12-2014

Determining Thresholds of Annoyance to Tones in Noise

Jennifer Marie Francis
University of Nebraska-Lincoln, jfrancis@unomaha.edu

DETERMINING THRESHOLDS OF ANNOYANCE TO TONES IN NOISE

by

Jennifer Marie Francis

A THESIS

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of Master of Science

Major: Architectural Engineering

Under the Supervision of Professor Lily M. Wang

Lincoln, Nebraska

December, 2014

DETERMINING THRESHOLDS OF ANNOYANCE TO TONES IN NOISE

Jennifer Marie Francis, M.S.

University of Nebraska, 2014

Advisor: Lily M. Wang

Building services equipment often produce noise with prominent tones that can lead to complaints from occupants in the built environment. Previous studies have investigated human perception to tones in noise but it is still unclear at what threshold of prominence these tones lead to human annoyance. The goal of this research is to apply two different methods towards defining thresholds of annoyance to two tonal frequencies: 125 Hz and 500 Hz. In Method I – Direct Assessment with Task, subjects are exposed to 10 minutes of broadband noise with a tonal frequency set at a certain level of prominence while completing a task. They are subsequently asked to fill out a subjective questionnaire after exposure to each noise condition. Five prominence levels of each of the two tonal frequencies are tested above two different background noise levels (40 dBA and 55 dBA) for a total of 20 test trials. In Method II – Magnitude Adjustment, subjects are exposed to each of the two tonal frequencies set at a certain level above each of the two background noise levels and are then asked to adjust the level of the tone, up or down, until it becomes just annoying. The same two tonal frequencies (125 Hz and 500 Hz) and two background noise levels (40 dBA and 55 dBA) that were used in Method I are also used in Method II with one repetition for a total of eight trials.

The potential thresholds of annoyance that were found for both methods were right around thresholds of prominence (as defined by Prominence Ratio in ANSI S1.13-2005). Subjects rated annoyance higher for the louder background noise condition at same prominence levels for both of the tonal frequencies. Results as well as strength and weaknesses of both methods are compared.

Acknowledgements

This project would not have been possible without the help and support from numerous people. First, I would like to thank the Institute of Noise Control Engineering for their 2012 Undergraduate Research Award. I must also thank Paul Sim for his help with the creation of the JavaScript program that was used in Method II of this study. Thanks to my undergrad assistants, Adam Steinbach for his help with the design of this research and Kristin Hanna for her help with running test sessions and collecting data. Thank you to Ellen Peng for sharing your knowledge of acoustics and statistics with me and for always being willing to help me even though you were busy with your own research! A *very* big thank you to Joonhee Lee for all the hours you spent measuring, calibrating and calculating tonal metrics for all those test signals. A big thank you also goes out to my fellow (and former) acoustics students at the University of Nebraska, especially those who participated in this research. I really could not have done this without all of you!!

Thank you to my family and friends for their unwavering support and encouragement during my time in graduate school. A special thank you goes out to Brad Epstein; I could not have done this without your love, encouragement and support. I love you!

Thank you to my advisory committee, Dr. Walt Jesteadt and Dr. Josephine Lau, for their help and guidance through this process. Last but certainly not least, thank you to Dr. Lily Wang for being my advisor, teacher, roommate, therapist, nurse, and mentor over these past few years. I feel lucky to have had the opportunity to learn from you. I appreciate you and everything you've done for me more than you will ever know.

Table of Contents

List of Tables

List of Figures

Chapter 4

Chapter 1: Introduction

1.1 Introduction to Work

Modern mechanical systems in buildings often produce noise that contains prominent tones due to rotating parts such as fans, motors, etc. Additionally, the push towards more energy-efficient heating, ventilation and air conditioning (HVAC) equipment often results in the production of louder and more prominent tones. This noise often leads to discomfort and complaints from those in the built environment and surrounding communities. Current noise criteria guidelines for designing building mechanical systems do not apply well if the noise contains prominent tones. Sufficient data are not available at this time to provide further guidance on acceptable levels of tones in noise.

Much research has been done throughout the years to investigate the effects of different types of tones in noise on human perception and performance as there are many sources (aircraft, industrial machinery, HVAC equipment, computers and other office equipment) that produce this type of spectra. In the 1950's research focused on defining acceptable noise criteria for office buildings. Leo Beranek conducted one such study in 1957 which led to the development of the indoor noise criteria known as Noise Criterion (NC) Curves; these NC curves are still utilized today. With the development of the jet engine in the 1960's, much research was done relating human annoyance to perceptible tones in aircraft noise (Little 1961, Kryter and Pearsons 1965, Little and Mabry 1969). Results from these studies led to the development of a tonal noise correction factor to be added to the metric that had been used to quantify human perception to aircraft noise, Perceived Noise Level (PNL). In the 1980's, Rhona Hellman conducted a number of studies on human perception to noise containing prominent tones and found that tones in noise as well as the frequency of the tone do impact ratings of annoyance, loudness and noisiness (Hellman 1982, 1984, 1985).

Many studies have been done involving human perception and performance while being exposed to ventilation-like spectra. But in most of these studies, only six or less signals were compared which is not a large enough sample to be able to specify thresholds of annoyance to noise with tones. Results from these studies indicate that the presence of tones can impact perception and performance (Landström et al. 1991, 1993, 1994; Holmberg et al. 1993; Ryherd and Wang 2010), but none of these provided guidelines for what the threshold of annoyance for tones in noise should be across a wide range of frequencies.

A number of methods for quantifying the prominence of tones in noise have been developed. Annex A in ANSI S1.13-2005 describes the metric used in this study, Prominence Ratio (PR). PR is the ratio of power contained in the critical band centered on the tone to the average power of the two adjacent critical bands, above and below the critical band containing the tone. Hellweg and Nobile (2002) conducted a round robin test to compare PR to another tonal metric that is also defined in ANSI S1.13-2005, Tone-to-Noise Ratio. Results from this test led to the 2005 revision of prominence for frequencies less than 1000 Hz. Hellweg and Nobile's study was limited in that they extrapolated their findings based on testing only two tonal frequencies (250 Hz and 1000 Hz). Their study was aimed at finding thresholds of prominence as their subjective questionnaires focused on the subjects' perception of 'prominence' rather than annoyance.

The goal of this research is to determine thresholds of annoyance to tonal noise using two different methods. In Method I, subjects perform a task while being exposed to different signals. Subjects are then asked to complete a subjective questionnaire based on the task they complete and the noise they hear. In Method II, subjects are exposed to a signal and are asked to adjust that signal until it is just annoying. Threshold of annoyance results from both methods are compared and contrasted.

1.2 Outline of Thesis

Chapter 2 discusses previous research pertinent to this study and explains how the previous work led to the development of this study. Chapter 3 presents the methodologies used including creation of test sessions and signals as well as the statistical analyses that were used on the data gathered from Method I. Chapter 4 presents and discusses the results from both methods utilized in this study. Chapter 5 summarizes that results and suggests ideas for future work.

Chapter 2: Literature Review

This chapter discusses previous research pertinent to this thesis. Previous research is separated into subsections involving: (1) human perception and annoyance, (2) tonal noise, and (3) tonal metrics. The application of these previous studies to the current one will also be discussed.

2.1. Human Perception and Annoyance

Over the years, many researchers have investigated the effects of noise on human perception and performance. It is important to have an understanding of the effects different noises have on humans as this enables the development of proper metrics and standards to help reduce annoyance. Starting in the 1950's, much research focused on defining acceptable conditions commonly found in office buildings. In 1957, Beranek conducted a study on office noise; 300 employees from four different types of companies filled out questionnaires relating to the noise in their environment. The resulting ratings from these questionnaires were then plotted as a function of each of three sound level metrics: the overall Sound Pressure Level (SPL), the speech interference level (SIL), and the loudness level (LL). Beranek found that SPL did not correlate as well with subjective ratings as SIL and LL did. The correlation of the subjective ratings was better with LL than with SIL. This study also showed that neither metric alone was sufficient to characterize the noise though; both were needed. As a result of this research, Beranek developed an indoor noise criterion known as Noise Criterion (NC) Curves. NC curves are still utilized today to help measure and classify mechanical system noise in rooms.

A number of studies similar to the one conducted by Beranek were carried out in order to determine acceptable noise conditions in office buildings (Keighly 1966, 1970, Hay and Kemp 1972, Blazier 1981, Beranek 1989, Blazier 1997). These studies led to the development of new indoor noise criteria including Balanced Noise Criteria (NCB), Room Criteria (RC) and Room Criteria Mark II (RC-Mark II). These noise criteria are described in Chapter 48 of the ASHRAE Handbook – HVAC Applications (2011). Since the development of these indoor noise criteria, researchers have conducted many studies to try to determine which criterion correlates best with subjective perception. Tang and colleagues conducted a number of such studies where occupants in built offices as well as residential apartments were surveyed about the noise in their environment (Tang et al 1996, Tang 1997, Tang and C.T. Wong 1998, Tang and M.Y. Wong 2004). These subjective responses were then correlated to physical measurements that were taken in the spaces. Ayr and colleagues also conducted occupant surveys in office buildings and then correlated the responses to physical measurements and noise indices (Ayr et al 2001, Ayr et al 2003). Both Tang and colleagues and Ayr and colleagues found that the A-weighted equivalent sound pressure level (L_{Aeq}) correlated best with subjective perceptions of annoyance and loudness.

Most research is conducted with the goal of trying to predict annoyance in order to prevent it, but in some cases being able to predict annoyance is necessary when using sounds to purposefully generate annoyance. This was the case in 1981 in Waco, Texas when law enforcement used Tibetan chants and heavy metal during the siege of the Branch Davidian compound (Frank et al 2007). If sound is going to be used to create annoyance for the purpose of dispersing crowds, preventing individuals from entering restricted areas, or evicting individuals from buildings, it is important to understand what sounds are most annoying and what effects these sounds will produce in the targeted individuals. Frank and colleagues designed a study to obtain preliminary information concerning the annoyance of and tolerance to 10 different sounds that were presented at three different SPL's (80, 90, 100 dB). Subjects were exposed to each sound and then rated their level of annoyance caused by the sound as well as estimating how long they would be able to tolerate listening to the sound while standing, threading a needle, or doing a crossword puzzle. Results showed that SPL was a dominant factor in perceived annoyance; the higher the SPL the greater the annoyance and the less time the sound was able to be tolerated.

As found by Persson and colleagues (1985, 1988), sound level is not the only quality of noise that affects annoyance; spectral content also makes a difference. Persson and Bjorkman (1988) exposed subjects to four broadband noises centered at different frequencies (80, 250, 500, and 1,000 Hz) for 30 minutes each while studying their own textbooks; after exposure to each signal, subjects evaluated their degree of annoyance using a 10-point scale ranging from "not annoying" to "very annoying." Persson and Bjorkman found that subjects rated the noise centered around 80 Hz as more annoying than the other frequencies when the level was held constant. These results indicate that noise containing more energy in the lower frequencies is more annoying than noise with higher frequency content. Results from a subsequent study that was done by Persson Waye and Rylander (2001) also suggest that annoyance was more closely related to low frequency content than to overall sound pressure level.

2.2 Tonal Noise

Tones in noise, or more specifically human annoyance to tones in noise, have generated much interest over the years as there are many sources that generate this type of spectra. Modern mechanical systems for heating, ventilation, and air conditioning, aircraft, industrial machinery, computers and other office equipment can all produce noise with perceptible tonal components due to rotating parts such as motors, fans, propellers, etc. This tonal noise often leads to discomfort and complaints by building occupants as well as those in the surrounding community.

2.2.1 Early Aircraft Work

In the late 1950's, Boeing introduced a new aircraft with jet engines as opposed to the propeller-driven aircrafts that had been utilized throughout the decade (Beranek 2007). These new aircrafts were deemed noisier than their propeller-driven counterparts and so it was necessary for Boeing to engage in a major noise-reduction program which led to many studies investigating annoyance to aircraft noise throughout the 1960's. Perceived noise level (PNL) is the metric that had been used to quantify the perceived noisiness of aircrafts by observers on the ground; many of these studies proposed correction factors to PNL in order to account for the prominent tones that are often heard in aircraft noise, especially those with jet engines. In 1961, Little found that the judged noisiness of aircraft noises containing tonal components was greater than what was predicted using PNL. Based on these results, Little proposed a correction factor be added to PNL in order to account for strong pure-tone components. Kryter and Pearsons conducted a study in 1965 where subjects made paired comparisons of a pure-tone in noise with the same noise without the tone; the signals with the tone and noise were found to be more annoying than those without the tone. Based on these results, Kryter and Pearsons also proposed a pure-tone correction factor be added to PNL.

Throughout the 1980's, Hellman conducted a number of studies on human perception to noise that contains prominent tones; much of this work was supported by funding from the NASA Langley Research Center in order to better understand the effects of community noise on people. Hellman was interested in investigating how tonal components contribute to the perceived annoyance, loudness and noisiness of broadband noise. In 1982, Hellman used absolute magnitude estimation (AME) to obtain judgments of overall loudness, annoyance, and noisiness of single tones centered within broadband noise; Hellman conducted a similar study in 1984 as well as one in 1985 using two-tone noise complexes. Based on results from these studies, Hellman found that perception of annoyance was related to the frequency of the tone as well as the number of prominent tones contained in the noise. Although this work done by Hellman has contributed to what is known about human perception to tones in noise, these findings are limited because only a few tonal frequencies were explored. These studies were also conducted before many of the tonal metrics that are used today were developed.

2.2.2 Tonal Metrics

A number of methods have been developed in order to objectively quantify the prominence of a tone in noise, including Tone-to-Noise Ratio (ANSI S1.13-2005), Prominence Ratio (ANSI S1.13-2005), 1/3 Octave Band Analysis (ISO 1996-2), and Aures' Tonalness metric (1985). Annex A in ANSI S1.13-2005 defines two of these metrics: Tone-to-Noise Ratio (TNR) and Prominence Ratio (PR). The TNR method compares the level of the tone to the level of the broadband noise within the critical band centered on the frequency of the tone. The PR method compares the level of the critical band centered on the tone to the average level of the two adjacent critical bands, above and below the tone. These metrics were first standardized in the 1995 revision of ANSI S1.13 and have since been revised after a large study was carried out to compare and contrast both metrics with each other and with subjective perception.

Although TNR and PR were found to generally correlate well with each other as well as with subjective ratings, in some cases, the procedures gave contradicting results. Because of this, a task group of the Inter-Committee Working Group on Noise from Information Technology and Telecommunications Equipment (ITTE) was formed in order to further understand and resolve these conflicts; the goal of the task group was to optimize one method to introduce into international standards (Balant et al, 1999). The main study was to be a round robin test involving 40 signals (both synthetic and recordings from machinery) which would be subjectively rated and also objectively rated using both the TNR and PR methods. While this database of signals was compiled, Balant and colleagues conducted pilot studies (Balant et al, 1999, 2000). The methodology for these pilot studies was the same as for the main round robin test; subjects listened to a number of signals, which all contained at least one tone, and rated the prominence of the tone(s) using a 7 point scale (0=inaudible and 6=extremely prominent). Results from both studies were similar. Subjects tended to rate the signals with lower frequency tones as less prominent (Balant et al, 2000); therefore tone frequency seemed to be a factor in determining prominence.

The main test had 28 participants who were all ITTE acoustics engineers (Balant et al, 2000; Hellweg et al, 2002). Subjects were asked to listen to 40 signals (22 synthetic samples and 18 recordings of machinery noise) which all contained one or more prominent tones and subjectively rate each one using the same 7 point scale as in the pilot studies. Subjects were also asked to calculate both the TNR and PR of 20 specific signals from the original 40 (10 synthetic and 10 real). Objective results were good; PR showed slightly less variability than TNR (Hellweg et al, 2002). As with the previous studies, subjects tended to rate the signals with lower frequency tones as less prominent than those with higher frequencies, even if they had equivalent PR and/or TNR values.

Based on these results, Hellweg and colleagues recommended a number of changes be made to the existing standards for both the Tone-to-Noise Ratio (TNR) and Prominence Ratio (PR) methods. First, a low frequency correction was suggested for both TNR and PR for frequencies lower than 1000 Hz. When this correction was applied to the signals used in the main study, the correlation coefficient between the subjective ratings and both (corrected) TNR and (corrected) PR improved significantly (Hellweg et al, 2002). At the time of this study, current standards for evaluating prominence had the criteria set at 6 dB for TNR and 7 dB for PR, meaning an audible tone in noise with a calculated TNR of 6 dB or PR of 7 dB would be considered prominent. Based on the results of the round robin test, Hellweg and colleagues suggested the current criteria for prominence be increased by 2 dB at all frequencies for both TNR and PR, in addition to the low frequency correction that was also suggested. Figure 2.1 shows these recommended changes compared to the current criteria at that time.

Modifications to the PR method were also recommended based on results from this study. At the time of this work, ANSI S1.13 used a simplification for the determination of the lower and upper critical bands which were approximations of the actual Zwicker critical bands; these simplifications could lead to errors. Calculating the actual lower and upper critical bands was recommended and could be done easily with the use of modern spreadsheets and computer programs. The final recommendation was to truncate the lower critical band at 20 Hz when evaluating frequencies lower than 142 Hz.

