SOUND QUALITY EVALUATION OF HVAC&R EQUIPMENT

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Weonchan Sung

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

August 2020

Purdue University

West Lafayette, Indiana

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF DISSERTATION APPROVAL

Dr. Patricia Davies, Chair

School of Mechanical Engineering

Dr. J. Stuart Bolton

School of Mechanical Engineering

Dr. George T. Chiu

School of Mechanical Engineering

Dr. Kai M. Li

School of Mechanical Engineering

Dr. Hong Z. Tan

School of Electrical and Computer Engineering

Approved by:

Dr. Nicole L. Key

Head of the School Graduate Program

TABLE OF CONTENTS

LIST OF TABLES

Table Page

Table Page

LIST OF FIGURES

Figure Page

Figure Page

Figure Page

ACKNOWLEDGMENTS

Thanks to Professor Patricia Davies and Professor J. Stuart Bolton for their guidance and support. Without their insightful advice, it would be impossible to come this far. I'm grateful to committee members Professor George Chiu, Professor Hong Z. Tan, and Professor Kai Ming Li for their advice. Thanks to my industrial sponsors Asad Sardar, Trent Marshall and Robert Chopko who have helped me and provided me with valuable advice. Thanks to the friends who helped me make precious memories at Purdue University. Thanks to my parents and parents-in-law for trusting me with their endless support. Thanks to Ihan and Jihan for being healthy and kind. Thank my lovely wife, Jieun Han, who always trusts and encourages me with love and patience.

ABSTRACT

Sung, Weonchan PhD, Purdue University, August 2020. Sound Quality Evaluation of HVAC&R Equipment. Major Professor: Patricia Davies, School of Mechanical Engineering.

Characteristics of heating, ventilation, air conditioning and refrigeration (HVAC&R) equipment sounds and people's responses to them were studied in order to develop models to predict annoyance from recordings of the sound. These models are intended to address shortcomings of currently used methods for HVAC&R product sound assessment. Coupled with sound prediction models, the annoyance models will be used to monitor and guide improvements to HVAC&R equipment sound quality throughout the product design process: from virtual early design, through to the prototyping and product refinement stages. Responses to residential and refrigerated truck product noise was studied; both produce broadband random noise and families of harmonics related to rotating and reciprocating components within the system. Tests were conducted to determine how people describe HVAC&R equipment sounds; how their descriptions relate to sound characteristics and overall assessments; and to develop models that relate predicted strengths of sound characteristics to the overall assessment. Annoyance models were developed for each types of product. Loudness and spectral balance metrics are included in models for both types of product. Inclusion of a tonalness metric improved models for residential units, and roughness and impulsiveness metrics improved models for refrigerated truck units. The models developed were used to predict responses in the other tests and there was good agreement between predicted and measured responses. An illustration of the use of the annoyance models, in conjunction with sound visualization and signal modification techniques, to guide improvements to product sound quality is given.

1. INTRODUCTION

In the United States, most houses have residential HVAC&R units with components consisting of fans, motors and compressors that operate for a long time. The sounds from these units can propagate to nearby houses. This has led to reported cases of people either outside or in the house complaining about the noise [1]. Studies have also shown that that residential HVAC&R noise exposure can cause sleep deprivation [2] or decreased work productivity [3]. Refrigerated trucks transport products that can be spoiled when they are not kept under a certain temperature. These trucks can generate noise that affects people in nearby dwellings while the vehicle is moving or stationary and parked for the night [4]. Reducing refrigeration noise in these trucks can also reduce driver fatigue, help the driver identify dangerous situations and reduce pedestrian and bike-related accidents [5]. Metrics such as A-weighted sound pressure level and predictions from loudness models, such as Zwicker's and Moore and Glasberg's loudness model [6, 7], have been used to quantify noise effects, but there has been dissatisfaction with their performance because some sounds which appear to be in a good sound level range can still be complained about due to other sound characteristics which the sound level does not capture. It is possible to do better at predicting which sounds will be perceived as being more acceptable, less annoying or more pleasant. If the development of an improved sound perception model is achieved, it can be used at all stages of machine design, prototyping, and in any subsequent noise trouble shooting, to help engineers optimize the sounds of their products.

HVAC&R equipment is composed of various mechanical components which contribute to its sound quality. Both residential and refrigerated truck HVAC&R units have fans and compressors and refrigerated trucks also generate diesel engine noise. Example spectra of refrigerated truck and residential sounds are shown in Figure 1.1 and Figure 1.2, respectively. As seen in Figure 1.1 and Figure 1.2, the sounds of

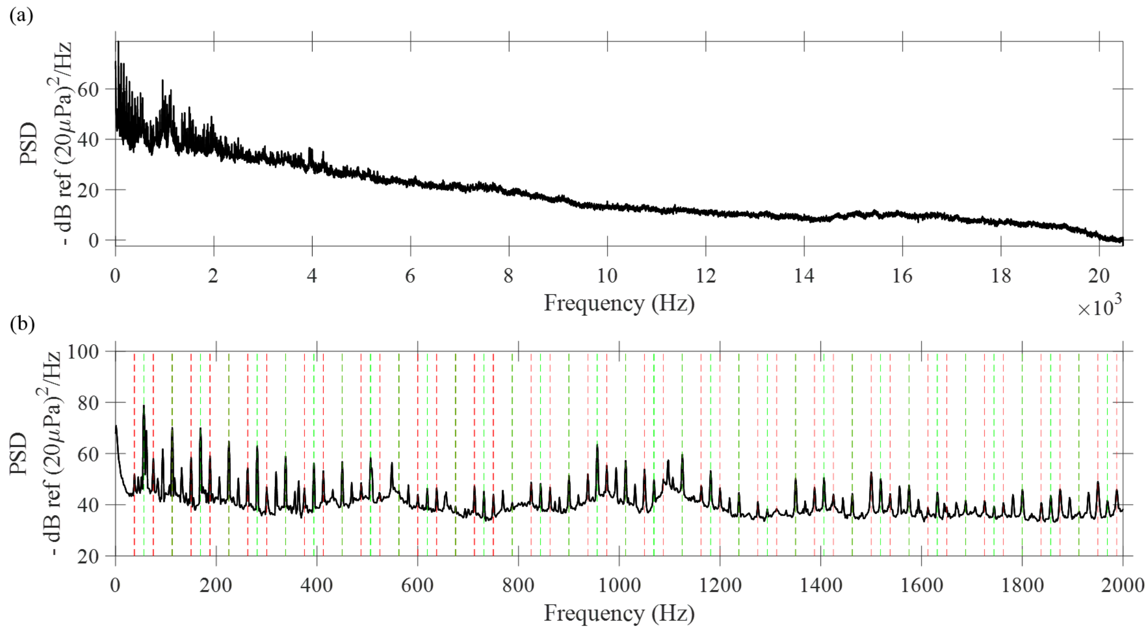


Figure 1.1. An example of the power spectral density of noise from a refrigerated truck HVAC&R unit. Power spectral density over: (a) a broad frequency range (0-20,480 Hz) and (b) a narrower frequency range (0-2000 Hz). The green and red lines indicate the locations of the compressor and diesel engine harmonics, respectively. During the recording, the compressor fundamental frequency was 56.3 Hz, and diesel engine fundamental frequency was 37.5 Hz.

residential and refrigerated truck HVAC&R units contain many tonal components and a broadband component. The tonal components mainly are due to the motor rotation rate, the blade passage frequency (BPF), and the compressor rotation rate. Both the fundamental frequencies and harmonics are present. Sounds with high amplitude tonal components or high levels of broadband components may be loud, but the former increases the tonality of the sound and the latter normally decreases it. Sometimes, two adjacent tones within a critical band create a beating phenomenon which can be tracked (tone separation $\leq$ 16 Hz) or is too fast to be tracked (tone separation $>$ 16 Hz). These effects produce, respectively, fluctuation and roughness characteristics to the overall sound. In Figure 1.3 are shown examples of loudness

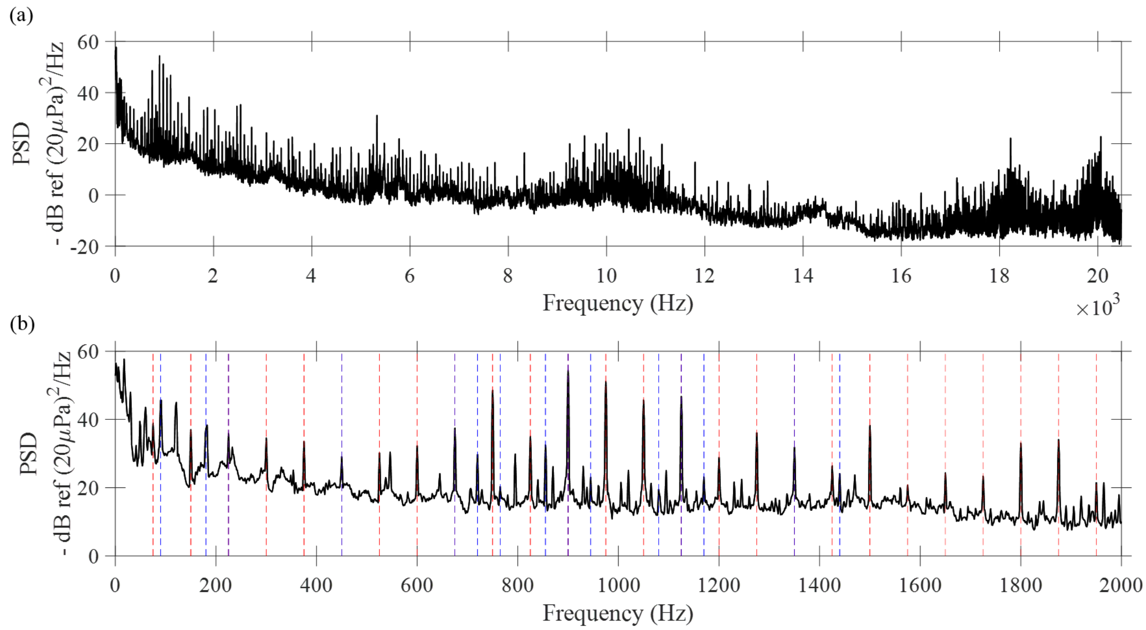


Figure 1.2. An example power spectral density of noise from a residential HVAC&R unit. Power spectral density on over: broad frequency range (0-20,480 Hz) and (b) a narrower frequency range (0-2000 Hz). The red and blue dotted lines indicate the location of the compressor and fan harmonics, respectively. During the recording, the compressor fundamental frequency was 75 Hz, and the 3 bladed fan rotating at 900 rpm (15 Hz) giving a blade passage frequency of 45 Hz.

time histories with fast and slow fluctuating sounds. Zwicker's model of time varying loudness [6] was used to produce these time histories. Hereafter it will be just referred to as Zwicker Loudness. Short transient sounds can be perceived as impulsive or having a "pounding" character when they are repeated consecutively at regular time intervals. A sound can be musical or unmusical depending on the ratio of fundamental frequencies. In the present research, we found that even a small frequency variation of a tonal component at high frequencies can affect annoyance ratings. For example, sounds with spectra represented by the red and black lines in Figure 1.4 have similar loudness levels but the annoyance ratings from subjects were significantly different. The signal with the red spectrum was found to be significantly more an-

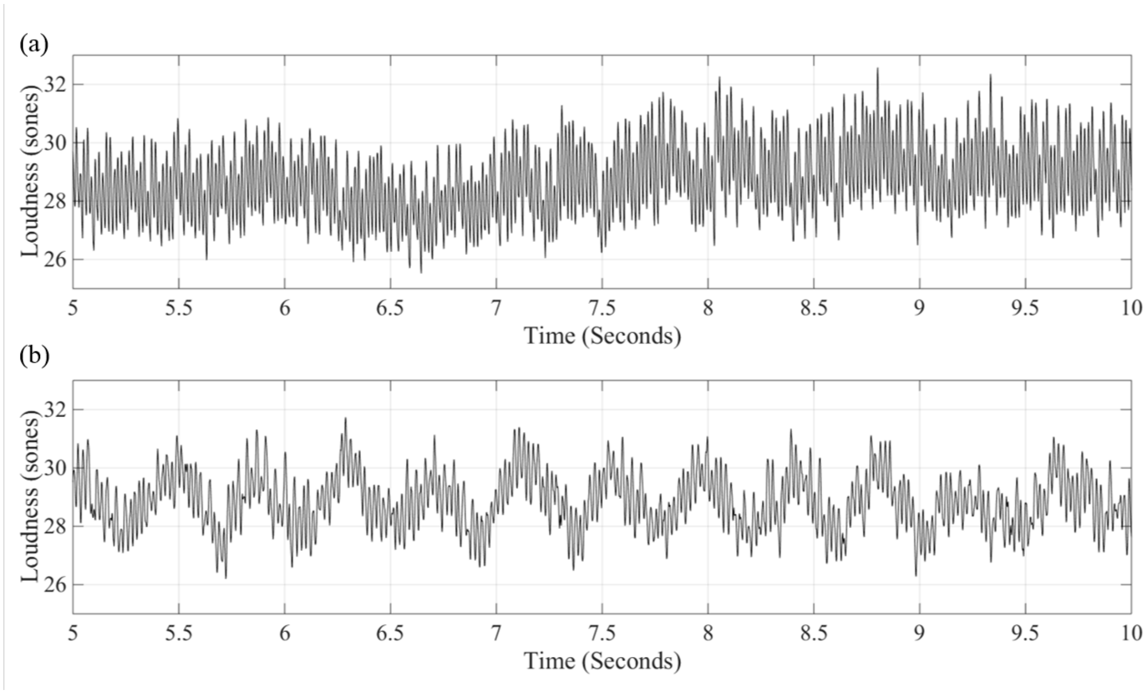


Figure 1.3. An example of Zwicker Loudness time histories for sounds with (a) mostly fast fluctuations and (b) both slow (around 2 Hz) and fast (> 40 Hz) fluctuations.

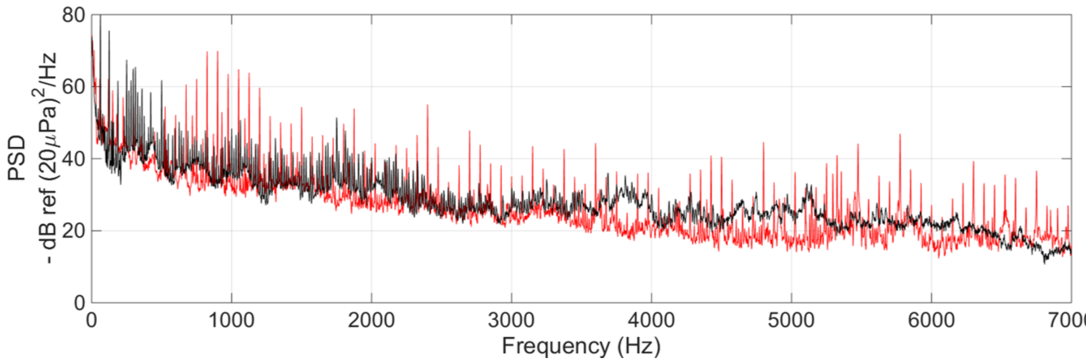


Figure 1.4. Power spectral densities of sounds with similar loudness levels but different annoyance ratings. Red - visible tonal components at both low and high frequencies, black - visible tonal components at low frequencies only.

noying that the sound with the black spectrum. The broadband components arise from air motion, turbulence, and fluid pulsation over a broad range of frequencies [8]. Broadband components in HVAC&R equipment sounds can make sounds loud, rough and sometimes increase fluctuation (Figure 1.3). Depending on the energy distribution across the spectrum, the sound can be sharp. The sound of HVAC&R equipment may have various sound characteristics and these characteristics affect perception of these sounds. Researchers have developed models to predict the perceived strength of sound characteristics; these are often referred to as sound quality metrics or sound quality models. Table 1.1 is an overview of the most well known sound quality models.

People's overall judgements of sounds, for example annoyance and preference, depend on many factors including sound characteristics. Other factors include a person's noise sensitivity [20], experience of similar sounds and context, the relationship between the noise generating system and individuals. For example, an HVAC system controls temperature in an environment, while it is providing much needed cooling on individual may have more tolerance for the noise. Aircraft noise is more annoying after accidents at airports [21]. If noise affects cognitive ability, then it interferes with productivity and could be therefore more annoying. Perceptions of loudness and other sound attributes vary from person to person, i.e., there is a distribution of responses, and metrics are often used to predict the average perceptions of the attribute strength. Similarly with overall judgments such as an annoyance or acceptability rating but, because of the large number of factors affecting overall judgements of sounds, the distributions associated with, e.g., annoyance, tend to be broader, than those for individual sound characteristics. In product noise assessment, typically trends associated with the average of annoyance ratings of a population of test subjects are modeled. While annoyance in a natural setting may be different to the annoyance ratings given in more sterile laboratory settings, it is hypothesized that the general trends identified from the tests in the laboratory are likely to play a role

Table 1.1. Summary of sound quality models and annoyance models that adjust level metrics to account for increased annoyance due to tonalness of sounds.

Sound Quality Metric	**Sound Characteristics**	**Reference**
Stevens Loudness (N_S)	Level → loudness	[9]
Zwicker Loudness (N_Z)		[6]
Moore and Glasberg Loudness (N_{MG})		[7]
A/C weighted Sound Pressure Level (dBA/dBC)		[10]
Tone-to-Noise Ratio (TNR)	Tonalness	[11]
Prominence Ratio (PR)		[11]
Aures' Tonality (T_{A5})		[12]
Tone Audibility (TA)		[13]
von Bismark Sharpness (S_{VB})	Spectral Balance	[14]
Aures' Sharpness (S_A)		[12]
Fluctuation Strength (FS)	Fluctuation	[15]
Roughness (R)		[15]
Rate of Change of the Loudness (RCL)	Impulsiveness	[16]
Kurtosis (K)		[17]
Annoyance Models	**Based on**	
Tone Corrected Perceived Noise Level ($PNLT$)	Sound Pressure & Tonal Prominence assessed from third-octave bands	[18]
Effective Perceived Noise Level ($EPNL$)	$PNLT$ & duration of aircraft flyover	[18]
Sound Quality Indicator (SQI, SQI^*)	Zwicker Loudness & Tonal Prominence assessed from third-octave bands	[19]
Psychoacoustic Annoyance	Loudness, Sharpness, Roughness, & Fluctuation Strength	[15]
Pleasantness	Aures' Tonality & Sharpness	[12]

in the annoyance experienced in the more natural setting. Annoyance predictions are made using models that are functions of sound attribute strengths.

In Table 1.1 is a list of annoyance models that combine a level metric with an assessment of how tonalness increases annoyance beyond that predicted by level alone. Because of non-acoustic factors including the context in which a sound is heard, annoyance models are developed for particular products, or groups of products. Fine-tuning a sound quality or annoyance model for a particular market sector may be more useful in product design. However, it is an attractive option to have models that could be used in design of a broader group of products. One question in this research is whether one model can be used in assessment of both refrigerated truck noise and residential HVAC&R equipment noise, or are two separate models better? The process for developing the annoyance models in this research is shown in Figure 1.5.

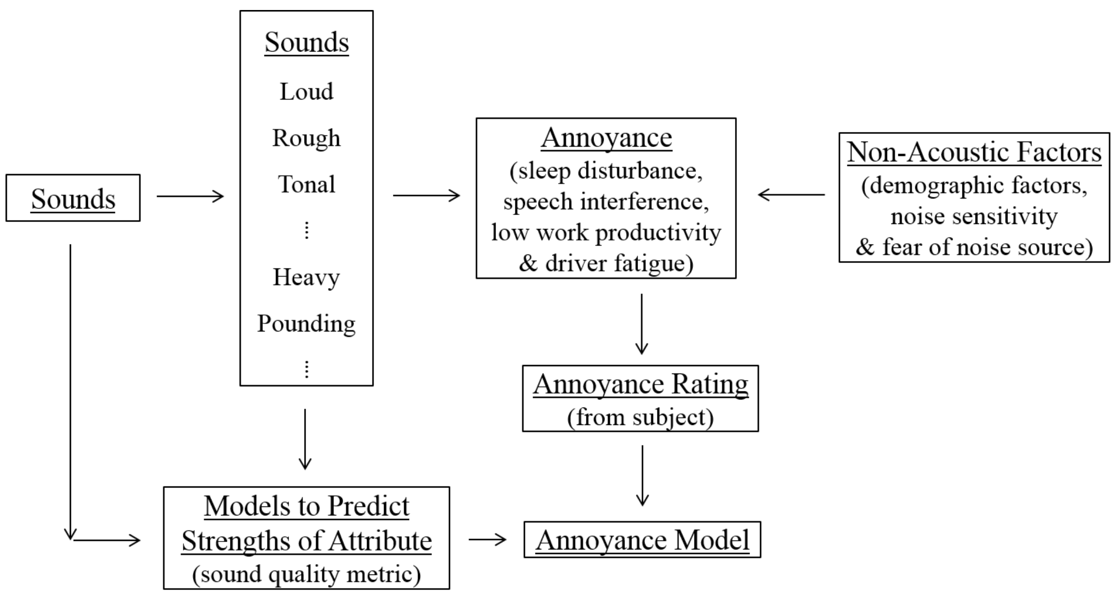


Figure 1.5. Process illustrating the annoyance model development.

In the following sections, research published in the literature on sound quality of HVAC&R equipment is reviewed and the chapter ends with a list of objectives for this research and a description of the organization of the remainder of this thesis.

1.1 Previous Research on Sound Quality Studies of HVAC&R Equipment

There have been many studies on sound quality evaluation of HVAC&R equipment noise. It has been found that level related metrics such as A-weighted sound pressure level [1, 22] and loudness metrics [23, 24] are correlated with the annoyance of or preference for the sound. Thus, researchers have focused on the noise level reduction [22, 25–28] even though some have pointed out the importance of other sound attributes.

Researchers have studied sound quality evaluation of HVAC&R equipment components such as fans, compressors and also for diesel engines. Zwicker's Loudness metrics are highly correlated with annoyance scores of fan noise [1,23,29–33]. Moore and Glasberg, in the development of their time varying loudness model [34], suggested that it might be useful for HVAC noise because of temporal variations in loudness of such equipment. Zwicker's time varying loudness model has also been used to assess refrigeration noise [35]. Typically, Loudness exceeded 5% of the time (N_5) is used as the level metric, but for more impulsive noise, Berry and Zwicker [36], e.g., proposed use of Loudness exceeded 2 or 3% of the time as a predictor of loudness and/or annoyance. This may be relevant for the refrigerated truck noise in this study. Zwicker and Fastl's Psychoacoustic Annoyance model includes N_5 [15]. In addition, Topken [37] has developed metrics that are based on ratio between different frequency ranges from the Specific Loudness Spectrum (an outcome of Zwicker Loudness calculation), and those metrics were used to understand the characteristics of the fan noise. Leita, Paul and Gerges [38] and Hohls, Biermeier, Balschke, and Becker [39] studied sound quality of HVAC systems in vehicles and they found that Zwicker's Loudness exceeded 5% of the time (N_5) is highly correlated with the average of test subjects' annoyance ratings [38, 39].

A few researchers have focused on the tonalness of fan noise [32, 40, 41]. Lemire and Vo [28] pointed out the significance of tonality, but did not specify a metric for it. As shown in Table 1.1, there are tone adjusted level metrics that quantify the increase in annoyance due to a sound being tonal beyond the increase in level due to the presence of the tones [18, 19, 42, 43]. Yamaguchi, Minorikawa and Kihara [40] found that the tonalness metrics Prominence Ratio (PR) and Tone-to-Noise-Ratio (TNR) were highly correlated with annoyance ratings. Others have studied sound quality of fan noise. Sottek and Genuit [41] propose using their hearing model to assess tonal fan noise.

For compressor noise, which contributes to a significant part of the HVAC&R system noise, researchers have found that loudness and sharpness are the most significant sound attributes that affect sound quality [44–46]. Wang [44] also found that the beating phenomenon in a sound negatively affects the sound quality of compressors.

Measured refrigerated truck sound in this study contains diesel engine sound, which is an important factor in refrigerated truck annoyance model development. Researchers have found that the level [47–50], fluctuation [48, 50], tonalness [50] and impulsiveness [50–59] of the sound affect the annoyance ratings of diesel engine sounds. However, impulsiveness metrics proposed by various researchers are usually tailored to only one application and those are not generally applicable to other applications. Hohls *et al.* [39] found that people's preference ratings of automobile HVAC noise are correlated with the Articulation Index, Sharpness and Roughness. Articulation Index [60], now superseded by Speech Intelligibility Index [61] is a measure of speech intelligibility. The Sound Quality Indicator (SQI) [19], a tone adjusted loudness metric was developed for the refrigeration industry, but its use in that industry is very limited. It was originally combined with a sound power calculation, but in this research it was modified to use sound pressure recordings instead of sound power [62], and, as will be shown later, is correlated with average annoyance ratings of residential unit sounds.

Despite many studies in this research area, including both published and industry in-house studies, the only sound evaluations widely used are level-based metrics such as sound power level, A-weighted sound pressure level or loudness metrics. A summary of references are shown in Table 1.2. While several sound evaluation models are developed from the test results, model robustness may be an issue, due to the limited number of signals used in the test and the variety of signal types, correlation between metrics and small numbers of test subjects. Correlation between metrics was not reported in most studies. In some studies, a small number of test subjects (less than 20) participated in the subjective test. In these cases, often the standard error is higher than the difference between average annoyance ratings so the differences are not significant. Also in a few studies, a small number of sounds (less than 10) were used in the test which is a problem when estimating multiple linear regression models with several parameters. A high correlation between the independent variables in the regression model can lead to ill-conditioning of the system of equations used to estimate model parameters, leading to large uncertainty in the predictions from the sound quality or annoyance model. Also missing in many of these studies is model validation.

Model validation involves several stages. Initially estimated model predictions should be compared to averages of annoyance ratings for signals not used in the model estimation. The model predictions should also be compared to average responses for sounds (of similar types) used in other tests. Over the long term, the success of the model in guiding product sound design should be assessed and improvements made as the types of sounds products make evolve.

The research described in this document is an attempt to address many of these issues. For example, to decorrelate metrics and increase the variety of sound types used in tests, signal modification techniques were developed. In each test a variety of signals were used and a large number of subjects took the test. The focus on particular sound attributes was driven by how people described the sounds.

Table 1.2. Summary of subjective tests focused on sound quality of HVAC&R equipment noise.

Authors	Ref. Number	Number of Signals	Variety of Signal Types	Number of Subjects	Parameters in Model	Metric Correlation Reported?	Validation Model Tested on New Data?	Application
Susini *et al.*	[23]	19	13	50	Zwicker Loudness Spectral Centroid Noise-to-Harmonic Ratio	YES	NO	Air-conditioning unit
Sato *et al.*	[24]	24	6	9	Zwicker Loudness Roughness	NO	NO	Refrigerator
You *et al.*	[29]	5	NM	100	Zwicker Loudness Fluctuation Strength	NO	NO	Refrigerator
Schneider & Feldmann	[30]	Test1 - 12 Test2 - 7	Test1 - 3 Test2 - 1	Test1 - 25 Test2 - 8	Zwicker Loudness	NO	NO	Fan
Topken & Van de Par	[37]	30	3	40	Specific Loudness Ratio	YES	NO	Fan
Yamaguchi *et al.*	[40]	5	1	20	Tone-to-Noise Ratio Prominence Ratio	NO	NO	Fan
Cho *et al.*	[45]	2	2	Not Mentioned	Zwicker Loudness Sharpness Fluctuation Strength	NO	NO	Compressor
Champagne & Shiau	[52]	8	2	168	Zwicker Loudness Kurtosis	NO	NO	Diesel Truck
Payri *et al.*	[55]	5	5	8	Proposed Combustion Indicator	NO	YES	Diesel Engine
Leita *et al.*	[38]	5	1	27	Zwicker Loudness	YES	NO	Vehicle HVAC
Hohls *et al.*	[39]	7	7	34	Zwicker Loudness Articulation Index	NO	NO	Vehicle HVAC

1.2 Summary and Research Objectives

In summary, in many HVAC&R equipment and diesel engine noise studies people have noted that sound characteristics, in addition to the level of the sounds, impact the sound quality and assessments of the sounds. These characteristics include tonalness, spectral balance, fluctuation, roughness and impulsiveness. Despite these observations over many years, the only metrics in widespread use in the HVAC&R industry are level based metrics. In reviewing the literature on studies aimed at developing assessment models, it was found that there were many impediments to robust model development and model validation (e.g., too few subjects, lack of variety of sounds, correlated metrics). In addition, there is still on-going work in development of models that predict strength of perceived sound characteristics (sound quality metrics). Of particular note, is the large number of proposed tonalness metrics and impulsiveness metrics; the complexity of calculated metrics for roughness and fluctuation strength; and lack of agreement on which loudness model (Zwicker or Moore and Glasberg) is most appropriate.

With a good annoyance prediction model, the sound quality of HVAC&R equipment can be evaluated with a single sound measurement. This allows engineers to know which sound attributes make a sound unsatisfactory or enhance the sound quality. It is possible to predict the sound people will be exposed to if the physical noise source locations and strengths are known and a source-path model is developed. Having the two types of models (sound prediction and sound quality) gives the designers the ability to optimize the sound signature of the HVAC&R equipment. These procedures are illustrated in Figure 1.6.

The main objectives of this research are to: (1) understand the characteristics of sounds subjects are listening to; (2) understand and model how people perceive the strength of sound attributes; (3) combine the predicted sound attribute strengths to predict annoyance due to HVAC&R equipment noise, and (4) assess whether both types of equipment should have the same annoyance model or difference models.

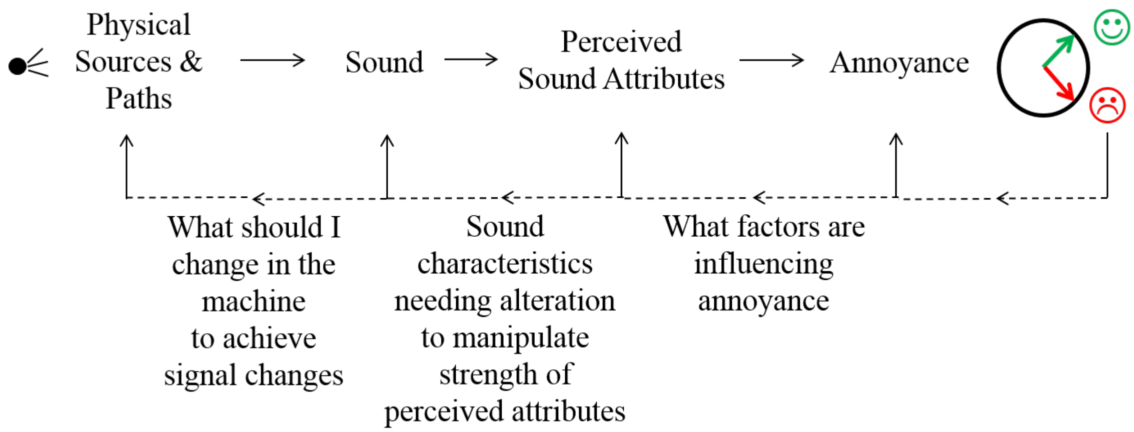


Figure 1.6. Process illustrating the product design guide.

1.3 Thesis Outline

The sound quality metrics and the types of psychoacoustic tests used are described in Chapter 2. The detail of the calculation procedure and formulae used to calculate the metric values from a signal are given in Appendix A. Chapter 3 is focused on preparations for the test. It includes a description of sound measurements, signal analysis, and the signal decomposition and modification processes, as well as the environment used for testing, and the general form of the psychoacoustic tests described in Chapter 4, 5, and 6. In Chapter 4, Chapter 5 and Chapter 6, three sets of tests (Test 1, Test 2 and Test 3, respectively) are described. Test results and analysis of results are also given. The performance of the proposed annoyance prediction models is also described in Chapter 6. Then, a brief example of a noise annoyance reduction strategy for residential HVAC&R equipment is described in Chapter 7. This includes an example of the use of acoustical holography to understand source contributions. Finally, Chapter 8 includes a summary of the findings and recommendations for future work.

In addition to the sound quality metric model descriptions in Appendix A, Appendix B contains MATLAB program code for signal modification, Appendix C contains a description of all the signals used in the research and the sound metric values for these signals, Appendix D contains the instructions given to subjects in the psychoacoustics tests, Appendix E contains a summary of the subjects' responses in the tests, and Appendix F contains a sample check list for the researcher conducting a particular test.

2. SOUND ATTRIBUTE MODELS AND PSYCHOACOUSTIC TESTING

HVAC&R equipment sounds have many sound characteristics which affect the annoyance rating from a subject participating in the test. To understand how those sound characteristics are affecting the annoyance rating, it is important to have models that calculate the strength of perceived sound attributes. In this chapter, a few models which predicts the strength of various sound attributes will be described. In addition, psychoacoustic testing methodologies and how they are applied in this study will be described.

2.1 Level Metrics

Many of the methodologies to predict level of a sound that is perceived have been standardized. The most commonly used methods are A and C-weighted sound pressure levels (dBA and dBC) which are both calculated based on equal loudness contours [10]. Even though A-weighted sound pressure level was developed for the use of relatively quiet sounds, unfortunately it is widely used for loud community sounds as well.

Loudness metrics model (in average sense) how people perceive the sound's strength. Stevens proposed the relationship between a perceived loudness and a stimuli intensity, called the "power law" [9]. Stevens' loudness model (various versions) are methods to predict the loudness of simple sounds relatively well but has some limitations on the calculation of loudness of complex sounds. Several improvements to his models have been proposed, such as Zwicker's [63] and Moore & Glasberg's [64] models, which are based on Stevens' work and equal loudness contours to predict loudness better by considering frequency selectivity and frequency and temporal masking. These models predict the loudness of complex sounds better than previous loudness models. Zwicker's model for stationary sounds was standardized in ISO 532B [65] and DIN 45631 [66]. The definition of critical bands at low frequencies is different in the two models and there are also differences at high frequencies. Zwicker's model has constant low frequency bandwidth, while the low frequency bandwidth of Moore and Glasberg's model decreases following similar trends similar to 1/3rd Octave bands at low frequencies. The two models are now incorporated into the latest ISO standards: Zwicker's Stationary and Non-stationary Loudness is in ISO 532-1 [6] and Moore & Glasberg Stationary Loudness is shown in ISO 532-2 [7]. The main update for the latter, over what is included in the ANSI S3.4 standard, is the change in how binaural summarion in calculated. Sottek [67] suggested use of a hearing model loudness estimation, which evaluates the spectral and temporal patterns of a sound, and attempts to fix the limitations of the existing loudness models which may overestimate or underestimate the loudness of impulsive sounds such as ratchet wheel or hammer sounds.

2.2 Other Sound Quality Metrics

Include here are models used or proposed to predict the perceived strength of other sound characteristics. These characteristics related to tonalness, spectral balance, impulsiveness, fluctuation and roughness.

2.2.1 Tonalness Measures

Tonal perception or tonality is not easy to describe. It is related to the detection of pitch or multiple pitches within a sound and the strength of those pitches. Pitches may not correspond to the frequency of tonal component in a spectrum, e.g., a harmonic series without the fundamental frequency could have a pitch at the fundamental frequency. Narrowband random noise can have a pitch associated with it, as can other quickly changing spectral features [15]. Pitch strength [68], which is related to tonalness, is difficult to interpret with complex signals. Most measures of the tonalness of a sound that are used are not sophisticated enough to deal with tonalness of a family of harmonics, and are measures the prominence of a pure tone in the surrounding broadband noise. There is research on complex tone perception but those results are, for the most part, yet to incorporated into the tonal measures used in engineering noise control. See, for example [50, 69–71]. Following is a description of tonalness metrics, used today.

The most simple and widely used tonality models are Tone-to-Noise Ratio (TNR) and Prominence Ratio (PR) which are described in ECMA-74 [11]. In both models, locations of the tones are identified from narrow-band spectra and the tonality of all identified tones are calculated, but only the highest tonality value is used. However, TNR and PR calculations are not suitable when multiple tones exist in a critical band. Compared to TNR and PR, the Aures' tonality model [12] sums all identified tonal components. In his model, the effect of bandwidth, center frequency of the tonal feature, prominence of the tone above the broadband component and the ratio of additional loudness due to tones are considered. Even though the calculation of Aures' Tonality is complicated, it is more applicable to a wider range of sounds. For frequency modulated or amplitude modulated sounds, there are still challenges with calculating the tonality metric values. A model referred to as psychoacoustic tonality is described in ECMA-74 [11] which is based on an improved hearing model by Sottek [72]. This model separates the loudness of tonal and broadband components and uses tonal components for the psychoacoustic tonality calculation.

2.2.2 Spectral Balance - Sharpness and Heaviness

The sound is sharper when the sound has more high frequency content than low frequency content. Sharpness metrics quantify the spectral balance of the sound. von Bismarck proposed a sharpness model [14] and Zwicker modified the model [15]. It weights the loudness spectrum with a function that is flat below 1000 Hz but increases above that. It then essentially finds the centroid of the spectrum. One acum is attributed to a narrow-band noise at 1 kHz with a bandwidth smaller than 150 Hz and a level of 60 dB. Aures' sharpness model [12] is another type of model that further emphasizes high frequencies (over von Bismark's model). Compared with the von Bismarck's sharpness model, the Aures' model has a higher degree of correlation with the loudness metric, but it is generally thought to be a more accurate model of people's perception of the sharpness of the sound. However, the lower correlation with Loudness in the von Bismark model makes it attractive when building models that are functions of several metrics.

2.2.3 Impulsiveness Models

Many researchers have studied the perception of impulsive sounds. C-weighted sound pressure level [73], perceived noise level (PNL) [74], a crest factor, a sound exposure level (SEL) [75], and Kurtosis of sound pressure [57] have been used to quantify the impulsiveness of a sound, or at least to have a better level related metric that quantifies perceived levels for impulse sounds. In a recent study on impulsive sound, Sottek and Moll found that existing loudness models strongly under or over estimate the loudness of impulsive sounds but the newer loudness model in standard ISO 532-1 performs better [76].

A few of the many suggested impulsiveness metrics were chosen and investigated in this research. The rate of change of the Zwicker Loudness [77] , Kurtosis of the Zwicker Loudness time histories and pressure time histories, the Heaviness ($dBC - dBA$), and the Crest Factor metrics were examined.

Kurtosis is a statistical measurement of a set of data and is the normalized fourth moment. Kurtosis has usually been calculated using the sampled pressure time history as in [51]. Kurtosis is not the metric developed based on the human hearing perception, but in this research it was calculated from the Zwicker Loudness time history, rather than the pressure time history, i.e., as a measurement of the presence of large excursions from an otherwise much lower level of loudness.

The Crest Factor measures the degree of prominence of an event in a signal and is defined by the absolute peak amplitude of the signal to the root mean square value of the signal. Similar to Kurtosis, the Crest Factor is not a psycho-acoustic metric but is a measure of the characteristics of the sound pressure time history, and it sensitive to large deviations from the standard deviation of the time history.

Rate of change of loudness is another metric that has been used to characterize suddenness of events. Marshall [16] used it when examining transient sounds. Marshall used it when examining transient sounds. Marshall applied it to the time history of loudness predicted by using Moore and Glasberg's model of loudness, but in the research described in this document, it is derived from the Zwicker's Time-Varying Loudness time history. The Zwicker Loudness time histories are calculated by using the Head ArtemiS software and the sampling internal is 0.002 seconds. A 121-point FIR filter differentiator that works over the range $0 - 0.2f_s$ Hz was used to calculate the derivatives. The filter was designed using the firpm program in MATLAB. The rate of change of loudness exceeded 2% of the time was examined as a potential impulsiveness measure.

"Pounding", a word used by subjects, may be a combination of impulsiveness and low frequency content in the signal, so a heaviness metric was also considered. Heaviness is the difference between C-weighted and A-weighted sound pressure levels and is a measure of low frequency energy in the signal. As noted earlier in this impulsiveness section, the Heaviness metric was also examined to account for low frequency contributions in HVAC&R sounds.

2.2.4 Fluctuation

Slow variations in loudness that can be tracked result in the sound being perceived as fluctuating and the fluctuation Strength metric quantifies the strength of this fluctuation ([15], Chapter 10). The fluctuation sensation is frequency dependent and it is a maximum at approximately 4 Hz. Zwicker developed two fluctuation strength models, one is for fluctuating broadband noise while the other is for tonal sounds [15]. Similar to the Roughness metric calculation, determination of a modulation frequency and the modulation depth for realistic sounds is problematic. A Fourier analysis of the specific loudness through time in a critical might reveal several contributing frequencies and it is not clear how to proceed from this observation. However, for sinusoidal variations in level, it can be straightforward to calculate. A 60 dB tone at 1 kHz center frequency with a 100% amplitude modulation at 4 Hz results in a Fluctuation Strength of 1 *vacil*. Modulations above 16 Hz are very difficult to track and so Fluctuation Strength drops rapidly as the modulation frequency increases.

2.2.5 Roughness

Sound where there are fast fluctuations in loudness are perceived as rough, and roughness is strongest when the rate of fluctuation is between 60 and 70 times per seconds. In the model the depth of the fluctuation is explicitly included and the frequency of the fluctuation is included indirectly through the temporal smoothing of the loudness spectrum. The roughness metric (Roughness) calculation is well defined for a modulating simple tone, but for complex sounds, there may be several modulation frequencies present and it is not clear how these should be combined to predict the perceived roughness of the sound. Zwicker proposed two roughness models [15] of which one asper is attributed to a sound with a 1 kHz center frequency, a 100%, 70 Hz amplitude modulation, and an overall level of 60 dB.

2.3 Level Metrics Adjusted for Tonal Presence

Some metrics were developed to estimate annoyance due to exposure to a sound that includes tonal components. Because people are familiar with level measures, the additional annoyance due to the presence of tones is expressed as an addition to the level measure. A sound with no tones at this level would be as annoying as a quieter sound with tonal components. The level metrics used in these models are either A-weighted sound pressure level or a loudness metric, see Table 1.1, for a list of such adjusted level models. The calculation of Tone-corrected Perceived Noise Level ($PNLT$) and Effective Perceived Noise Level ($EPNL$) is given in Federal Aviation Regulations [18]. Both metrics are used for aircraft noise quantification. The Perceived Noise Level (PNL) is obtained from on the third-octave sound pressure levels. The sound pressure levels are converted to noy values to calculate the PNL. $PNLT$ adds a tone correction, and then $EPNL$ is calculated based on $PNLT$ over time and takes into account the perceived level of an event.

The Sound Quality Indicator (SQI) is a standardized measure developed for the the Air-conditioning, Heating and Refrigeration Institute to quantify the quality of sounds from air-conditioning and refrigeration equipment [19]. The sound quality indicator calculation procedure is based on Zwicker stationary loudness model (ISO 532B) calculated from third-octave data. The level of the third-octave band is adjusted when it is prominent relative to adjacent third-octave band levels. Similarly in environmental noise standards, a level adjustment to A-weighted sound pressure level is recommended when a tone is prominent [42, 43]. The tonal audibility calculation in this standard is similar to Tone-to-Noies Ratio (TNR).

2.4 Other Annoyance Models

Zwicker and Fastl proposed a Psychoacoustic Annoyance model which incorporates Zwicker Loudness exceeded 5% of the time, Roughness, Fluctuation Strength and von Bismark Sharpness [15]. The Psychoacoustic Annoyance (PA) model was developed from studies on transportation noise but it is not widely used in HVAC&R industries. It does not contain a tonalness related metric, which is of concern for HVAC&R machinery. When studying responses to diesel engines Hastings also developed a modification to include tonalness [50]. More and Davies also modified Psychoacoustic Annoyance to include tonalness [78] and tuned the model to predict aircraft noise annoyance. The additional tonality terms are composed of Zwicker

Loudness exceeded 5% of the time and Aures' Tonality exceeded 5% of the time.

In the present research, multiple linear regression models were used to predict the annoyance of HVAC&R equipment noise. A linear combination of sound quality metrics is used to predict an average annoyance rating of HVAC&R equipment sound, although some individual metrics in the model are modified by limiting their contributions to predicting annoyance at low and high values. Limiting metric contributions in annoyance models is not new. The tone penalties to level metrics mentioned above all have lower and upper limits, and in the Psychoacoustic Annoyance model the Sharpness metric only contributes to annoyance predictions after exceeding a lower limit.

2.5 Psychoacoustic Testing

The design of subjective tests to provide sufficient data to build robust models, which can be used to predict peoples responses to similar types of sounds, can be challenging. There are many things to consider: the information sought from the test results; the sounds used in the test; the test environment; the quality of the sound reproduction; the testing methodology; the subject pool; protection of subjects; Institutional Review Board approvals, etc. Test design is never perfect because each design is a balance of conflicting requirements. One example is wanting to include a great variety of sounds that span the acoustic range of the types of products being studied, but subjects can get bored or tired if tests are too long. Sounds from a particular product type may span only part of this range of sounds, but a test subject listening to the larger group of sounds may only focus on the level of the sound, even though other sound characteristics may be important when people are exposed to sounds with a smaller loudness range, which is usually be the case in the real life situation. Generating a model for each group of products may be solution, but this can become confusing for people generating new products particularly at the interface between two product types - which model should be used? In the next few sections, some of the methodologies used in this research and issues that arise when designing a test are briefly described.

2.5.1 The Sounds

Sounds used in this research were either measurements of operating equipment or modified measurements. Algorithms were developed to control the strength of sound attributes, so that those attributes' effects on overall judgements (e.g., annoyance) could be examined. Sounds on a desktop computer were played through a LynxONE sound card connected to a Furman SP-20AB stereo amplifier, connected to ER-14A research headphones. The headphones produce a sound at the eardrum, thus desired sounds were filtered to simulate the sounds that would be measured at the "average subject"'s ear drum. The outer-ear frequency response in ISO 532-Part 2 was used as the basis of the FIR filter design [7]. Sounds were calibrated by using a Bruel and Kjaer Type 4946 coupler.

2.5.2 Subject Experience

A subject's responses are affected by many things, such as: the computer screen, content of messages, and other participants if multiple subjects are present simultaneously. It is important to control the test procedures and the test environments to ensure all subjects experience the same sound evaluation process during the test. In this research a checklist was developed to ensure that exactly the same procedure was followed for each subject to ensure test procedure and subject experience consistency. A sample of a checklist used in this research is given in Appendix F. Providing a consistent, detailed and understandable set of test instructions to all subjects is important to give the same information to all test subjects and to ensure they understand what is being asked of them. Test Instructions used in this research are given in Appendix D.

In this study, only one subject was tested at a time. The test took place in an

IAC double walled quiet room (Ford Sound Quality Chamber) in the Acoustic Wing at the Ray W. Herrick Laboratories at Purdue University. During the main parts of the tests when subjects were evaluating sounds, each subject heard a set of sounds in a randomized order, and a different random order was used for each subject who participated in the test. Because subjects' responses are averaged, this also helps to reduce ordering effects of sounds because the number of presentation orders in the test is higher than when subjects are tested in groups. Also, when subjects are tested individually, there is very little risk of people influencing each other's responses.

In the next subsection, different types of psychoacoustic tests (rating test, paired comparison test, and semantic differential test) that were used in this research will be briefly explained.

2.5.3 Psychoacoustic Tests Used in This Research

In this research, rating tests using a Likert scale [79], a paired comparison test [80], and a semantic differential test were used [81]. In a Likert test [79], subjects are asked to evaluate the sounds on a scale marked with a level of sensation, for example, "Not at all Annoying", "Slightly Annoying", "Moderately Annoying", "Very Annoying", and "Extremely Annoying" as shown in Figure 4.1 in Chapter 4. These words on the scale can be changed by application. Words on the scale are directly related to the question the subjects are being asked, e.g., "How annoying is the sound?". It is important to note that the sound evaluation is affected by the previous sound heard and so the order of sounds for each subject within a test is usually randomized so these ordering effects are reduced when the ratings are averaged over the subject pool. Also, there are some subjects who avoid using on the extreme ends of the scales while others tend to answer on the extreme ends of the scales; both of these types of subject practice can significantly affect the results particularly when there are very few subjects participating in the test. Usually the mean of the ratings and the standard deviation of the estimated mean are calculated. The distributions of the data across the subject group and for each signal are also examined. Typically the subject ratings' distributions vary, some occupy small ranges either at the top, middle or bottom of the range, and others use the full range. Despite this variability in individual responses, repeatability of the average of the responses in a well designed test can be quite high.

In a paired comparison test, subjects compare sounds in pairs and choose one in response to each question, such as, "Which sound is more annoying?". From the subjects' choices the probability of choosing one sound over another is estimated. Both orders of presentation A B and B A are included, so any tendency to choose the first or second presented sound can be averaged out of the probability estimates. These estimated probabilities are transformed by using e.g., the Bradley, Terry, Luce (BTL) transform [82], which was used in this research to give an estimate of relative sensation levels. In the transformation a reference must be used because only relative information is collected. This was done in this research usually by setting the average of the BTL values to 0, but it could be done by setting one of the sounds to have a particular rating. The estimated probabilities should not be 0 or 1, so some adjustment is required if they are, for example, by modifying 0 or 1 to $1/(2n)$ or $1 - 1/(2n)$, respectively, where n is the number of subjects. The paired comparison test is the most simple test for a novice test subject and usually only requires a short practice session for the subject to become comfortable making their choices. However, when the number of sounds are increased in the tests, the paired comparison tests takes much longer time to complete $2(N^2 - N)$ presentations involving N sounds as opposed to Likert test (N presentations of sounds). Analysis of the results of the paired comparison test can provide a ranking of the sounds, and by using the BTL (or some other transformation) the relative perceived sensation level. Thus, it is not as straightforward to interpret the BTL values, unlike the average ratings in a Likert test. However, even in a Likert test as used in this research, with the words on the scales, subjects sometimes expand their use of the scale on the set of sounds they are exposed to in the test.

