TOWSON UNIVERSITY

OFFICE OF GRADUATE STUDIES

ENVIRONMENTAL SOUND SIMILARITY AND IDENTIFICATION PERFORMANCE

By

Annamarie Rosen

A thesis proposal

presented to the faculty of

Towson University

in partial fulfillment

of the requirement for the degree of

Doctor of Audiology

Department of Audiology, Speech Language Pathology and Deaf Studies

Towson University
Towson, Maryland 21252

May 2017

TOWSON UNIVERSITY

Approval Page

**TOWSON UNIVERSITY
OFFICE OF GRADUATE
STUDIES**

THESIS APPROVAL PAGE

This is to certify that the thesis prepared by Annamarie Rosen, B.S. entitled "Environmental Sound Similarity and Identification Performance" has been approved by the thesis committee as satisfactorily completing the thesis requirement for the degree of Doctor of Audiology.

Nirmal Srinivasan, Ph.D.
Chairperson, Thesis Committee

11 May, 2017
Date

Jeremy Gaston, Ph.D.
Committee Member

9 May, 2017
Date

Brandon Perelman, Ph.D.
Committee Member

9 May, 2017
Date

Janet V DeLany, DEd

Janet DeLany
Dean of Graduate Studies

May 15, 2017
Date

Acknowledgments

I wish to thank my classmates for their unwavering support as we all went through this process together. Thank you to my family for their love and support during my journey of becoming an audiologist. Also, I would like to acknowledge my thesis committee members for the immeasurable hours where they provided feedback and guidance over the past year and a half. A special thanks to my fiancé for always listening and being a calming presence in my life.

ABSTRACT

Environmental Sound Similarity and Identification Performance

Anna Rosen

Prior to completing the experiments, listeners received a quick hearing test. Two experiments investigated listeners' performance related to classification and similarity ratings of environmental sounds. In Experiment 1, listeners were presented with a single stimulus to identify out of a closed set of 18 stimuli. Classification scores were above chance for all stimuli. In Experiment 2, listeners were presented with all possible pairs of stimuli from Experiment 1, after which they were prompted to choose a similarity rating using a 7-point Likert rating scale where 1 represented "as similar as possible" and 7 represented "as different as possible". The ratings from Experiment 2 were useful in developing a two-dimensional multidimensional scaling (MDS) solution. The Euclidean map showed separation between the three broad categories of sounds: power tools, household appliances, and vehicles. Sounds from similar sources were plotted closer together than sounds from different sources. A short survey was given upon completion of the experiments for listeners to report their familiarity with and frequency of use related to each stimulus. The survey used a 7-point Likert rating scale where 1 represented "not familiar" and "never use" and 7 represented "very familiar" and "use daily". The survey results indicated push mower, upright vacuum, and bathroom fan were the most familiar sounds; Mosquito, Mustang Piper, and Bell 212 were the least familiar. Bathroom fan, upright vacuum, and box fan were the most used while Bell 212, Mosquito, and Jet Ranger were the least used. Correlational analyses of familiarly and frequency of use indicated there was a significant moderately-strong positive correlation for all stimuli. Frequency of use and classification data from Experiment 1 had a significant and moderately positive correlation for

bathroom fan, hand vacuum, and sawzall. Familiarity and classification data from Experiment 1 had a significant and moderately positive correlation for bathroom fan, Bell 212, box fan, circular saw, and weed eater.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

KEY TO ABBREVIATIONS

ANSI: American National Standards Institute

ARL: Army Research Laboratory

dB: Decibel

GSI-61: Grason Stadler Instrument-61

Hz: Hertz

IRB: Institution of Research Board

ISI: Interstimulus interval

MDS: Multidimensional Scaling analysis

PTA: Pure tone average, average of thresholds at 500, 1000, and 2000 Hz.

VB: Van Bokkelen Hall

CHAPTER 1

INTRODUCTION

To reach the level of sound identification, listeners go through a complex process. The process begins with detection, where listeners determine the presence or absence of sound. Next, listeners develop perceptions related to detected sounds. Finally, listeners categorize what they hear and decide how to group incoming sounds. It is through this process that listeners use acoustic (i.e., frequency, intensity, duration) and semantic cues (i.e. context, experience, knowledge) that ultimately result in correct or incorrect identification of stimuli. Sound identification is possible through perception of psychoacoustic characteristics such as pitch, loudness, duration, and timbre. In some circumstances, sounds may share the same psychoacoustic characteristics leading listeners to interpret sounds as being similar.

Recent studies suggest that similarity between sound events can be used to predict accuracy in perceptual tasks (Dickerson & Gaston, 2014; Dickerson, Gaston, Perelman, Mermagen, & Foots, 2015; Gaston & Letowski, 2012; Gygi, Kidd, & Watson, 2007). One method of examining similarity is to use pair-wise similarity ratings and then analyze those ratings using a multidimensional scaling (MDS) analysis. MDS can then be used to create a Euclidean mapping, which is a 2-Dimensional representation of sounds grouped based on similarity (i.e., similar sounds close vs. dissimilar sounds further apart). Gaston and Letowski (2012) and Dickerson and colleagues (2015) showed that listeners made more errors in perceptual tasks when items were plotted closer, as opposed to further away from each other, in MDS space. These findings suggest that more perceptually similar sounds are harder to differentiate, resulting in a higher number of identification errors. More research examining the relationship between perceptual similarity and perceptual performance (e.g. discrimination, identification) may help explain why these listener errors occur.

The goal of this study is to extend recent findings related to perceptual performance to a new task with a novel set of environmental stimuli and to then examine individual differences in the relationship between perceptual similarity and performance in an identification task. Specifically, the aim of this study is to evaluate MDS distances between sound events to predict sound identification accuracy and the pattern of listener errors. The second purpose of this study is to determine if models based on individual similarly spaces would better predict performance than models based on the average similarity space.

CHAPTER 2

REVIEW OF LITERATURE

Past auditory research focused on studying the perception of complex stimuli using laboratory-created stimuli, speech, or music (Chan, Ho, & Cheung, 1998; Gygi et al., 2007; Holt & Lotto, 2006; Hyde et al., 2009; Jentschke & Koelsch, 2009; Kaan, & Swaab, 2002; Lively, Logan, & Pisoni, 1993; Pastore, Flint, Gaston, & Solomon, 2008). Fewer studies examined listener perception of environmental sounds (Gygi et al., 2007; Gygi, Kidd, and Watson, 2004; Marcell, Borella, Greene, Kerr, & Rogers, 2000; Vanderveer, 1980). Specifically, two of those studies presented listeners with more than 50 different environmental sounds encountered in daily listening, including nonverbal human sounds, animals, machinery, weather-related sounds, and sounds created by human activities (Gygi et al., 2007; Gygi et al., 2004). Fewer studies examined using perceptual similarity to predict classification scores (Dickerson & Gaston, 2014; Gaston & Letowski, 2012; Gygi et al., 2007). To date, there is little research that assessed listeners' abilities to correctly discriminate between similar environmental sounds; hence, the aim of this study is to add to the current research in this area.

The following literature review contains background information needed to grasp key points about psychoacoustic properties and the necessary steps for listeners to achieve accurate sound identification. The effects of perceptual similarity, as well as acoustic event characteristics, are described as they relate to identification performance. This review also discusses the usefulness of Likert rating scales and multidimensional scaling (MDS) analysis to create predictive mappings of listener errors to provide necessary information related to the design of this study.

Psychoacoustics

Psychoacoustics is the study of human psychological response to physical stimuli (Emanuel & Letowski, 2009). In other words, psychoacoustics is the study of how humans perceive sound. Sound characteristics such as pitch, loudness, duration, and timbre fall under the area of psychoacoustics. Listeners rely on these characteristics to make judgments of auditory similarity (Bonebright, 2001; Gygi et al, 2007). Each of these perceptual characteristics has an acoustic correlate that plays an integral part in sound perception and is further discussed here.

Pitch. Pitch is the perceptual correlate of frequency, perceived from low to high (ANSI, 1994). Pitch perception is highly dependent on stimulus frequency; however, pitch perception also depends on intensity, duration, and spectral complexity.

Frequency most commonly affects pitch perception; however, listeners can perceive a pitch that is acoustically absent from a stimulus. One example of this is the missing fundamental, a phenomenon where listeners perceive a pitch that is not acoustically represented in the frequency components of the actual sound, but is mathematically the greatest common factor in a frequency series (White, 1941). For example, if a complex stimulus consisting of 700, 800, and 900 Hz pure tones was presented to a listener, the listener would perceive a 100 Hz tone. The 100 Hz tone is the fundamental frequency, or greatest common frequency component of the complex stimulus in this example. The fundamental frequency, although missing in this example, would greatly influence pitch perception.

Stimulus intensity also impacts pitch perception; however, intensity is dependent on frequency and duration. A study by Stevens (1935) evaluated pitch perception of pure tones and found that when frequencies below 1000 Hz were increased in intensity, pitch perception effectively decreased. In the same study, pitch perception remained stable for pure tones between 1000 Hz and 2000 Hz, regardless of intensity changes; however, pure tones greater than 2000 Hz increased in pitch when intensity was increased (Stevens, 1935). It was then discovered that pitch perception was unique to the listener, yet listeners' pitch perception differed less than 2% based on frequency (Morgan, Garner, & Galambos, 1951). In general, as intensity is increased, listeners perceive pitch to decrease for low frequencies and increase for high frequencies (Morgan, Garner, & Galambos, 1951). In addition to frequency, stimulus duration impacts intensity, resultantly effecting pitch perception. When stimulus duration is less than 40 milliseconds (ms) and increased in intensity, listeners perceive decreases in pitch, regardless of frequency (Rossing & Houtsma, 1986).

As previously mentioned, stimulus duration also influences pitch perception. Pitch can be perceived in short tones; however, the duration required to perceive pitch depends on the frequency and intensity of the stimulus. For a stimulus about 4 ms in duration, a mid-ranged frequency (i.e., 1000 Hz) can be perceived to have pitch; whereas, a low frequency (i.e., 250 Hz) requires a longer duration of approximately 11 ms for pitch perception (Doughty & Garner, 1948). It was also reported that low frequencies with high intensities were perceived to decrease in pitch when stimulus duration was decreased; in contrast, high frequencies with low intensities were perceived to slightly increase in pitch when stimulus duration was decreased (Doughty & Garner, 1948).

Spectral complexity also influences perception of pitch. Pitch of complex stimuli, particularly frequencies of 350 Hz to 700 Hz, is usually perceived as slightly higher than pitch of simple stimuli with the same fundamental frequency (Lichte, 1941; Platt & Racine, 1985). Listeners use spectral cues, such as harmonics (integer multiples of the fundamental frequency) and partial overtones (harmonics other than the fundamental frequency), to develop pitch perception (Platt & Racine, 1985).

Research has shown that listeners' ability to discriminate pitch differences can be measured by presenting two tones that differ slightly in frequency, in immediate succession (Houtsma, 1995). The extent of change needed for a listener to detect a difference between two stimuli is known as the difference limen (DL), just noticeable difference (JND), or difference threshold (Guilford, 1954). The difference limen changes as a function of frequency and duration. For a pure tone in the mid to high frequencies, the DL can range from 0.1 to 1%; whereas, the DL is larger for frequencies below 500 Hz. Pitch discrimination becomes poorer as stimulus duration is decreased; therefore, the DL for pitch increases as stimulus duration is decreased (Turnbull, 1944). For example, an 800 Hz stimulus lasting 350 ms has a DL of 0.002, but the DL increases to 0.098 when the duration of the same stimulus is decreased to 12 ms.

Loudness. Loudness is the perceptual correlate of intensity, where listeners judge acoustic stimuli on a scale from soft to loud (ISO, 2006). This judgment is typically based on the intensity, or amplitude of the sound wave, but also depends on several other acoustic variables, such as stimulus duration, frequency, presentation mode, waveform shape, and spectral bandwidth (Röhl & Uppenkamp, 2012; Zahorik & Wightman, 2001).

Loudness most often is perceived from stimulus intensity or waveform amplitude cues. Intensity can increase with increased number of sound sources, as well as with decreased distance. When two random noise sources with identical intensity levels are in phase, the resultant intensity is 6 dB greater than each individual noise source. Listeners perceive the increase in intensity as an increase in loudness. Intensity of a stimulus is also related to the distance between the source and the listener. The intensity of direct sound decreases by approximately 6 dB every time the distance from the sound source is doubled in a free field; this is known as the Inverse Square Law (Scharine, Cave & Letowski, 2009). This law supports the statement that loudness perception increases with decreased distance.

Stimuli of short duration can be judged for loudness; however, a minimum duration of approximately 200 ms is required for a listener to perceive a change in loudness (Munson, 1947). The change in duration needed to perceive a change in loudness is specific to stimulus frequency and intensity (Ekman, Berglund, & Berglund, 1966). Stimuli with durations of 150 ms have small DLs until the duration exceeds 200 to 300 ms, in which case the DL is no longer dependent on stimulus duration (Zwicker & Fastl, 1999).

Stimulus frequency influences loudness perception, as indicated by equal loudness contours. The reference at each contour is a 1000 Hz tone at a specified SPL level. The contours represent the change in SPL needed to judge a tone of different frequency as the same loudness as the 1000 Hz reference. Of note is how the loudness contours change with increasing amplitude, specifically, the flattening of the contour in the low frequency region. The human ear is most sensitive to sounds between 3000 and 4000 Hz (Ballachanda, 1997). A sound within this frequency range is perceived as louder than a 1000 Hz stimulus of the same intensity. The ear is less sensitive to frequencies at the low and high ends of the frequency spectrum. For example, a 300 Hz stimulus would need to be delivered at a higher intensity to be perceived as equally loud as the 1000 Hz reference stimulus.

Additionally, loudness is dependent on stimulus presentation mode. Monaural stimulus presentations have different loudness perceptions than binaural presentations. Binaural stimulus presentations are perceived as louder than monaural presentations when the same intensity is presented to listeners. This increase in loudness perception, also known as binaural summation or binaural advantage, means that binaural presentations require lower stimulus intensity levels than monaural presentations to be perceived as equally loud (Reynolds & Stevens, 1960).

