8-2006

Selection of Attractor Sounds for an Audio-Based Navigation System for the Visually Impaired

Brad Salisbury
Western Michigan University

Follow this and additional works at: http://scholarworks.wmich.edu/masters_theses
Part of the Aerospace Engineering Commons, and the Mechanical Engineering Commons

Recommended Citation
http://scholarworks.wmich.edu/masters_theses/1407

This Masters Thesis-Open Access is brought to you for free and open access by the Graduate College at ScholarWorks at WMU. It has been accepted for inclusion in Master’s Theses by an authorized administrator of ScholarWorks at WMU. For more information, please contact maira.bundza@wmich.edu.
SELECTION OF ATTRACTOR SOUNDS FOR AN AUDIO-BASED NAVIGATION SYSTEM FOR THE VISUALLY IMPAIRED

by

Brad Salisbury

A Thesis
Submitted to the
Faculty of The Graduate College
in partial fulfillment of the
requirements for the
Degree of Master of Science in Engineering (Mechanical)
Department of Mechanical and Aeronautical Engineering

Western Michigan University
Kalamazoo, Michigan
August 2006
SELECTION OF ATTRACTOR SOUNDS FOR AN AUDIO-BASED NAVIGATION SYSTEM FOR THE VISUALLY IMPAIRED

Brad Salisbury, M.S.E.
Western Michigan University, 2006

Research into human hearing has been very ‘laboratory’ oriented in the sense that the test environments do not replicate most ‘real world’ situations [J.C. Middlebrooks and D.M. Green, Ann. Rev. of Psychology, 42, 135-159 (1991)]. While very useful information is gained from these types of tests, it is difficult to see how ‘real world’ situations affect sound source localization, recognition, and navigation (walking/way finding) performance. Such information is especially important to people who are visually impaired and dependent on prior knowledge of the environment or audio cues for travel. The research reported here was conducted during the development of an audio-based navigation system. The question that arose was, “what constitutes ‘good’ attractor sounds versus ‘bad’ sounds?” A series of physical tests were developed to identify sounds that performed best from within a group of ‘real-world’ attractor sounds. Testing was conducted with the aid of participants who were blind or visually impaired. The attractor sounds were compared in the time- and frequency-domains to identify common characteristics. Results of the experiments were consistent with those of Landau, et al. [S. Landau, et. al., Asst. Technology, 17, 133-143 (2005)].
ACKNOWLEDGEMENTS

I would like to start by thanking the people from the Noise and Vibration Laboratory at Western Michigan University who have assisted throughout the course of this work. I would like to thank John Stahl for his hours of LabView programming and Jay Pliskow for his time spent organizing the large amount of data collected.

Thanks are also extended to my thesis committee, Dr. Koorosh Naghshineh, Dr. Philip Guichelaar, and Dr. William Wiener for their time spent reviewing my work and providing advice. Dr. Naghshineh deserves special thanks for providing me the opportunity to work in the Noise and Vibration Lab over the past few years. Without that experience, this work would not have been possible.

I would also like to thank the staff of Resource Sharing (Interlibrary Loan) at Waldo Library for obtaining the numerous journal articles needed during the literature review phase of this thesis. I especially like to thank Ms. Eliana Lopez-Grez, who was always prompt to respond to my requests for, sometimes difficult to obtain, articles.

Finally, I would like to thank my parents for their encouragement and support throughout this process and for proofreading my work.

Brad Salisbury
# TABLE OF CONTENTS

ACKNOWLEDGEMENTS ........................................................................................................ ii

LIST OF TABLES ............................................................................................................... vi

LIST OF FIGURES ........................................................................................................... vii

INTRODUCTION .............................................................................................................. 1

CHAPTER

1. REVIEW OF RELEVANT LITERATURE ............................................................. 3
   1.1 LOCALIZATION .......................................................................................... 3
       1.1.1 AN INTRODUCTION TO SOUND SOURCE LOCALIZATION BY HUMANS .......... 4
       1.1.2 SIGNIFICANT LOCALIZATION RESEARCH PRIOR TO 1991 ......................................................... 8
       1.1.3 SIGNIFICANT LOCALIZATION RESEARCH SINCE 1991 ................................................................. 9
       1.1.4 LOCALIZATION RESEARCH INVOLVING ‘REAL WORLD’ ENVIRONMENTS ......................... 15
   1.2 RECOGNITION ............................................................................................. 24
   1.3 NAVIGATION ............................................................................................... 28

2. EXPERIMENTAL DESIGN ................................................................................... 31
   2.1 PREPARATION .......................................................................................... 31
   2.2 PARTICIPANT SELECTION ....................................................................... 34
   2.3 PROCEDURES ........................................................................................... 35
       2.3.1 INITIAL SOUND SELECTION ................................................................. 38
Table of Contents – Continued

CHAPTER

2.3.2 INDOOR NAVIGABILITY .............................................. 42
2.3.3 INDOOR LOCALIZATION AND SOUND LEVELS .... 51
2.3.4 OUTDOOR SOUND LEVELS ................................. 54
2.3.5 SOUND IDENTIFICATION AND RECOGNITION ...... 55
2.4 METHODS OF DATA COLLECTION ................................. 58

3. DATA ANALYSIS ......................................................................................... 59

3.1 INITIAL SOUND SELECTION ...................................................... 59
3.2 INDOOR NAVIGABILITY ............................................................ 59
3.3 SUBJECTIVE RESPONSE .............................................................. 62
3.4 INDOOR SOUND LOCALIZATION AND SOUND LEVEL ...... 63
3.5 SOUND IDENTIFICATION AND RECOGNITION ...................... 64
3.6 COMPARISON OF DATA .............................................................. 66
3.7 TIME- AND FREQUENCY-DOMAIN DATA ............................... 67

3.7.1 TIME-DOMAIN CHARACTERISTICS ......................... 68
3.7.2 FREQUENCY-DOMAIN CHARACTERISTICS ............ 70

4. CONCLUSIONS REGARDING THE CURRENT RESEARCH .......... 72

5. RECOMMENDATIONS FOR FUTURE RESEARCH ....................... 76

REFERENCES ................................................................................................... 78

APPENDICES

A. HSIRB RESEARCH PROTOCOL CLEARANCE ......................... 81
Table of Contents – Continued

APPENDICES

B. SOUND NORMALIZATION PROGRAM (MATLAB) .................. 83
C. PRELIMINARY TEST DATA .......................................................... 85
D. TRAVEL ENVELOPES OF PARTICIPANTS IN THE NAVIGATION TEST ........................................................................ 89
E. SOUND IDENTIFICATION AND RECOGNITION ANALYSIS TABLES ............................................................................................. 102
F. PROGRAM FOR DATA ANALYSIS (MATLAB) .......................... 109
G. DATA PLOTS ................................................................................... 116
   G1. COMPARATIVE OVERLAYED PLOTS: MAGNITUDE v. TIME .......................................................................................... 117
   G2. COMPARATIVE OVERLAYED PLOTS: SPECTROGRAMS .......................................................................................... 128
   G3. COMPARATIVE OVERLAYED PLOTS: OCTAVE BANDS .................................................................................. 139
LIST OF TABLES

1.1 A-B sound pairs ........................................................................................................ 25
2.1 Possible sounds .......................................................................................................... 32
2.2 Table of sounds and their descriptions as used in testing ..................................... 41
3.1 Spreadsheet analysis of the navigability test ......................................................... 61
3.2 Data analysis for the localization and sound level test ......................................... 64
3.3 Example of the analysis of the sound identification and recognition test ......... 66
3.4 Comparison of sound rankings .............................................................................. 66
3.5 Comparison of the top six sounds by test .............................................................. 67
4.1 Recommended sound numbers and levels for addition to the PING catalog ...... 75
LIST OF FIGURES

1.1 Simplified spherical head model ................................................................. 5
1.2 ITD confusion .............................................................................................. 6
1.3 The head-related transfer function ............................................................. 7
1.4 Experimental setup of Hofman and Van Opstal ........................................ 11
1.5 Stimuli examples ....................................................................................... 12
1.6 Effects of DTF modifications ..................................................................... 14
1.7 ESPRO Experimental Setup ..................................................................... 17
1.8 Loudspeaker array schematic ................................................................... 24
1.9 Walking trajectories .................................................................................. 30
2.1 Reverberation chamber test setup ............................................................. 39
2.2 Overhead diagram of the reverberation chamber test setup .................... 39
2.3 Schematic diagram of the equipment setup used during the initial sound selection .................................................................................................................. 40
2.4 Diagram of the area in Kohrman Hall used for the indoor navigability test . 43
2.5 Experimental space diagram (a) ............................................................... 45
2.6 Experimental space diagram (b) ............................................................... 46
2.7 Schematic diagram of the equipment setup used in the indoor navigation study .................................................................................................................. 47
2.8 Participant walking path during the indoor navigability test ..................... 48
2.9 CAD drawing of an example path (sound 7) ............................................. 50
2.10 Experimental setup for the indoor localization and sound level test .......... 51
List of Figures - Continued

2.11 Overhead diagram of the experimental space for the indoor localization and sound level test ................................................................. 52

2.12 Participant using a laser pointer to indicate direction of sound origination 54

2.13 Participant seated in reverberation chamber during the sound identification and recognition test ......................................................... 56

2.14 Representation of the equipment setup for the identification and recognition test ................................................................. 57

3.1 Linear FFT plot of 1000 Hz calibration tone ............................................. 68

3.2 Example of a repeated time-domain signal (sound 5) ............................. 69

3.3 Example of a ‘winged’ time domain signal (sound 1) ............................. 69

3.4 Example plot showing high SPLs in the 500 and 1000 Hz bands (sound 5) 71
INTRODUCTION

The Americans With Disabilities Act has brought about many changes that have improved the accessibility of public spaces for people with disabilities. Improvements such as wheelchair ramps and wheelchair height appropriate amenities have increased access by those with mobility limitations and written text options or amplified headsets exist for patrons who are deaf or hearing impaired. However, accommodations for people who are blind or visually impaired have lagged behind because of the technical difficulties of adapting visually oriented environments to those who cannot distinguish visual cues [Landau, et al. 2005]. Part of this problem is the three types of information used by humans in navigation - position, velocity, and acceleration – only acceleration-based information does not require any external visual or audio cues [Loomis, et al. 1993]. As a result, people with visual impairments are dependent on prior knowledge of the environment (from previous experience, visual descriptions, or raised-line maps) or audio cues for navigation.

In an effort to increase the accessibility of public spaces for people who are blind, an interactive audio-based navigation system, known as PING, is being developed by Touch Graphics, Inc. of New York, New York. The system is interfaced with the user via a portable telephone that is utilized to select a destination from a menu and trigger a pathway of audio beacons, in sequence, from the user’s location to the desired destination. Users of the system must select an attractor sound, then listen for that sound
played through a beacon, and then travel towards the originating beacon. The sound can be triggered by the user as often as is necessary. When the beacon is reached, the user activates the next beacon in the path and continues, repeating the process until they reach the final destination. For a more thorough description of the system, see Landau, et al. [2005].

Since the system must be capable of supporting multiple users there must be multiple attractor sounds available for selection. The sounds are stored in a type of library, where once a sound is selected by one user, ‘checked out’, it is unavailable to other users until it is ‘checked in’ at the end of the visit. The question that arose during development of the system was, “what are ‘good’ sounds and what are ‘bad’ sounds?”

The research reported here was conducted with three goals in mind. First, to develop a series of tests that can be used to identify ‘good’ sounds for inclusion in the library. Second, to recommend a collection of sounds for the library. Third, to identify any characteristics common to the ‘good’ sounds that could aid in selecting future sounds for the library. A ‘good’ sound was defined to be one that was easily recognizable to users, well liked by users, provided enhanced localization, and was non-obtrusive to other people in the environment.
1. REVIEW OF RELEVANT LITERATURE

Human hearing is a complex process that is only beginning to be understood, although much has been published on the topic. Most of the focus has been on sound localization with less emphasis is placed on sound recognition and navigation using sound. The research has been very ‘laboratory’ oriented in the sense that the test environments do not replicate most ‘real world’ situations, a fact recognized by Middlebrooks and Green [1991]. While very useful information is gained from these types of tests as far as the mechanism of human hearing, it is difficult to see how ‘real world’ situations are treated by the nervous system. Of notable exception is a four part series Localization of Sound in Rooms by W.M. Hartmann and Brad Rakerd (of Michigan State University) and the work of Guettler, et al. [2000].

It is important at this point to clarify several terms. For the purpose of this work, a sound is one of a group being tested or considered as an attractor for inclusion in the library for the audio navigation system, or, a similar attractor used in other research. In a more physical sense, a sound consists of mechanical vibrations and waves of an elastic medium within the range of human hearing. Terms that relate to the physical definition will always be used as a compound phrase, such as sound source, sound event, sound signal, etc. [Blauert 1997]. The distinction is important. For example, one sound source may be capable of producing many different sounds.

1.1 LOCALIZATION

The PING system relies on the assumption that users will be able to identify correctly the location of beacons based upon the sound emanating from them. Hence, it
is necessary to understand the processes by which users of the system identify the sound source location.

1.1.1 AN INTRODUCTION TO SOUND SOURCE LOCALIZATION BY HUMANS

It is generally accepted that human hearing utilizes two types of cues: monaural (information received by one ear) and binaural (information received by both ears) to localize sound events. Each cue provides unique details about a sound and it is a combination of the two cues that allow humans to recognize and locate sound events in an environment. Localization is thus, the ability of a listener to identify the origin of a sound event.

Binaural hearing cues consist of differences in time of arrival, or phase, between the two ears and level of a sound event between the two ears. These are known as interaural level differences (ILDs) and interaural time difference (ITDs) and were first reported by Lord Rayleigh [Rayleigh 1907]. The term interaural time difference is also known as the interaural phase difference, since the “phase can be measured in units of time or angles” [Blasch, Wiener, & Welsh 1997, pg. 108]. The origin of the ILD can be explained using a simplified model of the head. If one considers the head to be a sphere with the ears located at opposite poles on the sphere, then the ratio of the acoustical intensity of an incoming sound wave between the two ears, expressed in decibels, is the level difference. A visual description would be that the head casts a shadow over the ear at the pole opposite to the incoming wave, thus reducing the intensity at that ear. The ILD is frequency dependent and becomes less dominant for frequencies above 500 Hz where the wavelength is great enough for the wave to be diffracted by the head. The
cause of the ITD is simply the difference in the arrival time of the sound wave to each ear. Using the simplified sphere model of the head, the ITD can be expressed mathematically as:

\[
\Delta t = \frac{3a}{c} \sin \theta \quad \text{Equation 1}
\]

where, \(a\) is the radius of the head, \(c\) is the speed of sound, and \(\theta\) is the angle from the median plane as shown in Figure 1.1. However, if the wavelength is less than twice the diameter of the head, then in the steady-state situation, the sound source will be localized as being opposite of its actual location, see Figure 1.2 [Hartmann 1999]. The two cues gave rise to Lord Rayleigh’s *duplex theory*, which states that azimuth location consists mainly of high-frequency information derived from ILDs and low-frequency information derive from ITDs. Multiple studies have been conducted, the results of which, have been consistent with the duplex theory. [Middlebrooks & Green 1991]

![Figure 1.1. Simplified spherical head model.](image-url)
Monaural hearing cues (also called spectral or pinnea cues) are the result of the scattering of sound waves by the listener’s head, outer ears, shoulders, and torso [Hartmann 1999, Hofman & Van Opstal 1998]. This scattering can be expressed as a complex transfer function of the incoming sound wave and the sound wave that reaches the eardrum, known as the head-related transfer function (HRTF), which can be measured for individuals (see Blauert 1997, pgs. 78-84 and Ch. 4 for a listing of such experiments.

