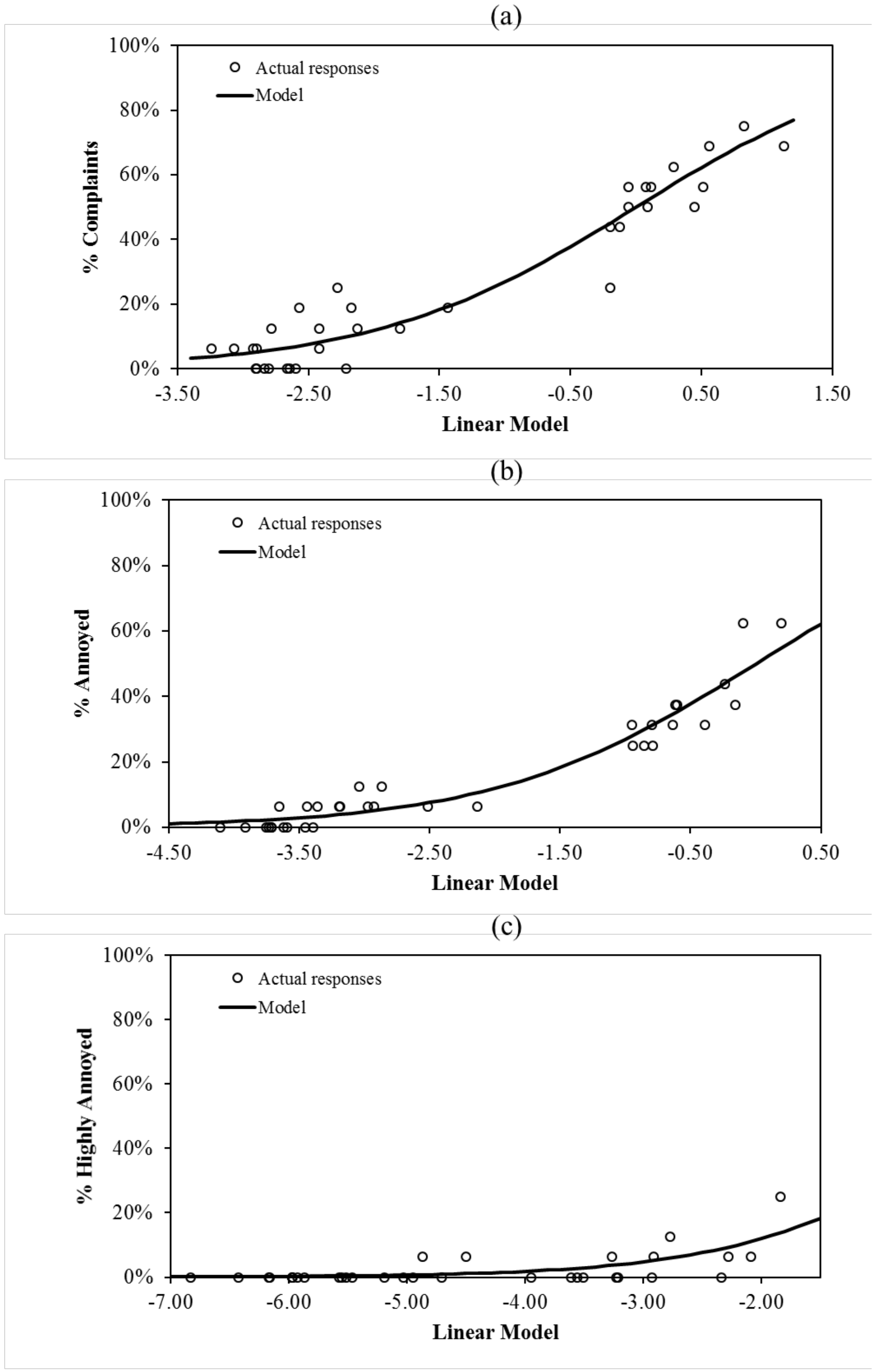


Figure 5.6 Dose response models of percentage of (a) Complaints from Equation 5.1, (b) Annoyed from Equation 5.2 (c) Highly Annoyed from Equation 5.3. The dots represent calculated percentages based on actual responses for each of 40 noise signals across parcitipants.

To suggest allowable tonality limits, the points at which 30%, 40% or 50% of participants would complain were determined from the logistic regression model to determine maximum Tonal Audibility, for a given ANSI Loudness level in phons (Fig. 5.7). The criteria lines in the figure demonstrate that the thresholds of acceptable tonality decrease as overall background noise level increases. The results mean that low levels of tonal components may not be acceptable when the overall background noise is loud. However, recommendations in Figure 5.7 are not practically applicable yet due to the small number of samples; the confidence intervals for the % Complaint model are still rather wide to generalize. This result should be verified with greater number of noise samples.

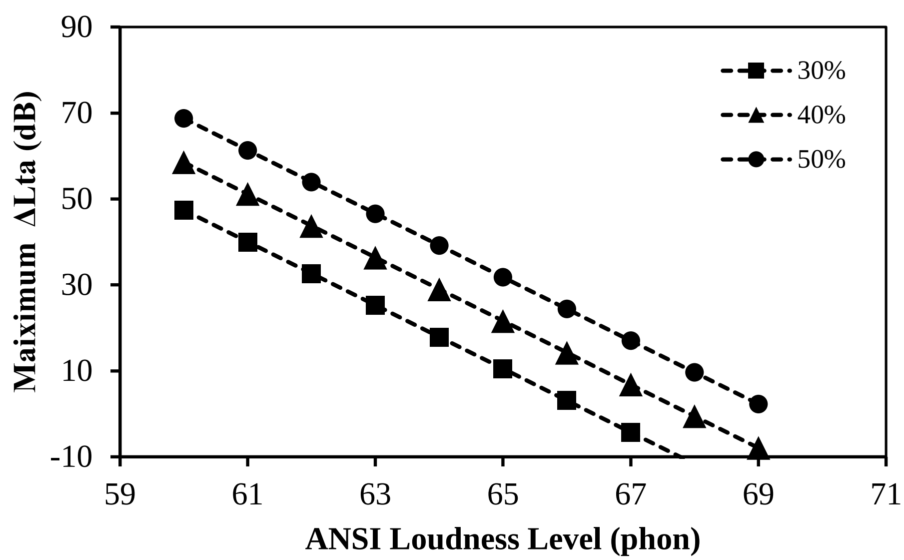


Figure 5.7 Maximum allowable Tonal Audibility criteria for given ANSI Loudness (phon). 30%, 40% and 50% of Complaints are chosen as guidelines.

5.4 Summary

The digit span task results reveal that even the least prominent tonal signals increased the time it took for participant to complete the digit span task, compared to broadband noise alone. Additionally, the level of tone affected the task performance in terms of completion times whereas a louder background noise level (RC-38 versus RC-30) and varying tone frequencies (from 125 Hz to 1 kHz) did not.

Based on the annoyance and likelihood-to-complain responses, a dose-response relationship has been developed. The reliability of the dose-response relationship depends on the selected noise metrics, which should correlate strongly to the perception of the noise. Based on correlational analyses from Section 5.3.2, the loudness metric ANSI Loudness Level showed the highest correlation overall to the annoyance responses, while the tonality metric Tonal Audibility also demonstrated significant correlation with the annoyance. Thus, these two noise parameters for loudness and tonality respectively were chosen to develop the dose-response relationship. Binary logistic regression models of the % Complaint, % Annoyed and % Highly Annoyed responses were developed. The % Complaint model fits the actual responses the best with the least wide confidence interval among the models, suggesting that similar studies in the future should focus on asking about the likelihood of subjects to complain due to a noise condition, rather than asking subjects to rate their annoyance. The % Complaint dose-response model is subsequently used to suggest maximum allowable tonality limits for a given ANSI Loudness level in phons based on the points at which 30%, 40% or 50% of participants would complain.

6. Chapter Six

Multidimensional Perception of Building Mechanical Noise

6.1 Introduction

This chapter presents a subjective study to investigate multidimensional perception of building mechanical system noise, including signals with tones. Unlike the previous studies presented in this dissertation which used synthesized noise signals with a single tone, actual noise recordings taken from building systems are included in this investigation. The motivation for this study is to explore other perceptual aspects besides loudness and tonality, which affect annoyance perception. Results are used to improve the annoyance model developed in the previous studies by adding in other significant acoustic characteristics.

6.2 Methods

6.2.1 Multidimensional Scaling Analysis

Multidimensional Scaling (MDS) analysis investigates how human subjects evaluate objects with a number of potentially unknown perceptual dimensions. Participants are usually asked to judge how similar a pair of objects is or how preferred one is over the other. MDS can be used for exploratory data analysis when the perceptions related to objects are not fully understood. In this dissertation, the MDS method is utilized to investigate other aspects of building mechanical system noises (if any) that impact their perception, outside of the loudness and tonality of the signal.

6.2.1.1 Obtaining proximity data

The data for multidimensional scaling analysis is called proximity data and is often handled in matrix form. There are mainly two methods to obtain proximity data: either directly from questions or transformed from other types of data. Proximity data can be directly derived by asking a question comparing all possible pairs of objects. The question is in most cases about how two objects are perceived to be similar or how much one signal is preferred over the other. Generally, a 5 to 9 point Likert scale is used with anchors that are labeled from "Very different" to "Very similar" for the similarity question.

The full matrix proximity data can be directly derived with all of the possible pairs of objects. To investigate n objects, $n*(n-1)/2$ paired comparisons are required, assuming the response is symmetrical. Asymmetrical MDS design and analysis is possible but the number of pairs to be answered increases. The advantage of directly obtaining proximity data is that it does not require any additional process to analyze the response. However, completing direct comparisons between all possible pairs can be a time-consuming task and may be fatiguing for some participants, especially when the number of objects under investigation is huge. Alternatively, researchers can randomly or systematically choose a portion of all possible pairs to reduce the number of trials per participant.

There are other indirect ways to obtain proximity data derived from other measures like confusion data or subjective clustering (Wickelmaier, 2003; Borg et al., 2012). These indirect methods are useful when using existing data. Because the indirect method is not utilized in this dissertation, further details are not discussed.

The required number of objects to obtain a reliable MDS solution is not explicit. The rule of thumb is to have more than four times the number of objects than dimensions under investigation (Green et al., 1989). It is also essential that the objects should demonstrate some differences for subjects to compare.

6.2.1.2 Multidimensional Scaling Analysis Algorithm

There are a number of algorithms for MDS, as developed in previous studies. Although there are differences in how they process the data, all MDS algorithms aim to derive the MDS solution with an optimal number of dimensions which have *distances* as close as possible to the raw proximity data. In MDS, the *distance* is a function that assigns values between two objects.

Each MDS algorithm differs in how it locates objects onto perceptual maps. The algorithm can use the aggregate values before the algorithm process, or it can average the individual results after the process. A combination approach, which is called Individual Differences Scaling (INDSAL) assumes that all test subjects share common dimensions but have different weighting values for each of the dimensions. Individual weight mapping then indicates how perceptual weights are different on each dimension amongst participants. In this study, INDSCAL was chosen because it can investigate individual differences but still obtain common perception mapping solutions.