In a semantic differential test [81] a subject rates sounds on many word scales. An example of a rating sheet is given with the subject instructions for Test 2 which is given in Appendix D. The ends of scale have words with opposite meanings. Scale ends are associated with adjectives or phrases such as "soft – loud", or "not at all annoying – extremely annoying". Thus, it is very important to have scales which can cover most of the sounds' attributes and also have words at the ends of the scales that subjects understand. In this research it will be seen that the "not tonal – highly tonal" scale seemed problematic for subjects. Some training or examples may be needed for such scales. The end-scale words can be determined by a literature review of related fields particularly those that include semantic differential tests for sound evaluation. Sometimes a simple description test can help in the selection the words [16, 83]. The results of the semantic differential test is interpreted by using factor analysis to find the number of independent factors affecting ratings. The careful choice of end-scale words provides results specific to a particular application. However, it is more difficult for subjects to understand how to properly complete the test when compared to understanding how to do the the rating test or a paired comparison test. That is, there may be 20 scales and subjects need to gain familiarity with the meaning of the end points of the scales. This is a lot more complicated than answering, e.g., which is more annoying? Also the time the evaluation takes for one sound limits the number of sounds that can be evaluated before subjects are tired. This means that model development is difficult using these types of tests, as is the case with paired comparison tests. Ideally, even with just a simple 3-metric linear annoyance model it would be desirable to have greater than 40 sounds evaluated ($>$ 10 times the number of estimated parameters in the model). For the development of a robust model, the set of sounds must cover the range of sound attribute strengths that could be present in the product sounds, and combinations of metrics (measures of sound attribute strengths) must be chosen so correlation between metrics is low for the set of sounds chosen. Robust model development points towards the need for more sounds than is practical in a semantic differential or a paired comparison test. Including the desired number of sounds in these types of tests, would be very time-consuming, and would involve either multiple test sessions or a much larger subject pool, where subjects only took parts of the test.

2.6 Chapter Summary

A number of level-based metrics, other sound quality metrics, and annoyance models have been described and a brief overview of psychoacoustic testing was given. Preparations for subjective testing are described in the next chapter. This includes the methodologies to vary the sound attributes measured by using the loudness, sharpness, roughness, and tonality metrics described earlier in this chapter, and sound metric decorrelation. Also given at the end of Chapter 3 is an overview of the three subjective tests described in Chapters 4-6, and the parts of the testing procedures that are common to all three tests.

3. PREPARATIONS FOR TESTS: SIGNAL ANALYSIS, SIGNAL MODIFICATION, AND TESTING PROTOCOLS

A good measurement is consistent and repeatable, and it is important to have sets of high quality recordings free of artifacts and are representative of the range of sounds typically encountered for the application of interest. To make good measurements, any measurement test standard for the sounds must be strictly adhered to. Analyzing the properly measured signal is an important step in understanding the sound's characteristics. For steady-state sounds that were used in this research, frequency domain analysis provides useful information: e.g. harmonic families related to repetition rates of machine components can be identified; the width of tonal features reflecting presence of variations in repetition rates; and the frequency contents of the broadband components. Temporal features are also important and must be examined. It is well known that averaged spectra that look very similar can originate from very different sounds. To develop robust annoyance prediction models, it is important to be able to vary different sound attributes to observe how individual attributes affect a subject's response. To create signals that have different strengths of important sound attributes, sound modification techniques were developed. In the following sections, detailed descriptions of sound measurement, sound analysis, sound decomposition, and sound modification will be described. The Chapter ends with a description of the features common to the three main psychoacoustic tests conducted in this research.

3.1 Sound Measurement

The sponsor had provided many sound measurements for this project but there was a need to supplement this set of sounds. Specifically, adding to the database of

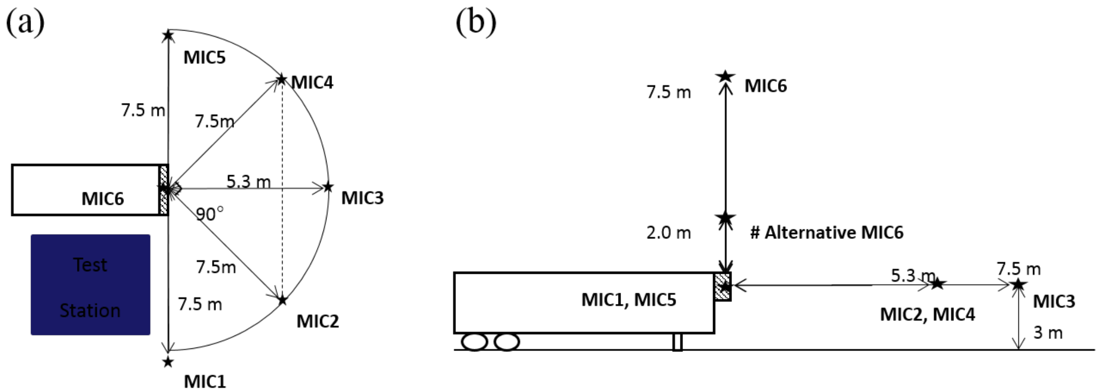


Figure 3.1. (a) Top view and (b) side view of refrigerated truck unit showing six microphone locations as recommended in the TNO report [85].

refrigerated truck sound measurements (pressure versus time). The sound measurement procedures followed the procedures described in two measurement standards ANSI/AHRI 1120 [84] and European PIEK standard [85]. Test setups were modified in accordance with the test purpose and the environment. Measurements were taken of sounds from 2 refrigerated truck units, the Vector 1350 and Vector 1950. For both units, there were two operating conditions. Figure 3.1 is a schematic of the unit and microphone locations as recommended in the TNO PIEK measurement procedure report [85]. To be compatible with the ANSI/AHRI Standard 1120, microphones 1, 3, and 5 were added at 3m above the ground. For safety issues, instead of locating microphone 6 at 7.5 m above the unit, the TNO Report included recommendations for alternative positions (2m above the unit), which were used. Figure 3.2 is a schematic of the truck and microphone locations recommended in ANSI/AHRI Standard 1120. The height of the microphones was 1.5 m above the ground. This was selected to locate the microphone closer to a person's ear position when standing. Microphone 6 (2 m above the unit) was added to be consistent with the above unit alternative microphone position recommended in the TNO report.

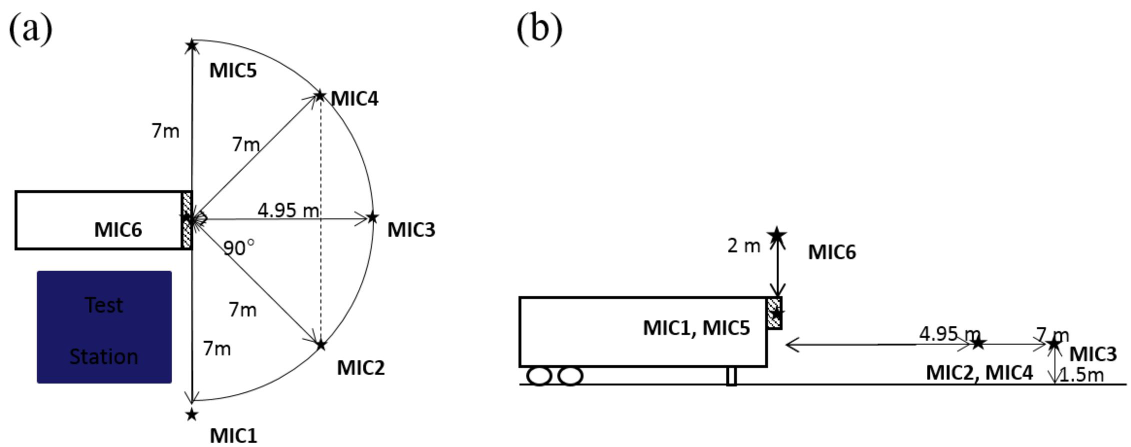


Figure 3.2. (a) Top view and (b) side view of refrigerated truck showing six microphone locations as recommended in ANSI/AHRI Standard 1120 [84].

An examples result showing a pressure time histories and corresponding power spectral density estimates are shown in Figure 3.3. The spectral resolution for the power spectral densities in Figure 3.3 is 1.25 Hz. The power spectral densities were estimated by using segment averaging and each segment was windowed using a Hann window (sometimes referred to as a Hanning window). 50% overlapped segments were used and the raw power spectral densities from each of the 38 segments were averaged. Finally, the estimate of the power spectral density was compensated due to the attenuation of the signal when using the Hann window (divided by 0.375). In Figures 3.3 (b) and (c), the spectral power distribution is shown and some prominent peak frequency locations can be identified. Components such as fans, motors, compressor, and a diesel engine, when operating at constant speed, produce harmonic families: a series of tones that are an integer multiple of a fundamental frequency that is directly related to the speed and repetition rates of events when the component is operating.

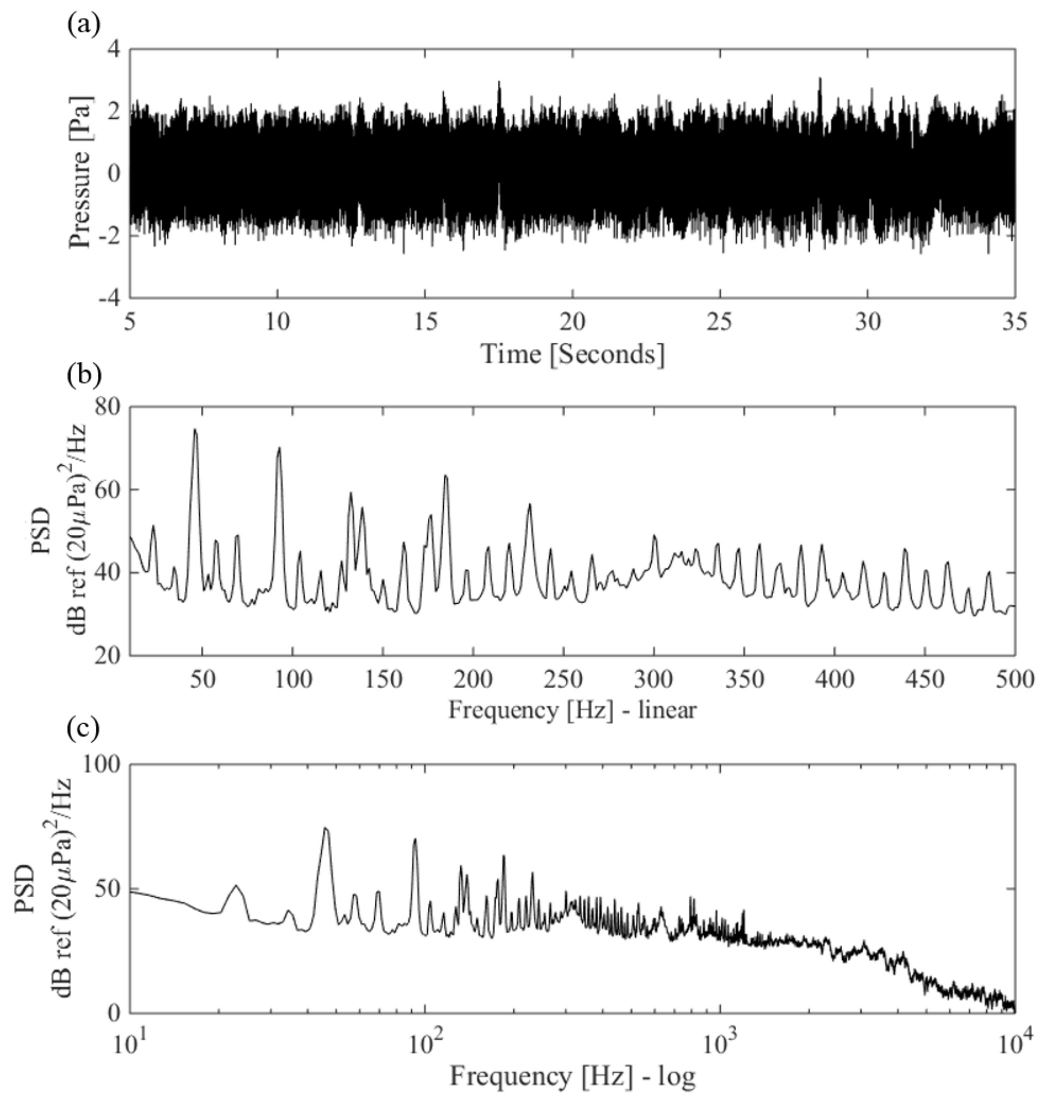


Figure 3.3. Example of a (a) pressure time history, and its power spectral density on (b) a linear (0-500 Hz) and (c) a logarithmic (10-10000 Hz) frequency scale. This measurement was from microphone 6 (Top) of a Vector 1350 unit running at 1800 rpm.

3.2 Sound Decomposition

As noted earlier, HVAC&R sounds consist at steady state of broadband and tonal components. To be able to vary the strength of sound attributes and relate them to physical components in the system, the tonal and broadband components are separated, modified and then reassembled to create a new sound. The harmonic families caused by the fans, the compressors or the diesel engine may not consist of pure tones, a tonal component's frequency and amplitude sometimes vary with time and theses variations need to be captured for proper signal decomposition, especially when these variations are high. Captured variations can be amplified or attenuated to change sound attributes. An overview of the tonal component extraction process is given in Figure 3.4. The approach is similar to the method taken by McMullen [16]. The outcomes of this process are an amplitude and frequency time history for each of the extracted tones. Further details of the properties of the filters used in this process are given in the following sections.

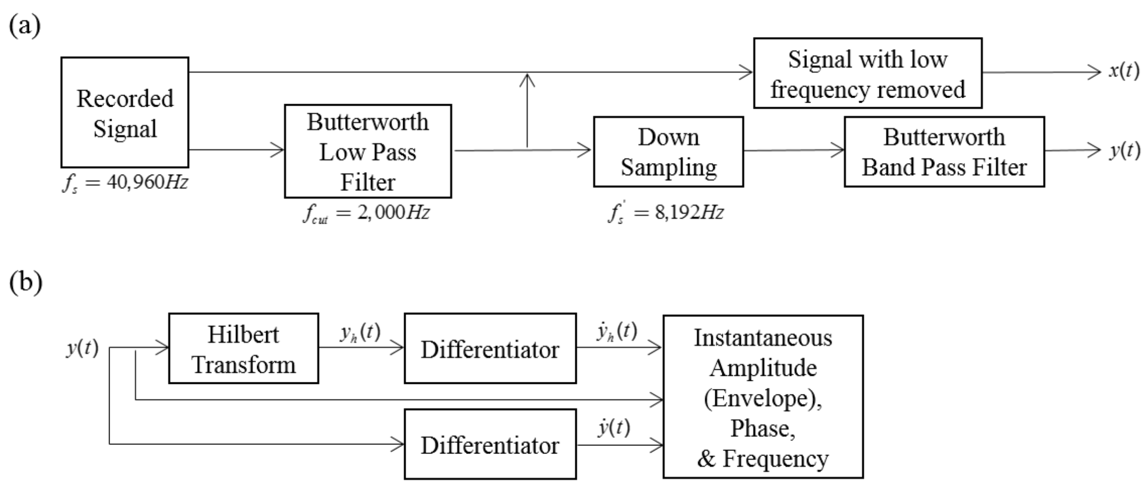


Figure 3.4. Flowchart of the tonal components extraction process and instantaneous amplitude, phase and frequency calculation.

3.2.1 Tonal Component Extraction

The sampling frequency of the recorded signals is 40.96 kHz. A 6th-order Butterworth digital low-pass filter with a cutoff frequency of 2000 Hz was used to avoid aliasing when the signal was down-sampled to 8192 Hz, a factor of 5 down from the original sample rate. The downsampling was helpful in the calculation of the instantaneous frequency and instantaneous frequency and instantaneous amplitude of tonal components. Tonal components of interest were in the 0 to 3000 Hz range, but tones above 2000 Hz are usually harmonics of lower frequencies and could be simulated by using the information from the lower frequency component analysis. Eighth order digital Butterworth bandpass filters were used to extract tonal components, and the choice of the bandwidth of these filters were based on the features of the power spectral density of the signal, depending on its apparent bandwidth and proximity of nearby tones. The bandwidths ranged from 5 to 50 Hz.

3.2.2 Instantaneous Phase, Amplitude and Frequency

In continuous-time theory, the extracted tone signal is modeled as:

$$y(t) = A(t)cos(\phi(t)) = A(t)cos(2\pi f_c t + \psi(t)), \tag{3.1}$$

where $A(t)$ is the instantaneous amplitude or envelope, $\phi(t)$ is the instantaneous phase, f_c is the center frequency of the tonal component and the derivative of $\psi(t)$ with respect to time gives the variation of the instantaneous frequency:

$$f(t) = \frac{1}{2\pi}\frac{d\phi(t)}{dt} = f_c + \frac{1}{2\pi}\frac{d\psi(t)}{dt}. \tag{3.2}$$

The instantaneous amplitude (envelope) and instantaneous frequency are calculated by using the original signal $y(t)$ and the Hilbert transform of the signal $y_h(t)$, which is given as:

$$y_h(t) = A(t)sin(\phi(t)), \tag{3.3}$$

and the instantaneous amplitude, phase, and frequency are defined as:

$$A(t) = \sqrt{y^2(t) + y_h^2(t)}, \phi(t) = tan^{-1}\frac{y_h(t)}{y(t)}, f(t) = \frac{1}{2\pi}\frac{\dot{y}_h(t)y(t) - \dot{y}(t)y_h(t)}{A^2(t)}, \tag{3.4}$$

where $\dot{y}_h(t)$ and $\dot{y}(t)$ are the time derivatives of $y_h(t)$ and $y(t)$, respectively.

Finite impulse response (FIR) filters were designed by using the *firpm* function in MATLAB program, which is an implementation of the Parks McClellan algorithm for optimal FIR filter design. A 255-point FIR filter was designed which emulates a Hilbert transformer in the frequency range of 50 to 4050 Hz. A 121-point FIR filter was designed which behaves as a differentiator over the frequency range 0 to 2000 Hz. This filter acts as a low-pass filter above 2000 Hz. The instantaneous phase, and instantaneous frequency were calculated by using Equation (3.4) with t=t_n=$n\Delta$, where n is the sample number, and Δ is the sampling interval in seconds. In this research, the frequency range of tones extracted and modified was in the range 0 to 2000 Hz, but this range can be extended by redesigning the FIR differentiator filter.

3.2.3 Estimated Amplitudes of Tonal Components

For most of the signals that have been analyzed in this research, the amplitudes of the tonal components changed very little over time and some of those variations observed mainly come from the noise floor in the signal. Thus, a constant amplitude model for the tonal components was used. The calculated instantaneous phase of each tonal component was used rather than assuming that the frequency of each component is constant. This was found to be more accurate when estimating the constant amplitudes of the tonal components in the HVAC&R equipment sound. While the model of the tonal component given in Equation (3.1) contains only a cosine term, a sinusoidal term with the same phase was also added to this constant amplitude model; this was found to improve the estimation of the tonal components in the signal. The constant amplitudes were thus calculated by solving, in a least squares sense, the following set of equations:

$$\begin{bmatrix} p(t_1) \\ \\ \vdots \\ \\ p(t_N) \end{bmatrix} = \begin{bmatrix} cos[\phi_1(t_1)] & sin[\phi_1(t_1)] & cos[\phi_2(t_1)] & \cdots & sin[\phi_{nc}(t_1)] \\ cos[\phi_1(t_1)] & sin[\phi_1(t_2)] & cos[\phi_2(t_2)] & \cdots & sin[\phi_{nc}(t_2)] \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ cos[\phi_1(t_N)] & sin[\phi_1(t_N)] & cos[\phi_2(t_N)] & \cdots & sin[\phi_{nc}(t_N)] \end{bmatrix} \begin{bmatrix} A_1 \\ B_1 \\ A_2 \\ \vdots \\ B_{nc} \end{bmatrix} \tag{3.5}$$

where A_k and B_k are the estimated constant amplitudes of the cosine and sine components, $p(t_n)$ are samples of the down-sampled low-pass filtered signal from which the tonal components were extracted, t_n are the sampling times, $\phi_k(t)$ is the instantaneous phase of the k-th tonal component, and nc is the number of tonal components being fitted to the down sampled signal. For signals where amplitudes are varying with time, an approach that estimates the tonal component amplitudes through time by using a subset of the signal around each time sample to estimate A_k and B_k can be used. This approach was used by McMullen [16] to model tonal components in rotorcraft noise.

The pressure-time histories of the tonal components were calculated by using the estimated constant amplitudes and calculated instantaneous phases. By subtracting the tonal signal from the down-sampled signal, the signal containing the broadband and un-selected tonal components was determined. Both tonal and broadband components with un-selected tonal components were then up-sampled to the measured sampling frequency of 40,960 samples per second and added to the high frequency part of the original signal ($x(t)$ in Figure 3.4). The estimated power spectral densities of the original signal (black) and the signal with tonal components removed (blue) are shown in Figure 3.5. The estimation of the constant amplitudes using Equation (3.5) helped to retain the noise floor where the tonal components were removed.

3.3 Sound Modification and Reconstruction

When developing sounds for subjective tests, it is important to fill in the sound attributes gaps that are normally present in sets of sound measurements. For example, if in a set of measurements there were only two levels of compressor tones one level very low and one level high, the signal modification method can be used to fill the gap so the set of signals contains many gradually increasing compressor tone levels. This enables examination of where compressor tone levels begin to strongly affect annoyance. Predictions of perceived sound attribute strengths (sound quality metric

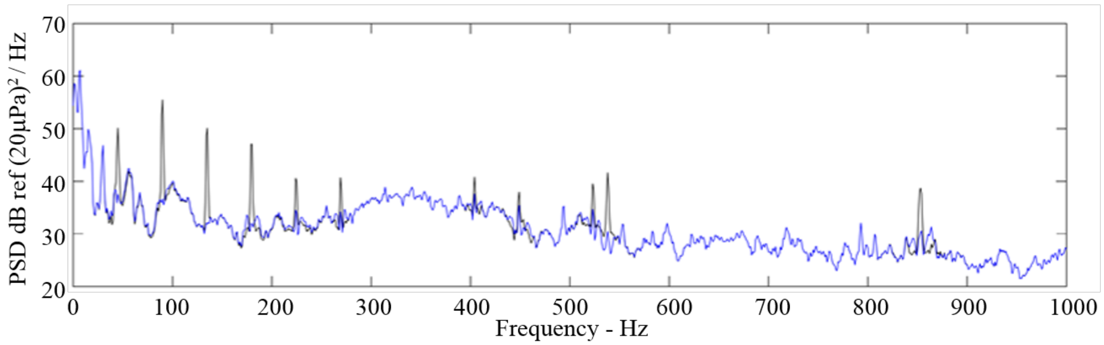


Figure 3.5. Power spectral densities of the recorded signal (black line) and the signal with selected tonal components removed (blue line). The original signal was a measurement of a residential unit (RA9) operating at fan 897 rpm (BPF = 44.85 Hz).

values), often exhibit gaps. Sound quality metrics include predictions from models of loudness (Zwicker [63], Moore&Glasberg [64], ISO532 Part 1&2), models of sharpness [12, 14], tonality [11], roughness [15] and fluctuation strength [15]. Impulsiveness is another sound attribute that was found to be important in this research. Annoyance models are functions of these metrics and for robust estimation of model coefficients there is a need to de-correlate metric values across the set of test signals.

An overview of the sound modification and reconstruction process is shown in Figure 3.6. The tonal and broadband components were manipulated individually by changing the amplitudes, changing spectral balance, and/or by adding modulation. The different methods of modifying signals are described in the following sections.

3.3.1 Adjustment of the Magnitude of Tonal and Broadband Components

Adjusting the level of the tonal and broadband components affects the quality of the sound which is reflected in the sound quality metric values that measure tonality, roughness, and loudness. Power spectral densities of tonal components amplified or broadband component amplified signals are shown in Figure 3.7. Metric values for

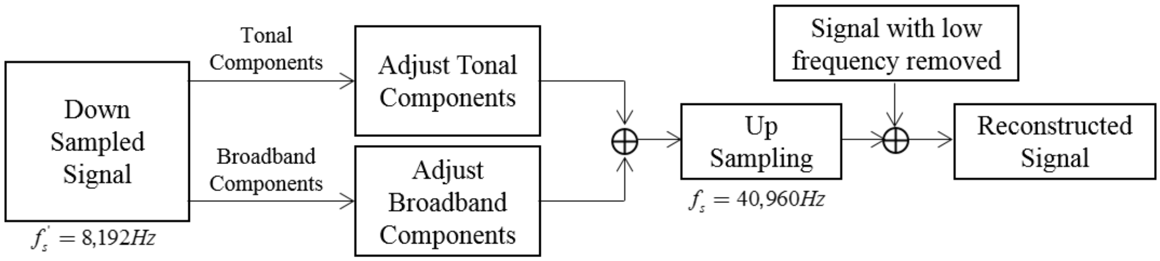


Figure 3.6. Flowchart of the sound modification and reconstruction process.

different amplification factors are shown in Figure 3.8. As tonal components are amplified, Zwicker Loudness exceeded 5% of the time and the DIN Tonality metric values increase proportionally, but the Roughness metric values did not change significantly. To produce realistic sounds, the amplification factor of the tonal components should

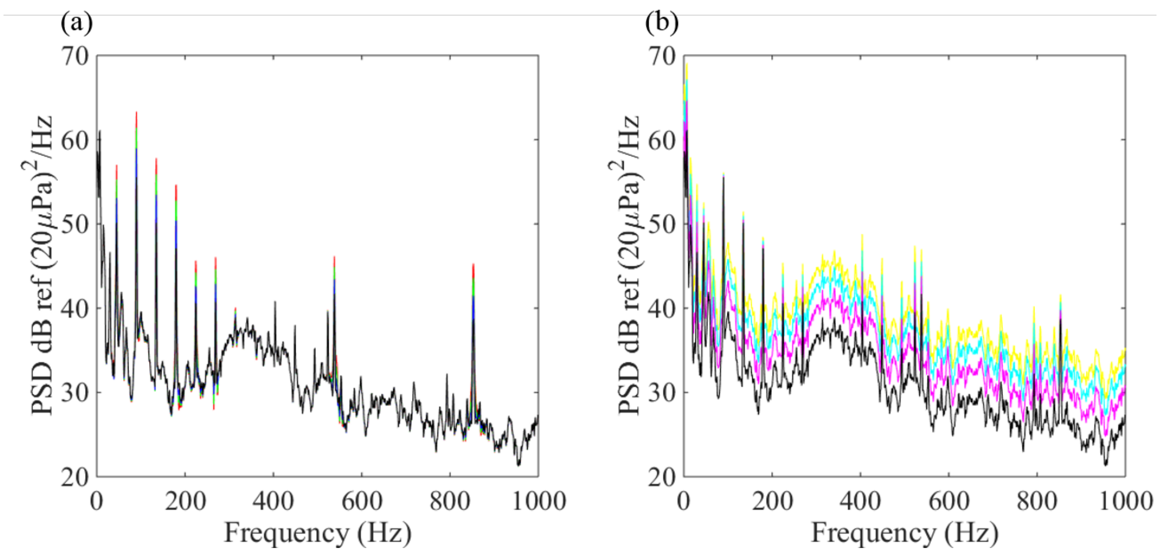


Figure 3.7. (a) Power spectral densities of the original signal (black line) and the tonal-component-amplified signals (amplification factors - blue: 1.5, green: 2, and red: 2.5), (b) Power spectral densities of the original signal (black line) and the broadband-component-amplified signals (amplification factors - magenta: 1.5, cyan: 2, and yellow: 2.5).

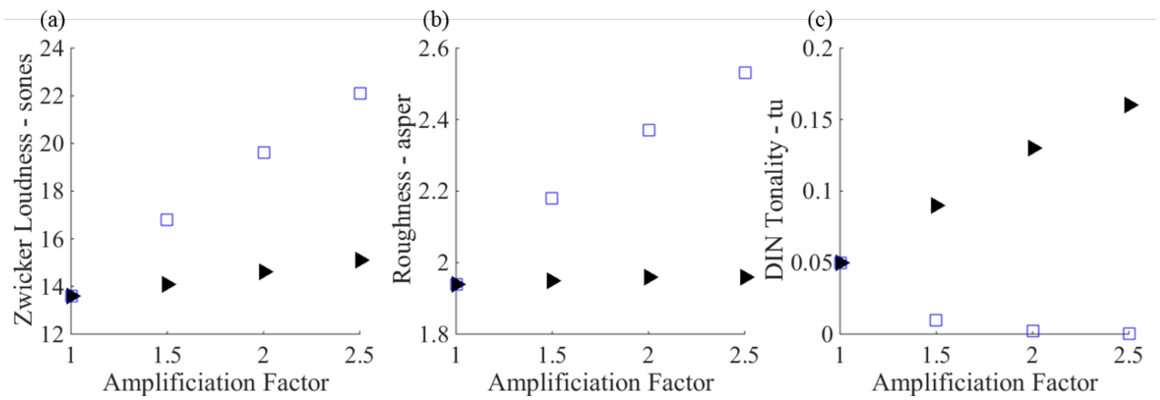


Figure 3.8. (a) Zwicker Loudness exceeded 5% of the time, (b) Roughness exceeded 5% of the time, and (c) DIN Tonality exceeded 5% of the time plotted against the amplification factor used for tonal components (►) and for broadband components (□).

not be too large. When the broadband components are amplified, Zwicker Loudness and Roughness metric values increase while the DIN Tonality metric values gradually decreases. In summary, increasing tonal or broadband components increases the Loudness metric values, though Tonality and Roughness metrics values do not always increase, they sometimes decrease. Increasing Tonality while Roughness and Loudness remain constant can be challenging.

3.3.2 Adding Modulation to Tonal and Broadband Components to Control Roughness Levels

To control Roughness, amplitude modulation was added to the tonal and/or the broadband components of the signal using a sampled version of the following equation:

$$x(t) := [1 + \gamma_1 cos\Psi(t)]x(t), \tag{3.6}$$

where $x(t)$ is either the tonal or the broadband components, γ_1 is a constant and $\Psi(t)$ is phase term. The modulation frequency is:

$$f_m(t) = \frac{1}{2\pi}\frac{d\Psi(t)}{dt} = 60 + \gamma_2\alpha(t) \text{ Hz}, \tag{3.7}$$

where $\alpha(t)$ is low pass filtered Gaussian distributed white noise (order = 1, cut-off frequency 200 Hz). The random component $\alpha(t)$ is adjusted to have a standard deviation of 1 and γ_2 is used to vary the range of the frequency variation about 60 Hz. The center frequency of the modulation was chosen to be 60 Hz because this is in the modulation frequency that affects the roughness sensation the strongest [15]. The phase, $\Psi(n\Delta)$ is calculated by digitally integrating the $2\pi\gamma_2\alpha(t)$ using MATLAB and adding the result to $2\pi 60t$. Both γ_1 and γ_2 affect the sound, and if too large, the sound does not sound realistic. Power spectral densities of the original signal and the signals where tonal and broadband components have been modulated are shown in Figure 3.9. The metric values for different values of γ_1 are shown in Figure 3.10. Zwicker Loudness and Roughness metric values increase as γ_1 increases, but the Tonality metric values vary depending on the signal's tonal and broadband components ratio. To increase a Roughness metric value, it appears to be more efficient to add modulation to the broadband components.

3.3.3 Adjusting the Spectrum Shape

Adjusting the balance of the spectrum can change the Sharpness metric values (Aures' or Von Bismark models). Often, the correlation between the Loudness and Sharpness metric values is quite high, so the goal in this modification is to have the Loudness metric value remain constant while the Sharpness metric values vary. Including these modifications in the set of signals will effectively de-correlate the Loudness and Sharpness metrics. The method developed by Carr [86] was used. Zero-phase finite impulse response (FIR) filters were designed that had flat frequency response regions below f_a Hz and above f_b Hz with a half-cosine wave shape in the transition region. f_a and f_b are chosen to be the centers of critical bands in regions around 1000 Hz. Depending on the nature of the sound and the critical frequency bands, the frequencies that define the flat regions can be adjusted. In this research $f_a = 1000$ Hz and $f_b = 2000$ Hz was used. By adjusting the gains in the flat regions, Sharpness values can be

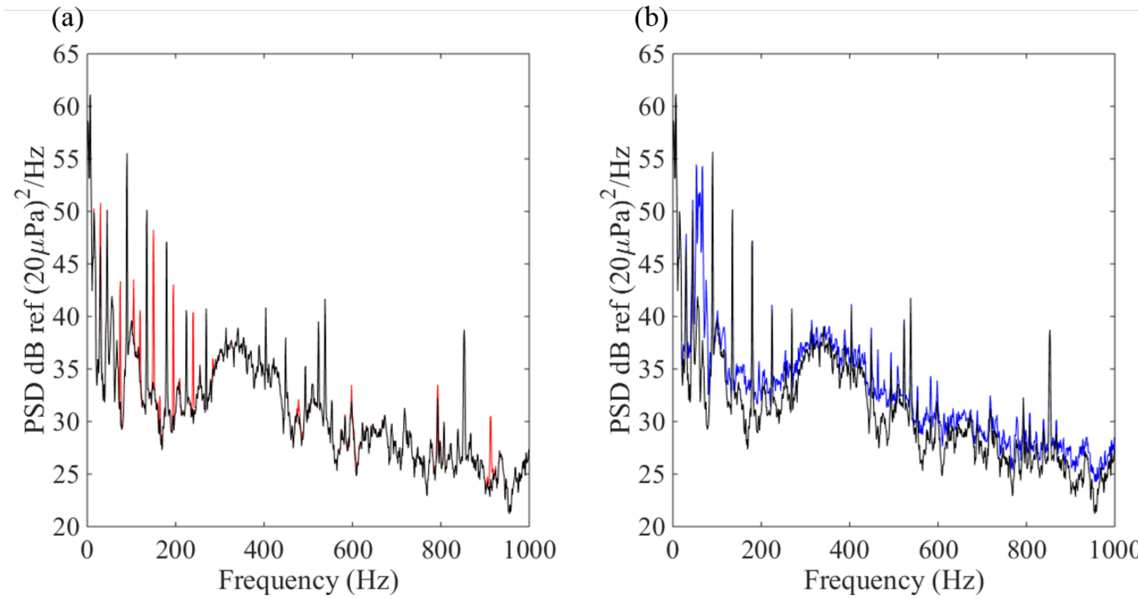


Figure 3.9. Power spectral densities of original signal (black line) and the modified signal, where the modification increased amplitude modulation of (a) the tonal components (red) and (b) the broadband components (blue). With reference to Equations (3.6) and (3.7), γ_1=0.9 and γ_2=1.0 for both (a) and (b).

modified while keeping Zwicker Loudness exceeded 5% of the time (N_5) unchanged. An example of the spectrum of a Sharpness increased signal is shown in Figure 3.11. The blue line is the recorded original refrigerated truck sound and the red line is the modified signal with an increased sharpness value. There is an attenuation at low frequencies and an amplification of high frequencies. This filter can be applied to the tonal and broadband components separately.

3.3.4 Sound Metric Decorrelation

By using the signal modification techniques described in the previous sub-sections, the recorded sounds were modified to expand the set of sounds. The correlation between the main sound quality metrics is illustrated in Figure 3.12. The blue filled circles denote the recorded sounds and the red 'x' denote the modified sounds. The R^2 values

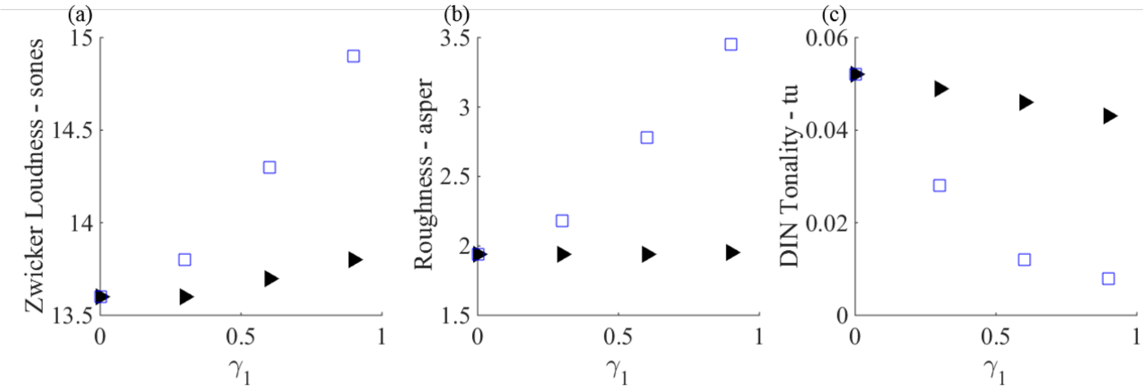


Figure 3.10. (a) Zwicker Loudness exceeded 5% of the time, (b) Roughness exceeded 5% of the time, and (c) DIN Tonality exceeded 5% of the time metric values, each plotted against the parameter γ_1 that controls the variation of modulation frequency and fixed γ_2 (1.0) with reference to Equation (3.6) and (3.7). Symbols: adding modulation to tonal components (▶) and broadband component (□).

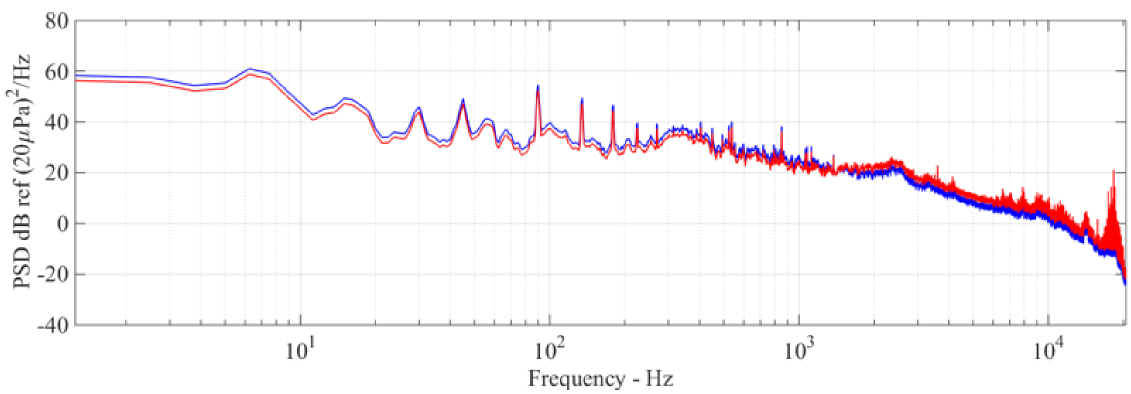


Figure 3.11. Power spectral densities of original signals (blue line) and the Sharpness modified signals (red line). N_5 = 13.6 sones, S_{VB5} =1.32 acum (blue), S_{A5} =1.47 acum (red), S_{A5} =2.19 acum (blue), and S_{A5} =2.49 acum (red).

between Zwicker Loudness values (N_5) and these other 4 sound quality metrics (R_5, T_5, S_{VB5} and FS_5) are 0.72, 0.22, 0.42, and 0.16, respectively for the original sounds and are 0.51, 0.01, 0.05, and 0.04, respectively after adding the modified signals to the set. Since Zwicker's roughness model is calculated based on the modulation depth

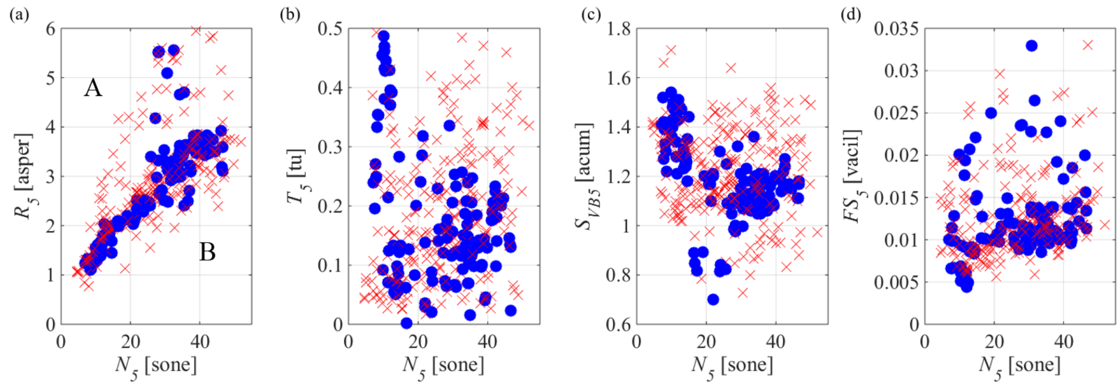


Figure 3.12. Variations in sound quality metric values plotted against N_5: (a) Roughness, (b) DIN Tonality, (c) Von Bismark Sharpness, and (d) Fluctuation Strength plotted against variations in Zwicker Loudness exceeded 5% of the time. This is for a set of original recordings (•) and modified recordings (x). A and B are 'gap' regions.

of the specific loudness spectrum through time, and the correlation between N_5 and R_5 for the measured sounds is quite high (R^2=0.72). By adding modulation to the tonal or the broadband components, Roughness was increased, which reduced this correlation. Even lower correlation can be achieved by adding modified sounds in regions A and B (Figure 3.12 (a)), but sounds generated in those regions did not sound realistic for this application (HVAC&R equipment sounds). The sounds used in the subjective tests described in the following Chapters were selected to have the lowest possible correlations of all the metric combinations for the set of signals chosen. This not only exposes the strength of various sound characteristics to the subject, but also helps in the estimate of a more robust annoyance prediction model.

3.4 Overview of Test 1 – 3

In the next three chapters the three subjective tests will be described. The signals used in each test were chosen to cover a range of signal characteristics, while de-correlating sound quality metrics as much as possible for the set of sounds chosen. As noted earlier, the sounds should be realistic, which sometime limits the range of some of the metrics.

In Table 3.1 are the numbers and demographics of subjects who participated in each test, and the number of sounds used. The number of sounds from refrigerated trucks or residential units, and whether they are modified or original recordings is also indicated. A detailed description of the test signals used in the three tests is given in Appendix C. The types of tests are also indicated in the Table.

Prior to conducting the test, Purdue Institutional Review Board reviewed and approved conducting of the tests (IRB number 150701634). Subjects were recruited by placing advertisements on notice boards in buildings on the Purdue campus and in the Purdue Today electronic newsletter, which is sent to all Purdue students, staff and faculty.

Prior to the subject arriving at the test site (Ray W. Herrick Labs.), the equipment was calibrated to ensure A-weighted sound pressure levels were within $\pm$ 0.4 dB of expected values. The test sounds are .wav files stored in a directory on the test computer. They are played to the subject over earphones, while the subject is sitting in front of a screen with a key board and a mouse in an IAC Double Walled Sound Booth. The sounds on the computer were played through a LynxOne soundcard connected to a Furman SP-20AB stereo amplified which drives the Etymotics ER-2 research earphones.

After subjects arrives, they are given a description of the different parts of the test session, i.e., given a testing session overview. Then they are given consent form to read, and if they wish to continue to take the test, they sign it. It is co-signed by the researcher conducting the test. An outline of all the parts of the test session is provided. This consists of the following.

- <u>A Hearing Test</u>

  A hearing test was performed to ensure that all subjects' hearing thresholds were no greater than 20 dB above threshold in the octave bands from 125 to 8000 Hz. If a subject's hearing threshold was higher than 20 dB, then the subject was compensated \$5 and did not proceed to the test. Subjects who passed the hearing test then proceeded to the next parts of the test which involve listening to sounds without rating them (familiarization).

- A Sound Familiarization Session

  In a familiarization part, the subject listened to sounds that researchers felt covered the range of characteristics of sounds in each test. Subjects did not do ratings in this part.

- Practice Test and Scenario

  In a practice test, subjects rated sounds and got used to using the scale and the hardware/software to enter their ratings. The following scenario was given to the subjects to read: "While you are listening, it may be helpful to imagine yourself in your garden, at any time during the day or evening, hearing these sounds continuously". Once the subject had completed the practice test and the researcher had answered any questions from the subject, the subject proceeded to take the main part of the test.

- Main Parts of the Test

  This may include additional familiarization and rating sessions, when there are several parts to the test. Orders of presentation of sounds or pairs of sounds were randomized with a different random order for each subject.

- Comment Sheet

  Subjects were asked to make comments about the test, the sounds, recommendations for changes in procedures, and intention of participating the future study.

- A 2nd Hearing Test

  This was conducted to check that hearing threshold levels had not changed.

- Payment

  The subject was compensated $10 for participating in the test.

After the subject left, the equipment was re-calibrated to ensure that sounds are still within $\pm$ 0.4 dB of the expected level.

3.5 Summary of Chapter 3

In this chapter the preparations for the subjective tests have been described. This included measurements, signal modification procedures, and decorrelation of sound quality metrics. In addition, a general description of a typical test session was given. The details of Test 1 to 3 are given in Chapters 4 to 6, respectively.

Table 3.1. Summary of signals and subjects. Appendix C contains a list of all the signals used in test and their associated metric values. Appendix E contains the results of all test.

<table>
<tr><th></th><th>Test Type</th><th>Signals</th><th>Subjects</th></tr>
<tr><td rowspan="4">Test 1</td><td>Part A
- Description</td><td>24 Sounds
- 11 original, 13 modified
- 14 residential, 10 refrigerated truck</td><td rowspan="2">42 subjects
- Average age: 26.8 yrs. (18-57)
- 23 males, 19 females
- 27 U.S., 14 Asia
- S.Q. test experienced: 3</td></tr>
<tr><td>Part B
- Rating</td><td>36 Sounds
- 12 original, 24 modified
- 22 residential, 14 refrigerated truck</td></tr>
<tr><td>Part C
- Description</td><td>4 Sounds
- 2 original, 2 modified
- 2 residential, 2 refrigerated truck</td><td rowspan="2">33 subjects
- Average age: 31.1 yrs. (22-62)
- 14 males, 19 females
- 16 U.S., 10 Asia, 7 South America
- S.Q. test experienced: 4</td></tr>
<tr><td>Part D
- Paired
Comparison</td><td>40 sounds
- 8 original, 32 modified
- 20 residential, 20 refrigerated truck</td></tr>
<tr><td>Test 2</td><td>Semantic
Differential
Test</td><td>22 Sounds
- 11 original, 11 modified
- 11 residential, 11 refrigerated truck</td><td>39 subjects
- Average age: 27.2 yrs. (19-51)
- 22 males, 17 females
- 21 U.S., 15 Asia, 3 South America,
- S.Q. test experienced: 4
- Musical event: 6</td></tr>
<tr><td rowspan="4">Test 3</td><td>Overall</td><td>120 Sounds
- 59 original, 61 modified
- 59 residential, 61 refrigerated truck
- 15 common sounds for
Part A, Part B and Part C</td><td rowspan="4">60 subjects
- Average age: 28.4 yrs. (18-62)
- 30 males, 30 females
- 32 U.S., 25 Asia, 1 South America,
2 Africa
- S.Q. test experienced: 5
- Musical event: 3</td></tr>
<tr><td>Part A
- Rating
(Quieter)</td><td>50 Sounds
- 28 original, 22 modified
- 38 residential, 12 refrigerated truck</td></tr>
<tr><td>Part B
- Rating
(Louder)</td><td>50 Sounds
- 30 original, 20 modified
- 11 residential, 39 refrigerated truck</td></tr>
<tr><td>Part C
- Rating
(Wider
Loudness
Range)</td><td>50 Sounds
- 19 original, 31 modified
- 26 residential, 24 refrigerated truck</td></tr>
</table>

4. TEST 1: WHAT DO PEOPLE NOTICE IN HVAC&R SOUNDS?

The main part of Test 1 consists of a description test (Part A) and a rating test (Part B). Part A focused on gathering subjects' descriptions of HVAC&R sounds. In Part B, subjects rated annoyance of the sounds from HVAC&R equipment sounds. Both original recordings and modified recordings were used in the test. A follow-up test was conducted. It also had two parts: a paired comparison test (Part C) and a description test (Part D). These additional tests were conducted to investigate perception of sounds with interesting characteristics not quantified by commonly used sound quality metrics and to study the tonality and roughness perception.