Figure 1.2. ITD confusion.

“Interaural time differences, given by the difference in arrival times of waveform features at the two ears, are useful localization cues only for long wavelengths. In (a), the signal comes from the right, and waveform features such as the peak numbered 1 arrive at the right ear before arriving at the left. Because the wavelength is greater than twice the head diameter, no confusion is caused by other peaks of the waveform, such as peaks 0 or 2. In (b), the signal again comes from the right, but the wavelength is shorter than twice the head diameter. As a result, every feature of cycle 2 arriving at the right ear is immediately preceded by a corresponding feature from cycle 1 at the left ear. The listener naturally concludes that the source is on the left, contrary to fact.”

(credit: Hartmann 1999, pg. 4)
and further information on the process). Figure 1.3 shows an example of measured HRTFs for different sound source locations. The HRTF is individualistic, but typically amplifies particular frequency bands depending on the origin of the sound source. For example, a sound source to the rear of the listener is typically amplified in the 1000 Hz band while a sound source in front of the listener would be amplified in the 3000 Hz band. Monaural cues aid in solving front/back and vertical localization confusion of a sound source that can arise from binaural cues. This is not to say that the monaural cues cannot be misinterpreted. A sound source located in front of the listener may have a strong component in the 1000 Hz band that could be interpreted as locating the sound

---

**Figure 1.3. The head-related transfer function.**

“The [head related] transfer function, which incorporates the effects of secondary scatterers such as the outer ears, assists in eliminating front–back confusion. The curves show the spectrum of a small loudspeaker as heard in the left ear of a manikin when the speaker is in front (red), overhead (blue), and in back (green). A comparison of the curves reveals the relative gains of the anatomical transfer function.”

(credit: Hartmann 1999, pg. 6)
source to the rear. [Hartmann 1999, Middlebrooks & Green 1991, Rakerd, Hartmann & McCaskey 1999]

Since each localization cue, taken alone, can be misinterpreted and lead to localization confusion, it is some combination of the binaural and monaural cues that fully explain localization of sound sources by humans. The available literature will often claim that one cue is more important than another, however, many times it is unclear as to whether this is a result of the test procedure, the test environment, or truly the result of the mechanism of human hearing. [Middlebrooks & Green 1991]

1.1.2 SIGNIFICANT LOCALIZATION RESEARCH PRIOR TO 1991

In 1974, the first comprehensive review of spatial hearing literature, Raechumliches Hoeren was published in German by Jens Blauert. In 1983, the book was translated into English and additional material was added that covered the advancements made since the 1974 edition was published. The book, in its English translation, was reprinted in 1997. Since its first publication, Spatial Hearing has served as the basis for much of the research into spatial hearing, including localization. The book is regularly cited as a reference, in both this work and others. However, since the book does not focus on localization, its direct usefulness in the research undertaken here is limited.

Of more interest to this research is the review by Middlebrooks & Green [1991] that focuses, as the title would indicate, on Sound Localization by Human Listeners. In addition to the information on binaural and monaural cues for localization previously mentioned, there is also discussion of dynamic cues for localization. Most of the dynamic cues available (or unavailable) are the result of (or the lack of) head movements
by the listener. Early experiments, reported in 1939 and 1940 by Wallach indicated a strong dependence on dynamic cues over monaural cues. On the other hand, experiments conducted by multiple research teams in the late 1960s showed dynamic cues having a far less significant role in localization [Middlebrooks & Green 1991]. That is not to say that the dynamic cues were unimportant, but that they provided only 10-30% better localization than without any head movement. The same research showed increasing the “duration of the sound source from 0.03 sec to 1 sec reduced the average error from 10° to 2°, an improvement of 500%!” [Middlebrooks & Green 1991, pg. 19]. From the publications dealing with dynamic cues, Middlebrooks and Green conclude that,

“Head motion is probably not a critical part of the localization process, except in cases where time permits a very detailed assessment of location, in which case one tries to localize the source by turning the head toward the putative location. Sound localization is only moderately more precise when the listener points directly toward the source. The process is not analogous to localizing a visual source on the fovea of the retina. Thus, head motion provides only a moderate increase in localization accuracy.” [pg. 22]

1.1.3 SIGNIFICANT LOCALIZATION RESEARCH SINCE 1991

More recently, much attention has been paid to the effects of monaural (spectral) cues on localization. Two papers are relevant to the research being conducted here. One paper deals with localization of sound sources where the spectral cues of the sounds are modified by various sub-band manipulation of broadband noise at the source. The second paper makes similar modifications to the sounds, but presents the information to the listener through modified HRTFs via headphones.
Hofman and Van Opstal [1998] examined the effect three types of stimuli had on the localization of the sound source. The three stimuli had “identically shaped broadband averaged power spectra” [pg. 2634] but were of very different spectral-temporal composition. The authors attempted to determine whether the auditory system uses a time-integration approach over a long time period (100 ms or greater) or a ‘multiple-look’ approach over multiple short time periods (several ms) to determine vertical localization and front-back confusions. Observations regarding localization in the horizontal plane, while not the original goal of the study, resulted from the manner in which the testing was done.

Seven subjects participated in the research, which took place in a sound attenuated room with dimensions 3 meters cube. The background noise level was 30 dBA. Each subject was given orientation training on sound source localization or had previous experience. Localization was measured by tracking eye movements. This method had been shown to be accurate in prior research. The sound source was a single speaker, the position of which was adjustable within a half hemisphere of the listener (Figure 1.4). Three types of sound stimuli were presented: broadband noise of durations $D = 3, 5, 10, 20, 40, 80$ ms; successions of $3$ ms of white noise cycling at $D = 3, 10, 20, 40, 80$ ms and having a total duration of about 500 ms; and FM sweeps with periods of $T = 1.28, 2.56, 5.12, 10.24, 20.48, 40.96$ ms and a total duration of about 500 ms. Examples of the three types of stimuli are shown in Figure 1.5. [Hofman & Van Opstal 1998]
Results of testing showed the following. Elevation gain increased as a function of stimulus duration, up to 80 ms, after which no significant improvement was seen. The latency between stimulus onset and eye movement response was increased by 20 ms for the short-burst successions of noise over that of the long-burst successions, but the data did not explain the delay. Duration length had almost no effect on azimuth localization with short durations producing localization results as accurate as long durations. Testing with the FM sweeps resulted in consistent results over all the variations, but was most accurate for periods $T < 5$ ms. The authors made the following conclusions regarding localization. First, elevation localization works on a ‘multiple-look’ strategy where

Figure 1.4. Experimental setup of Hofman and Van Opstal.

“Experimental setup for delivering acoustic stimuli at various spatial locations. Two stepping motors, M1 and M2, independently control the rotation angles $\phi_1$ and $\phi_2$, respectively. This construction ensures a fixed distance from the speaker to the center of the subjects’ eyes (0.9m) for any stimulus direction ($\phi_1$, $\phi_2$).”

(credit: Hofman & Van Opstal 1998, pg. 2636)
information is examined within intervals of about 5ms and combined for a final location estimate. This requires the sound to have a broadband spectrum and be presented in short-term intervals. Second, long duration sounds of at least 80 ms are necessary to provide enough ‘multiple-looks’ to give the best localization in the vertical plane.

Figure 1.5. Stimuli examples.

“Examples of the three stimulus types used in the localization experiments: a noise burst with $D = 520$ ms (a), a pulse train with $\Delta T = 20$ ms (b) and an FM sweep with $\Delta T = 20$ ms (c). The graphs show the stimuli from 20 ms before stimulus onset until 80 ms after stimulus onset. The large panels show the sonograms, which describe the spectro-temporal behavior of the power spectrum; spectral power is coded by the gray scale, where bright corresponds to low power, dark to high power. The panels on the right contain the time-integrated power spectra (all on the same scale). Finally, the lower panels show the stimulus waveforms.”

(credit: Hofman & Van Opstal 1998, pg. 2637)
Horizontal localization can be achieved in far fewer ‘looks’. Third, long periods of silence between sound bursts reduce the accuracy of elevation localization. Finally, the data did not support the concept that localization is the result of long term (100 ms or greater) time-integration, but instead, a short-term ‘multiple-look’ averaging [Hofman & Van Opstal 1998].

A similar examination of spectral cue influence on localization was conducted by Langendijk and Bronkhorst [2002]. In their study, various cues were removed from an individual’s HRTF and replaced by an average value for frequencies above 4 kHz. The researchers sub-divided the HRTF into an average transfer function (ATF) and a directional transfer function (DTF). The ATF remains constant for all source positions, whereas the DTF varies with source location. Therefore, to modify the HRTF, only the DTF needed to be changed. Figure 1.6 shows an example of the effect on the DTF for each test condition. A 200 ms burst of white noise was used as the stimuli. By modifying the HRTF, the new virtual sounds were presented over headphones to listeners who were then asked to locate the sounds’ origin. Many previous studies are cited by Langendijk and Bronkhorst that support the use of virtual sounds presented via headphones for localization experiments. Eight listeners participated in the study. All had previous localization experience. The virtual sound sources were limited to 23 equally spaced locations on the right hemisphere, of a virtual sphere, centered on the participants head. Participants were instructed to keep their eyes closed during the tests, which were conducted in a soundproof booth. [Langendijk & Bronkhorst 2002] As was expected, localization for the baseline condition was accurate for all of the participants over most of the locations. The 2/1-octave (four times the width of an octave band) and
1/1-octave conditions resulted in severely decreased localization accuracy for the majority of participants. All four, 1/2-octave conditions, though, resulted in very little change over the baseline. [Langendijk & Bronkhorst 2002]

**Figure 1.6. Effects of DTF modifications.**

“Example of the effect of the removal of spectral cues from the left ear DTF for one listener for one direction (0°, -56°) for each condition. From top to bottom: Baseline, 2-octave, 1-octave (low, middle, high), and 1/2-octave (low, middle-low, middle-high, high) conditions.”

(credit: Langendijk & Bronkhorst 2002, pg. 1585)
The results of the study led the authors to conclude that the vertical localization cues are located in the middle 1-octave band (5.7-11.3 kHz) while front-back localization cues are in the high 1-octave band (8-16 kHz), but that the exact frequencies used varied within these bands among the participants. It is proposed that peaks in the DTF within these bands, while correlating with the localization results, would be too narrow-band to aid in localization, since the peaks are much less than ½-octave in width, and the ½-octave manipulations tested had little effect on accurate localization [Langendijk & Bronkhorst 2002]. While these conclusions may be accurate over the frequency range tested, the exclusion of low frequencies leaves open the possibility that there are other low frequency cues acting to aid the localization process or resolve front-back confusions. The possibility of such cues has been suggested by Blauert [1969/70].

1.1.4 LOCALIZATION RESEARCH INVOLVING ‘REAL WORLD’ ENVIRONMENTS

The majority of the research discussed so far was conducted in sound attenuated environments. Such an environment allows for easy control of the variables to be tested by eliminating extraneous noise, but is a poor replication of the environments encountered on a daily basis by humans. Most ‘real world’ environments have some degree of reverberation or background noise that is likely to make localization more difficult than in an attenuated environment. This section covers research involving environments that are more realistic.

As mentioned previously, the work of Hartmann and Rakerd is some of the most significant in this area. The first of a four part series by Hartmann was published in 1983. The goal of the work was to examine the effects of the reverberant characteristics
of a room on sound localization. The study was conducted in the Espace de Projection (ESPRO) in Paris, France which is a medium sized concert hall of which the absorbent properties of the walls and ceiling can be modified as well as the ceiling height. This allows the room to be modified from reflective to absorbent [Hartmann, 1983]. The experimental setup used, shown in Figure 1.7, examined localization only in the horizontal plane. Three experiments were conducted.

The first experiment examined the subjects’ ability to localize a 50 ms pulsed 500 Hz sine tone in the ESPRO. The room was arranged in four conditions: absorbing, reflecting, low ceiling (reverberation time being half of the reflecting condition), and mirror reversed (the experimental setup was mirrored along the room centerline and then reversed in direction). During the experiment, the pulsed tone was directed to one of the eight speakers and the subject responded via the microphone as to which of the speakers was the sound source. Multiple runs of 80 trials were made for each subject in each room condition. Results of the first experiment showed that the effect of absorptive versus reflective conditions in the range produced (7 dB differences in states) had no significant effect on localization performance. A change in ceiling height from high to low did produce better localization performance. This is most likely because reflected sound from the ceiling agrees in direction with the direct sound from the source and reinforces the location, while sound reflected from the walls is received by the subject from a different direction than the source. Lowering the ceiling height caused the reflected sound from the ceiling to reach the listener sooner than reflected sound from the walls. [Hartmann 1983]
Experiment two was designed to test what effect the lack of an onset transient in the sound would have on localization. In experiment one, the pulsed tones had a strong attack transient, while in the second experiment the level of the sound was increased over a time of 6 to 10 seconds to a maximum intensity and then remained at the maximum

Figure 1.7. ESPRO experimental setup.
“Location of the eight-target speaker, circles 1-8: the subject \(x\); and talk-back microphone and speaker, \(m\) and \(s\). Positions 1 and 8 on the walls indicate the origins of reflections, calculated by the image method for an ideal rectangular room. Reflections from all speakers appear, in numerical order, between positions 1 and 8.”

(credit: Hartmann 1983, pg. 1381)

Experiment two was designed to test what effect the lack of an onset transient in the sound would have on localization. In experiment one, the pulsed tones had a strong attack transient, while in the second experiment the level of the sound was increased over a time of 6 to 10 seconds to a maximum intensity and then remained at the maximum
until the subject gave a response. Two sounds were examined – a continuous 500 Hz sine tone and broadband noise. Two room configurations, absorptive and reflective, were used. Localization performance for the sine tone was near to random guessing (average rms error of 12.9°) among all subjects. On the other hand, subjects were able to localize the broadband noise in the absorptive room with an average rms error of only 2.3° and in the reflective room with an average rms error of 3.2°. The results of the second experiment clearly show the advantage of broad frequency content in a sound for localization purposes. It also shows that a reverberant environment complicated the ability of the subjects to localize the sound source. One point raised by the author is that this degradation in localization performance could be a result of a lower signal-to-noise ratio (direct-to-reflected sound). [Hartmann 1983]

Good and Gilkey [1996] addressed the issue of signal-to-noise ratio as it pertains to sound localization and concluded that horizontal and vertical localization were only slightly affected by a decrease in the signal-to-noise ratio, while front-back localization was strongly affected. Differences in the experimental setups were present. Most significant is that Good and Gilkey used a broadband click-train as the target sound and broadband noise as masking to change the signal-to-noise ratio. Despite this difference, the horizontal rms error reported by Good and Gilkey for the signal-to-noise ratio in the range of approximately -6 to -12 dB are very similar (about 0°-5°) to those found by Hartmann. However, Giguere and Abel [1993] did not find a strong relationship between the direct signal-to-reverberant noises as it relates to localization performance.