Metric algorithm methods assume that the respondents' dissimilarity responses are metric data like interval and ratio level data while a non-metric MDS algorithm uses non-metric input data like rank order. The latter does not assume any type of relationship between distance and the input data. For the INDSCAL algorithm, both metric and non-metric methods are available. In this study, metric INDSCAL was used because the measured dissimilarity data were measured at the interval level.

6.2.1.3 Goodness-of-fit of MDS solutions

Once a solution is derived with the MDS algorithm, the goodness-of-fit of the solution should be evaluated. Since the MDS solution coordinates values with a certain number of dimensions, the goodness-of-fit evaluates how close distances of the coordinates values are to the proximity data.

There are mainly two methods that are applied to evaluate the goodness-of-fit. The first method is the Shepard diagram, which is a type of scatter plot with the proximity data along the x-axis and the distance data along the y-axis. The second method to assess the goodness-of-fit of a MDS solution is by calculating stress, or the squared difference between the proximities and the distances (Wickelmaier, 2003). The basic equation for stress is expressed as:

$$Stress = \sqrt{\frac{\sum (d(X) - \hat{d})^2}{\sum d(X)^2}} \tag{6.1}$$

where $d(X)$ and $\hat{d}$ indicate a distance function and raw proximity data, respectively. The stress indicates the amount of information loss from the proximity data when the raw data are represented by the MDS solution.

When determining the optimal number of dimensions for the MDS solution, a Scree plot is widely used. The Scree plot presents how the stress function changes as the number of dimensions increases. The lower the stress value is, the closer the MDS solution is to the original raw data. There is no strict rule to determine the number of dimensions needed for an MDS solution. Previous studies recommend the “elbow” point at which point including higher dimensions may represent only random components of the data (Borg et al., 2012) or the point where the stress value is below .05 (Wickelmaier, 2003).

6.2.1.4 Interpreting the MDS

There is no specific method for interpreting the meaning of each dimension from an MDS analysis. In this paper, correlation analysis with noise metrics are conducted with the MDS solutions to identify the perceptual meaning of each dimension; this method has been commonly used by others in noise studies (Susini et al., 2004; Woodcock et al., 2014) .

6.2.2 Noise Stimuli and Equipment

Fifteen actual audio recordings from building mechanical equipment and three artificially synthesized signals were used in this laboratory experiment. Assorted building mechanical equipment were included to have a wide range of noise stimuli in the tests. Three artificially synthesized broadband stimuli without tonal components were also included. Two of these followed the neutral Room Criteria contours of RC-38 and RC-51. The third stimulus had levels that were 12 dB higher in the 125 Hz octave band, above the RC-38 contour, giving a rumbly impression without any tonal components.

The sound levels of all signals were manually adjusted to be in the range of 45 dBA to 60 dBA while maintaining frequency spectrum. The tonality of the noise signals ranged from barely heard to prominent according to Tonal Audibility criteria (ISO, 2007). A few of them contained only a single tone characteristic, while others had fluctuating tonal characteristics, harmonic spectra, or inharmonic complex tone spectra. Table 6.1 lists each signal by its noise source, A-weighted equivalent noise level, the most dominant tone frequency, and general noise description. All noise stimuli were measured in the testing chamber at the listeners' ear position with a Larson Davis Sound Level Meter Model 831 for a minute.

Table 6.1 Description of noise signals

No.	Primary noise source	Noise Level (L_{Aeq})	Tone frequency	Noise description
1	condenser water pump	50.5	294	Single tone
2	radial blade pressure blower	57.2	313	Harmonics
3	water cooled screw chiller	51	297	Complex tone
4	vane axial fan	55.3	313	Complex tone
5	tube axial fan	50.1	155	Harmonics
6	heat pump	51.5	120	Single tone, fluctuating
7	outdoor condensing unit	54.9	41	Harmonics
8	digital compressor	54	95	Complex tone, fluctuating
9	heat pump	59.4	47	Harmonics, fluctuating
10	rooftop unit	48.6	119	Complex tone, fluctuating
11	heat pump	46.2	719	Complex tone
12	heat pump	46.8	119	Complex tone
13	lab fume hood	46.4	566	Complex tone
14	lab fume hood	47.5	234	Complex tone
15	screw compressor	47	593	Complex tone
16	Artificially synthesized	45.2	n/a	RC-38 neutral spectrum
17	Artificially synthesized	58.4	n/a	RC-51 neutral spectrum
18	Artificially synthesized	51.2	n/a	RC-38 rumbly spectrum

6.2.3 Subjective Testing Procedure

The test consisted of a half-hour orientation session and two half-hour main sessions, conducted on different days. In the orientation session, participants were informed briefly about the objective and methodology of the study, and they practiced the main task after completing a hearing screening test. During the main experiment, the participants completed a series of paired comparison tasks. They were asked to judge how two sound stimuli presented in a pair were similar in the first session, and which of a pair was perceived to be more annoying than the other in the second session. The responses were measured on 9-point Likert scales. Testing was administrated by a custom-coded program using a Matlab Graphic User Interface (GUI). Figure 6.1 illustrates the main display of the program for the subjective testing.

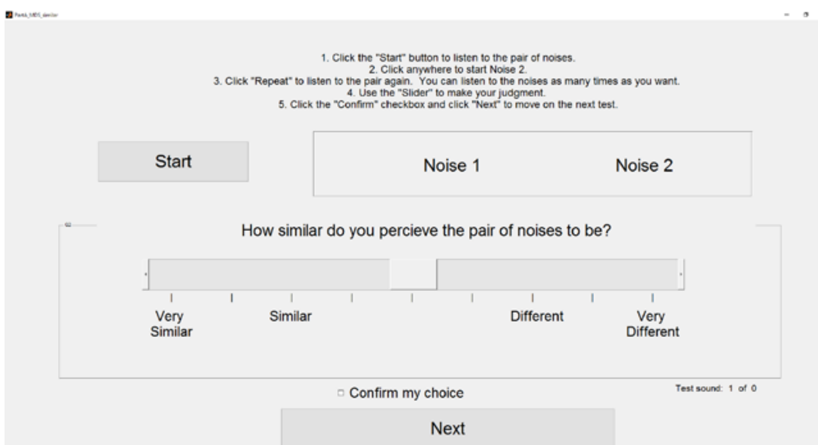


Figure 6.1 Subjective testing program interface for multidimensional scaling analysis

An incomplete cyclic test design was implemented instead of a complete set of paired comparisons to reduce the time it took to complete the sessions (Spence and Domoney, 1974). 72 trials were conducted in a session, which is 47% of a complete set of all possible pairs (153 for 18 noise signals). Efficiency, which is highly correlated with recovery measures, of the test design was 0.92 according to John et al. (1972).

6.2.4 Participants

Twenty adults (10 females, 10 males) were paid to participate in this study. They were recruited mainly from the University of Nebraska at Omaha north campus. The average age of the participants was 23.9 years with a standard deviation of 4.5 years. All subjects completed an orientation session with a hearing screening test and had lower hearing thresholds than 25 dB HL from 125 Hz to 8000 Hz for both ears. The noise sensitivity of each participant was also gathered at the orientation session by using a reduced NoiSeQ scale as in the previous two studies.

The consistency of participants' responses to the paired comparison tasks was checked before data analysis. Circular error rates (Parizet, 2002) were calculated to check each participants' responses consistency. The circular error rate counts inconsistent responses among multiple paired compression tasks. For example, the error occurs when a subject answers Signal A is more annoying than Signal B, Signal B is more annoying than Signal C, and Signal C is more annoying than Signal A. One participant whose error rates were above 15 percent (20%) was excluded from analysis.

6.3 Results and Discussions

The similarity responses were analyzed with the metric INDSCAL algorithm. First, the optimal number of dimensions for the MDS solution is determined by investigating the scree plot which plots the stress function against a number of MDS dimensions. Secondly, perceptual mapping with the obtained MDS solution is presented. Lastly, interpretation of each dimension is conducted with correlation analysis.

6.3.1 Similarity Task Results

Metric multidimensional scaling analysis was conducted with the similarity responses from the test. The individual difference scaling (INDSCAL) algorithm with individual weighting functions was used. First of all, the stress values were investigated with increasing number of dimensions. Figure 6.2 presents a Scree plot of how the stress function changes as the number of dimensions increases. Even though the elbow point is not very obvious in Figure 6.2, three or four dimensions appear to be the adequate choices to explain the raw data sufficiently. In this study, the MDS solution with four dimensions is chosen. The normalized raw stress value with four dimensions was 0.032.

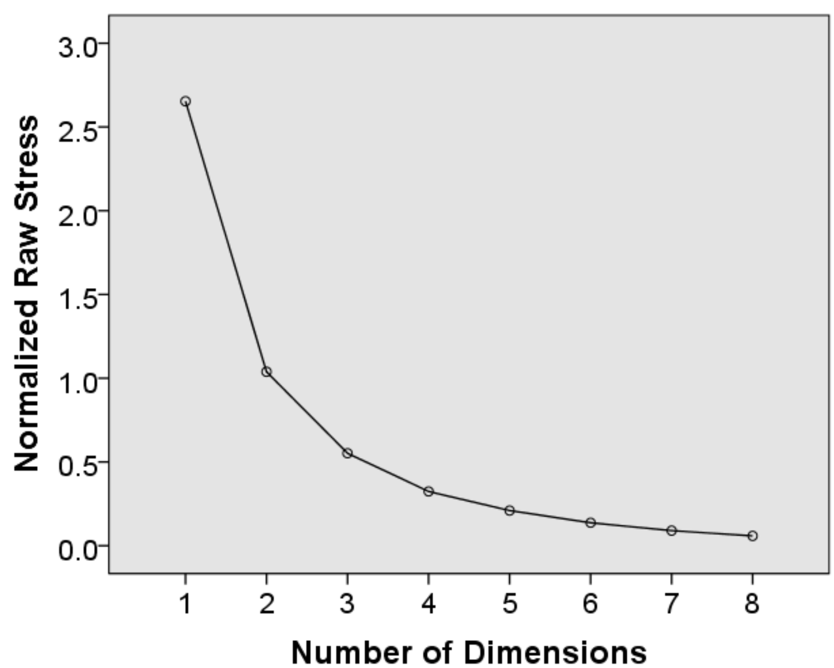


Figure 6.2 Scree plot of stress function with a number of dimensions for similarity task

Figure 6.2 presents the derived MDS solution with four dimensions, expressed through graphs with two dimensions each. The x-axis is dimension 1 for all plots while the y-axes are dimension 2, dimension 3, and dimension 4 in descending order.