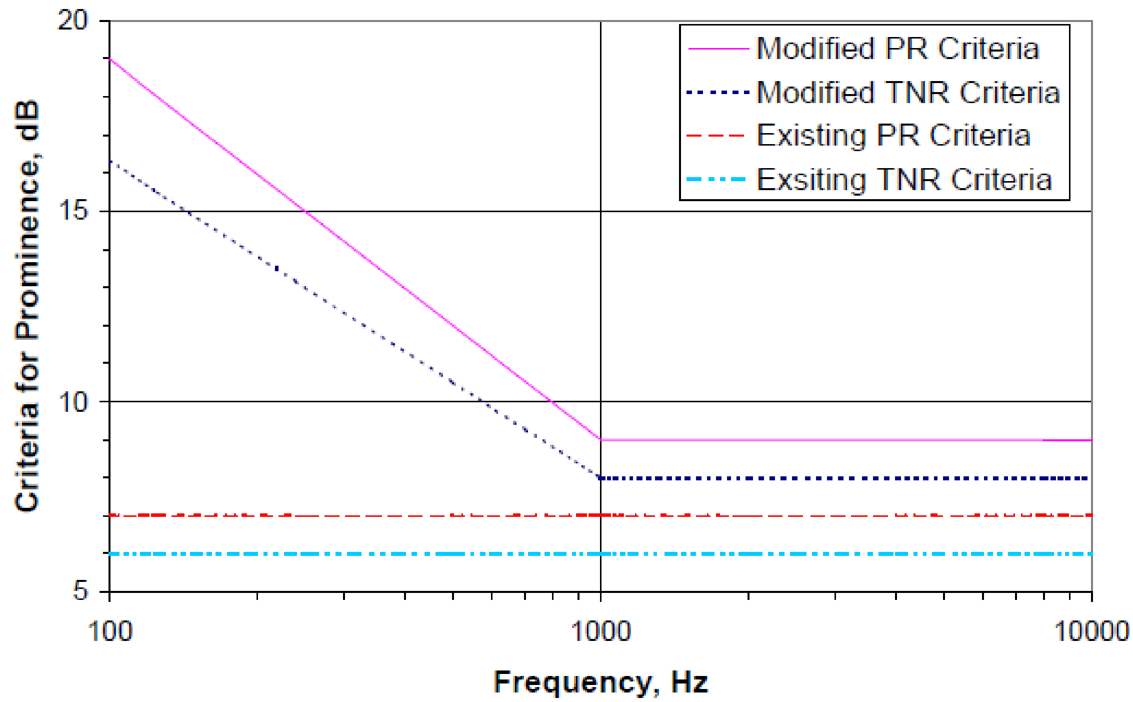

Figure 2.1 Recommended changes to criteria for prominence for TNR and PR applied to ANSI S1.13-2005. (Figure from Hellweg et al 2002).

As a result of this work and the recommendations that were made, ANSI S1.13 was revised in 2005 to take the aforementioned changes into account. The current version of ANSI S1.13-2005 lists higher criteria for prominence when using both TNR and PR as well as the modifications that were suggested for the calculations of the PR method. Hellweg and colleagues found that both TNR and PR correlated well with subjective ratings and so were unable to determine if one tonal metric was better suited than the other.

2.2.3 Recent Ventilation Noise Work

More recently, a number of studies have been conducted to investigate the effects that ventilation-like noise spectra which contain prominent tonal components have on human perception and performance. This source of tonal noise is of much interest since people are exposed to these types of conditions in most office environments. Landström, Holmberg, and colleagues conducted a number of studies on the effects of ventilation noise on annoyance and performance (Landström et al. 1991, 1993, 1995, Holmberg et al. 1993). In most of these studies, subjects were exposed to different types of ventilation-like noise while performing a task and were then asked to rate how annoying the noise was (Landström et al. 1991, 1995, Holmberg et al. 1993). Results from these studies showed that those signals containing tonal components were generally rated as more annoying than those without; annoyance increased with the presence of multiple tonal components. Annoyance ratings were higher for those signals which contained tonal components at higher frequencies; annoyance also increased when exposed to louder signals. Ratings of annoyance increased when subjects were exposed to signals for longer periods of time (60 minutes vs 30 minutes).

One study by Landström and colleagues to note was one in which subjects did not rate their annoyance but controlled the level of the noise to which they were being exposed (1993); two separate experiments were conducted within this one study. In the first, a total of 48 subjects were asked to adjust either a 100 Hz or a 1000 Hz tone to an acceptable level while working on a simple reaction time test and a more difficult grammatical reasoning task. Each subject completed both tasks and adjusted the sound level to a lower (L1) and a higher (L2) level. L1 was defined as the highest sound level where it was not necessary to exert any extra effort in order to maintain an unaffected performance; this lower limit would therefore be the threshold of annoyance. L2 was defined as the highest sound level at which it was possible to maintain unaffected performance while exerting extra effort; this upper limit would be the tolerance limit value. Subjects also rated their effort during the task as well as their performance. The second experiment replicated the first except that subjects were exposed to broadband stimuli, centered around 100 Hz or 1000 Hz, instead of a pure tone. Results from both experiments corresponded well with each other; there were no significant differences found in performance (both self-ratings and actual task performance) between the two tasks at the two different levels in both experiments. Effort was rated higher at the L2 level than at the L1 level in both experiment 1 and 2. In experiment 1, the tolerance levels for the 100 Hz tone were 33 dB above the tolerance levels for the 1000 Hz tone; in experiment 2, the tolerance levels for the noise centered around 100 Hz was 6.3 dB above the tolerance levels for the 1000 Hz noise. These results imply that tones and broadband noise need to be assessed differently and that the effect of tonality appears to be frequency dependent (Landström et al. 1993).

Within the last 10 years, a number of studies have been conducted to try to relate human annoyance to current indoor noise criteria systems (Bowden and Wang, 2005; Ryherd and Wang, 2008) in order to determine which noise criteria applies best in a variety of different ambient noise situations. In a pilot study, Bowden and Wang (2005) examined the ability of indoor noise criteria to relate to productivity scores and to the auditory perception of noise. Eleven subjects were exposed to 12 different background noises while completing two productivity tasks (a typing task and a proofreading task); subjects also provided subjective ratings of loudness, annoyance and spectral quality of the background noises. This study looked at the effectiveness of five indoor noise criteria systems in predicting human responses to background noise: Noise Criteria (NC), Balanced Noise Criteria (NCB), Room Criteria (RC), Room Criteria Mark II (RC Mark II), and A-weighted Equivalent Sound Pressure Level (L_{Aeq}). No significant correlations were found between productivity scores and subject ratings of loudness and annoyance or between productivity scores and criteria predictions of sound level. Significant correlations were found between subjective perceptions of loudness and annoyance and criteria predictions of level; a significant correlation was also found between the subjective ratings of loudness and annoyance. As expected, the signals with higher noise level ratings, as given by the five noise criteria that were studied, were perceived as louder and more annoying by the subjects. Perceptions of rumble and roar were also significantly correlated with subjective perception of loudness and annoyance, so as subjects perceived a noise to be more rumbly or roaring, they also perceived it to be louder and more annoying.

Ryherd and Wang (2008) expanded upon the aforementioned pilot study to investigate whether significant changes in subjective perception existed when comparing six noise conditions with similar criteria ratings but which had different tonality. In this study, thirty subjects completed three different tasks (typing, reasoning, and math) while being exposed to each of the six noise conditions. Subjects also completed subjective questionnaires about their perception to each noise exposure. The six noise signals consisted of one neutral condition which was broadband noise only and five tonal conditions which were broadband noise that contained a prominent tone. The tonal conditions explored 2 frequencies, 120 Hz and 235 Hz, at two Prominence Ratios (PR), PR=5 and PR=9, as well as a 595 Hz tone at PR=9. Subjects were exposed to only one noise condition per session to try to reduce biasing effects. The subjective questionnaires asked about loudness, rumble, roar, hiss, tonality, fluctuations over time, distraction, and annoyance. The authors were interested in whether significant differences in task performance and perception ratings existed between the six noise conditions. No significant differences in performance scores were found among the different noise conditions; however, there were some significant results when looking at perception ratings across the noise conditions. In general, the conditions with PR=9 were perceived as louder, more tonal, more annoying, and more distracting than the conditions with PR=5.

2.3 Applications to This Research

The goal of the current research is to expand on this previous research and try to determine thresholds of annoyance to tones in noise. This is different from the goal of Balant, Hellweg and colleagues who were focused on determining thresholds of prominence of the tone. Two methods are utilized to determine thresholds of annoyance. The first is a direct assessment with task which is similar to the method used by Ryherd (nee Bowden) and Wang (2005, 2008) although the tasks used in the two studies differ. This study aims to expand on previous research done by exploring different tonal frequencies, at a number of PR levels above different background noise levels. The tonal metric that was used in order to determine prominence is the same as was used by Ryherd and Wang; however the current study used the revised criteria for prominence and the former did not. The second method used is a magnitude adjustment method.

The results from both methods will be compared and contrasted and the pros and cons of each method will also be discussed. Results from this research will help in determining not only thresholds of annoyance but also the better method for obtaining those thresholds.

Chapter 3: Methodology

The purpose of this study was to determine the threshold of human annoyance to prominent tones in broadband noise. Two different methods were used and compared to explore human perception of two different tonal frequencies: 125 Hz and 500 Hz. In the first method, which will be referred to as Method I, subjects were asked to complete Sudoku puzzles while being exposed to broadband noise with a tonal component which was set at a specific level above the noise. They were then asked to give subjective reviews related to the noise they had just experienced. In the second method, which will be referred to as Method II, subjects were asked to listen to broadband noise with a tonal component and adjust the tone only until it became just annoying. The full study consisted of twelve thirty minute sessions with one orientation session, ten sessions in Method I and one session in Method II; all subjects completed the sessions in this order.

This chapter discusses the methodology of the study and is separated into two sections: (1) facilities used for testing and (2) experimental methods.

3.1 Facilities

3.1.1 Nebraska Test Chamber

All testing was completed at the Peter Kiewit Institute on the Pacific campus of the University of Nebraska – Omaha. Test sessions were held in a test chamber that is acoustically isolated from nearby spaces with a FSTC rating of 30 between the test chamber and monitor room. The test chamber resembles an ordinary office with carpet, gypsum board walls and acoustical ceiling tiles. The chamber measures approximately 9’ x 13’ with the short wall slanted approximately 8 degrees off angle and the long wall slanted approximately 6 degrees off angle. The test chamber is further acoustically treated with four bass traps – one in the front right corner, one in the rear right corner and two in the rear left corner – as well as 2 one-inch thick 4’x8’ Tectum panels hung on the left and rear walls. The average mid-frequency reverberation time from 500 Hz to 1000 Hz was measured to be 0.31 seconds. A layout of the test chamber is shown in Figure 3.1.

The test room contains a chair with a built-in desk, a 23.5” computer monitor and a wireless mouse; the monitor and wireless mouse were used only in Method II of the study. The chair was situated in the room so that the subject’s head was 4’3” from the wall that is shared with the monitor room, 5’2” from the back wall and 4’5” from the computer monitor. A height of 3’6” was used as the distance from the ground to the subject’s head. This height was considered to be the average height of the subjects’ ears as they completed the tasks in each method. This is the height at which the microphone was placed when making the measurements described in Section 3.2.2.

The loudspeakers that were used to play the test signals were an Armstrong i-Ceiling loudspeaker and a JBL Northridge ESeries Subwoofer. The i-Ceiling loudspeaker is made to resemble an acoustical ceiling tile and was situated next to a diffuser in the ceiling. The subwoofer was covered in fabric and placed in the corner of the room; this was necessary in order to produce the lower frequency content of the background noise. Two additional loudspeakers, used in an unrelated study, were also in the room located on each side of the computer monitor. A third loudspeaker, located in the back of the room behind the subject, was in the room for some weeks. This was also used in an unrelated study and was removed from the room when that study was completed. All non-utilized loudspeakers were covered in fabric and subjects were told to ignore them during their test. Two photographs of the interior of the test room can be seen in Figure 3.2.

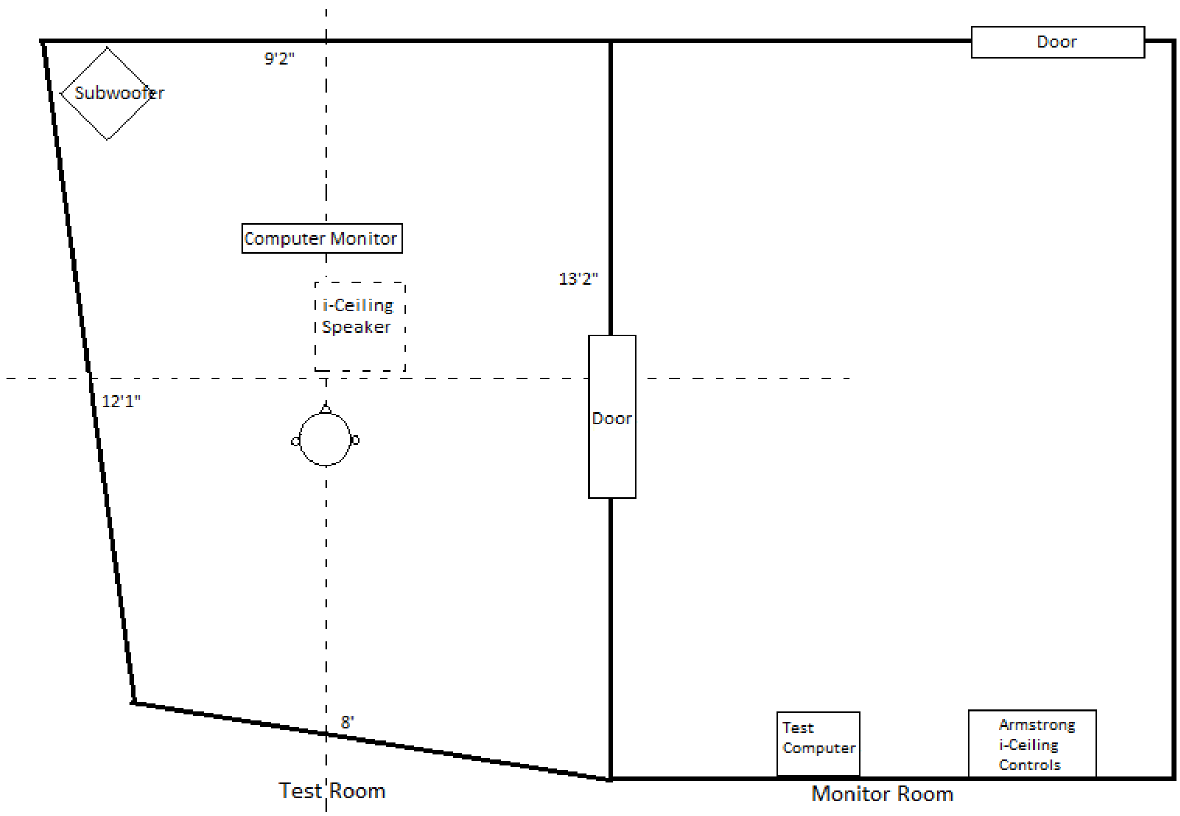

Figure 3.1 The layout of the Nebraska Test Chambers showing locations of the subject, loudspeakers and test equipment used in this study (not to scale). The height of the test room is 8'5".

The monitor room is located next to the test room. This room houses the test computer which is used for playing the signals as well as the power amplifier for the loudspeakers; the test monitor ran the test sessions from this room. The temperature of the test room was also controlled as best as possible from a thermostat in the monitor room; an average of 74.6 ^{o}F was measured across all test sessions.

Figure 3.2 Interior photographs of the test room.

3.1.2 Sound and Computer Systems

Configurations of the loudspeakers and other test equipment for both Method I and Method II of the study are shown in Figure 3.3. The test computer generated all signals in Method I and also ran the JavaScript program that was used in Method II. The computer monitor (located in the test room) and the wireless mouse used in Method II were connected to the test computer. By placing the test computer and loudspeaker controls in the monitor room, the only additional background noise sources in the test room were the loudspeakers and the HVAC system.

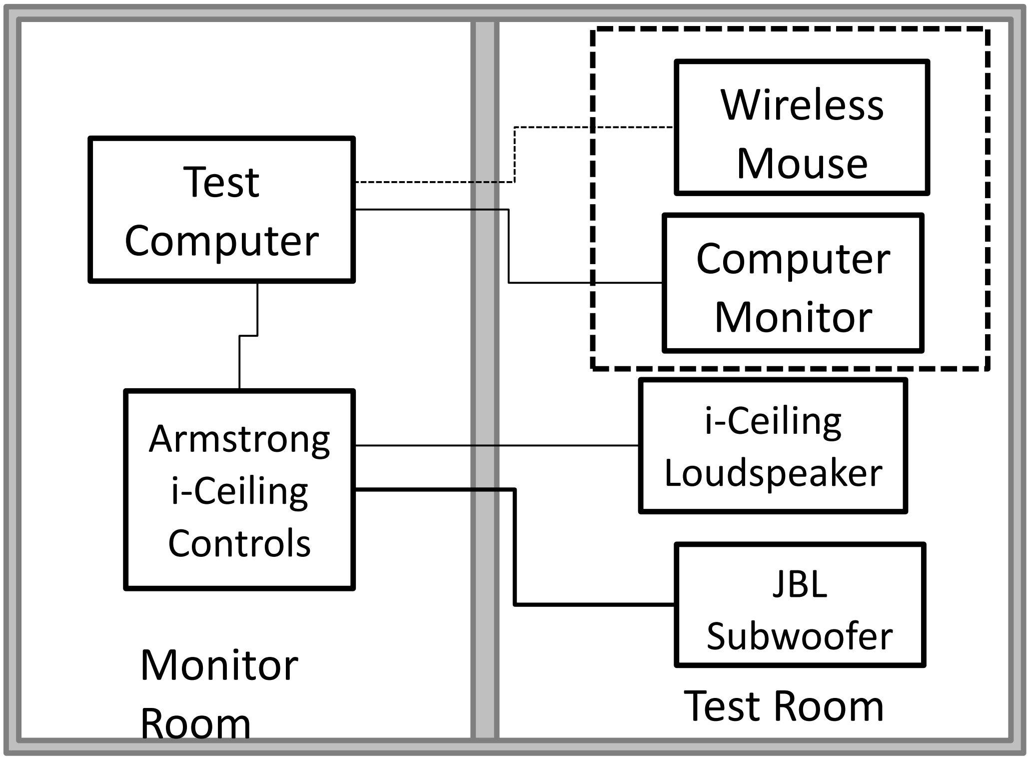

Figure 3.3 A diagram of the Nebraska Test Chamber system showing both the testing system and the sound system. The equipment within the dotted line box were only used in Method II.