Hartmann’s third experiment was designed to compare the performance of a spectrally complex sound with a sound of much less spectral density. The results of the
experiment showed that the spectrally complex tone was slightly more localizable, but exactly what characteristics in the spectrum that contributed to this could not be identified. [Hartmann 1983]

Further research was then conducted in the second part of the series to examine two of the findings from Hartmann’s first study; that lowering of the ceiling in the ESPRO improved localization accuracy and “that listeners had great difficulty localizing low-frequency tones with slow onsets” [Rakerd & Hartmann, 1985, pg.524]. Two experiments were conducted, the first to examine the effects of a single reflecting surface on impulsive tones, and the second, to examine low-frequency tones with a slow onset. The impulsive tone used was a 500 Hz sine tone pulse at a level of 50 dBA and with duration of 50 ms. The slow onset tone was again a 500 Hz sine tone at a level of 50 dBA, but with a seven-second onset. The tests were conducted in an anechoic chamber where the floor, ceiling, the left-side wall, or right-side wall (relative to the listener) could be covered with an acoustically reflective panel with a reflection coefficient of 100% at 500 Hz. The remainder of the setup (loudspeaker location, listener location, etc.) was similar to that shown in Figure 1.7. The localization procedure used for both tests was the same as that used by Hartmann [1983] in the first paper of the series. [Rakerd & Hartmann 1985]

In the test using impulsive tones there were significant differences in localization ability, depending on which surface was reflective (floor, ceiling, or wall), with the reflective wall condition showing the greatest error. While a reflective wall resulted in the lower accuracy than the empty (no reflective surface) room, having the reflective surface on the floor or ceiling did little to improve the accuracy of the listeners [Rakerd
This conflicts with the results of the first study that showed significant localization improvement for the reflective ceiling arrangement. For the reflective wall conditions, the authors hypothesized that, “the azimuth of a reflection will determine the nature of its influence” and therefore, “one expects that when the azimuth of a reflection competes with the azimuth of a direct sound subjects’ responses will be biased in the direction of the reflection” [pg. 526], however, this was not the case. Bias was uniform among the listeners and there was not a significant difference between the left- and right-wall conditions. This may have been influenced by the forced response method used in the study, which forces the listener to localize the sound at a limited number of finite locations. For the test of slow-onset tones, participants showed great difficulty in localizing the sound source. In fact, the errors were larger than would be found for random guessing, but were not random – the standard deviation was far below what would be found for random guessing [Rakerd & Hartmann 1985]. This is consistent with the results of the first study.

The third paper in the series further examines the effects of onset on localization and the effects of duration. The authors sought answers to four questions:

1. How does localization accuracy change as the onset of a tone is made more gradual, i.e., as the duration of the onset is increased?
2. Does localization accuracy depend upon onset duration per se or upon the onset rate?
3. Do onset effects vary with changes in the direction of room reflections and/or the delay of those reflections?
4. Do onset effects vary with the frequency of a tone?

[Rakerd & Hartmann 1986, pgs.1695-1696]

The experiments were conducted in the same anechoic chamber, with a moveable reflective wall, as in the second paper in the series, Rakerd & Hartmann, 1985. Four room arrangements were used - a reflective wall with brief reflection delay, a reflective ceiling with brief reflection delay, a reflective wall with long reflection delay, and the empty room (anechoic chamber without reflective wall). The loudspeaker arrangement relative to the listener was similar to that of Figure 1.7. Each test consisted of the subject completing one run of 100 trials. As in the previous works, the subject was asked to give a forced-response location of the sound source. A change from the previous studies now allowed the listener to give a location corresponding either to one of the loudspeaker locations, or, to marked locations left and right of the loudspeaker array. [Rakerd & Hartmann 1986]

Five tests were conducted in which onset duration was the focus. The first test examined the effects of onset duration in the ‘standard’ room configuration, which was the anechoic chamber with a reflective wall with brief reflection delay. The stimuli used were a 500 Hz sine tone at 65 dBA with a linear ramp onset of varying duration (0, 5, 10, 50, 100, 500, 1000, and 5000 ms). The other five tests were compared to this ‘standard’. The chamber was then altered to the reflective ceiling with brief reflection delay configuration for the second test. The stimuli tested were the same, except for the omission of the 5000 ms onset. The third test was the same as the second, but used the reflective wall with long reflection delay layout. To look at what the frequency of the
stimulus may contribute, the fourth test used a 1500 Hz sine tone while keeping all other aspects the same as the ‘standard’ configuration. For the fifth test, the peak amplitude of the stimulus was decreased to 40 dBA while keeping the ‘standard’ design. [Rakerd & Hartmann 1986]

Three further tests were conducted that examined the impact of tone duration on localization. The three tests were identical in their design except for the placement of the reflective panel in the room (brief wall delay, brief ceiling delay, and long wall delay). The stimuli used were 500 Hz sine tone pulses at 65 dBA. The pulse durations examined were 5, 50, 500, 1000, and 2000 ms. [Rakerd & Hartmann 1986]

Results of the onset testing showed conclusively that rapid stimulus onset is highly beneficial to sound source localization in reverberant environments. A rapid onset increases the usefulness of the precedence effect when locating the sound source while longer onsets detract from the precedence effect because of the near steady state sound field that is created. The acceptable length of a stimulus’ onset was found to be determined by its sound pressure level. Higher sound pressure levels resulted in longer acceptable onset times than low sound pressure levels. As found in the first study [1983], vertical reflections (caused by a reflective ceiling) were less disruptive than horizontal reflections (caused by a reflective wall). Finally, frequency did not have a significant effect on localization. These results were again verified by a series of similar tests conducted more recently [Rakerd & Hartmann 2005]. The tone duration tests showed no effect on localization performance for any of the room configurations [Rakerd & Hartmann 1986]. So, to answer the authors’ original four questions:
1. As the onset of a tone is made more gradual, i.e., as the duration of the onset is increased, the localization accuracy is decreased.

2. Localization accuracy depends upon the onset rate, for example Pa/sec, and not upon the onset duration. However, for a specific sound pressure level, localization accuracy is dependent upon duration.

3. Onset effects do vary with changes in the direction of room reflections and the delay of those reflections. Vertical reflections are less disruptive than horizontal deflections and long reflection delays allow for longer onset duration.

4. Onset effects do not vary with the frequency of a tone or the tone’s overall duration.

The results of the work by Hartmann and Rakerd prompted further research by Giguere and Abel [1993]. They considered how reverberation time, stimulus center frequency, stimulus rise and decay time, and sound source location affected localization. Testing was carried out in a reverberation chamber that could be configured as either reverberant or absorbent by adding lining materials. Two loudspeaker arrays, shown in Figure 1.8, were compared. Stimuli were 1/3-octave band noise with center frequencies at 500, 1000, 2000, and 4000 Hz. These stimuli had either a 5 ms or a 200 ms linear rise/decay. They were played at a level of 70 dB. For each room configuration (reverberant/absorbent) participants completed three sets of 120 trials for each of the 16 possible conditions. Subjects were asked to identify the sound source location using the forced location method. Results of the testing were consistent with the results of the work by Hartmann and Rakerd, (such as, an increase in reverberation time resulting in a decrease in localization performance independent of frequency, the effects of reflection
arrival time on localization, and the importance of sound onset in reverberant rooms) with the exception being the signal-to-noise relationship discussed earlier.

![Figure 1.8. Loudspeaker array schematic.](image)

“Schematic of the two loudspeaker arrays” used by Giguere and Abel.

(credit: Giguere and Abel 1993, pg. 770)

1.2 **RECOGNITION**

Unlike the process of sound source localization, there has been much less research into sound identification. Several experiments conducted by Rakerd, Hartmann, and McCaskey [1999] studied sound recognition in the median sagittal plane. In the first experiment, listeners were tasked with identifying a sound played as being either sound A
or sound B. The sounds were 0.5 seconds of broadband noise with a 2/3-octave bandwidth intensified by 10 dB (a bump) or attenuated by 10 dB (a dip). The sounds were presented at 72 dBA. There were 10 pairs of sounds, where a pair consisted of a bump and a dip on either side of a mean frequency, some of which are shown in Table 1.1. The sounds were played via one of five loudspeakers placed every 45 degrees in an arc along the median sagittal plane. The arc began directly in front of the listener’s head and ended directly behind. Each listener participated in 50 trials, of which, 25 trials had sounds presented at a fixed location directly in front of the listener and 25 trials had sounds presented at a random location. [Rakerd, Hartmann & McCaskey 1999]

<table>
<thead>
<tr>
<th>Center frequency (kHz)</th>
<th>Pair mean frequency (kHz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>1.4</td>
</tr>
<tr>
<td>1.8</td>
<td>2.3</td>
</tr>
<tr>
<td>2.8</td>
<td>3.5</td>
</tr>
<tr>
<td>4.5</td>
<td>5.7</td>
</tr>
<tr>
<td>7.1</td>
<td>8.9</td>
</tr>
<tr>
<td>11.2</td>
<td></td>
</tr>
</tbody>
</table>

**Table 1.1. A-B sound pairs.**

“Center frequencies of 2/3-octave bumps and dips used in experiment 1 are given in the first column. A neighboring pair of bumps/dips, used as an A-B pair in experiment 1, is characterized by a pair mean frequency, given in the second column.”

(credit: Rakerd, Hartmann, & McCaskey 1999, pg. 2813)
Results of the first experiment revealed similar results for all participants. Most participants found bumps easier to identify than dips. It was also noted that recognition was almost perfect for low frequency sounds (around 1-2 kHz) for both the fixed and random trials. Higher frequency sounds, however, resulted in recognition rates significantly worse than low frequency sounds for the fixed location trials and near to random guessing for the random location trials. [Rakerd, Hartmann & McCaskey 1999]

A second experiment was designed to see if the results of the first experiment would hold true for sounds that are more complex. The same process was used as in the first experiment, but with fewer trials. Sounds in the second experiment included 2/3-octave and 1/3-octave bumps or dips with 2/3-octave or 4/3 octave separations between sound pairs. There were 38 pairs tested in total with 19 having bumps and 19 having dips. [Rakerd, Hartmann & McCaskey 1999]

Experiment 2 reinforced the results of Experiment 1 and provided further insight into the effects of frequency. For sounds with low frequency structure, 1-3 kHz, listeners were able to identify the sound correctly. Sounds with the main frequency content above 4 kHz became increasingly difficult to identify correctly with increasing frequency. Above 8 kHz in the random location trials, it was nearly impossible for participants to identify the sounds correctly. [Rakerd, Hartmann & McCaskey 1999]

The third experiment compared sound recognition to sound source localization ability. Sounds were selected for each participant based on their performance in the previous tests. Two sound pairs that were easily identified by the participant and two pairs that were difficult to identify were included. The participants began by repeating the process of the first experiment. The test was then repeated, but listeners were asked
to identify from which loudspeaker location the sound originated. The results of the two tasks were then compared to see if a listener’s ability to localize a sound influences the ability to recognize a sound. It was found that correct localization does not necessarily lead to correct recognition. [Rakerd, Hartmann & McCaskey 1999]

A fourth and final experiment combined either two bumps or two dips at different ends of the frequency spectrum within one sound. Since most ‘real world’ sounds contain frequencies across the spectrum, the authors wanted to see if the high recognition that resulted from low frequency content is canceled by the poor recognition resulting from high frequency content. The procedure was the same as in Experiment 1. Analysis showed that the combination of frequencies produced identification accuracies as good as the low frequency only sounds. [Rakerd, Hartmann & McCaskey 1999]

The four experiments clearly show that sound recognition in the median sagittal plane is primarily linked to frequency content below 3 kHz. This holds true whether the sound originates from a known, or unknown, location. Furthermore, inclusion of higher frequency content does not degrade recognition [Rakerd, Hartmann & McCaskey 1999]. Whether this holds for recognition in the horizontal plane is a question that remains.

The fourth experiment conducted by Rakerd, Hartmann & McCaskey attempted to replicate ‘real world’ sounds while maintaining consistency with previous tests. It is clear though, that the sounds generated do not contain the complexities that are present in ‘real world’ sounds. Research by Guettler, Bolia, and Nelson [2000] is perhaps, the most realistic testing conducted in the area of sound recognition. Concerned with developing audio ‘displays’ for aircraft cockpits, Guettler, Bolia, and Nelson used ten sounds, including: a “baby crying, bees, birds, barking dog, frog, helicopter, horse gallop, lawn
mower, police whistle, and wooden chimes” [unpublished, pg. 3] in their testing. Unfortunately, the only records that remain of this work are a citation for a poster presentation, an unpublished draft research plan, and the raw experimental data that without knowledge of the actual test procedures used cannot be analyzed. Thus, its usefulness is limited to showing a previous concern for ‘real world’ sounds and providing information on possible test procedures. The focus of the work by Guettler, Bolia, and Nelson is on the number of sounds and physical tasks that can be monitored by a single person, whereas the research conducted for this thesis is focused on the recognition of a single sound among many other distracters, making for slightly different goals.

1.3 NAVIGATION

Although audio cues are used in orientation and mobility training for the visually impaired, the focus is typically on the use of audio cues to provide information on location relative to a known path. For instance, a pedestrian who is blind traveling between point A and point B may associate the sound of a fountain with the \textit{a priori} knowledge that the fountain is two streets before point B. In another case, the sound of traffic at an intersection can aid in determining when it is safest to cross the street, but even if the intersection has an audio aid, the aid’s primary function is to indicate when it is safe to cross, not where to cross.

One study [Loomis, Hebert, and Cicinelli 1990] does stand out, because its sole purpose was to evaluate pedestrian navigation using only audio cues presented in a real and virtual environment. The study has the added benefit that one of the participants is blind. Unfortunately, it only uses one type of sound and so it is unknown whether the
results will be similar if a different sound is used. The experiment was conducted within a large gymnasium with a background sound level of 34 dB and in which the authors note, “Echoes from the gymnasium walls were quite noticeable” [pg. 1761]. The stimulus presented was a 5 Hz square-wave, modified by a high-pass filter to produce 10 pulses per second at a level of 40 dB. For the case of the real environment, the sound was presented via two small loudspeakers mounted on a moveable boom. One loudspeaker was in the horizontal plane and the other was at a 50-degree angle below horizontal. The boom could be moved radially around the participant with a radius of 15 meters. In the case of the virtual environment, the sound was manipulated by a computer system with analog-to-digital conversion, to replicate the binaural cues that would be heard in the real environment, and then presented over headphones to the participant. [Loomis, Hebert, and Cicinelli 1990]

Each participant took part in two tests, one real and one virtual. Each test consisted of 18 trials corresponding 18 different sound source locations around the listener. The listener was positioned facing the same direction prior to each trial. When the stimulus was activated, the subject proceeded to walk towards where he believed the sound to be originating. When confident that he had reached the source location, the subject indicated so by saying, “stop”. During the trial, the computer system recorded the participant’s location and head movement. The path trajectory could then be plotted as shown in Figure 1.9, and head movements analyzed. [Loomis, Hebert, and Cicinelli 1990]

The tests resulted in two important observations. First, head rotations showed that subjects easily identified sources as being to their left or right, and more so for extreme
locations. The trajectories support this with immediate body rotation at the start to face the sound source. This held for both real and virtual presentation [Loomis, Hebert, and Cicinelli, 1990]. Second, users were able to navigate successfully between their initial location and the sound source with little prior training and in most cases with only small deviations from the straight path. The trajectories did take on more of an arc under the virtual environment than in the real environment. This is seen in Figure 1.9. The importance of participant FG, who is blind, will become apparent in later chapters.