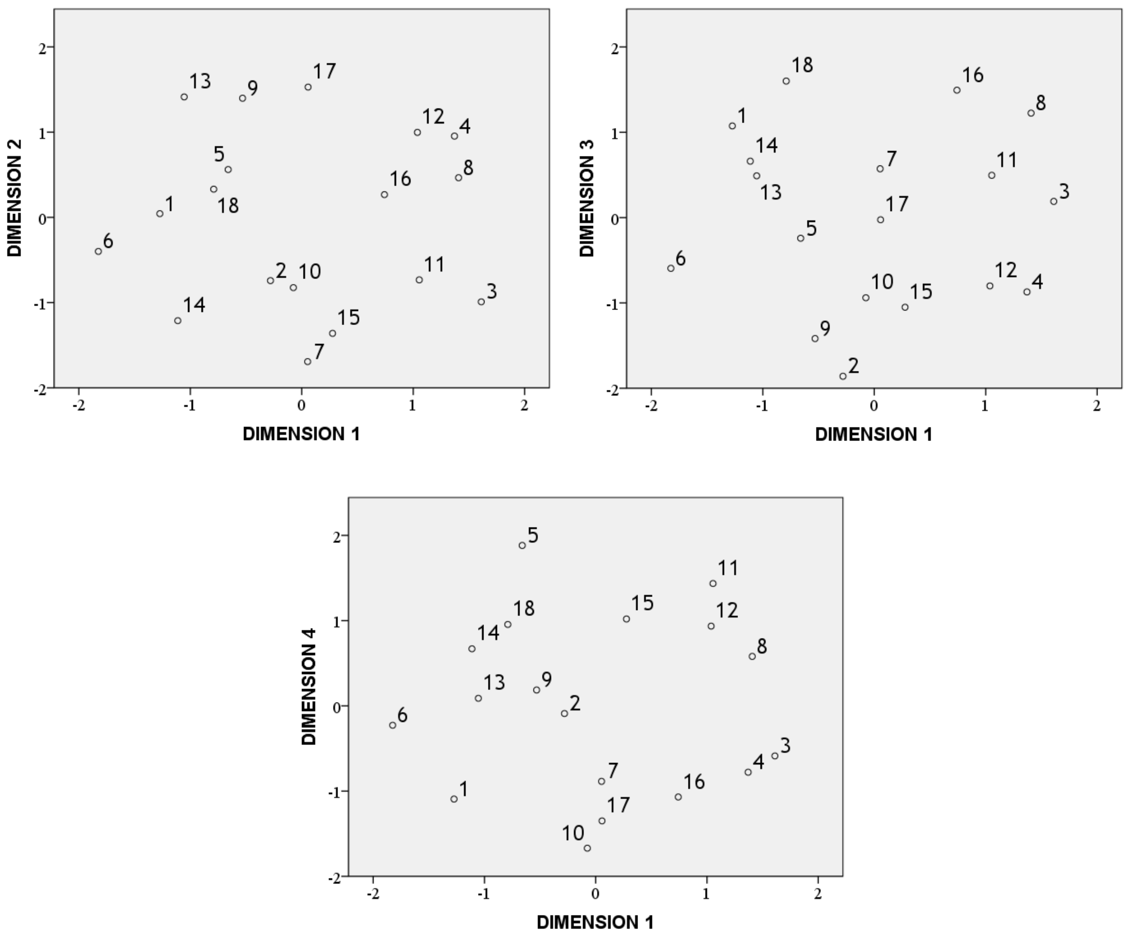


Figure 6.3 Signal coordinates expressed by four dimensions of MDS solution for similarity task

To interpret the dimensions as psychological structures, correlation analyses have been conducted with each dimension and assorted noise metrics describing the stimuli. All noise metrics introduced in Chapter Four and Five in addition to psychoacoustic parameters like Sharpness, Fluctuation Strength, and Roughness were calculated. Table 6.2 presents all correlation coefficients between noise metrics and perceptual dimensions. Dimension 1 was highly correlated with Sharpness perception; the correlation coefficient between the Sharpness measure and dimension 1 was .527 (p=0.25). Dimension 3 is highly correlated to tonality perception; the correlation coefficient between the dimension coordinates and Tonal Audibility metric was .651 (p=.003). Dimension 4 seems to be related to loudness perception even though the dimension coordinates are not significantly correlated with ANSI Loudness Level; the correlation coefficient was -.43 (p=0.07). The results indicate that the psychoacoustic characteristics of tonality and sharpness were more influential in determining the similarity results than the loudness perception.

Table 6.2 Spearman's correlation coefficients between psychoacoustic parameters and perceptual dimensions for similarity task

	Dimension 1	Dimension 2	Dimension 3	Dimension 4
PR	.092	-.052	.527*	.042
TNR	.120	.048	.582*	.043
ΔL_{ta}	.174	-.192	.651**	.039
dBA	-.028	.146	.282	-.278
ANSI Loudness	.282	.117	.373	-.430
ISO Loudness	.092	.063	.490*	-.397
PNL	-.088	.086	.352	-.340
PNLT	-.003	.088	.428	-.245
JNM	.096	-.026	.424	-.247
SQI	-.071	.117	.357	-.245
Sharpness	.525*	-.043	.321	-.220
Roughness	.117	-.042	-.044	.040
Fluctuation Strength	-.302	-.096	-.088	-.054

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).

There was no significant correlation between dimension 2 and any tested noise metrics. However, the noise signals with fluctuating tones (#6,8,9,10) were located closely along dimension 2 as shown in Figure 6.3. It suggests that dimension 2 may be related to fluctuating characteristics of the noise signal, although it was not significantly correlated with the psychoacoustic parameters like Roughness or Fluctuation Strength proposed by Fastl and Zwicker (2001). Further testing with more variety in fluctuating noise signals is required to identify the exact psychological structure with given dimensions.

6.3.2 Annoyance Regression Model

Since sharpness perception has been identified as one of the dominant perceptions in the MDS analysis on similarity data, the multiple regression model previously proposed in Section 4.3.3 is revised by including a sharpness metric in the model in addition to the loudness and tonality metrics.

To develop the annoyance regression model, the annoyance rating for each signal has to be determined. Because the subjective questions in this study followed the paired-comparison task of multidimensional scaling analysis methods, relative annoyance ratings can be calculated by methods used in previous research (Parizet et al., 2005; Woodcock et al., 2014):

$$A_{i,s} = \frac{1}{N_i}\sum_{j \neq i} P_{j,i,s} \tag{6.2}$$

where $A_{i,s}$ is the relative annoyance rating for participant *s* by signal *I*; N_i is the number of times signal *i* was asked in the subjective test; and $P_{j,i,s}$ is the relative paired comparison rating for signal *i* over *j* by participant *s* from preference tasks.

Annoyance ratings for each signal calculated from the obtained preference data were used to revise the annoyance prediction model with ANSI Loudness Level, Tonal Audibility and Sharpness. Table 6.3 presents standard error of coefficients, standardized coefficients, and statistical significance when ANSI Loudness Level and Tonal Audibility were only used (in step 1) and when Sharpness was also included (in step 2), in addition to the coefficient values for each predictor. Equation (6.3) presents the multivariate regression model with ANSI Loudness Level, Tonal Audibility, and Shaprness.

$$\begin{aligned} Annoyance \; &= .20 * ANSI\ \ Loudness(sone) + .01 * Tonal\ \ Audibility\left(dB\right) \\ &+ 1.11 * Sharpness(acum) - .19 \end{aligned} \tag{6.3}$$

Table 6.3 Linear regression model of predictors for annoyance perception with 95% bias corrected and accelerated confidence intervals reported in parentheses.

	b	Standard error B	β	p
		Step 1		
Constant	1.204	.540		
ANSI Loudness (sone)	.188 (.131, .253)	.034	.870	.001
Tonal Audibility (dB)	.043 (-.002, .081)	.20	.251	.048
		Step 2		
Constant	-.186	.575		
ANSI Loudness (sone)	.200 (.146, .266)	.034	.923	.001
Tonal Audibility (dB)	.008 (-.026, .048)	.018	.048	.630
Sharpness (acum)	1.112 (.472, 1.498)	.241	.329	.007

Note. R^2 *= .82 for Step 1;* ΔR^2 *= .06 for Step 2 (ps = .025).*

The R^2 value for the step 1 model is .82, which is a measure of goodness-of-fit of the linear regression, indicating that 82% of the annoyance rating variance can be explained by the ANSI Loudness and Tonal Audibility model only. When including Sharpness, the

R^2 value increased to .88. The revised multivariate regression model does significantly predict more variation in annoyance perception when including Sharpness perception. Figure 6.4 illustrates a regression line with the calculated linear model, as compared to Figure 4.4 with the former model developed in Chapter 4.

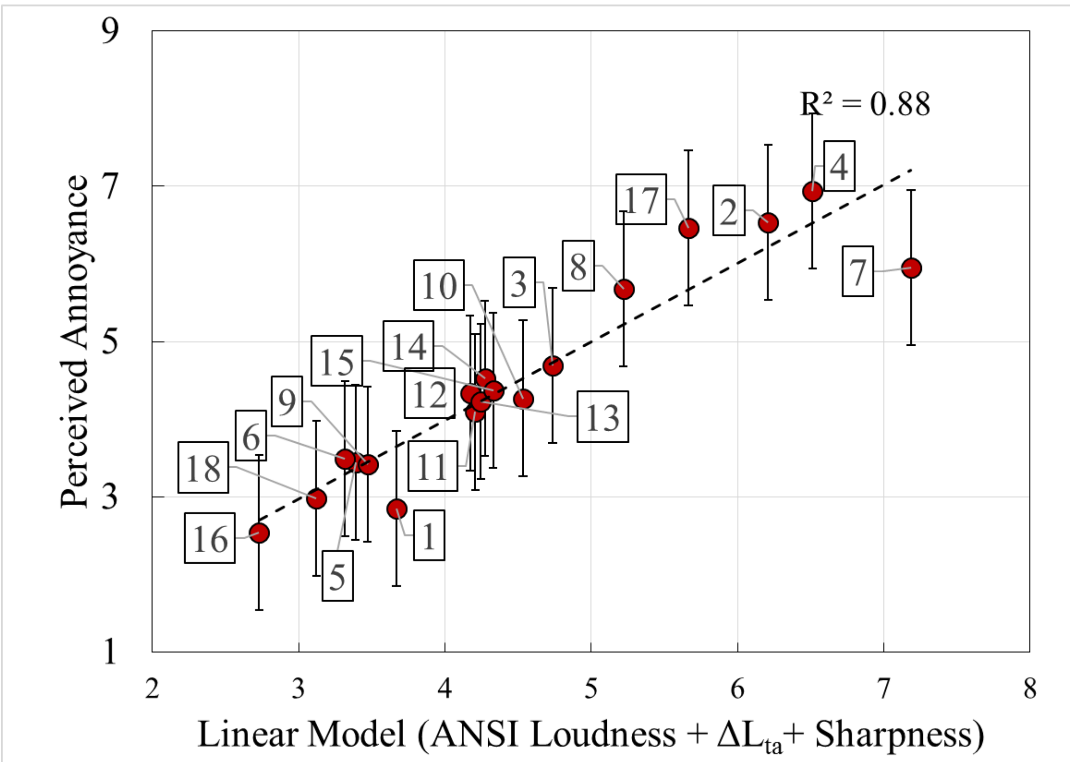


Figure 6.4 Averages (mark) and standard deviations (error bar) of the annoyance ratings across 19 participants for each noise stimulus plotted against the proposed linear regression model of annoyance perception from Equation 6.3 (dashed). The model is based on ANSI Loudness Level, Tonal Audibility and Sharpness (R^2=.88). The noise stimuli are labelled with assigned numbers from Table 6.1.