3.2 Experimental Methods

This section reviews the experimental methodology of the study and is separated into four subsections: (1) creation of the test signals, (2) measurement procedures and tonal metrics used to analyze the signals, (3) procedure involved with the creation and running of each test session and (4) statistical analyses used for data analysis.

3.2.1 Signal Creation

The signals used in both Method I and Method II were created using the same procedure; all of the signals that were used in Method I were also used in Method II. All signals consisted of broadband pink noise with a tonal component of either 125 Hz or 500 Hz set at a certain level so that it was audible. Signals were created using Esser Audio's

Test Tone Generator software. Pink noise was generated on the left channel at the same level for each signal and a sine wave was generated on the right channel at different levels, relative to the noise, at the two frequencies of interest. A screenshot of the interface of the software that was used is shown in Figure 3.4.

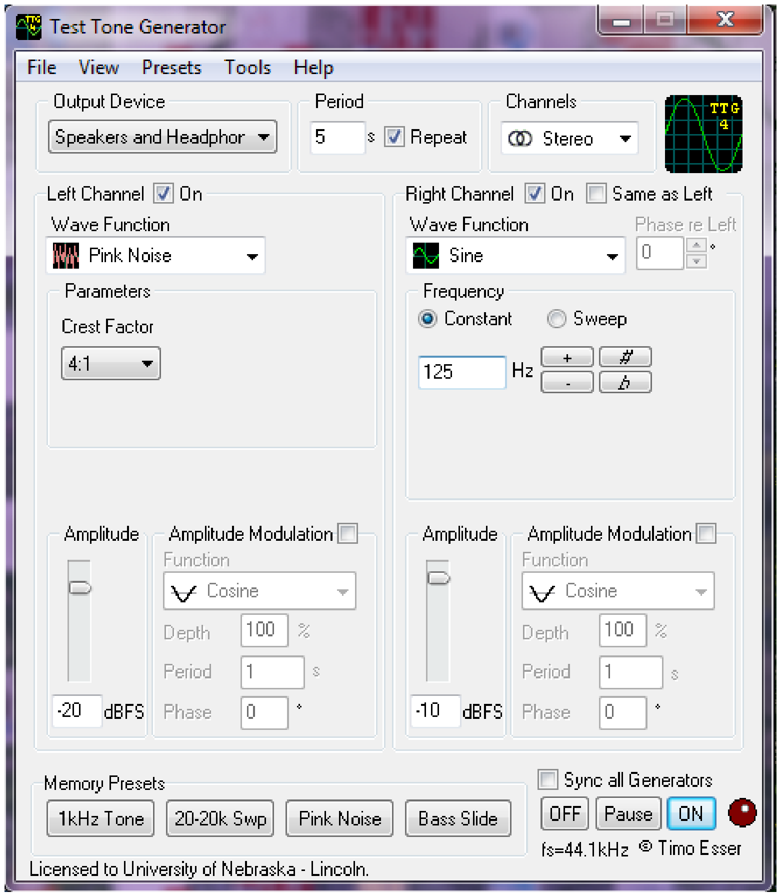

Figure 3.4 Screenshot of signal generation using Esser Audio's Test Tone Generator software. In this example, pink noise is generated on the left channel at -20 dBFS and a 125 Hz sine wave is generated on the right channel at -10 dBFS.

Each signal was looped and calibrated using the equalizer in CoolEdit until the pink noise was measured at 40 dBA or 55 dBA in the test chamber while being played over the i-Ceiling loudspeaker and subwoofer.

In Method I, five levels of each of the two tones were tested above two different background noise levels for a total of 20 test signals. In each thirty minute session subjects were exposed to three different signals for ten minutes each. The first and third signal in each consisted of broadband noise with a tonal component and the second signal was the broadband noise without a tone; the background noise level of the signals remained at a constant level throughout each session. The order of presentation of the background noise levels and signals was randomized using Latin squares. Subjects attempted to complete a packet of four Sudoku puzzles while being exposed to each signal for nine minutes. In the remaining minute, they were asked to fill out a short subjective questionnaire before the signal was changed. These steps were repeated twice more during each thirty minute session.

The same two background noise conditions and the same two frequencies that were explored in Method I were also used in Method II for a total of four background noise level/tonal frequency conditions. These four conditions were repeated once for a total of eight trials. In each trial, subjects “adjusted” the level of the tone using a JavaScript program that was created for this study. A total of 46 signals were created for each of the four conditions so that subjects would have a wide range of tonal levels to work with. The signals were created in the same way as described above with the level of the tone changed by 1 dB. When the subject adjusted the tone, the JavaScript program played the next signal, i.e. if the subject adjusted the signal ‘up’, the program would play the signal with the tonal component that is 1 dB louder than the previous one.

In Method I, signals were played over the two loudspeakers using WinAmp; signals were measured again to ensure they were still calibrated correctly. Signals were played through the JavaScript program in Method II; measurements were done to ensure calibration. In both parts, the stereo signals were merged into mono signals when played through the loudspeakers.

3.2.2 Signal Measurements and Tonal Metrics

In order to determine how prominent the tones in noise were, an objective tonal metric was necessary. For the purposes of this study, the Prominence Ratio (PR) was calculated for each signal. The PR was used as a guide to ensure the signals used contained tones that, by definition, were classified as prominent. All signal files were measured using Brüel and Kjær's Pulse analyzer system; the microphone was placed at the head position that was defined in Section 3.1.1. As each signal was measured, the Pulse system also recorded the signal as a .wav file as it played through the loudspeakers in the test chamber. This recording was saved for archival purposes. The data acquisition hardware for the Pulse system as well as the laptop that was used to run the necessary software was kept in the monitor room as to not add extra noise while measuring and recording the signals. A diagram of the measurement and recording setup is shown in Figure 3.5.

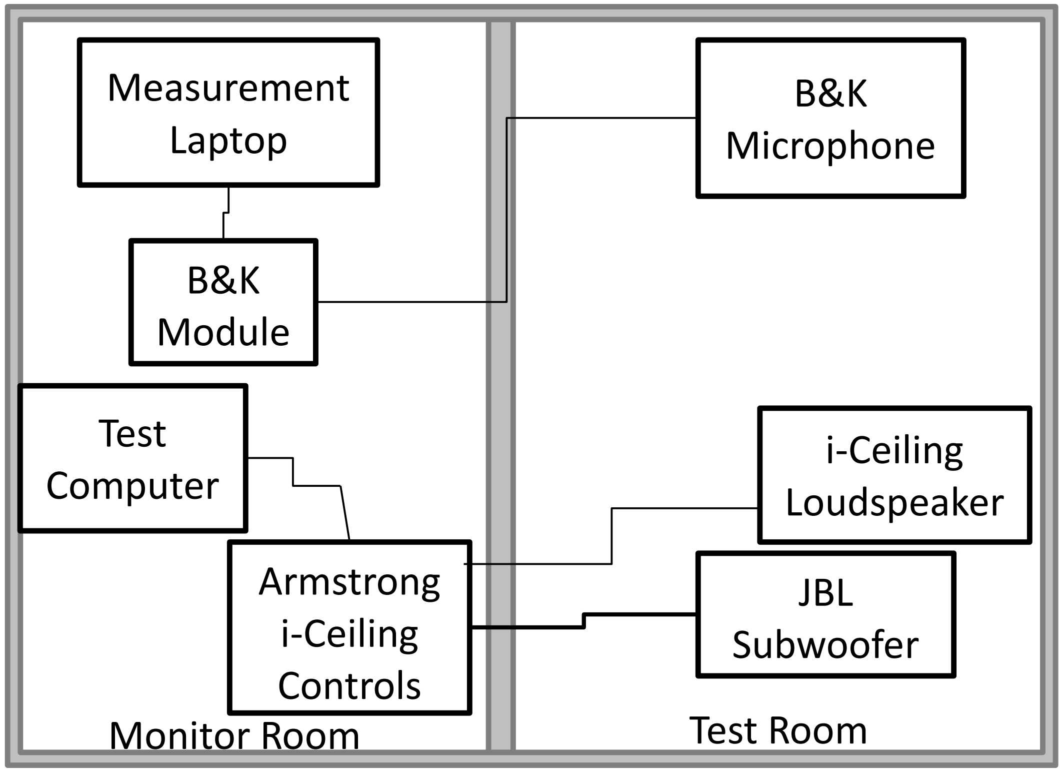

Figure 3.5 A diagram of the Nebraska Test Chamber system showing the equipment used for measuring and recording test signals in the test room.

The Prominence Ratio (PR) of each signal was calculated according to the procedure outlined in Annex A of American National Standard ANSI S1.13-2005. Prominence Ratio is an objective measure that can be used to characterize the prominence of a tone; that is, how much the tone stands out from the surrounding noise. A discrete tone which occurs in broadband noise is partially masked by the part of the noise contained in a narrow frequency band, called the critical band, which is centered at the frequency of the tone. The width of the critical band is a function of frequency. A tone is classified as prominent if the difference between the level of the critical band centered on the tone and the average level the adjacent bands is equal to or greater than 9 dB for a tonal frequency of 1,000 Hz and higher, and by a greater amount for tones at lower frequencies. For this study's two frequencies of interest, 500 Hz and 125 Hz, the PR would need to be greater than or equal to 12 dB and 18 dB, respectively, for the tone to be considered prominent.

The middle, upper and lower critical bands of each of the two frequencies were calculated according to the equations given in Sections A.8.2, A.8.3 and A.8.4 in ANSI S1.13-2005. The Fast Fourier Transform (FFT) was taken of each signal using the Pulse system. The level of each of the critical bands was determined from the FFT spectrum by bracketing the data points lying between the lower and upper band-edge frequencies and computing the sound pressure level of that band. These values were then plugged into an equation to determine the prominence ratio.

Table 3.1 gives PR's that were selected for use in Method I. These PR's were selected as they are above and below the threshold of prominence as defined by ANSI S1.13-2005. These 20 signals were also used in Method II along with 164 other signals that had both higher and lower PR values.

BNL	40dBA	55dBA	40dBA	55dBA
Tonal Component Frequency	125 Hz		500 Hz	
PR	15	13	9	6
	18*	15	12*	9
	21	18*	15	12*
	24	21	18	15
	27	24	21	18

Table 3.1 A table of the PR's that were used for the signals in Method I. The asterisk indicates prominence by definition of PR.

3.2.3 Test Session Procedure

The following section discusses the preparation and implementation of testing procedures. It contains three subsections: test session scheduling, test session design and procedure, and recruitment and orientation procedure.

3.2.3.1 Test Session Scheduling

The overall study consisted of an orientation session and eleven regular test sessions, each of which were 30 minutes long. There were two parts to the study: Method I consisting of 10 test sessions and Method II consisting of just one test session. All subjects completed the orientation session first, followed by all 10 sessions for Method I and then the one session for Method II. Subjects were only allowed to participate in one session per day.

The presentation of the test signals in Method I was determined using a Latin square design to avoid a test order bias. One 10x10 Latin square was used to determine the background noise level of each of the 10 sessions; subjects experienced 40 dBA for 5 sessions and 55 dBA for 5 sessions. Two more 10x10 Latin squares were used to randomize the order of the frequency component and its prominence.

The order of the eight trials in Method II of the study were randomized using a Latin square design. An 8x8 square was used for the first 8 subjects; the order for the remaining two subjects was determined using a random order function in Microsoft Excel.

3.2.3.2 Test Session Design and Procedure

The Sudoku puzzles used in Method I of the study were downloaded from www.veryfreesudoku.com. All puzzles were taken from the Novice level books available. Sudoku is a logic-based, number placement puzzle. Each puzzle consists of a 9x9 square grid that is subdivided into nine 3x3 boxes. The object is to fill the 9x9 grid with numbers so that each row, column and 3x3 box contains all the digits 1-9. Some numbers are already filled in on the puzzle and difficulty increases as the number of pre-printed numbers decreases. The novice level puzzles used for this study had 38-43 blank spaces with an average of 40.7 blank spaces per puzzle. An example of one puzzle is shown in Figure 3.6. Difficulty was held constant throughout all 10 sessions.

Subjects were given a packet containing four Sudoku puzzles to be completed while being exposed to each test signal. It was not expected that subjects would be able to complete all four puzzles in the allocated time and none did. After nine minutes of working on the Sudoku puzzles while being exposed to the signal, subjects then took one minute to complete a short paper-based 5-question subjective questionnaire (Figure 3.7).

A JavaScript program was designed exclusively for use in Method II of this study. The program was designed so that the test signals could easily be uploaded and sorted into 8 trials and then those trials were used to create tests. Each subject completed one test which contained 8 trials. A screenshot of the Java program while trials and tests were being created can be seen in Figure 3.8.

			2		4	8	1	
	4				8	2	6	3
3			1	6				4
1				4		5	8	
6	3	5	8	2				7
2			5	9		1		
9	1		7				4	
			6	8		7	2	1
8			4		3		5	

1_1.1

Figure 3.6 An example of a Sudoku puzzle.

ID_____ Session #______ Signal________ Date______ Time______ Temp____

Questionnaire

Please answer the following questions in relation to the task you just performed and the noise you were exposed to. Please do not skip any questions.

Please base your ratings only on the noise you just heard. Do not base your ratings on any previous noises you've heard or sessions you've attended.

1	How mentally demanding was the task?	not demanding	very demanding
2	How successful were you in accomplishing what you were asked to do?	perfect	unable to complete
3	How hard did you have to work to accomplish your level of performance?	not hard	very hard
4	How loud was the noise?	very quiet	very loud
5	How annoying was the noise?	not annoying	very annoying

Figure 3.7 A copy of the subjective questionnaire that subjects completed at the conclusion of each signal they were exposed to.

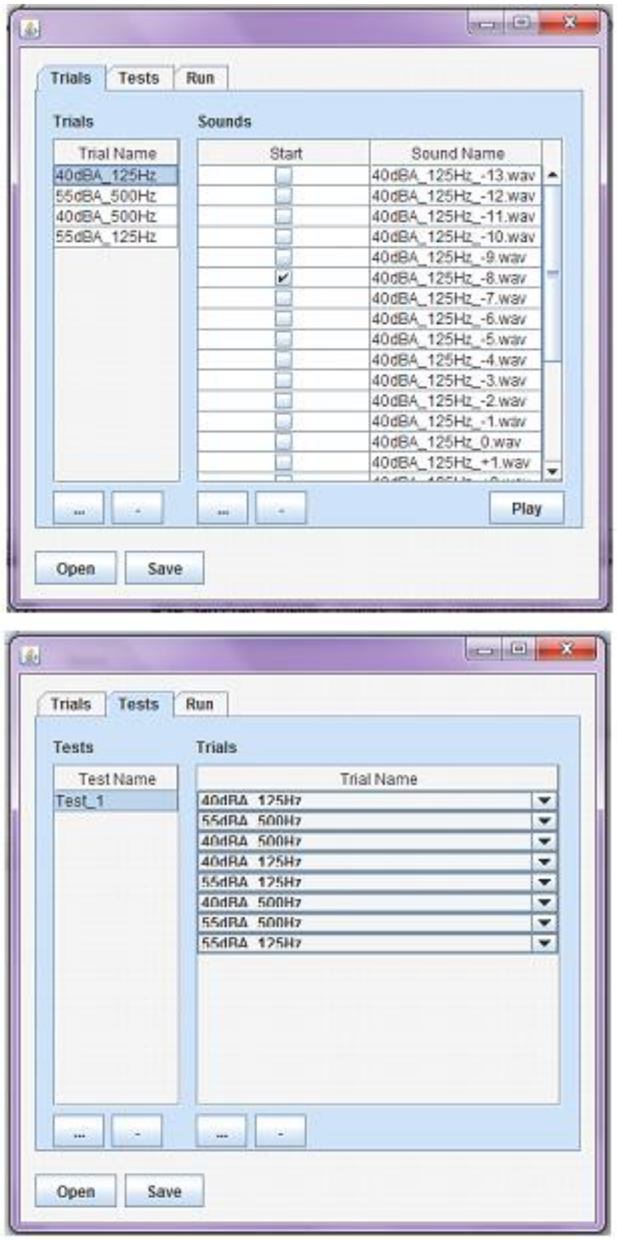

Figure 3.8 Screenshots of the creation of trials and tests in the JavaScript program that was used in Method II.

Subjects were given thirty minutes to complete Method II of the study, although most finished in less time. They were asked to listen to a signal and then to adjust it, up or down, until it became just annoying. When the signal was adjusted, the tonal component was changed, i.e. if the signal was adjusted up, the tonal component was increased by +1 dB relative to the noise. Subjects were not told how the signal would change when adjusted. Each trial started on a signal that was just below the threshold of prominence as defined by PR; for the 125 Hz conditions the start signal had a PR of 11 and for the 500 Hz conditions the start signal had a PR of 17. There were 46 signals in each trial. If the subject reached the first or last signal in a trial, a warning box would pop up to let them know that they could then continue to adjust the signal or select the minimum or maximum signal as being just annoying. The JavaScript program exported a .csv file with data after the completion of each test. A screenshot of the program while in use can be seen in Figure 3.9.