Loudness perception is influenced by waveform shape, which is primarily impacted by spectral bandwidth. Bandwidth (width of frequency region acoustically stimulated) encompasses the center frequency (frequency in the middle of the bandwidth of interest) and the surrounding frequencies above and below it. Spectral bandwidth varies across stimuli type, resulting in different loudness perceptions for narrowband, broadband, and speech stimuli. A higher intensity is needed for narrowband stimuli to be perceived as equally loud as broadband stimuli (Brand & Hohmann, 2001). Speech stimuli are perceived as louder than the narrowband stimuli at moderate intensity levels (e.g., 50 dB HL); however, loudness differences are less perceptible at very soft and very loud levels (Cox, Alexander, Taylor, & Gray, 1997). Increasing the bandwidth beyond the critical band (range of frequencies affected by acoustic stimulation, ultimately summed together by the auditory nervous system) effectively increases perceived loudness (Zwicker, Flottorp, & Stevens, 1957).

Duration. Duration is the perceived length of time of an acoustic stimulus. Listeners often perceive durations to be shorter or longer than the physical event (Eagleman, 2008). When stimulus duration is short, listeners tend to overestimate perceived duration; however, the degree of overestimation is inversely related to the physical stimuli duration, meaning the shorter the physical stimuli duration, the longer listeners perceive duration to be (Woodrow, 1951). When stimulus duration is between 200 ms and several seconds, perceived duration is fairly accurate (Scharine et al., 2009). When stimulus duration is longer than several seconds, listeners perceive duration based on their emotional state, expectations, and listening task (Angrilli, Cherubini, Pavese, & Mantredini, 1997; Droit-Volet, Brunot, & Niedenthal, 2004). Research has shown listeners perceive pleasant or familiar stimuli to have shorter durations compared to unpleasant or unfamiliar stimuli (Warm, Greenberg, & Dube, 1964; Warm & McCray, 1969).

Stimulus duration is also affected by stimulus intensity and frequency. Stimuli with high intensities are perceived to have longer durations (Goldstone et al., 1978). It has also been reported that stimuli of high frequency are perceptible at shorter durations than lower frequency stimuli (Watson & Gengel, 1969).

Timbre. Timbre has been classically difficult to define in terms of physical attributes (Patil, Pressnitzer, Shamma, & Elhilali, 2012). Timbre is the psychological sensation by which a listener can judge that two sounds are dissimilar when the sounds have the same duration, loudness, and pitch (ANSI, 1994). Timbre is determined using both spectral and temporal characteristics. Listeners use dynamic characteristics of sound including, but not limited to envelope shape and sound quality to aid in sound recognition (Gygi et al, 2007; Siedenburg, Jones-Mollerup, & McAdams, 2016).

Intensity changes over time form the temporal envelope of sound (Scharine et al., 2009). Temporal envelope can be subdivided into three main segments: the onset, steady state, and offset portions. Research has also shown that stimuli with up-ramp (rising) intensity changes are perceived to change more in loudness than stimuli with down-ramp (falling) intensity changes (Neuhoff, 2001). The onset of a stimulus is important for listeners' perception of timbre; listeners can perceive attack sharpness (how quickly the amplitude increases) at the onset of a sound (Lotto & Holt, 2011). Vowel production is an example of attack sharpness that produces timbre differences. When the vowel /u/ is produced (e.g., shoot), the oral cavity has an elongated shape and the bilabial articulators are constricted compared to when /ɑ/ is produced (e.g., father), the oral cavity is much wider and the bilabial articulators are not constricted. As the shape of the vocal tract changes to form different vowels, the amplitude of the harmonics varies. In the case of /u/ versus /ɑ/, /u/ has more attack sharpness as a result of greater constriction in production.

Sound quality is determined by listeners' emotional judgment of the stimulus and degree of listener satisfaction (Scherer & Oshinsky, 1977). Perceptual labels of sound quality include roughness, breathiness, degree of pleasantness, and naturalness. Music provides great examples of differences in sound quality; one guitarist and one flutist both play middle C, yet they sound completely different. Guitars are considered string instruments, while flutes are considered wind instruments; the different methods of sound production and the different resonances within the instrument cavities contribute to different perceived sound quality that can be perceived by listeners (Lotto & Holt, 2011).

Perception

After an acoustic stimulus is detected, it is then perceived. Auditory perception is the ability to identify, interpret, and attach meaning to sound (Scharine et al., 2009). For listeners to perceive auditory events, they must attend to stimuli, store the perception in memory, use context clues, develop expectations, and use relevant knowledge about incoming stimuli (Letowski & Letowski, 2012; Lotto & Holt, 2011). This section will describe how auditory perception is influenced by attention, memory, context, expectations and previous knowledge.

Auditory attention is required for listeners to detect changes in auditory stimuli when the behavioral task requires an explicit response. Additionally, attention is required for listeners to focus on target stimuli in an environment with competing stimuli (Shinn-Cunningham, 2008). If opposing stimuli are determined to be noise, the listener may need to attend to the target stimuli more than in situations where there are no opposing stimuli (Carlyon, Cusack, Foxton, & Robertson, 2001). Increased attention can also help listeners identify initially ambiguous stimuli (Shinn-Cunningham, 2008).

Auditory memory plays an important role in auditory perception, particularly the use of top-down processing (processing influenced by cognition and experience). Top-down processing allows listeners to fill in missing or ambiguous stimuli, allowing them to respond quicker (Schultz, Dayan, & Montague, l997; Summerfield & Egner, 2009). Numerous examples of memory and top-down processing can be found related to speech perception. One example of top down processing and speech perception is a fisherman who hears a word starting with “sh”. It can be any word beginning with that combination of sounds (i.e., shop, shoe, ship, etc.), but for a fisherman, it is more likely that the word is ship based on the context. The fisherman in this example would interpret the auditory stimulus using cognition and experience to fill in the missing parts of speech.

Listeners use context clues to aid in sound perception. Context can affect how individual sounds occurring in sequence are perceptually grouped. For example, when stimuli are presented in a logical series, listeners attribute the sounds to a single source (Winkler et al., 2003). Perceptual grouping then impacts listeners' expectations and vice versa (Gygi et al, 2007; Holt & Lotto, 2006). Listener expectations of auditory events and previous knowledge also influence auditory perception. A leading stimulus (stimulus that comes first) can affect the interpretation of all subsequent stimuli as listener expectations develop (Gygi et al, 2004; Snyder, Carter, Lee, Hannon, & Alain, 2008; Tsunada & Cohen, 2014). Music and language provide examples of listener expectation and knowledge. Research indicates prior exposure to unfamiliar music can add to listeners' knowledge and influence expectations (Bey & McAdams, 2002). Likewise, in written and oral language, listeners use syntactic rules to develop expectations for upcoming words based on context (Kaan & Swaab, 2002; McNamara, 2005; West & Stanovich, 1982).

Discrimination

Auditory discrimination is a method for measuring perception. Auditory discrimination is the ability to determine whether two acoustic stimuli are the same or different (Keith, 1988). Listeners use certain traits to discriminate between multiple stimuli; these traits can include consistency, harmonics, beginning and ending together, and sounding continuous unless otherwise disrupted (Gygi et al., 2007; Lotto & Holt, 2011).

Sometimes listeners fail to detect changes in auditory stimuli; this is known as the phenomenon of change deafness (Mitroff, Simons, & Levin, 2004). Like change blindness, change deafness is tested by presenting two auditory scenes (multiple auditory stimuli simultaneously) separated by a brief pause known as an interstimulus interval (ISI). The second scene can either be the same or different and participants are asked to determine if there was a change. Studies have created changes in the second auditory scene by replacing one sound with a new one or by removing a sound completely (Eramudugolla, Irvine, McAnally, Martin, & Mattingley, 2005; Gregg & Samuel, 2009). Gregg and Samuel (2009) reported that listeners noticed a change more often when the sound removed signaled a different acoustic or sematic category. It was then suggested by Dickerson and Gaston (2014), that change deafness errors were influenced by perceptual category and the similarity between the changed sound and background stimuli; Dickerson and colleagues (2015) later confirmed this relationship.

Perceptual Similarity. There are several techniques to determine listeners' discrimination capabilities, but this review will focus on Likert rating scales and MDS analysis. Likert rating scales provide a range of responses that can be used to describe sound similarity. For example, on a scale of 1 to 7, sounds described using a rating of 1 may represent the "most similar" versus a rating of 7 would be the "least similar". This type of scale can be used in conjunction with certain analysis techniques, such as MDS, to further evaluate listeners' discrimination abilities. MDS allows raw data to be analyzed and developed into perceptual dimensions of space. The data are plotted as points in space, where the distance between points represents the level of similarity. MDS analysis can be used for large data sets, hence their use to better understand the complex relationship between perceptual similarity and perceptual performance (Dickerson & Gaston, 2014). Recent research had listeners judge similarity of paired environmental stimuli using a Likert scale (Dickerson et al., 2015; Gygi, Kidd, and Watson, 2007). In both studies, the MDS analyses were used to analyze perceptual similarity data where listeners groups sounds based on both acoustic and semantic properties. Gaston and Letowski (2012) used MDS analysis to develop a Euclidean mapping of listeners' perception similarities of small arms. In Euclidean mapping, sounds perceived by listeners to be most similar appear grouped together, whereas sounds that are found to be dissimilar are separated spatially on the map. For example, in Gaston and Letowski's (2012) study, handgun shots were grouped together, but separated from rifle shots.

Localization

Sound localization is the ability to determine direction of a sound source (Bellis, 1996). When complex sounds are directly in front of the listener and at ear level, it has been reported that listeners can detect as small as a 1° to 2° change along the horizontal plane (Ford, 1942; Gardner, 1968). To achieve this precision of sound localization, listeners use two types of binaural cues to determine the direction of a sound along a horizontal plane: interaural time difference and interaural intensity difference (Letowski & Letowski, 2012; Lotto & Holt, 2011).

Interaural time difference refers to the difference in arrival time of the stimulus between ears. The interaural time difference is the result of two different phases (specific points in the cycle of a waveform measured by degree angle) of the same sound arriving at each ear (Lotto & Holt, 2011). When a sound source is equidistance to both ears, the sound will be perceived as coming from either directly in front or directly behind;

however, when the source is closer to one ear than the other, the arrival time and phases will cue a horizontal location (Emanuel & Letowski, 2009). This binaural difference cue is more profound for frequencies below 1500 Hz (Rayleigh, 1907).

Similarly, interaural intensity difference also aids in localization by providing the difference in sound intensity between both ears. Listeners experience what is known as the head shadow effect, which is an intensity-related phenomenon where listeners experience reduced loudness at the ear furthest from the source. This phenomenon is typically seen with frequencies above 1000 Hz (Emanuel & Letowski, 2009). Higher frequencies have smaller wavelengths; therefore, have a greater reduction in energy, or perceived loudness, as it crosses the head.

Sound localization becomes more difficult when the source is directly in front, behind, above, or below the listener because there are no interaural timing or intensity differences to compare for these source locations (Gardner & Gardner, 1973; Lotto & Holt, 2011). The ears, head, and shoulders help with monaural localization in the vertical plane through what is called the baffle effect. The baffle effect is an increase in sound pressure in front of an object due to reflected energy (Letowski & Letowski, 2012). Depending on the stimulus frequency, certain frequencies get enhanced, while others get attenuated. The pinnae enhance 3 to 4 kHz while the head and the shoulders enhance 2 to 3 kHz (Gardner, 1973; Roffler & Butler, 1978).

Most sound localization testing is performed in anechoic chambers where there are virtually no echoes, but this does not mimic most everyday listening situations. Typically, listening environments have reverberation where listeners hear indirect reflections of sound in addition to the original acoustic signal (Ihlefeld & Shinn-

Cunningham, 2011). Reverberation primarily affects the ear furthest from the source and low frequency information, resulting in interaural timing differences between ears (Shinn-Cunningham, Kopco, & Martin, 2005). Listeners perceive the source to be further away in reverberant conditions because they receive a distorted signal that is out of phase with the original acoustic signal; reverberations also aid listeners in binaural localization (Gardner & Gardner, 1973; Ihlefeld & Shinn-Cunningham, 2011; Kock, 1950).

Moreover, the distance between a listener and a sound source effects sound localization. In an open field, as the distance between the listener and the source decreases to within 0.5 to 1.0 meter of the listener's head, interaural intensity differences increase. While interaural intensity differences are dependent on the distance, interaural timing differences are independent of distance (Letowski & Letowski, 2012). Listeners perceive changes in distance of moving sources using the Doppler effect; as a sound source moves closer to listeners, the auditory stimulus intensity increases and is perceived as increasing in pitch versus decreasing in intensity and lowering in pitch as the source moves further away (Letowski & Letowski, 2012). Listeners make judgments about stimulus location using these cues.

Categorization

Acoustic categorization is the ultimate goal of listeners; it is the ability to classify and identify sounds (Tsunada & Cohen, 2014). Categorization is not the same as similarity judgments, used in discrimination tasks, although they are related concepts that can influence each other (Goldstone, 1994). Categorization is influenced by goals, theories, and within-category variability (Barsalou, 1991; Fried & Holyoak, 1984; Murphy & Medin, 1985).

Categorization tasks can take many forms, one being sorting. Several sorting tasks have been used to study how listeners categorize acoustic stimuli. Vandeveer (1980) reported that listeners sorted sounds based on temporal cues rather than spectral cues. Bonebright (2001) found that listeners used amplitude, frequency, and duration cues to make perceptual judgments. Another study suggested that listeners sorted sounds based on object type and context more than by event type and sound quality (Marcell, Borella, Greene, Kerr, & Rogers, 2000). This indicated that semantic information was highly important in environmental sound perception. Similarly, the results collected by Gygi and colleagues (2007) showed that listeners tended to sort stimuli based on source type. Each study gave listeners different directions and produced diverse results suggesting that sorting tasks are influenced by multiple factors.

Categorization can also be performed through identification tasks. Numerous researchers have studied listeners' identification abilities by varying the stimulus in some way. For example, some studies evaluated listeners' identification abilities of naturally occurring sounds such as footsteps or gait using posture and walker gender as variables (Li, Logan, & Pastore, 1991; Pastore, Flint, Gaston, & Solomon, 2008). Others evaluated identification of specific stimuli such as traffic noises (Cermak & Cornillon, 1976), broken versus bounced objects (Warren & Verbrugge, 1984), and weapons signatures (Fluitt, Gaston, Karna, & Letowski, 2010; Gaston & Letowski, 2012). The stimuli used in these studies had a variety of acoustic source properties, which allowed listeners to develop unique sound groupings.