4.1 Signals

An overview of the group of sounds used in Test 1 is given in Table 3.1. Twenty-four sounds were used in Part A, 11 of which were recorded sounds and 13 were modified sounds. Of the 13 modified sounds, 5 varied the level of the sound, 3 added modulation to the sound, 3 changed the DIN Tonality value, and 2 changed both Roughness and the DIN Tonality values. Twenty-four of Part B sounds are those used in Part A and one of the remaining 12 sounds is a recorded sound while 11 are modified sounds. Of the 11 modified sounds, 3 changed the level of the sound, 4 changed the Roughness value, and 4 changed both DIN Tonality and the Roughness values. Two of the four sounds in Part C were the same as those used in Part B, and the other two sounds had similar levels of Zwicker Loudness (N_5) as the first two sounds, but there were sounds that had received lower annoyance ratings. Table 4.5 and 4.6 contain descriptions of the sounds used in Part D in detail. Two residential and two refrigerated truck sounds were chosen, then DIN Tonality and Roughness values of each sound were varied while the Zwicker Loudness values were constant. A total of 40 sounds were used in Part D.

4.2 Subjects

Forty-two people participated in Parts A and B (average age = 27, median age = 25), and 33 participated in Parts C and D (average age = 31, median age = 32), the follow-up test. Eight of the 33 subjects who took the follow-up test also participated in Parts A and B. Most of the subjects were Purdue students and staff. The demographics of the subjects who participated in Test 1 are given in Table 3.1.

4.3 Test Parts

Parts A through D of Test 1 are described in the following subsections.

4.3.1 Description Test (Part A)

Prior to the sound description test, the subjects were told that they would hear a sound twice and then they should write down descriptions of the sounds in their own words. Subjects listened to 5 sounds in the familiarization session. The subject was given a list of 264 words to give an idea of the variety of words or phrases that could be used, but they were also told they could use other words if they wished. This list was created by a sound quality working group and from words used in sound quality tests published in the literature covering a variety of sound sources including HVAC&R equipment.

<u>The list of descriptive words or phrases provided was:</u>

Afar, abrupt bang, bark, bawl, bay, belling, bellow, blare, blatter, bleat, bong, boom, bowwow, brawl, bray, brushing, burning, buff, buzz, brief, burst, bouncing, beat cackle, caterwaul, caw, chafing, chatter, cheep, cheer, chirp, chirrup, chuck, chuckle, clack, clang, clank, clap, clash, clatter, click, clink, cluck, clunk, coarse, coo, crack, crackle, creak, croak, crow, crunch, cry, cuckoo, can dropped, complicated, crinkle drone, drumming, dropping, door opening (closing, shut), dull, distant, deep, dark echo fizz, fizzle, flutter, fritiniancy, falling object, flat, flexible, familiar, full gaggle, giggle, gobble, grate, grating, grinding, groan, growl, gruff, grum, grumble, grunt, gruntle, guffaw, guggle, gurgle, glass, gun fire halloa, halloo, harsh, hiss, hoarse, hollow, hoop, hoot, horrisonous, horse, howl, howl, high-pitched, howl, low-pitched, hum, heavy object, hitting the floor, high frequency, hard, high insect cry, itch jangle, jar, jingle knock latration, laugh, low, loud, long, light, low pitch, low frequency meow, mew, mewl, moan, moo, metallic, musical, medium pitched, muffled neigh, noisy, near oil canning, ooh-tone patter, peep, ping, pipe, pop, pounding, pule, purr, plastic container, paper on table, paper in it quack, quick rap, ratting, rattle, rebellow, reboation, ring, roar, rough, rumble, rustle, rapid, repetitive, resounding, rigid, rolling off, reverb, resonant screak, scream, screech, screech owl, scrub, sepulchral, shout, shriek, shrill, sizzle, slap, snap, snarl, sneeze, snigger, snore, snort, snuffle, squall, squash, squawk, squeak, squeal, stridulous, swish, swoosh, slow, short, slam, smooth, soft, simple, strange, sustained, sharp, stompting, scratched tapping, thrumming, thud, thump, tick, ting, tinkle, tittler, troat, twang, twirling, twitter, thunder ululation, unintelligible vibratory wheeze, whine, whirl, whirring, whistle, whiz, whoop, woodnote, whip, wiggle yap, yarr, yaup, yawl, yell, yelp

The list of words was removed before the subject started the practice test (2 sounds). Each sound was played for four seconds and repeated once after a four second pause. Subjects wrote down their responses (words, phrases, or sentences) on paper using a provided pencil as they listened to the sounds. After the practice test, the subject started the actual test (written descriptions of 24 sounds).

4.3.2 Rating Test (Part B)

In Test 1 Part B, each sound was played for four seconds. The scale that appeared on a computer screen is shown in Figure 4.1 and subjects moved a marker along the scale by using the mouse. The scale was marked at five equally-spaced pointed

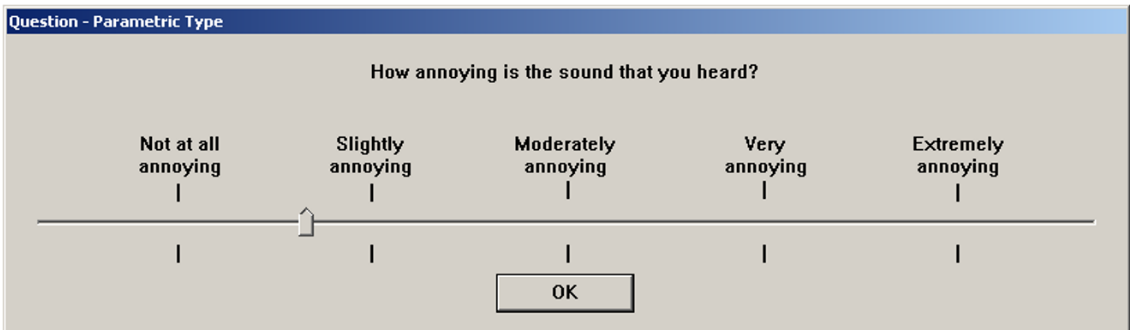


Figure 4.1. Annoyance scale and graphical user interface used in the test.

labeled: “Not at all annoying”, “Slightly annoying”, “Moderately annoying”, “Very annoying” and “Extremely annoying”. The ends of the line extended beyond first and the last markers to help avoid saturation effects. Subjects could rate at any point on the line. When the subject was satisfied with the rating, they clicked the “Next Sound” button on the screen. In the analysis, the five points marked on the scale were assigned numbers 2, 3.5, 5, 6.5 and 8, and the ends of the line were assigned 1 and 9. The order of the sounds was randomized for each subject. Subjects rated 3 sounds as a practice test prior to the actual rating test.

4.3.3 Description Test (Part C)

The two prominent outlier sounds in the linear regression analysis of Part B results were not included in the description test (Part A). Thus, an additional description test was conducted with different subjects. The two outlier sounds and two comparison sounds (similar loudness to the outlier sounds but with lower annoyance ratings) were included in the additional description test (Part C). The results from the descriptions of these 4 additional sounds are included in Sections 4.5 and 4.6.

4.3.4 Paired Comparison Test (Part D)

Eight sets of paired comparison tests were conducted to observe how annoyance ratings vary when there are changes in the DIN Tonality and Roughness values. In the paired comparison tests, 2 sounds were selected from loud region (signals used in Part A) and quiet region (4 in total). For each selected sound, DIN Tonality and Roughness values were varied and Loudness was kept constant. For quiet sounds N_5 = 19.4 ± 0.6 and 20.3 ± 0.3 sones and for loud sounds N_5 = 42.3 ± 0.1 sones. For each paired comparison test, 1 original sound and 4 modified sounds were included. In Figure 4.2, the structure of Test 1 Part C and the different test sections are shown. Quiet tests (4 sets) were conducted prior to loud tests (4 sets). Before conducting the quieter region actual test (Sets 1 – 4), 8 quieter sounds were played for the familiarization and 6 pairs of quieter sounds were presented in a practice tests. After the quieter region paired comparison test, familiarization and practice test were repeated with the louder sounds, then the louder region actual test (Sets 5 – 8) were conducted. To reduce ordering effects, the order of the 4 sets was randomized for each subject, and the order of the pairs of sounds presented within each set were also randomized with a different random order for each subject. The subject was asked to choose which sound was more annoying. The graphical used interface for the paired comparison tests is shown in Figure 4.3. In Table 4.1 and 4.2, signal information and sound quality metric values of the quieter and the louder region paired comparison test signals are given, respectively.

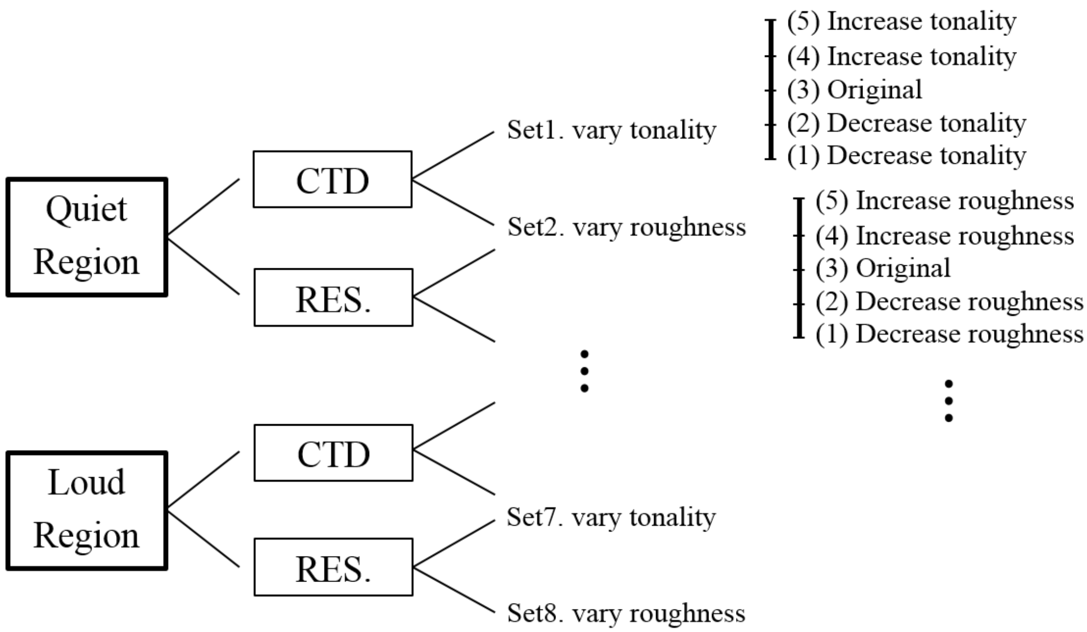


Figure 4.2. The sets of paired comparison test conducted in Test 1 Part D. Eight tests were conducted in total. CTD denotes refrigerated truck sound, and RES. denotes residential unit sounds.

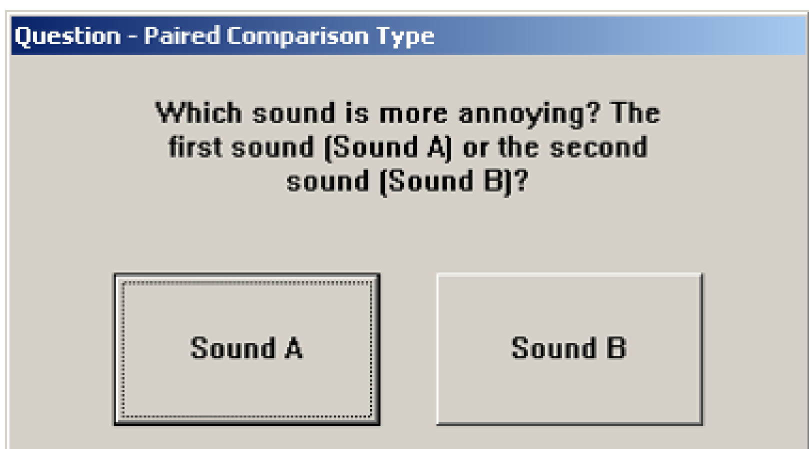


Figure 4.3. Graphical user interface used in the paired comparison test

4.4 Results of Part A and C, Sound Description

The most common descriptors used by the subjects are given in the second column of Table 4.3, where the number in parentheses indicates the number of times that

Table 4.1. Sound quality metric values of the quieter region paired comparison test sounds.

	Signal	Modification	N_5	T_5	R_5
Set 1	Refrigerated Truck (Vector 1350) Diesel Engine 2250 rpm	More Tonal 2	19.4	0.34	2.48
		More Tonal 1	19.4	0.21	2.49
		Original	19.4	0.13	2.47
		Less Tonal 1	19.4	0.12	2.40
		Less Tonal 2	19.4	0.10	2.45
Set 2		More Rough 3	18.8	0.13	4.24
		More Rough 2	19.8	0.13	3.83
		More Rough 1	19.4	0.13	2.71
		Original	19.4	0.13	2.47
		Less Rough 1	19.5	0.13	2.21
Set 3	Residential Fan 1100rpm Compressor 3600rpm	More Tonal 2	20.7	0.40	2.01
		More Tonal 1	20.5	0.33	1.98
		Original	20.8	0.24	2.03
		Less Tonal 1	20.7	0.18	2.02
		Less Tonal 2	20.5	0.16	2.15
Set 4		More Rough 4	20.7	0.24	3.18
		More Rough 3	20.8	0.24	2.90
		More Rough 2	20.7	0.24	2.50
		More Rough 1	20.8	0.24	2.26
		Original	20.8	0.24	2.03

the word was used by the subjects. Six of categories, among 9, “Soft – Loud”, “Tonal”, “Dull – Sharp”, “Smooth – Rough”, “Fluctuating”, and “Impulsive”, are related to sound characteristics. The “Pleasant – Annoying” category is related to

Table 4.2. Sound quality metric values of the louder region paired comparison test sounds.

	Signal	Modification	N_5	T_5	R_5
Set 5	Refrigerated Truck (Vector 1950) Diesel Engine 1850 rpm	More Tonal 2	42.2	0.30	3.65
		More Tonal 1	42.3	0.24	3.65
		Original	42.3	0.21	3.69
		Less Tonal 1	42.3	0.18	3.69
		Less Tonal 2	42.3	0.12	3.69
Set 6		More Rough 3	42.3	0.21	5.54
		More Rough 2	42.3	0.21	5.03
		More Rough 1	42.2	0.21	4.38
		Original	42.3	0.21	3.69
		Less Rough 1	42.3	0.21	3.40
Set 7	Residential Fan 452 rpm Compressor 1800 rpm	More Tonal 2	42.3	0.26	2.79
		More Tonal 1	42.3	0.21	2.81
		Original	42.3	0.19	2.81
		Less Tonal 1	42.3	0.16	2.83
		Less Tonal 2	42.3	0.1	2.81
Set 8		More Rough 3	42.3	0.19	3.80
		More Rough 2	42.3	0.19	3.45
		More Rough 1	42.3	0.19	3.06
		Original	42.3	0.19	2.81
		Less Rough 1	42.3	0.19	2.57

a summative judgment of the sound. The two other categories were named "Emotional Response" and "Functionality". The words and phrases were grouped and classified by the author. The opposites, e.g., "soft" and "loud" were placed in the

Table 4.3. Examples of the words collected and the classifications.

Classification	**Descriptors (number of times used)**
Soft/ Loud	Soft (56), Quiet (29), Muffled (16), Mild (10), Faint (7), Gentle (3) Medium,(19), Moderate (17) Loud (210), Powerful (11), Intense (9), Strong (5), Vigorous (2), Not Soft (3)
Not Tonal/ Tonal	Low (252), Low Frequency (12) Medium Frequency (10), High Pitch (54), Hum (43) High Frequency (17), High (17), Heavy Tone (6), Prominent (3)
Dull / Sharp	High Frequency (17), Dull (3) / Metallic (21), Scratching (14), Sawing (12), Sharp (11), Heavy Tone (6), Squeal (6)
Smooth/ Rough	Smooth (26), Even (5), Not Harsh (2) Whirling (25), Buzz (24), Harsh (23), Rough (15), Grinding (17), Rumble (16)
Fluctuating/ Not Fluctuating	Vibration (67), Pulsating (7), Uneven (6), Shaking (5), Echo (3), Beating (2), Oscillating (2)/ Constant (7), Even (5)
Impulsive	Drill (42), Choppy (25), Rattle (16), Repetitive (12), Drumming (6), Thudding (6), Thumping (4)
Pleasant/ Annoying	Pleasant (4), Not Irritating (7), Not Annoying (3) / Annoying (86), Irritating (26), Noisy (19), Disturbing (18)
Emotional Response	Calm (16), Relaxing (5) / Hurt Ears (12), Scary (6), Headache (5), Painful (4)
Functionality	Safe (7), Efficient (4), High Performance (3), Properly Working / Old (15), Broken (4), Rusty (4), Ineffective (3), Dangerous (3), Unsafe (2)

Table 4.4. Descriptors using the name of an other machine.

	Descriptors
Residential Sounds	Air-conditioner, fan, drone, spinning machine, vacuum, dryer, refrigerant, compressor
Refrigerated Truck Sounds	Helicoptor, drill, generator, engine, motorcycle, propeller, jack hammer, machine gun

same group. These classifications were reviewed by the research team and modified. Most descriptors were clear to classify, but some words and phrases, e.g. 'harsh' and 'high frequency', fit into more than one group. "Jack Hammer" is an also example of a descriptor that can fall into two categories: "Soft – Loud" and "Impulsive". The complete list of subjects' description of sounds are in Appendix E, Table E.1. Even though subjects did not study sound quality or related fields, sound attributes such as level, pitch, slow/fast fluctuation and spectral balance were described. Participants also used descriptions related to feelings or machine malfunctions. Sometimes people used other machines to describe the sound such as "Jack Hammer", "Helicopter", and "Vacuum", or used words to express the sound, e.g., "wo sound".

Expressions related to tonalness were often given for recorded and modified residential unit sounds, whereas expressions related to roughness and impulsiveness were often given to modified and recorded refrigerated truck HVAC&R sounds. Table 4.4 shows the descriptors that involve other machines from the recorded residential and the recorded refrigerated truck HVAC&R unit sounds, respectively. The expressions from the subjects for the two units are completely different. For the residential sounds, participants mainly used descriptors related to fans or rotating machines. For the refrigerated truck sounds, subjects described machines such as engines or motors that had repetitive, impulsive, and pounding characteristics. From these results, it is concluded that participants perceive the sounds from the two types of units as very differently.

4.5 Analysis and Discussion of Parts A and C

The anlysis of the word data followed a procedure used by Bi, Reid and Davies in [83]. In each group shown in Table 4.3, words or expressions were classified as"positive" (+) or "negative" (−). Examples of positive words are "quiet", "pleasant", and "working well" and negative words are "loud", "noisy", and "broken". The number of positive and negative descriptors that are in each group were counted. A positive

Table 4.5. Word score classifications indicating mostly negative (−− or -) or mostly positive (++ or +) word scores in each category for the 24 sounds in the word test, along with annoyance rating. NM indicates not mentioned. 0 means the numbers of positive and negative words are the same.

Signal	**(+) Soft/ Loud (-)**	**(+) Not Tonal/ Tonal (-)**	**(+) Dull/ Sharp (-)**	**(+) Smooth/ Rough (-)**	**Fluctuating (-)**	**Impulsive (-)**	**(+) Pleasant/ Annoying (-)**	**Annoyance Rating Test B [Bold >"Very Annoying"]**	**(+) Emotional Response (-)**	**(+) Functionality (-)**
2	−	−	NM	−	−−	−−	−	5.72	NM	NM
3	−	−	−	−	−	−−	−	5.70	−	NM
4	−	−	−	−	−	−−	−	5.29	NM	NM
5	−	−	−	−	−	−−	−	5.96	NM	NM
7	−	−−	−	−−	−	−	−	**6.65**	NM	−
8	−−	+	−−	−−	−	−	−	**6.92**	NM	−
9	−−	−	−	−−	−	−	−−	**7.19**	−	−
10	++	−	−	−	−	−	+	3.52	+	+
11	++	−	+	+	−	−	+	2.64	++	NM
13	++	NM	+	+	NM	NM	+	2.20	++	+
17	−−	−	−	−−	−−	−−	−−	**6.59**	−	NM
18	−	+	NM	−	−	−−	−	5.84	−	NM
19	−−	−	−	−−	−	−	−−	**7.47**	−	−
20	+	−	−	+	−	+	−	5.50	−	NM
22	−−	−−	−−	−	−	−	−−	**7.44**	−	NM
23	+	−	0	−	−	−−	−	4.02	−	NM
25	−−	−	−−	−−	−	−	−−	**7.90**	−−	−
26	+	−	NM	−	−	NM	−	5.06	0	+
28	−−	−	−	−	−	NM	−−	**7.37**	−	NM
30	++	−	+	−	0	NM	+	3.92	+	−
31	+	−	NM	+	−	NM	+	3.84	+	−
32	−−	−−	−−	+	−	NM	−	5.99	−	−
33	−−	−−	NM	−	−	−	−	6.02	NM	NM
35	++	−−	+	−	−−	−	+	3.48	+	+

descriptor was given +1 point and a negative descriptor was given -1 point. Then, the word score for each signal in each group was calculated. In Table 4.5, word scores are indicated as "++", "+", "0", "−", "−−" or "NM". "++" and "+" means more positive descriptors than negative descriptors, and vice versa for "−−" and "−". If descriptors in a category are not mentioned in that particular signal, "NM" is assigned. If the number of positive words are more than twice the negative words, "++" sign was used. If the number of positive words are more than negative words but less than twice the negative words, "+" sign was used. Similarly, that is how "−−" and "−" signs were determined. The results are shown in Table 4.5. Signals 10, 11 and 12 are described as very soft, mostly dull, smooth and pleasant, whereas signals 7, 8, 9 and 25 are mostly given negative descriptions across all classification.

4.6 Test 1 Part B Results: Annoyance Ratings

The annoyance ratings for the signals and the average of the annoyance ratings (and the corresponding standard error) for the signals are shown in Figure 4.4. The average annoyance ratings for the 24 sounds that were also included in the description test are given in Table 4.5. The standard deviations of the estimated means (standard error) are less than 0.3 (3.75% of entire scale). As expected there is a wide distribution of responses. Each subject's responses is denoted by a different symbol. for example the subject whose responses are denoted by the blue triangle symbol tended to have higher annoyance scores, whereas the subject whose response are denoted by the purple filled triangle tended to give lower annoyance ratings. There are significant differences in the average annoyance ratings with, e.g. sounds 11, 13 and 21 having average ratings close to 2 (not at all annoying) and sounds (9, 14, 19 and 25) having average ratings between 6.5 (very annoying) and 8 (extremely annoying) It is observed that signals with a prevalence of negative descriptors (− or −−) throughout the 9 groups in Table 4.5, received higher annoyance ratings (see Figure 4.4 and columns 8 and 9 of Table 4.5). For example, the results for signals 9, 19, 22, 25 and 28 in Table 4.5 and Figure 4.4. In contrast, signals that were described with more positive descriptors were given relatively lower annoyance ratings. For example, see results for signals 10, 11 and 13 in Table 4.5 and in Figure 4.4. This shows that results from the description test (Part A) and the rating test (Part B), despite being two different types of tests, are providing consistent results.

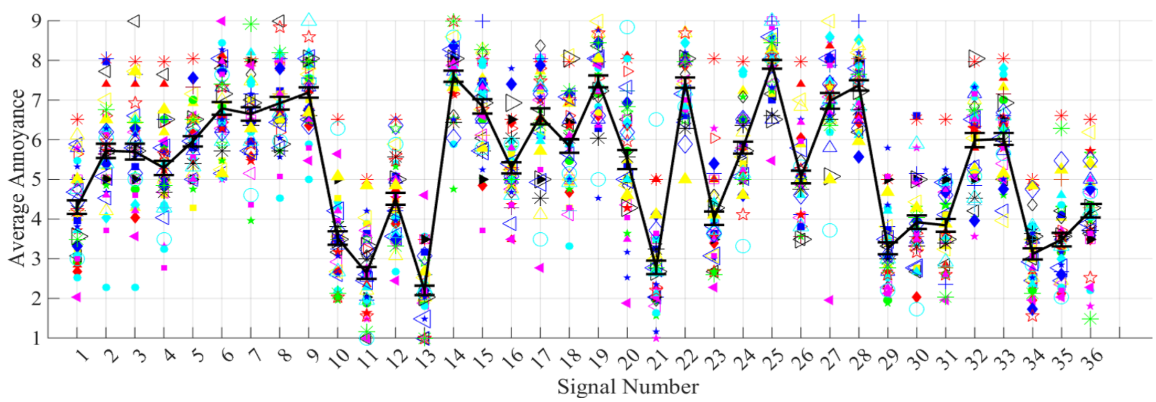


Figure 4.4. Subjects' annoyance ratings, and the average annoyance ratings for all sounds (black line joins average of annoyance ratings). Bars show the standard deviation of the estimated mean. Each type of marker is associated with a particular subject's response. Numbers correspond to: 2= "Not at All Annoying", 3.5="Slightly Annoying", 5="Moderately Annoying", 6.5="Very Annoying", and 8="Extremely Annoying".

4.7 Analysis of Test 1 Part B Results

Average annoyance prediction models were generated by using linear regression as a preliminary analysis. These results are limited by the number and variety of sounds used in this test. The results for two single-metric models (N_5 and SQI^*) are shown in Figure 4.5. The most accurate single-metric predictor is N_5 (R^2=0.93, p <0.001), $A-SPL_{Sones}$ and $A-SPL$ are also highly correlated with average annoyance ratings with R^2 =0.92 (p <0.001) and 0.83 (p <0.001), respectively. The correlation between the Roughness metric (R_5) and the average annoyance ratings is relatively low (R^2 = 0.30, p <0.001) but significant. Surprisingly, because tonality is expected to be an important contributor to annoyance the correlation between Aures' Tonality and the average annoyance ratings is not significant (R^2 =0.01, p=0.543). Multi-metric models were also examined but only a small improvement was found over the N_5 only model. The R^2 value was increased to 0.95 when R_5, and DIN Tonality (calculated by using Head ArtemiS) were included with N_5 in the model.

There are two prominent outliers in the multi-metric model predictions. These are the same outliers in the single metric model predictions based on N_5 (see the two open blue diamonds in Figure 4.5 far above the trend line). These two signals are Loudness modified residential unit sounds, amplified to lie in the N_5 loudness range of refrigerated truck HVAC&R sounds. Including Aures' (or DIN 45681) Tonality or von Bismarck (or Aures') Sharpness metrics in the annoyance prediction models did not move annoyance prediction for outlier signals, toward the trend line, even though the outlier signals are more sharp and tonal than sounds of similar loudness.

4.8 Results of Part D: the Paired Comparison Tests

Bradley-Terry-Luce (BTL) values were calculated from the estimated probability that one sound would be chosen as being more annoying than another. BTL values are relative annoyance ratings. The BTL values for Part D tests are shown in Figure 4.6.

When sounds become more tonal, relative annoyance rating increased. However,

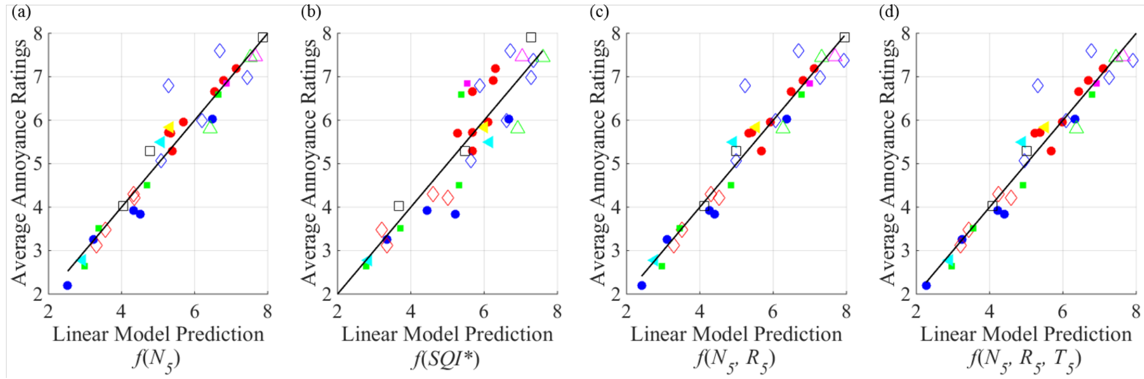


Figure 4.5. Linear regression models of annoyance using metrics (a) N_5, (b) SQI^*, (c) N_5 and R_5 and (d) N_5, R_5 and T_5. Circular markers represent annoyance rating of original recordings, while other markers represent annoyance of modified sounds. Different marker colors represent sounds from various HVAC&R units. Black line represents the best fit line to the data. R^2 values are (a) 0.93, (b) 0.84, (c) 0.94 and (d) 0.95.

DIN Tonality decreased sounds, were not always less annoying. For the group of high Loudness sounds (all sounds with approximately the same N_5 value), annoyance did not change with Roughness. From this test, it was confirmed that subjects' annoyance responses to HVAC&R sounds were varying depending on the degree of tonalness and the degree of roughness, as measured by the two metrics DIN Tonality and Roughness, respectively. In an additional description test, descriptors of the outlier sounds and the comparison sounds were different even though those sounds have the same loudness levels. Descriptors of the outlier sounds were "high pitch", "sharp", "metallic", and related words. However, similar words were not used by subjects when describing the comparison sounds (same Loudness but lower annoyance). This could explain why those outlier sounds were rated as being more annoying in the rating test (Part B).

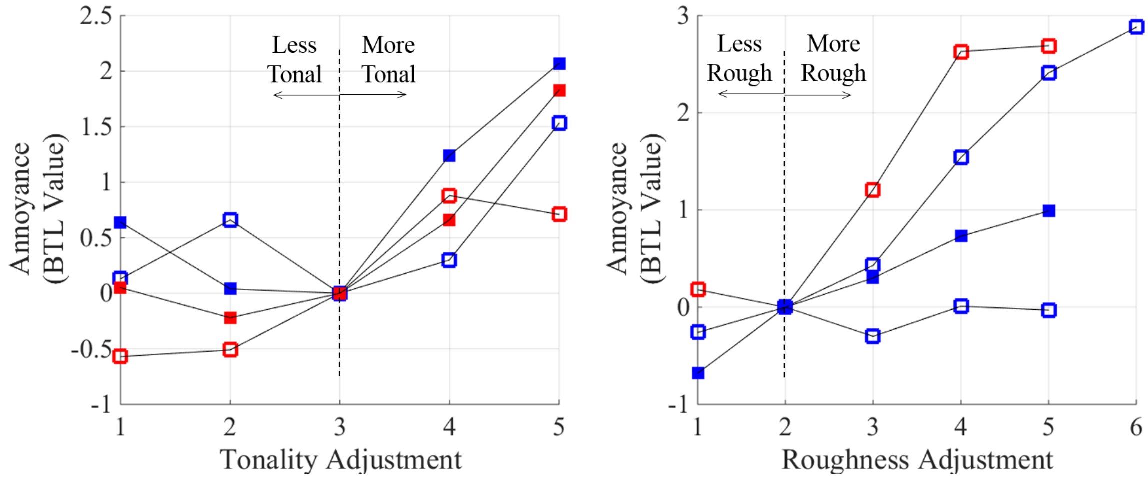


Figure 4.6. BTL values of (a) DIN Tonality adjusted sounds (Set 1, 3, 5, and 7) and (b) Roughness adjusted sounds (Set 2, 4, 6, and 8). The black dotted line indicates the sounds without DIN Tonality or Roughness modification. The markers indicate quieter refrigerated truck sound tests (□), quieter residential sound tests (□), louder refrigerated truck sound tests (■) and louder residential sound tests (■).

4.9 Test 1 Summary

In Test 1, subjects were asked to describe the HVAC&R sounds and to rate them in terms of annoyance. Descriptors from the subjects were categorized into 9 groups, and by assigning a –1 to "bad" descriptors and a +1 to "good" descriptors within the category, a numerical score was derived for each signal in each category. The word scores throughout the categories and the average annoyance ratings were highly correlated. Zwicker Loudness exceeded 5% of the time (N_5) was the metric that was most highly correlated with the average of subjects' annoyance ratings. Other level related metrics such as the Sound Quality Indicator calculated using sound pressure level (SQI^*), A-weighted sound pressure levels, and sound pressure level converted from dB to Sone are also highly correlated with the average annoyance ratings. Adding DIN Tonality, Sharpness, and Roughness metrics to the linear annoyance models did not yield much improvement to the annoyance predictions, even though subjects used words to describe the sounds that are relevant to the characteristics that should be measured by those metrics. It was shown that the Loudness range of test sounds as too broad (40 dB) so that subjects tended to rate annoyance just by the level of a sound rather than by other sound characteristics. An additional test (Part C) was conducted to supplement set of descriptors, with descriptors for the outlier sounds. Paired comparison tests were conducted to see how Tonalness and Roughness variation impact annoyance ratings. The outlier sounds, which were under-predicted by the model developed form responses in Part B, were described by sharpness related expressions. It was observed that subjects' annoyance ratings change when the DIN Tonalness and Roughness metric change.

5. TEST 2: SEMANTIC DIFFERENTIAL TEST

Test 2 was a semantic differential test, which was conducted to determine how many independent factors are contributing to people's judgements of the sounds.Thirty-nine subjects evaluated the 22 sounds on 17 word scales which were developed from the subjects' descriptors in Test 1. The seventeen word scales include "sound attributes", "system characteristics" and "summative judgments" scales. A factor analysis was conducted to determine how many independent factors were present and the nature of those factors, after rotation to align one of the factors with loudness.

5.1 Signals, Rating Scales and Subjects

Thirty-three subjects were asked to evaluate twenty-two sounds on 17 scales by making marks on a evaluation sheet using a provided pencil. The demographics of the subject group are given in the last column of Table 3.1. Most subjects were students or staff from Purdue University.

An example of the evaluation sheet used by subjects is shown in Figure 5.1. The word scales that were used in Test 2 are shown in Table 5.1 and examples of the words provided by subjects in Test 1 are also listed in the table (3rd column). The first 10 word pairs are "sound attribute" related words such as level, fluctuation, tonalness, spectral balance, regularity, and impulsiveness. Then the next 5 scales are related to system characteristics and the last two word scales are the summative judgment of the sounds. Sounds were repeated until the subject completed an evaluation sheet, There was a 3 seconds pause between each repetition. Word scales were positioned from "positive" to "negative" (left to right) such as "soft" to "loud", "smooth" to "rough". Subjects could mark the slash or 'v' mark at any place on the line. When subjects finished the evaluation of all 17 scales for a sound, they clicked the "Next

Sound" button on the screen and started evaluation of the next sound. The order of sounds played and word scales presented for each sound were randomized with a different random order of sounds and a different random order of word scales for each subject. Half of the test sounds were original recordings and half were modified recordings. Half of the signals were from refrigerated truck units and half were residential unit sounds. Table 3.1 contains an overview of the test, and Appendix C contains more details on the signals used in this (and other) test(s).

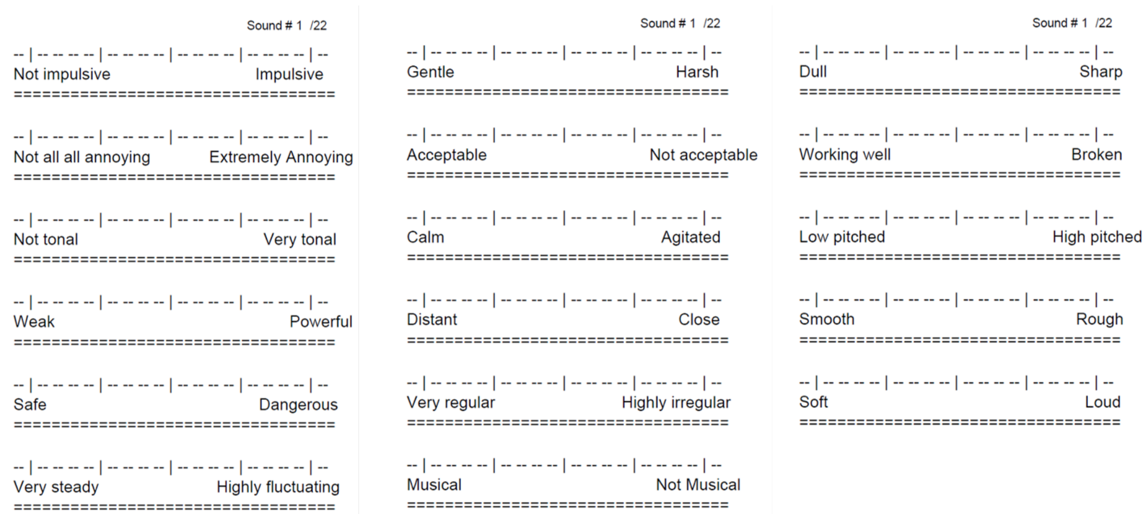


Figure 5.1. An example evaluation sheet used by subjects in the semantic differential test.

In subsequent plots, colors at the blue end of the spectrum are used for residential sounds, and colors at the red end of the spectrum are used for refrigerated truck sounds.

5.2 Results and Discussion

The marks on the scales were translated into numbers by assigning the numbers -10 and +10 to the left and right end points of each scale, respectively. The center mark then corresponded to a number of 0. The average of the 39 subjects' ratings and the

Table 5.1. Words scales used in Test 2 and example descriptors from Test 1.

Scale	Left-End of Scale	Right-End of Scale	Example Descriptors from Test 1 related to Word Scales
1	Soft	Loud	Soft, Quiet, Muffled, Mild / Loud, Intense, Not Soft
2	Low Pitched	High Pitched	Low Pitched / High Pitched, High Frequency
3	Dull	Sharp	Dull / Sharp, Metallic, Sawing, Squeal
4	Smooth	Rough	Smooth, Even / Rough, Grinding
5	Gentle	Harsh	Gentle, Not Harsh / Whirling, Harsh
6	Not Tonal	Very Tonal	Hum, Prominent, Tone
7	Very Steady	Highly Fluctuating	Even, Constant / Vibration, Pulsating, Beating
8	Not Impulsive	Impulsive	Drill, Choppy, Rattle, Repetitive, Drumming
9	Very Regular	Highly Irregular	Regular, Even / Irregular, Uneven
10	Musical	Not Musical	Musical
11	Calm	Agitated	Calm, Relaxing / Agitated, Headache
12	Weak	Powerful	Weak / Powerful, Strong
13	Safe	Dangerous	Safe / Dangerous, Scary
14	Distant	Close	Distant, Far / Close, Loud
15	Working Well	Broken	Properly Working, Efficient / Broken, Dangerous
16	Acceptable	Not Acceptable	Acceptable, Tolerable, Bearable / Not Acceptable
17	Not at all Annoying	Extremely Annoying	Pleasant, Not Irritating, Not Annoying / Annoying, Irritating, Noisy

standard deviation of the estimated mean (standard error) are given in Table E.4 and Table E.5 in Appendix E.

5.2.1 Average Ratings on Each Scale

The average ratings for each of the sounds that were original recordings are shown in Figure 5.2. Two distinguishable patterns are exemplified by the results of two sounds which are marked by thick blue and thick red lines. It was observed that same types of units (refrigerated truck or residential) have similar patterns, but the profile for one of the residential units (light blue line), which is a loud residential unit, tends to follow the trend of the refrigerated truck units on most of the scales. The full rating range was from -10 to 10, but the figure shows the average across all 39 subjects and extends from -8 to 8. The standard deviation of the estimates (standard error) were from 0.30 to 0.95. The standard errors were low on level related scales such as "Soft–Loud", "Weak–Powerful", and "Distant–Close" (0.48-0.52). On the other hand, the standard error was high on the "Not Tonal–Very Tonal" (0.76) scale.

In Figure 5.3, the average ratings of the residential HVAC&R unit sounds (recorded and modified) are shown. There are two strong patterns apparent in the average ratings, and those are emphasized as thick blue and purple lines. Lines that follow the trend of the thick blue lines are original recordings and line that follow the purple line are modified recordings. The standard error for the mean of ratings of residential HVAC&R units are from 0.32 to 0.95. Similar to the original recordings' ratings, the average of ratings on level related scales have lower standard errors and the average of ratings on the tonality scale has the highest standard error. The average ratings of the refrigerated truck HVAC&R unit sounds are shown in Figure 5.4. The pattern of the ratings is similar for all sounds, all following trend of the red line. Interestingly, the average standard errors were low on the "Calm–Agitated" and "Safe–Dangerous" scales (0.53 and 0.47, respectively). Standard errors associated with the average of ratings on those scales for residential sounds were higher (0.63 and 0.65). For many other scales, the average standard errors of refrigerated truck and residential ratings were similar, but for the impulsivness scale, the average standard error of refrigerated unit sounds was lower than that for the residential unit sounds (0.60 and 0.70). It is may be that the impulsive properties of diesel engine sound made the assessment easier. For this set of refrigerated truck sound the standard errors were high for the "Gentle–Harsh" (0.65 to 0.87) and the "Not Tonal–Very Tonal" (0.48 to 0.85) scales.

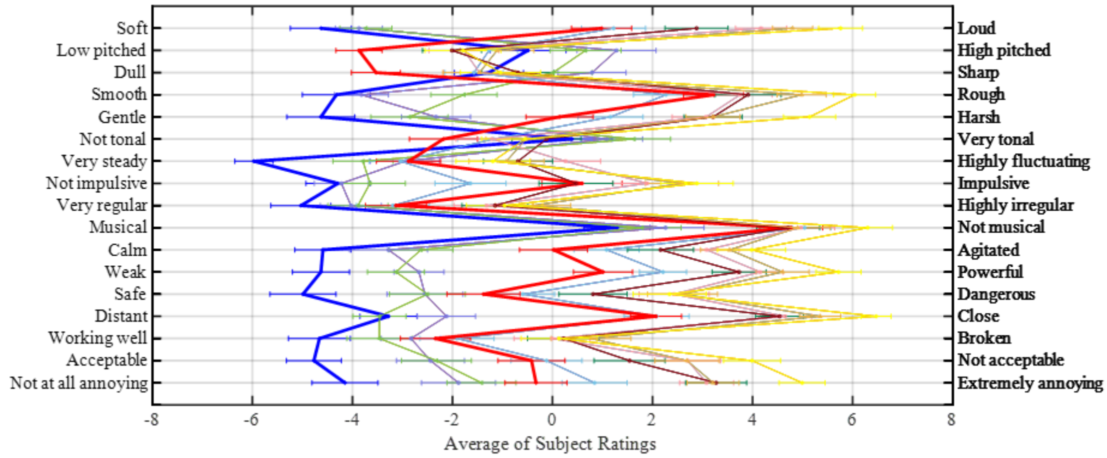


Figure 5.2. The average rating on each scale for the 9 sounds that were original recordings. Residential sounds are indicated by blue to green colors and refrigerated truck unit sounds are indicated by yellow to red colors. Error bars indicate ± the standard error of the estimated mean.

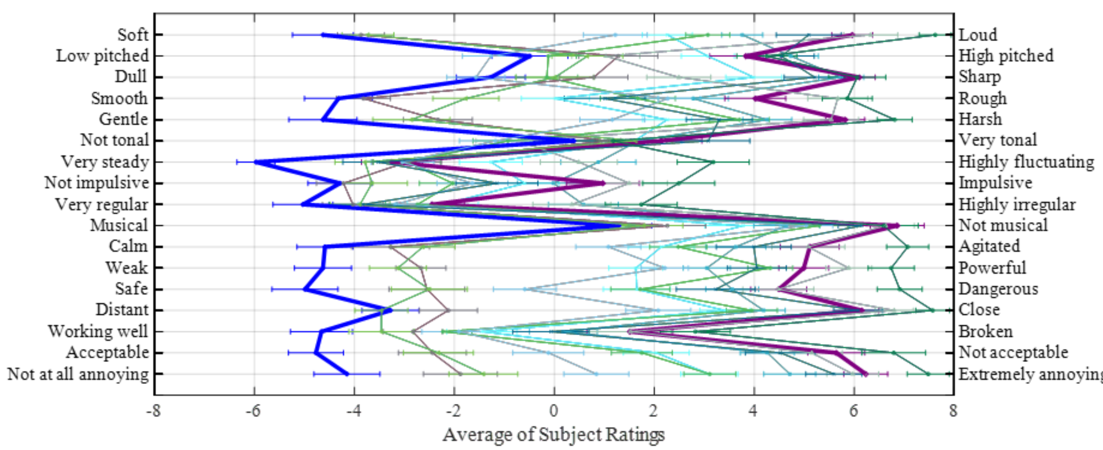


Figure 5.3. The average rating on each scale for the 11 sounds that were from residential HVAC&R units. Error bars indicate ± the standard error of the estimated mean.

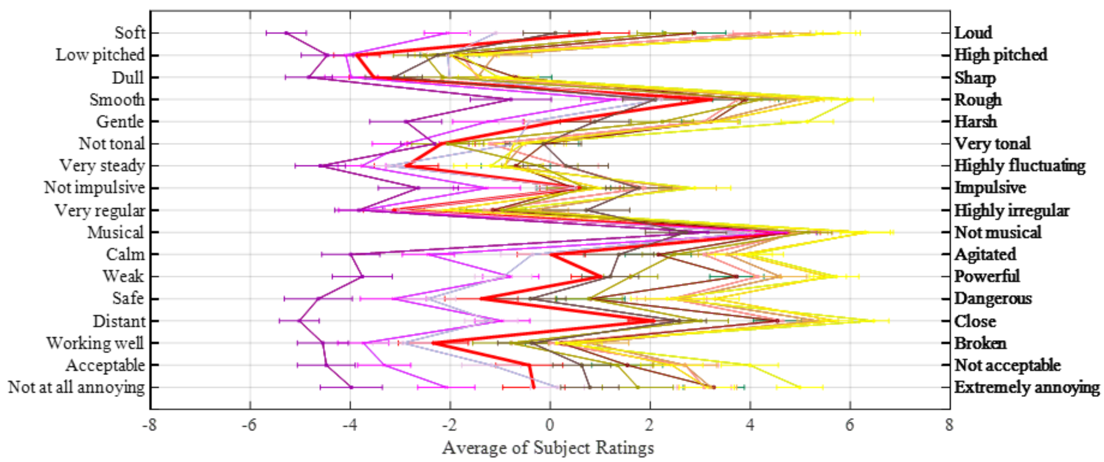


Figure 5.4. The average rating on each scale for the 11 sounds that were from refrigerated truck HVAC&R units. Error bars indicate ± the standard error of the estimated mean.

5.2.2 Correlation Between Sound Metrics and the Average Ratings

The correlation coefficients ($\rho = \sqrt{R^2}$) between the sound quality metrics and the average ratings are given in Table 5.2. In most cases, the highest correlation coefficients for metrics are the word scales that are associated with the property metric of the sound, but sometimes the absolute value of the highest correlation coefficient is quite low. For example, both tonality metrics (T_5 and T_{A5}) values are highly correlated with the average of ratings on the "Low Pitch–High Pitch" and "Not Tonal–Very Tonal" scales. Also, the values of the two sharpness metrics have high correlation coefficient values for average ratings on the "Low Pitch–High Pitch" and "Dull–Sharp" scales. The values of the Fluctuation Strength metric is the most highly correlated metric with average ratings on the "Very Steady–Highly Fluctuating", "Not Impulsive–Impulsive", and "Very Regular–Highly Irregular" scales, but absolute correlation coefficient values are low (0.37, 0.32, and 0.36), respectively. This could be because the model to estimate Fluctuation Strength is not accurate for these fairly complicated sounds; the model was developed for much simpler sounds with repeatable fluctuations. Also, Fluctuation Strength may only measure the strength of one of multiple sound characteristics that are used by subjects to decide on their rating on these scales.

Table 5.2. Correlation coefficients (ρ=$\sqrt{R^2}$) between the sound quality metrics and the average ratings on each scale.

	Correlation Coefficient (ρ)									
Word Scales	N_5	SQI^*	dBA	R_5	T_5	T_{A5}	FS_5	S_{VB5}	S_{A5}	RCL
Soft - Loud	0.95	0.96	0.97	0.57	0.16	-0.21	0.24	0.20	0.79	0.05
Low Pitched - High Pitched	0.34	0.45	0.37	-0.14	0.69	0.55	-0.02	0.82	0.74	-0.26
Dull - Sharp	0.46	0.56	0.51	0.02	0.68	0.45	0.01	0.81	0.82	-0.29
Smooth - Rough	0.81	0.75	0.84	0.79	-0.16	-0.54	0.26	-0.16	0.49	0.17
Gentle - Harsh	0.90	0.92	0.96	0.57	0.16	-0.24	0.20	0.27	0.82	0.01
Not Tonal - Very Tonal	0.21	0.29	0.15	-0.26	0.77	0.79	-0.06	0.62	0.49	-0.20
Very Steady - Highly Fluctuating	0.62	0.61	0.64	0.50	0.10	-0.21	0.37	0.12	0.52	0.37
Not Impulsive - Impulsive	0.72	0.70	0.74	0.69	-0.12	-0.42	0.32	-0.16	0.42	0.41
Very Regular - Highly Irregular	0.60	0.57	0.63	0.53	0.00	-0.29	0.36	0.10	0.51	0.39
Musical - Not Musical	0.85	0.86	0.92	0.63	0.03	-0.38	0.09	0.12	0.68	-0.12
Calm - Agitated	0.92	0.93	0.95	0.58	0.21	-0.17	0.21	0.24	0.80	0.10
Weak - Powerful	0.95	0.95	0.97	0.57	0.08	-0.29	0.26	0.12	0.74	0.01
Safe - Dangerous	0.91	0.91	0.91	0.47	0.30	-0.06	0.22	0.36	0.85	0.00
Distant - Close	0.94	0.97	0.97	0.54	0.16	-0.20	0.27	0.19	0.77	0.04
Working Well - Broken	0.88	0.88	0.88	0.56	0.20	-0.10	0.34	0.24	0.76	0.14
Acceptable - Not Acceptable	0.90	0.92	0.92	0.48	0.29	-0.06	0.19	0.37	0.86	0.03
Not at all Annoying - Extremely Annoying	0.91	0.93	0.93	0.49	0.30	-0.06	0.18	0.36	0.87	-0.01

5.2.3 Factor Analysis

A factor analysis was conducted to find the significant independent factors in the subjects' ratings. A schematic of the factor analysis process is shown in Figure 5.5. The MATLAB function *'factoran'* was used to determine the factors and the weightings on each factor that "explain" ratings on each scale. The orthogonal rotation was used to align factors with the "natural" sound characteristics such as loudness, sharpness, tonality, roughness, and impulsiveness. However, in the results shown below, the orthogonal rotation was to used to align one of the factors with the loudness scale. The results of a four-factor analysis on all 22 sounds (the fifth factor were not significant) are shown in Figure 5.6.