![Figure 1:9. Walking trajectories.](credit: Loomis, Hebert, and Cicinelli 1990, pg. 1760)

“Subjects walking trajectories to the 18 sound sources in the two conditions. The centers of the large circles represent the locations of the virtual sounds in the 15 x 15m workspace. Tie marks represent one second intervals. Workspace north is at the top of each panel. Subjects faced north prior to initiation of the sound.”

(credit: Loomis, Hebert, and Cicinelli 1990, pg. 1760)
2. EXPERIMENTAL DESIGN

The current research used a quantitative and qualitative research design. The data determined which sound stimuli are easiest to remember, most efficient for localization and navigation purposes in indoor and outdoor environments, and the range of sound intensities necessary in the different environments.

2.1 PREPARATION

Permission to conduct the study was granted by the Human Subjects Institutional Review Board (HSIRB) on July 28, 2004 (see Appendix A for a copy of the research protocol clearance). The following materials were reviewed by HSIRB: protocol description, consent form, and recording instruments.

Following approval by the HSIRB, the process of identifying potential sounds for testing began. The search for sounds focused on royalty-free *.wav and MP3 format sounds available from various sound effects websites. Chosen were twenty-six (26) total sounds, shown in Table 2.1, including the best ten (10) sounds from a prior phase of PING testing and sixteen (16) new sounds. The prior phase of testing only examined localization in the horizontal plane. The sounds under consideration were similar to those used by Guettler, Bolia, and Nelson [2000] such as bells, birdcalls, whistles, claps, etc. All sounds were normalized such that their time-domain peak levels were identical. Normalization was accomplished by use of a MatLab program (developed in the first phase of the PING tests), which can be found in Appendix B.
The program functioned in the following manner. First, the user was prompted to enter the name of the file to be normalized from the working directory. The program then loaded that file as a *.wav file. Next, the data were evaluated to identify the maximum and minimum amplitude in terms of an absolute value. Using the peak absolute value of the amplitude and a user specified scaling factor, the sound was normalized by multiplying each of the data points in the file by the ratio of the scaling factor.
factor to the absolute peak amplitude. The modified file was then saved as a new file named originalname2.wav where, originalname was the name of the file initially entered by the user.

The normalized sounds were then modified in the following manners using the free software program, Audacity 1.2.1 (http://audacity.sourceforge.net): pitch was doubled and halved, tempo was doubled and halved, amplitude was doubled, and the sound was repeated twice in a row. The result was 156 total sounds to be tested. These changes were made to examine the effect they might have had on sound localization and likeability.

This research was conducted in three university settings: 1) a laboratory reverberation chamber located in Western Michigan University’s (WMU’s) Noise and Vibration Laboratory, 2) a large university hallway and large empty classroom, and 3) a large open outdoor area on the University campus. It was necessary to identify an area that met the requirements of number two listed above. The goal was to find a room that would be similar in size and acoustic characteristics to a museum exhibit area. Under consideration were locations such as the WMU Student Recreation Center, Bernhard Center Ballroom, and College of Engineering and Applied Sciences. After researching the options, it was decided that a large unused room with high ceilings in WMU’s Kohrman Hall would be the best choice. Permission was obtained to use the room in Kohrman Hall as well as the reverberation chamber in the Noise and Vibration Lab.
2.2 PARTICIPANT SELECTION

This study included two sets of participants. The first set was comprised of five (sighted) volunteers from within the Mechanical and Aeronautical Engineering department at WMU. This group was utilized to reduce the large number of initial sounds under consideration to a number more manageable for further testing. These subjects did not receive a formal hearing screening, but were questioned about any known hearing problems. None of the participants in this part of the study reported any known problems. The second set of participants included five (blind or severely visually impaired) students identified by WMU’s Department of Blindness and Low Vision Studies as being interested in assisting with the study. They ranged in age from their twenties to their fifties and consisted of both congenitally blind and adventitiously blind and all participants in the second group were experienced cane travelers. Each of these participants was given a hearing screening assessment using a Maico model MA 41 audiometer. All participants in the second group possessed normal hearing, with pure tone hearing thresholds between 5 dB and 25 dB, i.e. the lowest sound pressure level at which they could identify the pure tone was between 5 dB and 25 dB.

Potential participants of the second group were made aware of a $20 payment for each session in which they participated. Those who began the research but decided to discontinue for any reason were paid for the individual sessions in which they participated. Recruitment of students was based on a first-come first-served basis. Those identified interested beyond the first five were placed on a waiting list should any of the original participants have later declined to participate. The researchers then met the students from both groups and explained the details of the project. The potential
participants had an opportunity to learn more about the study and have their questions answered before they agreed to participate.

In order to qualify, individuals were required to be 21 years of age or above, and possess normal hearing. Any person found to have an air conduction threshold greater than 25 dB was excluded from the study. In other words, if “the softest level at which a {participant} responds to a single frequency tone presented through earphones” [Beltone 2002] was greater than 25 dB, then they could not participate. The participants were informed before the screening that if they did not meet the above air conduction thresholds, they would be excluded from the study, and be given the opportunity of being referred for a more comprehensive auditory evaluation. The consent form for the participants was available in large print and Braille. The selection of participants in the study was performed concurrently with the selection of sounds to be tested.

2.3 PROCEDURES

The researchers met with each participant for a maximum of five hours over a week’s time. The details of the experiment were explained, a hearing screening provided, and data collected on subjects’ perception and performance with various sound stimuli. The researchers were responsible for tabulating and analyzing all of the data and statistics collected in the course of the user trials. A list of the research activities is given below in the sequence in which they were performed:

1. Meet with volunteer subjects and present the consent form.
2. Select the participants and explain the procedure.
3. Meet the participants at the arranged five sessions.
4. Provide a hearing screening and train to a criterion using a 1000Hz tone.

5. In an indoor setting, present various sound stimuli for testing memory for different sounds.

6. In an indoor setting, present sounds of various durations to determine ideal length of presentation.

7. In an indoor setting present sounds of various intensities to determine the decibel range for effective localization.

8. Evaluate efficiency of indoor travel routes that are directed by beacons using various sound stimuli.

9. Evaluate repetition rate of different sounds required for effective location of audio beacons along an indoor travel route.

10. In an outdoor setting present sounds of various intensities to determine the decibel range for effective localization.

11. Present sounds of various durations to determine the decibel range for effective localization in outdoor environments.

12. Evaluate efficiency of outdoor travel routes that are directed by beacons using various sound stimuli.

13. Evaluate repetition rate of different sounds required for effective location of audio beacons along an outdoor travel route.

14. Analyze the performance of the subjects with each set of sound characteristics in each environment.

15. Develop a set of recommended sounds.
Steps one through four were to provide the participants with information on the purpose, procedures, risks, and benefits of the study as described in the Protocol Outline presented for approval by the HSIRB. Step five familiarized the participants with the types of sounds under examination. Steps six through thirteen consisted of the data collection phase of the study and provided the data to be analyzed in step fourteen. Finally, in step fifteen, the analysis of the collected data was used to make recommendations on the sounds to be included in the PING catalog and their characteristics.

Steps six through thirteen, listed above, can be broken down into six different tests that were conducted to establish an additional library of sounds for the navigation system. Each experiment was designed to test a specific aspect of the sounds in relation to their use as a navigation aid.

In the following sub-sections, each of the six tests will be discussed in detail. The tests are initial sound selection, indoor navigability, indoor localization, indoor sound level, outdoor sound level, and sound identification and recognition. Participants from the Mechanical Engineering department, all of whom were sighted, only took part in the initial sound selection. All remaining tests were conducted with the help of the participants from the Department of Blindness and Low Vision Studies. The Indoor Navigability test had five participants. The remaining tests had four of those five participants. The fifth participant was unavailable during the later testing periods.
2.3.1 INITIAL SOUND SELECTION

The goal of the first test performed was to reduce the initial set of twenty-six (26) sounds and those that had been modified in the preparation stage (156 sounds total) to a group manageable for conducting further tests. The test was similar to that used by Giguere and Abel [1993], in that it was conducted in a reverberation chamber with an array of loudspeakers around the listener who made a forced-response location selection. The reverberation chamber was 21 feet long by 14 feet wide by 9 feet high and is located within the Noise and Vibration Laboratory at WMU (Figure 2.1). Within the chamber, a semi-circular array of fifteen (15) 3.5” full range Radio Shack loudspeakers was setup with a seating position at the circle center. The loudspeakers were located from 0 to 180 degrees at 36-degree increments and at distances of 3, 6, and 9 feet from the seating position. In a corner of the chamber, a JBL Decade Series Model D315 three-way floor-standing loudspeaker was placed. A chair was located at the circle center for participants to sit in during the test. The test setup is shown in the diagram of Figure 2.2.

Hoffman and Van Opstal [1998] used a similar test arrangement, but rather than the distance of the loudspeaker away from the listener being variable, elevation was variable in their test. Since the plan is for most of the PING beacons to be installed in their permanent locations at heights between waist and head high (for a person of average height) the elevation was not a concern in this project. All of the array loudspeakers were at a height of 45” from the floor, which was approximately chest to head height for the participants.
Figure 2.1. Reverberation chamber test setup.
Shows seating position and 15-loudspeaker array (picture taken from doorway).

Figure 2.2. Overhead diagram of the reverberation chamber test setup.
A masking sound of white noise was generated by a Larson-Davis Model 2900B Real-time analyzer, amplified by a Technics model SA-EX110 stereo receiver, and played through the JBL loudspeaker at a level approximately equal to that of the normalized sounds and the ambient conditions measured in the prior phase of the project (about 60 dB). This was the average level measured within the Kalamazoo Museum of Art using a handheld sound level meter. The sounds to be tested were played in *.wav format using the Winamp media player on a Dell Optiplex GX110 computer. The output from the soundcard was input to a Radio Shack Optimus Integrated Stereo Amplifier model SA-155 and then to a manually controlled switch that was used to select the particular loudspeaker from which the sound was to be played. A schematic of the setup is shown in Figure 2.3.

Figure 2:3. Schematic diagram of the equipment setup used during the initial sound selection. Solid red arrows represent wired connections between equipment and the direction of information flow.
Five (sighted) volunteers from within the Mechanical and Aeronautical Engineering department at WMU were selected to listen to each of the one-hundred fifty-six (156) sounds as they were presented in a random order through a randomly selected loudspeaker in the array. As each sound was played, the participants were asked to rank the sound’s pleasantness on a scale of 1-3, with 1 being low and 3 being high, and record from what loudspeaker they believed the sound originated. The test coordinator recorded the actual loudspeaker location used.

<table>
<thead>
<tr>
<th>Sound Number</th>
<th>Sound Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Chime1</td>
</tr>
<tr>
<td>2</td>
<td>Chime2</td>
</tr>
<tr>
<td>3</td>
<td>Clang1</td>
</tr>
<tr>
<td>4</td>
<td>Clang2</td>
</tr>
<tr>
<td>5</td>
<td>Cuckoo</td>
</tr>
<tr>
<td>6</td>
<td>Jetson’s doorbell</td>
</tr>
<tr>
<td>7</td>
<td>Owl</td>
</tr>
<tr>
<td>8</td>
<td>Computer Operation</td>
</tr>
<tr>
<td>9</td>
<td>Ding-Ding</td>
</tr>
<tr>
<td>10</td>
<td>Loon</td>
</tr>
<tr>
<td>11</td>
<td>Woodpecker</td>
</tr>
<tr>
<td>12</td>
<td>Chime3</td>
</tr>
</tbody>
</table>

Table 2.2. Table of sounds and their descriptions as used in testing.

Based on accuracy of localization and comments from participants (data included in Appendix C), twelve (12) sounds, listed in Table 2.2, were chosen by Touch Graphics (the company funding the research) to be used from this point forward for testing. The only modification to have a positive effect, and only in some instances, was repeating the
sound twice in a row. This appeared to be the preference for sounds that were originally very short.

2.3.2 INDOOR NAVIGABILITY

One of the most important aspects of the sounds used for the PING catalog is a person’s ability to navigate an intended course while localizing a chosen sound. An important aspect of the system is that users choose their own sound (it is not assigned) out of an array of available sounds in the PING catalog; therefore, all of the sounds in the catalog must lend themselves to successful navigation for all users, not just one user in particular. The indoor navigability test was designed to test each of the twelve (12) candidate sounds and reveal which are better suited for navigation in an indoor environment.

This test was carried out in Kohrman Hall on WMU’s main campus. Kohrman Hall, at the time, was undergoing minor renovations and many large rooms were being unused. The test area consisted of two intersecting hallways, a corridor, and two large empty rooms (see shaded area of Figure 2.4). The majority of the testing took place in room 2114, the ‘exhibit’ space. The dimensions of the room (29’ x 60’) were representative of an exhibit area in the museum. In their similar experiments on navigation using sound, Loomis, Hebert, and Cicinelli [1990] used a large gymnasium as the test location. Room 2214 in Kohrman Hall had a similar construction to a typical elementary school gymnasium.
Twelve different pathways, one for each sound, were created within the test area. All paths started at the intersection of the two hallways, indicated by the green mark (circle with cross) in Figure 2.4. From the starting point, six of the paths proceeded down the hallway (horizontally in Figure 2.4), through the corridor, and into room 2114. The other six paths led down the hallway (vertically in Figure 2.4) and directly into room 2114. Each path consisted of five straight legs of varying length, of which, at least three were located within the ‘exhibit’ space. At the end of each leg, a PING beacon was placed, thus, allowing the user to travel from one beacon to the next. The audio beacons ranged in height from 29” to 78”. Drawings of each path were generated using laser surveying and AutoCAD software. Figure 2.5 on the following page shows the beacon
locations, labeled as PING #, and a list of the twelve paths defined by beacon number. Figure 2.6 shows the different paths represented by colored lines. The PING # labeled in the figures correspond to the serial numbers of the unit. Some units were damaged during shipping and could not be used, so not all numbers between 1 and 13 appear in the figures.

The beacons were controlled by the PING computer system, provided by Touch Graphics, and were activated by a cordless telephone. After activating the system, the participants heard a recorded message providing instructions on how to use the phone to control the system. For the purpose of the experiment, it was only necessary for participants to press 1 to ‘ping’ the beacon (play sound) and press 2 when the destination had been reached to activate the next beacon along the path. In addition to the user-activated sounds, a separate laptop computer was utilized; along with an Optimus model SA-155 Integrated Stereo Amplifier and two Optimus model Pro-X55AV 2-way bookshelf loudspeakers to play a background soundtrack that would better replicate the ambient sound of a museum experience. The soundtrack consisted of two recordings, of about two minutes each, played in sequence and looped continuously. The recordings were found on the internet and were specifically designed to simulate background ‘conversations’, much like what would be heard in a public museum space. Figure 2.7 shows a schematic diagram of the equipment setup. The background sound level was set to 59 dBA, which is consistent with the level measured at the Kalamazoo Museum of Art in phase one of the project. The PING sounds were set to a level of approximately 60 dBA. This combination of levels provides a realistic representation of what can be expected in a museum.
Figure 2.5. Experimental space diagram (a). Showing beacon location (identified as PING #) and listing the 12 paths corresponding to the 12 sounds being tested. All paths begin at the point labeled START and end at the last beacon listed.
Figure 2.6. Experimental space diagram (b). Showing beacon location (identified as PING #) and the 12 paths corresponding to the 12 sounds being tested. All paths begin at the point labeled START. (Note: Some paths may overlap and some legs may not be visible. Refer to Figure 2.5 for beacon order.)
Five participants, who were blind, from the Department of Blindness and Low Vision Studies and selected in the Participant Selection phase of the study participated this experiment. To account for differences in participant ability and to provide multiple data sets, three different participants traveled each path. The paths each participant traveled were randomly selected as “drawn from a hat”. At the start of the experiment, a description of the test was given to each user and they were given a chance to become familiar with the PING system. The participant then activated the system, listened to the recorded instructions and proceeded to travel from one beacon to the next. As the participant walked the route, a researcher followed behind and marked the path traveled, on the floor, using a water-based marker. This method was effective, but not as

Figure 2.7. **Schematic diagram of the equipment setup used in the indoor navigation study.** Solid red arrows represent wired connections between equipment and the direction of information flow. Dashed red arrows represent wireless connections between devices and the direction of information flow.
sophisticated as the magnetic tracking method used by Loomis, Hebert, and Cicinelli [1990]. In Figure 2.8, a participant is shown walking one of the paths towards the beacon. Note that the lines on the floor were to aid the research team in identifying the different ‘ideal’ routes, but were not perceptible by participants using a cane. The marks made on the floor identified only the sound/path being tested and not the participant. The researcher making the marks also made sure there were no obstacles in the participant’s route that could cause injury.