Despite listeners' abilities to identify sound sources, listeners can make significant judgment errors. Recent research has tried to determine why these errors occur and if listener error can be predicted based on individual sound similarity (Dickerson & Gaston, 2014; Dickerson et al., 2015; Gaston & Letowski, 2012; Gygi et al., 2007). Euclidean mappings developed from MDS analysis can be used to better understand how listeners categorize sounds. By analyzing what types of sounds are grouped together, it can help reveal what auditory and semantic cues listeners might use for environmental sound identification, and in turn, then be used predict listener error.

Present Study

This study was designed to further evaluate the relationship between perceptual similarity and classification performance. Environmental sounds that differ from previous studies were used primarily; however, some sounds were repeated from previous studies for comparison. The experimental study consisted of two parts: a classification task and a similarity task. Classification performance was used to evaluate listeners' abilities to classify a range of environmental sounds across three levels from broad categories to individual stimuli. Similarity ratings collected for all paired stimuli from the classification task were used to develop a Euclidean mapping of the presented sounds. Then, the results from Experiment 1 were used to predict identification performance of Experiment 2 for both individual and averaged MDS spaces. Following the experimental part of the study, a short survey was completed by listeners using a 7-point Likert rating scale to collect data on their level of familiarity with each item and their frequency of use for each item. This study was approved by the Towson University Institutional Review Board (*See Appendix A*).

Hypotheses

1. Experiment 1 performance will be above chance. Chance performance is what

would be obtained if participants chose at random given the number of stimuli. Chance level was 5.56% for Level 1, 11.11% for Level 2, and 33.33% for Level 3.

2. Participants will give useful ratings for Experiment 2.

3. Experiment 2 will predict Experiment 1 performance for both the averaged and individual MDS spaces, but the individual distances will predict performance better than the averaged distances.

4. Survey results will show familiarity and frequency of use associated with each stimulus will be correlated. Familiarity will be correlated with classification performance, as will frequency of use.

CHAPTER 3

RESEARCH METHODOLOGY

Participants

A total of 40 adults, aged 18 to 30 years, with normal hearing were recruited from Towson University using a flyer (*See Appendix B*). An effort was made to have an equal number of female and male participants, but more females (N=23) participated than males (N=12). All participants signed a consent form (*See Appendix C*). Normal hearing was defined in this study as pure-tone air conduction hearing thresholds better than 20 dB HL for octave audiometric frequencies from 250 to 8000 Hz (i.e., 250, 500, 1000, 2000, 4000, 8000) and an absence of a self-reported history of otologic pathology (i.e., diagnosed hearing loss). Participants who did not meet the hearing criteria were still run, since hearing loss within a narrow frequency likely did not impair their ability to characterize complex environmental sounds at wider frequency bands, and their hearing deficit was annotated in their dataset. All information was kept secure with limited access only by experimental staff.

Equipment

A Roland R-44 field recorder and G.R.A.S. Sound and Vibration 40AF microphone was used to make new stimuli recordings (44.1 kHz, 16-bit) while pre-recorded stimuli were gathered from an existing database. A Grayson Stradler-61 audiometer was used with TDH-50 supra-aural headphones to do the hearing test for each participant in the study. Testing involved noise levels that complied with ANSI S3.1-1999 (R2008) standard for earphone/headphone listening. A custom computer program was designed by the U.S. Army Research Laboratory (ARL) in PEBL, was used to control the experimental program. Participants completed the experimental study on a laptop through Beyerdynamic T-90 stereo headphones coupled to an A-20 headphone amplifier.

Stimuli

Stimuli were recorded samples of vehicle, mechanical, and instrumental sounds consisting of items from multiple sound categories and multiple sub-categories. Some sounds were used from an existing library while other sounds were recorded specifically for this study. Sound recordings each lasted two-seconds in duration. Table 1 shows the stimuli that were used in this study. The stimulus list was arranged so that there were three levels of sound categories. Level 3 contained three broad classes of sounds, Level 2 contained three subcategories from Level 3, and Level 1 contained 18 individual sounds with two unique stimuli for each category from Level 2.

Procedure

The participants were seated in a sound-attenuating booth where the hearing screening was performed using a GSI-61 audiometer and TDH-50 supra-aural headphones. Participants completed two experiments on a laptop for this study following the hearing screening. The order of the experiments was randomized so that some participants began with Experiment 1 while others began with Experiment 2. Following the experimental design, listeners completed a short survey of their familiarity and frequency of use related to each stimulus (*See Appendix D*). This survey utilized a 7-point Likert rating scale where 1 represented “not familiar” or “never use” and a 7 represented “very familiar” or “use daily”.

Experiment 1. Participants performed a single-interval identification task to test their ability to classify and identify each of the stimuli. Each sound in the list was randomly presented a total of 10 times for a total of 180 trials. Listeners were asked to classify the sound across three levels as shown in Figure 1. By selecting a single icon for each presentation, the listener effectively classified the stimulus at levels 1, 2, and 3. For example, if the sound was a push mower (Level 1), the listener also knew it was a yard tool (Level 2), which falls under the umbrella of tools (Level 3).

Experiment 2. The stimuli from Experiment 1 were presented in a random pairwise fashion such that all possible pairs were presented in both directions (i.e., stimulus A then stimulus B, as well as stimulus B then stimulus A) for a total of 324 trials. The responses were collected to build the perceptual similarity spaces using Multidimensional scaling (MDS) analysis. The first sound in each pair was presented and then a 750 ms interstimulus interval (ISI) was between that and the presentation of the second sound. Following termination of the second sound, participants were prompted to rate how similar the two sounds were using a 7-point Likert rating scale where 1 represented “most similar” and 7 represented “least similar”. Participants were asked to try to use the entire rating scale. The response interval had no time limit meaning the experimental program did not automatically advance to the next trial if a response was not made within a certain time frame; otherwise, the program advanced to the next trial immediately following the participant’s response.

Analyses

Experiment 1 was analyzed by looking at classification and identification accuracy at each of the levels of analysis. Experiment 2 similarity distances were analyzed using a non-metric MDS analysis to create a Euclidean map of perceptual distances between sounds. The MDS space in Experiment 2 of perceptual similarity was based on average similarity ratings (i.e., AB and BA were averaged).

CHAPTER 4

Results

Analyses were completed to determine if Experiment 1 (i.e., classification task) performance was above chance and if the ratings given in Experiment 2 (i.e., similarity task) were useful. Analyses were also performed to better understand if the survey results correlated with the classification data. Listeners' performance on the classification task was above chance for all 18 individual stimuli. Additionally, the ratings given in the similarity task were used to generate a Euclidean map of the stimuli used in this study. The survey results significantly correlated with listeners' classification performance.

The data for 5 participants were excluded from data analysis because of missing information or errors in the outputs. Data from the remaining 35 participants were used for analysis. Listeners consisted of 12 males and 23 females, aged 18 to 29 years old (M = 23.31, SD = 2.07) with an average pure tone average (PTA) of 3.45 dB HL (SD = 4.00).

An independent samples t-test was performed to determine whether task order influenced classification performance amongst listeners. There was no significant difference in classification performance ($p > 0.05$) based on task order; therefore, all subsequent analyses were conducted using pooled data.

Results from the classification task showed that listeners correctly identified stimuli 47.4% of the time at Level 1, 66.5% at Level 2, and 84.9% at Level 3. The breakdown of classification scores is shown in Table 1, as well as Figures 2, 3, and 4.

Experiment 1

Level 3 Analysis. Level 3 analyses evaluated listeners' abilities to classify stimuli into three broad categories: power tools, vehicles, and household appliances. A paired-samples t-test was conducted to compare the classification performance of the Level 3 stimuli. There was a significant difference in the performance scores for blenders and vacuums (*M* = -9.86, *SD* = 21.30), *t*(34) = -2.74, *p* = 0.003, for blenders and fans (*M* = -28.42, *SD* = 30.79), *t*(34) = -5.46, *p* = 0.0003, and for vacuums and fans (*M* = -18.57, *SD* = 31.80), *t*(34)= -3.46, *p* = 0.0003. Vehicle sounds had the highest correct classification amongst listeners (*M* = 95.57, *SD* = 4.41), tools had the second highest classification scores (*M* = 87.00, *SD* = 6.98) and household appliances had the lowest classification scores (*M* = 72.10, *SD* = 10.49). The easiest stimuli for listeners to differentiate between were vehicles and household appliances (M_{diff} = 23.5). It was more difficult for listeners to differentiate between tools and household appliances (M_{diff} = 14.9); however, the hardest stimuli to differentiate between were vehicles and tools (M_{diff} = 8.6).

Level 2 Analysis. Level 2 analyses evaluated listeners' abilities to classify stimuli into nine subcategories from Level 3. A paired-samples t-test was conducted to compare the classification performance of the Level 2 stimuli based on the three broad categories from Level 3. There was a significant difference in the performance scores for helicopters and wheeled vehicles (*M* = 27.86, *SD* = 21.32), *t*(34) = 7.73, *p* = 0.0003 , for helicopters and prop planes (*M* = 3.86, *SD* = 9.24), *t*(34) = 2.47, *p* = 0.006, for wheeled vehicles and prop planes (*M* = -24.00, *SD* = 20.54), *t*(34) = -6.91, *p* = 0.0003, for drills and yard tools (*M* = -46.86, *SD* = 17.28), *t*(34) = -16.04, *p* = 0.0003, for drills and saws (*M* = 8.00, *SD* = 19.45), *t*(34) = 2.43, *p* = 0.007, for yard tools and saws (*M* = 54.86, *SD* = 15.41), *t*(34) = 21.06, *p* = 0.0003, for vacuums and blenders (*M* = 9.86, *SD* = 21.30), *t*(34) = 2.74, *p* = 0.003, for fans and vacuums (*M* = 18.57, *SD* = 31.80), *t*(34) = 3.45, *p* = 0.0003, and for blenders and fans (*M* = -28.42, *SD* = 3.08), *t*(34) = -5.46, *p* = 0.0003. Stimuli with the highest correct classification amongst listeners included helicopters (*M* = 96.14, *SD* = 8.23), yard tools (*M* = 94.71, *SD* = 8.13), and prop planes (*M* = 92.29, *SD* = 9.73). Stimuli with the lowest classification scores included saws (*M* =39.86, *SD* =14.77), blenders (*M* =40.29, *SD* = 14.85), and drills (*M* = 47.86, *SD* = 16.73). The easiest vehicle stimuli for listeners to differentiate between were helicopters and wheeled (M_{diff} = 27.9) and the hardest were wheeled and prop planes (M_{diff} = -24.0), while differentiation between helicopters and prop planes were non-significant (*p* = 0.7). The easiest power tool stimuli for listeners to differentiate between were yard tools and saws (M_{diff} = 54.9) and the hardest were drills and yard tools (M_{diff} = -46.9), while differentiation between drills and saws were non-significant (*p* = 0.7). The easiest household appliance stimuli for listeners to differentiate between were blenders and fans (M_{diff} = -28.43) while differentiation between blenders and fans (*p* = 0.4) and vacuums and fans were non-significant (*p* = 0.05). Table 2 shows all significant pairwise comparisons for Level 2.

Most of the classification errors were still part of the same Level 2. Some examples included bathroom fan for box fan, Bell 212 for Jet Ranger, circular saw for sawzall, and food processor for stand blender. There were also cases where the sound presented compared to listeners' responses did not fall under the same Level 3 classification; this was seen with box fan. Listeners confused box fan (which falls under household appliances) with Diesel SUV (which falls under vehicles). This same error was seen between saws and blenders, drills and blenders, and vacuums and saws.

Experiment 2

Similarity Ratings and MDS. Listeners' responses from the similarity task produced useful ratings for paired stimuli. Listeners chose a similarity rating score for each set of paired stimuli using a 7-point Likert rating scale where 1 represented "as similar as possible" and 7 represented "as different as possible". These ratings were then organized using Multidimensional Scaling (MDS) analysis, which was then used to generate a Euclidean map of the 18 stimuli presented to listeners during the similarity task. MDS generated a 2-dimentional map shown in Figure 5. Stimuli that were in close approximation to one another were judged as highly similar on the Likert rating scale by listeners during Experiment 2, whereas stimuli plotted further apart were judged as highly dissimilar by listeners.

Survey

Survey Results. Listeners used a 7-point Likert rating scale to report how familiar they were with each item and how frequently they used each item. Ratings at the lower end of the scale (i.e., 1) indicated they were not familiar or never used the item while ratings at the upper end of the scale (i.e., 7) indicated they were very familiar or used the item daily. Responses collected from the survey indicated that listeners were most familiar with push mower (M = 6.49, SD = 0.89), upright vacuum (M = 6.31, SD = 1.13), and bathroom fan (M =6.14, SD =1.26) out of the 18 individual stimuli. The least familiar stimuli were Mosquito (M =1.83, SD = 1.15), Mustang Piper (M =2.14, SD = 1.50), and Bell 212 (M = 2.31, SD = 1.66). Responses also showed that most used stimuli were bathroom fan (M = 6.2, SD = 1.32), upright vacuum (M = 5.03, SD = 1.38), and box fan (M = 4.0, SD = 1.88). Least used stimuli included Bell 212 (M = 1.26, SD = 1.07), Mosquito (M = 1.26, SD = 1.07), and Jet Ranger (M = 1.31, SD = 1.11). Table 3 shows the collective results from the survey.

Correlational Analyses. Correlational analysis showed that frequency of use, as collected from the survey, and performance scores from Experiment 1 were significant and moderately positively correlated for bathroom fan ($r = 0.445$, $p < 0.01$), hand vacuum ($r = 0.399$, $p < 0.05$), and sawzall ($r = 0.446$, $p < 0.01$). Table 4 shows all correlations and corresponding *p* values for frequency of use and performance scores. Familiarity with the stimulus, also collected from the survey, and performance scores from Experiment 1 were significant and moderately positively correlated for bathroom fan ($r = 0.645$, $p < 0.001$), Bell 212 ($r = 0.358$, $p < 0.05$), box fan ($r = 0.516$, $p < 0.01$), circular saw ($r = 0.343$, $p < 0.05$), sawzall ($r = 0.488$, $p < 0.01$), and weed eater ($r = 0.492$, $p < 0.01$). Correlations and corresponding *p* values for familiarity of sound and performance scores are in Table 5. Analysis also revealed that familiarity with the stimulus and frequency of use were significant and moderately to strongly positively correlated for all stimuli as shown in Table 6. A correlation for the aggregate survey data revealed there was a strong positive relationship between familiarity and frequency of use ($r = 0.808$, $p < 0.001$).