Based on the scales where loadings were high, the factors are named as follows:

"Loudness" (red), "Tonal/Sharpness" (green), "Irregular/Fluctuation" (black), and "Not Musical" (yellow). As noted earlier, subjects appeared to have a difficulty rating on the "Not Tonal–Very Tonal" scale, the standard error was high for the average on this scale. That is maybe why the tonality and sharpness attributes were combined in a one factor ("Tonal/Sharpness") even though those are two different sound characteristics. The weightings for the "Not at all Annoying–Extremely Annoying" scale were the strongest for the "Loudness" and "Tonal/Sharpness" factors, while the weightings on the "Working Well – Broken" scale were the strongest for the

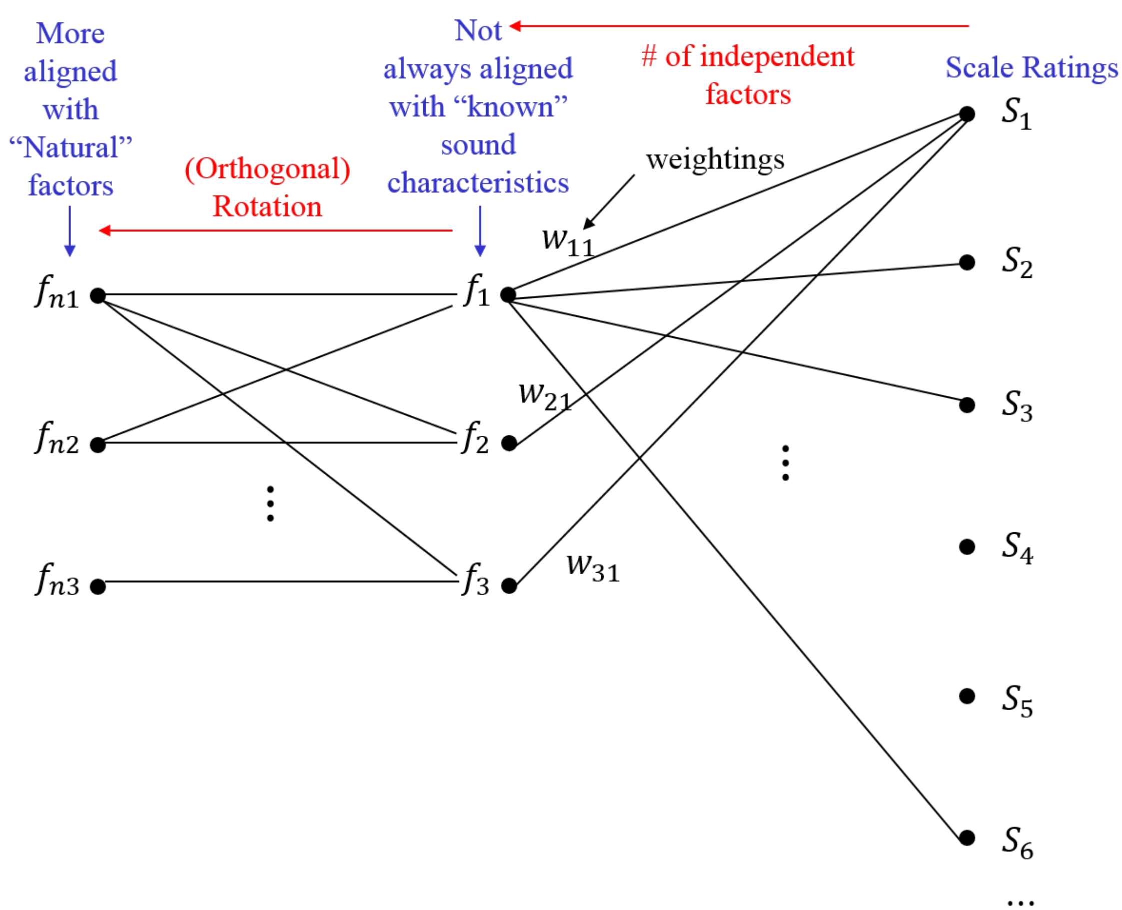


Figure 5.5. Schematic of the factor analysis.

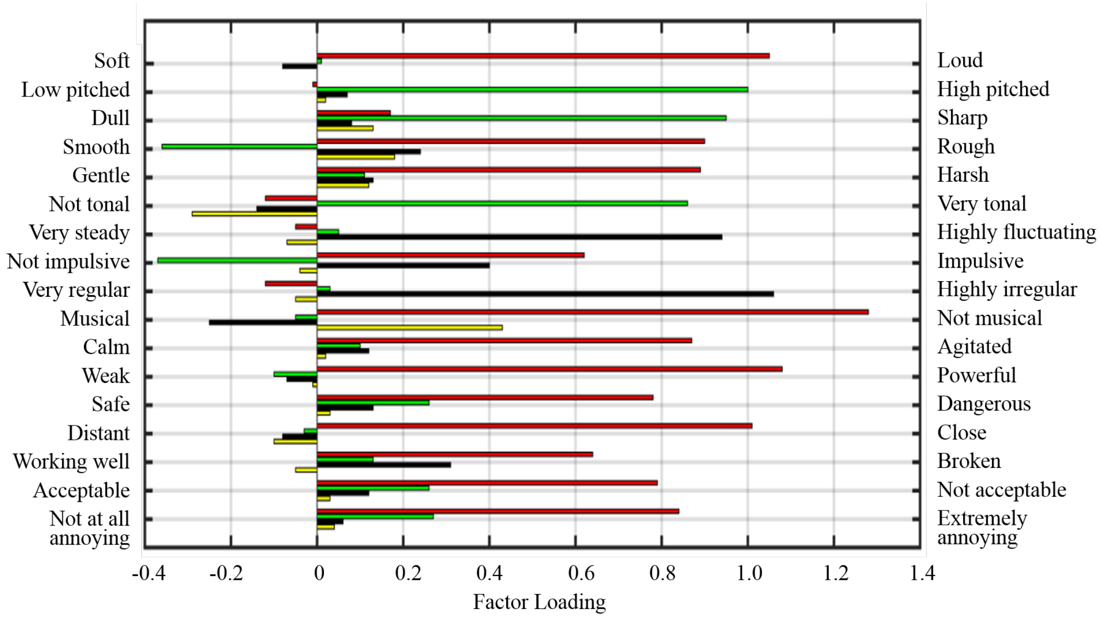


Figure 5.6. Results of a four factor analysis of all 22 sounds on 17 scales after orthogonal rotation. Bars show the factor loading: Factor 1 – Red: "Loudness" factor, Factor 2 – Green: "Tonal/Sharpness" factor, Factor 3 – Black: "Irregular/Fluctuation" factor, and Factor 4 – Yellow: "Not Musical" factor.

"Loudness" and "Irregular/Fluctuation" factors. There are three fairly strong factor loadings for the "Not Impulsive -– Impulsive" scale, showing that impulsive sounds are loud, irregular, but not sharp or not tonal.

The results of the factor analysis when using just the ratings on the 10 "sound attribute" scales are shown in Figure 5.7. In Figure 5.7(a), the analysis was based on the ratings of residential HVAC&R unit sounds and in Figure 5.7(b), the analysis was based on the ratings of refrigerated truck HVAC&R unit sounds. The first three factors in each case are similar to those in the all sounds factor analysis on all seventeen word scales shown in Figure 5.6, which included both sets of units, but in this unit-separated analysis, the fourth factor was more aligned with "Impulsiveness" rather than "Not Musical" factor, particularly for the refrigerated truck unit analysis. This could be because the refrigerated truck units' sounds have contributions from

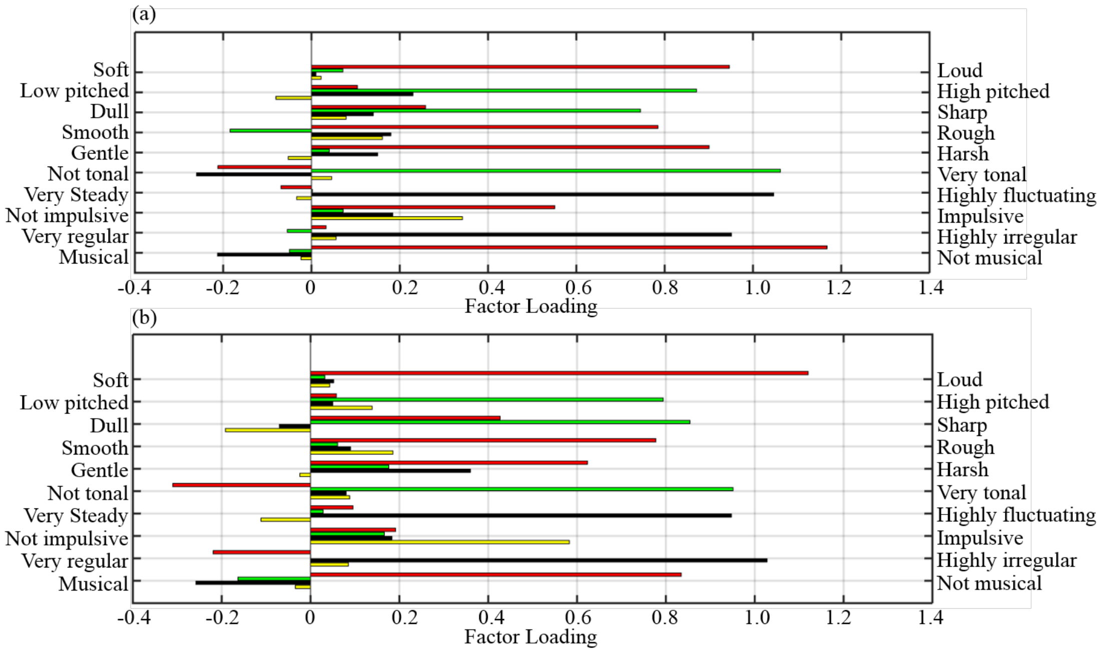


Figure 5.7. Results of a four factor analysis on 10 sound attributes scales after oblique rotation. Results are for the analysis of modified and recorded sounds from (a) residential and (b) refrigerated truck units. Bars show the factor loadings for: Factor 1 – Red: “Loudness” factor, Factor 2 – Green: “Tonal/Sharpness” factor, Factor 3 – Black: “Irregular/Fluctuation” factor, and Factor 4 – Yellow: “Impulsiveness” factor.

the diesel engine, which tend to be impulsive, see, for example, Russell [51], which is a characteristic that is not present typically in residential units’ sounds, and not in those 11 residential sound used in these tests.

The results of the analysis when using just the ratings on the 10 sound attribute scales and 9 sound quality metrics are shown in Figure 5.8. Unlike results in Figure 5.6 and Figure 5.7, “Tonal” factor and “Sharpness” factor were separated when adding sound quality metrics to the factor analysis. Here we can see the four factors as follows: “Loudness” (red), “Sharpness” (blue), “Impulsiveness” (black), and “Tonality” (green). So there is the possibility with this set of sound, to separate these as two independent characteristics, but subjects’ ratings appear to be a function of a combination of these two sound characteristics (tonal and sharpness). As noted

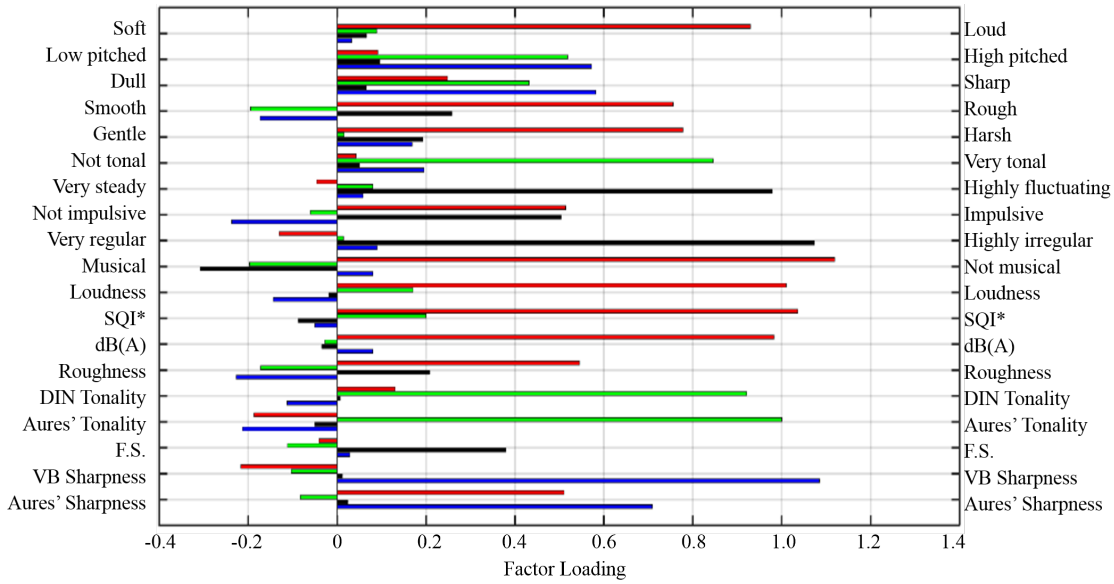


Figure 5.8. Results of four factor analysis of all 22 sounds on 10 sound attributes and 7 sound metrics after orthogonal rotation. Bars show the factor loadings for: Factor 1 – Red: "Loudness" factor, Factor 2 – Blue: "Sharpness" factor, Factor 3 – Black: "Irregular/Fluctuation" factor, and Factor 4 – Green: "Tonality" factor.

above, the ratings of the tonal scale had a large spread which also plays a role in the last of separation.

5.2.4 Annoyance Model

Linear combinations of sound quality metrics were examined as predictors of annoyance ratings. Here, the ratings of "Not at all Annoying–Extremely Annoying" scale were used as dependent variable and calculated sound quality metrics were the independent variables. The most accurate single-metric models were SQI^* (R^2=0.86, p <0.001) and N_5 (R^2=0.82, p <0.001) which are shown in Figure 5.9. Another level related metric, A-weighted sound pressure level, was also highly correlated with the average annoyance ratings (R^2=0.79, p <0.001). Two prominent outliers are shown in Figure 5.9 (filled cyan diamond above trend line). These outliers are more prominent

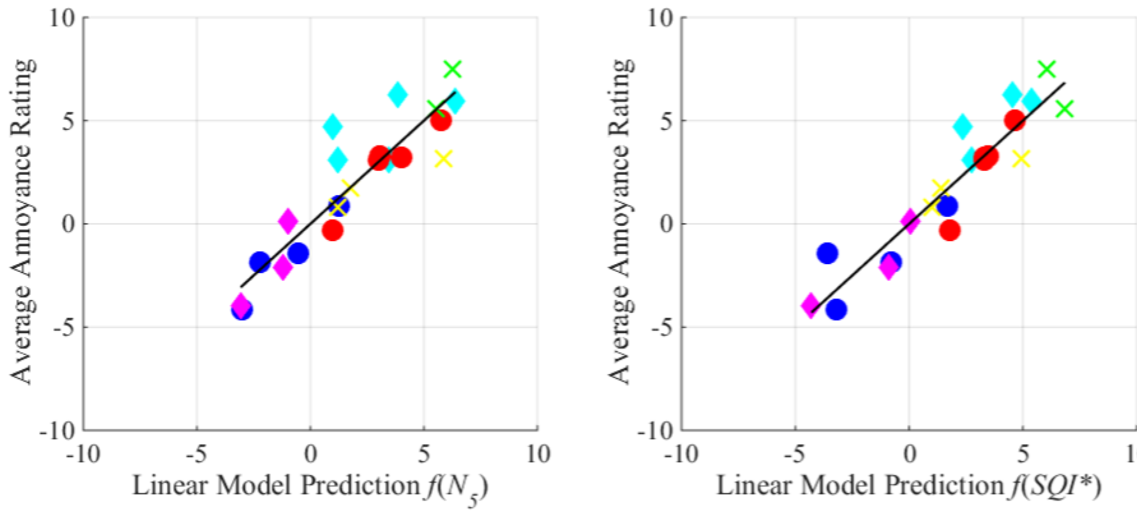


Figure 5.9. Linear regression models of annoyance using two metrics (a) N_5 (R^2=0.82) and (b) SQI^* (R^2=0.86). Blue, cyan and green colors are residential units and red, magenta, yellow colors are refrigerated truck units. Circular markers are original recordings and diamond and cross markers are modified recordings.

in the N_5 model than in the SQI^* model predictions. The reason for the better performance in SQI^* model is maybe the tone penalty included in the SQI^* calculation.

The performance of two-metric models was also examined. In these models, one of the metrics was fixed to be either N_5 or SQI^* which have the highest correlation in a single-metric model. All other metrics were considered as candidates for the second metric. The average annoyance ratings plotted against predictions from the best two-metrics models are shown in Figure 5.10. The R^2 value was increased to 0.91 (p <0.001) and 0.93 (p <0.001) when Aures' Sharpness metric (S_{A5}) was added as the second metric in the N_5 model and the SQI^* model, respectively. The R^2 value was increased when adding more terms into the models, but additional terms were not significant. These results correspond with the results of the factor analysis: "Not at all Annoying -– Extremely Annoying" scale ratings were explained mostly by the "Loudness" factor and "Tonal/Sharpness" factor. Surprisingly, adding a tonality

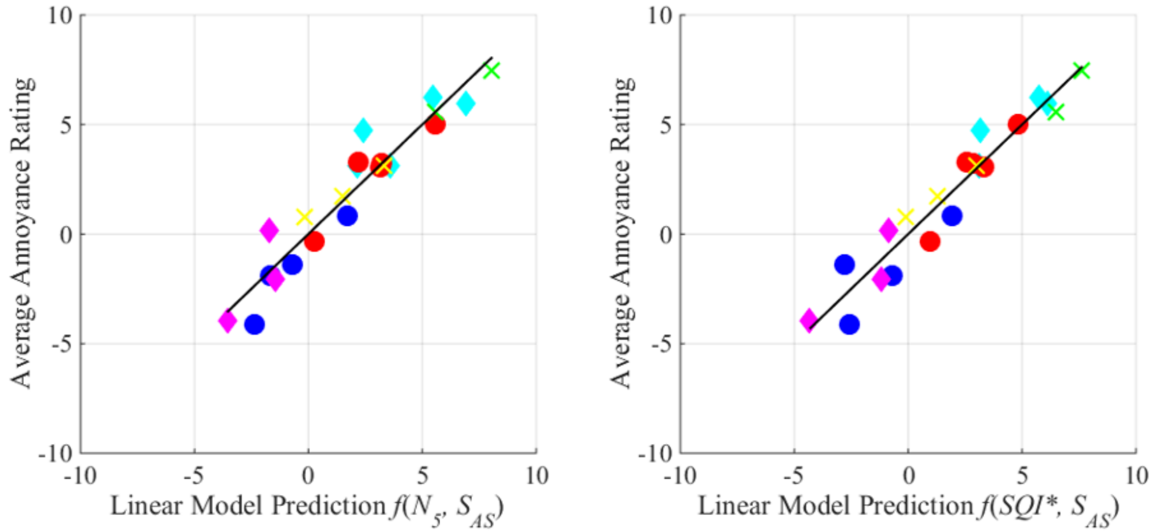


Figure 5.10. Linear regression models of annoyance using two metrics (a) N_5, S_{A5} (R^2=0.91) and (b) SQI^*, S_{A5} (R^2=0.93). Blue, cyan and green colors are residential units and red, magenta, yellow colors are refrigerated truck units. Circular markers are original recordings and diamond and cross markers are modified recordings.

metric did not improve the N_5 model significantly. However, the SQI^* model, which includes a tone correction, predicted annoyance better than the N_5 model.

5.3 Test 2 Summary

A semantic differential test was conducted to understand the independent factors that affect annoyance ratings due to exposure to HVAC&R sounds. Subjects evaluated HVAC&R recordings and modified recordings on 17 word scales. Two strong patterns were found in average rating profiles for the sets of original recordings. These profiles were mostly associated with a particular type of unit. Correlations between sound quality metrics and average ratings on each scale were the highest when the metric is supposed to be a measure of the attribute being rated. When analyzing the results of factor analysis for all sounds (modified and original sounds from both types of units) and all scales, three strong factors (“Loudness”, “Tonal/Sharpness”, "Irregular/Fluctuation") and one weaker factor ("Not Musical") were found. When ratings on the sound attribute scales were analyzed and the sounds from only one type of unit were included, the first three factor attributes were same as those in the all sounds factor analysis, but the fourth factor for the refrigerated truck sounds was more aligned with the "Impulsiveness" scale. When ratings on 10 sound attribute scales and 9 sound quality metrics were analyzed together, "Tonal" and "Sharpness" factors were separated and both became significant factors.

Factor loadings for the first ("Loudness") and second ("Tonal/Sharpness") factors were strong for the "Not at all Annoying -– Extremely Annoying" scale and these results are consistent with the best two-metric models for predicting annoyance that contained a loudness and a sharpness metric. Since SQI^* was the metric most highly correlated to the average annoyance ratings, the model containing SQI^* and Aures' Sharpness (S_{A5}) was found to be the most accurate two metric model. In Test 3, additional sound modification techniques were developed to manipulate Tonality and Sharpness characteristics independently and a few rating tests with a greater number of sounds and sound variations were conducted. This test is reported in Chapter 6.

6. TEST 3: ANNOYANCE RATING TEST

Three sets of annoyance rating tests were conducted in Test 3. These are referred to as Test 3 Parts A, B and C in Table 3.1, in which an overview of the test is given. Sixty subjects rated 120 sounds on the annoyance scale shown Figure 4.1. In Test 1 and Test 2, the Loudness ranges in each test were 40 sone. When the loudness range of the test sounds is broad, subjects tend to evaluate the sound mostly by the level of the sound. For customers and thus within the industry, other characteristics of sound may be more significant because the units under examination will compete with HVAC&R equipment of similar size, which means the loudness variations between units may not be broad. Thus, the tests (Parts A, B, and C) were designed to have different loudness ranges, as illustrated in the schematic shown in Figure 6.1. Part A included 50 quieter sounds (15 - 35 sone), Part B included 50 louder sounds (25 - 45 sone), and Part C consisted of 50 sounds with a wider range of loudness sounds (10 - 50 sone) which is similar to the ranges of Loudness in Test 1 and Test 2 sounds. Note that Loudness metric is Zwicker's time-varying Loudness exceeded 5% of the time (N_5). Fifteen common sounds were included in each of the parts the loudness range for these common sounds was from 25 to 35 sone. The three rating tests were first analyzed separately and then all together. Hence, the subjects rated 150 sounds, but only 120 of these were unique.

Half of the subjects, Group 1, did Part A first and then evaluated the sounds in Part B and then the sounds in in Part C, and the other half of the subject group, Group 2, rated sounds in Part B first and then rated the sounds in Part A, and then the sounds in Part C.

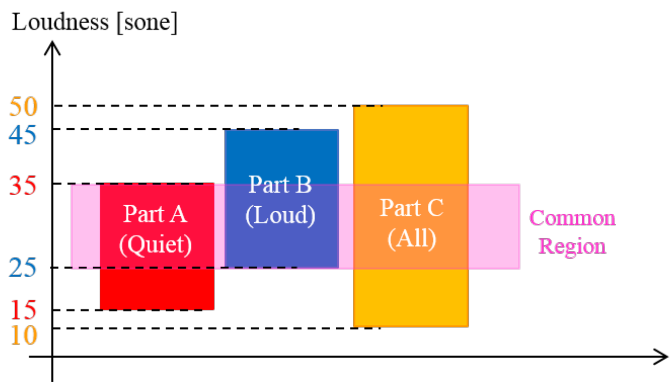


Figure 6.1. Test 3 schematic. Red - Part A (quieter test), blue - Part B (louder test), orange - Part C (wider loudness range test), and magenta - common sounds.

6.1 Selected Test Sounds and Metric Correlation

One hundred and twenty sounds were chosen from the combination of measured sounds and sounds that were synthesized through a signal decomposition and reconstruction as described in Chapter 3. A number of factors were taken into account when choosing the test sounds. First of all, for the set of sounds selected there should be low correlations between the sound metrics. At the same time, there was an effort to choose the same number of recorded and modified sounds. In addition, because Test 3 was conducted in three different parts, the Zwicker Loudness exceeded 5% of the time of the test sounds were evenly distributed across the loudness range. The set of sounds for Test 3 are described in Table C.1 in Appendix C. Metric values for these sounds are given in Appendix C Table C.2.

The examined sound metrics are listed in Table 6.1. Examples of relationships between the metrics for the groups of selected sounds are in Figure 6.2. The x-axis of Figure 6.2 (a)-(c) is N_5 and for (d)-(f) is SQI^*, which are metrics related to the level of the sound and the y-axes are Roughness exceeded 5% of the time (R_5), DIN Tonality exceeded 5% of the time (T_5) and Aures' Sharpness exceeded 5% of the time (S_{A5}). The filled circles are the recorded signals and the rest of the symbols indicate modified sounds. Blue, cyan, and green colored symbols are the sounds of residen-

Table 6.1. Sound metrics examined in the annoyance model prediction.

Metric Symbol	Metric Description	Threshold for Adjusted Metric
N_5	Zwicker time varying Loudness exceeded 5% of the time	N/A
S_{VB5}	Von Bismarck Sharpness exceeded 5% of the time	N/A
$S_{A5} \rightarrow S_{A5adj}$	Aures Sharpness exceeded 5% of the time	2.5 acum
$T_5 \rightarrow T_{5adj}$	Tonality metric exceeded 5% of the time	0.27 tu
RCL_5	Loudness derivative exceeded 2% of the time	N/A
K	Kurtosis	N/A
R_5	Roughness exceeded 5% of the time	N/A
FS_5	Fluctuation Strength exceeded 5% of the time	N/A
dBA, dBC	A or C weighted sound pressure level	N/A
SQI^*	SQI calculated from sound pressure not sound power	N/A
PA	Psychoacoustic Annoyance	N/A

Table 6.2. Correlation coefficients (ρ=$\sqrt{R^2}$) between metrics for the group of sounds selected for Test 3, Parts A, B and C.

	SQI^*	N_5	R_5	T_5	S_{A5}	RCL_5
SQI^*	1.00	0.89	0.46	0.35	0.59	-0.17
N_5	0.89	1.00	0.48	0.20	0.55	0.08
R_5	0.46	0.48	1.00	-0.02	0.01	0.01
T_5	0.35	0.20	-0.02	1.00	0.12	-0.33
S_{A5}	0.59	0.55	0.01	0.12	1.00	-0.43
RCL_5	-0.17	0.08	0.01	-0.33	-0.43	1.00

tial units and red, orange, and magenta are the refrigerated truck units. Before the adding the modified signals, the correlations between the metrics were higher especially between the S_{A5} and N_5, but after the additions, the correlations became much lower. It is important in model building to obtain a set of sounds where strengths of

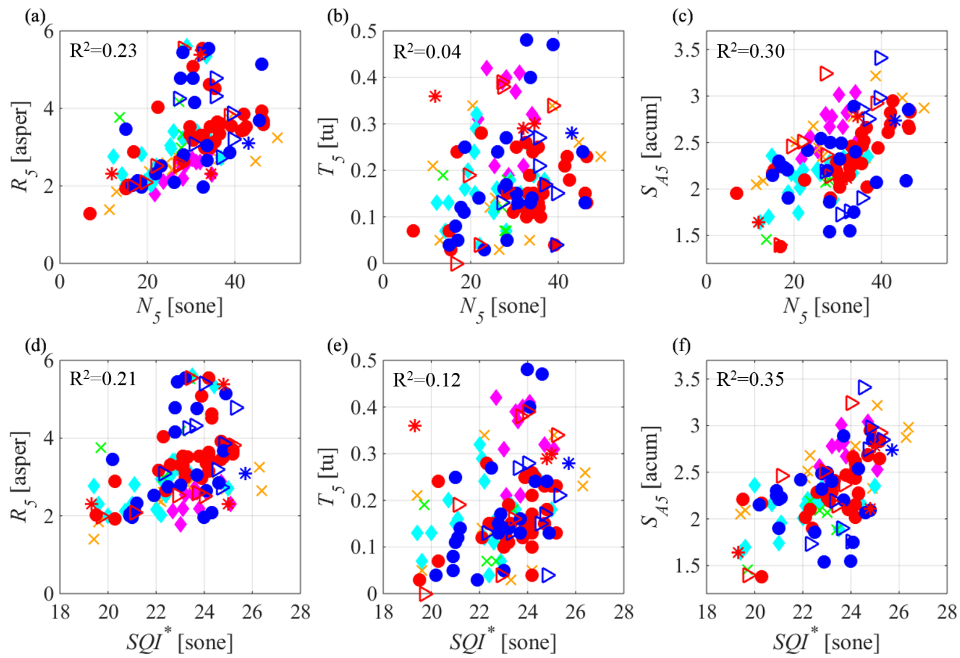


Figure 6.2. Correlations between sound metrics of Test 3 selected sounds. The markers indicate original residential(•), original refrigerated truck(•), loudness modified residential(◇), loudness modified refrigerated truck(◇), roughness or tonality metrics modified residential(x), roughness or tonality modified refrigerated truck(x), sharpness modified residential(▷), sharpness modified refrigerated truck(◁), , tonality modified residential(∗), and tonality modified refrigerated truck(∗).

important attributes have an even distribution and are uncorrelated. When lowering the metric correlations, by including modified sounds, the modified sounds were only included if they were assessed as sounding realistic, i.e., sounded like a residential or refrigerated truck unit.

Sixty subjects with normal hearing (hearing thresholds less than that or equal to 20 dB in octave bands 125 Hz to 8000 Hz) participated in Test 3. As noted earlier, 30 subjects were in Group 1 (test order: Part A, Part B, Part C) and 30 subjects were in Group 2 (test order: Part B, Part A, Part C). Part A contained quieter sounds and Part B contained louder sounds. The subjects were mostly students and staffs at Purdue University or from the local community in West Lafayette. Demographics of subjects are given in Table 3.1.

Subjects rated the sounds using the scale shown in Figure 4.1. In each part of the test the order of presentation of the sounds was randomized, with a different random order for each subject. Prior to each of the rating test parts (Parts A, B and C), 10 sounds were played for the familiarization (no rating) session and 2 practice ratings were done by the subjects.

The context provided for the ratings was the same as for Test 1 and 2, see Section 3.4 and Appendix D.

6.2 Results and Discussion

The rating mark position on the scale was translated to a number such that 2 corresponds to "not at all annoying" and 8 corresponds to "extremely annoying". The ends of the scale correspond to 1 and 9. Intermediate points on the scale correspond to 3.5 (slightly), 5 (moderately) and 6.5 (very). The subjects' ratings for each sound were averaged in 3 ways: all subjects ratings, Group 1 subjects' ratings and Group 2 subjects' ratings. For most of the analysis shown the first type of averaging is used. The second and third type of averaging was done to examine whether there are ordering effects present related to which part was done first.

The annoyance ratings from all subjects for each signal are plotted against the sound number for the three parts of Test 3 are shown in Figure 6.3. The average annoyance ratings of Part B are slightly higher than the average annoyance ratings of sounds in Part A. As expected, the average annoyance ratings from Part C are more widely distributed because of the wider distribution of Zwicker Loudness exceeded 5% of the time values (N_5). The average standard deviation of estimated means for: Part A is 0.15, Part B is 0.16 and Part C is 0.14. The range of the rating scale is 8 (1 to 9) and so these values are relatively small. The average annoyance ratings and the corresponding errors are given in Appendix E, Table E.7.

The range usage of each of the subjects is shown in Figure 6.4. Some subjects

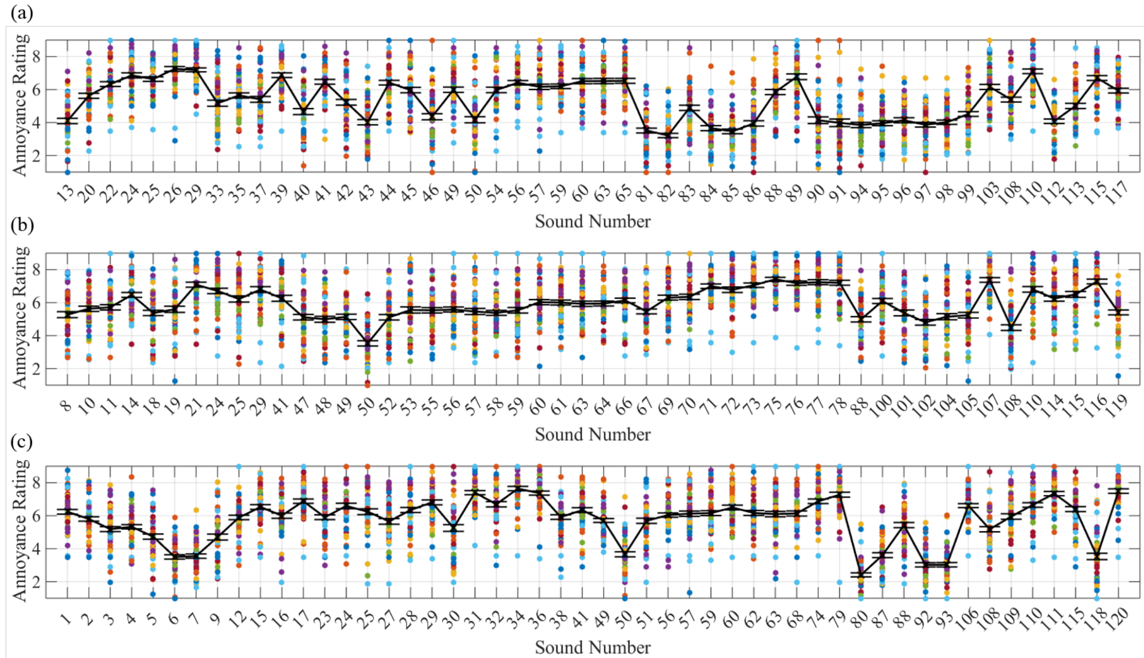


Figure 6.3. Annoyance ratings by sound number (a) Part A - quieter test, (b) Part B - louder test, and (c) Part C - wider loudness range test. Black line shows the average annoyance and bar shows the standard deviation of estimated means. Different marker shape and color indicates different subjects' responses to sounds. Numbers on the annoyance scale correspond to: 2="Not at All annoying", 3.5="Slightly Annoying", 5="Moderately Annoying", 6.5="Very Annoying", and 8="Extremely Annoying".

used a very narrow range on the rating scale (subject 3, 27, and 36) but most subjects used most of the rating scale range. There was a subject who was always rated above "moderately annoyed (5)" (subject 3) and a subject who was always rated below moderately annoyed (subject 27), and consequently they were the subjects who used only a part of the scale. The annoyance ratings plotted against the sequence of the sounds played in the test as shown in Figure 6.5. The average of these ratings by the test sequence are almost constant. This indicates that subjects did not need adjustment period in the actual test and that the familiarization and practice test were sufficient for the subjects to be comfortable doing the ratings.

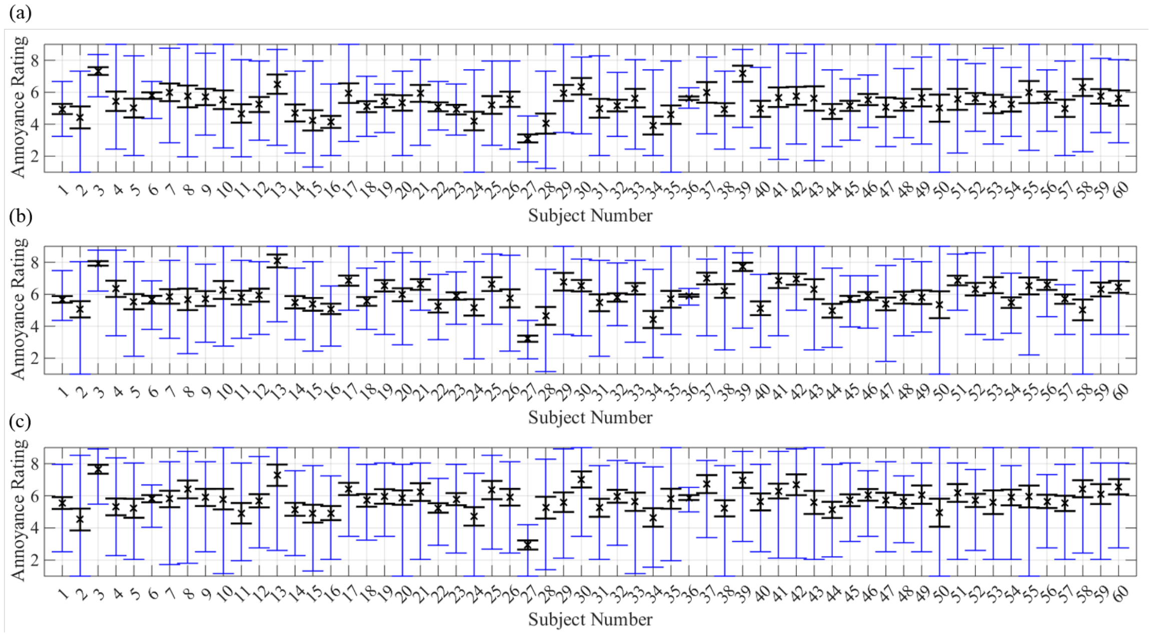


Figure 6.4. A annoyance scale range usage of the subjects (a) Part A - quieter test, (b) Part B - louder test, and (c) Part C - wider loudness range test. Black line shows the average annoyance and bar shows the standard deviation of estimated means. Numbers on the annoyance scale correspond to: 2="Not at All annoying", 3.5="Slightly Annoying", 5="Moderately Annoying", 6.5="Very Annoying", and 8="Extremely Annoying".

The average annoyance ratings plotted against the sound number for Group 1 and Group 2 are shown in Figure 6.6. The red line (Group 1) is the average annoyance ratings of the subjects who took the quieter test first (Part A) then Part B and Part C, and the blue line (Group 2) is the average annoyance ratings of the subjects who took the louder test first (Part B) then Part A and Part C. Even though the average ratings are following very similar trends, there are some significant differences in the average of the annoyance ratings. The differences are expected and are due to the ordering of the parts of the test. Group 1 will tend to rate Part B sounds higher, because they rated the quieter sounds (Part A) just before. Similarly, Group 2 will tend to rate Part A sounds lower because they rated the louder sounds (Part B) just before.

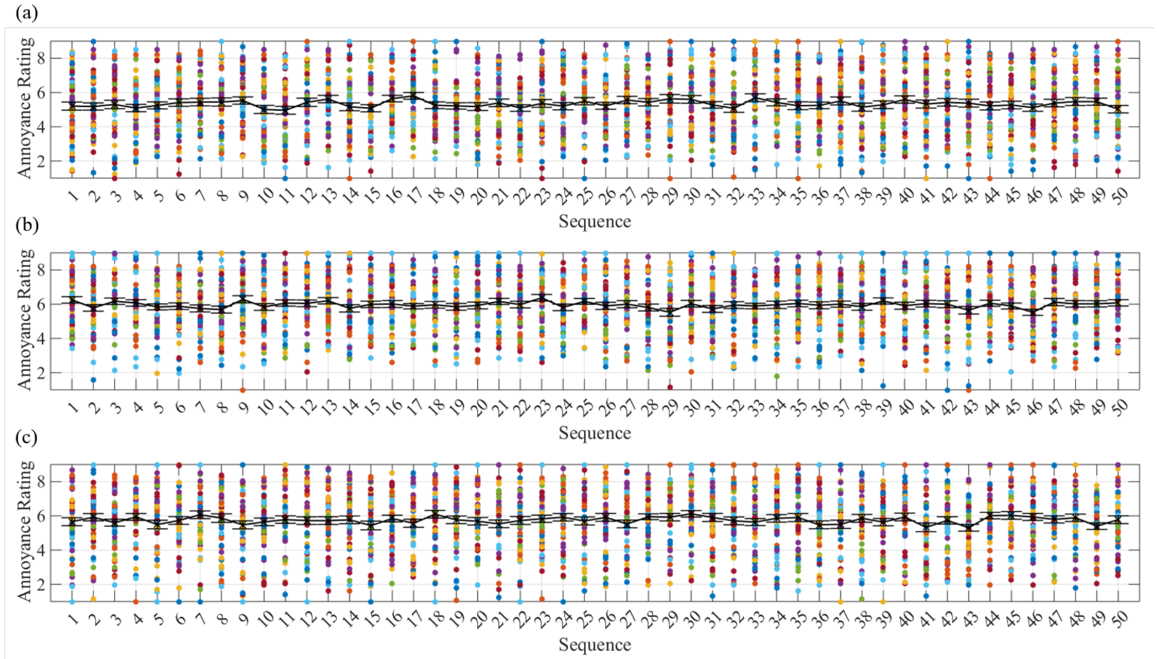


Figure 6.5. Annoyance ratings by the sequence of sounds played (a) Part A - quieter test, (b) Part B - louder test, and (c) Part C - wider loudness range test. Black line shows the average annoyance and bar shows the standard deviation of estimated means. Different marker shape and color indicate different subjects' responses to sounds. Numbers on the annoyance scale correspond to: 2="Not at All annoying", 3.5="Slightly Annoying", 5="Moderately Annoying", 6.5="Very Annoying", and 8="Extremely Annoying".

The average annoyance ratings of the 15 sounds common to all 3 parts are shown in Figure 6.7. For Group 1 the differences in average ratings are mostly not significant. For Group 2 the average ratings for the common sounds in the first part they took (Part B, the louder sounds) are significantly lower for sounds 49 to 57 and 108. The common sounds in Part B are at the lower end of the loudness range in that test part, whereas in Part A, they are at the higher end of the loudness range in that test part. For both Groups 1 and 2, where we might expect the ratings of the common sounds in Part A (black) to be slightly higher because they are in the louder end of the group of sounds, the average ratings are indeed consistently higher in Part A

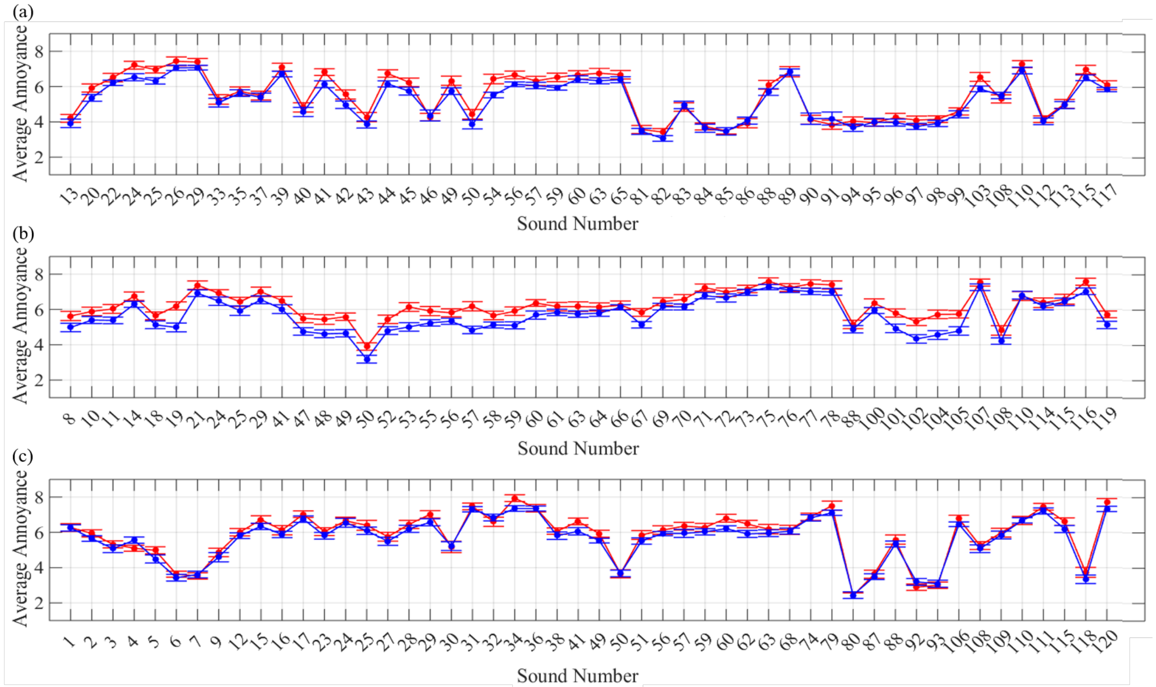


Figure 6.6. The average annoyance ratings with standard error bars. (a) Part A - quieter test, (b) Part B - louder test, and (c) Part C - wider loudness range test. Red - Group 1: Part A → Part B → Part C, blue - Group 2: Part B → Part A → Part C. Numbers on the annoyance scale correspond to: 2="Not at All annoying", 3.5="Slightly Annoying", 5="Moderately Annoying", 6.5="Very Annoying", and 8="Extremely Annoying".

across the sounds. Similarly, the ratings of the common sounds in Part B are lower because those sounds are in the group of lower loudness signals in Part B. The larger differences in Group 2 results are likely due to the slightly stronger ordering effects for Part B than for Part A, as shown in Figure 6.6 (b) and (c), respectively.

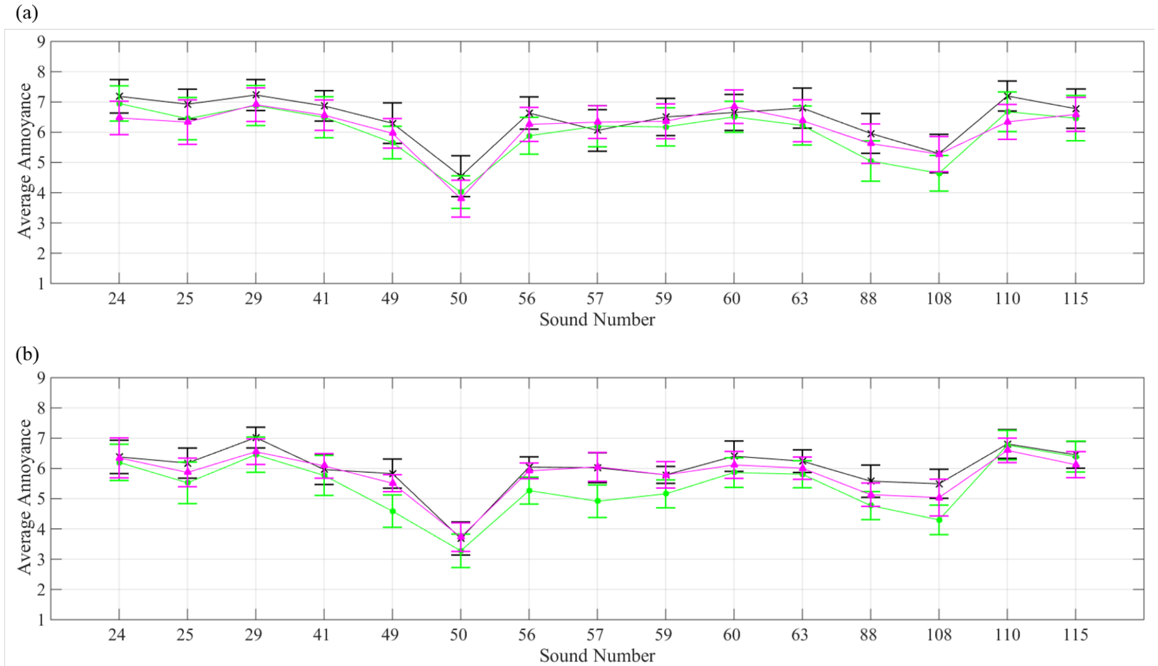


Figure 6.7. The average annoyance ratings with standard error bars of the common sounds in 3 tests. The line color correspond to: Black - Part A, Green - Part B, and Magenta - Part C. Group 1 – (a): Part A → Part B → Part C, Group 2 – (b): Part A → Part C → Part B. Numbers on the annoyance scale correspond to: 2="Not at All annoying", 3.5="Slightly Annoying", 5="Moderately Annoying", 6.5="Very Annoying", and 8="Extremely Annoying".