Figure 2.8. Participant walking path during the indoor navigability test.
After all five legs of the path had been completed; each participant was assisted in returning to the starting point. While the computer program was being setup for the next path, the participants were asked to rank subjectively the pleasantness of the sound and their ability to identify the location of the sound source. Any other comments by the participant regarding the sound were also noted. The process was then repeated for all sounds assigned to the participant and for all participants.

When all of the participants were finished, the research team followed the ideal paths, which had been marked on the floor prior to testing, and identified the extreme points of travel deviation along the right- and left-hand sides of the path by placing coded stickers on the floor. These extreme points marked an envelope of travel for each path leg. A professional land surveyor was then brought in to map the rooms, beacon locations, ideal paths, and path envelopes using a laser-surveying device manufactured by Leica Geosystems. Using the data gathered, the surveyor was able to create an AutoCAD drawing for each path. These are the drawings in Appendix D. An example is shown below in Figure 2.9, where the beacons are identified by PING #, the ideal path is identified by three (3) parallel lines between beacons, and the travel envelope is identified by the shaded region.
Figure 2.9. CAD drawing of an example path (sound 7). Showing ideal (purple, parallel lines) and actual (brown, shaded region) routes.
2.3.3 INDOOR LOCALIZATION AND SOUND LEVELS

The next step was to determine what sounds were most localizable and at what volumes (sound pressure levels). To accomplish this task, room 2114 shown in Figure 2.4 was used again. A layout similar to that of Hartmann [1983] was used. A listening position was identified at one end of the room and five beacons were placed in semi-circular fashion thirty-feet (30’) away and with four-foot (4’) spacing between beacons. The participant would stand at the listening position; use the cordless phone to ‘ping’ the beacon, and using a laser pointer would point in the direction from which he believed the sound to originate. The same background recordings used during the navigability test were also used for this test at sound pressure level of 60.8 dBA. Figures 2.10 and 2.11 show the experimental setup. The equipment was connected in the same manner as shown in Figure 2.7.

![Figure 2.10. Experimental setup for the indoor localization and sound level test.](image-url)
Figure 2.11. Overhead diagram of the experimental space for the indoor localization and sound level test.
Since the test also needed to determine the best volume level for each of the sounds, Adobe Audition 1.0 software was used to create three new sets of *.wav files representing a high volume (peak level = 65 dB), a medium volume (peak level = 55 dB), and low volume (peak level = 45 dB) for each of twelve (12) sounds. This is again similar to the tests by Hartmann [Hartmann and Rakerd, 1986] in which two sound pressure levels were compared, 65 dBA and 40 dBA. For each participant a unique playlist of sounds was created that randomized the sound heard, the volume, and the beacon location. The list assured that multiple participants would hear each sound to reduce human error as well.

On the day of testing, four of the five participants utilized in the indoor navigability study were available for this experiment. The purpose of the test was explained to the participants as well as instructions on how to use the laser pointer. Participants were given an opportunity to practice using a sound/volume combination not included on their playlist. They were allowed to ‘ping’ as many times as necessary until confident of the location. A record of the number of ‘pings’ was made.

Chalkboards were placed behind the beacons to provide a surface on which the researchers could see the laser beam. The chalkboards were marked with vertical lines in one-foot increments so that an approximate horizontal error could be measured. If the user identified the location within one-foot left or right of the beacon, it was considered accurate with no error. This was decided after noting the difficulty of maintaining the laser dot at an exact location due to slight hand motions. In Figure 2.12, a participant is seen pointing towards the beacons. For all participants the laser pointer was taped to their index finger to improve accuracy.
2.3.4 OUTDOOR SOUND LEVELS

The original test plan called for repeating the indoor trials outdoors. However, while setting up the outdoor test it was found that the background ambient sound pressure levels were significantly lower (48 dBA) than those used in the indoor test and that the researchers could clearly hear the ping beacon at a distance of over one hundred feet (100’). Outdoor areas will typically contain fewer reflective surfaces, such as walls and ceilings, than indoor spaces. The result is that the user will hear less indirect sound (reflections) in the outdoor area than indoors and more easily identify the beacon location. These factors being the case, it was decided that the indoor test represented the worst-case scenario and that it was unnecessary to test in the outdoor environment as long as the PING sounds are 10 dBA above the background level.

When there is a 10 dB difference, one can be assured that the sound is distinguishable from the background noise [Barron, 2002]. Without a 10 dB difference between the sound and any background noise there is no guarantee that the sound source,
the PING beacon being listened for, is actually the PING beacon and not the background noise. However, many people of normal hearing can clearly hear differences between sounds sources smaller than 10 dB. Although, if the difference is less than about 3 dB, the intensity level from the intended source becomes indistinguishable from the background [Barron, 2002].

2.3.5 SOUND IDENTIFICATION AND RECOGNITION

The final test was designed to see how well the participants could recognize an assigned ‘ping’ against a background of other ‘ping’ sounds. The need for the experiment arose from the fact that in a museum there could be multiple users activating the PING system at the same time. Therefore, the user must be able to pick out their ‘ping’ from among the other user’s ‘pings’. The reverberation chamber and loudspeaker array (shown in Figures 2.1 and 2.2 previously) was used again for this test. However, the setup changed from the initial sound selection test. A computer program was developed in LabView for automatically generating random sounds at a rapid pace. The program played the twelve (12) sounds in a random order, made a list of the order played and indicated if the participant responded to the sound. In addition, a background speaker played another set of random PING sounds so that the sounds overlapped. Due to a software glitch, sound 7 did not play at all. The program did play the remaining eleven (11) sounds (all at the medium volume used previously) in a random order and did record when the participant identified hearing their ping (by clicking a computer mouse).
The test was run in the following manner. The participant sat in the reverberation chamber and held a computer mouse (see Figure 2.13). The computer program would randomly select one of the twelve sounds to play. The computer, a Dell Optiplex GX110, soundcard was connected to five (5) Optimus Integrated Stereo Amplifiers (model SA-155). Each amplifier then split the signal two more times. A total of nine (9) of the array loudspeakers were used. All nine loudspeakers played at the same time creating a non-directional sound field relative to the participant seating location. Located at the rear of the chamber, directly in front of and facing away from the participant was a single JBL loudspeaker of the type previously mentioned. The JBL loudspeaker was connected via the Technics receiver to an Optiplex GX110 different from that running the LabView program. On the second computer, Windows Media Player was used to play another set of sounds in random order, minus the participant’s assigned ‘ping’. Since the second playlist did not contain the sound that the participant was listening for, it was assured that the only time the assigned sound was heard was if it had been played from the LabView program. A representation of the test equipment is shown in Figure 2.14.
Both computers began playing their respective playlists at the same time. When the participant heard his ‘ping’ played, he clicked the mouse, which recorded a response in the LabView program. Each participant listened for each of the sounds at a single volume level in separate tests. The volume levels for both playlists were set to be approximately equal, but no exact levels were recorded since it was not relevant to the test. The LabView program generated text files for each participant and each sound. These text files were then converted into Microsoft Excel spreadsheet files for analysis.
2.4 METHODS OF DATA COLLECTION

Collection of data was generated by faculty and students of WMU. Video tapes were made and examined to determine efficiency of subjects’ travel routes to the beacons during the navigation tests. In addition, a system of optical coordination was used to record the subjects’ deviation from the most ‘ideal’ path of travel. All video tapes were erased as soon as the information was encoded to protect the identity of participants involved. The other tests combined automated recording of data lists, by the computer programs, and hand recorded data, by the researchers. All data was collected using a coding system that did not directly identify the subjects. All data is reported in aggregate form.
3. DATA ANALYSIS

After each of the six tests, described in chapter two was concluded, the results were analyzed with the goal of developing a numerical score for each sound that would allow for comparison and ranking. Because each test was unique, the methods used to calculate the scores varied. In the end, the top 12 ranking sounds from each test were compared side-by-side to identify those that consistently ranked highest.

3.1 INITIAL SOUND SELECTION

A reduction of the original set of 156 sounds was necessary to allow for a more manageable group of sounds to be examined in the later tests. As was discussed in chapter 2.3.1, the data generated during this test phase was analyzed by Touch Graphics, which then selected the 12 sounds listed in Table 2.2 for further testing.

3.2 INDOOR NAVIGABILITY

To develop a ranking for navigability, a comparison of traveled areas was used. Analysis of the paths traveled by the participants was possible by utilizing the CAD drawings developed during the testing phase. First, the ‘ideal area’ of an ‘ideal path’ was measured as being one-foot on either side of a straight line between beacons for the length of the path leg - a leg being from one beacon to the next – unless the path included a slight obstacle, in which case the ‘ideal path’ was one-foot away from the obstacle (e.g., see Path/Sound 8 in Appendix D for such a situation). Second, the area outside of the ‘ideal path’ was calculated for the left and right sides (as one would walk the path) of each leg. Then the total area, for the left and right sides of each path, was calculated as
the sum of the individual (left or right side) leg areas. To find the areas, the AutoCAD area command was utilized. The command calculates the area bounded by a series of user-defined points. In this case, the points were the vertices of the participants’ travel envelope that lay outside of the ‘ideal path’. Since the first six paths contained three legs within room 2114 (Figure 2.3) and the second six paths contained four legs within room 2114, only the last three legs for each sound were considered. A composite area was also calculated as the sum of the left and right areas. Each path was then ranked based on the composite area.

Given that the paths were not of equal lengths, it was felt that using the composite area to rank the sounds would cause longer paths to rank lower than shorter paths. Therefore, the standard deviation of the area, of the last three legs, was calculated for the left and right sides, as well as a composite of both. A new composite rank was then calculated as the sum of the left and right standard deviations. The data are shown in Table 3.1, where the six highest composite ranked sounds are highlighted for each case. When the composite standard deviation of area rank was compared with the composite area rank, the results for the top 12 sounds closely agreed. The close agreement between the ranking of the sounds based on composite area and composite standard deviation of the areas argues against any concern that path length affected performance. A simple visual examination of the plots in Appendix D shows that short path legs did not necessarily produce less deviation from the ‘ideal path’ than long path legs, or vice versa.
<table>
<thead>
<tr>
<th>PATH/SOUND</th>
<th>AREA</th>
<th>TOTAL AREA</th>
<th>RANK (area)</th>
<th>STANDARD DEVIATION OF THE AREAS</th>
<th>RANK (std dev)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LEFT</td>
<td>RIGHT</td>
<td>LEFT</td>
<td>RIGHT</td>
<td>LEFT</td>
</tr>
<tr>
<td>1</td>
<td>4-13</td>
<td>0.77</td>
<td>0.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7-17</td>
<td>13.22</td>
<td>13.22</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7-9</td>
<td>0.12</td>
<td>1.64</td>
<td>1.64</td>
<td>1.64</td>
</tr>
<tr>
<td></td>
<td>4-13</td>
<td>15.15</td>
<td>2.54</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1-9</td>
<td>0.09</td>
<td>18.64</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>8-7</td>
<td>0.06</td>
<td>3.82</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>4-13</td>
<td>15.15</td>
<td>2.54</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1-9</td>
<td>0.09</td>
<td>18.64</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>8-7</td>
<td>0.06</td>
<td>3.82</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7-9</td>
<td>0.12</td>
<td>1.64</td>
<td>1.64</td>
<td>1.64</td>
</tr>
<tr>
<td>3</td>
<td>4-13</td>
<td>15.15</td>
<td>2.54</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1-9</td>
<td>0.09</td>
<td>18.64</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>8-7</td>
<td>0.06</td>
<td>3.82</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7-9</td>
<td>0.12</td>
<td>1.64</td>
<td>1.64</td>
<td>1.64</td>
</tr>
<tr>
<td>4</td>
<td>4-13</td>
<td>15.15</td>
<td>2.54</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1-9</td>
<td>0.09</td>
<td>18.64</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>8-7</td>
<td>0.06</td>
<td>3.82</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7-9</td>
<td>0.12</td>
<td>1.64</td>
<td>1.64</td>
<td>1.64</td>
</tr>
<tr>
<td>5</td>
<td>4-13</td>
<td>15.15</td>
<td>2.54</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1-9</td>
<td>0.09</td>
<td>18.64</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>8-7</td>
<td>0.06</td>
<td>3.82</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7-9</td>
<td>0.12</td>
<td>1.64</td>
<td>1.64</td>
<td>1.64</td>
</tr>
<tr>
<td>6</td>
<td>4-13</td>
<td>15.15</td>
<td>2.54</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1-9</td>
<td>0.09</td>
<td>18.64</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>8-7</td>
<td>0.06</td>
<td>3.82</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7-9</td>
<td>0.12</td>
<td>1.64</td>
<td>1.64</td>
<td>1.64</td>
</tr>
<tr>
<td>7</td>
<td>4-13</td>
<td>15.15</td>
<td>2.54</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1-9</td>
<td>0.09</td>
<td>18.64</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>8-7</td>
<td>0.06</td>
<td>3.82</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7-9</td>
<td>0.12</td>
<td>1.64</td>
<td>1.64</td>
<td>1.64</td>
</tr>
<tr>
<td>8</td>
<td>4-13</td>
<td>15.15</td>
<td>2.54</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1-9</td>
<td>0.09</td>
<td>18.64</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>8-7</td>
<td>0.06</td>
<td>3.82</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7-9</td>
<td>0.12</td>
<td>1.64</td>
<td>1.64</td>
<td>1.64</td>
</tr>
<tr>
<td>9</td>
<td>4-13</td>
<td>15.15</td>
<td>2.54</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1-9</td>
<td>0.09</td>
<td>18.64</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>8-7</td>
<td>0.06</td>
<td>3.82</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7-9</td>
<td>0.12</td>
<td>1.64</td>
<td>1.64</td>
<td>1.64</td>
</tr>
<tr>
<td>10</td>
<td>4-13</td>
<td>15.15</td>
<td>2.54</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1-9</td>
<td>0.09</td>
<td>18.64</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>8-7</td>
<td>0.06</td>
<td>3.82</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7-9</td>
<td>0.12</td>
<td>1.64</td>
<td>1.64</td>
<td>1.64</td>
</tr>
<tr>
<td>11</td>
<td>4-13</td>
<td>15.15</td>
<td>2.54</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1-9</td>
<td>0.09</td>
<td>18.64</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>8-7</td>
<td>0.06</td>
<td>3.82</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7-9</td>
<td>0.12</td>
<td>1.64</td>
<td>1.64</td>
<td>1.64</td>
</tr>
</tbody>
</table>

Table 3.1. Spreadsheet analysis of the navigability test. Pink highlighting indicates the six highest ranking sounds based on the area outside of the ‘ideal path’. Purple highlighting indicates the six highest ranking sounds based on the standard deviation of the area outside of the ‘ideal path’.
It is interesting to note, that in the majority of cases, users typically showed a distinct pattern to their movement: departure, correction, and arrival. Starting with their departure from a given location, the users would ‘ping’, or activate the system, and then begin traveling towards where they believed the next beacon was located, possibly ‘pinging’ several more times. At a point between beacons (often one-half to two-thirds the distance) users would stop walking, ‘ping’ again, correct their direction of travel, and continue towards the beacon possibly ‘pinging’ several more times along the way. A similar pattern is seen in Figure 1:9 and notably, participant FG in that study was blind. It appears that this is a common, and sensible, approach to navigation by sound for people with, or without, visual impairments. When the users arrived at where they believed the beacon to be located, in almost every case, they would ‘ping’ one final time and then feel for the beacon to assure themselves that they were in the correct location. While this pattern was not used to rank the sounds, it does show a common approach between participants in using the system. Furthermore, an examination of the paths in Appendix D shows that, although very long distances between beacons frequently required several corrections, all users did successfully reach their destinations.