CHAPTER 5

Discussion

Experiment 1

The results collected from the classification task showed listeners correctly identified stimuli above chance. Listeners correctly identified the presented stimulus nearly 48% of the time. When considering listeners choice of 18 different stimuli, they had a 5.56% chance of guessing correctly for each presentation. The instructions for this experiment are shown in Figure 6 while the exact page that listeners used to select their responses are shown in Figure 1 as mentioned previously.

There are complex relationships between frequency, intensity, and duration, which may have contributed to listeners' performance (Rossing & Houtsma, 1986; Stevens, 1935); however, determining which acoustic feature(s) influenced performance is complicated because each acoustic variable is somewhat influenced by the others. For example, frequency depends on intensity, duration, and spectral complexity. Intensity can depend on duration, frequency, presentation mode, waveform shape, and spectral bandwidth (Röhl & Uppenkamp, 2012; Zahorik & Wightman, 2001). Frequency, waveform shape, and spectral bandwidth could not be controlled, but intensity, duration, and presentation mode were controlled during stimuli recordings.

Categorizing acoustic stimuli is also influenced by psychoacoustics, goals, theories, and within-category variability (Barsalou, 1991; Fried &Holyoak, 1984; Murphy &Medin, 1985). Timbre is the psychological sensation by which listeners can judge two sounds as dissimilar when they have the same duration, loudness, and pitch (ANSI, 1994). Timbre is influenced by envelope shape and sound quality, which can aid in sound recognition (Gygi et al, 2007; Siedenburg, Jones-Mollerup, & McAdams, 2016). Listeners may have used intensity and spectral changes related to the envelope shape to make classifications and discriminations during the study (Scharine et al., 2009). This study controlled the within category variability to some degree, but it was not possible to fully control listeners' goals and theories related to the task or the study. Within category variability was effected by what stimuli were recorded and included in the study. It is important to note that additional stimuli such as drill press, miter saw, hedge trimmer, stand blender, and vent hood were considered, but were not used in the study due to recording difficulties or ease of access. The resultant stimuli presented in this study were novel and covered a range of sounds, some of which could be encountered in everyday life situations.

In general, the classification errors made by listeners indicated that they chose items that belonged to the same Level 3 (i.e., belonging to one of three broad categories of sounds) or Level 2 (i.e., subcategories of Level 3) most of the time. For example, bathroom fan was often confused with box fan, hand vacuum, and upright vacuum. In this example, the incorrect responses belonged to either the same Level 2 (i.e., fan) or Level 3 (i.e., household appliance). Some confusions were made where the item presented and the response belonged to different Level 3 categories. An example of this was when box fan was presented listeners thought it was a Diesel SUV; these two sounds belonged to two different Level 3 categories (i.e., household appliance vs. vehicle).

Experiment 2

Listeners rated or grouped paired stimuli without specific instructions in the similarity task. Listeners used the 7-point Likert rating to describe the relationship between two stimuli, including a stimulus paired with itself. Figure 7 is the exact page that listeners used to select their responses for this task. The listeners determined ratings based on acoustic, semantic features, a combination, or by other methods as no explicit instructions on how to rate the stimuli were given. There is also the possibility that listeners changed their rating method during the similarity task. In hindsight, it may have been advantageous to ask listeners how they chose similarity ratings upon completion of the experimental study.

As mentioned previously, timbre refers to listeners' abilities to judge that two sounds are dissimilar when the sounds have the same duration, loudness, and pitch (ANSI, 1994). The onset of a stimulus could have aided in timbre perception (Lotto & Holt, 2011). Listeners could have perceived attack sharpness (i.e., how quickly the amplitude increased) at the onset of a sound (Lotto & Holt, 2011). For example, the onset of the lawn mower compared to the bathroom fan might have aided listeners in determining the two stimuli were different. Listener expectation and previous knowledge also likely effected performance on this rating task. Expectation has been shown to impact perceptual grouping (Gygi et al, 2007; Holt & Lotto, 2006). Listeners expectations were controlled by providing the same instructions to each participant via the instruction page, but previous knowledge could not be controlled. The presentation of the paired stimuli could have effected ratings if not considered when designing the procedure for this study. For example, a leading stimulus could have affected the interpretation of subsequent stimuli as listener expectations developed (Gygi et al, 2004; Snyder, Carter, Lee, Hannon, & Alain, 2008; Tsunada & Cohen, 2014), but this was eliminated as a contributing factor by presenting stimuli in both orders (e.g., drill 1 then drill 2, drill 2 then drill 1).

The similarity rating task measured listeners' perception abilities by looking at their ability to discriminate between the paired stimuli. Auditory discrimination is the ability to determine whether two acoustic stimuli are the same or different (Keith, 1988). This study presented two stimuli separated by a brief ISI of 750 ms. In some cases, the two stimuli were in fact different, but in other cases the stimuli were identical. Past research showed listeners noticed a change when it involved a difference in acoustic or semantic groups (Dickerson et al., 2015; Gregg & Samuel, 2009; Gygi, Kidd, & Watson, 2007). Other work suggested listeners' discrimination skills and ultimately perceptual groupings were effected by object or source type and context more than by event type and sound quality (Gygi et al., 2007; Marcell, Borella, Greene, Kerr, & Rogers, 2000). The results of this study indicated listeners grouped stimuli based on source type, which was seen as separation between the vehicles, household appliances, and tools in the Euclidean map.

The Likert rating scale results also showed how listeners sorted stimuli. Vandeveer (1980) reported that listeners sorted sounds based on temporal cues rather than spectral cues. Bonebright (2001) found that listeners used amplitude, frequency, and duration cues to make perceptual judgments. The focus of this study did not evaluate the temporal or spectral components of the stimuli yet listeners' responses suggested they were influenced by these acoustic cues. Listeners may have used acoustic cues such as frequency to determine stimulus category. Table 7 shows that listeners made mistakes during the similarity task when they gave paired stimuli a rating of 1, meaning "as similar as possible", when the stimuli were not identical. Majority of these confusions were made for paired stimuli that belonged to different Level 2 categorizes; therefore, listener judgments were likely based on more than object source. Listener errors could have been the result of frequency or temporal commonalities. For example, blenders and vacuums both create suctioning sounds that could be close in pitch, which would lead to listener error. Further analyses would need to be done to confirm this speculation.

Multidimensional Scaling

Multidimensional scaling allowed the raw similarity data to be analyzed and developed into perceptual dimensions of space. MDS analyses suggested that listeners judged the stimuli to be grouped according to two dimensions. Figure 5 was used to make inferences about the relationship between stimuli. Stimuli plotted closer together were judged as more similar; therefore, presumably harder to differentiate than stimuli plotted further apart. For example, hand vacuum and upright vacuum were in close approximation on the Euclidean map, meaning listeners rated these two stimuli as being closely related. The average similarity rating for this set of paired stimuli was 2.63 on the 7-point Likert rating scale where a rating of 1 indicated the sounds were "as similar as possible" and 7 indicated they were as dissimilar as possible. Another example of this was seen with the helicopters, which were in close approximation to each other on the map and were given an average similarity rating of 2.69. Interestingly, listeners were familiar with fans (average rating of 5.74), but were unfamiliar with helicopters (average rating of 2.31). This may be due to familiarity with these items through other sources such as movies or television.

Past work has shown that when there was a high level of similarity between closely plotted stimuli, these perceptually similar stimuli were harder to differentiate leading to more classification error (Gaston & Letowski, 2012). The results of this study appear to support this speculation; there was an increase in listener error when stimuli were rated as similar and closely approximated on the Euclidean map. For example, listeners frequently confused bathroom fan and box fan when presented together in Experiment 2. The incorrect responses from the classification task showed listeners thought the fans were vacuums and wheeled vehicles. Listeners also confused one helicopter with the other helicopter (i.e. Bell 212 for Jet Ranger and vice versa).

Studies have tried to determine why listener errors occur and if errors can be predicted by looking at similarity data (Dickerson & Gaston, 2014; Dickerson et al., 2015; Gaston & Letowski, 2012; Gygi et al., 2007). Euclidean maps were utilized to graph the results from MDS to better understand how listeners categorize stimuli. The stimuli were then assessed to see if what commonalities and differences exist between the stimuli plotted close together compared to far apart. Close stimuli shared auditory and/or semantic cues that separated them from stimuli plotted further away. It can be concluded that listeners in this study grouped stimuli based on both acoustic and semantic features.

Survey

The survey results suggested the least familiar stimuli belonged to the helicopter and prop plane groups while the most familiar stimuli fell under general household appliances. These responses had some overlap with those collected for frequency of use. Most used items were those amongst the household appliances and least used were helicopters and prop planes. These collective results compared with the results from the classification task were interesting. Considering helicopters and planes were the least familiar and least used, it might be predicted that these stimuli would have lower classification scores, but the results were in fact the opposite. In fact, the classification scores were highest for vehicle sounds as a group. This finding may be explained by the unique acoustic features associated with the vehicle stimuli.

General Considerations and Future Directions

This study presented listeners with 18 different stimuli belonging to three different categories (i.e., household appliances, vehicles, and tools). Two previous studies used a larger set of stimuli, more than 50 sounds, that belonged to different categories than those used in this study (e.g., nonverbal human sounds, animals, machinery, weather-related sounds, and sounds created by human activities) (Gygi et al., 2007; Gygi et al., 2004). Considering the wide variety of stimuli that are potentially encountered in life, this study provided a small picture of classification performance and sound similarity data. It would be worthwhile to repeat this study using the same stimuli, as well as additional stimuli to get a broader picture of listeners' classification abilities related to environmental sounds.

There are a few variables that could have effected both classification performance and similarity ratings such as memory, attention, and previous knowledge (Letowski & Letowski, 2012; Lotto & Holt, 2011). Auditory memory, specifically top-down processing, and memory storage could have affected listener performance. Top-down processing could have helped listeners to identify ambiguous stimuli, which would have allowed them to respond quicker (Schultz, Dayan, & Montague, l997; Summerfield & Egner, 2009). Listeners would have had to store the perceived stimuli in memory to make judgments for identification and discrimination tasks. Attention was also an influential factor when considering listeners' performance. Increased attention has been shown to help listeners identify initially ambiguous stimuli (Shinn-Cunningham, 2008), but attention was not formally measured as a part of this study. Listener expectations and previous knowledge also influenced auditory perception. For example, a leading stimulus can affect the interpretation of all subsequent stimuli as listener expectations develop (Gygi et al, 2004; Snyder, Carter, Lee, Hannon, & Alain, 2008; Tsunada & Cohen, 2014).

Memory, attention, and previous knowledge could not be controlled for in this study; therefore, they could not be ruled out as contributory to the results. Implementing questionnaires to assess listeners' memory and attention capabilities may provide valuable information for future work. Considering the population used in this study (i.e., adults between 18 and 30 years old) it can be presumed that memory was not degraded, therefore, the results were not affected by this variable. However, the same cannot be assumed about attention. Attention may have influenced the results such that poorer attention lead to poorer classification performance and increased listener errors. Previous knowledge can be partially controlled for by using the survey results. If there were high reports of familiarity and frequency of use it would suggest higher levels of previous knowledge.

Familiarity and frequency of use were not controlled for, but they were evaluated for their contributions to the results of this study. Listeners reported their familiarity with each stimulus and their frequency of use in the survey given at the end of the experimental part of the study. They used a 7-point Likert rating scale to report if they were 1, “not familiar”, 7, “very familiar”, as well as 1, “never” and 7, “use daily”. These responses were used to determine what contributions they had on listeners’ performance on the classification task. The results of this study indicated there was a significant correlation between familiarity and classification performance for bathroom fan, box fan, hand vacuum, push mower, stand blender, upright vacuum, Diesel SUV, food processor, and Harley. These stimuli had ratings that corresponded with some level of familiarity (above 3.5). Additionally, there was a significant correlation between frequency of use and classification performance for bathroom fan, box fan, and upright vacuum. These stimuli were rated as frequently used (above 3.5). Overall, listeners were good at classifying some sounds they never used (e.g., helicopters, prop planes), but had a hard time classifying other sounds they used all the time (e.g., vacuums). This could be due to spectral and temporal differences versus commonalities between the sounds used in this study. Spectral and temporal analysis could add valuable information related to this speculation if the study were to be repeated in the future. A correlation of the aggregate survey data revealed there was a strong positive relationship between familiarity and frequency of use ($r = 0.808$, $p < 0.001$), which was as expected. Offering an open answer portion in the survey could add valuable information to better understand why listeners were familiar or frequently using items, which could then be used to better understand the classification performance.

Table 1

Classification Data from Experiment 1

Level 3	Percent Correct	Level 2	Percent Correct	Level 1	Percent Correct
Power Tool	87.0	Drill	47.9	Driver	44.9
				Impact Drill	23.7
		Saw	39.9	Circular Saw	26.9
				Sawzall	39.7
		Yard	94.7	Push Mower	87.1
				Weed Eater	64.9
Household Appliance	72.1	Blender	40.3	Stand Blender	24.9
				Food Processor	17.7
		Vacuum	50.1	Upright Vacuum	38.0
				Hand Vacuum	36.3
		Fan	68.7	Box Fan	41.7
				Bathroom Fan	74.9
Vehicle	95.6	Prop Plane	92.3	Mustang Piper	63.7
				Mosquito	27.4
		Helicopter	96.1	Bell 212	62.0
				Jet Ranger	64.6
		Wheeled	68.3	Harley	52.0
				Diesel SUV	62.0

Note. The results from Experiment 1 are shown next to each item for Levels 1, 2, and 3. The percentages represent correct classifications from all listeners (N=35). For example, listeners correctly identified tools 50.1% of the time, specifically drills 36.9% of the time, and a unique driver 45% of the time.