6.3 Annoyance Prediction Models

Annoyance prediction models that are linear functions of metrics were estimated. Prior to the model development, the average annoyance ratings from the subjects were adjusted to compensate for the slightly different average ratings observed for common signals in each part of the test. This process is described in Section 6.3.1. In the metric-annoyance analysis it was found to be advantageous to limit contribution of metrics in the annoyance model. This idea is not new. Tone adjustments to loudness in EPNL [18] and to A-weighted sound pressure level in the Danish Environmental Standard [87] are two examples of this, as is the form of the Sharpness contribution in the Unbiased Annoyance model [15]. This limiting of metric contributions is described in Section 6.3.2.

6.3.1 Aligning Average Ratings Over Subject Groups and Test Parts.

To align the average ratings in each part of the test to compensate for subject groups using scales in Part A and Part B slightly differently the following procedure was adopted. First, from Figure 6.6 (c) we see that the average ratings for Part C signals are consistent from both groups. So the approach adopted is to adjust Group 1 and Group 2 average ratings in Parts A and B to align with average ratings in Part C. The adjustment is assumed to be linear so the adjusted average rating is related to the average rating by:

$$\widetilde{A}_{pg} = \alpha_{pg} + \beta_{pg} A_{pgk}, \tag{6.1}$$

where the subscript p denotes Part A or Part B, the subscript g denotes Group 1 or Group 2, and k denotes the signal number. Eight coefficients are estimated, two for each group-part combination. $A_{p,g,k}$ denotes the average annoyance rating in Part p from Group g responses for signal k, and $\widetilde{A}_{p,g,k}$ is the corresponding adjusted ratings. α_{pg} and β_{pg} are estimated by solving the follow set of equations, in a least squares sense, for each group-part combination and using only the average ratings of the common signals:

$$\begin{bmatrix} A_{C_1} \\ A_{C_2} \\ \vdots \\ A_{C_{15}} \end{bmatrix} = \begin{bmatrix} 1, A_{pg_1} \\ 1, A_{pg_2} \\ \vdots \\ 1, A_{pg_{15}} \end{bmatrix} \begin{bmatrix} \alpha_{pg} \\ \beta_{pg} \end{bmatrix}, \tag{6.2}$$

where A_{C_k} are the average annoyance ratings for signal k in Part C of the test. The four equations for adjusting the average annoyance ratings are shown in Table 6.3. The largest offset if for Group 2 Part B, as would be expected from Figure 6.7 (c) where the green results are significantly lower than the red results for many of

Table 6.3. Adjustments to ratings to align ratings of Group 1 and 2 in Parts A and B to ratings of both groups in Part C. k=1,2,...,50. Coefficients estimated by using average ratings for the signals in Parts A, B, and C.

Group 1 Part A	$\widetilde{A}_{A,1,k} = -0.52 + 1.03\, A_{A,1,k}$
Group 1 Part B	$\widetilde{A}_{B,1,k} = 0.64 + 0.91\, A_{B,1,k}$
Group 2 Part A	$\widetilde{A}_{A,2,k} = -0.07 + 0.97\, A_{A,2,k}$
Group 2 Part B	$\widetilde{A}_{B,2,k} = 2.05 + 0.69\, A_{B,2,k}$

the signals. The greatest expansion/contraction is also Group 2 Part B which is a contraction of the range by 0.69.

6.3.2 Tonality and Sharpness Adjustment

In Test 2, a semantic different test, a combination of spectral balance and tonalness characteristics of the sound (Factor 2) affected the ratings of the "Not at all Annoying–Extremely Annoying" scale. But in this test, it was observed that metrics that predict the strength of those characteristics were not significant in annoyance prediction models as an independent variable. From this, the thresholds contributions of sound metrics such as Tonality and Aures' Sharpness were estimated and the terms adjusted under the hypothesis that the differences in values at the lower end of the range of metric values do not result in differences in the average annoyance ratings. Thresholds for each of the metrics in Table 6.1 were examined to first check the effectiveness of introducing thresholds. Two metrics, T_5 and S_{A5}, were found to perform more significantly in the annoyance model after thresholding them. Thus, the adjusted Tonality (T_{5adj}) and the adjusted Sharpness (S_{A5adj}) were introduced into the model instead of the original metrics. Both the residential and the refrigerated truck unit sounds were used to calculate thresholds.

The process for determination of thresholds is as follows.

1) Find a sharpness threshold (K_{s1}) which makes the R^2 value for a linear annoyance model ($f(N_5, S_{A5adj}, T_5, R_5)$) the maximum.

2) Find a tonality threshold (K_{t1}) which makes the R^2 value of linear model ($f(N_5, S_{A5adj}, T_{5adj}, R_5)$) the maximum.

3) Find a sharpness threshold (K_{s2}) again which makes the R^2 value of linear model ($f(N_5, S_{A5adj}, T_{5adj}, R_5)$) the maximum.

4) Find a tonality threshold (K_{t2}) which makes the R^2 value of linear model ($f(N_5, S_{A5adj}, T_{5adj}, R_5)$) the maximum.

5) Repeat steps 3) and 4) until the tonality threshold and the sharpness threshold converge.

Figure 6.8 shows the R^2 values of the linear model ($f(N_5, S_{A5adj}, T_{5adj}, R_5)$) after 10th iteration of the procedures explained above. The Sharpness threshold was 2.5 acum and the Tonality threshold was 0.27 tu. When tonality metric value is smaller

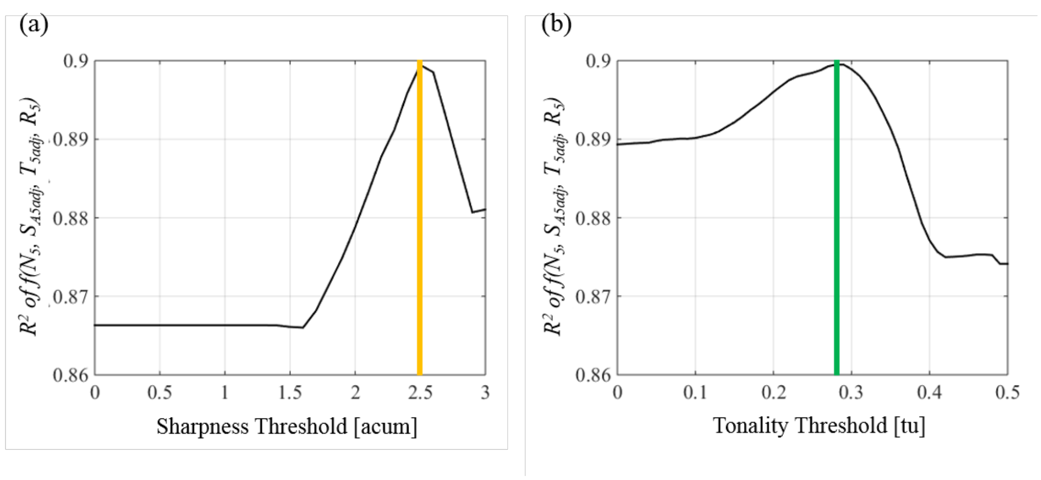


Figure 6.8. The R^2 values for models with different term thresholds. (a) Aures' Sharpness (S_{A5}) and (b) Tonality (T_5) thresholds. The orange line indicates the sharpness thresholds at the maximum R^2 values that were chosen as thresholds values: 2.5 acum and the green line indicates the tonality thresholds: 0.27 tu. This is after convergence (11th iteration of the threshold determination).

than 0.27 tu, T_{5adj} value is zero. When tonality metric value T_5 tu is greater than 0.27, the adjusted tonality value is $(T_5 - 0.27)$ tu. The calculation for the adjusted Aures' Sharpness value is following the same methodology. The adjusted Tonality and Aures' Sharpness values of all signals were calculated.

6.3.3 Annoyance Prediction Model Development

Annoyance prediction models for the HVAC&R equipment sounds were developed and the plots of the average annoyance ratings against the predicted annoyance ratings for Part A, Part B, and Part C are shown in Figure 6.9, 6.10, and 6.11, respectively. For all three Figures, plots in the first row are the N_5 loudness based models and the plots in the second row are the SQI^* based models. In Parts A, B, and C, the best single-metric models are functions of level related metrics such as the Zwicker time varying Loudness exceeded 5% of the time (N_5) or the sound quality indicator (SQI^*), which includes a tone penalty. Furthermore, adding the adjusted Aures' Sharpness metric significantly increases the accuracy of average annoyance predictions in all parts except for the Part C SQI^* based model.

There are some outlier sounds in each plots. For the Part A N_5 model, a modified residential sound (green 'x') is prominent outlier. But for the SQI^* model, recorded refrigerated truck sounds (red filled 'o') are distinct outlier sounds. Predictions under estimate annoyance in both cases. Some outlier sounds with the Part B N_5 model: a Sharpness modified residential sound (blue right triangle) and a Loudness modified residential sounds (cyan diamond). These moved towards the best-fit line after adding adjusted Aures' Sharpness to the models. However, even after these additions to the model there are still a few outlier sounds in both models' predictions. There are also a few outlier sounds (blue filled 'o') in the last plot for the Part C N_5 model. These are original residential recordings. There are a few outlier sounds (red filled 'o') in the last plot for the four-metric SQI^* model, which are original refrigerated truck recordings. Compared to the R^2 values of single metric models ($f(N_5)$ and

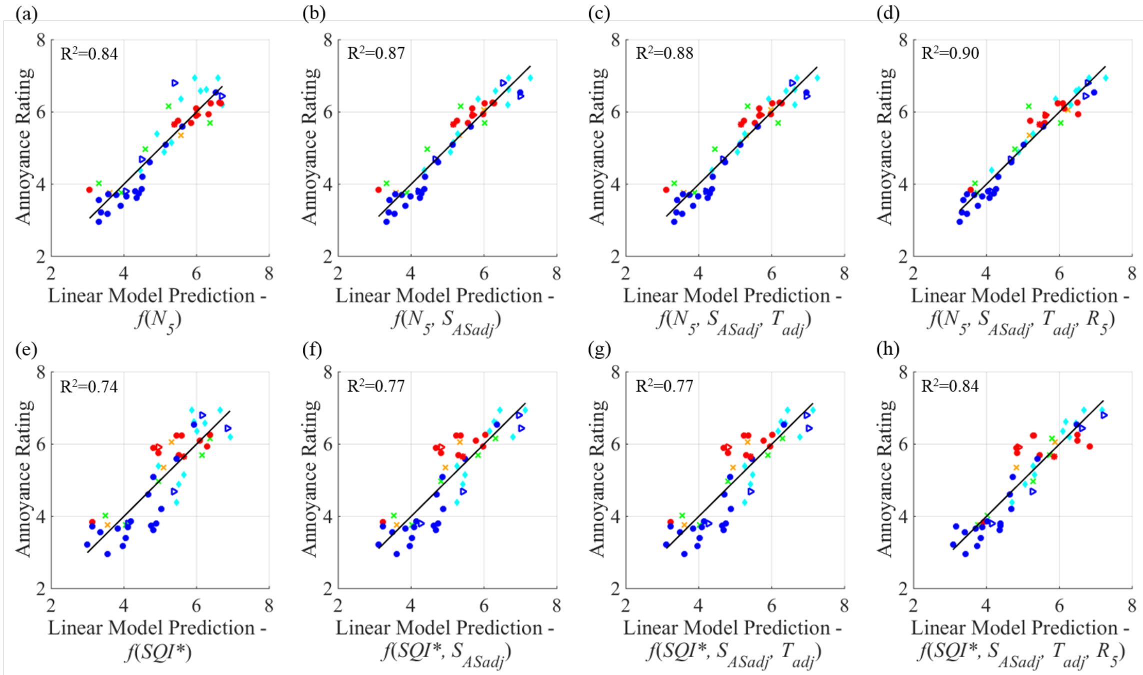


Figure 6.9. Average annoyance ratings plotted against annoyance model predictions for Part A (quieter sound test). The black line indicates best fit line to the data. The markers indicate original residential(●), original refrigerated truck(●), loudness modified residential(◇), loudness modified refrigerated truck(◇), roughness or tonality metrics modified residential(x), roughness or tonality modified refrigerated truck(x), sharpness modified residential(▷), sharpness modified refrigerated truck(◁), , tonality modified residential($*$), and tonality modified refrigerated truck($*$). (a)-(d) one to four metric models based on N_5 and (e)-(h) one to four metric models based on SQI^*. Models estimated from Part A sounds and ratings.

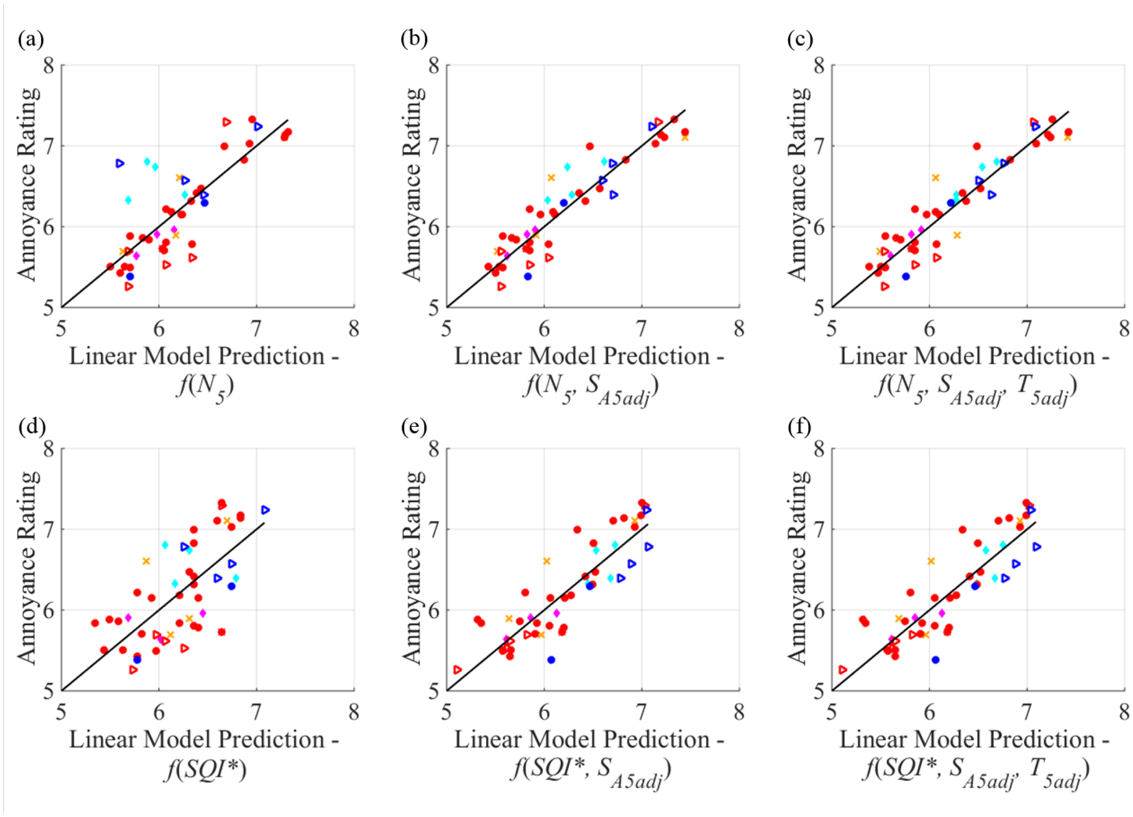


Figure 6.10. Average annoyance ratings plotted against annoyance model predictions for Part B (louder sound test). The black line indicates best fit line to the data. The markers representations are same as Figure 6.9. (a)-(c) One to three metric models based on N_5, and (d)-(f) one to three metric models based on SQI^*. Models estimated from Part B sounds and ratings.

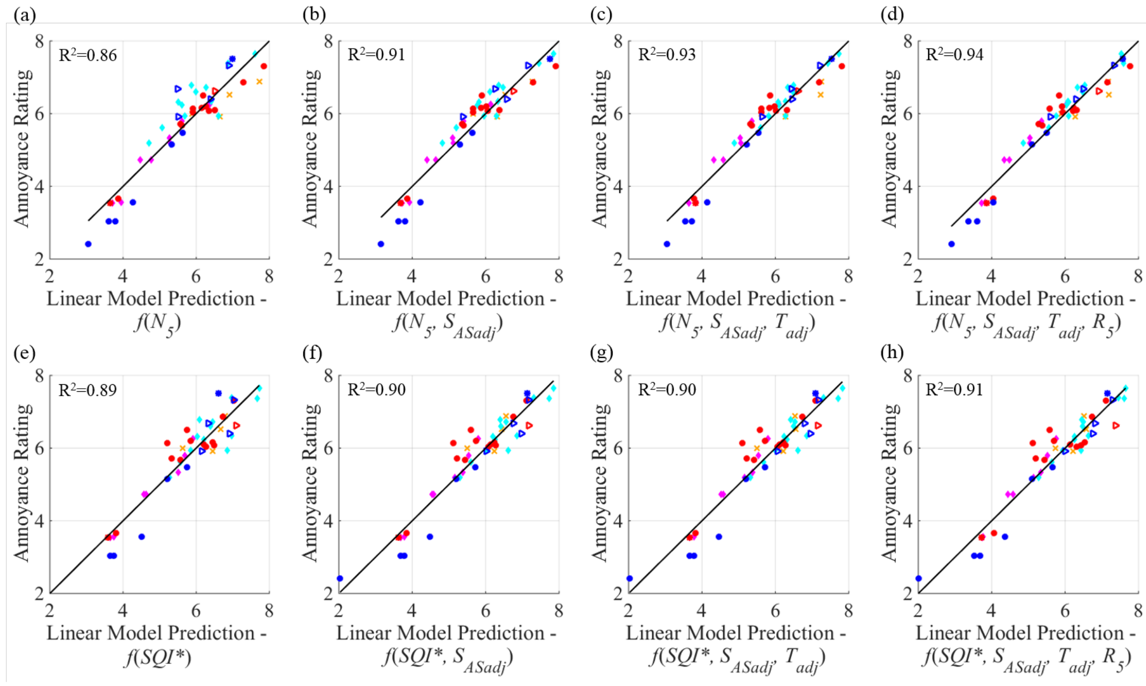


Figure 6.11. Average annoyance ratings plotted against annoyance model predictions for Part C (wider loudness range test). The black line indicates best fit line to the data. The markers representations are same as Figure 6.9. (a)-(d) one to four metric models based on N_5, and (e)-(h) one to four metric models based on SQI^*. Models estimated from Part C sounds and ratings.

$f(SQI^*)$) in Part A and Part B, that of Part C is higher because the wider loudness range makes subject's rating affected mostly by the level of a sound.

The annoyance model predictions for all sounds (150 sounds) are shown in Figure 6.12. The most accurate single-metric predictor is N_5. The best two-metric model is a function of N_5 and S_{A5adj}. Similar to the results of Part A, B, and C, adding adjusted Aures' Sharpness to the model makes some of the predictions for the Sharpness modified recordings move closer to the trend line, and thus significantly increase the R^2 value. However, the adjusted Tonality metric term was not significant in the SQI^* model, perhaps because it already includes a tonal adjustment. It was also observed that the many of the residential sounds with lower average annoyance ratings were often over predicted (below the trend line). In contrast, at higher levels of annoyance, the refrigerated truck sound results are more scattered than the results for the residential sounds. The annoyance prediction model generated from results of all 3 parts is:

$$A_{Combined,N_5} = 1.87 + 0.10N_5 + 1.15S_{A5adj} + 5.12T_{5adj} + 0.16R_5. \tag{6.3}$$

6.4 Separated Annoyance Prediction Models

The sound characteristics of a residential and a refrigerated truck units are different because their mechanical system and components are not exactly the same, especially because the sound from the refrigerated truck includes diesel engine noise. Therefore, development of two separate models for the different types of units was explored.

6.4.1 Residential Unit Annoyance Models

The average annoyance ratings for the residential HVAC&R unit sounds plotted against predicted annoyance ratings, obtained from a model estimated using residential sounds and ratings from all 3 parts of Test 3 are shown in Figure 6.13. SQI^* is the most accurate predictor among sound metrics. In previous tests and with a unit

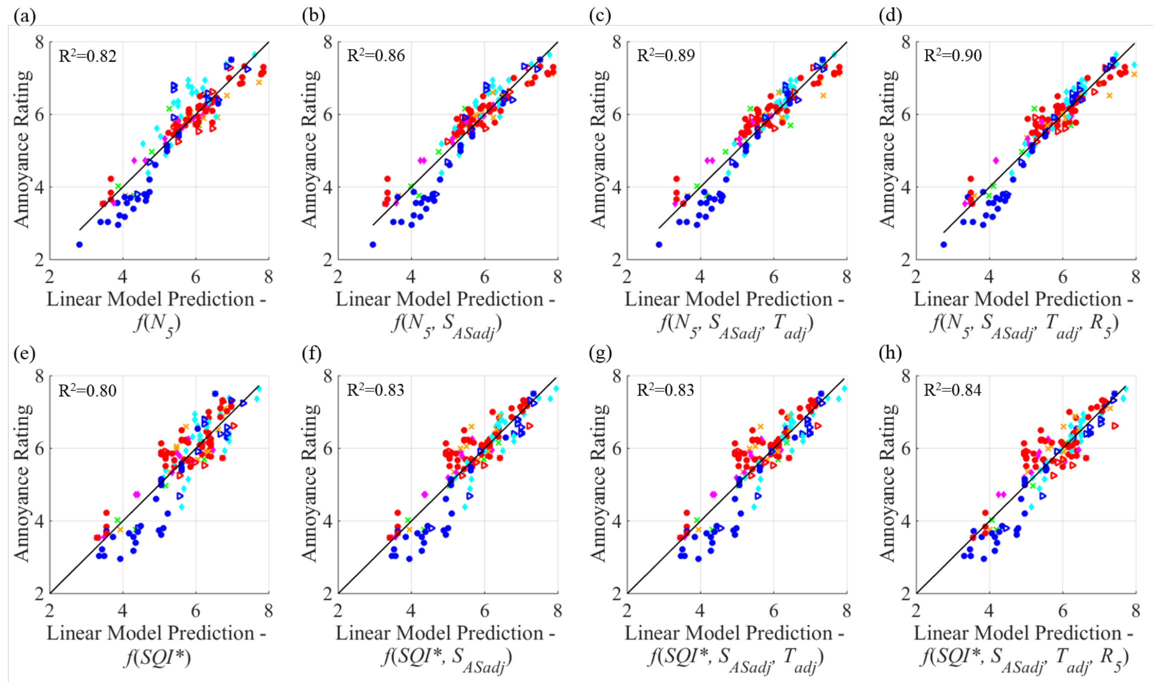


Figure 6.12. Average annoyance ratings plotted against annoyance model predictions for all sounds. The black line indicates best fit line to the data. The markers representations are same as Figure 6.9. (a)-(d) one to four metric models based on N_5 and (e)-(h) one to four metric models based on SQI^*. Models estimated from sounds and ratings in Parts A, B, and C combined.

combined analysis for Test 3, N_5 was mostly a better predictor than SQI^*. Adding an adjusted Sharpness metric (S_{A5adj}) to both N_5 and SQI^* single-metric models significantly increases the accuracy of the annoyance rating predictions, in particular, the predictions for Part B sounds (red) were greatly improved. Adding an adjusted Tonality (T_{5adj}) metric improves the predictions but R^2 increments were much smaller (but still significant), than the increment after adding an adjusted Sharpness metric. Adding an adjusted Tonality metric to the N_5 model was more effective than including it in the SQI^* model and this is probably due to the tone compensation in the SQI^* calculation. The fourth metric that improved predictions was Roughness (R_5), which increased the accuracy only by a small amount, but it was significant. In Figure 6.13 (c) and (d), there are two prominent outlier sounds they are a Tonality modified sound. This sound has an audible high frequency pitch but the Tonality metrics (T_5, T_{A5}) were not as high as expected. This sound is not an outlier in SQI^* models and this might indicate that the tone correction in SQI^* calculation works better than including the Tonality metric in models of sounds from residential HVAC&R units. The two four-metric annoyance prediction models for residential HVAC&R sounds are indicated in Equation (6.4) and Equation (6.5):

$$A_{Residential,SQI^*} = -6.21 + 0.45SQI^* + 1.60S_{A5adj} + 3.98T_{5adj} + 0.34R_5, \tag{6.4}$$

$$A_{Residential,N_5} = -0.77 + 0.11N_5 + 1.02S_{A5adj} + 7.17T_{5adj} + 0.23R_5. \tag{6.5}$$

In Table 6.4 and Table 6.5 standard error of each of the estimated coefficients and the range of contributions of terms in the annoyance model, given the range of metrics for the signals used in the test, are shown.

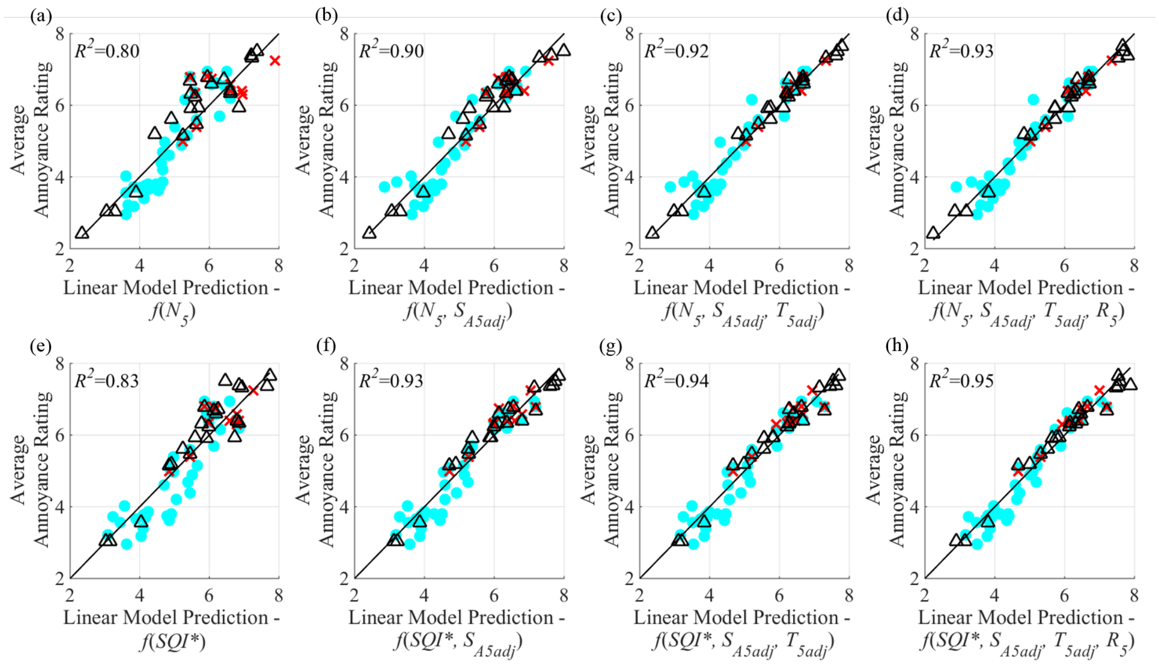


Figure 6.13. Average annoyance ratings plotted against annoyance model predictions for only residential HVAC&R sounds. Model estimated from ratings of residential sounds only. The black line indicates best fit line of the data. The markers indicate Part A (●), Part B(x), Part C(△) sounds. (a)-(d) models based on N_5 and (e)-(h) models based on SQI^*.

Table 6.4. Standard errors for estimated coefficients in the annoyance models for the residential sounds. Models estimated using the residential sounds and (adjusted) ratings and sound characteristics from all three parts of Test 3. Results for four-metric models given in Equation (6.4) and (6.5).

Models based on	Constant	Level Term	S_{A5adj} Term	T_{adj} Term	R_5 Term
N_5	0.26	0.01	0.12	0.92	0.10
SQI^*	0.58	0.03	0.18	0.98	0.09

Table 6.5. Contributions of of terms in the annoyance models over the set of residential sounds. Models estimated using the residential sounds and (adjusted) ratings and sound characteristics from all three parts of Test 3. Results for four-metric models given in Equation (6.4) and (6.5).

Models based on	Constant + Level Term	S_{A5adj} Term	T_{adj} Term	R_5 Term
N_5	0.2 to 3.4	0.0 to 0.8	0.0 to 1.5	0.2 to 0.9
SQI^*	1.3 to 6.5	0.0 to 1.0	0.0 to 0.8	0.4 to 1.2

6.4.2 Refrigerated Truck Unit Annoyance Prediction Model Development

In Test 2, a semantic differential test, the 4th factor in the factor analysis of the ratings to refrigerated truck HVAC&R sounds was “Impulsiveness” (Figure 5.7). In addition, subjects used the words or sentences related to impulsiveness such as “drilling”, “drumming” and “repetitive” to describe the refrigerated truck HVAC&R sounds in Test 1. A few impulsiveness metrics (Rate change of Loudness (RCL) and Kurtosis(K)) were examined when developing an annoyance prediction model in addition to more the commonly used sound metrics. In Figure 6.14, the average annoyance ratings of refrigerated truck sounds were plotted against annoyance predictions from the most accurate 1, 2, and 3 metric models. Responses from all three parts were used to develop models.

N_5 is the most accurate single metric predictor amongst all the sound metrics (R^2=0.91). SQI^*, which is the best predictor for residential HVAC&R sounds, was not as accurate (R^2=0.78) as N_5 when predicting annoyance ratings of refrigerated truck HVAC&R sounds. Adding S_{A5adj} in the model increased the R^2 value by 0.02. Perhaps surprisingly, adding T_{5adj} and R_5 in the model did not increase the accuracy of average annoyance rating predictions. However, the R^2 value increased by 0.01 when the rate change of the Loudness (RCL) exceeded 2% of the time was added to the two-metric model. For the SQI^* based three-metric model, adding RCL increase the R^2 value by 0.06. This mostly affected the accuracy of predictions for sounds where the annoyance ratings were under predicted by the two-metric model. The three-metric annoyance prediction model for refrigerated truck HVAC&R sounds is indicated Equation (6.6) and (6.7):

$$A_{RefrigeratedTruck,N_5} = 1.75 + 0.079N_5 + 0.76S_{A5adj} + 0.00029RCL, \tag{6.6}$$

$$A_{RefrigeratedTruck,SQI^*} = -5.69 + 0.49SQI^* + 1.17S_{A5adj} + 0.00068RCL. \tag{6.7}$$

In Table 6.6 and Table 6.7 standard error of each terms and contribution of terms in the annoyance model are shown.

Table 6.6. Standard errors for estimated coefficients in the annoyance models for the refrigerated truck sounds. . Models estimated using the refrigerated truck sounds and (adjusted) ratings and sound characteristics from all three parts of Test 3. Results for three-metric models given in Equation (6.6) and (6.7).

Models based on	Constant	Level Term	S_{A5adj} Term	RCL Term
N_5	0.14	0.005	0.12	0.00006
SQI^*	0.52	0.03	0.18	0.00011

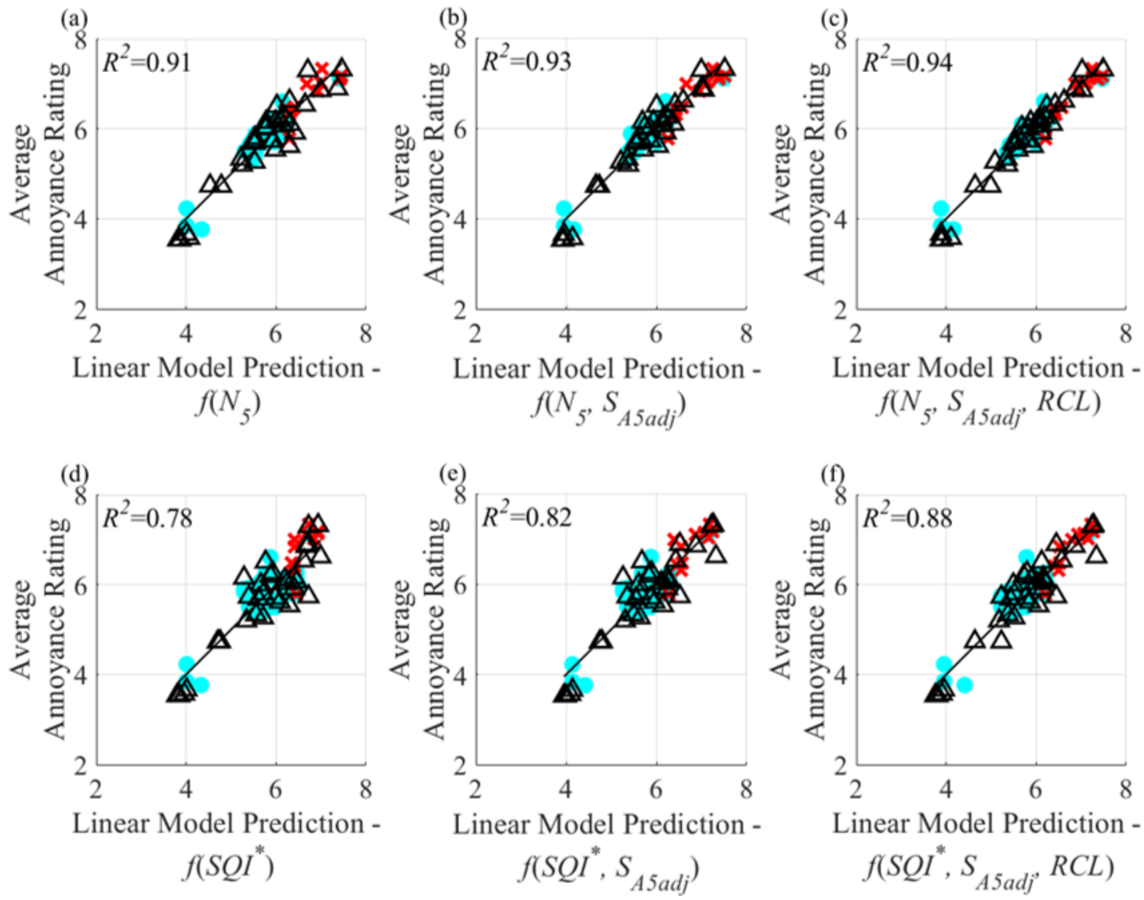


Figure 6.14. Average annoyance ratings plotted against annoyance model predictions for only refrigerated truck HVAC&R sounds. Model estimated from ratings of refrigerated truck sounds only. The black line indicates best fit line of the data. The markers indicate Part A (•), Part B(x), Part C(△) sounds. (a)-(c) models based on N_5 and (d)-(f) models based on SQI^*.

Table 6.7. Contributions of of terms in the annoyance models over the set of refrigerated truck sounds. Models estimated using the refrigerated truck sounds and (adjusted) ratings and sound characteristics from all three parts of Test 3. Results for four-metric models given in Equation (6.6) and (6.7).

Models based on	Constant + Level Term	S_{A5adj} Term	RCL Term
N_5	2.6 to 6.8	0.0 to 1.0	0.1 to 0.8
SQI^*	2.2 to 6.1	0.0 to 1.3	0.3 to 1.6

6.5 Annoyance Models' Predictions for Other Test Data

Two proposed annoyance models which are 4-metric residential unit models (Equation (6.4) and (6.5)) and one 3-metric refrigerated truck unit model(Equation (6.6)) were proposed. In this section, the performance of these models when predicting ratings for sounds not used to estimate the models' coefficients is examined. A summary of the data for each of the plots of annoyance predictions in this section is given in Table 6.8.

The annoyance ratings of Test 1 and Test 2 residential HVAC&R sounds were

Table 6.8. Overview of tests to validate estimated models' performance. The residential model is SQI^*-based and the refrigerated truck in N_5-based.

Models (development from Test 3 data)	**Sound Set**	**Figure**
Residential (Equation (6.4))	Test 1 Residential (22)	Figure 6.15 (a)
Residential (Equation (6.4))	Test 2 Residential (11)	Figure 6.15 (b)
Residential (Equation (6.4))	Test 3 Refrigerated Truck (61)	Figure 6.15 (c)
Residential (Equation (6.4))	Test 3 Residential (59)	Figure 6.15 (d)
Refrigerated Truck (Equation (6.6))	Test 1 Refrigerated Truck (14)	Figure 6.16 (a)
Refrigerated Truck (Equation (6.6))	Test 2 Refrigerated Truck (11)	Figure 6.16 (b)
Refrigerated Truck (Equation (6.6))	Test 3 Residential (59)	Figure 6.16 (c)
Refrigerated Truck (Equation (6.6))	Test 3 Refrigerated Truck (61)	Figure 6.16 (d)

predicted using the SQI^*-based model in Equation (6.4) and results are shown in Figure 6.15 (a) and (b). The model predicted the annoyance ratings accurately despite the different subjects and the different tests. These predictions were better than predictions using the N_5-based residential model (Equation (6.5)), which resulted in R^2 values of 0.88 and 0.91 for Test 1 and Test 2, respectively. The prominent outlier in predictions of annoyance for Test 1 sounds that received higher average rating than the prediction (green 'x') is a Tonality and a Roughness modified sound with audible high frequency pitch, but its associated metric values were not as high as expected. This sound contained a rich harmonic structure clearly visible in the spectrum from low to high frequencies. This is indication of the presence of a rotating or reciprocating component with very little frequency variation. Frequency variation scales with frequency so when it is present, higher harmonic features are lower and more spread out in the spectrum. There was no evidence of variation in the spectrum for this sound. This characteristic is usually associated with sounds from the compressor in residential units. Fan noise tones usually (but not always) have more frequency variations. Tonality metrics for sounds with strong harmonics families do not always work well. This is still the subject of research [70]. If the fundamental frequency is low, the sound may not sound strongly tonal even though the spectrum contains a lot of clear harmonics. Most metrics treat each identified tonal feature and pick only the one the produces the maximum "tonalness", when the harmonic family is heard as a single complex tone its strength is a function of all members of the family not just one.

In Figure 6.15 (c) and (d), Test 3 refrigerated truck and residential HVAC&R sound predictions using the residential model are shown. The predictions of the refrigerated truck sounds using SQI^*-based model were poor. This is because the sound quality indicator was designed and developed to predict the sound quality of the air-conditioning, heating, and refrigeration unit sounds not sounds that include diesel engine noise. Diesel engines often produce sounds with many harmonics, but the fundamental frequency and the pitch of the sound is low and again the discussion

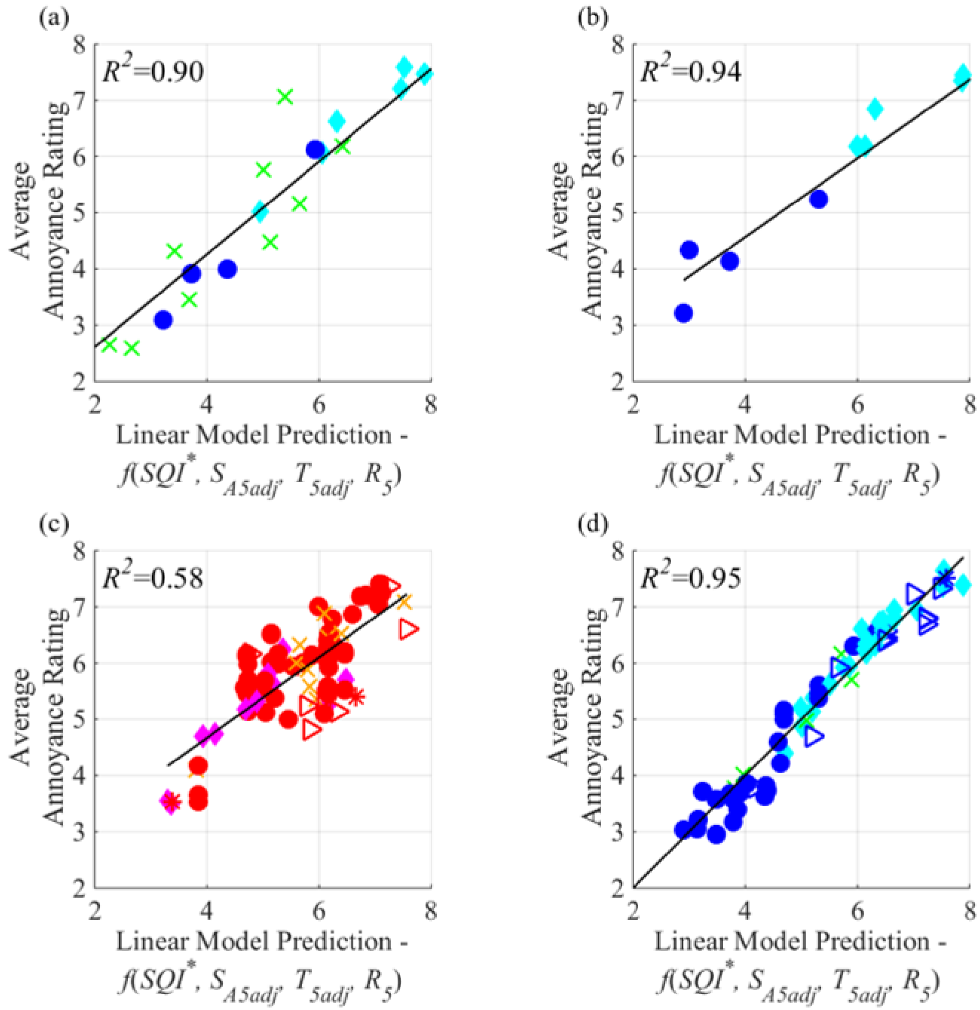


Figure 6.15. The average of annoyance ratings plotted against annoyance predictions from the SQI^*-based residential unit sound model developed using Test 3 data. Results for: (a) Test 1 residential sounds, (b) Test 2 residential sounds, (c) Test 3 refrigerated truck sounds, and (d) Test 3 residential sounds. The markers representations are same as those used in Figure 6.9.

on the outlier in Figure 6.15 (a) is relevant. However, it was clear that using SQI^* for residential sounds works well.

Similarly, the annoyance ratings of Test 1 and Test 2 refrigerated truck sounds were predicted using Equation (6.6) and results are shown in Figure 6.16 (a) and (b). In Figure 6.16 (c) and (d), Test 3 residential and refrigerated truck HVAC&R sounds were predicted using the refrigerated truck model. Unlike predicting the

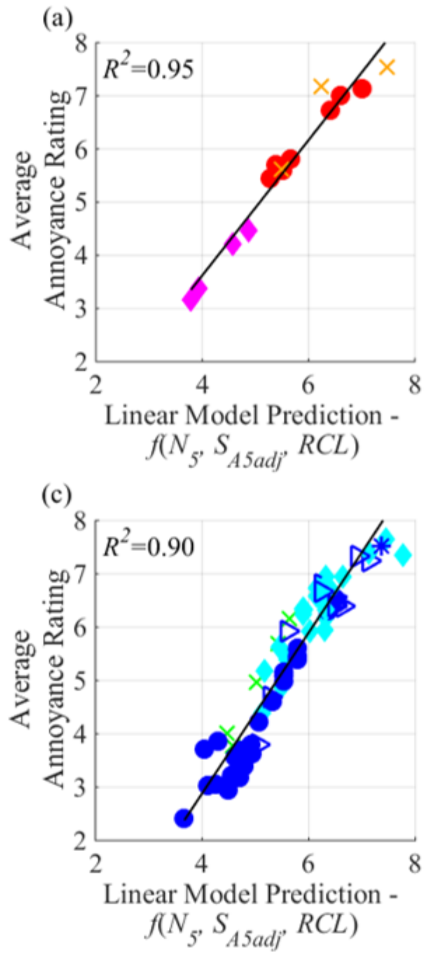


Figure 6.16. The average of annoyance ratings plotted against annoyance predictions from the N_5-based refrigerated truck unit sound model developed using Test 3 data. Results for: (a) Test 1 refrigerated truck sounds, (b) Test 2 refrigerated truck sounds, (c) Test 3 residential sounds, and (d) Test 3 refrigerated truck sounds. The markers representations are same as those used in Figure 6.9.

refrigerated truck sounds using the residential model, predicting the residential sounds using the refrigerated truck model was fairly accurate (R^2=0.90). This is because the N_5 metric was the second best predictor for residential sounds. From all these model validations, it was concluded that the SQI^* model is suitable for prediction of the residential sounds, but not for the refrigerated truck sounds. To have the best annoyance predictions of HVAC&R sounds, it is therefore beneficial to have separate models for the two types of units.

6.6 Test 3 Summary

Three rating tests were conducted to develop models that predict the annoyance ratings of HVAC&R equipment noise. Due to the wide loudness range (40 sone) of the sounds used in Test 1 and Test 2, participants' annoyance ratings were highly influenced by the level of the sound. Thus, Test 3 was sub-divided into three parts according to the loudness range of the sounds. Part A was a quieter sound test with mostly the residential HVAC&R unit sounds while Part B was a louder sound test which had many refrigerated truck HVAC&R unit sounds. Part C sounds were chosen to have a wider loudness range like Test 1 and Test 2. Fifteen common sounds were included in all three parts. Half of the participant conducted a quieter part (Part A) first and the other half conducted a louder part (Part B) first in order to compensate if should an ordering effect be present. The ratings of the common signals in the three parts were similar but there were some differences, so the average annoyance ratings in Parts A and B were adjusted to combine all data sets together when estimating models.

Several sound metrics were examined, and among them, the Aures' Sharpness and the DIN Tonality metrics were modified by thresholding the metric values. These modifications were based on the idea that subjects may only use that sound characteristic in annoyance judgements when it has reached a certain level. The annoyance prediction model, which was developed using data from both residential and refrigerated truck units together, is a function of Zwicker Loudness exceeded 5% of the time (N_5), adjusted Aures' Sharpness (S_{A5adj}), adjusted DIN Tonality (T_{5adj}), and Zwicker Roughness exceeded 5% of the time (R_5). All terms are significant and R^2 =0.90. Models for the each unit were also developed. The metrics included in the residential model were: the sound quality indicator calculated using the sound pressure (SQI^*), S_{A5adj}, T_{5adj}, and R_5; and the metrics for the refrigerated truck models were N_5, S_{A5adj}, and the rate of change of the time-varying Zwicker Loudness (RCL) exceeded 2% of the time. Both models show that the first and second most significant metrics are the level related metrics and the sharpness metric. For the residential models, N_5 based models worked well, but SQI^* based models were more accurate. However, the annoyance predictions of refrigerated truck HVAC&R sounds using the SQI^*-based model was poor (R^2 =0.58). For the refrigerated truck model, RCL was significant while the tonality and roughness metrics did not improve the model, which may be that RCL reflected the annoyance due to the impulsiveness characteristic of diesel engine sound. The developed models successfully predicted the sounds from Test 1 and Test 2, especially when using the separate models.

There are a few problems associated with some of the metrics examined. he most significant is the problem of assessing the tonality or "tonalness" of a sound when complex tones (harmonic families with many components) are present, and how the fundamental frequency (which affects the pitch of the sound) impacts the perceived tonalness of the sound. Both diesel engines and compressors can produce sounds with rich harmonic structure so this is important for these types of HVAC&R units. While we proposed using rate of change of Zwicker Loudness exceeded 2% of the time as a measure of impulsiveness, there is a need to develop this impulsiveness metric further. Repetition rates and the spectral balance may also affect perception of the impulsiveness of sounds. Subjects mentioned "pounding", "drilling", and "jack hammer" so it is a characteristic that is noticed. Other proposed metrics for impulsiveness (Kurtosis, N_{99}-N_1) were examined but RCL performed better.

7. APPLICATION: ANNOYANCE REDUCTION STRATEGY OF RESIDENTIAL HVAC&R UNIT

In the previous chapters, the annoyance prediction models were developed and validated by conducting the series of subjective tests. With the model, an annoyance rating of HVAC&R equipment sound can be predicted with a simple measurement of a sound. This allows engineers to know which sound attributes contribute to making the sound unsatisfactory. Models of the relationship between sound sources and the measured sound away from the unit provide engineers with the capability of predicting changes to the sound produced by the unit due to potential modifications of components within the unit. Coupling these sound prediction models with the annoyance prediction model gives design engineers the tools they need to reduce the likelihood of developing a unit that produces an annoying sound. Understanding how sound signal components impact sound quality metrics connects the physical characteristics of the signal to the sound characteristics perceived by the listener.

Developing source-path models is challenging. However, sound visualization techniques can be used to see which part of the machine is contributing to the sound at different frequencies, and help inform source-path model development. The visualization technique can also help in deciding which signal characteristics are related to different components in the unit, and can guide signal modification strategies coupled to machine behavior. Techniques described in Chapter 3 can then be used to modify sounds, and the sound quality metrics and the annoyance model can guide a unit sound improvement strategy. Described in this chapter is the application of such techniques to an HVAC&R equipment unit.

7.1 Noise Source Identification

It is complex and difficult to find noise sources in a machine. Many techniques such as the intensity probe measurement, Near-field Acoustical Holography (NAH), and beamforming have been developed. When a noise source is identified, it is possible to examine how it might affect annoyance. In this study, NAH which uses a microphone array of measurements was chosen to find the noise sources in the residential HVAC&R equipment.