3.3 SUBJECTIVE RESPONSE

Participants in the indoor navigability test were asked to score subjectively each sound they heard on a scale of 1-3, with one being low and three being high, for likeability. The scores given by the participants were averaged and constituted a ranking for the sound. The subjective responses are important because they relate to how comfortable the participants were in their ability to navigate using the different sounds.
For instance, one particular sound may have resulted in good navigation for a particular participant, but the participant himself may have felt unsure or unclear that he was on the right path, and thus, uncomfortable using the system.

3.4 INDOOR SOUND LOCALIZATION AND SOUND LEVEL

The scores for this test were based upon the number of ‘pings’ the participant required to be confident of the sound source location, the accuracy of the participant in localizing the sound source, and the sound level. The first steps were to calculate the average number of ‘pings’ required by the participants for the sounds they heard and their average localization accuracy (in feet). Then a raw score for each sound at each sound level was calculated as the sum of the two averages. Using the raw score the sounds were ranked based on sound pressure level (volume) in the categories of high volume, medium volume, and low volume (where a low score was better than a high score) and overall with respect to all sounds and levels.

In the museum environment, the PING system must not be disturbing to other patrons, hence, the lower the PING volume the better. In order to account for the differences in sound pressure levels between the sounds, the equivalent sound pressure level ($L_{eq}$) in dBA was measured for each sound over 20 seconds. As a means to penalize louder sounds, the raw score was multiplied by the $L_{eq}$ value to create a new weighted score. The weighted score was then used to create a weighted rank for each sound and level. Table 3.2 shows the analysis and ranking for the localization and level testing.
3.5 SOUND IDENTIFICATION AND RECOGNITION

Again, a composite score was developed. In this case, the total number of times the sound was played for all participants was found, from the LabView output files generated during the testing described in section 2.3.5, along with the total number of
correct, missed, and incorrect responses. A correct response was one to which a participant indicated a response when the desired sound was played, a missed response was one to which the participant did not respond when the desired sound was played, and an incorrect response was one to which the participant responded to a different sound.

The percent and correct responses, \( C \), and missed responses, \( M \), were calculated based on the total number of times the sound being tested was played for all participants,

\[
C = \left( \frac{n_C}{N} \right) \times 100 \\
M = \left( \frac{n_M}{N} \right) \times 100
\]

where, \( n_C \) is the total number of correct responses, \( n_M \) is the total number of missed responses, and \( N \) is the total number of times the sound being tested was played. The percent of incorrect responses, \( I \), was calculated based on the equation,

\[
I = \frac{n_I}{(t \cdot p - N)} \times 100
\]

where, \( n_I \) is the total number of incorrect responses, \( t \) is the total number of sounds played during an individual test, \( p \) is the number of participants to test the sound, and \( N \) is the total number of times the sound being tested was played. After calculating \( C, M, \) and \( I \) the composite score, \( K \), for each sound was found from Equation 3.4. An example of one calculation is shown in Table 3.3, all of the calculations for the identification and recognition test are located in Appendix E.

\[
K = C - M - I
\]
3.6 COMPARISON OF DATA

Once the sounds were ranked in each test, the ranks were compared to one another to determine which sounds consistently ranked highest. Table 3.4 shows the sounds ordered by rank for each test. The blue area identifies the top six ranked sounds in each test and the yellow area identifies the bottom six sounds as ranked. Next, the top six highest-ranking sounds from each test were compared to see which ranked in multiple tests. The comparison is shown in Table 3.5.

<table>
<thead>
<tr>
<th>SUBJECT 1</th>
<th>SUBJECT 2</th>
<th>SUBJECT 3</th>
<th>SUBJECT 4</th>
</tr>
</thead>
<tbody>
<tr>
<td># PLAYED</td>
<td>7</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td># CORRECT</td>
<td>7</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td># MISSED</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td># INCORRECT</td>
<td>8</td>
<td>7</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th># PLAYED</th>
<th># CORRECT</th>
<th># MISSED</th>
<th># INCORRECT</th>
<th>COMPOSITE</th>
<th>SCORE</th>
</tr>
</thead>
<tbody>
<tr>
<td>23</td>
<td>21</td>
<td>2</td>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% CORRECT</td>
<td>% MISSED</td>
<td>% INCORRECT</td>
<td></td>
<td>73.22</td>
<td></td>
</tr>
<tr>
<td>91.30</td>
<td>8.70</td>
<td>9.90</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3.3. Example of the analysis of the sound identification and recognition test.

Table 3.4. Comparison of sound rankings. The top six sounds based on rank are shown in the blue area. The six bottom ranked sounds are shown in the yellow area.
3.7 TIME- AND FREQUENCY-DOMAIN DATA

As an extension of the originally planned physical testing, the sounds were examined and compared within their time- and frequency-domains. A MatLab program (see Appendix F) was written that would plot the magnitude (on a relative scale) of the selected sound versus time. It also computed the FFT of the sound and displayed a spectrogram plot for the sound and an octave-band bar graph. The axes were scaled equally for all sounds to allow for easier comparison. Using a 1000 Hz sine wave as the input signal the program was verified to produce accurate results as shown in Figure 3.1. Utilizing Adobe Photoshop’s layer properties, plots of different sounds could be overlaid to make comparisons and observe any trends within the data. Comparisons were grouped according to the four tests – navigability, localization and level, recognition, and subjective response – and the group of recommended sounds. Each grouping was split into the top performing and bottom performing sounds. This allowed each sound to be examined based on test type, rank, and overall performance. The comparative plots are

<table>
<thead>
<tr>
<th>TOP SIX SOUNDS</th>
<th>Navigability</th>
<th>Localization &amp; Level</th>
<th>Subjective</th>
<th>Recognition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td>5</td>
<td>5 M</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>5 M</td>
<td>6</td>
<td>6 M, 6 L</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td>9</td>
<td>9 L</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td>10</td>
<td>10 L</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td>11</td>
<td>11 M</td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
<td>12</td>
<td></td>
</tr>
</tbody>
</table>

Table 3.5. Comparison of the top six sounds by test.
included as part of Appendix G. It was hoped that trends between sounds within a given test, or overall, would emerge from the plots that would distinguish ‘good’ sounds from ‘bad’ sounds.

3.7.1 TIME-DOMAIN CHARACTERISTICS

Sounds that ranked within the top six typically showed one of two characteristics in the time-domain signal. First, a repeat of the signal, such as seen in sound 5 in Figure 3.2, and second, a signal structure that will be described as ‘winged’, like that of sound 1 in Figure 3.3 However, these time signals are not necessarily indicative of a ‘good’
sound since some sounds with these characteristics ranked poorly in some tests. For instance, sound 1 ranked well in the navigation and subjective tests, but poorly in the recognition test.

Figure 3.2. Example of a repeated time-domain signal (sound 5).

Figure 3.3. Example of a ‘winged’ time-domain signal (sound 1).
3.7.2 FREQUENCY-DOMAIN CHARACTERISTICS

Looking at the frequency content of the tested sounds is limited by the FFT, which is affected by the sampling rate of the initial recording. In some cases, the sampling rate was only 8000 Hz, which limits the useful FFT data to roughly 3200 Hz. All sounds were limited in frequency to 10000 Hz. Therefore, the effect of high frequency content cannot be judged. Using the information available, all of the tested sounds, like sound 1 in Figure 3.4, have the greatest SPL in the 500 – 2000 Hz octave-bands. This is consistent with the work of Rakerd, Hartmann and McCaskey [1999] who concluded that recognition in the median sagittal plane was best for sounds low frequency content (1-3 kHz). It cannot be said with certainty, however, that this is the cause of accurate recognition in the horizontal plane. Unfortunately, neither the octave-band nor the spectrograms showed any distinctive differences between ‘good’ and ‘bad’ sounds.
Figure 3.4. Example plot showing high SPLs in the 500 and 1000 Hz bands (sound 5).
4. CONCLUSIONS REGARDING THE CURRENT RESEARCH

As was stated previously, the ultimate goal of this research was threefold: to develop a series of tests to be used in identifying ‘good’ attractor sounds based on navigability, localization, recognition, and likeability; second, to use these tests to recommend a collection of sounds for the PING library; third, to identify any common characteristics that could aid in selecting future sounds for the library. As the background information has shown, there has been very little research that examines a sound’s characteristics for the combined purposes of navigation, localization, and recognition – especially for ‘real-world’ sounds. However, by combining aspects of the tests that have been previously conducted by other researchers, a series of physical tests - tests requiring the physical participation of volunteers - were developed that are consistent in design. Data collection during physical testing and its subsequent analysis resulted in a ranking system that identifies ‘good’ sounds from within a larger group.

The small number of participants involved in the study limits the statistical significance of the measurements made during the experiments. That being said, the tests were approached with this in mind and provisions were made to reduce error. This does make it more difficult to identify trends within the data that may identify characteristics of ‘good’ versus ‘bad’ sounds as applied to the PING system.

Examination of the time and frequency data for the 12 tested sounds reveals very little, if any, differences between ‘good’ and ‘bad’ sounds. There are two possible explanations for this: first, the initial reduction from 175 sounds to 12 for testing was sufficient to eliminate sounds that would have performed extremely poorly; second, all of the originally selected sounds were ‘good’ to begin with.
One trend that stands out amongst the majority of sounds (‘good’ and ‘bad’) is a concentration of intensity within the 500 Hz to 2000 Hz octave bands. The importance of low-frequency content for vertical localization [Blauert, 1969/70] and recognition within the median sagittal plane [Rakerd, Hartmann, & McCaskey, 1999] has been reported previously, but not for the horizontal plane. Since very few of the tested sounds exhibited frequency content above 10 kHz, it is believed (based on the data available) that the low frequencies are also essential to localization, recognition, and navigation in the horizontal plane for reverberant environments when head movements are allowed. Head movements (dynamic cues) are thought to be important in resolving the front/back confusions that can arise because of the influence of these frequencies in the head-related transfer function. While the work of Middlebrooks and Green [1991] concluded head movements were “probably not a critical part of the localization process” [pg. 19], the length of sounds tested here was much greater than the one-second used in their study and therefore “permits a very detailed assessment of location” [pg. 19] required for dynamic cues to be effective.

Another trend is that many of the top ranked sounds had quick onsets and quick decays. The benefit of a rapid onset in a reverberant environment was shown by Rakerd and Hartmann [1986] and participants reported during this research that sounds with a shorter decay tended to cause fewer reflections from surrounding surfaces and, thus, were easier to localize. The navigation experiments conducted by Loomis, Hebert, and Cicinelli [1990] also used a sound with a rapid onset and decay (a square wave) with positive results.
The research presented here showed a close agreement between sounds that were ranked highly in localization and recognition. Previous work found that correct localization does not necessarily result in correct recognition in the vertical plane [Rakerd, Hartmann & McCaskey, 1999]. The current work does not invalidate that conclusion, but does raise the question of whether correct localization results in correct recognition (or vice versa) in the horizontal plane. However, this question was not addressed in the design of the current experiments.

After having completed all tests and analysis, and using the information in Tables 3.4 and 3.5, seven sounds were identified as suitable for addition to the PING catalog. Although all twelve sounds performed well, the seven selected did show slightly better performance than the other five. The selected sounds can be organized into three groups based on their rate of recurrence of high ranks. Group A includes sounds that ranked within the top six in all three objective experiments (sounds 5, 6, and 10). Group B includes sounds that ranked within the top six in the subjective experiment and multiple objective experiments (sound 3). Sounds that ranked within the top six in any two tests are included in Group C (sounds 1, 9, and 11). Sound 7 was not included in the final selection, despite its high ranking in some tests, because all researchers involved and most participants found it to be exceedingly unpleasant. Therefore, it is recommended that the sounds identified in Table 4.1 be added to the PING catalog at the volume levels indicated. Three sounds are birdcalls, one is a familiar tune from television (the Jetson’s doorbell), and three are some other type of bell.

The research discussed here has been applied by Touch Graphics to the latest version of the PING audio navigation system. Trials of the PING system that incorporate
the recommended sounds will begin in the fall of 2006 at the New York Hall of Science, a hands-on science museum in New York City.

Table 4.1. Recommended sound numbers and levels for addition to the PING catalog.
5. RECOMMENDATIONS FOR FUTURE RESEARCH

One of the greatest limitations of this research involved the amount of time needed to conduct the physical testing and the availability of volunteers to participate in the tests. It was hoped that analysis of the data would lead to clear trends that would distinguish the ‘good’ from the ‘bad’ sounds and that the information could be used in the future to better select sounds at the outset, thus reducing the amount of physical testing required. That was not the case, although some trends within the data were observed. Future research should consider three areas to address this issue: examination of the ‘winged’ time-domain structure; investigation into frequency characteristics within the 500 – 2000 Hz octave bands; and the relationship between localization and recognition.

Many of the top ranked sounds showed a ‘winged’ time domain signal like that of sound number one, part of the time-domain of which is shown in Figure 3.3. This signal shape could be an indicator of performance that could reduce the amount of physical testing required of the sound. While sounds in this study showed a similarity in shape, they were not identical. The current work did not look at this structure in detail (time envelopes of less than one second). Such an analysis may reveal information (rise/decay, time/rate, amplitude, etc.) that could allow the synthesis of attractor sounds in the future.

Most sounds examined here contained strong frequency content within the 500 – 2000 Hz bands. The small number of sounds tested and participants, however, limits the statistical significance of the information that can be gained. It is risky to draw broad conclusions based on the spectrum of such a small sample set. An examination focusing on the frequency structure within the aforementioned bands of a larger number sounds
with a larger number of participants could yield highly useful information for the selection of attractor sounds in the future.

The current study showed a strong, positive relationship between localization and recognition in the horizontal plane. In other words, sounds that ranked well for localization also tended to rank well for recognition. Five of the top six ranked sounds were the same for each of the two tests. Such a relationship has not been previously reported in the literature. This may render one test or the other (localization or recognition) unnecessary, but further tests are needed with a larger sample set before that decision can be made.