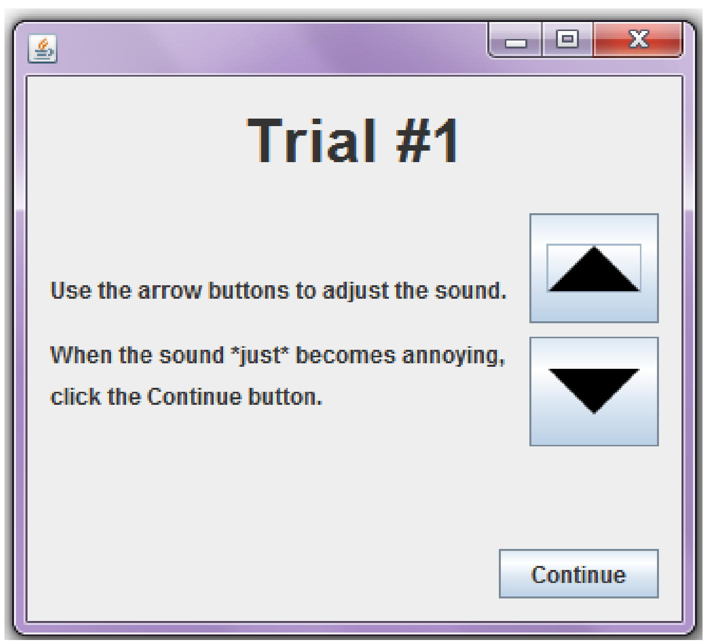

Figure 3.9 Screenshot of the JavaScript program used in Method II.

3.2.3.3 Recruitment and Orientation Procedure

Subjects were recruited by fliers posted on the University of Nebraska – Omaha campus (see Figure 3.10). The first session in which each subject participated was an orientation session. At this session, subjects completed necessary paperwork, participated in a hearing screen and were presented with a PowerPoint presentation covering the instructions for testing.

Figure 3.10 A copy of the recruitment flyer that was posted on the campus of the University of Nebraska – Omaha.

An audiometer was used to test hearing thresholds of both ears individually and was administered in the test room. Pure tones of each octave band between 125 Hz and 8,000 Hz were individually presented first at 30 dB hearing level (HL). If the subject failed to respond at this level, the level was increased by 5 dB; if the subject responded to this level, the level was decreased by 5 dB. This continued until 20 dB HL was reached. Subjects were required to have a hearing threshold at or below 25 dB HL to participate; all subjects passed the hearing screen.

The subject then viewed a training PowerPoint presentation which explained procedures for both Method I and Method II of the study. The instructions for completing the Sudoku puzzles were also explained and each subject completed a sample Sudoku puzzle to ensure understanding of the objective. After completion of the training presentation, subjects were asked to complete a demographic/noise sensitivity questionnaire. The questionnaire was taken from the reduced version of the Noise Sensitivity Questionnaire (NoiseQ) developed by Schutte et al (2007) the questionnaire is shown in Figure 3.11. Total noise sensitivity for each participant was calculated based on the information provided on the questionnaire.

Demographic Questionnaire

Please provide your gender and age.

Gender: Male Female

Age: _______________

Please rate each statement in order. Please do not skip any questions. If possible, imagine yourself in each situation and response accordingly without spending too much time considering if you agree or disagree with a given statement. We are looking for your personal opinions. There are no correct or incorrect responses.

		Strongly agree	Slightly agree	Slightly disagree	Strongly disagree
1	I need an absolutely quiet environment to get a good night's sleep.	1	2	3	4
2	I need quiet surroundings to be able to work on new tasks.	1	2	3	4
3	When I am at home, I habituate to noise quickly.	1	2	3	4
4	I become very agitated if I can hear someone talking while I am trying to fall asleep.	1	2	3	4
5	I am very sensitive to neighborhood noise.	1	2	3	4
6	When people around me are noisy I don't get on with my work.	1	2	3	4
7	I am sensitive to noise.	1	2	3	4
8	My performance is much worse in noisy places.	1	2	3	4
9	I do not feel well rested if there has been a lot of noise the night before.	1	2	3	4
10	It would not bother me to live on a noisy street.	1	2	3	4
11	For a quiet place to live I would accept other disadvantages.	1	2	3	4
12	I need peace and quiet to do difficult work.	1	2	3	4
13	I can fall asleep even when it is noisy.	1	2	3	4

Figure 3.11 A copy of the demographic/noise sensitivity questionnaire that subjects completed during the orientation session.

3.2.4 Statistical Analysis

The data gathered in this study were statistically analyzed using Microsoft Excel and SPSS. Both perception data and performance data were analyzed, although this study was mainly concerned with the results from the perception data. Perception data were gathered from the subjective questionnaires. Performance data were collected in the form of the number of puzzles completed (including partially completed puzzles) as well as accuracy. Accuracy was calculated by taking the number of correct answers divided by the number of attempted answers; this included both complete and incomplete puzzles.

Most of the data gathered was suitable for the use of parametric tests for data analysis; however, both parametric and non-parametric tests are used and presented. Data may be considered suitable for parametric tests if it meets the following conditions: data are measured at an interval or ratio level, data sets have roughly the same variances, and data are distributed normally. Normal distribution of the data sets was determined by using the Shapiro-Wilk test. This test compares the set of data in the sample to a normally distributed data set with the same mean and standard deviation (Field and Hole, 2003). If the test is significant ($p<0.05$) then the data are not normally distributed.

3.2.4.1 Standard Error of the Mean

The standard error of the mean (SE) is the standard deviation of the sample means and is reported as error bars in the results graphs in the following chapter. Large values of SE mean that there is more variability between sample means and, therefore, a given sample may not be representative of the population. SE is found by Equation 3.1:

$$SE = \frac{s}{\sqrt{N}} \qquad (3.1)$$

where s is the sample standard deviation and N is the sample size (Field and Hole, 2003).

3.2.4.2 Parametric Tests

General relationships between a dependent variable and an independent variable were determined using Pearson Product Moment Correlations and linear mixed model analysis. An example of a general relationship is the relationship between perceived annoyance rating and signal presented. Any significant relationships were reported using these two statistical tests. The Pearson Product Moment Correlation reported the correlation, r, between the two variables and the linear mixed model reported the F value with degree of freedom, df, from the numerator and denominator. These tests are reported, along with their respective significances, in the following format: $F_{dfn,dfd}$ = ____, r = ____, where dfn is numerator degrees of freedom and dfd is denominator degrees of freedom.

Repeated measures analysis of variance (ANOVA) was used to compare a single dependent variable across multiple independent variables. An example of this comparison was the relationship between perceived annoyance ratings and the different levels of tonalness (or PR). Each repeated measures ANOVA test statistic, F, is reported with significance in the following format: F(df,N) = ____, where df is degrees of freedom and N is the sample size. The effect size,ω, was found by Equation 3.2:

$$\omega = \sqrt{\frac{MS_M - MS_R}{MS_M + ((n-1) \times MS_R)}} \qquad (3.2)$$

where MS_M is the mean sum of squares, MS_R is the mean squared error and n is the sample size. When the F statistic was significant, Bonferroni ad hoc tests were used to find significant differences between group means (Field and Hole, 2003)

3.2.4.3 Non-Parametric Tests

In most cases, data were found to be non-normally distributed so non-parametric tests were appropriate for data analysis. The parametric tests described in the previous subsection may not provide accurate results when used on data that are not normally distributed. The non-parametric equivalents of the tests described above were used and compared to the results of the parametric tests. Spearman Correlation, r, is used in place of the Pearson Product Moment Correlation to find the general relationship between a single dependent and independent variable.

Friedman's ANOVA is used in place of the repeated measures ANOVA to compare a single dependent variable across multiple independent variables. Each Friedman's ANOVA test statistic, χ^2, is reported with significance in the following format: χ^2(df) = ____, where df is the degrees of freedom. To find exactly where differences between group means lie, a Wilcoxon test was used with a Bonferroni correction. The Wilcoxon test statistic, T, was reported with the effect size, r. The effect size was calculated using Equation 3.3:

$$r = \frac{Z}{\sqrt{N}} \tag{3.3}$$

where Z is the z-score produced by SPSS and N is the total number of observations compared. The z-score, or standard score, is a measure of how far a data point is above or below the population mean that is expressed in terms of standard deviations (Field and Hole, 2003).

Chapter 4: Results and Discussion

This chapter presents results from analyses of the data collected from the subjective questionnaires and task performance scores in Method I and the PR of the signal selected to be just annoying in Method II. The effects of prominence ratio, tone frequency and background noise level are also investigated. All results are reported and analyzed using the statistical analysis methods described in the previous chapter. The results from both methods are compared and contrasted.

4.1 Demographic Results

Ten subjects participated in this study: four females and six males. Subjects ranged in age from 25 to 43 with an average age of 31.

During the orientation session, all subjects completed a NoiseQ-R survey (Schutte et al, 2007) which is a noise sensitivity questionnaire; results from the questionnaire are shown by question in Figure 4.1. These results were weighted and calculated into sleep, work, residential and total noise sensitivity percentages. These calculations were based on work done by Schutte et al. (2007); each question was weighted and then an average calculated to find the overall sensitivity percentage for each subject. Overall noise sensitivities ranged from 33% (slightly sensitive) to 92% (very sensitive) with an average overall noise sensitivity of 66.9% and a standard error of the mean of 6.2%. A histogram of overall noise sensitivities is shown in Figure 4.2.

4.2 Method I – Direct Assessment with Task Results

The results from Method I are divided into three subsections: (1) task performance results, (2) subjective rating results, and (3) thresholds of annoyance.

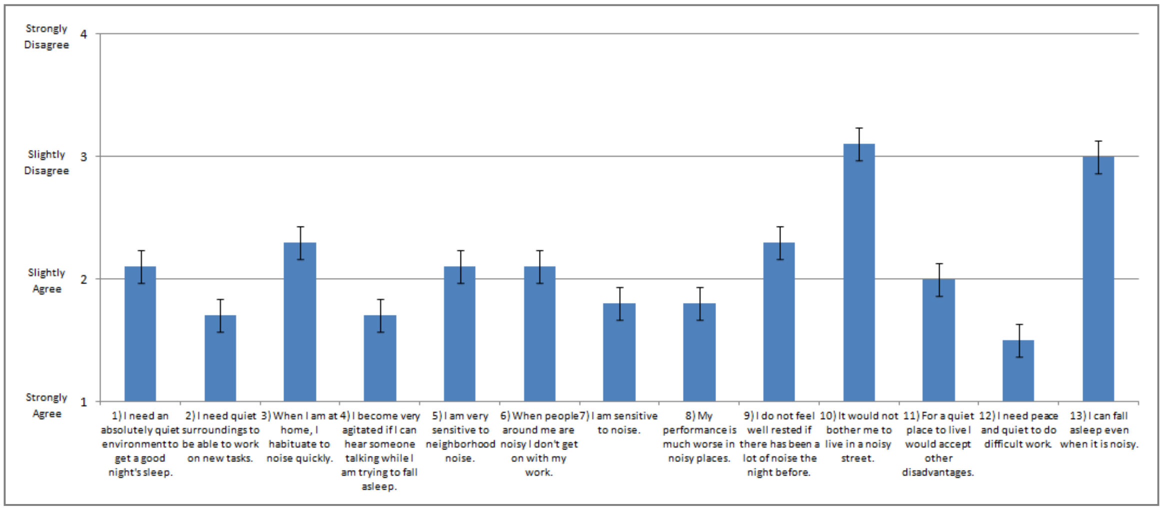

Figure 4.1 Results of the noise sensitivity questionnaires averaged across all subjects. Error bars represent the standard error of the mean.

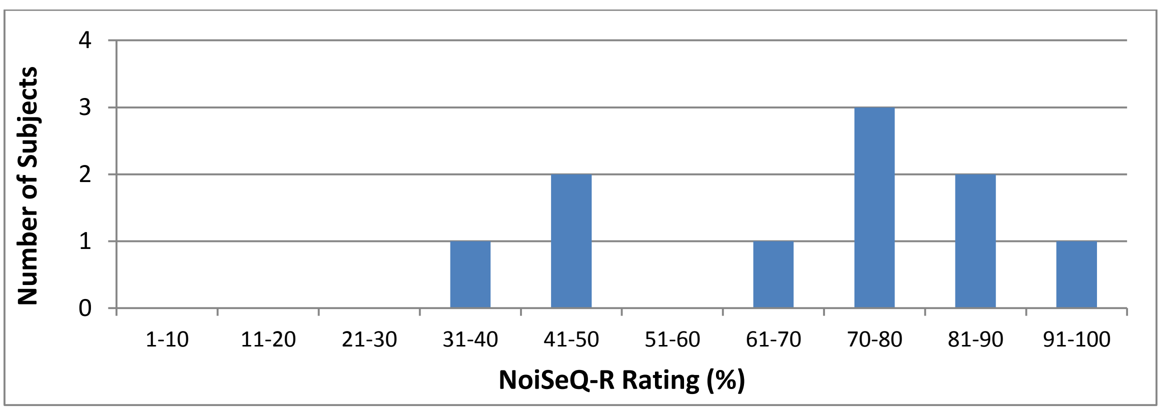

Figure 4.2 Histogram of subjects' overall noise sensitivity ratings.

4.2.1. Task Performance Results

Task performance was measured in terms of the number of Sudoku puzzles completed within each 9 minute tonal noise exposure (including started but not complete puzzles) and accuracy (also including incomplete puzzles). Statistical analyses using SPSS were performed as described in Section 3.2.4. Results were tested for normal distribution using the Shapiro-Wilk test. Data for the number of puzzles completed were found to be normally distributed; however, the accuracy data were found to be non-normally distributed. Because of this, both parametric and non-parametric test results are reported for this data.

Data from the number of completed puzzles can be seen in Figure 4.3. Figure 4.3 shows the number of completed puzzles across all tonal noise conditions, including the background noise only condition which is separated from the tonal conditions by a black line. Frequency and background noise level are distinguished by color with blue for the 125 Hz condition and green for the 500 Hz condition; the lower background noise condition (40 dBA) is represented by the lighter shade of each color and the higher background noise condition (55 dBA) is represented by the darker shade.

Repeated measures ANOVA was used to find any differences among the number of completed puzzles for each BNL and tone frequency condition. Looking at only the data to the right of the line in Figure 4.3, it can be seen there is not much variance among the number of completed puzzles. This was confirmed through the ANOVA tests that were run; no significant differences were found for the number of completed puzzles among the different groups. Although the data were found to be normally distributed, a Friedman's ANOVA was also used to confirm the results of its parametric equivalent.

The number of puzzles completed while being exposed to the signals that contained tones was also compared to the number of puzzles that were completed while being exposed to the background noise only signals, to see if there was an effect of prominent tones on performance. These results can be seen in Figure 4.3. The average number of completed puzzles for the background noise only conditions was found to be the same for both background noise levels with subjects completing an average of 1.39 puzzles during both the 40 dBA background noise and the 55 dBA background noise condition.

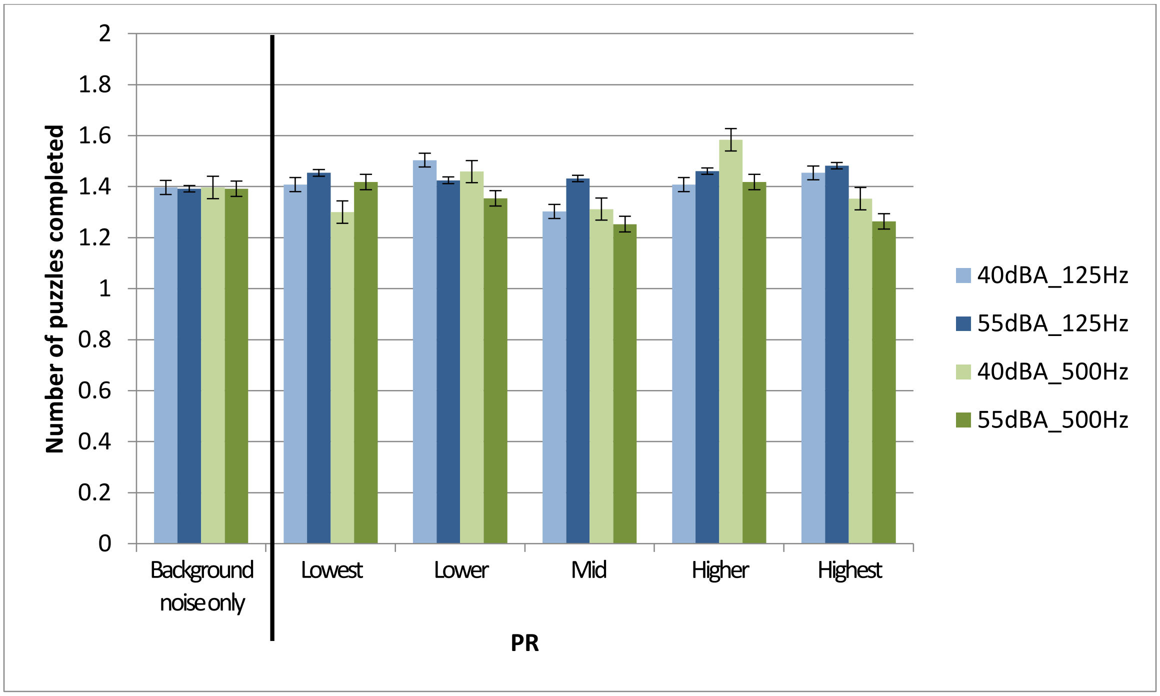

Figure 4.3 Number of completed Sudoku puzzles across all noise conditions, including the background noise only signals, and PR's. Error bars represent the standard error of the mean.