Table 2

Significant Pairwise Comparisons for Level 2 Stimuli

Mean Comparison	Mean Diff.	Lower	Upper	Sig.
Blender/Fan	-5.686	-9.31	-2.061	0.001
Blender/Helicopter	-11.171	-13.068	-9.275	0.001
Blender/Power Tool	-10.886	-12.696	-9.076	0.001
Blender/Prop Plane	-10.4	-12.379	-8.421	0.001
Blender/Wheeled	-5.6	-8.712	-2.488	0.001
Drill/Fan	-4.171	-7.45	-0.893	0.003
Drill/Helicopter	-9.657	-11.54	-7.775	0.001
Drill/Power Tool	-9.371	-11.406	-7.337	0.001
Drill/Prop Plane	-8.886	-10.718	-7.053	0.001
Drill/Wheeled	-4.086	-6.604	-1.567	0.001
Fan/Helicopter	-5.486	-8.562	-2.41	0.001
Fan/Power Tool	-5.2	-8.104	-2.296	0.001
Fan/Prop Plane	-4.714	-7.791	-1.638	0.001
Fan/Saw	5.771	2.501	9.042	0.001
Helicopter/Saw	11.257	9.609	12.905	0.001
Helicopter/Vacuum	9.2	6.569	11.831	0.001
Helicopter/Wheeled	5.571	3.061	8.082	0.001
Power Tool/Saw	10.971	9.158	12.785	0.001
Power Tool/Vacuum	8.914	6.44	11.388	0.001
Power Tool/Wheeled	5.286	2.687	7.885	0.001
Prop Plane/Saw	10.486	8.537	12.434	0.001
Prop Plane/Vacuum	8.429	5.984	10.873	0.001
Prop Plane/Wheeled	4.8	2.383	7.217	0.001
Saw/Wheeled	-5.686	-8.404	-2.967	0.001
Vacuum/Wheeled	-3.629	-7.113	-0.144	0.034

Note. Significant pairwise comparisons are shown from Experiment 1 Level 2 data. Mean difference (Mean diff.), lower and upper boundaries for 95% confidence interval, and significance level (Sig.) are shown for each stimulus pair.

Table 3

Survey Results for Individual Stimuli

Stimulus	Frequency	Familiarity
Bathroom Fan	6.2	6.14
Bell 212	1.26	2.31
Box Fan	4	5.34
Circular Saw	1.69	3.49
Diesel SUV	2.14	4.6
Driver	2.03	3.37
Food Processor	2.57	4.4
Hand Vacuum	3.23	5.54
Harley	1.66	4.69
Impact Drill	1.74	2.71
Jet Ranger	1.31	2.31
Mosquito	1.26	1.83
Mustang Piper	1.34	2.14
Push Mower	3.11	6.49
Sawzall	1.66	2.6
Stand Blender	3.71	5.37
Upright Vacuum	5.03	6.31
Weed Eater	2.63	5.91

Note. Listeners completed a survey following the experimental part of the study. A 7-point Likert rating scale was used to evaluate how often listeners used each stimulus with a score of 1 represented "never use" and 7 represented "daily". Listeners also used the scale to report how familiar they were with each stimulus where 1 represented "not familiar" and 7 represented "very familiar". Average ratings were used for each stimulus.

Table 4

Correlation of Frequency and Classification of Stimuli

Stimulus	*r*	*p*
Bathroom Fan	**0.445**	**0.007**
Bell 212	0.247	0.153
Box Fan	-0.013	0.942
Circular Saw	0.115	0.512
Diesel SUV	-0.271	0.116
Driver	0.265	0.124
Food Processor	0.281	0.102
Hand Vacuum	**0.399**	**0.018**
Harley	0.166	0.341
Impact Drill	0.23	0.184
Jet Ranger	0.133	0.446
Mosquito	-0.08	0.648
Mustang Piper	-0.088	0.614
Push Mower	-0.284	0.098
Sawzall	**0.446**	**0.007**
Stand Blender	-0.176	0.312
Upright Vacuum	0.032	0.853
Weed Eater	0.192	0.27

Note. Listeners reported their frequency of use for each stimulus using a Likert rating scale of 1 to 7, where 1 represented "never use" and 7 was "daily". Survey responses were analyzed with the classification data from Experiment 1. Values in bold font were significant.

Table 5

Correlation of Familiarity and Classification of Stimuli

Stimulus	*r*	*p*
Bathroom Fan	**0.645**	**0.001**
Bell 212	**0.358**	**0.035**
Box Fan	**0.516**	**0.002**
Circular Saw	**0.343**	**0.044**
Diesel SUV	0.017	0.923
Driver	0.256	0.138
Food Processor	0.23	0.185
Hand Vacuum	0.203	0.243
Harley	0.288	0.093
Impact Drill	0.072	0.683
Jet Ranger	0.119	0.497
Mosquito	-0.138	0.431
Mustang Piper	-0.028	0.873
Push Mower	-0.012	0.945
Sawzall	**0.488**	**0.003**
Stand Blender	-0.17	0.329
Upright Vacuum	0.183	0.294
Weed Eater	**0.492**	**0.003**

Note. Listeners reported their familiarity with each stimulus using a Likert rating scale of 1 to 7, where 1 represented "not familiar" and 7 was "very familiar". Survey responses were analyzed with the classification data from Experiment 1. Values in bold font were significant.

Table 6

Correlation of Familiarity and Frequency of Stimuli

Stimulus	*r*	*p*
Bathroom Fan	**0.703**	**0.001**
Bell 212	**0.568**	**0.001**
Box Fan	**0.462**	**0.005**
Circular Saw	**0.382**	**0.024**
Diesel SUV	**0.431**	**0.01**
Driver	**0.735**	**0.001**
Food Processor	**0.567**	**0.001**
Hand Vacuum	**0.624**	**0.001**
Harley	**0.363**	**0.032**
Impact Drill	**0.805**	**0.001**
Jet Ranger	**0.404**	**0.016**
Mosquito	**0.445**	**0.007**
Mustang Piper	**0.609**	**0.001**
Push Mower	**0.436**	**0.009**
Sawzall	**0.589**	**0.001**
Stand Blender	**0.509**	**0.002**
Upright Vacuum	**0.577**	**0.001**
Weed Eater	**0.525**	**0.001**

Note. Listeners reported their familiarity with each stimulus using a Likert rating scale of 1 to 7, where 1 was "not familiar" and 7 was "very familiar". Listeners also reported their frequency of use of each item using a Likert rating scale of 1 to 7, where 1 represented "never use" and 7 was "daily". Values in bold font were significant.

Table 7

Confusion Matrix for Incorrect Rating of 1 on Experiment 2

Paired Stimuli	Frequency	Both Orders	Level 2	Level 3	Similarity
bathroom fan/box fan	9	YES	YES	YES	2.23, 2.74
food processor/stand blender	9	YES	YES	YES	2.34, 2.51
food processor/driver	7	YES	NO	NO	2.91, 2.51
Jet ranger/bell 212	5	YES	YES	YES	2.86, 2.51
driver/stand blender	4	YES	NO	NO	2.77, 3.14
bathroom fan/driver	3	YES	NO	NO	5.00, 4.91
circular saw/food processor	3	YES	NO	NO	2.54, 2.91
hand vacuum/food processor	3	YES	NO	YES	2.86, 2.80
impact drill/sawzall	3	YES	NO	YES	3.23, 3.40
upright vacuum/circular saw	3	YES	NO	NO	2.69, 2.77
upright vacuum/hand vacuum	3	YES	YES	YES	2.60, 2.66
circular saw/driver	2	NO	NO	YES	3.46
circular saw/stand blender	2	NO	NO	NO	2.77
driver/box fan	2	YES	NO	NO	4.86, 4.46
driver/impact drill	2	YES	YES	YES	4.49, 3.51
hand vacuum/circular saw	2	YES	NO	NO	2.91, 2.51
mosquito/bathroom fan	2	NO	NO	NO	3.71
mustang piper/mosquito	2	YES	YES	YES	3.20, 3.34
bathroom fan/circular saw	1	NO	NO	NO	4.51
bathroom fan/diesel SUV	1	NO	NO	NO	3.97
bathroom fan/upright vacuum	1	NO	NO	YES	3.26
box fan/food processor	1	NO	NO	YES	3.4
circular saw/mustang piper	1	NO	NO	NO	5.8
circular saw/weed eater	1	NO	NO	YES	5
diesel SUV/bell 212	1	NO	NO	YES	4.97
driver/sawzall	1	NO	NO	YES	4.14
driver/upright vacuum	1	NO	NO	NO	3.34
food processor/impact drill	1	NO	NO	NO	4.89
hand vacuum/box fan	1	NO	NO	YES	3.6
hand vacuum/diesel SUV	1	NO	NO	NO	4.77
harley/box fan	1	NO	NO	NO	3.37
harley/diesel SUV	1	NO	YES	YES	3.83
jet ranger/diesel SUV	1	NO	NO	YES	4.4
jet ranger/mosquito	1	NO	NO	YES	3.97
mosquito/hand vacuum	1	NO	NO	NO	4.89

mustang piper/diesel SUV	1	NO	NO	YES	4.06
push mower/weed eater	1	NO	YES	YES	2.86
sawzall/food processor	1	NO	NO	NO	3.57
stand blender/jet ranger	1	NO	NO	NO	5.57
upright vacuum/box fan	1	NO	NO	YES	3.63
upright vacuum/stand blender	1	NO	NO	YES	2.69
weed eater/box fan	1	NO	NO	NO	3.66
weed eater/stand blender	1	NO	NO	NO	2.91
hand vacuum/bathroom fan	1	NO	NO	YES	3.89

Note. Listeners provided similarity ratings for paired stimuli during Experiment 2. Listeners used a 7-point Likert rating scale, where 1 represented "similar" and 7 represented “dissimilar”. This confusion matrix includes all paired stimuli given a rating of 1 that were not identical stimuli. For example, if presented with bathroom fan/bathroom fan, the correct rating would be 1 because the stimuli were identical. This table also includes the frequency of the error per stimuli pair, if the error was seen in both presentation orders (e.g. bathroom fan/box fan vs. box fan/bathroom fan), if the stimuli belonged to the same Level 2 and 3, and the average similarity rating for those stimuli.

Stimulus Selection Screen for Experiment 1

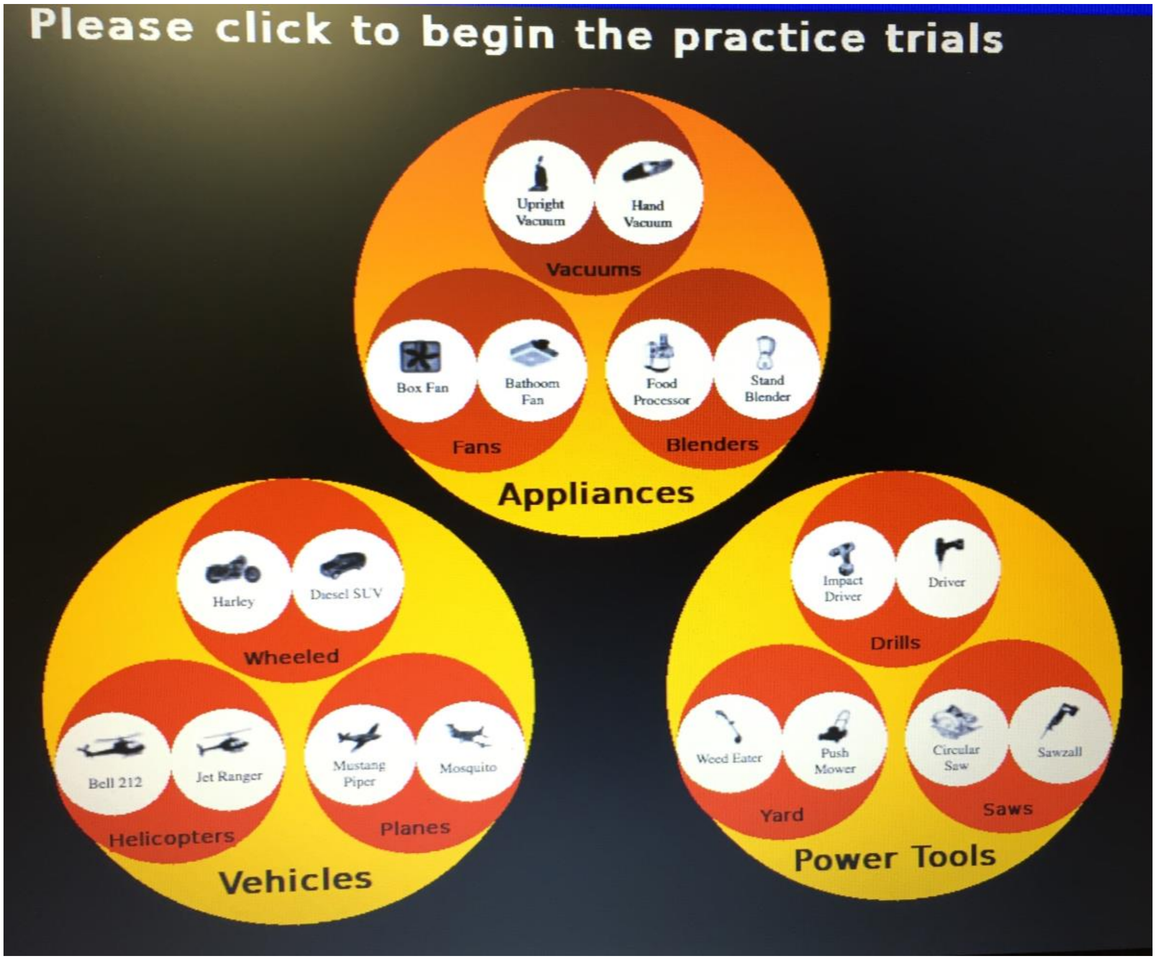

Figure 1. This is the screen that listeners used to provide their responses for all stimuli presentations in Experiment 1. Listeners were instructed to click a single icon that they felt represented the presented stimulus. After clicking an icon, the program automatically advanced to the next presentation.