Finally, it would be desirable to develop a model similar to those discussed in Part VII of *Binaural and Spatial Hearing in Real and Virtual Environments* by Gilkey and Anderson [1997]. Such a model could be used to develop a computer program that could be used to predict accurately the performance of a test sound, without the need for human participants. A similar program would save significant time compared to physical testing. To create such a program will require a better understanding of the processes of localization, recognition, and navigability of ‘real-world’ sounds that this work has begun to explore.
REFERENCES


Guettler, D. L., Bolia, R.S. & Nelson, W. T. (unpublished draft). Multiple sound source identification project: Research plan. RAIVE lab, Wright-Patterson AFB. Note: This paper is a draft copy of a research plan used in the year 2000 work by the authors cited above. It was generously provided by Ms. Guettler.


APPENDIX A

HSIRB RESEARCH PROTOCOL CLEARANCE
Date: July 28, 2004

To: William Wiener, Principal Investigator  
Koorosh Nagshineh, Student Investigator

From: Amy Naugle, Ph.D., Interim Chair

Re: HSIRB Project Number: 04-07-15

This letter will serve as confirmation that your research project entitled “Ping: The Location of Museum Exhibits” has been approved under the expedited category of review by the Human Subjects Institutional Review Board. The conditions and duration of this approval are specified in the Policies of Western Michigan University. You may now begin to implement the research as described in the application.

Please note that you may only conduct this research exactly in the form it was approved. You must seek specific board approval for any changes in this project. You must also seek reapproval if the project extends beyond the termination date noted below. In addition if there are any unanticipated adverse reactions or unanticipated events associated with the conduct of this research, you should immediately suspend the project and contact the Chair of the HSIRB for consultation.

The Board wishes you success in the pursuit of your research goals.

Approval Termination: July 28, 2005
APPENDIX B

SOUND NORMALIZATION PROGRAM (MATLAB)
This program will normalize the input sound based on peak amplitude and the user specified scale factor.

```matlab
clear
FileToLoad=input('Enter filename ... ','s');
FileToLoad1=strcat(FileToLoad,'.wav');
ReadinCommand=['[data,fs,nbits]=wavread(''',FileToLoad1,''');'];
eval(ReadinCommand);
factor=max([abs(max(data)),abs(min(data))])
sound(data,fs);
pause
ScaleFactor=0.5;
sound((ScaleFactor/factor)*data,fs)
SaveoutCommand=['wavwrite((.5/factor)*data,fs,nbits,"',FileToLoad,'2.wav");'];
eval(SaveoutCommand);
```
APPENDIX C

PRELIMINARY TEST DATA
LEGEND:
Q1 - corresponds to the average pleasantness response
Q2 - corresponds to the average loudspeaker response (localization)
Angular – average error (number of speaker locations)
Radial – average front/back (+/-) error (speaker row location)
Unmarked, hand written – Touch Graphics notes

<table>
<thead>
<tr>
<th>Block 1</th>
<th>Q1</th>
<th>Q2</th>
<th>Angular</th>
<th>Radial</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>owl2_edit2_amplitude.wav</td>
<td>1 12</td>
<td>0 0</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>jetson_edit2_pitch2.wav</td>
<td>2 2</td>
<td>0 0</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>E5-broadwingedhawk2.wav</td>
<td>2 4</td>
<td>0 -1</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>bird_edit2_repeat.wav</td>
<td>3 15</td>
<td>0 0</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>clang1_edit2_amplitude.wav</td>
<td>2 10</td>
<td>1 1</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>l2-horn2_amplitude.wav</td>
<td>2 1</td>
<td>0 0</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>bell1_edit2_tempo2.wav</td>
<td>2 1</td>
<td>0 -2</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>bell2_edit2_tempo2.wav</td>
<td>2 1</td>
<td>0 -2</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>bell1_edit2_repeat.wav</td>
<td>2 1</td>
<td>0 -2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>F6-canadagoose2.wav</td>
<td>2 3</td>
<td>0 0</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>bell1_edit2.wav</td>
<td>2 1</td>
<td>0 0</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>bell2_edit2_amplitude.wav</td>
<td>2 1</td>
<td>0 0</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>bird_edit2.wav</td>
<td>2 1</td>
<td>0 0</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>bell1_edit2_amplitude.wav</td>
<td>2 8</td>
<td>2 0</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Block 2</th>
<th>Q1</th>
<th>Q2</th>
<th>Angular</th>
<th>Radial</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>s19-ping32_pitch2.wav</td>
<td>2 15</td>
<td>0 1</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>r18-ping12_tempo2.wav</td>
<td>1 7</td>
<td>1 1</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>r18-ping12_tempohalf.wav</td>
<td>1 9</td>
<td>0 0</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>u21-ping52_repeat.wav</td>
<td>2 15</td>
<td>0 1</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>clang2_edit2_repeat.wav</td>
<td>2 9</td>
<td>0 0</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>p16-operation2.wav</td>
<td>1 8</td>
<td>1 1</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>clang2_edit2_tempohalf.wav</td>
<td>2 7</td>
<td>0 -1</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>u21-ping52_pitchhalf.wav</td>
<td>2 14</td>
<td>0 0</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>hawk_edit2_amplitude.wav</td>
<td>2 2</td>
<td>1 -1</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>clang1_edit2_pitch2.wav</td>
<td>1 8</td>
<td>0 -1</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>bell2_edit2_pitch2.wav</td>
<td>2 15</td>
<td>0 2</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>jetson_edit2_amplitude.wav</td>
<td>3 7</td>
<td>0 0</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>s19-ping32_tempo2.wav</td>
<td>1 12</td>
<td>1 2</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>bell1_edit2_pitch2.wav</td>
<td>1 8</td>
<td>0 0</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>clang1_edit2_tempo2.wav</td>
<td>1 2</td>
<td>1 -2</td>
<td></td>
</tr>
</tbody>
</table>

C1. Preliminary test data sheet one.
C3. Preliminary test data sheet three.
APPENDIX D

TRAVEL ENVELOPES OF PARTICIPANTS IN THE NAVIGATION TEST
D1. Path/Sound 1

Path/Sound 1

- Speaker Location
- PING Beacon Location
- Travel Envelope
- Straight-line "ideal" Path
- "Ideal" Path Width
Path/Sound 5
D9. Path / Sound 9

Path/Sound 9

- Speaker Location
- PING Beacon Location
- Travel Envelope
- Straight-line 'Ideal' Path
- 'Ideal' Path Width
D10. Path / Sound 10
APPENDIX E

SOUND IDENTIFICATION AND RECOGNITION ANALYSIS TABLES
### E1. Sound 1

<table>
<thead>
<tr>
<th></th>
<th>SUBJECT 1</th>
<th>SUBJECT 2</th>
<th>SUBJECT 3</th>
<th>SUBJECT 4</th>
</tr>
</thead>
<tbody>
<tr>
<td># PLAYED</td>
<td>10</td>
<td>6</td>
<td>13</td>
<td>10</td>
</tr>
<tr>
<td># CORRECT</td>
<td>10</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td># MISSED</td>
<td>0</td>
<td>3</td>
<td>9</td>
<td>5</td>
</tr>
<tr>
<td># INCORRECT</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th># PLAYED</th>
<th># CORRECT</th>
<th># MISSED</th>
<th># INCORRECT</th>
</tr>
</thead>
<tbody>
<tr>
<td>39</td>
<td>22</td>
<td>17</td>
<td>7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>% CORRECT</th>
<th>% MISSED</th>
<th>% INCORRECT</th>
</tr>
</thead>
<tbody>
<tr>
<td>56.41</td>
<td>43.59</td>
<td>3.55</td>
<td></td>
</tr>
</tbody>
</table>

**SCORE**

**COMPOSITE**

9.27

### E2. Sound 2

<table>
<thead>
<tr>
<th></th>
<th>SUBJECT 1</th>
<th>SUBJECT 2</th>
<th>SUBJECT 3</th>
<th>SUBJECT 4</th>
</tr>
</thead>
<tbody>
<tr>
<td># PLAYED</td>
<td>3</td>
<td>8</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td># CORRECT</td>
<td>3</td>
<td>7</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td># MISSED</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td># INCORRECT</td>
<td>2</td>
<td>1</td>
<td>5</td>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th># PLAYED</th>
<th># CORRECT</th>
<th># MISSED</th>
<th># INCORRECT</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>14</td>
<td>6</td>
<td>11</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>% CORRECT</th>
<th>% MISSED</th>
<th>% INCORRECT</th>
</tr>
</thead>
<tbody>
<tr>
<td>70.00</td>
<td>30.00</td>
<td>5.09</td>
<td></td>
</tr>
</tbody>
</table>

**SCORE**

**COMPOSITE**

34.91
<table>
<thead>
<tr>
<th></th>
<th>SUBJECT 1</th>
<th>SUBJECT 2</th>
<th>SUBJECT 3</th>
<th>SUBJECT 4</th>
</tr>
</thead>
<tbody>
<tr>
<td># PLAYED</td>
<td>7</td>
<td>5</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td># CORRECT</td>
<td>7</td>
<td>4</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td># MISSED</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td># INCORRECT</td>
<td>8</td>
<td>7</td>
<td>1</td>
<td>4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th># PLAYED</th>
<th># CORRECT</th>
<th># MISSED</th>
<th># INCORRECT</th>
<th>COMPOSITE</th>
</tr>
</thead>
<tbody>
<tr>
<td>23</td>
<td>21</td>
<td>2</td>
<td>20</td>
<td>SCORE</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>% CORRECT</th>
<th>% MISSED</th>
<th>% INCORRECT</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>91.30</td>
<td>8.70</td>
<td>9.39</td>
<td>73.22</td>
</tr>
</tbody>
</table>

E3. Sound 3

<table>
<thead>
<tr>
<th></th>
<th>SUBJECT 1</th>
<th>SUBJECT 2</th>
<th>SUBJECT 3</th>
<th>SUBJECT 4</th>
</tr>
</thead>
<tbody>
<tr>
<td># PLAYED</td>
<td>1</td>
<td>6</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td># CORRECT</td>
<td>1</td>
<td>6</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td># MISSED</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td># INCORRECT</td>
<td>9</td>
<td>6</td>
<td>4</td>
<td>9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th># PLAYED</th>
<th># CORRECT</th>
<th># MISSED</th>
<th># INCORRECT</th>
<th>COMPOSITE</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>15</td>
<td>5</td>
<td>28</td>
<td>SCORE</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>% CORRECT</th>
<th>% MISSED</th>
<th>% INCORRECT</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>75.00</td>
<td>25.00</td>
<td>12.96</td>
<td>37.04</td>
</tr>
</tbody>
</table>

E4. Sound 4
<table>
<thead>
<tr>
<th>SUBJECT 1</th>
<th>SUBJECT 2</th>
<th>SUBJECT 3</th>
<th>SUBJECT 4</th>
</tr>
</thead>
<tbody>
<tr>
<td># PLAYED</td>
<td>8</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td># CORRECT</td>
<td>7</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td># MISSED</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td># INCORRECT</td>
<td>5</td>
<td>6</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th># PLAYED</th>
<th># CORRECT</th>
<th># MISSED</th>
<th># INCORRECT</th>
<th>COMPOSITE</th>
</tr>
</thead>
<tbody>
<tr>
<td>19</td>
<td>17</td>
<td>2</td>
<td>17</td>
<td>SCORE</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>% CORRECT</th>
<th>% MISSED</th>
<th>% INCORRECT</th>
<th>COMPOSITE</th>
</tr>
</thead>
<tbody>
<tr>
<td>89.47</td>
<td>10.53</td>
<td>7.83</td>
<td>71.11</td>
</tr>
</tbody>
</table>

E5. Sound 5

<table>
<thead>
<tr>
<th>SUBJECT 1</th>
<th>SUBJECT 2</th>
<th>SUBJECT 3</th>
<th>SUBJECT 4</th>
</tr>
</thead>
<tbody>
<tr>
<td># PLAYED</td>
<td>10</td>
<td>14</td>
<td>8</td>
</tr>
<tr>
<td># CORRECT</td>
<td>10</td>
<td>13</td>
<td>2</td>
</tr>
<tr>
<td># MISSED</td>
<td>0</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td># INCORRECT</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th># PLAYED</th>
<th># CORRECT</th>
<th># MISSED</th>
<th># INCORRECT</th>
<th>COMPOSITE</th>
</tr>
</thead>
<tbody>
<tr>
<td>40</td>
<td>33</td>
<td>7</td>
<td>3</td>
<td>SCORE</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>% CORRECT</th>
<th>% MISSED</th>
<th>% INCORRECT</th>
<th>COMPOSITE</th>
</tr>
</thead>
<tbody>
<tr>
<td>82.50</td>
<td>17.50</td>
<td>1.53</td>
<td>63.47</td>
</tr>
</tbody>
</table>

E6. Sound 6
E7. Sound 7

E8. Sound 8
<table>
<thead>
<tr>
<th></th>
<th>SUBJECT 1</th>
<th>SUBJECT 2</th>
<th>SUBJECT 3</th>
<th>SUBJECT 4</th>
</tr>
</thead>
<tbody>
<tr>
<td># PLAYED</td>
<td>4</td>
<td>6</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td># CORRECT</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td># MISSED</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td># INCORRECT</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th># PLAYED</th>
<th># CORRECT</th>
<th># MISSED</th>
<th># INCORRECT</th>
<th>COMPOSITE SCORE</th>
</tr>
</thead>
<tbody>
<tr>
<td>21</td>
<td>18</td>
<td>3</td>
<td>6</td>
<td>68.64</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>% CORRECT</th>
<th>% MISSED</th>
<th>% INCORRECT</th>
</tr>
</thead>
<tbody>
<tr>
<td>85.71</td>
<td>14.29</td>
<td>2.79</td>
</tr>
</tbody>
</table>

E9. Sound 9

<table>
<thead>
<tr>
<th></th>
<th>SUBJECT 1</th>
<th>SUBJECT 2</th>
<th>SUBJECT 3</th>
<th>SUBJECT 4</th>
</tr>
</thead>
<tbody>
<tr>
<td># PLAYED</td>
<td>4</td>
<td>3</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td># CORRECT</td>
<td>4</td>
<td>3</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td># MISSED</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td># INCORRECT</td>
<td>5</td>
<td>2</td>
<td>10</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th># PLAYED</th>
<th># CORRECT</th>
<th># MISSED</th>
<th># INCORRECT</th>
<th>COMPOSITE SCORE</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>11</td>
<td>0</td>
<td>17</td>
<td>92.44</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>% CORRECT</th>
<th>% MISSED</th>
<th>% INCORRECT</th>
</tr>
</thead>
<tbody>
<tr>
<td>100.00</td>
<td>0.00</td>
<td>7.56</td>
</tr>
</tbody>
</table>

E10. Sound 10
### E11. Sound 11

<table>
<thead>
<tr>
<th></th>
<th>SUBJECT 1</th>
<th>SUBJECT 2</th>
<th>SUBJECT 3</th>
<th>SUBJECT 4</th>
</tr>
</thead>
<tbody>
<tr>
<td># PLAYED</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td># CORRECT</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td># MISSED</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td># INCORRECT</td>
<td>0</td>
<td>5</td>
<td>2</td>
<td>5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>COMPOSITE</th>
<th>SCORE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>COMPOSITE</td>
<td>SCORE</td>
</tr>
</tbody>
</table>