In the NAH method, a model is developed that describes the acoustic field generated by the source [88]. Among the various NAH methods, the Equivalent Source Method (ESM) was used. The basic idea of ESM is that multiple equivalent sources are located on a plane and these sources can generate sound fields like the sound field generated by the physical sources. There are two types of equivalent source methods. In one, the sources are distributed in fixed positions and an optimization method is used to determine the source strength. In most cases the sources are assumed to be monopoles [89, 90]. In the other method the location and the strength of multi-pole sources are determined through optimization [91]. An Arbitrary Source Location (ASL) method developed by Liu [91] was used in this research to help develop robust initialization for the equivalent source method (wideband acoustical holography) described below. In this study, a fixed monopole source equivalent source method was used and this plane is referred to as the equivalent source plane, it is as shown in red in Figure 7.1. The reconstruction plane (yellow) is where the results of the optimization are used to predict the sound field. The measurement plane (blue) must be placed in front of the source plane, but can be between the equivalent source plane and the reconstruction plane, or as shown. Using the measured sound pressure level on the measurement plane, the monopole source strengths on the source plane are predicted and then the sound pressure, velocity, and acoustic intensity are reconstructed on the reconstruction plane.

There are usually a lot more monopoles on the equivalent source plane than measurements. So a strategy has to be developed that reduces the number of monopoles whose strengths are being calculated. Results are sensitive to initialization of this iterative optimization problem and also to the strategy used to turn sources off (or on) [92]. One problem is the appearance of ghost sources whereby small subtle artifacts in the measurements result in large pressure distributions on the reconstruction plane. To overcome this issue, a few regularization methods have been developed including Wideband Acoustical Holography (WBH) which was used in this research [89]. Wideband acoustical holography balances the source model accuracy and the sparsity of the system, so that the results are concentrated on reproducing the sound field from the major sources.

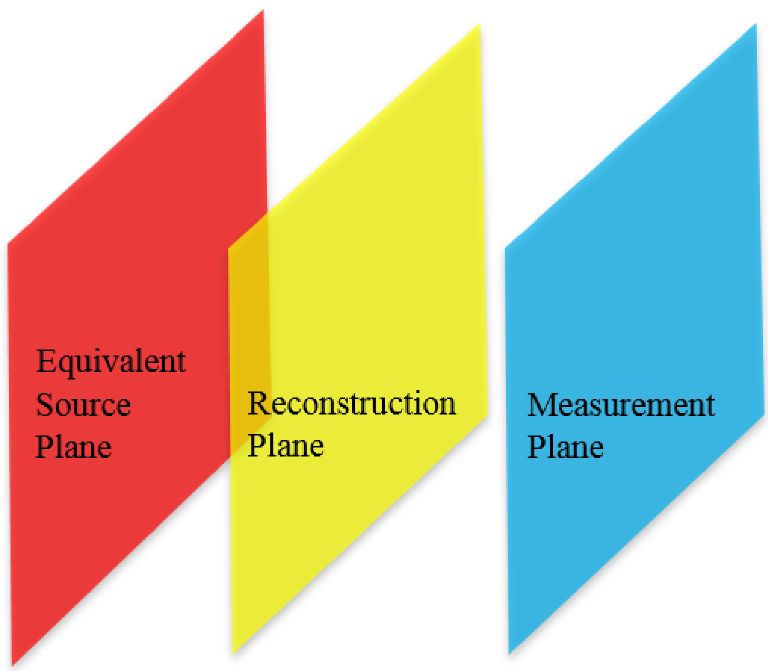


Figure 7.1. The positions of equivalent source plane (red), reconstruction plane (yellow) and measurement plane (blue) for wideband acoustical holography. The measurement plane could also be between equivalent source plane and the reconstruction plane.

The main elements of the measurement, signal processing, and estimation process are shown in Figure 7.2. The first step is the microphone array measurement. Then,

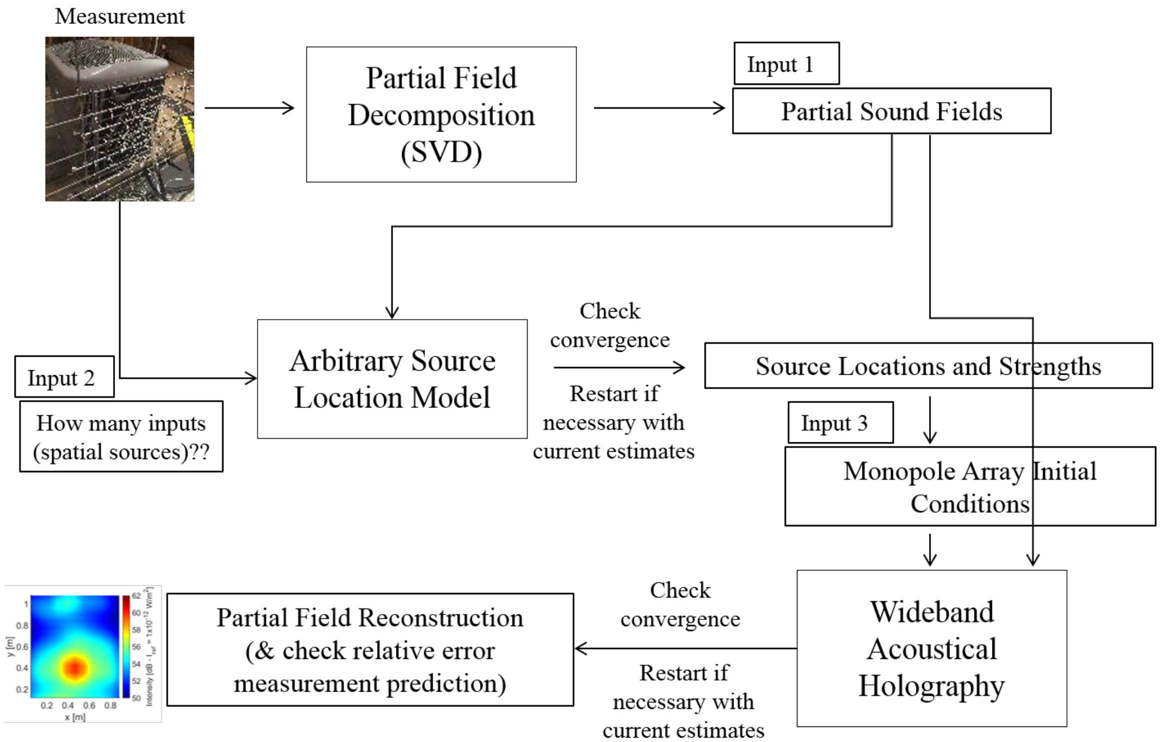


Figure 7.2. Schematic of the noise source identification process involving partial field decomposition, arbitrary source location model estimation, which provides initial estimates of the monopole source strengths for the wideband acoustical holography optimization.

using the measured pressure time histories, a cross-spectral density matrix is estimated. This is followed by a partial field decomposition that is performed to separate the contributions of different uncorrelated sound sources. To separate partial fields, a singular value decomposition was applied to the cross-spectral density matrix of measured pressure time histories. A detailed description of partial field decomposition is given is given in Lee's [93] work. Each partial field is an input to wideband acoustical holography. The wideband acoustical holography method is sensitive to the initial estimates of the source strengths in the monopole array, so Liu's arbitrary source location model [91] was used to estimate the location and strength of dominant sources and then translate those to initial estimates of the monopole source strengths in the equivalent source array. The wideband acoustic holography algorithm then optimizes the estimates of monopole strengths by minimizing the error between the measurements and the prediction of the sound field on the measurement plane from the characteristics of the monopole array. Sound pressure, velocity, and the intensity are then estimated on the reconstruction plane using the results of the wideband acoustical holography optimization.

A 55×55 array of 3025 equivalent monopole sources, 2 cm apart that covers a rectangular area 1.1 m by 1.1 m was placed on the source plane, which was 10 cm behind the reconstruction plane. The mesh of predictions' on the reconstruction plane was exactly the same as the that of the source plane (3025 predictions over a 1.1 x 1.1m area). As noted earlier there are 3025 equivalent source strengths that need to be determined from only 56 measurements. The settings in the optimization algorithm that were found to work best for the residential units studied in this research are:

$$\alpha = 0.7; \quad D_0 = 0.1\ dB; \quad \triangle D = 1.0\ dB; \quad D_{max} = 80\ dB; \quad e = 0.1, \tag{7.1}$$

where α is a relaxation factor, D_k is a dynamic range at the k-th iteration, $\triangle D$ is the step change of D_k in each iteration, D_{max} is the maximum allowable dynamic range, and the e is the desired averaged relative error between the predicted sound pressure and the measured sound pressure on the measurement plane.

7.2 Microphone Array Measurement and Results

An experiment was conducted in the semi-anechoic chamber at the Ray W. Herrick Laboratories. A picture of the residential HVAC&R equipment in the semi-anechoic chamber and also the measurement array is shown in the Figure 7.3. Fifty-six 1/4-inch microphones (PCB Piezotronics ICP Free-Field Array Microphone, Model 130F21) were used to measure the sound pressure on the measurement plane which was 10 cm away from the residential HVAC&R unit surface. The microphone readings are received through a National Instrument PXIe-4497 24-bit Sigma-Delta Analog-to-

Digital converter. Then, they are processed through National Instrument PXIe-8840 Quad-Core Data Acquisition System (DAQ) using Labview. The distance between each row of microphones was 16.5 cm and the distance between each column of microphones was 10 cm. This spacing limits the frequency range of the analysis to below 6860 Hz. After 20 minutes warm-up of the unit, the compressor speed was set to 4250 rpm (f_0 = 70.8 Hz) and the three-bladed fan speed to 900 rpm (motor rotation frequency = 15 Hz, Blade Passage Frequency = 45 Hz). The sampling frequency was 40,906 Hz and 60 seconds of data was collected to allow sufficient averaging (more than 100 raw spectral density estimates averaged, data segmented with 0.8 second Hann windows and 50% overlap of segments) when calculating the spectral densities used as the inputs to the partial field decomposition, the first stage in the wideband acoustical holography estimation process (see Figure 7.2). The spectral line spacing was thus 1.25 Hz.

Figure 7.3. 8×7 microphone array (56 microphones) used to measurd the sound from the residential HVAC&R unit. The measurement plane is 10 cm away from the face of the unit.

7.2.1 Partial Field Decomposition

The power spectral densities of the 56 array microphones and the results of partial field decomposition are shown in Figure 7.4. In Figure 7.4 (b), the thicker red line above the other lines is the total sound field which is the sum of the other 56 partial

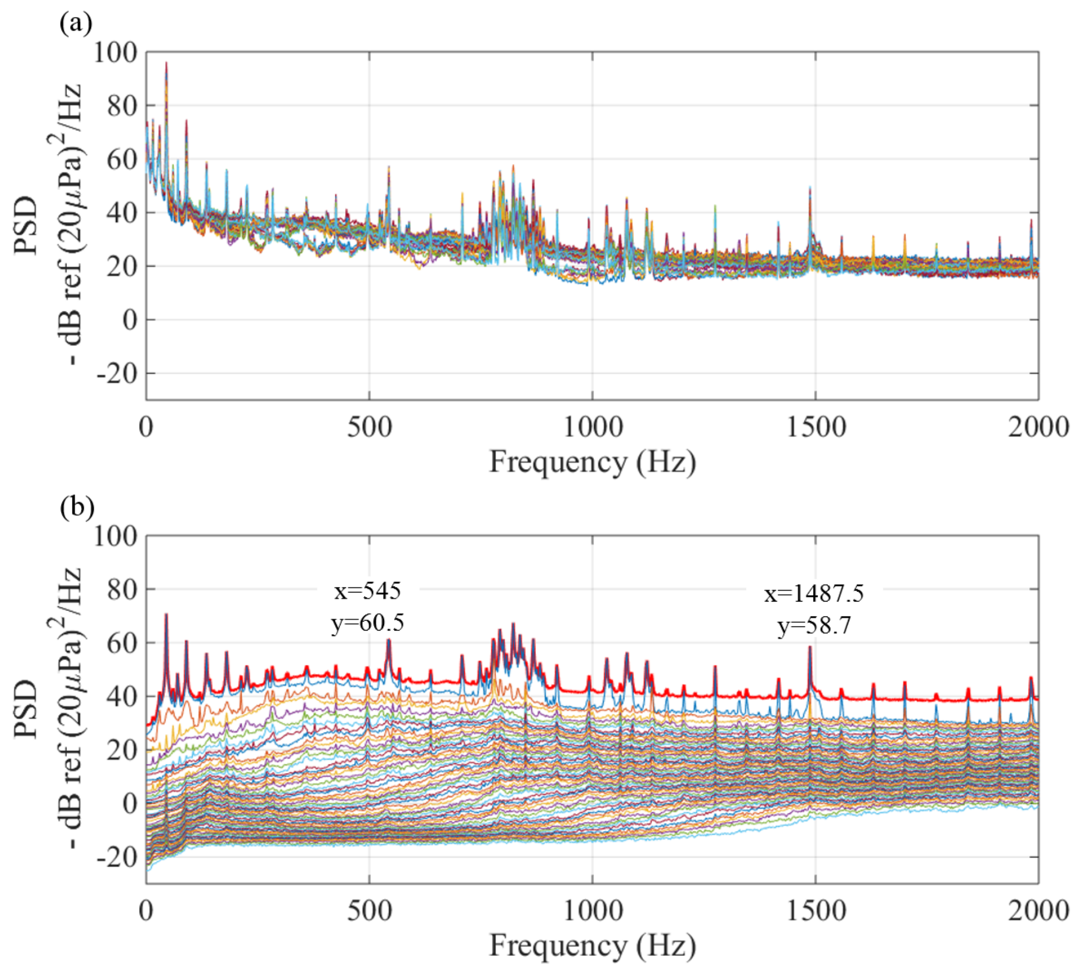


Figure 7.4. (a) The estimated power spectral densities of array microphones measurements, and (b) the results of the partial field decomposition. Fan speed was 900 rpm (BPF = 45 Hz) and compressor speed was 4250 rpm (f_0 = 70.8 Hz). [Hann Window (0.8 seconds), 50% overlap, spectral resolution - 1.25 Hz, 151 raw spectral density estimates averaged.]

fields. The first blue line right below the thick red line is the 1st (and strongest) partial field; it is close to the total sound field at many of the harmonics of the compressor and the fan motor speeds. This indicates that one uncorrelated source is dominant at these harmonic frequencies. In the broadband region, the total sound field is the sum of many partial fields. This indicates that in the broadband region, many uncorrelated sound sources are affecting the total sound field. In Figure 7.5, an example partial field decomposition result and the number of partial fields to get 99% (black) or 95% (red) of the total power are given. At tonal frequencies, the number of partial fields to have 99% or 95% is very small, 1 or 2 (Figure 7.5 (b)). At broadband frequencies, 10 or above partial fields need to be added to reach 99% of the total power.

7.2.2 Example Results Using the Equivalent Source Method with Wideband Acoustical Holography

An example sound intensity reconstruction at the 36th harmonic of the fan motor (545 Hz) and the 21st harmonic of the compressor (1487.5 Hz) are shown in Figure 7.6. At these frequencies, only the 1st partial field was used to reconstruct acoustic intensity because the peaks at 545 Hz and 1487.5 Hz could be represented by the 1st partial field with 99% accuracy (see Figure 7.5). At 545 Hz, the intensity distribution is localized in three places: the fan motor location, near the compressor shield, and at the bottom of the unit. A possible explanation is that the torque pulsation of the fan motor at the top of the unit is exciting the structure of the unit which then radiates the sound. This is currently under investigation. At 1487.5 Hz, the reconstructed intensity distribution is localized on the compressor shield.

When carefully applied, the acoustical holography method can provide accurate location predictions of the noise source at each frequency and also indicates which machine component(s) generates most noise at each frequency.

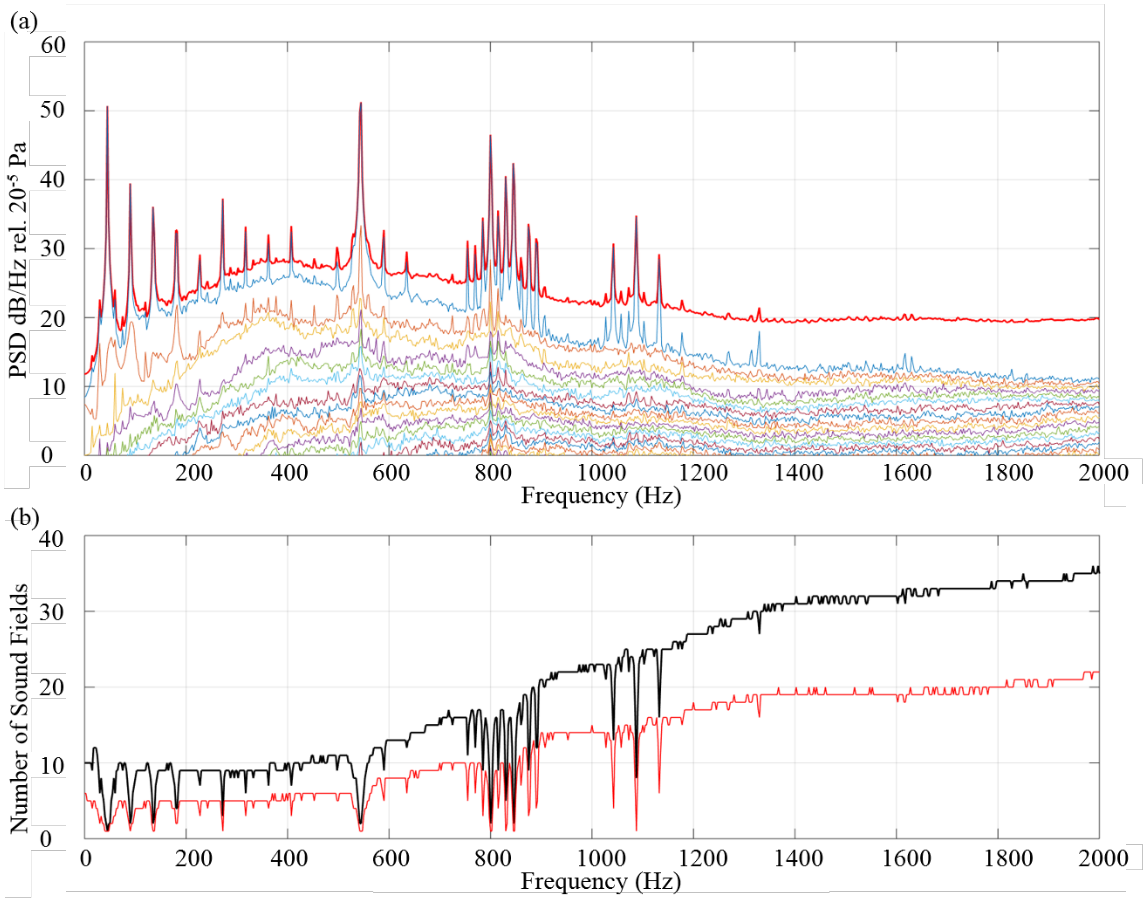


Figure 7.5. (a) The partial field decomposition 0 to 2000 Hz showing the total field red and the stronger partial fields. (b) The number of virtual sources required to predict 99% (black) or 95% (red) of the total power.

7.3 Effects on Sound Metrics and an Annoyance Prediction

The far-field sound can be predicted from the nearfield acoustical holography results or by using the nearfield acoustic holography results to guide development of a source-transfer path model and this is one of the future steps in this research [94]. However, it is significant, at this stage, to examine how the sound metrics and the annoyance prediction change when the characteristics of identified noise sources are changed. The near-field acoustical holography method was used to find the dominant noise

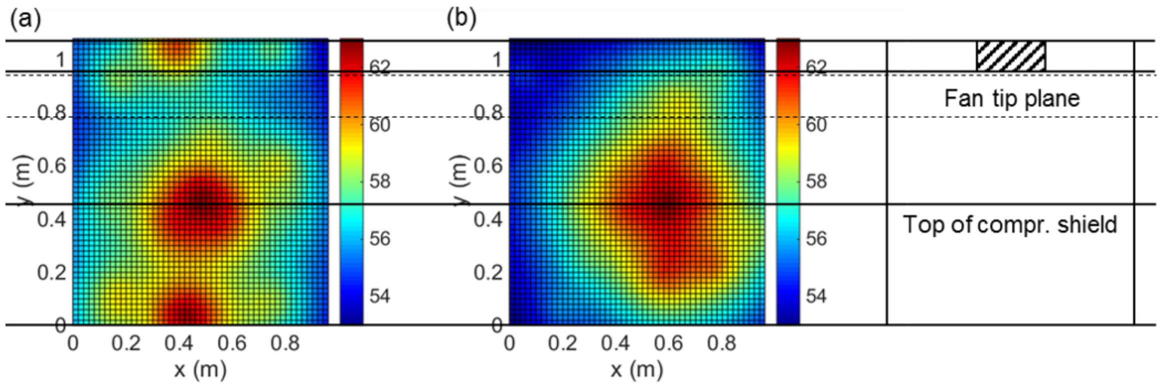


Figure 7.6. Reconstructed intensity distribution at (a) 545 Hz (36th harmonic of fan motor rotation rate) and (b) 1487.5 Hz (21st harmonic of compressor speed).

sources and those, unsurprisingly, were the fan and the compressor. The harmonics corresponding to the noise source can be identified from the spectra of the signals and from the virtual source spectra. The power spectral densities of the residential HVAC&R sound recorded in the semi-anechoic chamber, the compressor harmonics removed signal (red), and the fan harmonics removed signal (cyan) are shown in Figure 7.7.

The sound metrics and the annoyance predictions (made by using Equation (6.4)) for the original, compressor-tones-removed and the fan-removed-tones sounds are given in Table 7.1, where the number in the bracket indicates the term value in Equation (6.4). The predicted annoyance of the recorded sound was 5.00 which corresponds to "Moderately Annoying" in the rating test. The sound metrics of the residential HVAC&R sound with the compressor tones removed did not change significantly overall, but the annoyance prediction was decreased 0.19 due to a slight decrease in the T_{5adj} value. In the rating test, the numerical value of the interval between the scales (e.g. "Slight Annoying" and "Moderately Annoying") was 1.5. When the fan harmonics were removed, the SQI^* and the T_{5adj} were reduced, but the S_{A5adj} was increased because the fan harmonics were mostly located at the lower frequencies. The annoyance prediction decreased 0.38 when the fan tones were removed

Table 7.1. Summary of the sound metric values and the annoyance predictions using Equation (6.4).

	SQI^*	S_{A5adj}	T_{5adj}	R_5	Predicted Annoyance
Recorded Sound	22.4 (10.10)	0.00 (0.00)	0.10 (0.40)	2.14 (0.73)	5.00
Compressor Harmonics Removed Sound (14 tones)	22.2 (10.00)	0.00 (0.00)	0.07 (0.28)	2.22 (0.75)	4.81
Fan Harmonics Removed Sound (16 tones)	21.5 (9.70)	0.28 (0.45)	0.00 (0.00)	2.10 (0.73)	4.62

and this seems more effective than eliminating the compressor harmonics. Addressing the Sharpness increase in combination with fan tone reduction could lead to an even better result.

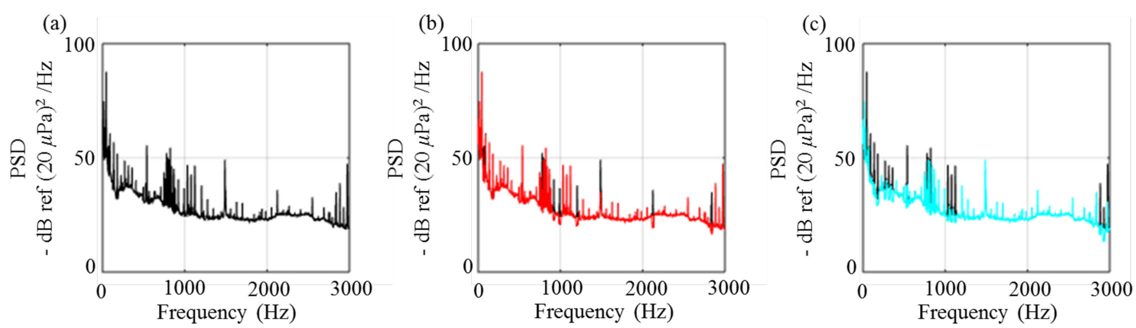


Figure 7.7. Power spectral densities of (a) the recorded signal (black), (b) the compressor harmonics removed signal (red) and (c) the fan harmonics removed signal (cyan). In (b) and (c) the power spectral density of the recorded signal (black) is plotted first as a reference and where it is visible highlights where tones have been removed.

7.4 Annoyance Reduction Strategy Summary

A sound visualization method that was used to identify locations of major noise sources was described. This will be used in future research to guide development of source-path models, which can be manipulated to explore how unit design can impact overall sound quality. To use this holography technique effectively in this HVAC&R application, it was necessary to use another technique to set up initial conditions so that the method converged. Within the nearfield acoustic holography optimization algorithm, it was also very important to find good settings for parameters such as those given in Equation (7.1). The results were particularly sensitive to the α and D_k parameters that control turning potential monopole sources on and off.

Two significant spectral features were examined in detail, one related to a fan motor harmonic and another related to a compressor harmonic. Using techniques developed to modify sounds for subjective tests (Chapter 3), compressor and fan harmonics were removed from the sound to see how that affected annoyance predictions from the model described in Chapter 6. Improvements due to reductions in tonality when fan harmonics were removed were offset by an increase in sharpness. However, it may be possible to reduce sharpness by including more sound absorption - something to explore in the development of source-path models.

8. CONCLUSIONS AND FUTURE WORK

The study focused on measuring the perceived annoyance of HVAC&R equipment sound. To build the annoyance prediction models, three sets of the subjective tests were conducted. Because of the limited numbers and lack of variety of recorded sounds, a sound decomposition and sound modification techniques were applied to the recorded sounds. The level, spectral shape, tonalness, and roughness of recorded HVAC&R equipment sound were modified to form a set of sounds with sounds including various attributes.

Test 1 consisted of a description test (Part A), a preliminary rating test (Part B) and a paired comparison test (Part C). It was observed that subjects focused on not only the level of the sound but also the other sound attributes such as sharpness, tonality, roughness, and impulsiveness. Furthermore, it was observed that the subjects' annoyance ratings were influenced by the variation of a tonality and roughness (as measured by using DIN Tonality and Roughness exceeded 5% of the time metrics) even when the sounds were normalized to have the same loudness (Zwicker Loudness exceeded 5% of the time).

Test 2 was the semantic differential test where subjects evaluated sounds on 17 scales. Word pairs at the ends of the scales were developed by using the descriptors from Test 1. A factor analysis was conducted with a rotation to find the important independent factors, and orthogonal rotation was used to align the factors with known sound attributes such as loudness. When both the residential and the refrigerated truck sounds were analyzed together, three strong factors "Loudness", "Tonal/Sharpness", and "Irregular/Fluctuation", were found. When the residential and the refrigerated truck sounds were analyzed independently, the fourth factor was more aligned with the "Impulsiveness" scale and it has higher weighting in the refrigerated truck factor analysis.

With the results from Test 1 and Test 2, Test 3 was designed to develop the annoyance prediction model. The test was divided into three parts based on the level of the loudness. Fifteen common sounds were included in the three parts and those were used to compensate for the discrepancies between the average annoyance ratings due to ordering of test parts. Aures' Sharpness (S_{A5}) and the DIN Tonality (T_5) metric values were modified by calculating the thresholds above which they contributed to annoyance predictions. The thresholds were calculated to maximize the R^2 values of the 4 metric models using an iterative procedure. The thresholds for Aures' Sharpness and DIN Tonality were 2.5 acum and 0.27 tu, respectively. Two separate models and one unified model; i.e., the residential unit model, refrigerated truck unit model, and all unit model, were developed. For the residential unit sounds, a model based on four metrics (sound quality indicator (SQI^*), adjusted Aures' Sharpness (S_{A5adj}), adjusted Tonality (T_{5adj}), and Roughness (R_5)) was found to give the best results ($R^2 = 0.95$). For the refrigerated truck unit sounds, the annoyance model consisting of Zwicker Loudness (N_5), an adjusted Aures' Sharpness (S_{A5adj}), and the rate change of the time varying Zwicker Loudness (RCL) gave the best predictions ($R^2 = 0.94$). The models of the two units are distinctly different. Using the developed models, the annoyance ratings of the Test 1 and Test 2 sounds were successfully predicted. However, when the residential model was used to predict the refrigerated truck sounds the results were poor. This confirms again that using the appropriate model for the unit is necessary.

The purpose of developing the annoyance prediction model is not only to evaluate sound quality but also to predict changes in sound quality and subjects' annoyance ratings when machine components are replaced or damaged. As an example, methodologies to find the noise sources of a the residential HVAC&R unit and to reduce annoyance were examined. In this work, a specific method of Near-field Acoustical Holography (NAH), known as the Equivalent Source Method (ESM) using Wideband Acoustical Holography (WBH) regularization was used to localize the noise sources. A microphone array was used to measure the sound pressures. The acoustic intensity was reconstructed on the residential HVAC&R unit surface by using the measured sound pressures. The fan and the compressor were found to be the main noise sources of this particular residential unit. The measured signal was modified to remove the fan and compressor harmonics independently and the changes of sound metrics and annoyance predictions were calculated using the residential annoyance prediction model. For this unit, it was found to be more efficient to modified the fan rather than the compressor. Using the same methodology it is possible to evaluate how the independent machine components are affecting the annoyance of the HVAC&R sound.

Recommendations for future work include:

1. *Investigation on the Rate Change of the Loudness Metric*: It was observed that the rate change of the time varying Zwicker Loudness metric (RCL) improves the annoyance prediction for the refrigerated truck sounds. However, a more in-depth investigation of the relationship between the RCL and annoyance of the impulsive sound is necessary. Subjective tests on impulsiveness perception and the role that rate of change of loudness plays in that perception are recommended. The many descriptors provided by subjects for impulsive sounds suggest that low frequency content and repetition might also play important roles in their impulsiveness evaluation and this also requires further investigation. Of course, for any proposed new metric for impulsiveness, there should be a step-by-step comparison between its performance and the performance of other impulsiveness metrics available in commercial sound assessment software.

2. *Outlier Sounds in the Model Fitting*: In Test 3 model development, there was an outlier sound for which the model significantly under predicted the annoyance. The spectrum has many narrow peaks over a large frequency range and there is very little evidence of any frequency variation. This is indicative of a highly repetitive process in which the temporal shape of the sound time

history includes sharp transitions. Depending on the fundamental frequency of the harmonic family existing tonality sound quality metrics might under-rate (or over-rate) the tonalness of these sounds, and there are clearly sound characteristics present not being captured when using existing sound quality metrics in linear annoyance models. The relationship between the characteristics of these types sounds (and similar sounds with, e.g., more frequency variation, or fewer high frequencies), sound quality metrics, and perception of these sounds needs to be examined further to improve the annoyance prediction model.

3. *Rationale of Deciding the Thresholds Values and Nonlinear Models of Annoyance*: The sharpness and tonality metrics were modified by calculating the thresholds. The inclusion of thresholds significantly improved the annoyance model. There may also need to be a limitation introduced at higher levels of metric values (sound attribute strength predictions) where the contribution to annoyance might saturate. This is the case in the tone adjusted level metrics. This saturation should be investigated. Also, can the refrigerated truck annoyance model be improved by, e.g., emphasizing the contribution of Tonality and attenuating the contribution of Impulsiveness as sounds get quieter, and by doing so, is it possible to generate a model that works well for both types of units?

4. *Source Path Models and Optimization of Component Noise*: The fan rotation noise, the aerodynamic noise, the compressor noise, and the structure borne noise should be modeled independently and then, combined to produce the source transfer path model that predicts the sound of the unit. The quality of the sound at the listener location is a function of both the individual component sound attributes and also how they combine together. All components could sound good, but in combination may not. Thus when designing and modifying equipment, the ability to predict the total sound of the machine is important. Tuning the sound of a component in isolation can still lead to

disappointing unit sound.

Individuals working on detailed models of components should have access to the source-path model and have the ability to plug their sound source prediction into the source-path model, so they can hear the effects of their modifications on the whole unit sound. Having the annoyance prediction models developed in this research, that take the predicted unit sound and predicts how annoying it will be, while highlighting which sound attributes are being positively or negatively affected by the component changes, would be very useful in optimizing both the components and the overall unit sound.

There are still a few studies left to be done, but the current study clearly showed what people listen to and how they perceive HVAC&R sounds, what attributes of sounds from different types of units are important, what sound metrics can be used to measure the strength of these perceived sound attributes, and how to combine them to predict the annoyance ratings. Furthermore, the accuracy of the produced annoyance prediction models was demonstrated by using them to predict annoyance ratings from other tests.

REFERENCES

REFERENCES

[1] J.S. Bradley. Disturbance caused by residential air conditioner noise. *The Journal of the Acoustical Society of America*, 93(4):1978–1986, 1993.

[2] W. Passchier-Vermeer and W.F. Passchier. Noise exposure and public health. *Environmental Health Perspective*, 108(Suppl. 1):123, 2000.

[3] K. Holmberg. *Critical Noise Factors and Their Relation to Annoyance in Working Environments*. PhD Thesis, Lulea University of Technology, Lulea, Sweden, 1997.

[4] G.J. Thiessen and A.C. Lapointe. Effect of intermittent truck noise on percentage of deep sleep. *The Journal of the Acoustical Society of America*, 64(4):1078–1080, 1978.

[5] X. Wang, Y. Zhou, R. Geovaers, J. Holguin-Veras, J. Wojtowicz, S. Campbell, J. Miguel, and R. Webber. *Feasibility of Installing Noise Reduction Technologies on Commercial Vehicles to Support Off-hour Deliveries.* Technical Report C-11-13, U.S. Transportation Collection, U.S. Department of Transportation, Washington D.C., USA, 2014, New York State Energy Research and Development Authority (NYSERDA) Contract Agreement Number 25733.

[6] ISO 532-1. *Acoustics - Methods for Calculating Loudness - Part 1: Zwicker Method*. International Organization for Standardization, ISO 532-1, Geneva, Switzerland, 2017.

[7] ISO 532-2. *Acoustics - Methods for Calculating Loudness - Part 2: Moore-Glasberg Method*. International Organization for Standardization, ISO 532-2, Geneva, Switzerland, 2017.

[8] M. Roger, S. Moreau, and A. Guedel. Broadband fan noise prediction using single-airfoil theory. *Noise Control Engineering Journal*, 54(1):05–14, 2006.

[9] S.S. Stevens. Concerning the form of the loudness function. *The Journal of the Acoustical Society of America*, 29(5):603–606, 1957.

[10] ISO 226. *Acoustics - Normal Equal-loudness-level Contours*. International Organization for Standardization, ISO 226, Geneva, Switzerland, 2003.

[11] ECMA. *Measurement of Airborne Noise emitted by Information Technology and Telecommunications Equipment*. ECMA International, ECMA-74:2019, Geneva, Switzerland, 2019.

[12] W. Aures. Procedure for calculating the sensory euphony of arbitrary sound signals. *Acustica*, 59:130–141, 1985.

[13] T.H. Pedersen, M. Sondergaard, and B. Andersen. Objective method for assessing the audibility of tones in noise, joint nordic method - version 2. *Delta Acoustic and Vibration AV 1952/99*, pages 1–18, 2000.

[14] G.V. Bismarck. Sharpness as an attribute of the timbre of steady sounds. *Acoustica*, 30:159–172, 1974.

[15] E. Zwicker and H. Fastl. *Psychoacoustics: Facts and Models*. 3rd Edition, Springer Science & Business Media, 2008.

[16] A. McMullen. *Assessment of Noise Metrics for Application to Rotorcraft*. Master's Thesis, Purdue University, West Lafayette, Indiana, USA, 2014.

[17] K. Pearson. The error law and its generalizations by fechner and pearson. *A Rejoinder. Biometrika*, 4(1/2):169–212, 1905.

[18] FAA. *Calculation of Effective Perceived Noise Level from Measured Data*. Federal Aviation Administration Code of Federal Regulations, 14 CFR Part 36, Appendix A2, Section A36.4, 2006.

[19] ANSI/AHRI 1140. *Sound Quality Evaluation Procedures for Air-Conditioning and Refrigeration Equipment*. Air-Conditioning, Heating, and Refrigeration Institute, ANSI/AHRI 1140, Virginia, USA, 2014.

[20] H.M.E. Miedema and H. Vos. Demographic and attitudinal factors that modify annoyance from transportation noise. *The Journal of the Acoustical Society of America*, 105(6):3336–3344, 1999.

[21] B. Berglund and T. Lindvall. *Community Noise, Archives of the Center for Sensory Research*. Stockholm University and Korlinska Institute, 1995.

[22] A.F. Seybert, M.J. Crocker, J.W. Moore, and S.R. Jones. Reducing the noise of a residential air conditioner. *Noise Control Engineering Journal*, 1(2):79–85, 1973.

[23] P. Susini, S. McAdams, S. Winsberg, I. Perry, S. Vieillard, and X. Rodet. Characterizing the sound quality of air-conditioning noise. *Applied Acoustics*, 65(8):763–790, 2004.

[24] S. Sato, Y. Jin, and J.Y. Jeon. Sound quality characteristics of refrigerator noise in real living environments with relation to psychoacoustical and autocorrelation function parameters. *The Journal of the Acoustical Society of America*, 122(1):314–325, 2007.

[25] W. Neise and G. H. Koopmann. Reduction of centrifugal fan noise by use of resonators. *Journal of Sound and Vibration*, 73(2):297–308, 1980.

[26] E. Envia. Fan noise reduction: an overview. *International Journal of Aeroacoustics*, 1(1):43–64, 2002.

[27] M. V. Lowson. Reduction of compressor noise radiation. *The Journal of the Acoustical Society of America*, 43(1):37–50, 1986.

[28] S. Lemire and H.D. Vo. Reduction of fan and compressor wake defect using plasma actuation for tonal noise reduction. *Journal of Turbomachinery*, 133(1):11–17, 2011.

[29] J. You, H.Y. Chang, C.H. Lee, J.H. Jeong, and J.Y. Jeon. Sound quality characteristics of refrigerator noise in real living environments. *Transactions of the Korean Society for Noise and Vibration Engineering*, 16(8):797–805, 2006.

[30] M. Schneider and C. Feldmann. Psychoacoustic evaluation of fan noise. In *Proceedings of Fan Noise 2015: International Conference on Fan noise, Technology and Numerical Methods*, Lyon, France, 15–17 April, 2015.

[31] S. Naji and A. Sanon. Human perception and fan noise. In *Proceedings of Fan Noise 2015: International Conference on Fan Noise, Technology and Numerical Methods*, Lyon, France, 15–17 April, 2015.

[32] A. Gérard, A. Berry, and P. Masson. Control of tonal noise from subsonic axial fan. Part 1: reconstruction of aeroacoustic sources from far-field sound pressure. *Journal of Sound and Vibration*, 288(4):1049–1075, 2005.

[33] C. Feldmann, T. Carolus, and M. Schneider. A semantic differential for evaluating the sound quality of fan systems. In *Proceedings of ASME Turbo Expo 2017: Turbomachinery Technical Conference and Exposition*, Charlotte, NC, U.S.A, 26–30 June, 2017.

[34] B.R. Glasberg and C.J. Moore. A model of loudness applicable to time-varying sounds. *Journal of the Audio Engineering Society*, 50(5):331–342, 2002.

[35] P. May, P. Davies, and J.S. Bolton. Correlations between subjective and objective evaluations of refrigerator noise. In *Proceedings of INTER-NOISE 96 (Noise control: the next 25 years)*, pages 2257–2260, Liverpool, England, 30 July–2 August, 1996.

[36] B.F. Berry and E. Zwicker. Comparison of subjective evaluations of impulsive noise with objective measurement of the loudness-time function given by a loudness meter. In *Proceedings of INTER-NOISE 86*, pages 821–824, Cambridge, MA, USA, 21–23 July, 1986.

[37] S. Topken and S. Van de Par. Perceptual dimensions of fan noise and their relationship to indexes based on the specific loudness. *Acta Acustica united with Acustica*, 105(1):195–209, 2019.

[38] R.P. Leite, S. Paul, and S.N. Gerges. A sound quality-based investigation of the HVAC system noise of an automobile model. *Applied Acoustics*, 70(4):636–645, 2009.

[39] S. Hohls, T. Biermeier, R. Balschke, and S. Becker. Psychoacoustic analysis of HVAC noise with equal loudness. In *Proceedings of INTER-NOISE 2014*, pages 1800–1806, Melbourne, Australia, 16–19 November, 2014.

[40] T. Yamaguchi, G. Minorikawa, and M. Kihara. Study on evaluation method of the pure tone for small fan. In *Proceedings of INTER-NOISE 2014*, pages 1629–1633, Melbourne, Australia, 16–19 November, 2014.

[41] R. Sottek and K. Genuit. Sound quality evaluation of fan noise based on hearing-related parameters. In *Proceedings of the Third International INCE Symposium on Fan Noise*, Lyon, France, 17–19 September, 2007.

[42] L.S. Sondergaard and T.H. Pedersen. Tonality in wind turbine noise. iec 61400-11 ver. 2.1 and 3.0 and the Danish/Joint Nordic Method compared with listening tests. In *Proceedings of 5th International Conference on Wind Turbine Noise*, Denver, Colorado, USA, 28–30 August, 2013.

[43] ISO 1996-2:2017. *Acoustics - Description, Measurement and Assessment of Environmental Noise - Part 2: Determination of Sound Pressure Level*. International Organization for Standardization, ISO 1996-2:2017, Geneva, Switzerland, 2017.

[44] S.Y. Wang. Determination of sound quality of refrigerant compressors. In *Proceedings of the 12th International Compressor Engineering Conference at Purdue*, West Lafayette, Indiana, USA, 19–22 July, 1994.

[45] K.Y. Cho, S.B. Yang, H.W. Kim, and J.C. Kim. Improving sound quality of reciprocating compressor using random PWM. In *Proceedings 8th International Conference on Power Electronics and Variable Speed Drives*, pages 431–436, London, England, 18–19 September, 2000.

[46] S.G. Park, J.T. Park, K.W. Seo, and G.B. Lee. Comparison of the sound quality characteristics for the outdoor unit according to the compressor model. In *Proceedings of the 21st International Compressor Engineering Conference at Purdue*, West Lafayette, Indiana, USA, 18–19 September, Paper No. 2229, 2012.

[47] M.F. Russell. Subjective assessment of diesel vehicle noise. In *Proceedings of XXIV Fisita Congress*. London, England, 7–11 June, IMECHE NO C 389/004 and FISITA NO 925187, 1992.

[48] K. Genuit. The use of psychoacoustic parameters combined with A-weighted SPL in noise description. In *Proceedings of INTER-NOISE 1999*, pages 1887–1892, Fort Lauderdale, FL, USA, 06–08 December, 1999.

[49] C. Patsouras, H. Fastl, D. Patsouras, and K. Pfaffelhuber. Psychoacoustic sensation magnitudes and sound quality ratings of upper middle class cars' idling noise. In *Proceedings of the 17th International Conference on Acoustics*, Rome, Italy, 02–07 September, 2001.

[50] A. Hastings. *Sound Quality of Diesel Engines*. PhD Thesis, Purdue University, West Lafayette, Indiana, USA, 2004.

[51] M.F. Russell and R. Haworth. Combustion noise from high speed direct injection diesel engines. In *Proceedings of 1985 SAE Surface Vehicle Noise and Vibration Conference*, Traverse City, Michigan, USA, Paper No. 850973, 1985. The Society of Automotive Engineering.

[52] A.J. Champagne and N. Shiau. Commercial van diesel idle sound quality. In *Proceedings 1997 SAE Noise and Vibration Conference*, Traverse City, Michigan, USA, 15–17 May, Paper No. 971980, 1997. The Society of Automotive Engineering.

[53] M. Bodden and R. Heinrichs. Diesel sound quality analysis and evaluation. In *Proceedings of Forum Acusticum: 4th European Congress on Acoustics*, pages 1–4, Budapest, Hungry, 15 August–02 September, 2005.

[54] M. Bodden and R. Heinrichs. Diesel impulsiveness: technical process. In *Proceedings of INTER-NOISE 2007*, pages 2469–2477, Istanbul, Turkey, 28–31 August, 2007.

[55] F. Payri, A. Broatch, X. Margot, and L. Monelletta. Sound quality assessment of diesel combustion noise using in-cylinder pressure components. *Measurement Science and Technology*, 20(1):No.015107, 2009.

[56] A.M. Willemsen and M.D. Rao. Prediction of subjective annoyance for transient sounds using a novel loudness-based impulsiveness measure. In *Proceedings of INTER-NOISE 2010*, pages 159–170, Lisbon, Portugal, 13–16 June, 2010.

[57] S.E. Wright and A. Damongeot. Psychoacoustic studies of impulsive noise. Report Paper No. 55, Onera and Aerospatiale, 1977.

[58] R. Ingham, N. Otto, and T. McCollum. Sound quality metric for diesel engine. *SAE Transactions*, 108(6):3003–3007, 1999.

[59] P. Sellerbeck, C. Nettelbeck, R. Heinrichs, and T. Abels. Information on levels of environmental noise requisite to protect public health and welfare with an adequate margin of safety. Report 550-9-74-004, U.S. Environmental Protection Agency, Washington D.C., USA, 1974.

[60] ANSI S3.5. *Methods for the Calculation of the Articulation Index*. American National Standards Institute, ANSI S3.5, New York, USA, 1969.

[61] ANSI S3.5. *Methods for Calculation of the Speech Intelligibility Index*. American National Standards Institute, ANSI S3.5, New York, USA, 1997.

[62] W. Sung, P. Davies, and J.S. Bolton. Sound quality evaluation of residential HVAC&R equipment. In *Proceedings of INTER-NOISE and NOISE-CON 2018*, volume 258, pages 1066–1074, Chicago, Illinois, USA, 13–16 June, 2018.

[63] E. Zwicker. Procedure for calculating loudness of temporally variable sounds. *The Journal of the Acoustical Society of America*, 62(3):675–682, 1977.

[64] B. Moore and B. Glasberg. A revision of Zwicker's loudness model. *Acustica-Acta Acustica*, 82(2):335–345, 1996.

[65] ISO 532B. *Acoustics - Method for Calculating Loudness Level*. International Organization for Standardization, ISO 532B, Geneva, Switzerland, 1975.

[66] DIN45631. *Acoustics - Method for Calculating Loudness Level*. Deutsches Institut fur Normung, DIN45631/A1, Berlin, Germany, 2010.

[67] R. Sottek. A hearing model approach to time-varying loudness. *Acta Acustica united with Acustica*, 102(4):725–744, 2016.

[68] E. Terhardt, G. Stoll, and M. Seewann. Algorithm for extraction of pitch and pitch salience from complex tonal signals. *The Journal of the Acoustical Society of America*, 71(3):679–688, 1982.

[69] K.H. Lee. *Perception of Tones in Machinery Noise and Its Influence on Annoyance*. PhD Thesis, Purdue University, West Lafayette, Indiana, USA, 2006.

[70] K.H. Lee, P. Davies, and A.M. Surprenant. Tonal strength of harmonic complex tones in machinery noise. *The Journal of the Acoustical Society of America*, 118(3):1921, 2005.

[71] S. Topken. *Sound Characteristics of Multi-tone Sounds: Measurement Concepts and Model Approach*. PhD Thesis, University of Oldenburg, Oldenburg, Germany, 2017.

[72] J. Becker and R. Sottek. Psychoacoustic tonality analysis. In *Proceedings of INTER-NOISE and NOISE-CON 2018*, Chicago, Illinois, USA, 13–16 June, 2018.

[73] P.D. Schomer. High-energy impulsive noise assessment. *The Journal of the Acoustical Society of America*, 79(182):182–186, 1986.

[74] J.W. Leverton. Helicopter noise: Can it be adequately rated? *Journal of Sound and Vibration*, 43(2):351–361, 1975.

[75] J.M. Fields and C.A. Powell. Community reactions to helicopter noise: Results from an experimental study. *The Journal of the Acoustical Society of America*, 82(2):479–492, 1987.

[76] R. Sottek and M. Tobias. Loudness perception and modeling of impulsive sounds. In *Proceedings of EuroNoise 2015*, pages 1919–1924, Maastricht, Netherlands, 31 May–3 June, 2015.

[77] D.J. Carr. *Two Laboratory Studies of Peoples Responses to Sonic Booms and Other Transient Sounds as Heard Indoors*. Master's Thesis, Purdue University, West Lafayette, Indiana, USA, 2015.

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

[79] R. Likert. A technique for the measurement of attitudes. *Archives of Psychology*, 22(140):5–55, 1932.

[80] L.L. Thurstone. A law of comparative judgement. *Psychological Review*, 34(4):278–286, 1927.

[81] J.G. Snider and C.E. Osgood. *Semantic Differential Technique: A Source Book*. Aldine Pub. Co., 1969.

[82] R.A. Bradley. *Paired Comparisons: Some Basic Procedures and Examples, Vol. 4 of Handbook of Statistics*. Elsevier Science Publishers, 1984.

[83] Y. Bi, T. Reid, and P. Davies. An exploratory study on proposed new sounds for future products. *Noise Control Engineering Journal*, 65(3):244–260, 2017.

[84] ANSI/AHRI 1120. *Acoustical Test Methods and Sound Power Rating Procedures for Transport Refrigeration Equipment*. Air-Conditioning, Heating, and Refrigeration Institute, ANSI/AHRI 1120, Virginia, USA, 2012.