6.4 Summary

This chapter investigated the multidimensional characteristics of tonal noise from HVAC systems with the MDS method. Paired comparison tasks were conducted to gather both similarity and preference data using both actual HVAC recordings and artificially synthesized signals. The test results show that the latent psychological structures were related to the tonality, loudness and sharpness perceptions of the noise stimulus. A revised annoyance prediction model including the sharpness perception showed better performance against annoyance ratings. The noise signals with fluctuating tones were located closely in MDS dimension coordinates but there was no statistically significant relation with noise metrics such as Roughness and Fluctuation Strength due to small sizes of noise samples.

7. Chapter Seven

Perceptual Weight of Multi-tone Complex of Annoyance Perception

7.1 Introduction

Assorted building mechanical systems generate tonal components within the background noise of built environments. In most cases, this type of noise includes multiple tones in harmonic or inharmonic structures rather than a single tone. However, there is limited information on the comprehensive annoyance caused by multiple tones as perceived by human occupants. Two current standards, ISO 1996-2 and ANSI S1.13, propose calculation methods to address tones in noise but those methods only analyze the tones individually. These tonality metrics from the two standards can result in the inaccurate prediction of overall annoyance perception. This chapter aims to investigate how each tone contributes to overall annoyance perception when complex tones are present in background noise. A subjective study with two different structures (harmonic and inharmonic distribution) of five tone components was conducted. Perceptual weighting analysis is applied to the results to compute a spectral weighting function for overall annoyance. The performance of the derived spectral weighting function is examined against annoyance ratings of actual building mechanical noises.

7.2 Methods

7.2.1 Perceptual Weighting Analysis

Perceptual weight analysis (or molecular psychophysics) method provides the relative weights of each component of perceptual features such as loudness by trial-by-trial analysis (Berg and Green, 1990; Lutfi and Jesteadt, 2006). As the level or magnitude of each component varies randomly, subjects are usually asked to choose a noise stimulus in a pair based on loudness or preference perception. Relative weights and global perception can be modeled as:

$$D = \sum_{i=1}^{m} w_i x_i + C, \quad \sum_{i=1}^{m} w_i = 1 \tag{7.1}$$

where D is the participants' decision, w_i is the perceptual weight for the ith component, x_i is the magnitude difference between a pair of the noise stimuli, C is a constant, and m is the total number of components in the noise stimulus (Leibold et al., 2007). Because relative weights are under investigation in most cases, weighting values w_i are normalized to have unity when all components are summed up.

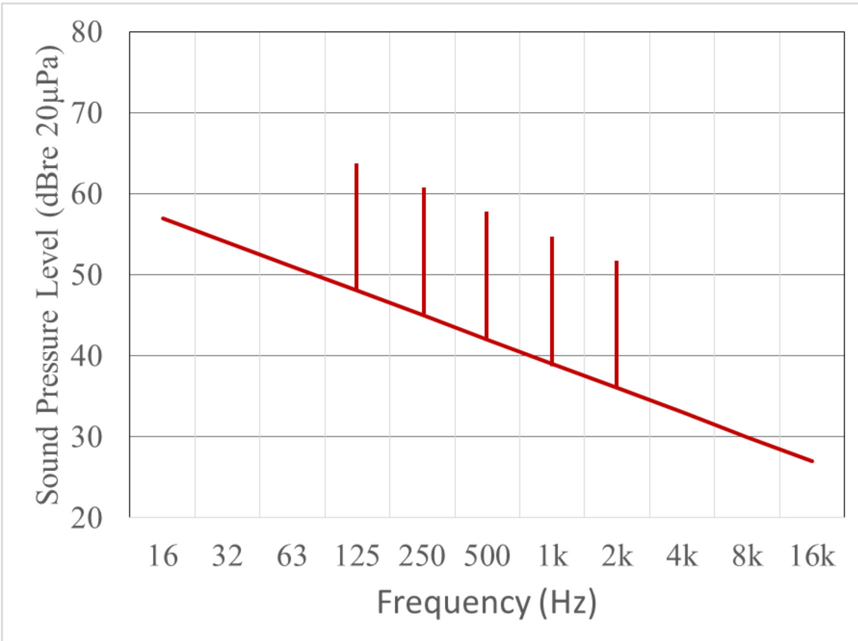
Multiple linear regression between variations of each component and responses provides the relative weighting of components (Leibold et al., 2007; Jesteadt et al., 2014). The main research area of perceptual weight analysis is investigating spectral components (Leibold et al., 2007; Jesteadt et al., 2014) or temporal components (Oberfeld et al., 2012) of complex noises contributing to overall loudness.

Perceptual weight analysis has not been widely used in annoyance studies. Dittrich and Oberfeld (2009) first adopted this method to an annoyance study. They investigated annoyance and loudness perception of temporal varying stimuli. They found that temporal weighting improved the prediction of loudness and annoyance, and the annoyance responses by the listeners were significantly different from the loudness responses. One of the biggest challenges in tonal noise is to investigate effects of harmonics on overall tonality and annoyance perception, and this methodology is the ideal method to explore this research question.

7.2.2 Noise Stimuli and Equipment

The signals were pink noise with added five tone complexes. The broadband pink noise spectrum signal was generated by using the program Test Tone Generator by Esser Audio. The overall level of the pink noise signal was 57 dB SPL, and the frequency spectrum decreased at the rate of 3 dB per octave. Five tone complexes were added to the pink noise to generate test signals. Two different frequency structures were used. For the harmonic structure utilized in the main session 1 and 2, tones of 125, 250, 500, 1000, 2000 Hz were used. For the inharmonic structure utilized in the main session 3 and 4, tones of 125, 200, 430, 910 and 1890 Hz were used by slightly shifting the tones to be heard separately. The level of all individual tones in a reference signal was set to be 12 dB above from the pink noise octave level of the center frequency. For the comparison signal, levels of each tone were randomly varied from a rectangular distribution with a range of 16 dB and a step size of 4 dB, centered on 12 dB above the pink noise octave spectrum for each trial. That is, individual tones could vary from +4 dB to +20 dB in steps of 4 dB above the pink noise level. Figure 7.1 illustrates sample frequency spectra of the reference and comparison stimuli and shows calculated Tonal Audibility values of tone components for the reference signal.

(a)



(b)

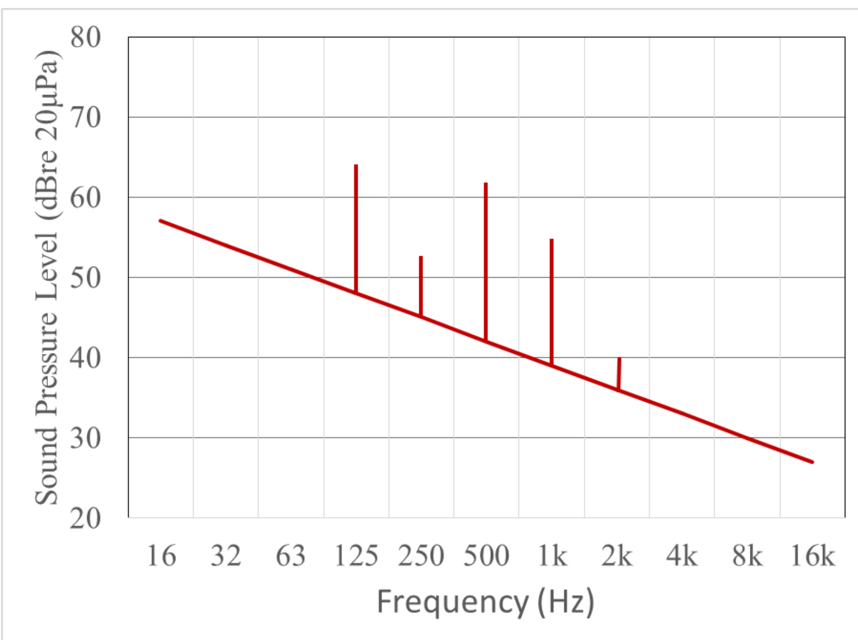


Figure 7.1 Frequency spectrum illustration of (a) the reference signal and (b) comparison signal

Table 7.1 Tonal Audibility values of tone components for the reference signal used in the subjective test. Tonal Audibility values of comparison signals vary randomly from -8 dB to +8 dB with a step size of 4 dB from the reference signal values.

Frequency (Hz)	125	250	500	1000	2000
Tonal Audibility (dB)	15.0	14.1	14.3	14.8	15.5

7.2.3 Subjective Testing Procedure

The subjective test aims to determine perceptual weighting functions of the noise signals by using a two-interval, annoyance judgment task. Participants first took an orientation session for an hour. In the orientation session, participants filled out a questionnaire on their musical experiences and noise sensitivity questionnaires. Participants were informed about the definition of annoyance and how it is different from loudness. The participants were also asked to imagine themselves hearing the noises in their office while working as the context of the study. The noise sensitivity survey applied the NoiseEQ scale used in the previous studies. In the main study, participants completed 4 hour-long sessions of paired-comparison tasks to choose more annoying noise stimuli. In each session, participants completed 500 paired comparison tasks with 2-minute breaks for every 100 trials. Figure 7.2 presents the computer test program interface in the main session.

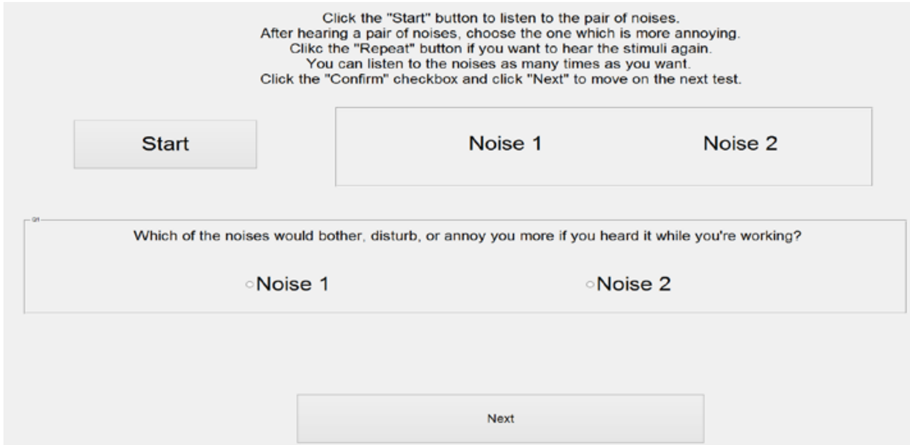


Figure 7.2 Main session test program interface

7.2.4 Participants

Ten participants (4 males, 6 females) with at least 3 years musical experiences were recruited in this study through recruiting flyers. They were recruited mainly from the University of Nebraska at Omaha campus. The average age of the participants was 25.8 years with a standard deviation of 9.6 years. The average musical experience period of the participants was 14.5 years with a standard deviation of 13.2 years. The participants completed an orientation session with a hearing screening test. All participants had a normal hearing sensitivity with thresholds below 25 dB HL from 125 Hz to 8000 Hz for both ears.