Repeated measures ANOVA was used to find differences between the average number of completed puzzles while being exposed to the tonal signals and while being exposed to the background noise only signals. No significant differences were found. This was again confirmed with a Friedman's ANOVA which did not produce any significant results.

The second task performance measure was accuracy, or percent correct, of the Sudoku puzzles. Both complete and incomplete puzzles were graded. Subjects were not penalized for blank squares, however; only incorrect answers counted against them. Accuracy data can be seen in Figure 4.4. Figure 4.4 shows the accuracy across all tonal noise conditions including the background noise only conditions with a black line separating the tonal conditions from the background noise only conditions. Frequency and background noise level are distinguished by color with blue for the 125 Hz condition and green for the 500 Hz condition; the lower background noise condition (40 dBA) is represented by the lighter shade of each color and the higher background noise condition (55 dBA) is represented by the darker shade.

Because these accuracy scores were found to be non-normally distributed, a Friedman's ANOVA was used to find any differences in the scores among the four BNL and tone frequency conditions. No significant differences were found between any of the group means. A repeated measures ANOVA was also run to support the results of the non-parametric test and no significant results were found.

The accuracy data while being exposed to the signals that contained tones were also compared to the accuracy that was achieved while being exposed to the background noise only signals to see if there was an effect of the tones on performance. These results can be seen in Figure 4.4. As with the average number of completed puzzles, the average accuracy for the background noise only conditions was found to be the same around 95% correct for both background noise levels. However, there is much greater variability among the accuracy scores when subjects were exposed to the tonal signals which may suggest tonal noise does impact performance.

Friedman's ANOVA was used to find any differences between the accuracy while being exposed to the tonal signals and the accuracy while being exposed to the background noise only signals. No significant differences were found. A repeated measures ANOVA was also run to support these results and no significant findings were produced.

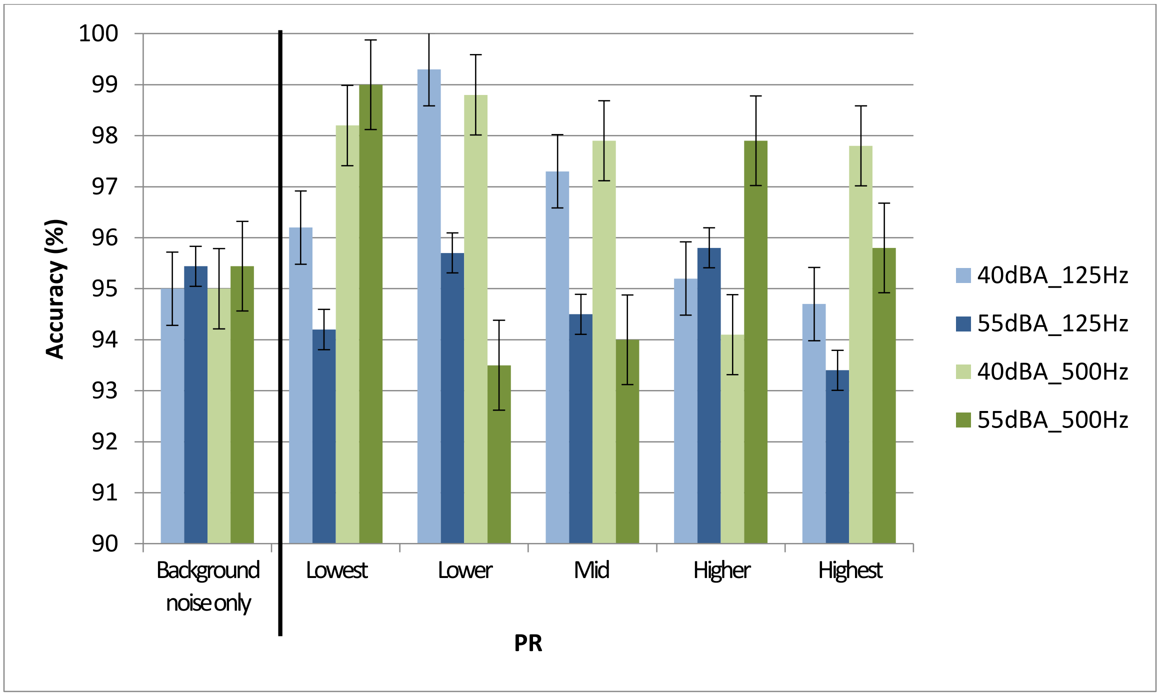

Figure 4.4 Overall percentage of correct answers across all tonal noise conditions and PR's as well as both background noise conditions. Error bars represent the standard error of the mean.

4.2.2 Subjective Perception Results

After being exposed to each of the three signals in each session in Method I, subjects rated their perception of mental demand of the task, overall success in accomplishing the task, effort put forth in completing the task, loudness of the noise, and annoyance to the noise. A total of thirty subjective questionnaires were completed by each subject. The results from the completed subjective questionnaires after being exposed to the signals with prominent tones are shown in Figures 4.5-4.9.

The data were tested for normality using the Shapiro-Wilk test and some data exhibited a normal distribution while some were found to be non-normal. Because of this, both parametric and non-parametric analyses were utilized to determine if there were any significant differences in the ratings for the different tonal, background noise level and PR conditions.

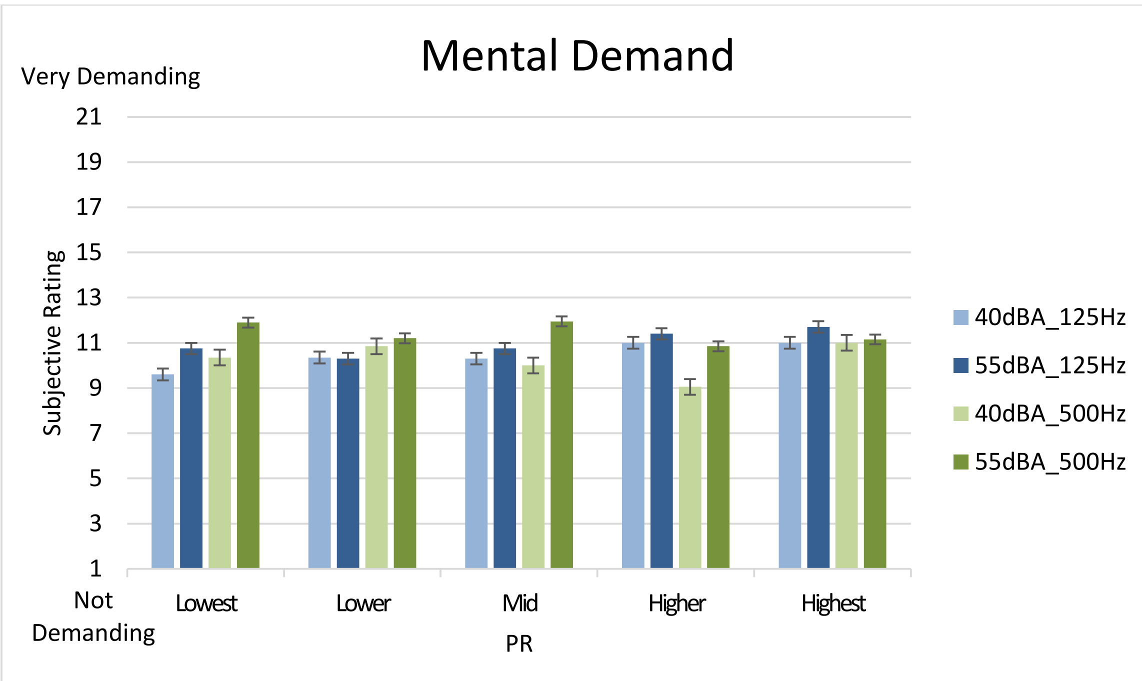

Figure 4.5 Subjective rating results of mental demand for all background noise level, tonal frequency and PR conditions. The error bars represent the standard error of the mean.

Figure 4.5 indicates that there seem to be some variance in mental demand rating among the different signals and this was supported by the ANOVA tests. Significant differences in mental demand ratings between the two background noise levels were found, $F(1,10) = 6.699$, $p<.05$. Subjects rated their mental demand while being exposed to the quieter background noise signals as slightly lower than the mental demand for the louder background noise levels. No other significant differences were found; all results from this test can be found in Table 4.1.

Mental Demand				
Source	**SS**	**df**	**MS**	**F**
BNL	35.701	1	35.701	6.699*
Freq	0.661	1	0.661	0.098
PR	10.305	1.845	5.584	0.253
BNL*Freq	4.961	1	4.961	1.679
BNL*PR	11.055	2.438	4.535	0.458
Freq*PR	31.67	2.082	15.213	0.881
BNL*Freq*PR	7.12	1.785	3.989	0.223

Table 4.1 ANOVA results for the subjective measure mental demand. The asterisk indicates significance at the $p<0.05$ level.

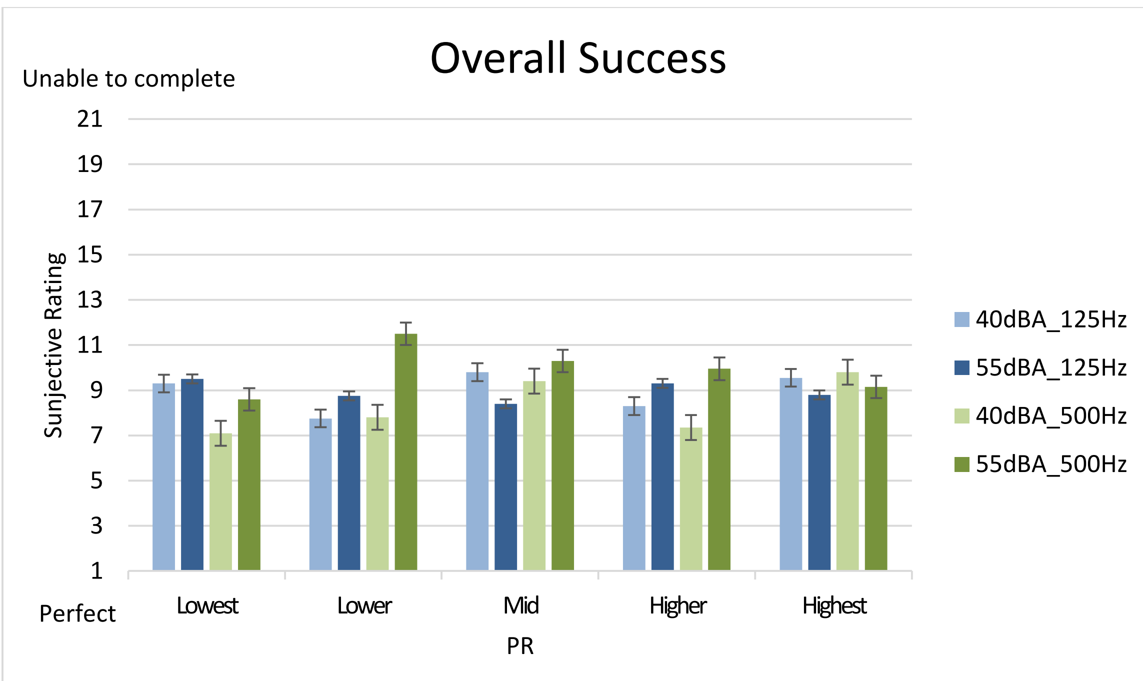

Figure 4.6 Subjective rating results of overall success for all background noise level, tonal frequency and PR conditions. The error bars represent the standard error of the mean.

Although Figure 4.6 seems to indicate some differences in overall success ratings, no statistical differences were found. Table 4.2 gives the full results of the test.

Overall Success				
Source	**SS**	**df**	**MS**	**F**
BNL	32.401	1	32.401	2.763
Freq	0.781	1	0.781	0.028
PR	24.668	2.622	9.406	2.73
BNL*Freq	30.031	1	30.031	2.763
BNL*PR	61.493	2.798	21.98	0.454
Freq*PR	48.963	2.766	17.699	0.689
BNL*Freq*PR	9.588	1.401	6.845	0.103

Table 4.2 ANOVA results for the subjective measure overall success.

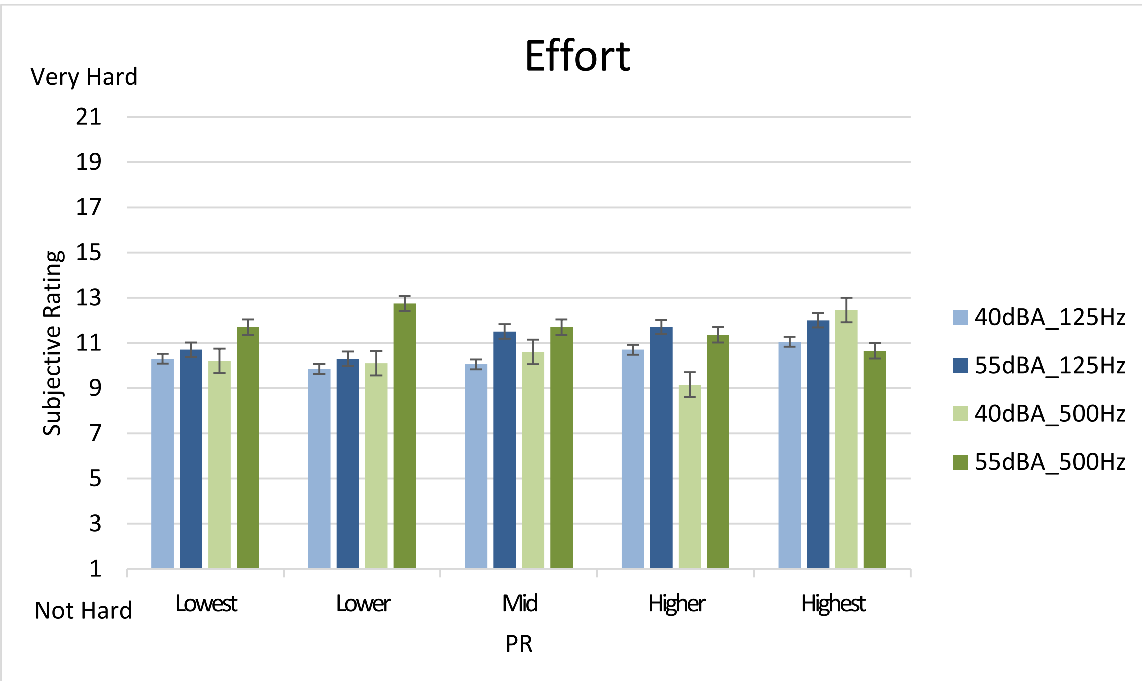

Figure 4.7 Subjective rating results of effort for all background noise level, tonal frequency and PR conditions. The error bars represent the standard error of the mean.

Figure 4.7 indicates some differences in the ratings of effort among the different signal conditions and this was confirmed through ANOVA tests. There were significant differences in the ratings of effort for the two background noise levels, $F(1,10) = 7.168$, $p<0.05$. Subjects rated their effort as slightly higher when being exposed to the signals with the louder background noise level. No other statistically significant differences were found; a complete table with all results can be found in Table 4.3.

Effort				
Source	**SS**	**df**	**MS**	**F**
BNL	47.531	1	47.531	7.168*
Freq	3.251	1	3.251	0.255
PR	20.887	2.04	10.238	0.45
BNL*Freq	1.201	1	1.201	0.166
BNL*PR	28.287	2.646	11.479	0.95
Freq*PR	26.892	2.534	10.612	0.589
BNL*Freq*PR	36.492	1.97	18.525	0.906

Table 4.3 ANOVA results for the subjective measure effort. The asterisk indicates significance at the $p<0.05$ level.

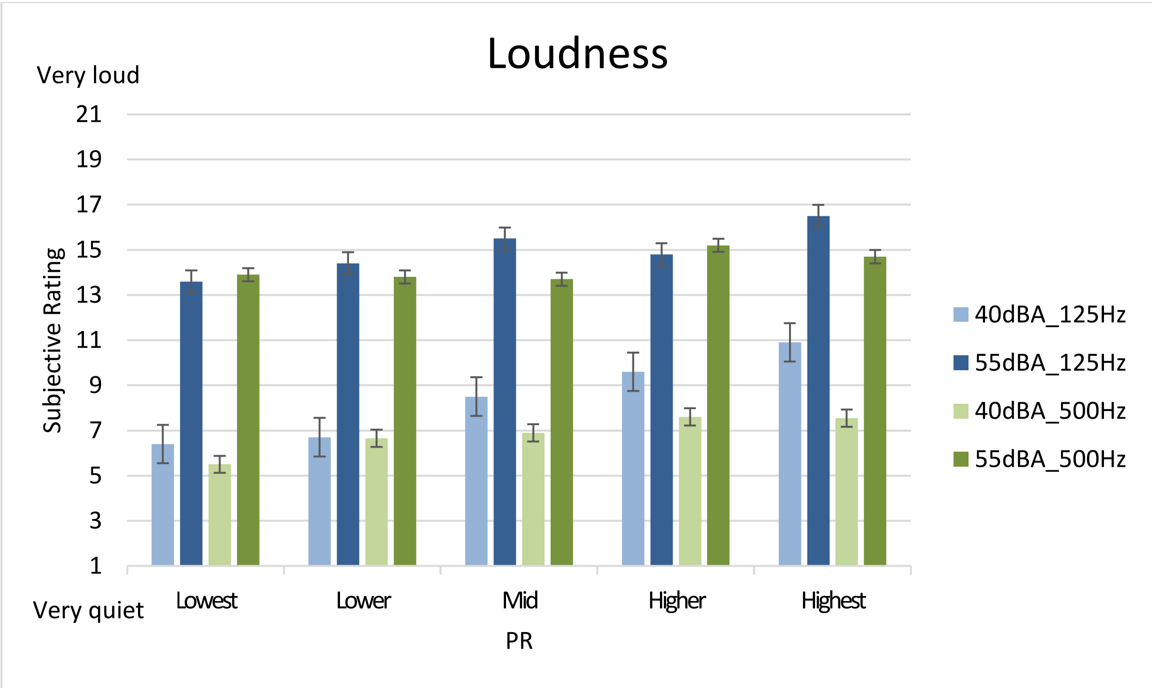

Figure 4.8 Subjective rating results of Loudness of Noise for all background noise level, tonal frequency and PR conditions. The error bars represent the standard error of the mean.