### E12. Sound 12

<table>
<thead>
<tr>
<th></th>
<th>SUBJECT 1</th>
<th>SUBJECT 2</th>
<th>SUBJECT 3</th>
<th>SUBJECT 4</th>
</tr>
</thead>
<tbody>
<tr>
<td># PLAYED</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>12</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>COMPOSITE</th>
<th>SCORE</th>
<th></th>
<th>COMPOSITE</th>
<th>SCORE</th>
</tr>
</thead>
</table>

### Score Details

- **E11. Sound 11**
  - Correct: 22, Missed: 5, Incorrect: 1
  - Score: 62.48%

- **E12. Sound 12**
  - Correct: 1, Missed: 2, Incorrect: 12
  - Score: -38.48%
APPENDIX F

PROGRAM FOR DATA ANALYSIS (MATLAB)
clear all

% ++++ START .WAV FILE LOAD ++++ %
FileToLoad=input('Enter filename... ','s');
FileToLoad1=strcat(FileToLoad,'.wav');
ReadinCommand=['[data,fs,nbits]=wavread(''',FileToLoad1,''');'];
eval(ReadinCommand);
% ---- END .WAV FILE LOAD ---- %

% ++++ START TIME DOMAIN PLOT ++++ %
time in ten-thousandth of a second to plot (5 seconds)
t1 = 5e4;
T1 = 1:t1;
T1 = T1/10000;

% calculate number of 0 entries to make sound 5 sec
number_x = numel(data);
needed_x = t1 - number_x;
if needed_x >= 0;
    Z = zeros(needed_x,1);
mag_time = [data; Z];
else
    for ab = 1:number_x+needed_x;
        mag_time = [data(1:ab)];
    end
end

% magnitude of sound in time domain
figure(1)
plot(T1,mag_time);
axis([0 max(T1) -0.8 0.8]);
whitebg('w')
title(['Linear Time-Magnitude Plot of ',eval(''FileToLoad''')]);
xlabel('Time (sec)');
ylabel('Magnitude');
grid on;
% pause
% ---- END TIME DOMAIN PLOT ---- %

% ++++ START FFT PLOTTING ++++ %
% Following code modified from Ch. 6.2 of
% http://www.utexas.edu/math/Matlab/tec6.2.html

Fn=fs/2;                         % Nyquist frequency
x = mag_time;                    % inputed data
NFFT=2.^(ceil(log(length(x))/log(2)))); % Next highest power of 2
% greater than length(x).
FFTX=fft(x,NFFT);                % Take FFT, padding with zeros.
% length(FFTX)==NFFT
NumUniquePts = ceil((NFFT+1)/2);
FFTX=FFTX(1:NumUniquePts);            % FFT is symmetric, throw away % second half
MX=abs(FFTX);
MX=MX*2;                              % Take magnitude of X
% Multiply by 2 to take into % account the fact that we % threw out second half of %FFTX above
MX(1)=MX(1)/2;                        % Account for endpoint % uniqueness
MX(length(MX))=MX(length(MX))/2;      % We know NFFT is even % Scale the FFT so that it is % not a function of the length % of x.

% Setting fixed frequency range 0-10000 Hz based on sample freq.
deltaf = fs/NFFT;
needed_mx = 10000/deltaf - numel(MX);
for ac = 1:10000/deltaf
    f(ac) = deltaf*(ac-1);
end
if needed_mx >= 0;
    Z2 = zeros(needed_mx, 1);
    mx = [MX; Z2];
else
    for ac = 1:numel(MX)+needed_mx;
        mx = [MX(1:ac)];
    end
end

% plot fft on linear scale
figure(2)
plot(f,mx);
axis([0 10000 0 0.02]);
grid on
whitebg('w')
title(['Linear Frequency-Magnitude Plot of ' eval('FileToLoad')]);
xlabel('Frequency (Hz)');
ylabel('Magnitude');
% pause

% plot fft in dB SPL
if needed_mx >= 0;
    dBSPL_freq = 20*log10(MX/(2e-5));
    dBSPL_freq = [dBSPL_freq; Z2];
else
    dBSPL_freq = 20*log10(mx/(2e-5));
end

figure(3)
plot(f,dBSPL_freq);
axis([0 10000 -100 60]);
grid on
whitebg('w')
title(['Frequency-Magnitude Plot of ' eval('FileToLoad')]);
xlabel('Frequency (Hz)');
ylabel('Magnitude (dB SPL)');
grid on;
% pause
%----- END FFT PLOT ---- %
% Waterfall plots do not provide useful info above what can be found in
% the spectrogram

++++ START SPECTROGRAM ++++
% requires function "spectrogram.m"

bw = 20;

figure(8)
spectrogram2(mag_time, fs, bw, T1);
axis([0 max(T1) 0 10]);
grid on
title(['Spectrogram of ', eval('FileToLoad')]);

% pause

++++ START 1/1-OCTAVE ANALYSIS ++++
% define cutoff & center frequencies

cutoff = [0 22.4 45 90 180 355 710 1400 2800 5600 10000];
center = [31.5; 63; 125; 250; 500; 1000; 2000; 4000; 8000];

% calculate data in each octave band, put into new matrices

m1 = mx;
while numel(m1) > ceil(cutoff(2)/deltaf)
    m1(numel(m1)) = [];
end

m2 = mx;
while numel(m2) > ceil(cutoff(3)/deltaf)
    m2(numel(m2)) = [];
end
    while numel(m2) > (ceil(cutoff(3)/deltaf) - ceil(cutoff(2)/deltaf))
        m2(1) = [];
end

m3 = mx;
while numel(m3) > ceil(cutoff(4)/deltaf)
    m3(numel(m3)) = [];
end
    while numel(m3) > (ceil(cutoff(4)/deltaf) - ceil(cutoff(3)/deltaf))
        m3(1) = [];
end

m4 = mx;
while numel(m4) > ceil(cutoff(5)/deltaf)
    m4(numel(m4)) = [];
end
    while numel(m4) > (ceil(cutoff(5)/deltaf) - ceil(cutoff(4)/deltaf))
        m4(1) = [];
end
m5 = mx;
while numel(m5) > ceil(cutoff(6)/deltaf)
    m5(numel(m5)) = [];
end
while numel(m5) > (ceil(cutoff(6)/deltaf) - ceil(cutoff(5)/deltaf))
    m5(1) = [];
end
m6 = mx;
while numel(m6) > ceil(cutoff(7)/deltaf)
    m6(numel(m6)) = [];
end
while numel(m6) > (ceil(cutoff(7)/deltaf) - ceil(cutoff(6)/deltaf))
    m6(1) = [];
end
m7 = mx;
while numel(m7) > ceil(cutoff(8)/deltaf)
    m7(numel(m7)) = [];
end
while numel(m7) > (ceil(cutoff(8)/deltaf) - ceil(cutoff(7)/deltaf))
    m7(1) = [];
end
m8 = mx;
while numel(m8) > ceil(cutoff(9)/deltaf)
    m8(numel(m8)) = [];
end
while numel(m8) > (ceil(cutoff(9)/deltaf) - ceil(cutoff(8)/deltaf))
    m8(1) = [];
end
m9 = mx;
while numel(m9) > ceil(cutoff(10)/deltaf)
    m9(numel(m9)) = [];
end
while numel(m9) > (floor(cutoff(10)/deltaf) - ceil(cutoff(9)/deltaf))
    m9(1) = [];
end
m10 = mx;
while numel(m10) > floor((cutoff(11) - cutoff(10))/deltaf)
    m10(1) = [];
end
m = [m1; m2; m3; m4; m5; m6; m7; m8; m9; m10];

% calculate band values
%sum1 = (sum(m1.^2))/((2e-5)^2);
sum2 = (sum(m2.^2))/((2e-5)^2);
sum3 = (sum(m3.^2))/((2e-5)^2);
sum4 = (sum(m4.^2))/((2e-5)^2);
sum5 = (sum(m5.^2))/((2e-5)^2);
sum6 = (sum(m6.^2))/((2e-5)^2);
sum7 = (sum(m7.^2))/((2e-5)^2);
sum8 = (sum(m8.^2))/((2e-5)^2);
sum9 = (sum(m9.^2))/((2e-5)^2);
sum10 = (sum(m10.^2))/((2e-5)^2);
%dBSPL_b1 = 10*log10(sum1);
dBSPL_b2 = 10*log10(sum2);
dBSPL_b3 = 10*log10(sum3);
dBSPL_b4 = 10*log10(sum4);
dBSPL_b5 = 10*log10(sum5);
dBSPL_b6 = 10*log10(sum6);
dBSPL_b7 = 10*log10(sum7);
dBSPL_b8 = 10*log10(sum8);
dBSPL_b9 = 10*log10(sum9);
if sum10 == 0
    dBSPL_b10 = 0;
else
    dBSPL_b10 = 10*log10(sum10);
end

% pause

band = [dBSPL_b2; dBSPL_b3; dBSPL_b4; dBSPL_b5; dBSPL_b6;...
        dBSPL_b7; dBSPL_b8; dBSPL_b9; dBSPL_b10];

% create plot
figure9 = figure;
%whitebg('w')
axes1 = axes(...
    'XTick', [1 2 3 4 5 6 7 8 9], ...
    'XTickLabel', {'31.5', '63', '125', '250', '500', '1000', '2000', '4000', ...
                  '8000'}, ...
    'YGrid', 'on', 'Parent', figure9);
ylim(axes1, [-20 80]);
box(axes1, 'on');
hold(axes1, 'all');
title(['Octave-Band Plot of ', eval('FileToLoad')]);
xlabel(' Center Frequency (Hz)');
ylabel(' SPL (dB)');
bar1 = bar(band);

%---- END 1/1-OCTAVE ANALYSIS ----%

function spectrogram2(x,fs,bw,T1)
% Plot spectrogram with variable sampling rate and bandwidth
% Modified from the code written by Evan Ruzanski, ECE259, 2/26/2003

% Set minimum FFT length
fftmin = 256;

% Set 2*fs/bw variable window length for good resolution with long block length
% will provide higher frequency resolution as main-lobe of the window
% function will be narrow and short block length will provide higher time resolution as less averaging across samples is performed for each
% STFT value
winlen = floor(2*fs/bw);

% Get FFT length
ftflen = max([winlen fftmin]);
% Create window (Hamming for favorable sidelobe attenuation) and zero pad accordingly
win = [hamming(winlen) ; zeros(fftlen-winlen,1)];
win = win/sum(win);
windel = (0:(length(win)-1)) * win;

% Set overlap (Effects expansion of spectrogram display as no overlap plots
% fftlen data points and maximum overlap gives 1 data point)
ntime = 500; % Choose based on trial-and-error for best looking plot
overlap = floor(max(fftlen/2, (ntime*fftlen-length(x))/(ntime-1)));
ntime = floor((length(x)-overlap)/(fftlen-overlap));

% Create arrays
c1=(1:fftlen)';
r1=(0:ntime-1)*(fftlen-overlap);

% Take FFT of real data
b = fft(x(c1(:,ones(1,ntime))+r1(ones(fftlen,1),:)).*win(:,ones(1,ntime)));
if size(b,1) == 1
    m = length(b1);
    b(floor((m+4)/2):m) = [];
else
    m=size(b,1);
    b(floor((m+4)/2):m,:) = [];
end
b = b.*conj(b);

% Setup pixel locations for plot on frequency axis and time axis
f=(0:fftlen)*fs/fftlen; % Point spacing
t = (r1+windel)/fs;

% Set limit for dB scale
lim = max(b(:))*0.0001;

% Set dB scale
b=2.5*log10(max(b,lim));

% Plot
imh = imagesc(t,f/1000,b);

% Set up axis and labels for plot
axis('xy');
title('Spectrogram');
xlabel('Time (s)');
ylabel('Frequency (kHz)');

% Set and apply grayscale levels
colormap(jet);

% Apply color legend to plot
orient landscape;
APPENDIX G
DATA PLOTS
**G1. COMPARATIVE OVERLAYED PLOTS: MAGNITUDE v. TIME**

KEY: *The following graphs utilize the sound/color relationships listed below.*

<table>
<thead>
<tr>
<th>Sound #</th>
<th>Color</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>red</td>
</tr>
<tr>
<td>2</td>
<td>magenta</td>
</tr>
<tr>
<td>3</td>
<td>blue</td>
</tr>
<tr>
<td>4</td>
<td>cyan</td>
</tr>
<tr>
<td>5</td>
<td>green</td>
</tr>
<tr>
<td>6</td>
<td>yellow</td>
</tr>
<tr>
<td>7</td>
<td>orange</td>
</tr>
<tr>
<td>8</td>
<td>dark green</td>
</tr>
<tr>
<td>9</td>
<td>dark cyan</td>
</tr>
<tr>
<td>10</td>
<td>purple</td>
</tr>
<tr>
<td>11</td>
<td>brown</td>
</tr>
<tr>
<td>12</td>
<td>mustard</td>
</tr>
</tbody>
</table>
G1.1. Top six sounds for navigability
G1.2. Bottom six ranked sounds for navigability
G1.3. Top six ranked sounds for localization and level
G1.4. Bottom six ranked sounds for localization and level
G1.5. Top six ranked sounds for recognition
G1.6. Bottom six ranked sounds for recognition
G1.7. Top six ranked sounds for subjective response
G1.8. Bottom six ranked sounds for subjective response
G1.9. Sounds recommended for addition to the library
G1.10. Sounds not recommend for addition to the library
G2. COMPARATIVE OVERLAYED PLOTS: SPECTROGRAMS
G2.1. Top six sounds for navigability
G2.2. Bottom six sounds for navigability
G2.3. Top six ranked sounds for localization and level
G2.4. Bottom six ranked sounds for localization and level
G2.5. Top six ranked sounds for recognition
G2.6. Bottom six ranked sounds for recognition
G2.7. Top six ranked sounds for subjective response
G2.8. Bottom six ranked sounds for subjective response
G2.9. Sounds recommended for addition to the library
G2.10. Sounds not recommended for addition to the library
G3. COMPARATIVE OVERLAYED PLOTS: OCTAVE-BANDS

*KEY*: The following graphs utilize the sound/color relationships listed below. *NOTE*: The overlapping of data may make color differentiation difficult.

<table>
<thead>
<tr>
<th>Sound #</th>
<th>Color</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>red</td>
</tr>
<tr>
<td>2</td>
<td>magenta</td>
</tr>
<tr>
<td>3</td>
<td>blue</td>
</tr>
<tr>
<td>4</td>
<td>cyan</td>
</tr>
<tr>
<td>5</td>
<td>green</td>
</tr>
<tr>
<td>6</td>
<td>yellow</td>
</tr>
<tr>
<td>7</td>
<td>orange</td>
</tr>
<tr>
<td>8</td>
<td>dark green</td>
</tr>
<tr>
<td>9</td>
<td>dark cyan</td>
</tr>
<tr>
<td>10</td>
<td>purple</td>
</tr>
<tr>
<td>11</td>
<td>brown</td>
</tr>
<tr>
<td>12</td>
<td>mustard</td>
</tr>
</tbody>
</table>
G3.1. Top six ranked sounds for navigability
G3.2. Bottom six ranked sounds for navigability
G3.3. Top six ranked sounds for localization and level
G3.4. Bottom six ranked sounds for localization and level
G3.5. Top six ranked sounds for recognition
G3.6. Bottom six ranked sounds for recognition
G3.7. Top six ranked sounds for subjective response
G3.8. Bottom six ranked sounds for subjective response
G3.9. Sounds recommended for addition to the library
G3.10. Sounds not recommended for addition to the library