[85] M.G. Dittrich, H.W. Jansen, and A.M. van Noort. *Methods of Measurement for Peak Noise During Loading and Unloading*. TNO Report MON-RPT-2010-00466, Amsterdam, Netherlands, 2010.

[86] D. Carr and P. Davies. The influence of sharpness on people's responses to wind noise inside automobiles. In *Proceedings of NOISE-CON 2017*, pages 749–757, Grand Rapids, Michigan, USA, 12–14 June, 2017.

[87] EPA. *Environmental Noise Regulation in Denmark*. Danish Environmental Protection Agency, Odense, Denmark, 2012.

[88] P. Nelson and S.H. Yoon. Estimation of acoustic source strength by inverse methods: Part I, conditioning of the inverse problem. *Journal of Sound and Vibration*, 233(4):639–664, 2000.

[89] J. Hald. Fast Wideband Acoustical Holography. *The Journal of the Acoustical Society of America*, 139(4):1508–1517, 2016.

[90] J.S. Bolton T. Shi, Y. Liu. Spatially sparse sound source localization in an under-determined system by using a hybrid compressive sensing method. *The Journal of the Acoustical Society of America*, 146(2):1219–1229, 2019.

[91] Y. Liu. *Efficient Modeling of Sound Source Radiation in Free-space and Room Environments*. PhD Thesis, Purdue University, West Lafayette, Indiana, USA, 2016.

[92] T. Shi. *Sound Field Reconstruction for an Under-determined System and Its Application*. PhD Thesis, Purdue University, West Lafayette, Indiana, USA, 2019.

[93] M.H. Lee and J.S. Bolton. Scan-based near-field acoustical holography and partial field decomposition in the presence of noise and source level variation. *The Journal of the Acoustical Society of America*, 119(1):382–393, 2006.

[94] M.D. Hayward. *Identification and Modification of Dominant Noise Sources in Diesel Engines*. Master's Thesis, Purdue University, West Lafayette, Indiana, USA, 2013.

[95] ANSI S1.13. *Measurement of Sound Pressure Levels in Air*. American National Standards Institute, ANSI S1.13, New York, USA, 1995.

APPENDICES

A. SOUND QUALITY METRICS AND OTHER MODELS

Researchers have developed sound quality metrics such as loudness, sharpness, roughness, fluctuation strength and tonality to quantify the strengths of sound attributes. These metrics were developed by observation of the human hearing system. The following sections describe detail calculation of sound quality metrics, a sound quality indicator, and psychoacoustic annoyance models.

A.1 Loudness

The "power law" is shown in Equation (A.1),

$$L = kI^p, \tag{A.1}$$

where L is perceived loudness, I is stimuli intensity, and k and p are constant. Based on Stevens' model and equal loudness contour, Zwicker's [63] and Glasberg's [64] have proposed the loudness models. Those models are standardized in ISO 532-1 [6] and ISO 532-2 [7]. In this research, both Zwicker's and Moore's models were calculated using Head ArtemiS software and the Zwicker's model was more accurate in the most of analysis.

A.2 Sharpness

Von Bismarck proposed the sharpness model [14] and Zwicker modified the model [15] and it is given by:

$$S = 0.11 \frac{\int_0^{24} N'(z)g(z)zdz}{N} \quad acum, \tag{A.2}$$

where N is overall loudness, $N^{'}$ is the specific loudness, and $g(z)$ is a weighting factor calculated as below:

$$g(z) = \left\{ \begin{matrix} 1 & for \;\; z \leq 16 \\ 0.066e^{0.171z} & for \;\; z > 16 \end{matrix} \right\}, \tag{A.3}$$

where z is the critical band rate in Bark. Aures' sharpness model [12] that emphasize high frequency and a weighting factor is calculated as below:

$$g(z) = 0.078 \frac{e^{0.0171z}}{z} \cdot \frac{N}{log_e(0.05N+1)}. \tag{A.4}$$

In this study, effects of both von Bismarck and Aures' model was examined.

A.3 Roughness

Zwicker proposed the model and it is given below [15]:

$$R = 0.3 f_{mod} \int_0^{24} \triangle L(z) dz \quad asper, \tag{A.5}$$

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

$$\triangle L(z) = 20log_{10}\left(\frac{N^{'}max(z)}{F_{N^{'}min(z)}}\right) \quad or \quad \triangle L(z) = 20log_{10}\left(\frac{N_1^{'}(z)}{F_{N_{99}^{'}(z)}}\right), \tag{A.6}$$

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

A.4 Fluctuation Strength

Equation (A.7) shows the broadband noise fluctuation strength model developed by Zwicker [15].

$$F_{BBN} = \frac{5.8(1.25m - 0.25)(0.05L_{BBN} - 1)}{(f_{mod}/5)^2 + (4/f_{mod}) + 1.5} \quad vacil, \tag{A.7}$$

where m is the modulation factor, L_{BBN} is the broadband noise level, and f_{mod} is the modulation frequency. The pure tone fluctuation strength model also developed by Zwicker is shown below [15]:

$$F = \frac{0.008 \int_0^{24} (\triangle L(z) dz)}{f_{mod}/4 + 4/f_{mod}} \quad vacil. \tag{A.8}$$

In this study, fluctuation strength metrics were calculated by both B&K sound quality package and Head ArtemiS program, but none of them were significant in the annoyance prediction models.

A.5 Tonality

Tone-to-Noise Ratio is the ratio of the power contained in the tone to the power contained in the critical band centered on that tone. The Tone-to-Noise Ratio can be calculated by using:

$$TNR = 10log_{10}(W_t/W_n) \quad dB, \tag{A.9}$$

where W_t is the power of the tone and W_n is the power of the masking noise. The power of the masking noise W_n can be defined by:

$$W_n = (W_{tot} - W_t)\frac{\triangle f_c}{\triangle f_{tot} - \triangle f_t}, \tag{A.10}$$

where W_{tot} is the total power in the critical band centered on the tone, $\triangle f_c$ is the bandwidth of the tonal component, and $\triangle f_{tot}$ is the width of the frequency band used to compute W_{tot}. The bandwidth of the tonal component $\triangle f_c$ is given as follow:

$$\triangle f_c = 25 + 75[1 + 1.4(f_t/1000)^2]^{0.69} \quad Hz. \tag{A.11}$$

The start (Cut-on) and end (Cut-off) point of the critical band can be calculated by using:

$$f_1 = -\frac{\triangle f_c}{2} + \frac{\sqrt{(\triangle f_c)^2 + 4f_t^2}}{2}, \tag{A.12}$$

and

$$f_2 = f_1 + \triangle f_c. \tag{A.13}$$

Additional calculation procedures are required if several tones exist in one critical band. The power of the tone must be removed from the masking power for accurate calculation. A tone is considered prominent if its TNR is greater than 6 dB [95].

Prominence Ratio (PR) is the ratio of the power contained in the critical band centered on the tone under investigation to the average power contained in the two adjacent critical bands. The Prominence Ratio is defined by:

$$PR = 10log_{10}\left(\frac{W_M}{(W_L + W_U)/2}\right), \quad dB \tag{A.14}$$

where W_M is the power in the critical band centered on the tone of interest, W_L is the power in the lower adjacent critical band, and W_U is the power in the higher adjacent critical band. The bandwidth can be calculated using Equation (A.11). A tone is considered to be prominent when the Prominence Ratio exceeds 7 dB. If multiple tones are in the sound, the Prominence Ratio needs to be calculated individually and choose the highest one as Prominence Ratio.

Tonal Audibility measures the prominence of tones in sounds and it is defined as:

$$\triangle L_{ta} = L_{pt} - L_{pn} + 2 + log_{10}\left(1 + \left(\frac{f_c}{502}\right)^2\right) \quad dB, \tag{A.15}$$

where L_{pt} is the total sound pressure level of tones in the critical band of interest, 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 in Hz. The following equation is used for calculating the total sound pressure level L_{pt},

$$L_{pt} = 10log_{10}\sum 10^{\frac{L_{pti}}{10}} \quad dB, \tag{A.16}$$

where L_{pti} is the sound pressure level of i-th tone in the critical band. The masking noise total sound pressure level (L_{pn}) is determined from the average sound pressure level within the band and the equation is shown below:

$$L_{pn} = L_{pn,avg} + 10log_{10}\left(\frac{CBW}{EAB}\right) \quad dB, \tag{A.17}$$

where EAB stands for effective analysis bandwidth and is 1.5 times the frequency resolution in the spectrum when a Hann window is used. CBW is the critical bandwidth and is dependent on the center frequencies of the critical bands. Also, the average noise level ($L_{pn,avg}$) can be calculated by disregarding the tonal components in the spectrum. Tonal Audibility measurement is used to correct the average A-weighted sound pressure level.

Aures proposed a model for the tonalness of the sound [12]. In his model, the effect of bandwidth, center frequency, level and the ratio of tonal loudness to noise loudness are considered. The equation is as follow:

$$T = c \cdot w_T^{0.29} \cdot w_{Gr}^{0.79}, \tag{A.18}$$

where c is a constant, w_T is a weighting function that represents the contribution of the tonal components, and w_{Gr} is a weighting that reflects the influence of noise on the overall sound. The tonal component weighting function w_T includes the influence of bandwidth, center frequency, and level. w_T can be expressed as:

$$w_T = \sqrt{\sum_{i=1}^{n}[w_1'(\triangle z_i)w_2'(f_i)w_3'(\triangle L_i)]^2}, \tag{A.19}$$

where $w_1^{'} = w_1^{1/0.29}$, $w_2^{'} = w_2^{1/0.29}$, and $w_3^{'} = w_3^{1/0.29}$. The influence of bandwidth is defined as:

$$w_1(\triangle z) = \frac{0.13}{\triangle z + 0.13}, \tag{A.20}$$

where $\triangle z$ is the bandwidth of the tone in Bark. The influence of center frequency is calculated by using:

$$w_2(f) = \left(\frac{1}{\sqrt{1 + 0.2\left(\frac{f}{700} + \frac{700}{f}\right)}} \right)^{0.29}, \tag{A.21}$$

where f is the center frequency of the tonal component in Hz. The influence of level is expressed as:

$$w_3(\triangle L) = (1 - e^{-\frac{\triangle L}{15}})^{0.29}, \tag{A.22}$$

where $\triangle L$ is the excess level of the tonal component in dB. The excess level of the i-th component with frequency of f_i is calculated using following equation.

$$\triangle L_t = L_t - log_{10}\left\{\left[\sum_{k\neq i}^{n} A_{Ek}(f_i))\right]^2 + E_{Gr}(f_i) + E_{Hs}(f_i)\right\} \quad dB, \tag{A.23}$$

where $A_{E_k}(f_i)$ is the excitation level which is produced at the frequency f_i by the i-th tone, $E_{Gr}(f_i)$ is the noise intensity present in the critical band around the considered tonal component, and $E_{Hs}(f_i)$ indicates hearing threshold at the frequency f_i. The weighting that reflects the influence of noise, w_{Gr} is defined as:

$$w_{Gr} = 1 - \frac{N_{Gr}}{N}, \tag{A.24}$$

where N_{Gr} is the loudness of the noise component and N is the total loudness of the sound.

A.6 Rate of Change of the Loudness

Marshall [16] used it when examining transient sounds. The Zwicker time varying Loudness are calculated from the Head ArtemiS software and a MATLAB 121-point FIR filter differentiator was used to calculate the derivatives. The rate of change of Loudness exceeded 2% of the time was examined as a potential impulsiveness measure. In this research, RCL exceeded 2% of the time improved the prediction of the refrigerated truck sounds' annoyance ratings.

A.7 Sound Quality Indicator

The first step of the sound quality indicator calculation is to get one-third octave band data. The level of one-third octave band data which exceeds more than 1.5 dB to the average of two adjacent bands is adjusted arithmetically. The tone adjusted sound level are calculated using this equation:

$$L^{'} = L - P + 10log_{10}(10^{(D+B)} + 1), \tag{A.25}$$

where, P is a projection above the average of the two adjacent bands in dB, L is the measured sound level for the band in dB, $L^{'}$ is a tone adjusted sound level for the band in dB, and F is a the octave band center frequency from 125 Hz to 8000 Hz. And,

$$\begin{aligned} &D = log_{10}(10^{(P/10)} - 1), \\ &B = 76.28 - 75.74Y + 29.98Y^2 - 6.14Y^3 + 0.69Y^4 - 0.04Y^5 + 0.001Y^6, \\ &Y = ln(F). \end{aligned} \tag{A.26}$$

Adjusted levels are converted to rating indices based on conversion table in the ANSI/AHRI 1140 standard. The formula for SQI is expressed by:

$$\begin{aligned} &SQI = K + 10 \times log \sum_{i=100Hz}^{10000Hz} I_i, \\ &K = 11.83888 - 4.94569lnX + 0.614812(lnX)^2, \\ &X = \frac{\sum I_i}{I_m}, \end{aligned} \tag{A.27}$$

where I_m is the maximum one-third octave band rating index from 100 to 10000 Hz. The SQI can be calculated by using both sound power and sound pressure level. In this report, the Sound Quality Indicator calculated using sound pressure level is expressed as SQI*. Even though a Sound Quality Indicator is not widely used in HVAC&R industry, this metric is a good predictor especially in annoyance prediction of residential units (Chapter 6).

A.8 Psychoacoustic Annoyance Model

Zwicker proposed psychoacoustic annoyance model which is used in transportation noise [15]. The psychoacoustic annoyance model is calculated by using Equation (A.28):

$$PA = N_5[1 + \sqrt{w_s^2 + w_{FR}^2}], \tag{A.28}$$

where,

$$w_s = \begin{cases} 0.25(S - 1.75)log_{10}(N_5 + 10) & for \quad S > 1.75 \\ 0 & for \quad S < 1.75 \end{cases}, \tag{A.29}$$

and,

$$w_{FR} = \frac{2.18}{(N_5)^{0.4}}(0.4F + 0.6R). \tag{A.30}$$

S is Sharpness, F is Fluctuation Strength, R is Roughness and N_5 is Loudness exceeded 5% of the time. More and Davies have modified model to consider tonality metrics and it is shown as follow [78]:

$$PA_{mod} = N_5(1 + \sqrt{\gamma_0 + \gamma_1 w_s^2 + \gamma_2 w_{FR}^2 + \gamma_3 w_T^2}), \tag{A.31}$$

where the additional tonality term is defined as:

$$w_T^2 = [(1 - e^{-\gamma_4 N_5})^2 (1 - e^{-\gamma_5 K_5})^2]. \tag{A.32}$$

K_5 is Aures' Tonality and the coefficients were developed based on responses from subjective tests.

B. SIGNAL MODIFICATION PROGRAM

This Appendix contains the programs written in MATLAB for signal decomposition and modification. The function signaldecomp.m gives the tone removed signal and addmodulation.m adds modulation to the input signal.

```
function [toneremovedsignal]=signaldecomp(pres,t,fs, MPD, MPH
    , MPP, NT)
% This function decompose HVACR signal
% into tone and broadband components

% Inputs:
%       pres - pressure time histories
%       t - time
%       fs - sampling frequency
%       Tone selection parameters
%           MPD - min. peak distance
%           MPH - min. peak height
%           MPP - min. peak prominence
%       NT - number of tones want to remove
% Outputs
%       toneremovedsignal - tonal component removed signal
% Written by: Weonchan Sung

% low pass filter - anti-aliasing filter
Fp=(3000)/(fs/2);
Fs=(4500)/(fs/2);
```

```
[n_LP,Wn_LP]=buttord(Fp,Fs,2,15);
[B,A]=butter(n_LP,Wn_LP);
pres_low=filtfilt(B,A,pres);

% high frequency component of the signal
pres_high=pres-pres_low;

% re-sampling - avoid stability issue in IIR filter design
n_sampleratio=5;
pres_low_down=downsample(pres_low,n_sampleratio);
t_down=downsample(t,n_sampleratio);
fsnew=fs/n_sampleratio;
dtnew=1/fsnew;

% power spectral density
Nfftnew=2^round(log2(1.*fsnew)); %NFFT
[Pxx_PSD,Freq_PSD]=pwelch(pres_low_down,hann(Nfftnew),...
    Nfftnew/2,Nfftnew,fsnew); % 50% overlap
Pxx_PSD_dB=10*log10(Pxx_PSD./(4e-10)); %Unit of PDS : dB ref
    4e-10 Pa^2

% find peaks of the signal
% minpeakdistance (MPD), minpeakheight (MPH)
% and minpeakprominence (MPP)
% should be decided by the spectral shape
[pks,locs]=findpeaks(Pxx_PSD_dB, 'MinPeakDistance', MPD,...
    'MinPeakHeight', MPH, 'MinPeakProminence', MPP);
Nf=length(locs);
f=zeros(Nf,1);
```

```
f_locs=zeros(Nf,1);
    for i=1:Nf
        f_locs(i)=locs(i);
        f(i)=Freq_PSD(locs(i));
    end

% remove tone below 40Hz
jj=1;
    for kk=1:length(f)
        if f(kk)>=40
            newf(jj)=f(kk);
            newpks(jj)=pks(kk);
            newloc(jj)=f_locs(kk);
            jj=jj+1;
        else
            jj=jj;
        end
    end
pksdecend=sort(newpks,'descend');kkk=1;
    for iii=1:length(pksdecend)
        for jjj=1:length(pksdecend)
            if pksdecend(iii)==Pxx_PSD_dB(newloc(jjj))
                newfdecend(kkk)=Freq_PSD(newloc(jjj));
                kkk=kkk+1;
            else
                kkk=kkk;
            end
        end
    end
```

```
% bandpass filter and calculate instantaneous phase
% for tone reconstruction
    for ntc=1:NT
        [n,Wn] = myBPfilter(aa(ntc),bb(ntc),cc(ntc),...
            dd(ntc),fsnew,Rp(ntc),Rs(ntc));
        [B,A]=butter(n,Wn);
        pres_low_down_bp=filtfilt(B,A,pres_low_down);

        % Hilbert transform
        n=255;a1=.01;a2=.99;f0=1;f1=450;
        h=firpm(n-1,[a1 a2],[1 1],'hilbert');
        % Show the shape of Hilbert tranform
        N2=(n-1)/2;
        tn=(-N2:N2)/fsnew;
        f_hil=[0 a1*(fsnew/2) a2*(fsnew/2) fsnew/2];
        Alt=[.0001 1 1 .0001];
        % Check frequency response
        NP2=round(log2(8*n)); NP2=2^NP2;
        H_freq=fft(h, NP2);
        freq=(0:NP2-1)*fsnew/NP2;

        %Compensate for -nst*DELTA start
        H_freq=H_freq.*exp(1i*2*pi*freq*N2/fsnew);

        % Get x_hat and Instantaneous Amplitude
        x_hat=conv(h,pres_low_down_bp);
        x_hat=x_hat(((n-1)/2)+1:end-((n-1)/2));
```

```
            Amp=sqrt((pres_low_down_bp.^2)+(x_hat.^2)); %
                INSTANTANEOUS AMPLITUDE
            a=pres_low_down_bp+1i*x_hat;
            phi=angle(a);
            phi=unwrap(phi);

            instamp(:,ntc)=Amp;
            phimat(:,ntc)=phi;
        end
% tone estimation
Aloop9=zeros(length(phimat),2*NT);
for ii=1:NT
            Aloop9(:,2*ii-1)=sin(phimat(:,ii));
            Aloop9(:,2*ii)=cos(phimat(:,ii));
end
Py9=regress(pres_low_down,Aloop9,0.01);
tear=Aloop9*Py9;
toneremovedsignal=pres_low_down-tear;
end
```

```
function [RoughSig]=addmodulation(pres,fc,fs,gamma)
% This function adds modulation to the signal

% Inputs:
%       pres - pressure time histories
%       fc - center frequency [Hz]
%       fs - sampling frequency [Hz]
%       gamma - amplitude of modulation frequency
% Outputs
%       RoughSig - Modulation added signal
% Written by: Weonchan Sung

% Center frequency
for ii=1:length(pres)
    fr(ii,1)=fc;
end
% time vector
t=0:1/fs:(length(pres)-1)/fs; t=t';

% Add modulation
RoughSig=pres.*(gamma*cos(2*pi*fr.*t)+1);

% PSD
Nfft=2^round(log2(1.*fs)); %NFFT
[Pxx,Freq]=pwelch(pres,hann(Nfft),Nfft/2,Nfft,fs); % 50%
    overlap
Pxx_dB=10*log10(Pxx./(4e-10)); %Unit of PDS : dB ref 4e-10 Pa
    ^2
```

```
[Pxx_newRN,Freq_newRN]=pwelch(RoughSig,hann(Nfft),Nfft/2,Nfft
    ,fs); % 50% overlap
Pxx_newRN_dB=10*log10(Pxx_newRN./(4e-10)); %Unit of PDS : dB
    ref 4e-10 Pa^2

% Plot results
plot(Freq_newRN,Pxx_newRN_dB, 'r');hold on; plot(Freq,Pxx_dB,
    'k');
xlim([0 fs/2]);
title('Red-Rougher');
ylabel('PSD dB ref {(20\muPa)}^{2}/Hz');
xlabel('Frequency (Hz)');
end
```

C. TEST SIGNALS

Descriptions of sounds used in Test 1, Test 2 and Test 3 are shown in the Table C.1. The signal number in each test, signal name, unit type, microphone location, modification type, and corresponding wave file names are given in the Table C.1. In Table C.2, sound quality metric values are given. The second line of the 1st row is the unit of the metric.

Table C.1. Test Sounds. Column 1 to 5 show the signal number used in each test and Column 6 is description of the signal. Column 7 shows the modification type and Column 8 shows the figure number in this report.

Signal Number					**Signal**	**Unit Type**	**Mod.**	**Wavefile**
T1	**T2**	**T3 A**	**T3 B**	**T3 C**	**Name**		**Type**	**Names**
				13	001	Residential, fan 452 rpm, compr. 1800 rpm, MIC 3	Loudness	S1
1		8			003	Residential, fan 452 rpm, compr. 1800 rpm, MIC 1	Loudness	S3
		11			005	Residential, fan 568 rpm, compr. 2475 rpm, MIC 1	Loudness	S5
		12			007	Residential, fan 452 rpm, compr. 1800 rpm, MIC 2	Loudness	S7
2		13	11	25	008	Residential, fan 452 rpm, compr. 1800 rpm, MIC 4	Loudness	S8
		4	8	14	011	Residential, fan 692 rpm, compr. 3150 rpm, MIC 4	Loudness	S11
	1	5	9	15	012	Residential, fan 692 rpm, compr. 3150 rpm, MIC 3	Loudness	S12
		6			013	Residential, fan 895 rpm, compr. 3825 rpm, MIC 4	Loudness	S13
				16	015	Residential, fan 692 rpm, compr. 3150 rpm, MIC 1	Loudness	S15
3	2			17	017	Residential, fan 895 rpm, compr. 3825 rpm, MIC 1	Loudness	S17
		7	10	18	019	Residential, fan 1100 rpm, compr. 4500 rpm,, MIC 4	Loudness	S19
				19	021	Residential, fan 1100 rpm, compr. 4500 rpm, MIC 3	Loudness	S21
4	3			20	022	Residential, fan 1100 rpm, compr. 4500 rpm, MIC 1	Loudness	S22
				21	025	Residential, fan 814 rpm, compr. 3600 rpm, MIC 2	Loudness	S25
	4				030	Residential, fan 814 rpm, compr. 3600 rpm, MIC 3	Loudness	S30
5				22	032	Residential, fan 692 rpm, compr. 3150 rpm, MIC 5	Loudness	S32

Table C.1. Continued from previous page.

		9			033	Residential, fan 1100 rpm, compr. 4500 rpm, MIC 5	Loudness	S33
	5				034	Residential, fan 1100 rpm, compr. 3600 rpm, MIC 4	Loudness	S34
6				23	035	Residential, fan 1100 rpm, compr. 3600 rpm, MIC 2	Loudness	S35
		10			039	Residential, fan 814 rpm, compr. 3600 rpm, MIC 5	Loudness	S39
				24	040	Residential, fan 1100 rpm, compr. 3600 rpm, MIC 5	Loudness	S40
		45	45	44	042	Residential, SPP Units, MIC 2	-	S42
7					057	Residential, fan 692 rpm, compr. 3150 rpm, MIC 1	-	S57
	6			41	060	Residential, fan 452 rpm, compr. 1800 rpm, MIC 5	-	S60
				42	066	Residential, fan 814 rpm, compr. 3600 rpm, MIC 2	-	S66
	7				074	Residential, fan 568 rpm, compr. 2475 rpm, MIC 5	-	S74
		38			075	Residential, fan 1100 rpm, compr. 4500 rpm, MIC5	-	S75
8		39			076	Residential, fan 1100 rpm, compr. 3600 rpm, MIC 4	-	S76
9		40			077	Residential, fan 1100 rpm, compr. 3600 rpm, MIC 2	-	S77
	8	41			078	Residential, fan 1100 rpm, compr. 3600 rpm, MIC 3	-	S78
		42			079	Residential, fan 1100 rpm, compr. 3600 rpm, MIC 1	-	S79
		43			081	Residential, fan 814 rpm, compr. 3600 rpm, MIC 5	-	S81
10			39		082	Residential, fan 1100 rpm, compr. 3600 rpm, MIC 5	-	S82
	9			5	083	CTD, Vector 1950, ANSI/AHRI 1120, 1450 rpm, MIC 6	Loudness	S83
				6	084	CTD, Vector 1950, TNO, 1450 rpm, MIC 6	Loudness	S84

Table C.1. Continued from previous page.

				7	087	CTD, Vector 1350, TNO, 1800 rpm, MIC 5	Loudness	S87
			1		088	CTD, Vector 1350, ANSI/AHRI 1120, 1800 rpm, MIC 6	Loudness	S88
				9	089	CTD, Vector 1950, TNO, 1450 rpm, MIC 1	Loudness	S89
11	10				090	CTD, Vector 1350, ANSI/AHRI 1120, 1800 rpm, MIC 5	Loudness	S90
			2		091	CTD, Vector 1350, ANSI/AHRI 1120, 1800 rpm, MIC 1	Loudness	S91
			3		093	CTD, Vector 1350, ANSI/AHRI 1120, 2250 rpm, MIC 6	Loudness	S93
12					094	CTD, Vector 1950, ANSI/AHRI 1120, 1450 rpm, MIC 6	Loudness	S94
13	11				097	CTD, Vector 1350, TNO, 2250 rpm, MIC 6	Loudness	S97
				1	100	CTD, Vector 1950, ANSI/AHRI 1120, 1850 rpm, MIC 1	Loudness	S100
				2	105	CTD, Vector 1350, TNO, 1800 rpm, MIC 4	Loudness	S105
				3	114	CTD, Vector 1350, ANSI/AHRI 1120, 1800 rpm, MIC 4	Loudness	S114
				4	121	CTD, Vector 1350, TNO, 1800 rpm, MIC 3	Loudness	S121
14					134	CTD, Vector 1350, ANSI/AHRI 1120, standby motor, 2250 rpm, MIC 1	Loudness	S134
			12		141	CTD, Vector 1950, ANSI/AHRI 1120, 1450 rpm, MIC 6	-	S141
15	12		13		142	CTD, Vector 1950, TNO, 1450 rpm, MIC 6	-	S142
16		19	14	26	143	CTD, Vector 1950, TNO, 1450 rpm, MIC 5	-	S143
17		20	15	27	144	CTD, Vector 1350, TNO, 1800 rpm, MIC 6	-	S144
				28	145	CTD, Vector 1350, TNO, 1800 rpm, MIC 5	-	S145

Table C.1. Continued from previous page.

			16		146	CTD, Vector 1350, ANSI/AHRI 1120, 1800 rpm, MIC 6	-	S146
			17		147	CTD, Vector 1950, TNO, 1450 rpm, MIC 1	-	S147
		21			148	CTD, Vector 1350, ANSI/AHRI 1120, 1800 rpm, MIC 5	-	S148
			18		149	CTD, Vector 1350, ANSI/AHRI 1120, 1800 rpm, MIC 1	-	S149
18		22	19	29	151	CTD, Vector 1350, ANSI/AHRI 1120, 2250 rpm, MIC 6	-	S151
		23	20	30	152	CTD, Vector 1950, ANSI/AHRI 1120, 1450 rpm, MIC 5	-	S152
			21		154	CTD, Vector 1350, TNO, 2250 rpm, MIC 5	-	S154
		24	22	31	155	CTD, Vector 1350, TNO, 2250 rpm, MIC 6	-	S155
		25	23	32	156	CTD, Vector 1950, ANSI/AHRI 1120, 1450 rpm, MIC 1	-	S156
			24		157	CTD, Vector 1350, TNO, 2250 rpm, MIC 1	-	S157
				33	158	CTD, Vector 1950, ANSI/AHRI 1120, 1850 rpm, MIC 1	-	S158
		26	25	34	162	CTD, Vector 1950, ANSI/AHRI 1120, 1850 rpm, MIC 6	-	S162
			26		164	CTD, Vector 1950, ANSI/AHRI 1120, 1850 rpm, MIC 5	-	S164
		27			166	CTD, Vector 1950, ANSI/AHRI 1120, 1450 rpm, MIC 4	-	S166
	13		27		170	CTD, RG Unit E, ANSI/AHRI 1120, 50 Hz full load, MIC 1	-	S170
	14		28		171	CTD, Vector 1950, TNO, 1850 rpm, MIC 6	-	S171
				35	172	CTD, Vector 1350, ANSI/AHRI 1120, 1800 rpm, MIC 4	-	S172
			29		173	CTD, Vector 1350, ANSI/AHRI 1120, 1800 rpm, MIC 2	-	S173
			30		175	CTD, RG Unit D, ANSI/AHRI 1120, 50 Hz full load, MIC 1	-	S175

Table C.1. Continued from previous page.

19					177	CTD, Vector 1950, ANSI/AHRI 1120, 1450 rpm, MIC 2	-	S177
			31		181	CTD, Vector 1350, ANSI/AHRI 1120, road only, 1600 rpm, MIC 1	-	S181
20	15				182	CTD, Vector 1350, ANSI/AHRI 1120, road only, 1800 rpm, MIC 1	-	S182
			32		186	CTD, Vector 1950, ANSI/AHRI 1120, 1850 rpm, MIC 2	-	S186
			33		187	CTD, Vector 1350, ANSI/AHRI 1120, 2250 rpm, MIC 2	-	S187
				36	188	CTD, Vector 1350, ANSI/AHRI 1120, 2250 rpm, MIC 4	-	S188
21					189	CTD, Vector 1950, TNO, 1850 rpm, MIC 2	-	S189
			34		191	CTD, Vector 1350, ANSI/AHRI 1120, 2250 rpm, MIC 3	-	S191
			35		195	CTD, Vector 1350, ANSI/AHRI 1120, road only, 2250 rpm, MIC 1	-	S195
	16		36		196	CTD, RG Unit D, ANSI/AHRI 1120, 60 Hz full load, MIC 1	-	S196
			37		197	CTD, RG Unit E, ANSI/AHRI 1120, 60 Hz full load, MIC 1	-	S197
				37	198	CTD, Vector 1350, TNO, 2250 rpm, MIC 1	-	S198
22					199	Residential, fan 692 rpm, compr. 3150 rpm, MIC 3	Roughness	S199
23					200	Residential, fan 1100 rpm, compr. 1800 rpm, MIC 1	Roughness	S200
				9	204	CTD, Vector 1350, ANSI/AHRI 1120, 1800 rpm, MIC 5	Roughness, Tonality	S204
		1			207	CTD, Vector 1950, TNO, 1450 rpm, MIC 6	Roughness, Tonality	S207

Table C.1. Continued from previous page.

24		14			209	Residential, fan 1100 rpm, compr. 3600 rpm, MIC 3	Roughness	S209
25					212	Residential, fan 447 rpm, MIC 2	-	S212
				38	217	Residential, fan 570 rpm, MIC 2	-	S217
		28			225	Residential, fan 697 rpm, MIC 5	-	S225
		29			226	Residential, fan 897 rpm, MIC 1	-	S226
		30			230	Residential, fan 897 rpm, MIC 5	-	S230
		31			231	Residential, fan 997 rpm, MIC 1	-	S231
		32			232	Residential, fan 997 rpm, MIC 2	-	S232
		33			233	Residential, fan 997 rpm, MIC 3	-	S233
				39	234	Residential, fan 997 rpm, MIC 4	-	S234
	18	34	38	40	235	Residential, fan 997 rpm, MIC 5	-	S235
26					237	CTD, Vector 1350, TNO. 1800 rpm, MIC 1	Tonality, Roughness	S237
			4		241	CTD, Vector 1950, ANSI/AHRI 1120, 1850 rpm, MIC 4	Tonality, Roughness	S241
				10	242	CTD, RG Unit D, ANSI/AHRI 1120, 60 Hz full load, MIC 1	Tonality, Roughness	S242
		19		11	243	CTD, RG Unit D, ANSI/AHRI 1120, 60 Hz full load, MIC 1	Tonality, Roughness	S243

Table C.1. Continued from previous page.

27					245	Residential, fan 1100 rpm, compr. 4500 rpm, MIC 3	Tonality, Roughness	S245
28					247	Residential, fan 895 rpm, compr. 3825 rpm, MIC 5	Roughness	S247
		15			257	Residential, fan 897 rpm, MIC 3	Tonality, Roughness	S257
				12	264	CTD, Vector 1350, TNO, 2250 rpm, MIC 5	Tonality, Roughness	S264
29			5		266	CTD, Vector 1350, ANSI/AHRI 1120, 2250 rpm, MIC 6	Tonality	S266
30					269	CTD, Vector 1350, ANSI/AHRI 1120, 2250 rpm, MIC 6	Tonality	S269
		16			277	Residential, fan 895 rpm, compr. 3825 rpm, MIC 5	Tonality, Roughness	S277
31					278	CTD, Vector 1350, ANSI/AHRI 1120, 2250 rpm, MIC 2	Roughness	S278
		17			279	Residential, fan 897 rpm, MIC 3	Tonality, Roughness	S279
32					281	Residential, fan 447 rpm, MIC 3	Tonality	S281
			6		282	CTD, Vector 1350, TNO, 1800 rpm, MIC 1	Tonality, Roughness	S282
	20	2			283	CTD, Vector 1350, TNO, 1800 rpm, MIC 1	Tonality, Roughness	S283

Table C.1. Continued from previous page.

			7		285	CTD, RG Unit D, ANSI/AHRI 1120, 60 Hz full load, MIC 1	Tonality, Roughness	S285
33	21				287	Residential, fan 692 rpm, compr. 3150 rpm, MIC 5	Tonality, Roughness	S287
		18			288	Residential, fan 897 rpm, MIC 1	Tonality, Roughness	S288
34					290	Residential, fan 1100 rpm, compr. 4500 rpm, MIC 5	Tonality, Roughness	S290
35					292	Residential, fan 1100 rpm, compr. 3600 rpm, MIC 5	Tonality, Roughness	S292
36	22				293	Residential, fan 895 rpm, compr. 3825 rpm, MIC 2	Tonality, Roughness	S293
		3			294	CTD, Vector 1950, ANSI/AHRI 1120, 1850 rpm, MIC 5	Tonality, Roughness	S294
		35			301	Residential, Unit D, MIC 3	-	S301
		36			306	Residential, Unit E, MIC 3	-	S306
		37			307	Residential, Unit F, MIC 1	-	S307
				45	311	Residential, fan 895 rpm, compr. 3825 rpm, MIC 1	Sharpness	S311
		46	46	46	312	Residential, fan 895 rpm, compr. 3825 rpm, MIC 1	Sharpness	S312

Table C.1. Continued from previous page.

				47	314	Residential, fan 1100 rpm, compr. 4500 rpm, MIC 1	Sharpness	S314
		47			316	Residential, fan 1100 rpm, compr. 3600 rpm, MIC 4	Sharpness	S316
		48			317	Residential, fan 814 rpm, compr. 3600 rpm, MIC 5	Sharpness	S317
			40		318	CTD, Vector 1350, TNO, 1800 rpm, MIC 6	Sharpness	S318
			41		319	CTD, Vector 1350, TNO, 1800 rpm, MIC 6	Sharpness	S319
		44			320	CTD, Vector 1950, ANSI/AHRI 1120, 1450 rpm, MIC 5	Sharpness	S320
			42		321	CTD, Vector 1350, TNO, 2250 rpm, MIC 6	Sharpness	S321
			43		322	CTD, Vector 1950, TNO, 1850 rpm, MIC 6	Sharpness	S322
				43	323	CTD, Vector 1950, TNO, 1850 rpm, MIC 6	Sharpness	S323
			44		325	CTD, Vector 1350, ANSI/AHRI 1120, road only, 1800 rpm, MIC 1	Sharpness	S325
			47		330	Residential, fan 1100 rpm, compr. 3600 rpm, MIC 5	Sharpness	S330
				50	339	Residential, fan 568 rpm, compr. 2475 rpm, MIC 3	Sharpness	S339
		50			340	CTD, Vector 1350, ANSI/AHRI 1120, 2250 rpm, MIC 6	Sharpness	S340
				49	342	CTD, Vector 1950, TNO, 1850 rpm, MIC 6	Sharpness	S342
			50		343	CTD, Vector 1350, TNO, 2250 rpm, MIC 6	Sharpness	S343
		49	48	48	351	Residential, fan 452 rpm, compr. 1800 rpm, MIC 4	Sharpness	S351
			49		352	Residential, fan 568 rpm, compr. 2475 rpm, MIC 4	Sharpness	S352

Table C.2. Values of the main sound quality metrics for the signals used in this research. Most metric values calculatd by using Head Acoustics ArtemiS software, except for SQI^*, T_{A5}, and RCL

Signal Number	N_5 sone	SPL_A dBA	SPL_C dBC	SQI^*	R_5 asper	T_5 tu	T_{A5}	S_{A5} acum	S_{VB5} acum	FS_5 vacil	RCL
001	28.7	70.1	78.9	23.2	2.54	0.19	0.00	2.78	1.42	0.018	134.1
003	25.5	68.8	77.1	23.1	2.19	0.21	0.22	2.41	1.29	0.019	32.9
005	32.1	72.5	78.7	23.7	2.61	0.21	0.29	2.82	1.40	0.012	338.6
007	21.7	67.6	76.0	23.0	1.79	0.31	0.35	2.32	1.31	0.019	87.0
008	34.7	75.0	82.9	25.1	2.30	0.31	0.38	2.53	1.23	0.018	378.5
011	31.2	73.8	78.4	24.1	2.61	0.41	0.40	2.67	1.37	0.011	142.6
012	28.1	71.7	76.6	23.8	2.19	0.40	0.39	2.67	1.40	0.011	352.2
013	34.0	76.0	79.3	24.7	2.86	0.32	0.30	3.04	1.47	0.009	78.4
015	23.6	68.2	72.7	22.7	2.12	0.42	0.40	2.56	1.40	0.010	197.0
017	27.2	72.0	74.9	23.5	2.59	0.39	0.29	2.80	1.46	0.009	444.4
019	30.3	73.1	79.3	23.6	2.70	0.37	0.31	3.01	1.50	0.008	46.6
021	20.5	68.8	71.1	22.2	2.37	0.34	0.37	2.51	1.43	0.006	245.6
022	38.6	78.3	80.2	25.1	3.75	0.34	0.27	3.22	1.48	0.008	87.8
025	33.5	73.0	78.7	24.2	3.10	0.05	0.08	2.78	1.32	0.014	162.2
030	48.8	78.2	84.1	25.7	3.76	0.05	0.00	3.14	1.28	0.015	92.6
032	44.7	79.7	86.3	26.4	2.64	0.26	0.40	2.98	1.35	0.024	328.9

Table C.2. Continued from previous page.

033	24.3	68.5	77.6	22.3	2.60	0.12	0.17	2.68	1.44	0.019	147.9
034	37.0	73.3	82.0	23.8	3.01	0.19	0.14	2.66	1.21	0.013	54.2
035	49.7	78.7	85.8	26.3	3.25	0.23	0.30	2.87	1.17	0.016	130.8
039	26.6	68.4	76.9	23.3	2.79	0.03	0.09	2.35	1.20	0.013	191.3
040	36.3	73.7	85.9	24.9	2.69	0.18	0.29	2.33	1.08	0.014	183.4
042	25.7	67.1	79.2	22.1	2.41	0.14	0.14	2.44	1.24	0.010	327.9
057	10.3	55.8	60.2	18.6	1.38	0.47	0.41	2.11	1.41	0.007	544.2
060	11.4	56.7	66.9	19.4	1.39	0.21	0.29	2.05	1.34	0.018	216.9
066	13.0	58.1	64.5	19.6	1.84	0.05	0.09	2.09	1.29	0.009	158.0
074	14.6	62.4	690	21.2	1.45	0.28	0.44	2.11	1.35	0.015	78.1
075	15.0	61.1	70.2	20.0	2.01	0.13	0.18	2.36	1.44	0.015	180.0
076	19.4	63.2	71.9	20.7	2.12	0.20	0.18	2.16	1.20	0.010	180.6
077	20.8	65.1	72.1	22.2	2.03	0.24	0.31	2.12	1.17	0.011	178.2
078	21.1	64.9	72.2	22.1	2.08	0.29	0.29	2.14	1.17	0.010	655.8
079	21.4	64.8	73.7	22.0	2.16	0.32	0.26	2.01	1.10	0.010	398.1
081	22.0	65.5	74.0	22.4	2.52	0.04	0.09	2.21	1.20	0.012	290.2
082	36.9	73.9	86.2	25.0	2.71	0.18	0.29	2.35	1.08	0.014	432.3
083	18.6	62.2	75.6	21.1	2.81	0.16	0.19	1.96	1.11	0.008	120.5
084	12.3	56.1	67.8	19.4	2.76	0.13	0.19	1.64	1.03	0.006	441.1

Table C.2. Continued from previous page.

087	14.2	60.0	69.8	19.6	2.30	0.07	0.13	1.70	1.07	0.017	98.1
088	29.0	70.7	79.9	23.5	5.60	0.18	0.21	1.89	1.00	0.012	243.7
089	21.0	64.5	83.4	21.0	2.31	0.15	0.24	1.74	0.98	0.010	2674.1
090	11.1	56.6	66.6	18.6	2.11	0.12	0.17	1.65	1.08	0.007	180.6
091	31.4	71.2	83.3	22.8	3.41	0.15	0.19	2.34	1.15	0.012	315.5
093	33.4	73.3	79.6	24.4	5.34	0.25	0.23	2.17	1.08	0.010	119.2
094	13.0	56.9	77.0	18.3	2.09	0.13	0.12	1.64	1.02	0.010	225.6
097	19.5	64.8	71.7	21.8	4.21	0.26	0.18	1.82	1.06	0.009	454.8
100	34.4	72.5	81.9	23.3	3.17	0.19	0.18	2.37	1.13	0.011	154.2
105	28.1	69.6	78.2	22.9	3.23	0.07	0.11	2.29	1.17	0.011	140.7
114	25.4	68.2	76.3	22.2	2.95	0.12	0.11	2.21	1.17	0.009	490.8
121	25.2	68.0	76.9	22.6	3.04	0.11	0.08	2.21	1.16	0.009	127.5
134	19.4	64.3	72.3	21.0	2.47	0.13	0.12	2.20	1.24	0.013	176.8
141	26.0	67.4	80.8	22.7	3.41	0.16	0.19	2.16	1.11	0.010	311.0
142	27.2	68.1	79.8	23.0	4.18	0.12	0.18	2.07	1.06	0.009	18.84
143	27.7	69.0	80.9	22.3	2.98	0.07	0.15	2.19	1.11	0.024	606.6
144	13.6	59.1	68.7	19.7	3.76	0.19	0.21	1.46	0.95	0.009	293.3
145	28.1	70.5	80.2	22.7	3.32	0.07	0.14	2.09	1.09	0.024	294.6
146	28.3	70.3	79.6	23.4	5.53	0.18	0.21	1.88	1.00	0.012	183.7

Table C.2. Continued from previous page.

147	28.3	68.9	87.8	22.4	2.64	0.15	0.24	1.90	0.98	0.011	2714.0
148	29.7	71.5	81.5	23.1	3.53	0.11	0.14	2.19	1.11	0.010	335.6
149	29.8	70.4	82.5	22.6	3.32	0.15	0.19	2.29	1.15	0.011	141.2
151	30.5	71.9	78.2	23.9	5.08	0.25	0.23	2.10	1.08	0.009	178.9
152	30.5	69.0	89.0	22.1	3.18	0.12	0.22	2.02	1.00	0.014	3247.8
154	32.3	72.4	81.0	23.1	3.31	0.12	0.21	2.31	1.13	0.010	297.9
155	32.5	72.8	79.7	24.2	5.55	0.26	0.23	2.15	1.07	0.011	292.4
156	32.8	70.1	88.4	23.0	3.31	0.10	0.21	2.23	1.06	0.014	2257.4
157	33.1	72.1	81.6	23.9	2.97	0.15	0.20	2.46	1.18	0.013	120.1
158	33.5	72.1	81.4	23.2	3.12	0.19	0.18	2.35	1.13	0.011	542.6
162	34.2	71.5	83.0	24.3	4.62	0.24	0.21	2.28	1.08	0.011	323.9
164	34.3	72.7	81.4	23.3	3.52	0.15	0.17	2.41	1.14	0.012	56.8
166	34.5	71.3	88.6	23.2	3.00	0.11	0.20	2.29	1.07	0.012	1941.8
170	35.4	72.8	85.4	24.2	3.15	0.13	0.20	2.60	1.21	0.013	287.0
171	35.5	71.8	83.6	24.3	4.52	0.25	0.22	2.27	1.07	0.012	217.1
172	35.6	73.4	81.5	23.8	3.53	0.12	0.11	2.47	1.16	0.010	177.9
173	36.0	73.3	84.3	24.2	3.64	0.10	0.16	2.52	1.18	0.010	120.4
175	36.5	72.8	84.6	24.1	3.35	0.15	0.22	2.66	1.21	0.014	282.6
177	37.5	72.3	91.0	23.1	3.10	0.14	0.14	2.35	1.06	0.011	283.8

Table C.2. Continued from previous page.

181	39.2	73.3	91.0	24.2	3.54	0.04	0.22	2.44	1.09	0.024	2122.4
182	39.4	74.2	91.0	24.2	3.84	0.05	0.14	2.40	1.08	0.040	111.9
186	41.5	75.0	84.2	24.2	3.42	0.21	0.19	2.61	1.14	0.012	244.7
187	42.1	75.8	84.9	25.0	3.84	0.23	0.21	2.82	1.22	0.010	239.8
188	42.1	75.8	83.8	24.7	3.45	0.23	0.22	2.69	1.17	0.011	360.5
189	42.3	75.6	83.4	24.3	3.68	0.21	0.11	2.63	1.15	0.014	171.3
191	42.5	76.4	83.6	24.8	3.52	0.18	0.19	2.95	1.27	0.011	123.8
195	46.2	76.5	91.4	24.7	3.92	0.15	0.25	2.68	1.11	0.020	850.5
196	46.6	76.2	88.8	25.2	3.66	0.13	0.23	2.83	1.17	0.013	500.0
197	46.3	77.0	89.1	25.2	3.70	0.13	0.28	2.65	1.11	0.016	197.2
198	46.6	76.9	87.5	25.2	3.59	0.23	0.25	2.84	1.18	0.011	181.2
199	11.6	58.1	65.2	19.3	2.89	0.38	0.31	1.94	1.29	0.016	124.4
200	8.4	52.0	62.5	17.5	1.97	0.24	0.21	1.74	1.20	0.012	225.8
204	36.7	74.3	85.9	24.2	3.60	0.17	0.24	2.02	0.97	0.011	215.4
207	16.8	61.6	77.0	20.3	2.88	0.24	0.29	1.38	0.82	0.007	1083.5
209	22.4	66.2	73.8	22.3	4.03	0.28	0.20	2.10	1.14	0.012	390.5
212	4.7	42.6	59.0	15.4	1.06	0.05	0.00	2.05	1.42	0.008	178.6
217	6.8	47.9	59.2	16.3	1.28	0.07	0.00	1.95	1.34	0.014	314.0
225	15.4	59.9	70.8	19.5	2.02	0.03	0.09	2.21	1.33	0.013	59.4

Table C.2. Continued from previous page.