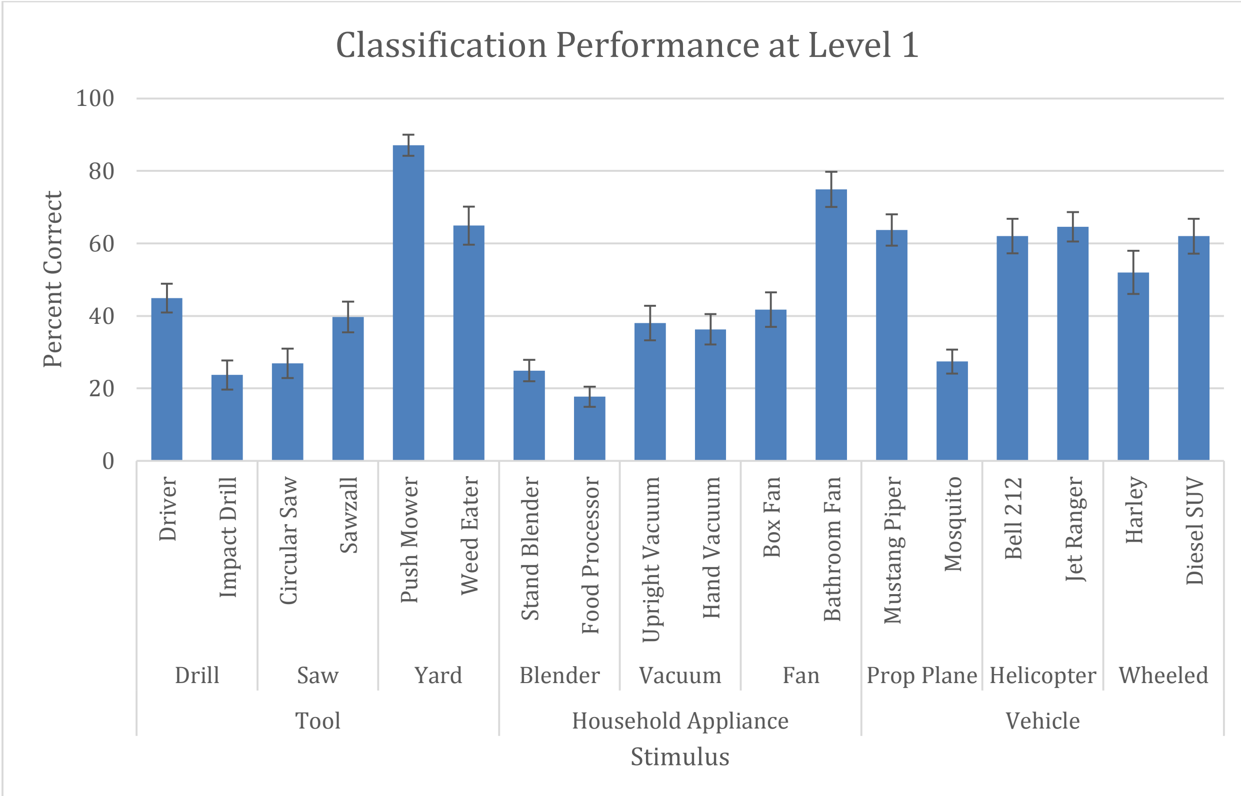

Figure 2. The results from Experiment 1 are shown for Level 1 stimuli. Each bar represents the average percent of correct identifications of each stimulus for all listeners. Error bars are included to show the standard error of the mean for each stimulus.

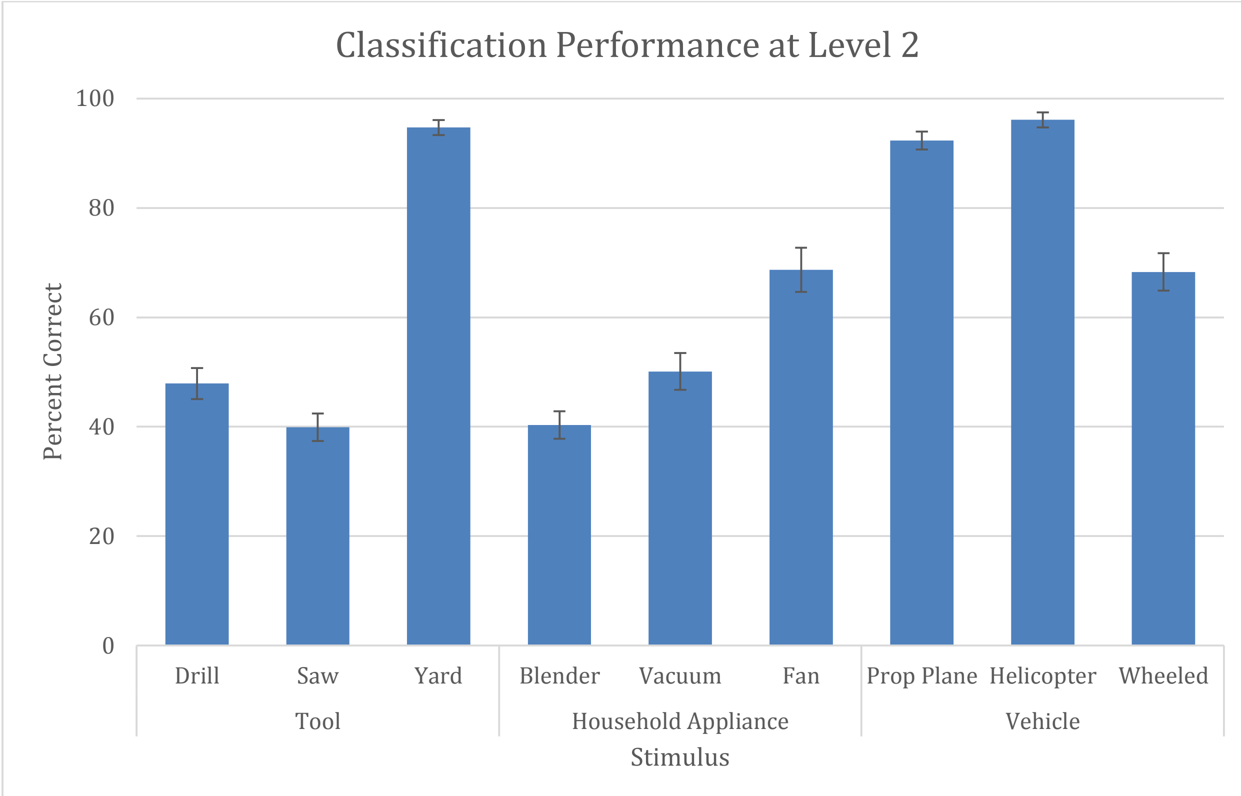

Figure 3. The results from Experiment 1 are shown for Level 2 stimuli categories. Each bar represents the average percent of correct identifications of each stimulus for all listeners. Error bars are included to show the standard error of the mean for each stimulus.

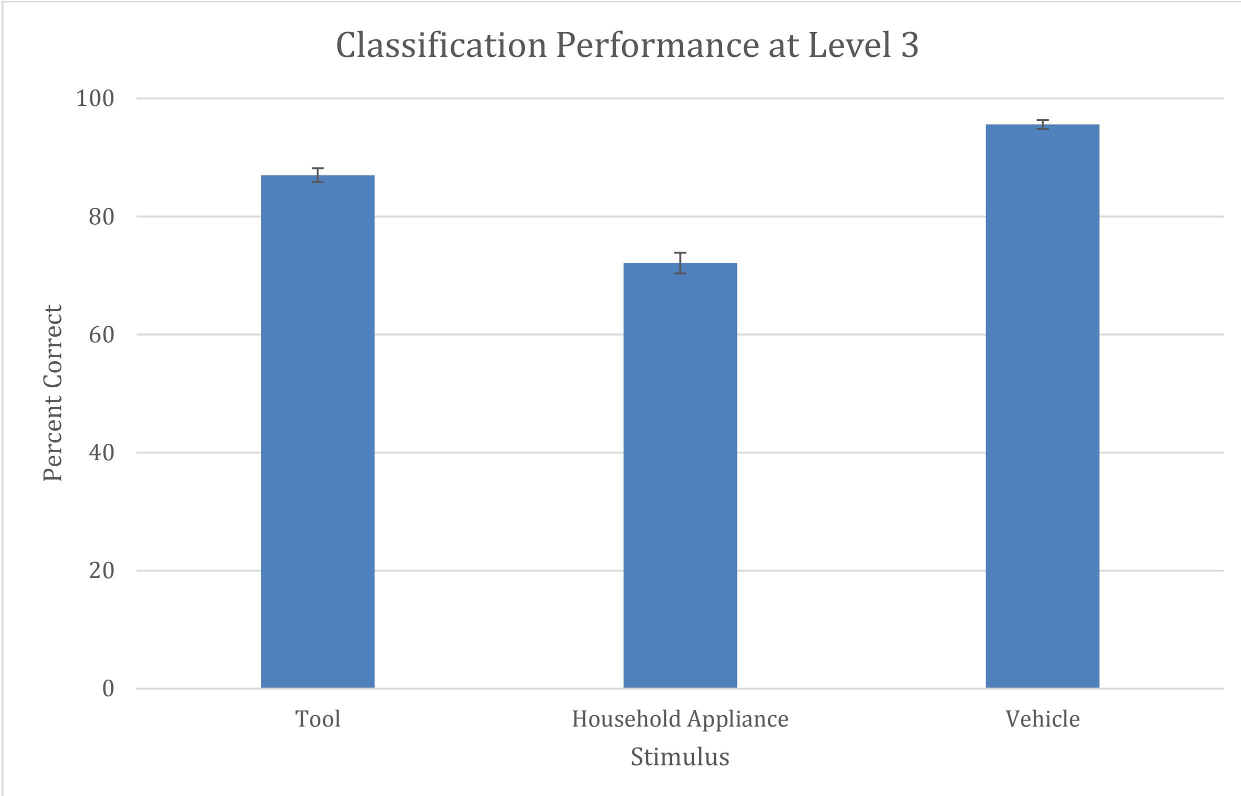

Figure 4. The results from Experiment 1 are shown for Level 3 stimuli categories. Each bar represents the average percent of correct identifications of each stimulus for all listeners. Error bars are included to show the standard error of the mean for each stimulus.

Euclidean Map Generated from MDS for Experiment 2

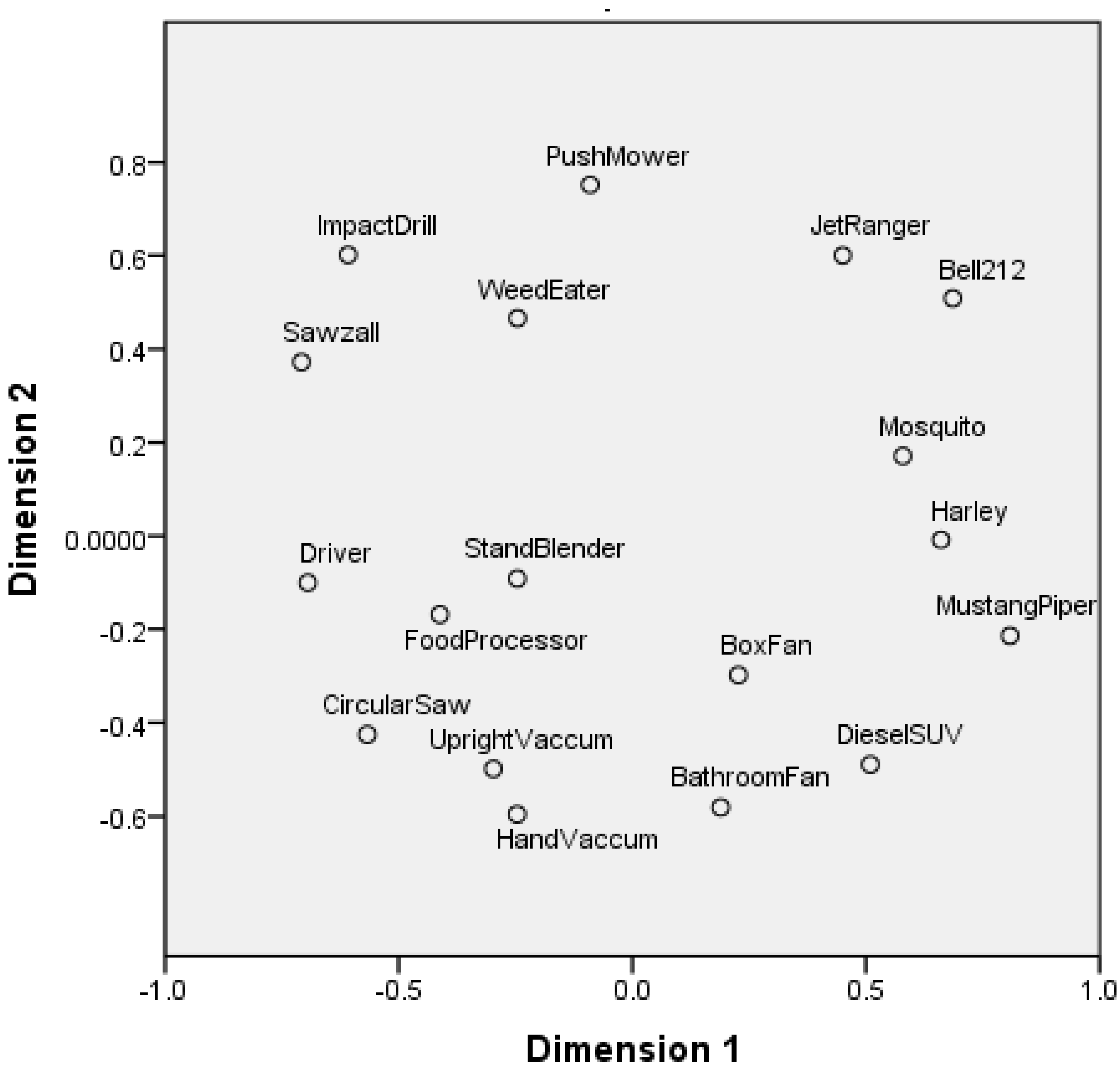

Figure 5. Multidimensional scaling analysis of the results from Experiment 2 were used to generate this 2-D Euclidean map. Stimuli that are close together (e.g., bathroom fan and box fan) were given rating scores that represented high degree of similarity compared to stimuli with greater distance between them were considered highly dissimilar (e.g., Jet Ranger and circular saw).