7.3 Results and Discussions

7.3.1 Reliability of Data

Prior to analyzing the perceptual weighting functions, the reliability of the participants' responses are examined by calculating split-half reliability (Jesteadt et al., 2014). The individual responses were divided into halves by separating odd and even numbered responses. The perceptual weights were then calculated with the odd or even numbered responses separately. The weight values were calculated with 250 responses 10 perceptual weights per subject were used to calculate correlation coefficients for the split-half reliability. Table 7.2 presents all participants' response reliabilities along with their self-reported number of years of musical education experience and noise sensitivity calculated from the NoiseEQ scale. Generally, a coefficient value above .8 is considered to be reliable. All participants' showed reliability above .9. Thus, all participants' perceptual weight results are included in the following analyses. There was no statistically significant correlation found between the reliability, musical experience and noise sensitivity.

Table 7.2 Split-half reliability of each participant's responses.

Participant	Split half, r	Musical Experience (years)	Noise Sensitivity, NoiseEQ
1	.98	44	2.38
2	.99	13	3.62
3	.94	3	2.77
4	.95	25	2.77
5	.96	9	2.85
6	.98	20	3.08
7	.99	5	3.08
8	.97	3	2.46
9	.94	3	3.15
10	.97	20	3.08

7.3.2 Perceptual Weighting Function

Perceptual weight functions are derived for each participant by calculating multiple linear regression models between level differences of each tone and dichotomous subjects' responses. The regression coefficients of the tones were then normalized to sum to unity. Table 7.3 presents all perceptual weights calculated from each participant's responses. There was a statistically significant relationship in the multiple regression models between tone components and participants' response except for the first tone component ($p < .05$) for subjects 3, 4, 5. For subject 1, all tone components were statistically significant. For subject 2, the third tone component in the harmonic structure, and the first and the second tone in the inharmonic structure were not statistically significant. For subject 7, the second tone component was not statistically significant in the harmonic structure, and the first and fourth tone were not statistically significant in the inharmonic structure. For subject 10, the third and the fourth tone in the harmonic structure and the second and the fourth tone in the inharmonic structure were not statistically significant. Figure 7.3 illustrates average perceptual weights across all participants for the harmonic and inharmonic conditions separately.

Table 7.3 Normalized perceptual weights of five tone components for each participant. P values of each weight, average across participants and standard errors of average weights are also presented.

Harmonic Structure										
	125Hz Tone		250Hz Tone		500Hz Tone		1000Hz Tone		2000Hz Tone	
Subject	weight	p	weight	p	weight	p	weight	p	weight	p
1	0.09	.00	0.24	.00	0.32	.00	0.15	.00	0.21	.00
2	0.08	.00	0.13	.00	0.00	.83	0.05	.03	0.73	.00
3	0.03	.13	0.15	.00	0.30	.00	0.25	.00	0.27	.00
4	0.03	.11	0.16	.00	0.34	.00	0.15	.00	0.31	.00
5	0.01	.62	0.12	.00	0.35	.00	0.17	.00	0.34	.00
6	0.03	.34	0.14	.00	0.34	.00	0.12	.00	0.37	.00
7	0.07	.00	0.03	.07	0.29	.00	0.14	.00	0.47	.00
8	0.06	.01	0.16	.00	0.31	.00	0.13	.00	0.34	.00
9	0.03	.33	0.22	.00	0.28	.00	0.33	.00	0.14	.00
10	0.18	.00	0.09	.00	0.05	.06	0.04	.20	0.64	.00
Mean	0.06		0.14		0.26		0.15		0.38	
SE	0.02		0.02		0.04		0.03		0.06	
Inharmonic Structure										
	125Hz Tone		200Hz Tone		430Hz Tone		910Hz Tone		1890Hz Tone	
Subject	weight	p	weight	p	weight	p	weight	p	weight	p
1	0.08	.00	0.33	.00	0.33	.00	0.10	.00	0.16	.00
2	0.03	.20	0.00	.91	0.05	.01	0.11	.00	0.81	.00
3	0.02	.54	0.34	.00	0.23	.00	0.11	.00	0.30	.00
4	0.03	.32	0.25	.00	0.22	.00	0.09	.00	0.40	.00
5	0.04	.13	0.15	.00	0.33	.00	0.14	.00	0.35	.00
6	0.10	.00	0.29	.00	0.24	.00	0.08	.00	0.29	.00
7	0.01	.56	0.17	.00	0.14	.00	0.05	.06	0.62	.00
8	0.01	.74	0.33	.00	0.25	.00	0.05	.04	0.37	.00
9	0.09	.01	0.33	.00	0.25	.00	0.14	.00	0.19	.00
10	0.08	.00	0.05	.08	0.08	.00	0.03	.22	0.76	.00
Mean	0.05		0.22		0.21		0.09		0.43	
SE	0.01		0.04		0.03		0.01		0.07	

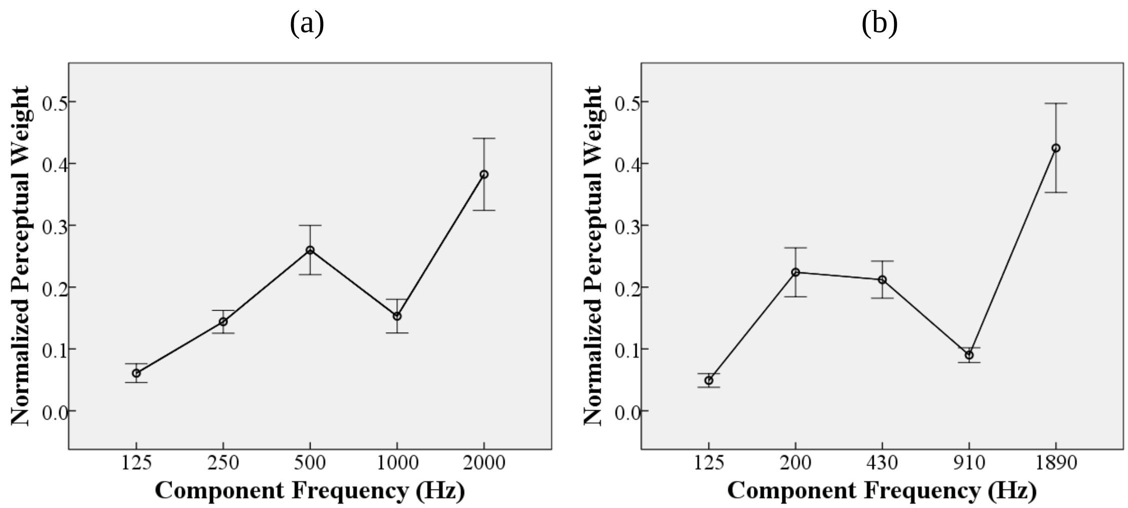


Figure 7.3 The mean perceptual weights for each tone component across participants. The perceptual weights are normalized to have a total sum of one. (a) The left graph shows the perceptual weight values for harmonic structure stimuli in the first two session and (b) the right graph shows the values for inharmonic structure stimuli in the last two session. Error bars represent +/-1 standard error.

The first tone component was not significant or had nearly zero weight values across participants. The range of perceptual weight values was wider for the higher frequency tone components. For both harmonic and inharmonic structures, the highest weight is observed at the highest frequency. A prime difference between the two structures was found at the second tone components of 250 Hz and 200 Hz. In the inharmonic structure, the subjects assigned higher weights to the second, 200Hz, tone component unlike the weight assigned to the 250 Hz tone component in the harmonic structure.

Repeated-measure factorial ANOVA confirmed the trend in Figure 7.3. The two structure types and five tone components were taken as independent variables, and the regression coefficients were taken as dependent variables. Mauchly's test indicated that the assumption of sphericity for repeated-measure ANOVA was violated for the effect of tone ($\chi^2(9)=49.87$, $p<.001$) and for the structure and tone interaction ($\chi^2(9)=30.98$,

$p<.001$). Thus, Greenhouse-Geisser corrected degree of freedom was used. The repeated-measure ANOVA confirmed that there was no significant main effect of structure, but there was a significant effect of tone [$F(1.15,10.34)=20.47$, $p=.001$] and significant interaction of structure and tone [$F(1.37,12.29)=5.03$, $p=.035$] on the annoyance perceptual weights.

7.3.3 Application of Weighting Functions

A new way of calculating tonality metrics is proposed using the obtained perceptual weighting functions. The existing Tonal Audibility metrics use the most prominent single tone to calculate a single number rating, even for a complex tone stimulus. It means that all the other harmonic and inharmonic tone information is not considered. The obtained perceptual weighting functions can be used to obtain the comprehensive annoyance rating for complex tones.

The Weighted-sum Tonal Audibility ($\Delta L_{ta,w}$) is developed by revising the Tonal Audibility calculation method. Figure 7.4 illustrates the process of calculating the Weighted-sum Tonal Audibility. First, the frequency spectrum of the noise stimulus is analyzed by using FFT. Tonal Audibility values are then calculated for all prominent tones. Then, a normalized perceptual weighting function is applied. Applying the harmonic or inharmonic perceptual functions depend on a prevailing tone structure of the noise stimulus. Lastly, all of the weighted Tonal Audibility values for each tone in the stimulus are summed to calculate a single number rating. Table 7.4 presents the previous Tonal Audibility and the new Weighted-sum Tonal Audibility values for all noise stimuli used in Chapter 6.

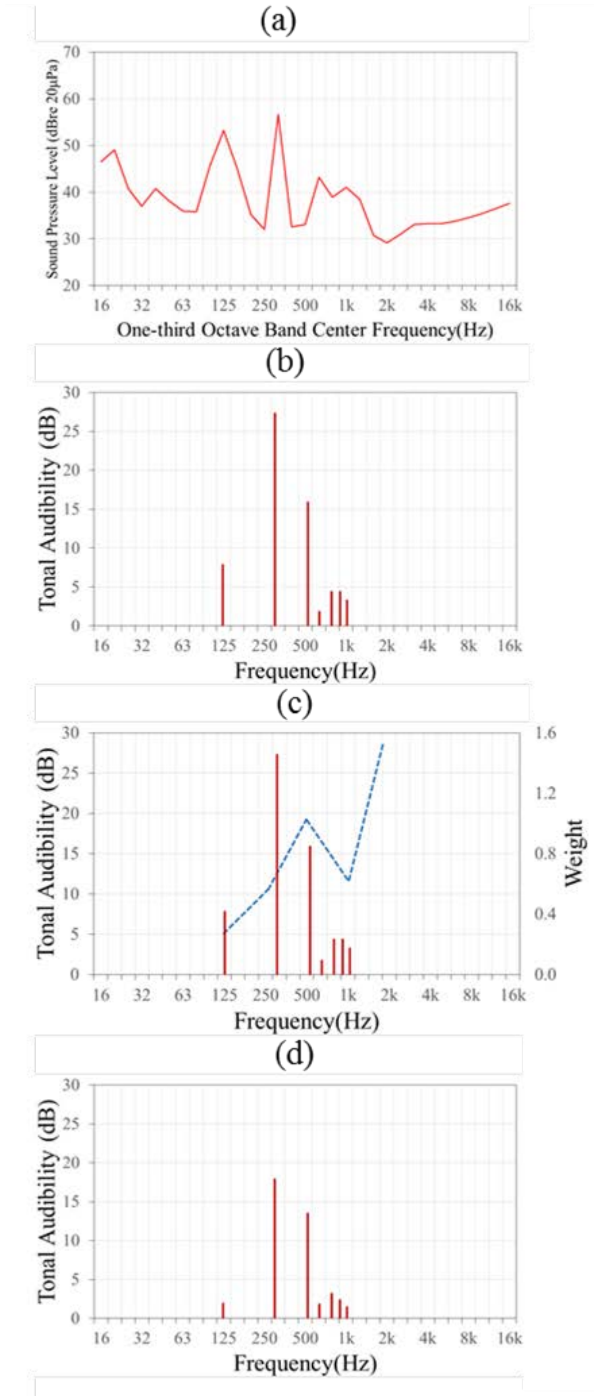


Figure 7.4 Example of process on applying perceptual weigthing functions (a) one-third octave band spectrum of noise signal, (b) tone extraction calculated by Tonal Audibility, (c) overlapping the perceptual weighting function with the Tonal Audibility values, and (d) the result of applying the weighting function to the individual Tonal Audibility values.