Figure 4.8 clearly indicates major differences in the ratings of loudness for the different background noise levels, tone frequency and PR conditions. Statistically significant differences in loudness ratings were found for background noise level (F(1,10) = 89.773, p<0.001), frequency (F(1,10) = 7.842, p<0.05), and PR (F(4,10) = 6.597, p<0.01). Subjects rated the louder background noise levels as much louder than the quieter background noise levels. The lower frequency (125 Hz) was rated as louder than the higher frequency (500 Hz). Ratings of loudness increased as the PR of the signals increased. There were no significant interactions among these three factors. Table 4.4 gives all results from these analyses.

Loudness				
Source	**SS**	**df**	**MS**	**F**
BNL	2422.08	1	2422.08	89.773***
Freq	62.72	1	62.72	7.842*
PR	171.732	2.393	71.764	6.597**
BNL*Freq	8	1	8	1.596
BNL*PR	17.233	2.95	5.842	0.676
Freq*PR	38.892	2.795	13.914	1.75
BNL*Freq*PR	17.712	2.881	6.148	0.904

Table 4.4 ANOVA results for the subjective measure loudness. The single asterisk indicates significance at the p<0.05 level, the double asterisk indicates significance at the p<0.01 level, and the triple asterisk indicates significance at the p<0.001 level.

Figure 4.9 also clearly indicates differences in the ratings for annoyance to noise among the different signal conditions. Statistically significant differences in annoyance ratings were found for background noise level (F(1,10) = 37.526, p<0.001) and PR (F(4,10) = 4.953, p<0.05). The louder background noise level (55 dBA) was rated as more annoying than the quieter background noise level (40 dBA) and annoyance ratings increased as PR of the tone increased. No significant differences in annoyance were found between the two frequency conditions; however there was a significant interaction between background noise level and frequency (F(1,10) = 8.805, p<0.05). When exposed to the quieter background noise level (40 dBA), subjects rated the lower frequency signals (125 Hz) as more annoying than the signals with the higher frequency tones (500 Hz). However, when subjects were exposed to the signals with the louder background noise levels (50 dBA), ratings of annoyance for the two tonal frequencies were very similar. Subjects rated annoyance to the louder background noise level (50 dBA) higher than for the quieter background noise level regardless of tonal frequency. The other three interaction terms did not produce significant results; all results can be found in Table 4.5.

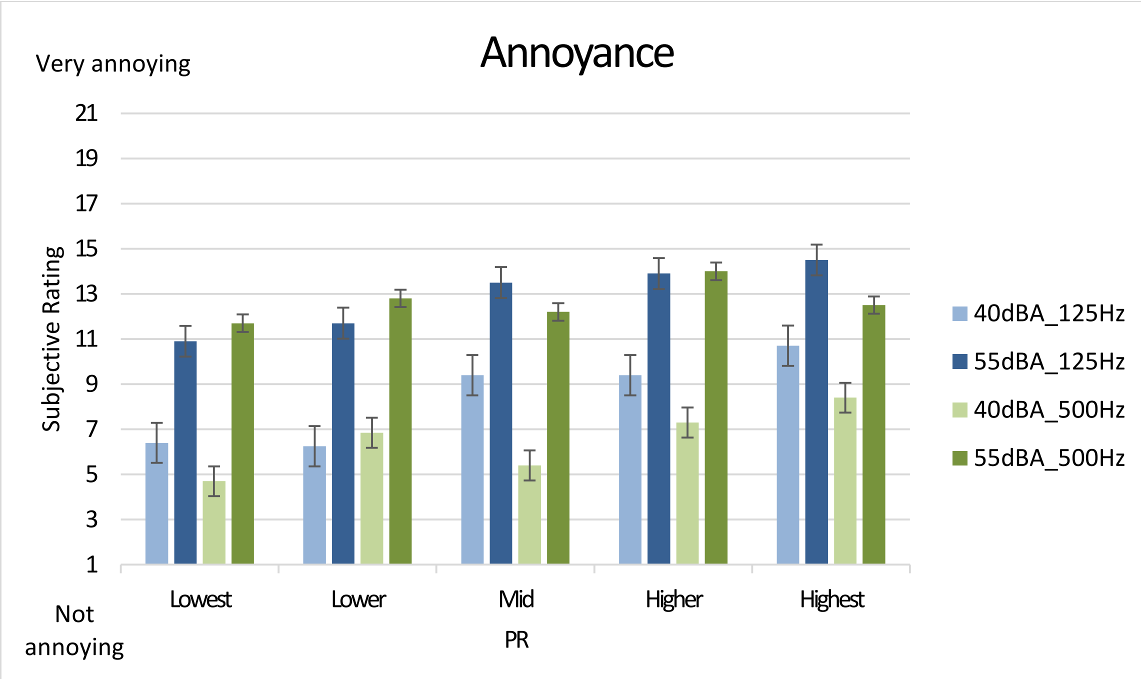

Figure 4.9 Subjective rating results of annoyance to noise for all background noise level, tonal frequency and PR conditions. The error bars represent the standard error of the mean.

Annoyance				
Source	**SS**	**df**	**MS**	**F**
BNL	1378.13	1	1378.13	37.526***
Freq	58.32	1	58.32	3.376
PR	260.47	2.431	107.139	4.953*
BNL*Freq	28.88	1	28.88	8.805*
BNL*PR	23.25	2.898	8.022	0.486
Freq*PR	79.63	2.364	33.682	1.485
BNL*Freq*PR	12.77	2.35	5.435	0.297

Table 4.5 ANOVA results for the subjective measure annoyance. The single asterisk indicates significance at the p<0.05 level and the triple asterisk indicates significance at the p<0.001 level.

Figure 4.10 shows a scatter plot of subjective ratings of annoyance and loudness. This figure shows there is a positive relationship between ratings of annoyance and loudness; as ratings of loudness increase so do ratings of annoyance. A Pearson product-moment correlation was run to determine how strong this relationship is, a statistically significant correlation was found between subjective ratings of annoyance and loudness, ($r = 0.807$, $p<0.001$). However, there is a substantial amount of scatter in this plot, supported by the low r^2 number ($r^2 = 0.65$), which could be due to the tonal conditions. A large amount of previous work (e.g. Hellman) has shown correlation between annoyance and loudness, but this low r^2 number lends support to impact of tonal level.

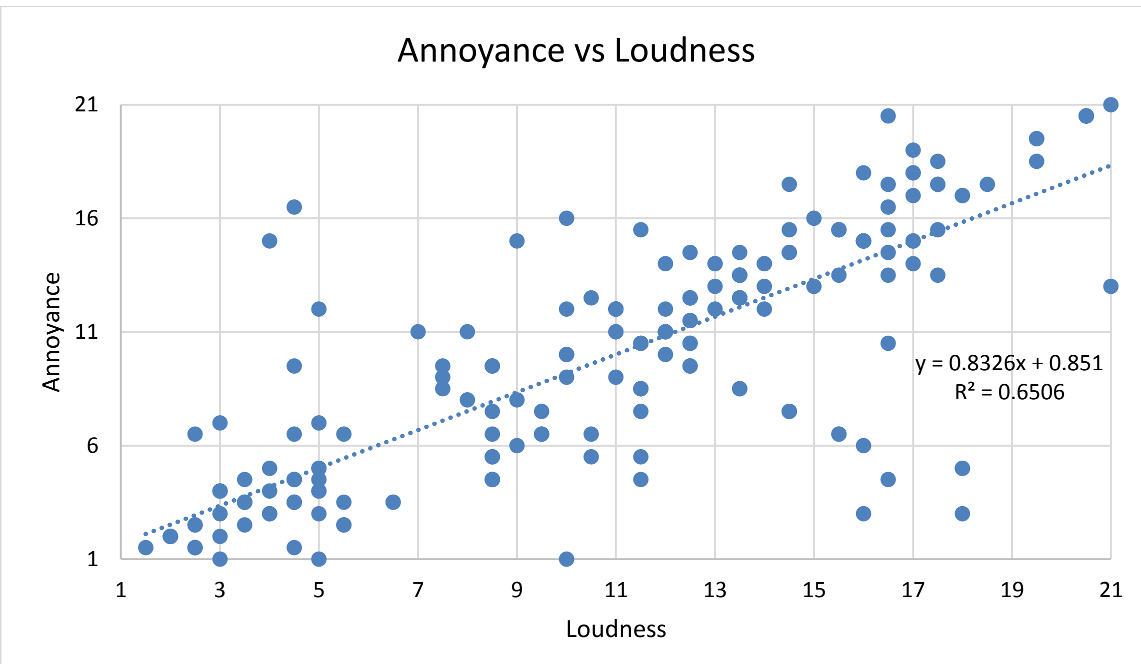

Figure 4.10 Scatter plot of subjective ratings of annoyance vs subjective ratings of loudness.

4.2.3 Thresholds of Annoyance

In order to determine thresholds of annoyance, repeated measures ANOVA tests were used to find any significant differences between the annoyance ratings of just the background noise signals and the annoyance ratings of the signals containing prominent tones. In other words, if a subject rated their annoyance of the background noise only condition and a tonal signal as the same (or close to it) then that tonal signal was determined to be below the threshold of annoyance. If annoyance to the tonal signal was rated as significantly higher than annoyance to the background noise only signal, then the PR of the annoying tonal signal was considered to be a threshold of annoyance. Figures 4.11-4.15 show annoyance ratings results of the tonal signals as well as the background noise only signals; the black line separates annoyance ratings of the background noise only condition from the tonal noise conditions. Figure 4.11 shows the annoyance ratings for all the noise conditions including those without tones; Figures 4.12-4.15 breakdown these ratings for each background noise level and frequency condition.

For the 40 dBA BNL case with a 125 Hz tone, no significant differences were found between the annoyance ratings of the background noise only signal and the two signals with the lowest PR's, 15 and 18 (Figure 4.12). However, there were significant differences between annoyance ratings of the background noise only condition and the signals with the three highest PR's: 21 ($F(1,10) = 11.100$, $p<0.01$), 24 ($F(1,10) = 13.669$, $p<0.01$) and 27 ($F(1,10) = 16.620$, $p<0.01$). These results indicate a potential threshold of annoyance to a 125 Hz tone (above a 40 dBA background noise) at a PR between 18 and 21. Recall that a 125 Hz tone is considered prominent by definition of Prominence Ratio when it has a PR of 18, according to ANSI S1.13-2005.

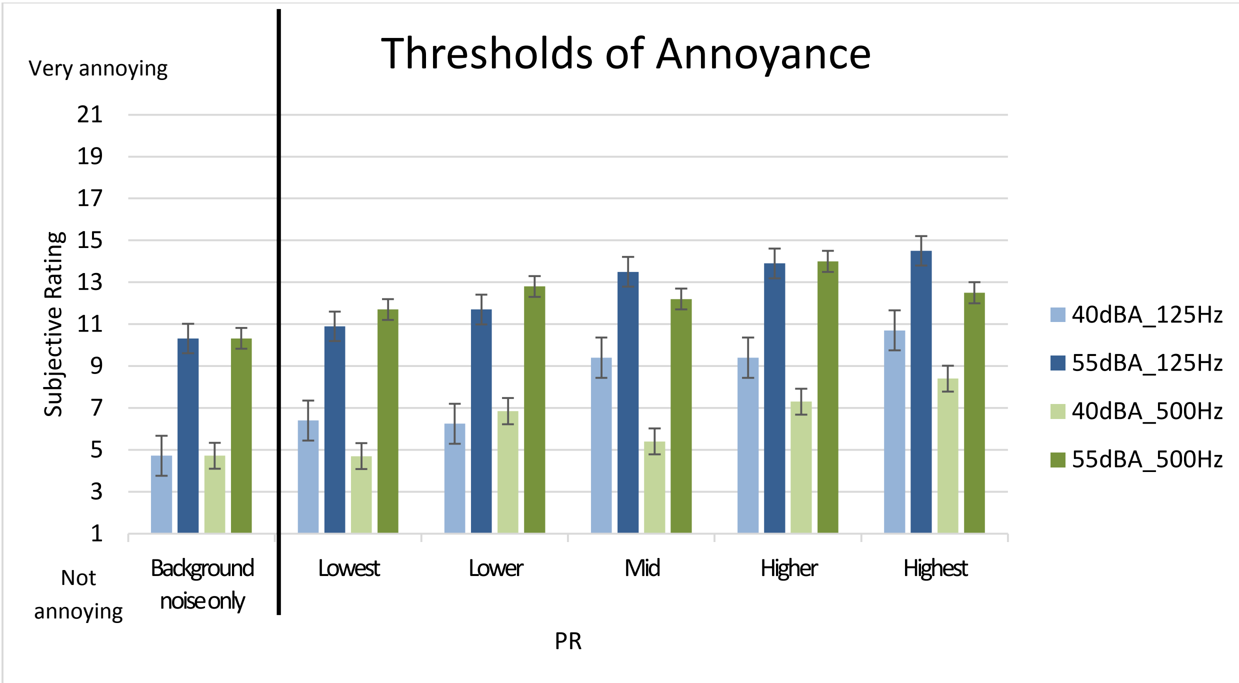

Figure 4.11 Annoyance ratings of the tonal signals compared to annoyance ratings of the background noise only signals. The error bars represent the standard error of the mean.

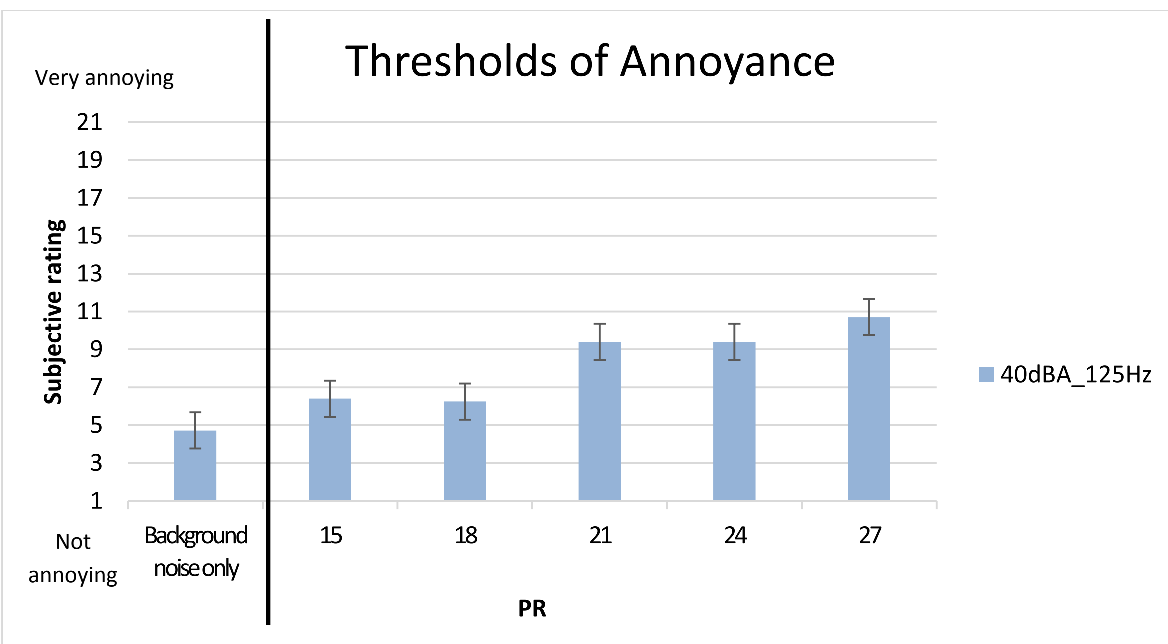

Figure 4.12 Thresholds of annoyance for the 40 dBA BNL and 125 Hz tone signals. The error bars represent the standard error of the mean.