226	15.1	60.0	67.8	20.3	1.93	0.07	0.12	2.17	1.30	0.009	258.0
230	23.2	66.4	77.1	21.9	2.52	0.03	0.10	2.42	1.30	0.012	90.3
231	18.5	63.0	71.7	21.0	2.11	0.11	0.16	2.21	1.26	0.009	98.3
232	16.5	61.5	67.4	20.9	2.01	0.08	0.11	2.30	1.34	0.009	71.9
233	17.8	62.6	69.4	21.1	2.13	0.12	0.11	2.23	1.28	0.009	68.9
234	17.0	61.8	68.7	20.9	2.05	0.05	0.13	2.25	1.31	0.009	293.4
235	28.3	69.5	79.6	23.0	2.79	0.05	0.10	2.50	1.26	0.013	161.1
237	40.1	70.3	90.2	22.8	4.71	0.08	0.18	1.85	0.93	0.027	132.0
241	34.0	72.4	80.8	23.2	5.55	0.14	0.15	2.40	1.16	0.080	185.6
242	38.8	74.0	89.8	24.6	2.85	0.47	0.40	2.07	0.94	0.012	364.8
243	30.4	69.4	82.9	22.8	4.78	0.15	0.21	2.32	1.14	0.012	180.3
245	23.1	70.5	74.5	22.7	4.61	0.19	0.21	2.42	1.35	0.016	621.0
247	38.3	70.4	90.8	22.5	4.64	0.22	0.22	2.24	1.09	0.027	215.2
257	18.7	63.4	73.2	21.0	1.96	0.25	0.34	1.90	1.09	0.009	157.8
264	45.7	78.5	90.1	24.8	3.68	0.24	0.35	2.09	0.92	0.012	521.7
266	27.5	70.5	74.9	23.7	4.77	0.16	0.17	2.20	1.16	0.009	384.1
269	46.6	78.2	86.1	25.7	4.10	0.17	0.11	2.83	1.19	0.010	80.5
277	26.2	74.4	78.6	24.3	2.09	0.24	0.38	2.54	1.37	0.014	311.1
278	25.6	74.2	80.0	24.0	1.55	0.26	0.57	1.43	0.85	0.043	294.7

Table C.2. Continued from previous page.

279	32.7	74.3	85.4	24.0	1.96	0.48	0.60	1.55	0.80	0.011	79.9
281	8.0	52.9	68.3	17.6	0.77	0.49	0.46	1.51	1.05	0.019	85.7
282	33.6	73.5	85.2	24.1	2.65	0.40	0.45	1.75	0.87	0.009	319.6
283	28.1	68.3	83.2	22.5	2.75	0.27	0.28	1.86	0.97	0.011	851.2
285	46.2	75.7	88.9	24.9	5.15	0.13	0.20	2.85	1.19	0.014	401.5
287	45.4	82.0	86.6	26.8	2.39	0.46	0.46	2.88	1.29	0.013	427.5
288	15.1	59.8	68.5	20.2	3.46	0.04	0.00	2.15	1.29	0.010	275.4
290	17.1	60.0	77.8	19.2	3.20	0.15	0.14	2.20	1.31	0.012	189.4
292	36.4	75.1	85.0	25.5	2.36	0.41	0.40	2.20	1.01	0.013	196.3
293	48.3	80.0	86.0	26.2	4.64	0.39	0.26	3.56	1.51	0.025	254.8
294	30.8	70.1	80.1	22.8	4.16	0.13	0.16	2.49	1.23	0.013	518.4
301	33.6	72.6	80.8	23.7	3.05	0.13	0.14	2.89	1.36	0.009	96.7
306	21.9	65.8	81.0	21.2	2.32	0.14	0.13	1.19	0.70	0.013	993.7
307	16.6	60.7	77.6	19.7	1.98	0.00	0.00	1.40	0.85	0.013	1802.4
311	27.2	72.6	77.4	23.7	2.58	0.38	0.29	2.37	1.26	0.009	417.2
312	27.1	70.9	73.1	24.0	2.51	0.39	0.26	3.24	1.64	0.008	75.4
314	38.7	79.2	81.2	25.2	3.81	0.34	0.26	2.92	1.37	0.009	258.6
316	19.3	62.9	70.3	21.1	2.09	0.19	0.17	2.46	1.34	0.010	217.5
317	21.9	66.0	72.0	22.9	2.52	0.04	0.09	2.51	1.34	0.010	340.1

Table C.2. Continued from previous page.

318	28.1	69.7	78.0	23.4	5.54	0.16	0.17	2.14	1.11	0.012	120.8
319	28.1	71.0	81.2	22.9	5.44	0.17	0.19	1.54	0.85	0.012	157.4
320	30.7	69.6	90.4	22.3	3.09	0.13	0.24	1.73	0.88	0.015	3522.1
321	32.5	73.5	81.7	24.0	5.41	0.28	0.25	1.76	0.91	0.013	615.4
322	35.5	72.5	85.6	23.6	4.31	0.27	0.24	1.90	0.92	0.013	427.8
323	35.5	72.7	80.1	25.3	4.78	0.21	0.21	2.85	1.31	0.011	152.4
325	39.4	74.0	88.2	24.8	3.86	0.04	0.19	2.98	1.30	0.037	132.9
330	36.8	73.2	84.3	24.7	2.73	0.17	0.27	2.75	1.23	0.012	157.8
339	39.6	74.8	82.8	24.5	3.20	0.15	0.14	3.41	1.55	0.014	419.9
340	27.1	70.5	73.8	23.3	4.26	0.13	0.00	2.19	1.16	0.010	237.5
342	11.8	55.2	65.1	19.3	2.31	0.36	0.32	1.64	1.04	0.007	154.8
343	32.0	73.2	79.4	24.8	5.39	0.29	0.26	2.11	1.06	0.011	92.5
351	34.6	74.2	81.9	25.0	2.31	0.30	0.37	2.78	1.34	0.018	218.9
352	43.1	78.2	85.3	25.7	3.10	0.28	0.35	2.74	1.25	0.015	158.7

D. INSTRUCTIONS PROVIDED TO SUBJECTS IN PSYCHOACOUSTIC TESTING

Instructions that were provided to the subjects prior to the subjective test are given in the following pages. The instructions include actual user interface and test sheet provided to subjects in the test are also given.

D.1 Rating Tests' Instruction

You will listen to 50 HVAC&R (Heating, Ventilating, Air-conditioning, and Refrigeration) equipment sound. Each sound will last 4 seconds. After each sound has played you will be asked to make a mark on a line that reflects how annoying you find the sound. The rating scale looks like Figure like D.1.

You can move the marker or click on a position on the line using the mouse.

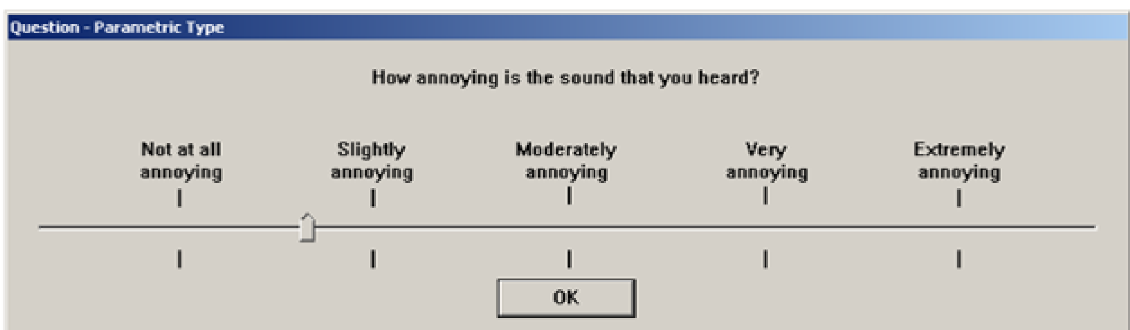


Figure D.1. Annoyance scale and graphical user interface used in the test.

While you are listening, it may be helpful to imagine yourself in your garden, at any time during the day or evening, hearing these sounds continuously. There are no right or wrong responses; we are only interested in your opinion of the sound. We recommend selecting an answer based on your "gut instinct" rather than thinking a lot about the sounds that you hear.

Before the actual test, you will be listening 10 familiarization sounds and a practice session where you will rate 2 sounds.

D.2 Description Test

You will hear sounds and after or while each sound is being played we want you to write down words or phrases that describe the sound. You will listen to 50 HVAC&R (Heating, Ventilating, Air-conditioning, and Refrigeration) equipment sounds. Each sound will last 4 seconds. There will be a 4 seconds pause, then you will hear the sound again.

As soon as the sound starts the first time, you can start writing down words that describe the sound in the sheet which will be provided to you (Figure D.2). Just describe sounds in your own words. You will have about 15 seconds to describe a sound, but can take more time if you wish. When you are done and the second playing is complete, click the NEXT SOUND button.


Subject # ________

Sound 1 / 24

Description

Figure D.2. Example of the header on the sheets used in the description test. The rest of sheet was blank. This sheet was used in Test 1.

D.3 Semantic Differential Test

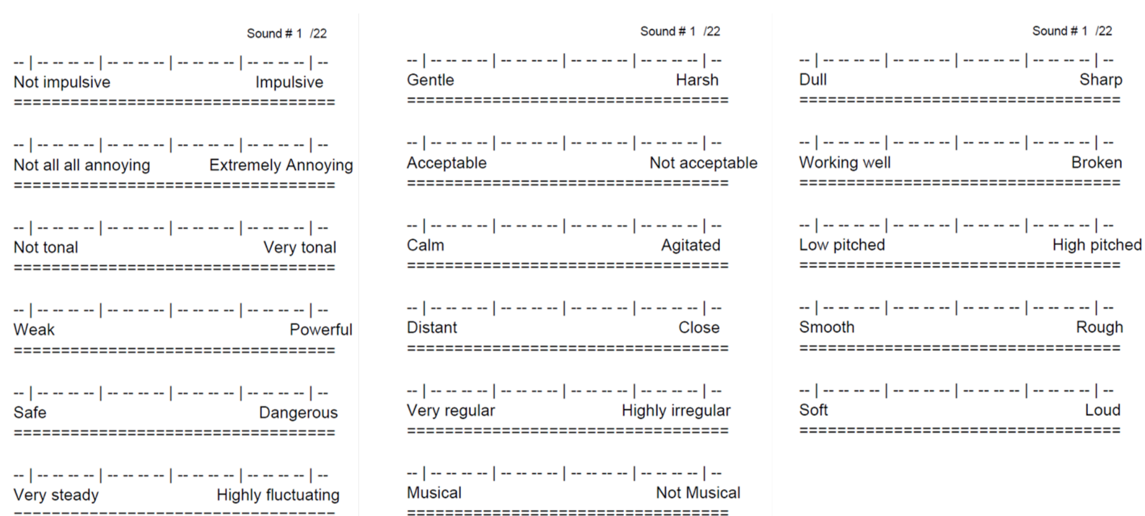
While you are listening, it may be helpful to imagine yourself in your garden, at any time during the day or evening, hearing these sounds continuously. There are no right or wrong responses; we are only interested in your descriptions of the sound. Also, don't worry about being grammatically correct or spelling, and go with your "gut instincts" about the sound.

Before actual test, we will have a practice session with 2 sounds. I will then check to make sure everything is clear and then we will proceed with the test.

You will listen to 22 HVAC&R (Heating, Ventilating, Air-conditioning, and Refrigeration) equipment sound. Each sound will last 10 seconds. A sound will be repeated until you complete an evaluation sheet consisting of 17 scales for the corresponding sound. Three seconds of silence will be included between each repetition. You will have two 3 minutes breaks after the 7th and 14th sounds. After the first time the sound is played you can start evaluating the sound by making a mark on each of the lines. The rating scale looks like the figure below.

While you are listening, it may be helpful to imagine yourself in your garden, at any time during the day or evening, hearing these sounds continuously. There are no right or wrong responses; we are only interested in your opinion of the sound. We recommend you to answer based on your "gut instinct" rather than thinking a lot about the sounds that you hear. We will give you the dictionary definition of word scales to help you understand the meaning of the word scales.

Before starting the actual test, we will play 10 sounds so you can hear the range of sounds. We will then have a practice session with 2 sounds. This will allow you to become familiar with the scoring and will give you the opportunity to ask any questions you may have about the test.


Sound # 1 /22

Not impulsive — Impulsive

Not all all annoying — Extremely Annoying

Not tonal — Very tonal

Weak — Powerful

Safe — Dangerous

Very steady — Highly fluctuating

Sound # 1 /22

Gentle — Harsh

Acceptable — Not acceptable

Calm — Agitated

Distant — Close

Very regular — Highly irregular

Musical — Not Musical

Sound # 1 /22

Dull — Sharp

Working well — Broken

Low pitched — High pitched

Smooth — Rough

Soft — Loud

E. SUBJECTIVE TEST RESULTS

This Appendix contains the results of Test 1 (Chapter 4), Test 2 (Chapter 5) and Test 3 (Chapter 6). In Table E.1 the descriptions of sounds provided by subjects in Test 1 are given. In Table E.2 and E.3 are the BTL values from the series of paired-comparison tests that were a follow-up to the description test. In Table E.4 are the averages of subjects' ratings in Test 2, and in Table E.5 are the corresponding standard errors. In Table E.6 are the end words for scales used in the semantic differential (Test 2). In Table E.7 are the average of subjects' ratings in Test 3, Parts A, B, and C, and the corresponding standard errors. The characteristics of the signals are given in Table C.1 and the corresponding sound quality metrics are given in Table C.2.

Table E.1. Words provided by subjects for each signal in Test 1, the description test. Numbers inside parentheses indicate the number of times the words were used by different subjects. Number of subjects = 42.

Signal Number	Descriptions
3	compressed, loud(3), vacuum(6), spinning,distant, pump, mild(2), bearable, familiar, typical, safe, unworried, medium(3), harsh, blender, vibration(3), constant(2), not heavy, smooth(6) , unpleasant, whirring(2), refrigerator, repetitive, heard, ringing, not loud(2), soft(3), rumble, buzz, tumbling, efficient, quiet, swirling, drone, motor(2), fan(2), tone, noisy, thick, static(2), sucking(2), interference, light, shredder(2), moderate, distractive, dryer(2), even, small drone, air movement(3), annoying, steady, spinning, dish washer, grrrr, hum
8	moving(3), saw, noisy(3), loud(14), vacuum(6), pump, heavy(2), high pitch(4), not bearable, scratching, metallic(2), disturbing(2), sharp, engine(2), above 80dB, high(2), hum(3), whistle(2), constant, strong, artificial, planted, altered, high frequency, ringing, smooth, irritating, itchy, rustle, spinning(2), tumbling, not bad, vibration, echoed, whirling(2), air blowing(10), motor, annoying(3), robust, dryer, even, unpleasant, medium tone, mechanical, washing, upsetting, sucking, shrill,

Table E.1. Continued from previous page.

35	buzzing, loud(27), disturbing(2), conveyor belt, heavy(2), alarming, annoying(5), empty, painful(2), harsh, rotating(2), unpleasant(2), digital, noisy(3), above 90 dB, constant(2), freezer sound, fan(5), uncomfortable(3),air blowing(10), forced, high, smooth, hum, irritating, whiny, high pitched, static(2), prominent, rough, washing, hurt ears, abrasive, close, not soft, temporary, penetrating, repetitive, harsh, vacuum(4), grating, one note, powerful, dryer
76	precessing, vacuum(3), light vehicle move, distant(3), white noise, bearable, light(4), familiar, safe, typical, low noise(2), blender, digital, bees, acceptable, fan(3), medium(3), muffled, dull, hum(3), buzz, quiet, calm(5), cool(2), relaxed(3), home, lightly rough, soft(9), air blowing(4) , itchy, uneven, spinning(3), long, washing(2), sucking, factory(2), systematic, problematic, inefficient, hurrr, powerful, generator, grinding, pleasant, dryer, whirring, distorted, faint, even, loud
77	fan(3), vacuum(12), normal(2), distant(4), rotating(2), not irritating, high speed, white noise, not pleasant, familiar, light, loud, disturbing(2), background(2), fine sound, airy (2), bees + television, not too noisy, buzz, hum(5), medium(2), consistent, refrigerant, cool, artificial, ancient, inefficient, middle high, spacious, soft(4), blowing, loud, sucking(3), whirring(2), calm(3), acceptable, washing(4), dryer, bothersome, smooth(3), hissing, unharmful, hollow, static, nasally, moderate, interference
82	bang, vibration(3), loud(17), engine(4), compressor, heavy(3), wide, bearable, high performance, drone, shredding tree, not harsh, digital, wo∼∼∼, constant, buzz(4), medium, rotating(3), muffled, hum(5), lower background, abrupt(2), noisy, apparent,air blow(8), empty, annoying, itchy, simple, whirly(2), irritating, drilling, static, vacuum(7), mild, one note, close(2), organized, powerful, distracting, fan,

Table E.1. Continued from previous page.

94	vibration(3), high peak, truck (4), low(8), low freq(2), tractor noise at distance,moderate, heavy(3), familiar, soothing(3), engine(2), heavy tone, intense, pulsing, muffled(2), quiet(3), whirring, subdued, old(2), rusty, smooth, small propeller(3), natural, not harsh but not soft(2), distant(3), gravelly, drumming, bumbling, relaxing, reassuring, hum(3), rattling, soft(5), subdued motorcycle, robust, train, faint, properly working, rotating(3), light, mild(2), constant, dangerous, beating
142	grinding(2), truck running(5), helicopter, exhaust fan, heavy(2), loud(6), generator(7), disturbing, scary, vibration(4), pulsating, grunting, motorcycle, medium(3), hum(3), long, irritating, old(3), whirring, prominent, choppy, propeller(2), deep(3), not irritating(2), rumble(2), low frequency(2), full, smooth, drilling(4), thudding, harsh(2), monotonous(2), distorted, moderate, synchoronized, powerful, distinct, cutting tree, rough, beating, annoying(2), repetitive(2), intimidating, boring
143	slow, dull, moderate(3), truck, engine(4), irritating(2), machine(2), heavy(3), grinding(3), tolerable, metallic(3), scary, scratching, rythmic, drilling(4), key tumbling, rotating(2), muffled, clatter, deep, clashing, idling, rumbling, vibration, two sounds, used, repetitive(2), rocky, bumpy(2), rattling(4), grumbling(2), loud(3), whirring, humming, tin can, rocky, mower(3), noisy, distorted, chain involved, powerful(2), snychronized, motor(3), old(2), tractor, rough, close, even, bothersome, cranking

Table E.1. Continued from previous page.

144	motor(3), annoying(4), drilling(8), low cc engine(4), vibrating(5), tolerable, scratching, disturbing, jack hammer, machine gun(2), heavy tone, deep(2), repetitive(2), low(3), muffled, helicopter, choppy(3), weed wacker, summer(2), outdoors, loud(3), high, buzzing, harsh, trembling(2), generator, rattling, propeller(2), thudding, chisel, thumping, fan(2), strong, rumbling, crackling, synchronozed, moderate(2), rotation, cutting tree, rough, low annoying, compresser, Harley Davidson, construction, invasive, grating, drumming, metallic
151	high frequency(2), irritating(2), racing car, helicopter(2), disturbing(2), drilling(6), shaking, heavy, loud(6), unbearable, annoying(4), old, pulsing, cutting(2), old tractor, motorcycle(3), choppy(3), rattle, intense, whirring, weed wacker, summer(3), rotation, bumpy, rough(2), deep, abrupt, consistent, harsh(2), near, propeller(3), grinding(2), chisel, very fast(2), helicopter, crackling, distorted(2), isolated, distinct, broken muffler, wood chipper, electrical saw, vibration, banging, grrr
177	engine(3), high frequency(2),big truck, heavy(6), hard(2), irritating(4), cooling fan(2), shaking, scary, annoying(5), noisy(3), rough(2), deep(9), vibration(13), large fan(2), loud(12),muffled, intense, tractor, hot, old(4), dark, bumpy, grumbling, rumble(2), thunder, rattling, spinning, buzzing, machinery(2), thudding, gravelly, coarse, twirling, moderate, interference, powerful, organized, abrasive, inefficient, whirring, choppy, garbage disposal, close, drill(2), garbling, grating, continuous

Table E.1. Continued from previous page.

182	chirp(2), vibrating(3), noisy(2), big truck(2), hard, metallic(3), drill(2), near(2), annoying(3), motor(3), scratching, unsafe, broken, loud(14), generator(2), construction(4),intense,heavy(3), deep(4), roar, choppy, squeal, grumble, artificial, abrupt, rough, bump, grumpy, rumble, tumbling, low freq.(3), hum, drumming(2), thumping , thudding, harsh(2), squeaky(5), interference, abrasive, grinding(2), rusty, cutting tree, outdoor, jack hammer, large, dangerous, polluting, intrusive
189	irritating, annoying(8), rotating(3), ventilation fan(2), heavy(3), shaking, loud(14), unpleasant(2), old(2), unsafe, drilling, pitchy(2), intense(2), choppy(3), hoarse, rumble(2), vibration(4), deep(2), rusty, propeller(2), rough(2), bumpy, harsh(3), near, gravelling, grating(2), low, worrisome, broken, fast, friction, thudding, uneven, wood chipper(3), powerful(2), twirling, machine gun, loose, close, vigorous, jack hammer, engine, destructive, hurt ears, constant, intrusive, crushing
199	chopper(6), light(6), distant(5), chainsaw, drill, thin(2), scratching(2), safe, medium(5), fine noise(2), deep, sharp, vibration(3), fast(2), rotating, hum(2), high pitch(4), unique, unheard, strange, high, soft(10), smooth, even, buzz(2), whirring(2), rattling, pleasant, calm, quiet(3), thumping, fan(2), subdued, echo, twirling, faint, disorganizing, interference, static(2), flag flapping, mild, small(2), muffled, grrr, airy, growl

Table E.1. Continued from previous page.

200	drone(3), light(4), normal, distant(5), not loud, not irritating(2), machine, low speed, bearable, low(2), safe, muffled, scratching, cover tv with blanket, not noisy, quiet(8), hum(4), chirp, rotating, calm(6), fan, background(2), clean, crisp, smooth(4), soft(12), ringing, vibration(2), reassuring, gentle, whirling(2), nice, subdued, natural, twiwling, faint, static(3), heavy resistence, pleasant, efficient, gentle, mild, staccato, peachful, growl, hissing
212	subtle(2), light(5), AC(2), not irritating, air leakage(3), small(2), mild, bearable, low(4), familiar, safe ,typical, muffled, relaxing, rubbed sound, fan(2), fire alarm cover with blanket, not annoying(2), quiet(12), calm(5), smooth(5), whisper, soft(10), hissing, gentle, itchy, flowing, reassuring, drone, faint(3), far, serene, reflective, complete, rainfall, stream(3), efficient(2), pleasant, hiss, white noise, soothing, help concentrate, air movement(2), distant, steady, blissful, nice, airy
247	loud(8), ear hurt, high frequency, high pressure exhaust, heavy(2), strong(2), irritating(2), unbearable, disturbing(3), scary, harsh(2), engine(3), helicopter(6) , drum(2), dancing in the night bar, pitchy(3), full, deep(6), variation, deep echo background(2), whirring, choppy, artificial, pulsing(3), abrupt, annoying(4), near(2), grainy, gravelly, hum, bumbling, rattling(2), grating(2), drill(2), screeching, high pitch, rumbling(2), rotation, moderate, organized, oscillation(2), rough, vigorous, sharp, uneven, repeat, vibration(3), intrusive, fearful, growl

Table E.1. Continued from previous page.

266	drone, light(2), normal(3), distant(2), not loud, not irritating(2), machine, low speed, quiet, bearable, low, safe, muffled, scratching, cover tv with blanket, not noisy, annoying(4), repetitive(2), loud(5), muffled, rapid(3), rotation(2), bumpy(2), grinding(3), grating, crunching, coarse, choppy(3), moderate(4), thumping, vibration(2), tapping, soft, drill, engine, propeller, staccato, blowing, crackling, mower(2), blender, clanky, broken, not too annoying, distracting, diesel engine, bad, mild
269	high speed, vibration(5), fan(2), loud(25), not pleasant, heavy(4), annoying(7), scratching, disturbing, coarse, tv display show snow, noisy(2), high frequency, muffled, choppy, rumbling(2), abrupt(2), strong, rough, clunky, harsh , grinding(2), bumbling, whirring, tractor, irritating, prominent, drilling(3), blending, ears hurt, garble, uneven, motor(2), construction(3), powerful, stationary, tractor, engine, high pitch, static, headache, intense, distractive, terrible, constant, grating, dangerous, sharp, dark, scary
278	creepy(2), alien(6), pumping water(2), light, telephone(8), whistling, CNC, annoying(2), resonance, metal, high pitch(4), musical(4), medium, muffled, scary, repetitive, instrument, hum(3), deep, smooth(3), hollow(2), ghostly, whirring(3), hover, soft(2), vibration, moderate, futuristic, bubbly, weird(2), ocean, ambiguous, electronic

Table E.1. Continued from previous page.

287	floor polishing, high speed rotation(2), loud(18), irritating(4), high frequency(3), drill(2), high pitch(7), annoying(6), resonance, disturbing, harsh(4), vibration, wo∼∼∼, medium, hum(3), persistant, constant train whistle, aggressive, high(3), rattling, noisy, hard, screech, whirring, obnoxious, sharp(2), hover, spinning, ears hurt(4), yell, powerful, train horn, dryer, random orbital sound, metal grinder(2), noticeable, headache, vacuum(3), foreign
290	gloomy, shaking(2), moderate, dish washer, low frequency(2), drill, irritating(2), muffled(2), high performance, safe, scratching, vibrating cloth, AC compressor(2), heavy tone, motor cycle, choppy(2), intense, hum, dull, rapid, rough(2), helicopter(6), deep(2), pulsating, uneven, rattle, ticking, whooshing, buzz, vibration(3), soft, small gun, airplane, moderate, thin blade, distant(3), wierd, horror, mild, not to distracting, flappy, rumbling, small, suppressed,
293	high frequency(2), headache, noisy(2), high speed, loud(14), rotating, compressor, annoying(8), high pitch(7), unbearable, disturbing, high performance, scratching(3), harsh(4), metal cut/saw(6), unpleasant, irritating(3), buzzing, intense, painful, grinding(2), bad, frightening, high(2), abrupt(2), rumbling, disruptive, grating(2), sawing, shredding, bumbling, metal blade spinning(2), jack hammer, irritating, shrill, friction, whining, cutting(2), jet engine, hurts ears(4), uneven, robust, choppy, generator, old, sharp(2), coarse, uncomfortable, growl

Table E.2. Bradley-Terry-Luce (BTL) values in Test 1, the paired comparison test quieter region. Number of subjects = 33.

Test Number	Signal	Modification	BTL
Set 1	Refrigerated Truck (Vector 1350) Diesel Engine 2250 rpm	More Tonal 2	0.71
		More Tonal	0.88
		Original	0.00
		Less Tonal 1	-0.51
		Less Tonal 2	-0.57
Set 2		More Rough 3	2.69
		More Rough 2	2.63
		More Rough 1	1.21
		Original	0.00
		Less Rough 1	0.18
Set 3	Residential Fan 1100 rpm Compressor 3600 rpm	More Tonal 2	1.53
		More Tonal 1	0.30
		Original	0.00
		Less Tonal 1	0.66
		Less Tonal 2	0.13
Set 4		More Rough 4	2.88
		More Rough 3	2.41
		More Rough 2	1.54
		More Rough 1	0.43
		Original	0.00

Table E.3. Bradley-Terry-Luce (BTL) values in Test 1, the paired comparison test louder region. Number of subjects = 33.

Test Number	Signal	Modification	BTL
Set 5	Refrigerated Truck (Vector 1950) Diesel Engine 1850 rpm	More Tonal 2	2.07
		More Tonal 1	1.24
		Original	0.00
		Less Tonal 1	0.04
		Less Tonal 2	0.64
Set 6		More Rough 3	0.99
		More Rough 2	0.73
		More Rough 1	0.30
		Original	0.00
		Less Rough 1	-0.68
Set 7	Residential Fan 452 rpm Compressor 1800 rpm	More Tonal 2	1.83
		More Tonal 1	0.66
		Original	0.00
		Less Tonal 1	-0.22
		Less Tonal 2	0.05
Set 8		More Rough 3	-0.03
		More Rough 2	0.01
		More Rough 1	-0.30
		Original	0.00
		Less Rough 1	-0.26

Table E.4. Average of subjects' ratings in Test 2, the semantic differential test. Number of subjects = 39.

Signal Number	Scale 1	Scale 4	Scale 5	Scale 6	Scale 8	Scale 11	Scale 13	Scale 15	Scale 16	Scale 17	Scale 10	Scale 2	Scale 9	Scale 12	Scale 14	Scale 7	Scale 3
12	2.26	3.07	3.99	0.02	2.26	1.60	-1.27	-0.59	-2.23	3.78	2.14	1.63	1.65	3.41	-1.45	2.04	3.10
17	3.76	4.73	5.74	2.78	4.26	1.68	0.88	-0.02	0.58	4.65	3.63	3.05	3.56	4.19	0.18	4.26	4.71
22	5.96	3.83	6.10	4.02	5.82	2.14	-3.25	0.97	-2.45	6.86	5.11	4.99	4.49	6.15	1.51	5.64	6.23
30	6.33	1.03	2.49	5.67	5.54	-0.72	0.49	1.50	0.35	6.63	5.08	5.89	4.48	6.80	1.52	5.15	5.97
34	3.07	-0.12	-0.15	1.67	3.70	-0.93	-2.95	-2.03	-2.74	5.30	2.48	4.32	1.72	3.95	-2.23	1.75	3.11
60	-4.63	-0.49	-1.27	-4.32	-4.63	0.39	-5.97	-4.29	-5.03	1.29	-4.59	-4.63	-4.99	-3.28	-4.66	-4.77	-4.15
74	-3.87	1.28	0.79	-3.82	-2.33	1.08	-2.95	-4.22	-4.03	2.26	-3.28	-2.67	-2.55	-2.12	-2.83	-2.44	-1.88
78	-3.73	0.66	0.01	-1.77	-2.86	1.63	-3.79	-3.65	-3.89	1.79	-2.63	-3.13	-2.50	-3.46	-3.46	-2.32	-1.42
83	-2.06	-4.09	-3.98	1.29	-1.25	-2.89	-3.77	-1.27	-3.77	4.73	-2.44	-0.79	-3.14	-0.96	-3.74	-3.32	-2.08
90	-5.28	-4.46	-4.83	-0.79	-2.89	-2.31	-4.60	-2.64	-3.83	3.15	-3.99	-3.76	-4.64	-5.02	-4.55	-4.48	-3.98
97	-1.09	-2.05	-2.02	2.22	-0.42	-0.70	-3.24	0.50	-3.03	4.05	-0.36	-0.94	-2.45	-1.18	-2.95	-1.18	0.14
142	0.98	-3.87	-3.53	3.22	0.14	-2.18	-2.88	0.59	-3.13	4.77	0.02	1.01	-1.38	2.07	-2.34	-0.42	-0.33
170	4.17	-1.77	-1.46	3.75	3.07	-1.21	0.26	1.89	-1.32	4.97	3.08	4.15	2.43	4.65	-0.01	2.70	3.08
171	2.88	-2.01	-0.69	3.91	3.21	-0.12	-0.69	0.47	-1.15	4.71	2.16	3.73	0.81	4.54	0.22	1.54	3.27
182	4.80	-1.11	-1.46	4.98	3.21	-0.53	-0.92	2.59	-0.39	4.90	3.54	4.60	2.60	5.35	0.77	2.69	3.21
196	5.76	-1.79	-1.04	6.04	5.14	-0.55	-1.19	2.89	-0.96	6.30	3.98	5.72	2.34	6.47	0.13	3.99	4.99
205	5.78	-1.99	-1.39	5.46	3.31	-0.73	-0.91	1.00	-2.13	6.34	3.80	5.55	3.28	5.71	0.38	2.43	3.13

Table E.4 continued from previous page

235	1.22	-1.27	-1.57	2.27	1.16	-1.18	-3.04	-1.64	-3.19	5.04	1.08	2.21	-0.59	2.08	-1.91	-0.12	0.84
243	2.29	-2.51	-2.15	4.22	2.24	-2.05	-0.21	0.70	-0.95	5.16	2.48	1.61	0.77	3.01	-0.79	1.34	1.75
283	0.10	-2.25	-3.13	2.10	0.89	-0.14	0.31	1.77	0.72	2.71	1.37	1.20	-0.40	2.58	-0.37	0.63	0.79
287	5.09	4.54	5.21	1.01	3.30	3.08	-3.52	-1.15	-3.83	6.15	3.99	4.05	3.23	6.08	-0.07	4.52	5.59
293	7.62	4.25	6.06	5.86	6.81	1.18	3.18	2.49	1.74	6.59	7.07	6.74	6.91	7.57	2.82	6.79	7.48

Table E.5. Standard deviation of the estimated mean of subjects' ratings in Test 2, the semantic differential test. Number of subjects = 39.

Signal Number	Scale 1	Scale 4	Scale 5	Scale 6	Scale 8	Scale 11	Scale 13	Scale 15	Scale 16	Scale 17	Scale 10	Scale 2	Scale 9	Scale 12	Scale 14	Scale 7	Scale 3
12	0.50	0.53	0.56	0.68	0.49	0.71	0.63	0.68	0.71	0.77	0.66	0.54	0.67	0.50	0.70	0.66	0.57
17	0.41	0.55	0.45	0.70	0.49	0.86	0.75	0.72	0.70	0.66	0.60	0.47	0.53	0.43	0.68	0.54	0.52
22	0.41	0.72	0.32	0.61	0.39	0.81	0.61	0.73	0.65	0.54	0.59	0.50	0.56	0.34	0.86	0.57	0.44
30	0.54	0.68	0.64	0.60	0.55	0.85	0.77	0.76	0.78	0.57	0.73	0.46	0.71	0.38	0.82	0.61	0.52
34	0.44	0.56	0.62	0.75	0.60	0.70	0.67	0.67	0.64	0.57	0.57	0.45	0.59	0.51	0.58	0.61	0.52
60	0.61	0.76	0.69	0.68	0.68	0.80	0.38	0.64	0.60	0.66	0.56	0.57	0.66	0.57	0.62	0.55	0.66
74	0.47	0.79	0.68	0.54	0.69	0.72	0.69	0.54	0.61	0.77	0.78	0.50	0.76	0.58	0.59	0.68	0.74
78	0.52	0.71	0.78	0.66	0.77	0.73	0.60	0.71	0.60	0.78	0.64	0.57	0.76	0.53	0.65	0.70	0.69
83	0.46	0.42	0.53	0.68	0.70	0.67	0.48	0.67	0.46	0.59	0.52	0.56	0.66	0.55	0.51	0.53	0.57
90	0.40	0.52	0.47	0.81	0.72	0.68	0.50	0.80	0.48	0.77	0.58	0.60	0.68	0.40	0.51	0.58	0.62
97	0.55	0.52	0.64	0.72	0.61	0.65	0.54	0.80	0.57	0.58	0.63	0.57	0.56	0.53	0.60	0.59	0.54
142	0.60	0.46	0.49	0.60	0.67	0.68	0.64	0.80	0.61	0.63	0.68	0.59	0.73	0.51	0.70	0.67	0.62
170	0.51	0.70	0.72	0.64	0.68	0.72	0.70	0.69	0.65	0.67	0.65	0.46	0.70	0.53	0.76	0.66	0.54
171	0.63	0.59	0.72	0.66	0.58	0.69	0.69	0.74	0.66	0.63	0.66	0.54	0.68	0.48	0.72	0.71	0.61
182	0.40	0.74	0.73	0.49	0.54	0.80	0.75	0.73	0.75	0.65	0.58	0.53	0.70	0.41	0.80	0.64	0.52
196	0.44	0.76	0.80	0.42	0.52	0.87	0.74	0.72	0.76	0.50	0.68	0.45	0.73	0.30	0.76	0.57	0.46

Table E.5 continued from previous page

205	0.33	0.72	0.76	0.48	0.48	0.82	0.77	0.70	0.68	0.52	0.64	0.37	0.50	0.33	0.65	0.62	0.55
235	0.64	0.57	0.58	0.65	0.65	0.72	0.63	0.71	0.73	0.53	0.65	0.47	0.63	0.65	0.74	0.71	0.65
243	0.55	0.63	0.77	0.49	0.62	0.72	0.76	0.77	0.73	0.47	0.64	0.54	0.67	0.55	0.76	0.72	0.71
283	0.64	0.60	0.57	0.65	0.69	0.76	0.85	0.66	0.87	0.81	0.70	0.57	0.71	0.54	0.68	0.58	0.58
287	0.65	0.66	0.61	0.81	0.66	0.83	0.71	0.82	0.68	0.47	0.75	0.59	0.79	0.49	0.76	0.70	0.56
293	0.32	0.60	0.57	0.50	0.35	0.95	0.72	0.72	0.72	0.69	0.42	0.46	0.45	0.34	0.70	0.64	0.42

Table E.6. Test 2 scale numbers and corresponding names used in Table E.4 and E.5.

End Scale Number	Scale Name
Scale 1	Soft – Loud
Scale 2	Weak – Powerful
Scale 3	Not at all Annoying – Extremely Annoying
Scale 4	Low Pitched – High Pitched
Scale 5	Dull – Sharp
Scale 6	Smooth – Rough
Scale 7	Acceptable – Not Acceptable
Scale 8	Gentle – Harsh
Scale 9	Safe – Dangerous
Scale 10	Calm – Agitate
Scale 11	Not Tonal – Very Tonal
Scale 12	Distant – Close
Scale 13	Very Steady – Highly Irregular
Scale 14	Working Well – Broken
Scale 15	Not Impulsive – Very Impulsive
Scale 16	Very Regular – Highly Irregular
Scale 17	Musical – Not Musical

Table E.7. Average and standard deviation of the estimated mean of responses in Test 3. Signals are described in Table C.1.

Part A			Part B			Part C		
Signal Number	Average	Standard Error	Signal Number	Average	Standard Error	Signal Number	Average	Standard Error
207	4.10	0.16	88	5.27	0.18	100	6.24	0.14
283	5.62	0.15	91	5.63	0.16	105	5.80	0.13
294	6.33	0.14	93	5.71	0.16	114	5.19	0.15
11	6.86	0.15	241	6.47	0.15	121	5.33	0.13
12	6.63	0.14	266	5.37	0.16	83	4.73	0.16
13	7.25	0.15	282	5.59	0.19	84	3.50	0.13
19	7.19	0.12	285	7.09	0.15	87	3.55	0.13
3	5.17	0.18	11	6.71	0.16	89	4.70	0.18
33	5.64	0.16	12	6.22	0.19	204	5.89	0.13
39	5.41	0.18	19	6.79	0.18	242	6.53	0.14
5	6.87	0.14	8	6.26	0.18	243	6.00	0.16
7	4.66	0.18	141	5.12	0.17	264	6.88	0.13
8	6.47	0.14	142	4.99	0.18	1	5.91	0.15
209	5.23	0.17	143	5.14	0.16	11	6.59	0.16
257	4.02	0.16	144	3.53	0.16	12	6.25	0.18

277	6.42	0.14	146	5.11	0.17	15	5.66	0.17
279	5.97	0.17	147	5.56	0.19	17	6.35	0.16
288	4.35	0.19	149	5.54	0.16	19	6.81	0.15
143	5.99	0.16	151	5.58	0.15	21	5.25	0.20
144	4.17	0.19	152	5.46	0.18	22	7.40	0.12
148	5.98	0.17	154	5.38	0.15	25	6.69	0.16
151	6.41	0.13	155	5.53	0.17	32	7.65	0.13
152	6.17	0.16	156	6.03	0.16	35	7.37	0.13
155	6.20	0.14	157	6.01	0.15	40	5.92	0.14
156	6.51	0.16	162	5.93	0.16	8	6.34	0.14
162	6.52	0.16	164	5.95	0.15	143	5.71	0.12
166	6.53	0.15	170	6.13	0.14	144	3.66	0.15
225	3.51	0.15	171	5.47	0.15	145	5.69	0.16
226	3.22	0.13	173	6.30	0.15	151	6.04	0.13
230	4.89	0.16	175	6.35	0.16	152	6.12	0.16
231	3.66	0.15	181	7.00	0.14	155	6.16	0.15
232	3.46	0.14	186	6.78	0.16	156	6.50	0.14
233	3.95	0.16	187	7.05	0.14	158	6.17	0.15
235	5.85	0.16	191	7.41	0.12	162	6.10	0.17
301	6.78	0.17	195	7.18	0.13	172	6.14	0.16

306	4.14	0.20	196	7.24	0.15	188	6.87	0.14
307	3.98	0.23	197	7.21	0.15	198	7.27	0.15
75	3.85	0.16	235	4.99	0.16	217	2.41	0.12
76	3.96	0.15	82	6.10	0.15	234	3.61	0.14
77	4.14	0.15	318	5.38	0.18	235	5.43	0.15
78	3.87	0.14	319	4.81	0.17	60	3.03	0.13
79	4.02	0.13	321	5.15	0.19	66	3.03	0.13
81	4.51	0.14	322	5.24	0.17	323	6.61	0.12
320	6.16	0.15	325	7.37	0.14	42	5.18	0.16
42	5.39	0.15	42	4.49	0.17	311	5.95	0.17
312	7.10	0.14	312	6.80	0.17	312	6.65	0.14
316	4.08	0.13	330	6.25	0.16	314	7.34	0.13
317	5.01	0.16	351	6.50	0.17	351	6.40	0.15
351	6.70	0.14	352	7.28	0.14	342	3.53	0.19
340	5.94	0.14	343	5.41	0.14	339	7.49	0.13

F. SAMPLE CHECKLIST

This Appendix contains a sample of the checklists used in the various tests.

APPENDIX F. SAMPLE CHECKLIST

Research Project: **Sound Quality of HVAC Equipment**

Maximum number of subjects approved: ___200___

Current subject total: _________
This subject's number: ________

Before Subject Arrival:

- ☐ Wash hands; make sure you have gloves available
- ☐ Turn on computer
- ☐ Turn on amplifier
- ☐ Check connections (to sound card, amplifier, patch panel, earphones and eye tracker)
- ☐ Check amplifier settings
    - Speaker outputs on
    - Left knob turned to ~ 6
    - Mode switch set to "Stereo"
    - Ground lift switch set to "Lift"
- ☐ Enter sound booth and close doors carefully
- ☐ Turn on audiometer
- ☐ Calibrate sound level meter
    - Meter, microphone, and calibrator stored in large black case in sound quality booth
    - Attach microphone to meter and turn meter on
    - Place calibrator over microphone and turn it on
    - Using stylus (stored in slot on side of meter), tap the menu button in the bottom left corner of the meter screen, and select 'Calibration' from the pop-up menu
    - Tap the 'Start calibration' button
    - When the dialog box pops up, check the numbers and tap 'Accept calibration' if the numbers look good (generally deviates by 0.03 dB or lower)
    - When the screen displays the 'Exit calibration' button, tap it
    - Remove and stow calibrator
- ☐ Assemble B&K coupler and place on sound level meter microphone
    - Parts stored in black plastic box on desk in sound booth
- ☐ Open program 'SubjTest'
    - Be sure to maximize window
- ☐ Play 'Signal_warm' several times to warm up the system
    - Signal located in 'Test Sound' directory
- ☐ Play 94 dB calibration tone through LEFT (blue tube) earphone and coupler to calibrate playback. Adjust Lynx levels as necessary.
    - ***Important note:*** *the meter, coupler, and earphone assembly is very sensitive to vibration or movement. When recording sounds, hold the meter in your hands and keep as still as possible.*

      Expected: __94 ± 0.3__ Actual: ________

- ☐ Play 94 dB calibration tone through RIGHT (red tube) earphone and coupler to calibrate playback. Adjust Lynx levels as necessary.

Initials _________
Date and Time _________

APPENDIX F. SAMPLE CHECKLIST

Expected: __94 ± 0.3__ Actual: ________

- ☐ Close Lynx window if open
- ☐ The instrumentation settings for this tone are the same as the settings for all the signals to be played in the test.
- ☐ Tick here if yes. If no, what are the new amplifier settings for playing the test signals? And confirm that they have been set.
  New Settings: ________________________Are Set? _____________________________
- ☐ Play 5 check signals through LEFT (blue tube) earphone and coupler. (These signals located in subdirectory “Signal check” under “Test Sound”.)
  - Press the Play/Pause button above the screen to begin recording, and to stop. Recorded levels will appear on the screen.
  - Clear recordings by pressing the button to the left of Play/Pause.
  - Record each signal, clearing the display after each time. (This will help to avoid rounding errors/corrupted data.)
  - Record fast average A-weighted levels below. Circle all that are acceptable:

EF3:	Expected __72.5 ± 0.4dBA__		Observed ________
________	________	________	________
EF77:	Expected __69.9 ± 0.4dBA__		Observed ________
________	________	________	________
EF189:	Expected __80.7 ± 0.4dBA__		Observed ________
________	________	________	________
EF199:	Expected __61.2 ± 0.4dBA__		Observed ________
________	________	________	________
EF287:	Expected __85.1 ± 0.4dBA__		Observed ________
________	________	________	________

- ☐ Play 5 check signals through RIGHT (red tube) earphone and coupler. (Same signals/location as in previous step.) Record fast average A-weighted levels below:

EF3:	Expected __72.2 ± 0.4dBA__		Observed ________
________	________	________	________
EF77:	Expected __69.2 ± 0.4dBA__		Observed ________
________	________	________	________
EF189:	Expected __80.3 ± 0.4dBA__		Observed ________
________	________	________	________
EF199:	Expected __61.2 ± 0.4dBA__		Observed ________
________	________	________	________
EF287:	Expected __85.0 ± 0.4dBA__		Observed ________

Checklist, page 2/5 Initials _________

Date and Time _________

APPENDIX F. SAMPLE CHECKLIST

________ ________ ________ ________

- Close program 'SubjTest'
- Open program'Test1_A'
- Turn off sound level meter; return meter, coupler, and microphone to their cases
- Wash hands
- Clean and wipe down testing area
  - Audiometer headphones
  - Mouse and Keyboard
  - Desk
- Hang 'Be Quiet—Human Subject Testing In Progress' signs on entrances to Acoustics wing
- Get subject packet (consent form, questionnaire) and fill in the cover sheet, which includes a check on the number of subjects tested so far under this IRB test protocol (#1507016324). Test stops here, if this subject would exceed the approved number for testing.

When subject arrives:

- Greet subject and give a brief test overview (outline major points of test procedure)
- Obtain informed consent (Appendix B)
  - Make sure participants sign on the last page of consent form
  - Make sure researcher sign on the last page of consent form after get participants' signature (initials on pages)
- Have subject fill out the questionnaire (Appendix C)
- Put on gloves
- Test subject hearing – If HL > 20 dB, provide info on Audiology clinic (Test stops here)
  - Explain how the test works and what the subject should expect
  - Subject should be facing away from the audiometer so they can't see you working the machine
  - Make sure turn amplitude down to zero before change the frequency
  - Cards containing Audiology clinic info are located in drawer on top of audiometer

During test:

- Wearing gloves, attach ear-tips in front of subject and give them ear-tips and show them how to insert the tips properly
  - Make sure hand ear-tips to subject before show them how to insert the tips
- Give subject test instructions (Appendix F) and sheet of sound descriptions
- Explain how test will work: typically e.g., sequences of sound familiarization; some practice; and taking the test, followed by writing down comments. Answer any questions
  - Mention that they can have more control by clicking and dragging the slide
- Remove sheet of sound descriptions from the subject.
- Familiarization (5 sounds)
- Practice (2 sounds)
- Check that subjects are comfortable, answer any questions
- Exit the sound booth and shut the doors carefully

Initials _________
Date and Time _________

APPENDIX F. SAMPLE CHECKLIST

- Give subject the signal that they may begin the test
  - Watch subject to make sure they do not displace their earphones
- Test1, Part A (verbal descriptions of the sounds)
- Re-enter the sound booth; give subject the option of a 15-minute break
- Close program 'Test1_A'
- Open program'SubjTest'
- Load test file
- Give subject test instructions (Appendix F)
- Practice (3 sounds)
- Check that subjects are comfortable, answer any questions
- Exit the sound booth and shut the doors carefully
- Test1, Part B (Rating the sounds)

After Test:

- Get general comments from the subject
- Retest hearing – If subject shows sign of threshold shift, provide info on Audiology clinic
- Escort subject to Donna in HLAB
- Pay subject
- Ask if subject would like a copy of the whole consent form
- Escort subject to the door
- Save test (click "Save")
- Check data (subject's responses) under "View Report"
  - Save report file under name "SubjectResponses##"
  - Close program
  - Transfer report file to directory "Results"
- Calibrate sound level meter.
  - See "Before Subject Arrival" for procedure
- Play calibration signals through LEFT (Blue tube) headphones and coupler to calibrate playback.

  Expected: __94 ± 0.3__ Actual: ________

- Play calibration signals through RIGHT (Red tube) headphones and coupler to calibrate playback.

  Expected: __94 ± 0.3__ Actual: ________
  - Test data accepted
  - Test data rejected

If expecting another subject immediately afterwards: (write yes/no here: ________)

- Take a fresh copy of this checklist
- Look at the calibration results above and copy them into the "Before Subject Arrival" section of the new checklist
- Gather previous subject's papers ***plus this checklist*** into complete packet; set aside in cabinet at testing station
- Continue with the new checklist from calibration on

 Initials _________

Date and Time _________

APPENDIX F. SAMPLE CHECKLIST

If not expecting another subject immediately afterwards: (check here if so ☐)

- ☐ Power down sound level meter
    - ☐ Return sound level meter, coupler, and microphone to their cases
- ☐ Power down audiometer
- ☐ Power down computer
- ☐ Power down amplifier
- ☐ Wipe down testing area again
    - ☐ Audiometer headphones
    - ☐ Mouse and keyboard
    - ☐ Desk
- ☐ Take down ‘Be Quiet—Human Subject Testing In Progress’ signs
- ☐ Deposit packet in Prof. Davies’ office

Initials _________
Date and Time _________