Table 7.4 Description of noise signals with additional Weighted-sum Tonal Audibility.

No.	Primary noise source	ANSI Loud-ness (sone)	Tone Frequency (Hz)	Tonal Audibility (dB)	Weighted-sum Tonal Audibility (dB)
1	condenser water pump	50.5	294	9.5	6
2	radial blade pressure blower	57.2	313	17.5	9.8
3	water cooled screw chiller	51	297	27.5	19.9
4	vane axial fan	55.3	313	21.7	15.5
5	tube axial fan	50.1	155	23.0	10.6
6	heat pump	51.5	120	15.6	5
7	outdoor condensing unit	54.9	41	14.0	6.3
8	digital compressor	54	95	11.9	6.1
9	heat pump	59.4	47	27.7	7.8
10	rooftop unit	48.6	119	23.0	8.6
11	heat pump	46.2	719	11.2	11.2
12	heat pump	46.8	119	14.7	11.7
13	lab fume hood	46.4	566	8.4	13.3
14	lab fume hood	47.5	234	11.5	13.9
15	screw compressor	47	593	12.4	14.3
16	RC-38 neutral spectrum	45.2	n/a	0	0
17	RC-51 neutral spectrum	58.4	n/a	0	0
18	RC-38 rumbly spectrum	51.2	n/a	0	0

The new Weight-sum Tonal Audibility should be used with a caution due to following limitations. Determining a perceptual weighting function between for the harmonic and the inharmonic structure is rather subjective because, in many noise stimulus, harmonic and inharmonic tones are blended in the same stimuli. Also, the suggested weighting functions use a linear interpolation for any tone frequencies between examined tones. More weighting functions should be examined with assorted scenarios of other frequency and sound level ranges.

The performance of the new Weighted-sum Tonal Audibility was compared to that using the previous Tonal Audibility with the annoyance regression model which was developed in Section 6.3.2. The same annoyance ratings and noise stimuli were used to test the new Weighted-sum Tonal Audibility. The developed annoyance model in Section 6.3.2 utilized ANSI Loudness Level, Tonal Audibility, and Sharpness. Three regression models were compared in this section. Model 1 included ANSI Loudness Level and Tonal Audibility, and Model 2 included ANSI Loudness Level, Tonal Audibility, and Sharpness as the same annoyance model in the previous Section. Model 3 included ANSI Loudness Level and the Weighted-sum Tonal Audibility. Figure 7.5 illustrates regression lines with these models.

As presented in the Table 6.3, the goodness-of-fit of the regression model (R^2) was .88 for the Model 2 with the three noise metrics. For the Model 1 without Sharpness, the R-square change (ΔR^2) by adding Tonal Audibility was only .03 (p=.047), and the goodness-of-fit of the regression model (R^2) is .82 for the Model 1. When using the new Weighted-sum Tonal Audibility metrics in the Model 3, the goodness-of-fit (R^2) was improved to .88 by increasing the R-square change (ΔR^2) to .06 (p=.002). The R^2 values for the Model 2 and the Model 3 were almost the same. Including Sharpness to the Model 3 didn't improve the goodness-of-fit. It was mainly because the Weighted-sum Tonal Audibility and Sharpness accounted for the same variances of the annoyance ratings. Equation 7.2 presents the regression model with ANSI Loudness Level and Weighted-sum Tonal Audibility.

$$\begin{aligned} Annoyance = &.20 * ANSI\ Loudness(sone) + .08 * Weighted\_sum\ Tonal\ Audibility(dB) \\ &+ .91 \end{aligned} \tag{7.2}$$

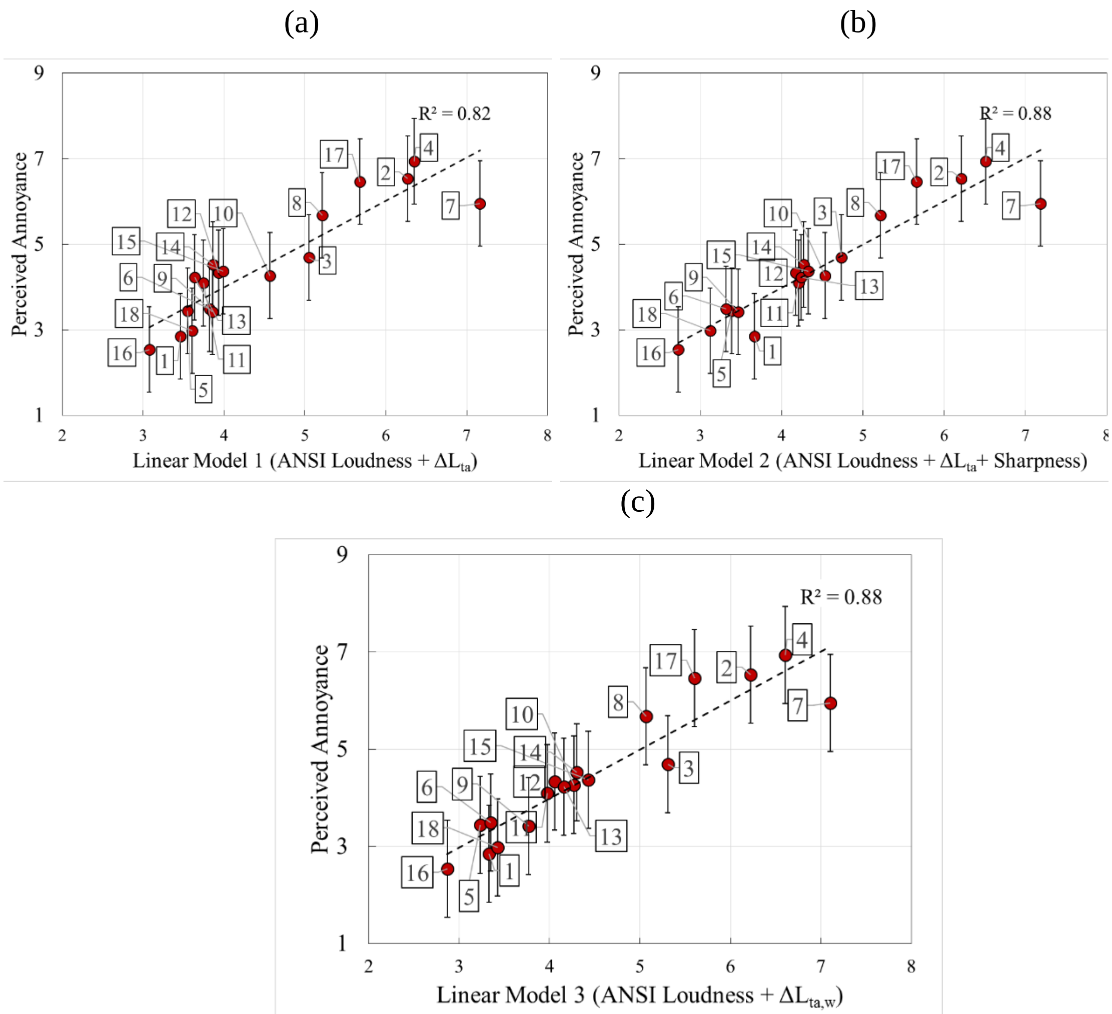


Figure 7.5 Averages (mark) and standard deviations (error bar) of the annoyance ratings across participants for each noise stimulus from Chapter 6. The dashed lines represent the linear regression models with ANSI Loudness Level and (a) Tonal Audibility (Model 1, R^2=.82), (b) Tonal Audibility and Sharpness (Model 2, R^2=.88), (c) Weighted-sum Tonal Audibility (Model 3, R^2=.88). The noise stimuli are labelled with assigned numbers from Table 7.4.

7.4 Summary

Noise stimuli with five-tone complexes between 125 Hz to 2 kHz were artificially generated for subjective testing to obtain the perceptual weighting function of complex tones. The levels of each tone were randomly adjusted for every trial, and both harmonic and inharmonic structured tone complexes were utilized. Ten musically-trained subjects participated in the subjective test involving paired comparisons. Each participant was asked to choose which noise stimulus was more annoying between two noise signals. Perceptual weighting analysis results were applied as a spectral weighting function to calculate a proposed Weighted-sum Tonal Audibility metric. The performance of the newly developed metric showed better annoyance prediction than that from the traditional Tonal Audibility metric. The revised annoyance regression model with the two noise metrics of ANSI Loudness Level and the Weighted-sum Tonal Audibility showed similar prediction performance to the regression model with ANSI Loudness Level, Tonal Audibility and Sharpness from Chapter 6.

8. Chapter Eight

Conclusion

8.1 Summary

This research aimed to investigate effects of tonal background noises on human annoyance perception and task performance in the built environment. The dissertation addressed three complementary objectives: 1) to examine the relationship between associated tonal noise metrics and annoyance perception, 2) to determine upper limits of acceptability for tonality, and 3) to identify effect of tones on human task performance. Four phases of subjective testing were conducted in an indoor acoustic testing chamber at the University of Nebraska to achieve the research objectives.

In the first study, subjects were asked to complete Sudoku puzzles while exposed to broadband noise with a tonal component set at a specific level above the noise. Participants then filled out a subjective rating questionnaire on the noise they had just experienced. Five levels of two tonal frequencies (125 Hz and 500 Hz) were tested above two different background noise levels for a total of 20 test signals. Results were used to develop an annoyance prediction model of tonal noise.

A factorial repeated measure ANOVA (Analysis of Variance) revealed that participants felt more annoyed with increasing background noise level, lower tone frequency and stronger prominence of the tone strength. Correlation analysis with noise parameters and subjective perception ratings suggested that the ANSI Loudness level among all other loudness metrics correlated most strongly with annoyance perception while assorted tonality metrics showed relatively weaker but still statistically significant correlations with annoyance. A multivariate regression model with ANSI Loudness Level and Tonal Audibility was subsequently developed.