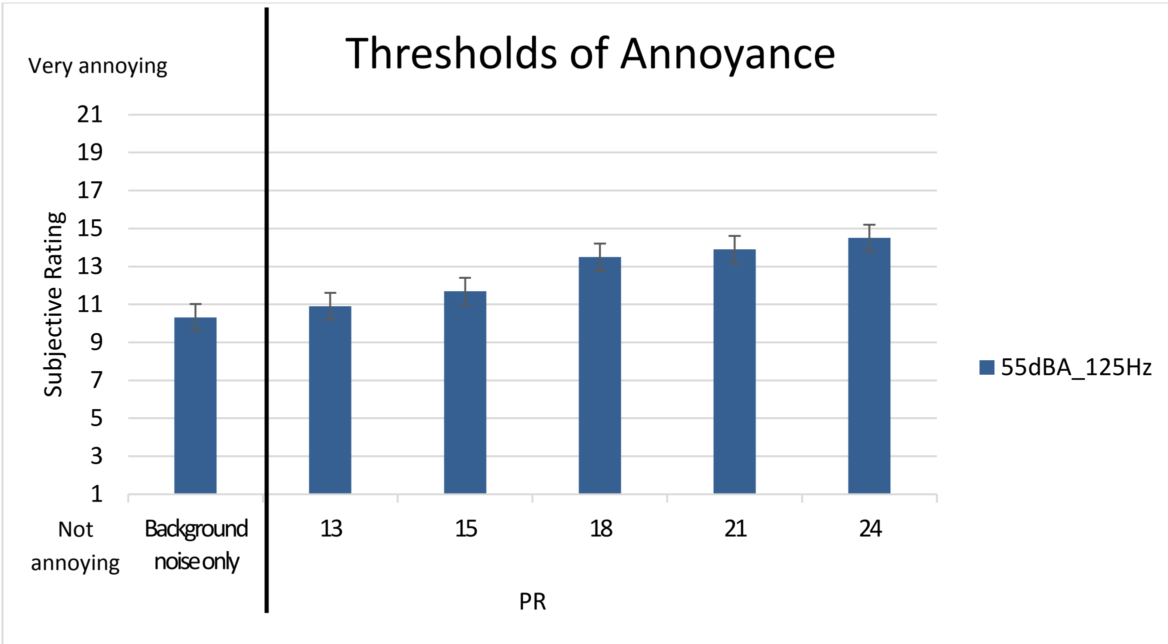

Figure 4.13 Thresholds of annoyance for the 55 dBA BNL and 125 Hz tone signals. The error bars represent the standard error of the mean.

Figure 4.13 shows the annoyance ratings for the same frequency as Figure 4.12 (125 Hz) but above the louder background noise signal (55 dBA). While there were no significant differences between annoyance ratings of the background noise only signal and the tonal signals with the two lowest PR's, there were some significant differences between the background noise only signal and the tonal signals with the higher levels of PR. The cases of a 125 Hz tone with a PR 18 or higher were found to have annoyance ratings that were significantly different than the annoyance rating of the 55 dBA background noise signal, PR=18 $F(1,10) = 14.182$, $p<0.01$, PR=21 $F(1,10) = 6.837$, $p<0.05$ and PR=24 $F(1,10) = 6.654$, $p<0.05$. Consequently, a potential threshold of annoyance for a 125 Hz tone above a 55 dBA background noise level may occur between a PR of 15 and 18. Again recall that prominence of a 125 Hz tone occurs when PR is equal to or greater than 18 (ANSI S1.13-2005).

Note that overall annoyance ratings of the 125 Hz tonal signals were higher for the louder background noise level (55 dBA) than they were for the quieter background noise level (40 dBA). Specifically, for a 125 Hz tone above a 40 dBA background noise level, the potential threshold of annoyance was found to be between PR of 18 and 21. Above the 55 dBA background noise level, though, the potential threshold of annoyance was found to be lower, between a PR of 15 and 18. This would indicate that subjects had a lower tolerance for annoyance when exposed to the 125 Hz tone above the louder background noise level.

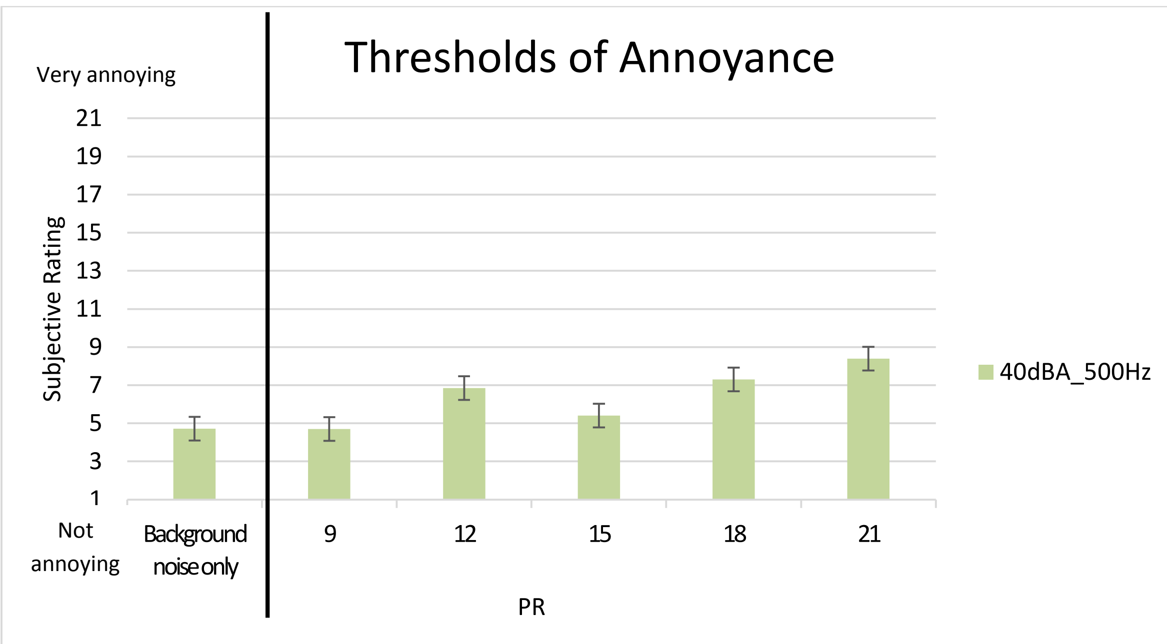

Figure 4.14 Thresholds of annoyance for the 40 dBA BNL and 500 Hz tone signals. The error bars represent the standard error of the mean.

Figure 4.14 shows the annoyance ratings for the 500 Hz tone above the 40 dBA background noise level and the 40 dBA background noise only signals. There were no significant differences between ratings of the background noise only signal and the first three PR levels of the tone. Significant differences were found between the background noise only signal and the 500 Hz tone signals with a PR of 18 ($F(1,10) = 7.555$, $p<0.05$) and 21 ($F(1,10) = 4.511$, $p<0.05$). This indicates that the potential threshold of annoyance for a 500 Hz tone above a 40 dBA background noise signal occurs between a PR of 15 and 18. A tone of 500 Hz is considered prominent when PR is equal to or greater than 12 (ANSI S1.13-2005) so here the potential threshold of annoyance is found to be slightly higher than the threshold of prominence.

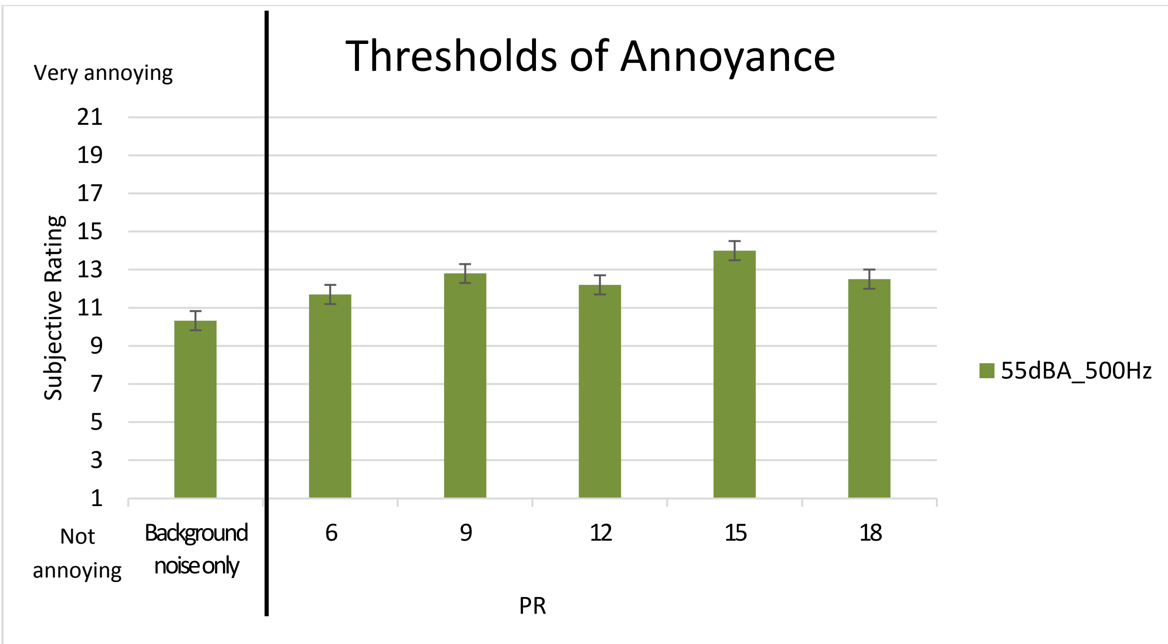

Figure 4.15 Thresholds of annoyance for the 55 dBA BNL and 500 Hz tone signals. The error bars represent the standard error of the mean.

Figure 4.15 shows the annoyance ratings for the final noise condition, the 500 Hz tone above a 55 dBA background noise level. There were no significant differences found between the annoyance ratings of the background noise only condition and the 500 Hz tone with a PR of 6. However there were significant differences between the background noise only condition and the signals with the 500 Hz tone and higher values of PR. Comparing the annoyance ratings of the background noise only with the ratings of the tonal signals with PR of 9 ($F(1,10) = 4.330$, $p<0.05$), 12 ($F(1,10) = 5.650$, $p<0.05$), 15 ($F(1,10) = 29.594$, $9<0.001$), and 18 ($F(1,10) = 4.188$, $p<0.05$), resulted in a potential threshold of annoyance for a 500 Hz tone above a 55 dBA background noise level between a PR of 6 and 9. This is lower than the threshold of prominence for a 500 Hz tone which occurs when PR is equal to or greater than 12 (ANSI S1.13-2005).

As with the annoyance ratings for the 125 Hz tonal signals, the overall annoyance ratings of the 500 Hz tonal signals were much higher for the louder background noise level (55 dBA) than for the quieter background noise level (40 dBA). This indicates that background noise level does have an effect on annoyance to noise that contains prominent tones.

4.3 Method II – Magnitude Adjustment Results

In Method II, subjects were asked to listen to a broadband noise with a tonal component and adjust the level of the tone until it became just annoying. The same background noise levels (40 dBA and 55 dBA) and tone frequencies (125 Hz and 500 Hz) that were explored in Method I were also used in Method II. Subjects were exposed to each of the four BNL and frequency conditions a total of two times. The data for this method come from the PR of the signal that was chosen to be just annoying. Figures 4.16-4.19 show the signals that were selected as being just annoying for each of the four conditions explored. The blue square represents the signal that was selected as just annoying for that condition from the subjects' first exposure, and the red diamond represents the signal that was selected from the subjects' second exposure for the same condition. The PR of the selected signals for each BNL and frequency condition was then averaged across all subjects in order to determine a threshold of annoyance; this average is indicated with a black line on the aforementioned figures. Although 10 subjects participated in this methodology, the data from one subject was not used in this analysis because this person selected the very lowest PR to be just annoying for each condition. Results from the remaining nine subjects were fairly consistent, with most subjects choosing the same PR, or close to it, both times they were exposed to that particular BNL and frequency condition.

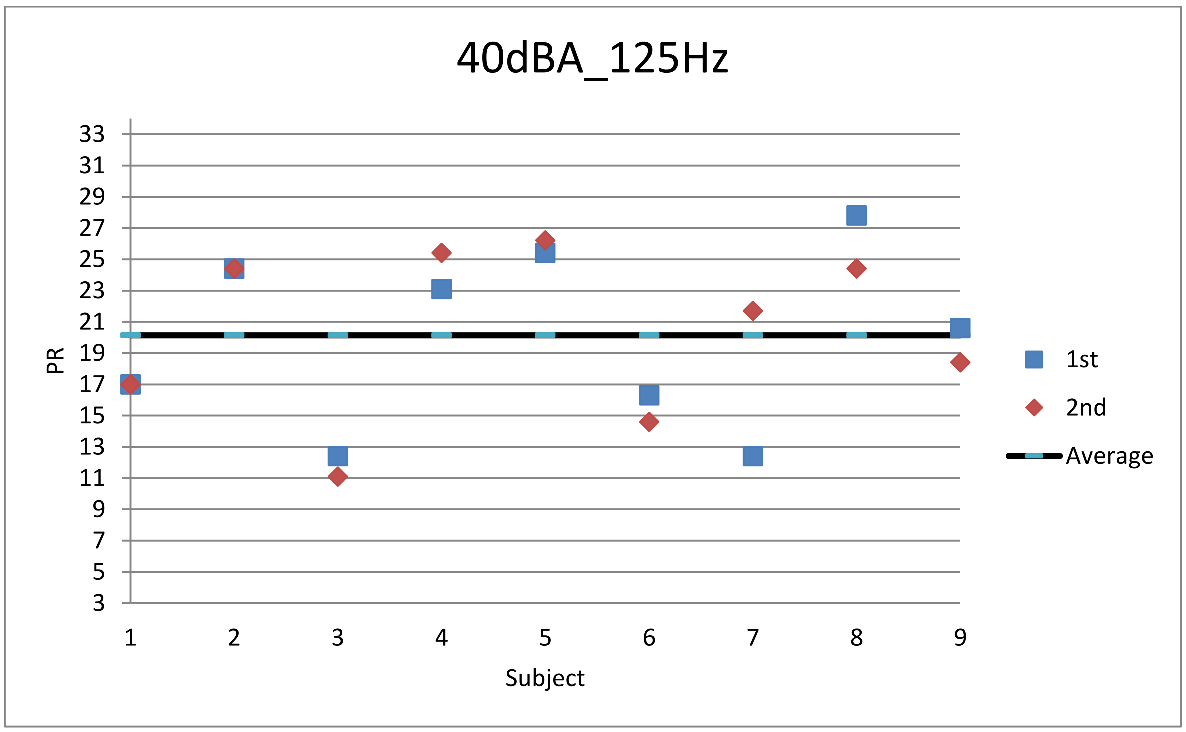

Figure 4.16 Signals that were selected to be just annoying for the 125 Hz tone above the 40 dBA background noise level. The blue square represents the signal that was selected as just annoying for the first exposure to this condition and the red diamond represents the signal that was selected during the second exposure to the same condition. The black line is the average signal selected from both exposures across all subjects.

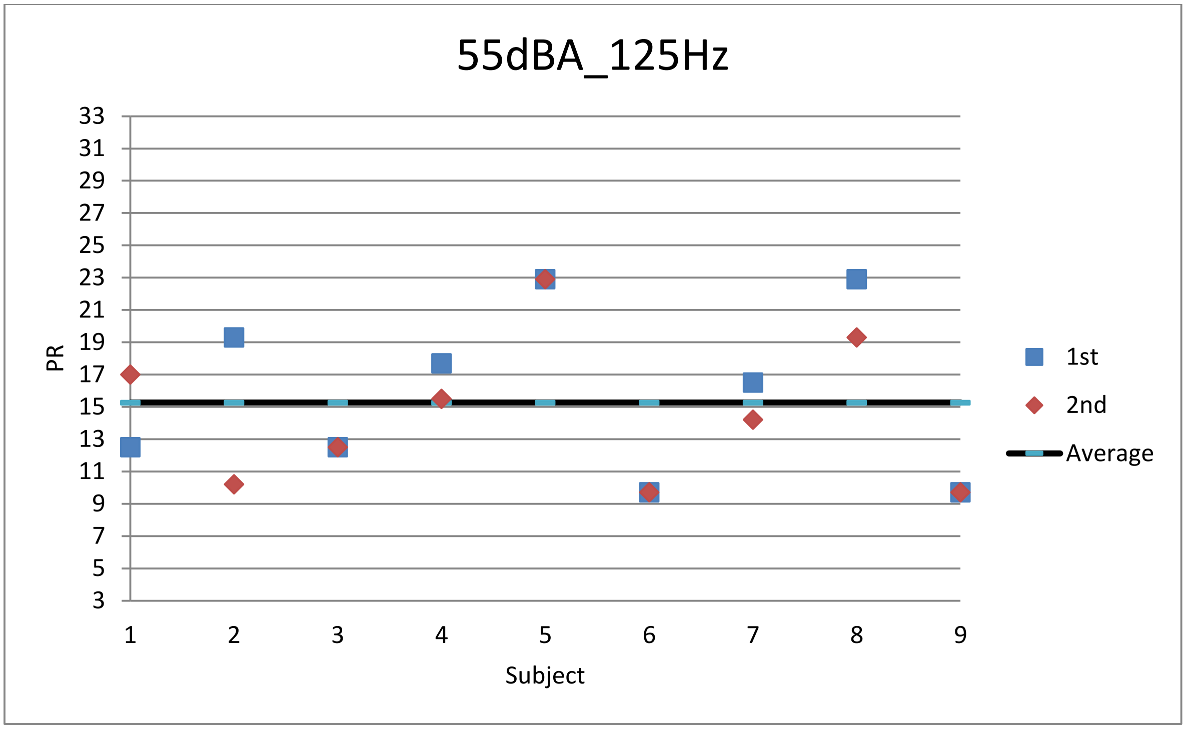

Figure 4.17 Signals that were selected to be just annoying for the 125 Hz tone above the 55 dBA background noise level. The blue square represents the signal that was selected as just annoying for the first exposure to this condition and the red diamond represents the signal that was selected during the second exposure to the same condition. The black line is the average signal selected from both exposures across all subjects.