Instruction Screen for Experiment 1

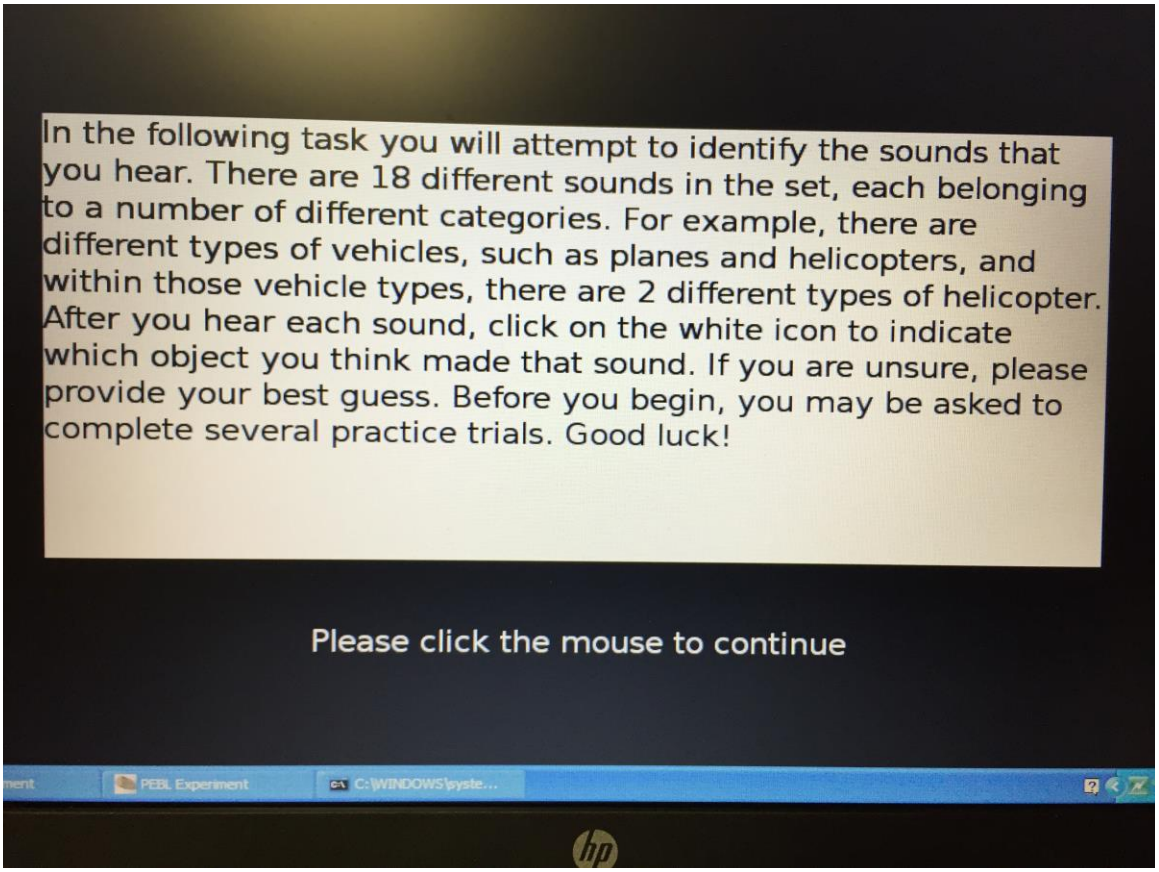

Figure 6. This is the instruction screen that listeners saw before beginning Experiment 1.

Instruction Screen for Experiment 2

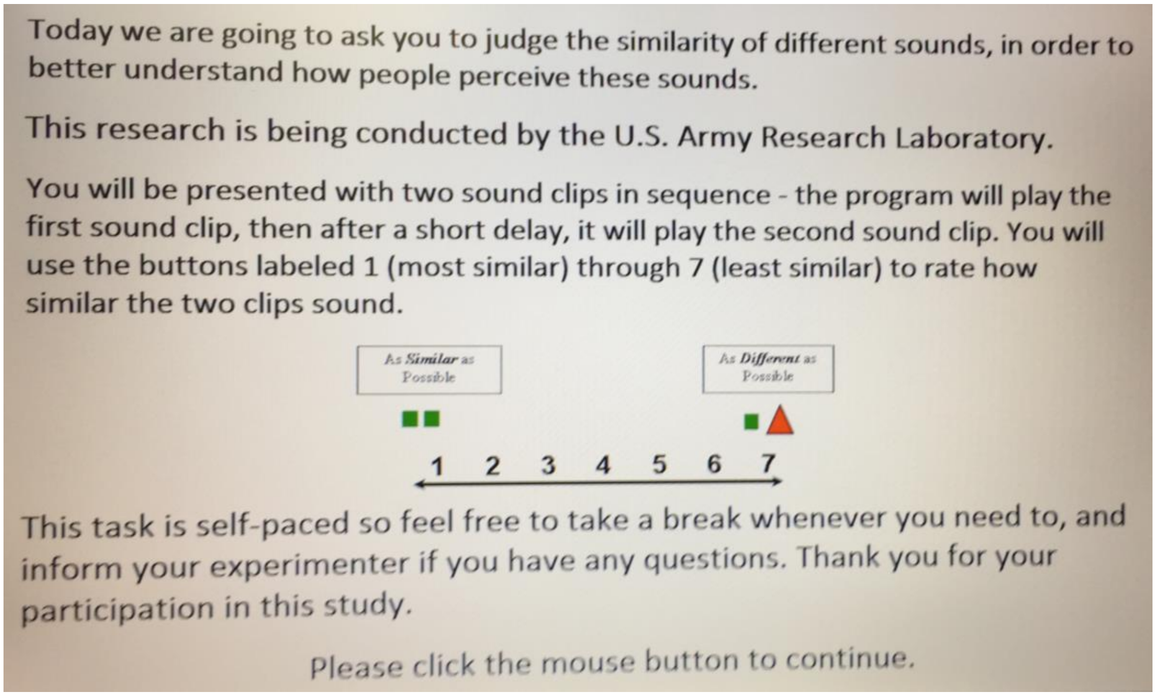

Figure 7. This is the instruction screen for Experiment 2, which included the Likert rating scale that listeners used to record their responses. A response of 1 represented items were completely similar or identical and a response of 7 represented items were highly dissimilar. After listeners chose a similarity rating, the program automatically advanced to the next set of paired stimuli.

REFERENCES

American National Standards Institute. (1994). Acoustical terminology. ANSI S1.1-1994 (R2004). New York: American National Standards Institute (ANSI).

Angrilli, A., Cherubini, P., Pavese, A., & Manfredini, S. (1997). The influence of affective factors on time perception. *Perception and Psychophysics, 59*(6), 972-982.

Ballachanda, B. B. (1997). Theoretical and applied external ear acoustics. *Journal of the American Academy of Audiology, 8*(6), 411-420.

Barsalou, L. (1991). Deriving categories to achieve goals. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 27, pp. 1-64). New York, NY: Academic Press.

Bey, C., & McAdams, S. (2002). Schema-based processing in auditory scene analysis. *Perception and Psychophysics, 64*(5), 844-854.

Bonebright., T. L. (2001). Perceptual structure of everyday sounds: A multidimensional scaling approach. *In Proceedings of the International Conference on Auditory Display* (pp.73-78). Espoo, Finland: ICAD.

Brand, T., & Hohmann, V. (2001). Effect of hearing loss, centre frequency, and bandwidth on the shape of loudness functions in categorical loudness scaling. *Audiology, 40*(2), 92-103.

Carello, C., Anderson, K. L., & Kunkler-Peck, A. J. (1998). Perception of object length by sound. *Psychological Science, 9*(3). 211-214.

Carlyon, R. P., Cusack, R., Foxton, J. M., & Robertson, I. H. (2001). Effects of attention and unilateral neglect on auditory stream segregation. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(1), 115-127.

Cermak, G. W., & Cornillon, P. C. (1976). Multidimensional analyses of judgments about traffic noise. *Journal of the Acoustical Society of America*, *59*(6), 1412-1420.

Chan, A. S., Ho, Y. C., & Cheung, M. C. (1998). Music training improves verbal memory. *Nature*, *396*(6707), 128.

Cox, R., Alexander, G., Taylor, I., & Gray, G. (1997). The contour test of loudness perception. *Ear and Hearing*, *18*(5), 388-400.

Dickerson, K., & Gaston, J. R. (2014). Did you hear that? The role of stimulus similarity and uncertainty in auditory change deafness. *Frontiers in Psychology*, *5*(1125), 1-5.

Dickerson, K., Gaston, J. R., Perelman, B. S., Mermagen, T., & Foots, A. N. (2015). Sound source similarity influences change perception in complex scenes. *Acoustical Society of America*, *23*, 1-14.

Doughty, J. M., & Garner, W. R. (1948). Pitch characteristics of short tones. II. Pitch as a function of tonal duration. *Journal of Experimental Psychology*, *38*(4), 478-494.

Droit-Volet, S., Brunot, S., & Niedenthal, P. (2004). Perception of the duration of emotional events. *Cognition and Emotion*, *18*(6), 849-858.

Eagleman, D. M. (2008). Human time perception and its illusions. *Current Opinion in Neurobiology*, *18*, 131-136.

Ekman, G., Berglund, B., & Berglund, U. (1966). Loudness as a function of the duration of auditory stimulation. *Scandinavian Journal of Psychology, 7*(1), 201.

Emanuel, D. C., & Letowski, T. (2009). *Hearing science*. Philadelphia, PA: Wolters Kluwer Health/Lippincott Williams and Wilkins.

Eramudugolla, R., Irvine, D. R., McAnally, K. I., Martin, R. L., & Mattingley, J. B. (2005). Directed attention eliminates 'change deafness' in complex auditory scenes. *Current Biology*, *15*(12), 1108-1113.

Fluitt, K., Gaston, J., Karna, V., & Letowski, L. (2010). Feasibility of audio training for identification of auditory signatures of small arms fire. Army Research Laboratory, ARL-TR-5413.

Ford, A. (1942). Dynamic auditory localization. *Journal of the Acoustical Society of America, 13*(4), 367-372.

Fried, L., & Holyoak, K. (1984). Induction of category distributions: A framework for classification learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(2), 234-257.

Gabrielsson, A. (1974). Performance of rhythm patterns. *Scandinavian Journal of Psychology*, *15*(1), 63-72.

Gardner, M. B. (1968). Lateral localization of 0 degrees-or near-0 degrees-oriented speech signals in anechoic space. *The Journal of the Acoustical Society of America, 44*(3), 797-802.

Gardner, M. B. (1973). Some monaural and binaural facets of median plane localization. *Journal of the Acoustical Society of America, 54*(6), 1489-1495.

Gardner, M. B., & Gardner, R. S. (1973). Problem of localization in the median plane: Effect of pinnae cavity occlusion. *Journal of the Acoustical Society of America, 53*(2), 400.

Gaston, J., & Letowski, T. (2012). Listener perception of single-shot small arms fire. *Noise Control Engineering Journal*, *60*(3), 236-245.

Goldberg, B., & Cannon-Bowers, J. (2015). Feedback source modality effects on training outcomes in a serious game: Pedagogical agents make a difference. *Computers in Human Behavior, 52,* 1-11.

Goldstone, R. (1994). The role of similarity in categorization: Providing a groundwork. *Cognition*, *52*(2), 125-157.

Goldstone, S., Lhamon, W. T., & Sechzer, J. (1978). Light intensity and judged duration. *Bulletin of the Psychonomic Society, 12*, 83-84.

Gregg, M. K., & Samuel, A. G. (2009). The importance of semantics in auditory representations. *Attention, Perception, & Psychophysics*, *71*(3), 607-619.

Guilford, J. P. (1954). *Psychometric methods*. New York, NY: McGraw-Hill.

Gygi, B., Kidd, G. R., & Watson, C. S. (2007). Similarity and categorization of environmental sounds. *Perception and Psychophysics, 69*(6), 839-855.

Gygi, B., Kidd, G., & Watson, C. (2004). Spectral-temporal factors in the identification of environmental sounds. *Journal of the Acoustical Society of America, 115*(3), 1252-1265.

Hazan, V., & Barrett, S. (2000). The development of phonemic categorization in children aged 6–12. *Journal of Phonetics*, *28*(4), 377-396.

Holt, L. L., & Lotto, A. J. (2006). Cue weighting in auditory categorization: Implications for first and second language acquisition. *Journal of the Acoustical Society of America, 119*(5), 3059-3071.

Houtsma, A. M. (1995). Pitch perception. In B. J. Moore, B. J. Moore (Eds.), *Hearing* (2nd ed., pp. 267-295). San Diego, CA, US: Academic Press.

Houtsma, A., & Rossing, T. (1987). Effects of signal envelope on the pitch of short complex tones. *Journal of the Acoustical Society of America, 81*(2), 439-444.

Hyde, K. L., Lerch, J., Norton, A., Forgeard, M., Winner, E., Evans, A. C., & Schlaug, G. (2009). Musical training shapes structural brain development. *Journal of Neuroscience, 29*(10), 3019–3025.

Ihlefeld, A., & Shinn-Cunningham, B. (2011). Effect of source spectrum on sound localization in an everyday reverberant room. *Journal of the Acoustical Society of America, 130*(1), 324-333.

International Organization for Standardization. (2006). ISO 16832, Acoustics - loudness scaling by means of categories. Geneva: International Organization for Standardization (ISO).

Jentschke, S., & Koelsch, S. (2009). Musical training modulates the development of syntax processing in children. *Neuroimage, 47*(2), 735-744.

Kaan, E., & Swaab, T. Y. (2002). The brain circuitry of syntactic comprehension. *Trends in Cognitive Science, 6*(8), 350–356.

Kock, W. E. (1950). Binaural localization and masking. *Journal of the Acoustical Society of America, 22*(6), 801.

Koelsch, S., Kilches, S., Steinbeis, N., & Schelinski, S. (2008). Effects of unexpected chords and of performer's expression on brain responses and electrodermal activity. *PLOS One, 3*(7), e2631.

Kunkler-Peck, A. J., & Turvey, M. (2000). Hearing shape. *Journal of Experimental Psychology: Human Perception and Performance*, 26(1), 279.

Letowski, T. & Amrein, K. (2005). The effects of auditory training on the listener's ability to detect and recognize sounds in noise. 12th International Congress on Sound and Vibration.

Letowski, T. R. & Letowski, S. T. (2012). Auditory spatial perception: Auditory localization. Army Research Laboratories, ARL-TR-6016.

Li, X., Logan, R., & Pastore, R. (1991). Perception of acoustic source characteristics: Walking sounds. *Journal of the Acoustical Society of America, 90*(6), 3036-3049.

Lichte, W. H. (1941). Attributes of complex tones. *Journal of Experimental Psychology, 28*(6), 455-480.

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *The Journal of the Acoustical Society of America, 94*(3), 1242-1255.

Lotto, A., & Holt, L. (2011). Psychology of auditory perception. *Wiley Interdisciplinary Reviews: Cognitive Science, 2*(5), 479-489.

Lutfi, R. A., & Liu, C. (2007). Individual differences in source identification from synthesized impact sounds. *The Journal of the Acoustical Society of America, 122*(2), 1017-1028.

Lutfi, R. A., & Oh, E. L. (1997). Auditory discrimination of material changes in a struck-clamped bar. *The Journal of the Acoustical Society of America, 102*(6), 3647-3656.

Magne, C., Schön, D., & Besson, M. (2006). Musician children detect pitch violations in both music and language better than non-musician children: Behavioral and electrophysiological approaches. *Journal of Cognitive Neuroscience, 18*(2), 199–211.

Marcell, M., Borella, D., Greene, M., Kerr, E., & Rogers, S. (2000). Confrontation naming of environmental sounds. *Journal of Clinical and Experimental Neuropsychology, 22*(6), 830-864.