An increased number of 40 tonal signals was generated for the second subjective testing study. Five levels of tones at four specific frequencies of 125, 250, 500, 1 kHz were added separately to broadband background noise signals. During each session, participants performed digit span tasks in which they memorized a series of numbers in the reverse order of presentation while exposed to assorted tonal signals. After each trial, the participants completed a subjective questionnaire with two items: how annoyed they were by the noise, and whether or not they would complain about the noise.

Results were analyzed to determine a threshold of acceptability for tonality. First, a dose-response model was formulated to predict the percentage of persons lodging complaints when both tonality and loudness are considered; a multivariate logistic regression model then indicates what the human annoyance thresholds are of tones in noise and reflects that thresholds vary, depending on the absolute level of the ambient background noise. Suggested threshold values of Tonal Audibility have been presented for given background noise levels. The results show that maximum allowable tonal components decrease when background noise level is high.

These repeated-measure subjective tests with mentally demanding tasks showed effects of the tones on human performance. Factorial repeated-measured ANOVA of test results have demonstrated that tonality has a crucial influence on completion time of subjects whereas loudness levels alone did not.

The third investigation aimed to improve the annoyance regression model by exploring multidimensional aspects of annoyance perception using actual building mechanical noise signals with tones and perceptual weighting of complex tones. Fifteen actual audio recordings from building mechanical equipment and three artificially synthesized signals were used in the experiment to investigate psycho-acoustical attributes, the presence of harmonics, and time fluctuation characteristics of the tones.

During the experiment, participants completed a series of paired comparison tasks about how two sound stimuli presented in a pair were similar and which one they perceived to be more annoying than the other. The dominant acoustic characteristics for annoyance perception were determined by multidimensional scaling analysis (MDS).

A non-metric, individual scaling difference (INDSCAL) algorithm was used to derive the MDS similarity solution. The goodness-of-fit of the derived solution indicated that four perceptual dimensions were appropriate for describing the solution. The results showed that the latent psychological structures for the similarity task were related to the sharpness, tonality and loudness of the noise stimulus. A revised multidimensional annoyance model, incorporating sharpness of noise in addition to loudness and tonality, was subsequently proposed based on these test results to improve the prediction accuracy of the annoyance model.

To improve the predictions by the comprehensive complex tones in the noise, a newly revised Weighted-sum Tonal Audibility was proposed against the traditional Tonal Audibility metrics. Perceptual weighting analysis was carried out with the harmonic and inharmonic tone-complexes to develop the Weighted-sum Tonal Audibility. The revised version showed better predictions of annoyance in the multivariate regression model. The annoyance prediction model with ANSI Loudness Level and Weighted-sum Tonal Audibility showed the comparable performance to the regression model with ANSI

Loudness Level, Tonal Audibility and Sharpness. Using Weighted-sum Tonal Audibility is recommend because one does not need to calculated additional Sharpness metric.

The research clearly indicates that tonality should be included in understanding annoyance responses from building mechanical noise. It has been found that the ANSI Loudness Level, Tonal Audibility, Sharpness contribute as significant predictors related to annoyance perception. A dose-response relationship was also developed to determine the upper limits of tonality in noises. The upper limit levels of tonality according to determined background noise level have been suggested. Lastly, the results showed that even just-audible tones can significantly increase the reaction time of participants to complete a cognitively demanding task.

8.2 Future Research

One limitation of this study is that all findings are from laboratory experiments. Even though the subjective testing in the laboratory has assorted advantages to test research hypotheses, the findings should be validated with in-situ measurements. This research also utilized a limited number of participants and noise signals. Thus, more data are needed to verify the suggested annoyance model.

Continuing research should investigate effects of time-fluctuating characteristics of tones on annoyance. The noise signals used in this test did not exhibit a wide range of fluctuation properties of tones in noises and failed to find any statistically significant effect of tone fluctuation characteristics. The weighting function of multi-tone complexes should also be expanded to wider frequency ranges with various scenarios to develop a more accurate ways to calculate the tonality metrics. The presented weighting functions in this dissertation were not practically applicable yet because they were only tested in limited multi-tone complexes with specific sound levels and frequencies.

Reference

AHRI. “AHRI/ANSI 1140-2012: Sound Quality Evaluation Procedures for Air-Conditioning and Refrigeration Equipment.” (2012).

ANSI. “ANSI S3.4-2007 Procedure for the Computation of Loudness of Steady Sounds.” 2007.

ANSI/ASA. “ANSI/ASA S12.10-2010/Part 1 Acoustics – Measurement of Airborne Noise Emitted by Information Technology and Telecommunications Equipment – Part 1 : Determination of Sound Power Level and Emission Sound Pressure Level.” 2010.

Aures, W. “The Sensory Euphony as a Function of Auditory Sensations.” *Acustica* 59 (1985): 130–414.

Bailey, BP, and JA Konstan. “On the Need for Attention-Aware Systems: Measuring Effects of Interruption on Task Performance, Error Rate, and Affective State.” *Computers in Human Behavior* 22.4 (2006): 685–708.

Berg, Bruce G., and D M Green. “Spectral Weights in Profile Listening.” *The Journal of the Acoustical Society of America* 88.2 (1990): 758–66.

Berglund, B., P. Hassmén, and A. Preis. “Annoyance and Spectral Contrast Are Cues for Similarity and Preference of Sounds.” *Journal of Sound and vibration* 250.1 (2002): 53–64.

Bisping, R. “Car Interior Sound Quality: Experimental Analysis by Synthesis.” *Acta Acustica united with Acustica* 83.5 (1997): 813–818.

Blazier, WE. “Revised Noise Criteria for Application in the Acoustical Design and Rating of HVAC Systems.” *Noise Control Engineering Journal* 16.2 (1981): 64–73.

Borg, Ingwer, Patrick JF Groenen, and Patrick Mair. *Applied Multidimensional Scaling*. Springer Science & Business Media, 2012.

Bradley, David. “Analysis of Parameter Effects on Sound Energy Decay in Coupled Volume Systems.” Doctoral dissertation, University of Nebraska - Lincoln, 2006.

Brambilla, G, and F Pedrielli. “Community Reactions to Noise and Vibrations from Railway Traffic.” *Proc. Inter-Noise, Liverpool, England* (1996).

Brocolini, L et al. “Prediction and Explanation of Sound Quality Indicators by Multiple Linear Regressions and Artificial Neural Networks.” *Acoustics 2012 Nantes*. Nantes, France: (2012): 2127–2132.

Choe, B. “Nonmetric Multidimensional Scaling of Complex Sounds.” Doctoral dissertation, Universität Oldenburg, 2001.

Cox, David Roxbee, and E. Joyce Snell. *Analysis of binary data*. Vol. 32. CRC Press, 1989.

David W. Hosmer, Jr., and Stanley Lemeshow. *Applied Logistic Regression*. John Wiley & Sons, 2004.

Dittrich, Kerstin, and Daniel Oberfeld. "A Comparison of the Temporal Weighting of Annoyance and Loudness." *The Journal of the Acoustical Society of America* 126.6 (2009): 3168–78.

Ebissou, Ange, Patrick Chevret, and Etienne Parizet. "Work Performance and Mental Workload in Multiple Talker Environments." *Proceedings of Meetings on Acoustics*. Vol. 19. Acoustical Society of America, 2013.

Esser, Timo. "Test Tone Generator 4.4." 2014.

FAA. "Part 36. Noise Standards: Aircraft Type and Airworthiness Certification." (1969).

Fastl, H, and E Zwicker. *Psychoacoustics: Facts and Models*., 2001.

Fidell, Sanford, David S. Barber, and Theodore J. Schultz. "Updating a Dosage–effect Relationship for the Prevalence of Annoyance due to General Transportation Noise." *The Journal of the Acoustical Society of America* 89.1 (1991): 221–233.

Field, A. *Discovering Statistics Using IBM SPSS Statistics*. Sage, 2013.

Francis, Jennifer M. "Determining Thresholds of Annoyance to Tones in Noise." University of Nebraska - Lincoln, 2014.

Frank, T, T. A. Brungart, and N. C. Nicholas. "Annoyance from and Tolerance to Sounds." *Noise Control Engineering Journal* 55.3 (2007): 357–366.

Green, PE, FJ Carmone, and SM Smith. "Multidimensional Scaling: Concepts and Applications." (1989).

Grjmaldi, John V. "Sensori-Motor Performance under Varying Noise Conditions." *Ergonomics* 2.1 (1958): 34–43.

Guski, R, U Felscher-Suhr, and R Schuemer. "The Concept of Noise Annoyance: How International Experts See It." *Journal of sound and vibration* 223 (1999): 513–527.

Haka, M et al. "Performance Effects and Subjective Disturbance of Speech in Acoustically Different Office Types-a Laboratory Experiment." *Indoor air* 19.6 (2009): 454–67.

Hart, S. G. "Nasa-Task Load Index (NASA-TLX); 20 Years Later." *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 50.9 (2006): 904–908.

Hasting, Aaron, et al. "Measurement of the attributes of complex tonal components commonly found in product sound." *Noise Control Engineering Journal* 51.4 (2003): 195-209.

Hellman, R P. “Growth Rate of Loudness, Annoyance, and Noisiness as a Function of Tone Location within the Noise Spectrum.” *The Journal of the Acoustical Society of America* 75.1 (1984): 209–18.

Hellman, R P. “Perceived Magnitude of Two-Tone-Noise Complexes: Loudness, Annoyance, and Noisiness.” *The Journal of the Acoustical Society of America* 77.4 (1985): 1497–504.

Hellweg, R et al. “A Comparison of Two Methods for the Evaluation of Prominent Discrete Tone-Recommendations.” *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*. Institute of Noise Control Engineering., 2001. 2286–2291.

Hellweg, R, M Nobile, and A. C Balant. “A COMPARISON OF TWO METHODS FOR THE EVALUATION OF PROMINENT DISCRETE TONES - PHASE 2.” *InterNoise 2000*. Institute of Noise Control Engineering, 2000. P

Holmberg, K, U Landstrom, and A Kjellberg. “Effects of Ventilation Noise due to Frequency Characteristic and Sound Level.” *Journal of low frequency noise & vibration* 12.4 (1993): 115–122.

ISO. “ISO 1996-2:2007 Acoustics -- Description, Measurement and Assessment of Environmental Noise -- Part 2: Determination of Environmental Noise Levels.” 2007.

ISO. “ISO 532:1975 Acoustics -- Method for Calculating Loudness Level.” 1975.

ISO. “ISO/TS 15666:2003 Acoustics - Assessment of Noise Annoyance by Means of Social and Socio-Acoustic Surveys.” 2003.

Jahanshahi, Marjan et al. “A Preliminary Investigation of the Running Digit Span as a Test of Working Memory.” *Behavioural neurology* 20.1-2 (2008): 17–25.

Janssen, S. A et al. “A Comparison between Exposure-Response Relationships for Wind Turbine Annoyance and Annoyance due to Other Noise Sources.” *The Journal of the Acoustical Society of America* 130.6 (2011): 3746–3753.

Jesteadt, Walt et al. “Perceptual Weights for Loudness Judgments of Six-Tone Complexes.” *The Journal of the Acoustical Society of America* 136.2 (2014): 728–735.