For the 125 Hz tone above the 40 dBA background noise level, the average PR chosen was 20. This is slightly higher than the threshold of prominence for a 125 Hz tone which is 18 (ANSI S1.13-2005). Looking at the same tone frequency above the 55 dBA background noise level, the average PR chosen to be just annoying was 15, just slightly lower than the threshold of prominence. This again suggests that background noise level does have an effect on perceived annoyance as the PR threshold of annoyance shifts lower with higher background noise levels.

The 500 Hz tone above the 40 dBA background noise level produced an average PR of 16 to be just annoying. This is slightly higher than the threshold of prominence for a 500 Hz tone which is 12 (ANSI S1.13-2005). For this tone above the 55 dBA background noise level, the average PR chosen to be just annoying was 11 which is slightly lower than the threshold of prominence. These results are consistent with what was found for the 125 Hz tone, in that the louder background noise levels produce a threshold of annoyance that is slightly lower than the threshold of prominence. This indicates that human annoyance to tonal noise may occur even when the tone is not considered prominent (by definition of Prominence Ratio).

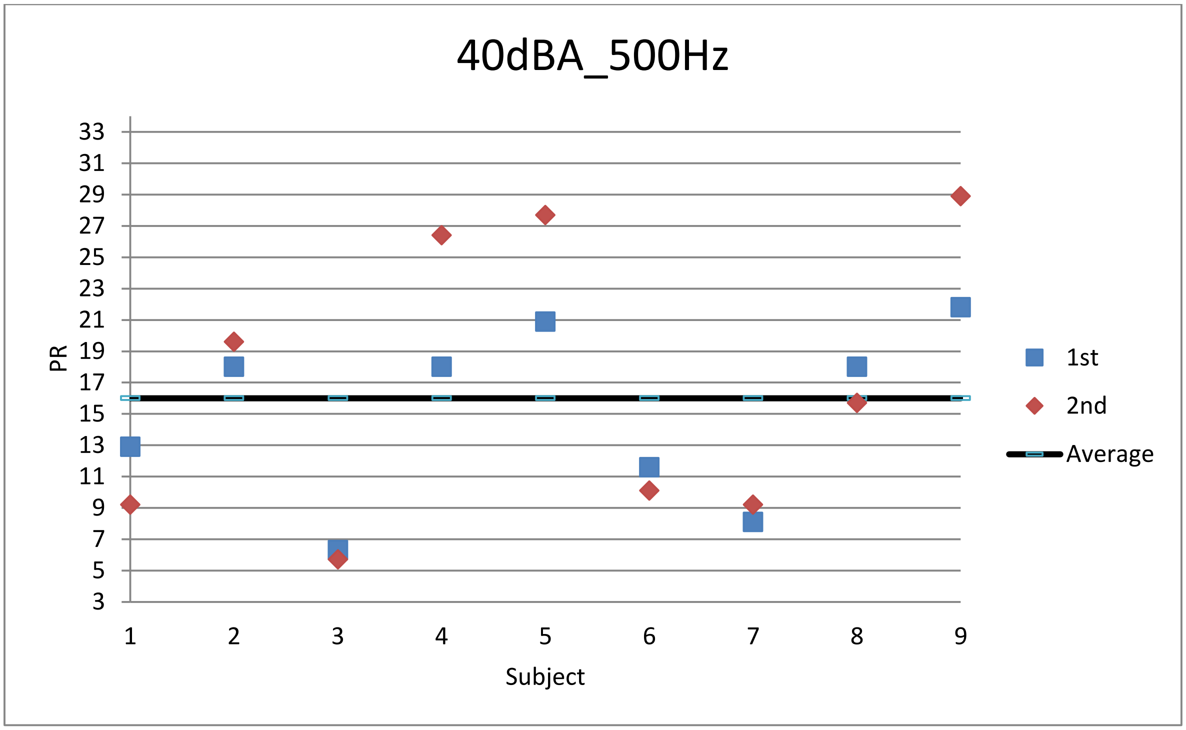

Figure 4.18 Signals that were selected to be just annoying for the 500 Hz tone above the 40 dBA background noise level. The blue square represents the signal that was selected as just annoying for the first exposure to this condition and the red diamond represents the signal that was selected during the second exposure to the same condition. The black line is the average signal selected from both exposures across all subjects.

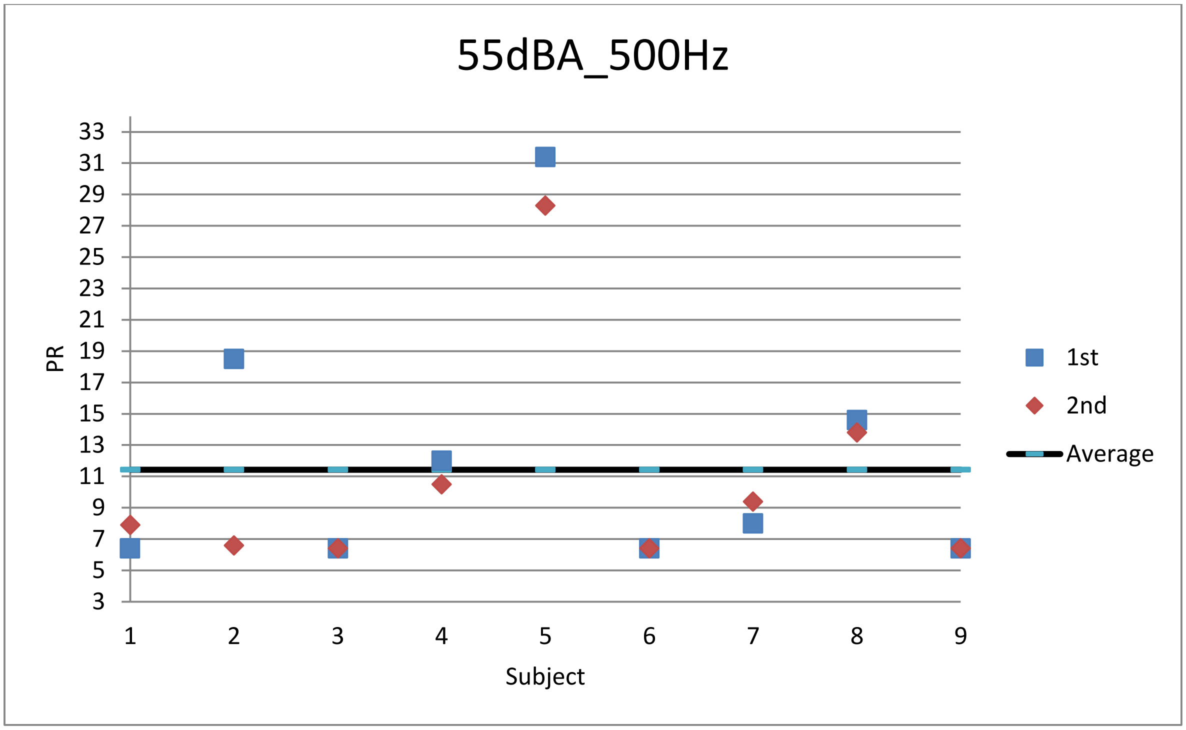

Figure 4.19 Signals that were selected to be just annoying for the 500 Hz tone above the 55 dBA background noise level. The blue square represents the signal that was selected as just annoying for the first exposure to this condition and the red diamond represents the signal that was selected during the second exposure to the same condition. The black line is the average signal selected from both exposures across all subjects.

4.4 Comparison of Results from Method I & Method II

The PR values that were taken to be potential thresholds of annoyance in Method I correspond closely to the PR values that were selected to be just annoying in Method II. A summary of the results from both of these methods can be seen in Table 4.6. In both methods, the PR values that were determined to be thresholds of annoyance were found to be right around the listed values of thresholds of prominence from ANSI S1.13-2005. In three out of four cases, when subjects were exposed to a tone above the louder background noise level, the threshold of annoyance was found to be slightly lower than the threshold of prominence. This is indicative that the level of background noise does impact human perception of annoyance. This is further supported by the fact that when exposed to the same tonal frequency above a louder background noise level, subjects had a lower threshold of annoyance than for the quieter background noise level.

Lower thresholds of annoyance were also determined for the 500 Hz tonal frequency compared to those for the 125 Hz tones. This indicates that frequency of the tone also contributes to perceived annoyance, with lower frequency tones allowed to be at higher PR's.

	Method I - Direct Assessment with Task	**Method II - Magnitude Adjustment**
40dBA_125Hz	18-21	20
55dBA_125Hz	15-18	15*
40dBA_500Hz	15-18	16
55dBA_500Hz	6-9*	11*

Table 4.6 Thresholds of annoyance determined from both methods in the study, given as prominence ratio (PR) ranges for Method I and values for Method II. The asterisk indicates that the PR value shown is below the threshold of prominence for that tonal frequency.

Chapter 5: Conclusions

This study aimed to determine the annoyance thresholds of tones in noise, using two different methods (a direct assessment with task, and a magnitude adjustment methodology) to test combinations of two tonal frequencies (125 Hz and 500 Hz) against two different background noise levels (40 dBA and 55 dBA). The PR values that were determined to be thresholds of annoyance in Method I closely correspond to the PR values that were selected to be just annoying in Method II with slightly greater differences found for the louder background noise cases. Those annoyance thresholds of tones in noise were found to be right around the listed thresholds of prominence. The thresholds of annoyance for the tones in the louder background noise levels were slightly lower than those for the quieter background noise level conditions. These results suggest that background noise level has an effect on perceptions of annoyance and therefore needs to be taken into consideration when trying to predict annoyance.

As Method II is much quicker to implement, it may be a better method to use in future tests of more tonal frequencies, background noise levels, and tonal levels. A field-test in an actual office environment would also be very beneficial to validate the findings from lab-based studies. Another idea for future work is to compare other existing tonal metrics besides PR to see how well they correlate with the results obtained here on human perception of annoyance.

References

ANSI. 2005. ANSI S1.13: Measurement of Sound Pressure Levels in Air. Melville, NY: Acoustical Society of America.

ASHRAE. 2011. “Noise and Vibration Control,” Ch. 48 in *2011 ASHRAE Handbook—HVAC Applications*. Atlanta: American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc.

Aures, W. 1985. The sensory euphony as a function of auditory sensations. *Acustica*, 49:130-141.

Ayr, U., E. Cirillo, and F. Martellotta. 2001. An experimental study on noise indices in air conditioned offices. *Applied Acoustics*, 62:633-643.

Ayr, U. E. Cirillo, I. Fato, and F. Martellotta. 2003. A new approach to assessing the performance of noise indices in buildings. *Applied Acoustics*, 64:129-145.

Balant, A.C., R.D. Hellweg, Jr., M.A. Nobile, and L. Wittig. 1999. A comparison of two methods for the evaluation of prominent discrete tones. Proceedings of Internoise, 106(1):1373-1378.

Balant, A.C., K. Barringer, and M. Nobile. 2000. Psychoacoustic comparison of two methods for the evaluation of prominent discrete tones. *The Journal of the Acoustical Society of America*, 107(5):2877.

Beranek, L. 1957. Revised criteria for noise in buildings. *Noise Control Engineering Journal*, 3:19-27.

Beranek, L. 1989. Balanced Noise-Criterion (NCB) curves. *The Journal of the Acoustical Society of America*, 86(2):650-664.

Beranek, L. 2007. The noisy dawn of the jet age. *Sound and Vibration*, 4:94-100.

Blazier, W. 1981. Revised noise criteria for application in the acoustical design and rating of HVAC systems. *Noise Control Engineering Journal,* 16(2):64-73.

Blazier, W. 1997. RC Mark II: A refined procedure for rating the noise of heating, ventilating, and air-conditioning (HVAC) systems in buildings. *Noise Control Engineering Journal,* 45(6):243-250.

Bowden, E.E., and L.M. Wang. 2005. Relating human productivity and annoyance to indoor noise criteria systems: a low frequency analysis. *ASHRAE Transactions,* 111(1):684-692.

Field, A. and G. Hole. 2003. *How to Design and Report Experiments.* London: Sage Productions

Frank, T., T.A. Brungart, and N.C. Nicholas. 2007. Annoyance from and tolerance to sounds. *Noise Control Engineering Journal,* 55(3):357-366.

Hastings, A., K. Lee, P. Davies, and A. Surprenant. 2003. Measurement of the attributes of complex tonal components commonly found in product sound. *Noise Control Engineering Journal,* 51(4):195–209.

Hay, B. and M.F. Kemp. 1972. Measurements of noise in air conditioned, landscaped offices. *Journal of Sound and Vibration,* 23(3):363-373.

Hellman, R.P. 1982. Loudness, annoyance, and noisiness produced by single-tone-noise complexes. *The Journal of the Acoustical Society of America,* 72(1):62–73.

Hellman, R.P. 1984. Growth rate of loudness, annoyance, and noisiness as a function of tone location within the noise spectrum. *The Journal of the Acoustical Society of America,* 75(1):209-218.

Hellman, R.P. 1985. Perceived magnitude of two-tone complexes: loudness, annoyance, and noisiness. *The Journal of the Acoustical Society of America,* 77(4):1497–1504.

Hellweg, R., M. Nobile, and A.C. Balant. 2000. A comparison of two methods for the evaluation of prominent discrete tones – phase 2. Proceedings of Internoise, 5:4078-4087.

Hellweg Jr., R.D., and M.A. Nobile. 2002. Modification to procedures for determining prominent discrete tones. Proceedings of Internoise, 111(1):1707-1714.

Holmberg, K., U. Landström, and A. Kjellberg. 1993. Effects of ventilation noise due to frequency characteristic and sound level. *Journal of Low Frequency Noise and Vibration,* 12(4):115–122.

ISO. 2007. *ISO 1996-2: Acoustics – Description, Measurement and Assessment of Environmental Noise Levels.* Geneva, Switzerland: ISO

Keighly, E.C. 1966. The determination of acceptability criteria for office noise. *Journal of Sound and Vibration,* 4(1):73-87.

Keighly, E.C. 1970. Acceptability criteria for noise in large offices. *Journal of Sound and Vibration,* 11(1):83-93.

Kryter, K., and K. Pearsons. 1965. Judged noisiness of a band of random noise containing an audible pure tone. *The Journal of the Acoustical Society of America,* 38:106–112.

Landström, U., A. Kjellberg, L. Söderberg, and B. Nordström. 1991. The effects of broadband, tonal, and masked ventilation noise on performance, wakefulness and annoyance. *Journal of Low Frequency Noise and Vibration,* 10:112–122.

Landström, U., A. Kjellberg, and M. Byström. 1993. Acceptable levels of sounds with different spectral characteristics during the performance of a simple and complex nonauditory task. *Journal of Sound and Vibration,* 160(3):533–542.

Landström, U., A. Kjellberg, L. Söderberg, and B. Nordström. 1994. Measures against ventilation noise – which tone frequencies are least and most annoying? *Journal of Low Frequency Noise and Vibration*, 13:81–88.

Landström, U., E. Akerlund, A. Kjellberg, M. Tesarz. 1995. Exposure levels, tonal components and noise annoyance in working environments. *Environment International*, 21(3):265-275.

Lee, K.H., P. Davies, and A.M. Suprenant. 2004. Quantification of the tonal prominence of complex tones in machinery noise. Proceedings of Noise-Con. 114(1):927-933.

Lee, K.H., P. Davies, and A.M. Suprenant. 2005. Tonal strength of harmonic complex tones in machinery noise. Proceedings of Noise-Con. 113(1):169-174.

Little, J.W. 1961. Human responses to jet engine noises. *Noise Control*, 7(3):11-13.

Little, J.W., and J.E. Mabry. 1969. Empirical comparisons of calculation procedures for estimating annoyance of jet aircraft flyovers. *Journal of Sound and Vibration*, 10(1):71-80.

Persson, K., and M. Björkman. 1988. Annoyance due to low frequency noise and the use of the dB(A) scale. *Journal of Sound and Vibration*, 127:491-197

Persson Waye, K., R. Rylander, S. Benton, and H.G. Leventhall. 1997. Effects on performance and work quality due to low frequency ventilation noise. *Journal of Sound and Vibration*, 205(4):467-474

Persson Waye, K., J. Bengtsson, A. Kjellberg, and S. Benton. 2001. Low frequency noise "pollution" interferes with performance. *Noise and Health,* 4(13):33-49.

Persson Waye, K. and R. Rylander. 2001. The prevalence of annoyance and effects after long-term exposure to low-frequency noise. *Journal of Sound and Vibration,* 240(3):483-487

Ryherd, E.E., and L.M. Wang. 2008. Implications of human performance and perception under tonal noise conditions on indoor noise criteria. *The Journal of the Acoustical Society of America,* 124(1):218-226.

Ryherd, E.E. and L.M. Wang. 2010. AB-10-018: The effects of noise from building mechanical systems with tonal components on human performance and perception (1322-RP). *ASHRAE Transactions,* 116, Part 2, 541-552.

Schutte, M., A. Marks, E. Wenning, and B. Griefahn 2007. The development of the noise sensitivity questionnaire. *Noise & Health, 9*(34):15-24.

Tang, S., J. Burnett and C. Poon. 1996. Aural environment survey in air-conditioned open-plan offices. *Building Services Engineering Research and Technology,* 17(2):97-100.

Tang, S. 1997. Performance of noise indices in air-conditioned landscaped office buildings. *The Journal of the Acoustical Society of America,* 102(3):1657-1663.

Tang, S., and C. Wong. 1998. Performance of noise indices in office environment dominated by noise from human speech. *Applied Acoustics*, 55(4):293-305.

Tang, S. and M. Wong. 2004. On noise indices for domestic air conditioners. *Journal of Sound and Vibration,* 274(1):1-12.