Marques, C., Moreno, S., Castro, S. L., & Besson, M. (2007). Musicians detect pitch violation in a foreign language better than nonmusicians: Behavioral and electrophysiological evidence. *Journal of Cognitive Neuroscience, 19*(9), 1453–1463.

McNamara, T. P. (2005). Semantic priming: Perspectives from memory and word recognition. Sussex, England: Psychology Press.

Mitroff, S. R., Simons, D. J., & Levin, D. T. (2004). Nothing compares 2 views: Change blindness can occur despite preserved access to the changed information. *Perception and Psychophysics, 66*(8), 1268-1281.

Moore, B. C., & Glasberg, B. R. (1990) Frequency discrimination of complex tones with overlapping and non-overlapping harmonics. *Journal of the Acoustical Society of America, 87*(5), 2163–2177.

Morgan, C. T., Garner, W. R., & Galambos, R. (1951). Pitch and intensity. *Journal of the Acoustical Society of America, 23*(6), 658-663.

Munson, W. A. (1947). The growth of auditory sensation. *Journal of the Acoustical Society of America, 19*(4), 584.

Murphy, G., & Medin, D. (1985). The role of theories in conceptual coherence. *Psychological Review, 92*(3), 289-316.

Neuhoff, J. (2001). An adaptive bias in the perception of looming auditory motion. Ecological Psychology, 13(2), 87-110.

Nittrouer, S. (2004). The role of temporal and dynamic signal components in the perception of syllable-final stop voicing by children and adults. *The Journal of the Acoustical Society of America, 115*(4), 1777-1790.

Oxenham, A. J. (2012). Pitch perception. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience, 32*(39), 13335–13338.

Pastore, R., Flint, J., Gaston, J., & Solomon, M. (2008). Auditory event perception: The source-perception loop for posture in human gait. *Perception And Psychophysics, 70*(1), 13-29.

Pearce, M. T., & Wiggins, G. A. (2012). Auditory expectation: The information dynamics of music perception and cognition. *Topics in Cognitive Science, 4*(4), 625.

Phillips, D. P., Quinlan, C. K., & Dingle, R. N. (2012). Stability of central binaural sound localization mechanisms in mammals, and the Heffner hypothesis. *Neuroscience and Biobehavioral Reviews, 36*(2) 889-900.

Platt, J., & Racine, R. (1985). Effect of frequency, timbre, experience, and feedback on musical tuning skills. *Perception and Psychophysics, 38*(6), 543-53.

Pollack, I. (1954). Intensity discrimination thresholds under several psychophysical procedures. *Journal of the Acoustical Society of America, 26*(6), 1056-1059.

Rayleigh, (1907). On our perception of sound direction. *Philosophical Magazine, 13,* 214-232.

Reynolds, G. S., & Stevens, S. S. (1960). Binaural summation of loudness. *Journal of the Acoustical Society of America, 32(*10), 1337.

Roffler, S. K., & Butler, R. A. (1978). Factors that influence the localization of sound in vertical plane. *Journal of the Acoustical Society of America, 43*(6), 1255-1259.

Rossing, T., & Houtsma, A. (1986). Effects of signal envelope on the pitch of short sinusoidal tones. *Journal of the Acoustical Society of America, 79*(6), 1926-1933.

Scharine, A., Cave, K., & Letowski, T. (2009). Auditory perception and cognitive performance. In C. E. Rash, M. B. Russo, T. Letowski, and E. T. Schmeisser (Eds.), *Helmet-Mounted Displays: Sensation, Perception, and Cognition Issues*, U.S. Army Aeromedical Research Laboratory, Ft. Rucker, AL.

Scherer, K., & Oshinsky, J. (1977). Cue utilization in emotion attribution from auditory stimuli. *Motivation and Emotion, 1*(4), 331-346.

Schlauch, R.S., DiGiovanni, J.J., & Ries, D.T. (1998). Basilar membrane nonlinearity and loudness. *Journal of the Acoustical Society of America, 103*(4), 2010-2020.

Schlaug, G. (2001). The brain of musicians. A model for functional and structural adaptation. *Annals New York Academy of Sciences, 930*, 281-299.

Schouten, B., Gerrits, E., & van Hessen, A. (2003). The end of categorical perception as we know it. *Speech Communication, 41*(1), 71-80.

Schultz, W., Dayan, P., & Montague, R. (1997). A neural substrate of prediction and reward. *Science, 275*(5306), 1593–1599.

Shinn-Cunningham, B. G. (2008). Opinion: Object-based auditory and visual attention. *Trends in Cognitive Sciences*, *12(5),* 182-186.

Shinn-Cunningham, B. G., Kopco, N., & Martin, T. J. (2005). Localizing nearby sound sources in a classroom: binaural room impulse responses. *The Journal of the Acoustical Society of America*, *117*(5), 3100-3115.

Sluming, V., Barrick, T., Howard, M., Cezayirli, E., Mayes, A., & Roberts, N. (2002). Voxel-based morphometry reveals increased gray matter density in Broca's area in male symphony orchestra musicians. *Neuroimage: A Journal of Brain Function, 17*(3), 1613-1622.

Smits, R., Sereno, J., & Jongman, A. (2006). Categorization of sounds. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(3), 733-754.

Snyder, J. S., Carter, O. L., Lee, S., Hannon, E. E., & Alain, C. (2008). Effects of context on auditory stream segregation. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(4), 1007-1016.

Steinbeis, N., Koelsch, S., & Sloboda, J. A. (2006). The role of harmonic expectancy violations in musical emotions: Evidence from subjective, physiological and neural responses. *Journal of Cognitive Neuroscience, 18*(8), 1380–1393.

Stevens, S. S. (1935). The Relation of Pitch to Intensity. *Journal of the Acoustical Society of America*, *6*(3), 150-154.

Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences*, *13*(9), 403-409.

Tillmann, B. (2012). Music and language perception: Expectations, structural integration, and cognitive sequencing. *Topics in Cognitive Science*, *4*(4), 568-584.

Tsunada, J., & Cohen, Y. E. (2014). Neural mechanisms of auditory categorization: From across brain areas to within local microcircuits. *Frontiers in Neuroscience, 8*, 161.

Turnbull, W. W. (1944). Pitch discrimination as a function of tonal duration. *Journal of Experimental Psychology, 34*(4), 302-316.

Vanderveer, N. J. (1980). Ecological acoustics: Human perception of environmental sounds. *Dissertation Abstracts International, 40,* 4543.

Warm, J. S., Greenberg, L. F., & Dube, C. S. (1964). Stimulus and motivational determinants in temporal perception. *Journal of Psychology*, *58*(2), 243-248.

Warm, J. S., & McCray, R. E. (1969). Influence of word frequency and length on the apparent duration of tachistoscopic presentations. *Journal of Experimental Psychology*, *79*(1), 56-58.

Warren, W., & Verbrugge, R. (1984). Auditory perception of breaking and bouncing events: A case study in ecological acoustics. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(5), 704-712.

Watson, C. S., & Gengel, R. W. (1969). Signal duration and signal frequency in relation to auditory sensitivity. *Journal of the Acoustical Society of America, 46*(4B), 989-997.

West, R. F., & Stanovich, K. E. (1982). Source of inhibition in experiments on the effect of sentence context on word recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition, 8*, 385–399.

White, W. B. (1941). Contribution offered towards a discussion of the paper 'pitch and the missing fundamental.'. *Journal of the Acoustical Society of America, 13*(1), 87-88.

Winkler, I., Horváth, J., Tervaniemi, M., Näätänen, R., Sussman, E., & Ritter, W. (2003). Preattentive auditory context effects. *Cognitive, Affective and Behavioral Neuroscience, 3*(1), 57-77.

Woodrow, H. (1951). Time perception. In: Stevens, S.S. (Ed.) *Handbook of Experimental Psychology* (pp. 1224-1236). New York, NY: Wiley and Sons.

Zahorik, P., & Wightman, F. L. (2001). Loudness constancy with varying sound source distance. *Nature Neuroscience, 4*(1), 78-83.

Zwicker, E., & Fastl, H. (1999). Psycho-acoustics: Facts and models. (3rd ed.). Berlin, China: Springer.

Zwicker, E., Flottorp, G., & Stevens, S. S. (1957). Critical band width in loudness summation. *Journal of the Acoustical Society of America, 29*(5), 548-557.

APPENDICES

Appendix A

Dear Ms. Rosen,

The IRB has approved your protocol "Environmental Sound Similarity and Identification Performance " effective 9/7/2016

Your IRB protocol can now be viewed by your faculty advisor in MyOSPR. For more information, please visit: http://www.towson.edu/academics/research/sponsored/myospr.html

If you should encounter any new risks, reactions, or injuries to subjects while conducting your research, please notify IRB@towson.edu. Should your research extend beyond one year in duration, or should there be substantive changes in your research protocol, you will need to submit another application.

We are offering training and orientation sessions for faculty in the fall, I encourage you to sign up for one of the sessions: http://fusion.towson.edu/www/signupGeneric/index.cfm?type=OSPR

Regards,
Towson IRB

Approval Number: 1609004502

Appendix B

VOLUNTEERS NEEDED FOR RESEARCH STUDY

Environmental Sound Similarity and Identification Performance

Adults 18-30 years old

We are conducting research to find out how listeners identify environmental sounds.

- Free **hearing screening!**
- Rate pairs of environmental sounds and identify the category with a computer-based interface
- Approximately **1-hour** session at **Van Bokkelen** on Towson University's campus

Appendix C

INFORMED CONSENT FORM

PRINCIPAL INVESTIGATOR: Anna Rosen PHONE: (610) 675-5610

Purpose of the Study:

The purpose of this study is to look at how listeners identify environmental sounds. While there is existing research that has analyzed listener sound identification, sound identification is still complex and needs to be further researched.

Procedures:

Participants will be given a quick hearing screening before beginning the two-part study. The first part of the study will ask participants to rate pairs of sounds using a 7-point similarity rating scale. The second part will have participants identify the category, as well as the actual item associated with each sound. All parts of the study will take place in Van Bokkelen at Towson University and will take no more than 1 hour. The study requires participants to sit in a sound-attenuating booth with insert earphones and use a computer-based interface to provide their answers.

Risks/Discomfort:

There are no known risks associated with participation in the study. Should the interview become distressing to you, it will be terminated immediately.

Benefits:

There are no perceived benefits associated with participation in the study.

Alternatives to Participation:

Participation in this study is voluntary. You are free to withdraw or discontinue participation at any time. Refusal to participate in this study will not penalize participants in anyway.

Cost Compensation:

Participation in this study will involve no costs or payments to you.

Confidentiality:

All information collected during the study period will be kept strictly confidential. You will be identified through identification numbers. No publications or reports from this project will include identifying information on any participant. If you agree to join this study, please sign your name below.

____ I have read and understood the information on this form.

____ I have had the information on this form explained to me.

______________________________________ ______________
Subject's Signature Date

______________________________________ ______________
Witness to Consent Procedures Date

______________________________________ ______________
Principal Investigator Date

This research has been reviewed by the Towson University Institutional Review Board. If you have any questions regarding this study please contact Dr. Emanuel at (410) 704-2417 or the Institutional Review Board Chairperson, Dr. Debi Gartland, Office of University Research Services, 8000 York Road, Towson University, Towson, Maryland 21252; phone (410) 704-2236.

Appendix D

Participant ID: ________________ Date: ________________

Environmental Sound Similarity and Identification Performance

Thesis Survey

Please rate your familiarity with the following items using a Likert rating scale of 1 to 7, with 1 being "not familiar" and 7 being "very familiar".

1. Impact Drill	1---2---3---4---5---6---7
2. Driver	1---2---3---4---5---6---7
3. Push Mower	1---2---3---4---5---6---7
4. Weed Eater	1---2---3---4---5---6---7
5. Diesel SUV	1---2---3---4---5---6---7
6. Harley Motorcycle	1---2---3---4---5---6---7
7. Mosquito Prop Plane	1---2---3---4---5---6---7
8. Bell 212 Helicopter	1---2---3---4---5---6---7
9. Jet Ranger Helicopter	1---2---3---4---5---6---7
10. Mustang Piper Prop Plane	1---2---3---4---5---6---7
11. Box Fan	1---2---3---4---5---6---7
12. Bathroom Fan	1---2---3---4---5---6---7
13. Upright Vacuum	1---2---3---4---5---6---7
14. Hand Vacuum	1---2---3---4---5---6---7
15. Stand Blender	1---2---3---4---5---6---7
16. Food Processor	1---2---3---4---5---6---7
17. Circular Saw	1---2---3---4---5---6---7
18. Sawzall	1---2---3---4---5---6---7

Please rate your frequency of use with the following items using a Likert rating scale of 1 to 7, with 1 being "never use" and 7 being "daily".

1. Impact Drill	1---2---3---4---5---6---7
2. Driver	1---2---3---4---5---6---7
3. Push Mower	1---2---3---4---5---6---7
4. Weed Eater	1---2---3---4---5---6---7
5. Diesel SUV	1---2---3---4---5---6---7
6. Harley Motorcycle	1---2---3---4---5---6---7
7. Mosquito Prop Plane	1---2---3---4---5---6---7
8. Bell 212 Helicopter	1---2---3---4---5---6---7
9. Jet Ranger Helicopter	1---2---3---4---5---6---7
10. Mustang Piper Prop Plane	1---2---3---4---5---6---7
11. Box Fan	1---2---3---4---5---6---7
12. Bathroom Fan	1---2---3---4---5---6---7
13. Upright Vacuum	1---2---3---4---5---6---7
14. Hand Vacuum	1---2---3---4---5---6---7
15. Stand Blender	1---2---3---4---5---6---7
16. Food Processor	1---2---3---4---5---6---7
17. Circular Saw	1---2---3---4---5---6---7
18. Sawzall	1---2---3---4---5---6---7

CURRICULUM VITA

NAME: Annamarie Rosen

[redacted]

PROGRAM OF STUDY: Audiology

DEGREE AND DATE TO BE CONFERRED: Doctorate of Audiology, 2018

Secondary education: Bloomsburg University, Bloomsburg, Pennsylvania, 2014

Collegiate institutions attended	Dates	Degree	Date of Degree
Towson University	August 2014-Present	Au.D	May 2018
Bloomsburg University	August 2010-May 2014	B.S.	May 2014

Major: Speech Language Pathology & Audiology

Professional publications: N/A

Professional positions held: N/A