Job, R. F. S. et al. “General Scales of Community Reaction to Noise (dissatisfaction and Perceived Affectedness) Are More Reliable than Scales of Annoyance.” *The Journal of the Acoustical Society of America* 110.2 (2001): 939-946.

Job, R.F. Soames. “Noise Sensitivity as a Factor Influencing Human Reaction to Noise.” *Noise & health* 1.3 (1999): 57–68.

Job, RFS, A Topple, and NL Carter. “Test-Retest Reliability of Psychological Scales (sensitivity and Attitude) Related to Community Reaction to Noise.” *International Congress on Noise Control Engineering*. 1996. 2413–2418.

John, J. A., F. W. Wolock, and H. A. David. *Cyclic Designs*. Washington, DC: National Bureau of Standards Applied Mathematics Series 62, 1972.

Joseph F. Hair Jr et al. *Multivariate Data Analysis*. 7th ed. Prentice Hall, 2009.

Kerry, G et al. “The Assessment of Noise from Low Flying Military Jet Aircraft in the Community.” *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*. Institute of Noise Control Engineering, 1998. 240–243.

Khan, M. S, O Johansson, and U Sundback. “Evaluation of Annoyance Response to Engine Sounds Using Different Rating Methods.” *International Congress on Noise Control Engineering*. 1996. 2517–2520.

Kim, R. “Burden of Disease from Environmental Noise.” *WHO International Workshop on “Combined Environmental Exposure: Noise, Air Pollutants and Chemicals.”* Ispra: 2007. Print.

Kroesen, Maarten et al. “Measuring Subjective Response to Aircraft Noise: The Effects of Survey Context.” *The Journal of the Acoustical Society of America* 133.1 (2013): 238–46.

Kryter, K. D. “The Meaning and Measurement of Perceived Noise Level.” *Noise Control* 6.5 (1960): 12–27.

Kryter, K D, and K S Pearsons. “Judged Noisiness of a Band of Random Noise Containing an Audible Pure Tone.” *The Journal of the Acoustical Society of America* 38 (1965): 106–12.

Kryter, K. D. “Community Annoyance from Aircraft and Ground Vehicle Noise.” *The Journal of the Acoustical Society of America* 72.4 (1982): 1222–1242.

Kryter, K. D. “Acoustical, Sensory, and Psychological Research Data and Procedures for Their Use in Predicting Effects of Environmental Noises.” *The Journal of the Acoustical Society of America* 122.5 (2007): 2601–2614.

Laird, D. “The Influence of Noise on Production and Fatigue, as Related to Pitch, Sensation Level, and Steadiness of the Noise.” *Journal of Applied Psychology* 17.3 (1933): 320 – 330.

Landström, Ulf et al. "Exposure Levels, Tonal Components, and Noise Annoyance in Working Environments." *Environment International* 21.3 (1995): 265–275.

Laux, P. C, P Davies, and G. R Long. "The Correlation of Subjective Response Data with Measured Noise Indices of Low Frequency Modulated Noise." *Noise Control Engineering Journal* 40.3 (1993): 241–253.

Lee, Kyoung Hoon, P Davies, and A. M Surprenant. "Quantification of the Tonal Prominence of Complex Tones in Machinery Noise." *Noise-Con 04. The 2004 National Conference on Noise Control Engineering.* 2004.

Lee, Kyoung Hoon, Patricia Davies, and Aimee M. Surprenant. "Tonal Strength of Harmonic Complex Tones in Machinery Noise." *The Journal of the Acoustical Society of America* 118.3 (2005): 1921–1921.

Lee, Pyoung Jik, and Jin Yong Jeon. "Relating Traffic, Construction, and Ventilation Noise to Cognitive Performances and Subjective Perceptions." *The Journal of the Acoustical Society of America* 134.4 (2013): 2765–72.

Leibold, Lori J. et al. "Contributions of Individual Components to the Overall Loudness of a Multitone Complex." *The Journal of the Acoustical Society of America* 121.5 (2007): 2822.

Little, J. W. "Human Response to Jet Engine Noises." *Noise Control* 32.11 (1961): 1516–1516.

Los Angeles County, CA. *Noise Control Ordinance of the County of Los Angeles*. 1978. Print.

Lutfi, Robert a., and Walt Jesteadt. "Molecular Analysis of the Effect of Relative Tone Level on Multitone Pattern Discrimination." *The Journal of the Acoustical Society of America* 120.6 (2006): 3853.

Marquis-Favre, C., E. Premat, and D. Aubrée. "Noise and Its Effects – a Review on Qualitative Aspects of Sound. Part II: Noise and Annoyance." *Acta Acustica united with Acustica* 91.4 (2005): 626–642.

MathWorks. "MATLAB 2013." 2013.

Miedema, H. M, and Henk Vos. "Exposure-Response Relationships for Transportation Noise." *The Journal of the Acoustical Society of America* 104.6 (1998): 3432–3445.

Miedema, Henk M. E., and Henk Vos. "Exposure-Response Relationships for Transportation Noise." *The Journal of the Acoustical Society of America* 104.6 (1998): 3432–3445.

Minnesota. *Ordinance of Minnesota Chapter 389. - Noise*. 2008.

Mølhave, L., B. Bach, and O.F. Pedersen. "Human Reactions to Low Concentrations of Volatile Organic Compounds." *Environment International* 12.1-4 (1986): 167–175.

More, Shashikant, and Patricia Davies. "Human Responses to the Tonalness of Aircraft Noise." *Noise Control Engineering Journal* 58.4 (2010): 420–440.

Nagelkerke, Nico JD. "A note on a general definition of the coefficient of determination." *Biometrika* 78.3 (1991): 691-692.

New York. *N.Y. ADC. LAW § 27-770 : NY Code - Section 27-770: Noise Control of Mechanical Equipment*. 2006.

Oberfeld, Daniel et al. "Spectro-Temporal Weighting of Loudness." *PloS one* 7.11 (2012): e50184.

Parizet, Etienne, Nacer Hamzaoui, and Guillaume Sabatie. "Comparison of some listening test methods: a case study." *Acta Acustica united with Acustica* 91.2 (2005): 356-364.

Parizet, Etienne. "Paired Comparison Listening Tests and Circular Error Rates." *Acta Acustica united with Acustica* 88.4 (2002): 5.

Pedersen, Eja et al. "Response to Noise from Modern Wind Farms in The Netherlands." *The Journal of the Acoustical Society of America* 126.2 (2009): 634–43.

Pedersen, Torben Holm. "The "Genlyd" Noise Annoyance Model: Dose-Response Relationships Modelled by Logistic Functions." *DELTA Acoustics & Electronics:* Hørsholm, Denmark (2007).

Pedersen, Torben Holm, Morten Søndergaard, and Bent Andersen. "Objective method for assessing the audibility of tones in Noise Joint Nordic Method–Version 2." *DELTA 2000* (2000): 04-14.

Persson, K, M Bjorkman, and R Rylander. "An Experimental Evaluation of Annoyance due to Low Frequency Noise." *Journal of low frequency noise & vibration* 4.4 (1985): 145–153.

Persson, K., and M. Björkman. "Annoyance due to Low Frequency Noise and the Use of the dB(A) Scale." *Journal of Sound and Vibration* 127.3 (1988): 491–497.

Ryherd, Erica E, and Lily M Wang. "Implications of Human Performance and Perception under Tonal Noise Conditions on Indoor Noise Criteria." *The Journal of the Acoustical Society of America* 124.1 (2008): 218–26.

Ryherd, Erica E, and Lily M Wang. "AB-10-018: The Effects of Noise from Building Mechanical Systems with Tonal Components on Human Performance and Perception

(1322-RP).” *ASHRAE Transactions* (2010).

Saeki, Tetsuro et al. “Effects of Acoustical Noise on Annoyance, Performance and Fatigue during Mental Memory Task.” *Applied Acoustics* 65.9 (2004): 913–921.

Sailer, U, and M Hassenzahl. “Assessing Noise Annoyance: An Improvement-Oriented Approach.” *Ergonomics* 43.11 (2000): 1920–38.

Schomer, Paul D. “A White Paper : ASSESSMENT OF NOISE ANNOYANCE.” Schomer and Associates, 2001.

Schomer, Paul D. “Criteria for Assessment of Noise Annoyance.” *Noise Control Engineering Journal* 53.4 (2005): 125-137.

Schultz, Theodore J. “Synthesis of Social Surveys on Noise Annoyance.” *The Journal of the Acoustical Society of America* 64.2 (1978): 377-405.

Schutte, Martin et al. “The Development of the Noise Sensitivity Questionnaire.” *Noise and Health* 9.34 (2007): 15.

Seattle. *Seattle Municipal Code Chapter 25.08 Noise Control*. 2007.

Spence, Ian, and Dennis W. Domoney. “Single Subject Incomplete Designs for Nonmetric Multidimensional Scaling.” *Psychometrika* 39.4 (1974): 469–490.

Starck, J, E Toppila, and I Pyykkö. “Impulse Noise and Risk Criteria.” *Noise & health* 5.20 (2003): 63–73.

Susini, Patrick et al. “Characterizing the Sound Quality of Air-Conditioning Noise.” *Applied Acoustics* 65.8 (2004): 763–790.

Tang, S.K. “A DISTRIBUTION FUNCTION APPLICABLE TO OFFICE NOISE STUDY.” *Journal of Sound and Vibration* 208.4 (1997): 603–615.

Terhardt, Ernst, Gerhard Stoll, and Manfred Seewann. “Algorithm for Extraction of Pitch and Pitch Salience from Complex Tonal Signals.” *The Journal of the Acoustical Society of America* 71.3 (1982): 679–688.

Töpken, Stephan, Jesko Verhey, and Reinhard Weber. “Perceptual Space of Multiple Complex Tones.” *Proceedings of the 20th International Congress on Acoustics*. ICA, 2010.

Trollé, A., C. Marquis-Favre, and A. Klein. “Short-Term Annoyance Due to Tramway Noise: Determination of an Acoustical Indicator of Annoyance Via Multilevel Regression Analysis.” *Acta Acustica united with Acustica* 100.1 (2014): 34–45.

von Hunerbein, S, A King, and J Hargreaves. *Perception of Noise from Large Wind*

Turbines. Greater Manchester, UK. 2010.

Weinstein, N. D. “Individual Differences in Critical Tendencies and Noise Annoyance.” *Journal of Sound and Vibration* 68.2 (1980): 241–248.

Wickelmaier, Florian. "An introduction to MDS." *Sound Quality Research Unit, Aalborg University*, Denmark (2003): 46.

Woodcock, J., A. T. Moorhouse, and D. C. Waddington. “A Multidimensional Evaluation of the Perception and Annoyance Caused by Railway Induced Groundborne Vibration.” *Acta Acustica united with Acustica* 100.4 (2014): 614–627.

Yanagisawa, Hideyoshi, Shunsuke Yokoo, and Tamotsu Murakami. “Effect of Tonal Harmonic Feature in Product Noise on Emotional Quality.” *Improving Complex Systems Today*. London: Springer, 2011. 469–476.